OPERATIONS AND MAINTENANCE OPTIMIZATION OF STOCHASTIC SYSTEMS: THREE ESSAYS

by

David T. Abdul-Malak

B.S., Arcadia University, 2012

Submitted to the Graduate Faculty of

the Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

David T. Abdul-Malak

It was defended on

November 14, 2018

and approved by

Jeffrey P. Kharoufeh, Ph.D., Professor, Department of Industrial Engineering

Lisa M. Maillart, Ph.D., Professor, Department of Industrial Engineering

Daniel R. Jiang, Ph.D., Assistant Professor, Department of Industrial Engineering

Zhi-Hong Mao, Ph.D., Professor, Department of Electrical and Computer Engineering

Dissertation Director: Jeffrey P. Kharoufeh, Ph.D., Professor, Department of Industrial Engineering

OPERATIONS AND MAINTENANCE OPTIMIZATION OF STOCHASTIC SYSTEMS: THREE ESSAYS

David T. Abdul-Malak, PhD

University of Pittsburgh, 2018

This dissertation presents three essays on topics related to optimally operating and maintaining systems that evolve randomly over time. Two primary areas are considered: (i) joint staffing and pricing strategies for call centers that use co-sourcing to improve service operations and reduce costs; and (ii) optimally maintaining stochastically degrading systems when either multiple systems are associated via a common environment, or when a single-unit system is maintained using a population of heterogeneous spare parts.

First we present a queueing and stochastic programming framework for optimally staffing a call center utilizing co-sourced service capacity. The interplay between the call center and external service provider is modeled as a leader-follower game in which the call center, acting as the follower, solves a two-stage stochastic integer program. The problem is reformulated as a quadratically-constrained linear program to obtain the optimal contract prices and the optimal staffing problem yields a closed-form solution. Numerically we demonstrate that significant cost reductions can be achieved, even in the presence of imperfect and asymmetric information. Second the problem of optimally replacing multiple stochastically degrading systems using condition-based maintenance is considered. Properties of the optimal value function and policy motivate a tractable, approximate model with state- and action-space transformations and a basis-function approximation of the action-value function. It is demonstrated that near optimal policies are attainable and significantly outperform heuristics. Finally, we consider the problem of optimally maintaining a stochastically degrading system using spares of varying quality. Conditions are provided under which the optimal value function exhibits monotonicity and the optimal policy is characterized. Numerically we demonstrate the utility of our proposed framework, and provide insights into the optimal policy as an exploration-exploitation type policy.

TABLE OF CONTENTS

PREFACE . . . viii

1.0 INTRODUCTION . . . 1

- 1.1 DISSERTATION OUTLINE AND CONTRIBUTIONS . . . 1

2.0 STAFFING AND PRICING IN CO-SOURCED CALL CENTERS . . . 4

- 2.1 INTRODUCTION . . . 4
- 2.2 MODEL FORMULATION . . . 8
  - 2.2.1 Call Center Staffing Problem . . . 8
  - 2.2.2 Contractor's Pricing Problem . . . 13
  - 2.2.3 Profit-Based Objective . . . 14
- 2.3 REFORMULATING THE PRICING PROBLEM . . . 15
- 2.4 NUMERICAL STUDY . . . 20
  - 2.4.1 Staffing Level Problem . . . 22
  - 2.4.2 Joint Staffing-Pricing Problem . . . 25
- 2.5 CONCLUSIONS . . . 28

3.0 MULTI-SYSTEM REPLACEMENT IN A SHARED ENVIRONMENT . . . 31

- 3.1 INTRODUCTION . . . 31
- 3.2 DEGRADATION MODEL AND PROBLEM FORMULATION . . . 34
- 3.3 STRUCTURAL RESULTS . . . 38
  - 3.3.1 Results for the General Case . . . 38
  - 3.3.2 Special Case: A Single System . . . 46
- 3.4 AN APPROXIMATE FORMULATION . . . 47
- 3.5 NUMERICAL EXAMPLES . . . 50
  - 3.5.1 Single-System Problems . . . 51
  - 3.5.2 Multiple-System Problems . . . 56

4.0 MAINTAINING SYSTEMS WITH HETEROGENEOUS SPARE PARTS . . 64
4.1 INTRODUCTION 64
4.2 MODEL FORMULATION 67
4.3 STRUCTURAL RESULTS 71
4.4 NUMERICAL EXAMPLES 86
4.4.1 Randomly-Generated Problem Instances 86
4.4.2 A Specific Two-quality Problem 89
4.4.3 A Specific Three-quality Problem 91
4.5 CONCLUSIONS AND FUTURE WORK 94
BIBLIOGRAPHY 96

LIST OF TABLES

1 Comparison of exact and learned thresholds. 53

2 Summary of policy comparison results. 89

LIST OF FIGURES

1 Square-root scale arrival rate forecasts. 21
2 Comparison of one-week staffing costs for varied c_0 and c_a. 23
3 Comparison of one-week staffing costs for varied c_0 and c_a (scaled parameters). 24
4 Probability of violating QoS constraints for unscaled (a) and scaled (b) call centers. . 25
5 Comparison of one-week staffing cost for varied ψ_1 on the unscaled call center (ψ_1). . 26
6 Comparison of one-week staffing cost for varied ψ_1 on the scaled call center (ψ_1). . . 26
7 Leader's objective function values (lighter shading indicates favorable performance). . 28
8 Effect of sample size on the optimal holding cost. 29
9 Effect of sample size on the optimal activation cost. 29
10 Boxplots of thresholds from learned policies (optimal thresholds given by circles). . . 52
11 Boxplots comparing cost using the optimal policy versus the learned policies. 54
12 Boxplots of percent and absolute differences. 55
13 Comparison between optimal and learned policy in environment state 1. 57
14 Boxplots comparing cost under different policies. 58
15 Comparison between optimal and learned value function in environment state 2. . . . 60
16 Comparison between optimal and learned policy in environment state 1. 61
17 Comparison between optimal and learned policy in environment state 2. 61
18 Boxplots comparing cost under different policies. 62
19 Boxplots comparing cost under different policies. 63
20 Boxplots comparing policy costs for problem instances $m = 10$ and $m = 12$. 90
21 Depiction of the optimal value function (dark colors indicate lower costs). 91
22 Comparison between the *MOMDP* and *Heuristic* policies. 92
23 Depiction of the optimal value function (dark colors indicate lower costs). 93
24 Depiction of the *MOMDP* policy (dark colors indicate lower thresholds). 94

PREFACE

Nobody ever achieves anything meaningful without the support of others. As such, I owe a debt of gratitude to those that have helped bring this dissertation to fruition. First, and foremost, I will forever be thankful for the tireless efforts of my advisor, Professor Jeff Kharoufeh. Without Dr. Kharoufeh's steadfast support, constant guidance, and fastidious care, this dissertation could simply never have been. Additionally, I would like to thank the members of my committee, Drs. Lisa M. Maillart, Daniel R. Jiang, and Zhi-Hong Mao, whose careful reading and insightful comments have significantly improved this dissertation.

I am also grateful to many people who have contributed to my love of curiosity and learning. To Mrs. Sherry Roses, Mr. J.P. Huebner, and Mr. Todd Curtis, whose passion for education drove me to a lifetime love of science, technology, engineering, and mathematics. To the Computer Science and Mathematics faculty at Arcadia University; particularly, Drs. Xizhong Zheng and Ned Wolff for sparking my love of research and encouraging me to pursue my Ph.D. To Arnab Bhattacharya, Juan Borrero, and Ruichen Sun, who have been my best friends at Pitt.

Finally, I am grateful for the love and support of my family, which has shaped me into the person I am today. To my siblings, Dan and Jen, for being a constant source of inspiration, motivation, joy, and pride. My parents, Denise and Tarek, whose many sacrifices have allowed me to achieve my dreams. And finally, Niquelle, my wife, for her continuous example of perfect love.

1.0 INTRODUCTION

This dissertation presents three essays on topics related to optimally operating and maintaining systems that evolve randomly over time. Two primary areas are considered: (i) joint staffing and pricing strategies for call centers that use co-sourcing to improve service operations and reduce costs; and (ii) optimally maintaining stochastically degrading systems when either multiple systems are associated via a common environment, or when a single-unit system is maintained using a population of heterogeneous spare parts.

The models developed and analyzed herein can all be viewed as important extensions of classical models considering enriched state and action spaces. Chapter 2 considers the classical problem of staffing a call center over multiple periods, but it involves an expanded state space that allows for demand rate uncertainty, and the action space includes decisions for both the call center and an external service provider. Chapter 3 considers optimal replacement under environment-driven degradation, but the state and action spaces considered are higher-dimensional spaces allowing for the consideration of multiple systems whose stochastic and economic dependencies cannot be ignored. Chapter 4 considers optimal maintenance of a single-unit system, but the state space allows for maintaining a belief about the quality of a system, and the action space considers both repair and replacement.

1.1 DISSERTATION OUTLINE AND CONTRIBUTIONS

Chapter 2 presents a joint queueing and stochastic programming framework for optimally staffing a single-class call center that utilizes co-sourced service capacity. The call center is faced with uncertain, time-varying demand and must meet specified quality-of-service (QoS) requirements. Before the arrival rates are realized, the call center responds to an external service provider's pricing strategy by setting its in-house staffing levels and choosing the number of co-sourced agents to place on call in order to minimize its expected total staffing costs. Once the arrival rates are realized, the call center activates the on-call agents to ensure that its QoS constraints are satisfied. The external service provider (or contractor) seeks to maximize expected total revenues over a finite contract period by setting per-agent holding and activation prices. We model the interplay between the call center and contractor as a leader-follower game in which the contractor plays the part of the leader, and the call center represents the follower. We show that the corresponding bilevel programming problem can be reformulated as a quadratically-constrained linear program to obtain the contractor's optimal per-agent holding and activation prices. The call center's optimal staffing problem – a two-stage stochastic integer program (SIP) with recourse – is shown to be highly tractable for a wide range of QoS constraints. A numerical study illustrates the advantages of using our joint optimization framework in the presence of imperfect and asymmetric information.

Chapter 3 considers the problem of optimally replacing multiple stochastically degrading systems using condition-based maintenance. Each system degrades continuously at a rate that is governed by the current state of the environment, and each fails once its own cumulative degradation threshold is reached. The objective is to minimize the sum of the expected total discounted setup, preventive replacement, reactive replacement, and downtime costs over an infinite horizon. For each environment state, we prove that the cost function is monotone nondecreasing in the cumulative degradation level. Additionally, under mild conditions, these monotonicity results are extended to the entire state space. In the case of a single system, we establish that monotone policies are optimal. The monotonicity results help facilitate a tractable, approximate model with state- and action-space transformations and a basis-function approximation of the action-value function. Our computational study demonstrates that high-quality, near-optimal policies are attainable and significantly outperform heuristic policies.

Chapter 4 considers the problem of optimally maintaining a stochastically degrading, single-unit system using heterogeneous spares of varying quality. The system's failures are unannounced; hence, it is inspected periodically to determine its status (functioning or failed). The system continues in operation until it is either preventively or correctively maintained. The available maintenance options include perfect repair, which restores the system to an as-good-as-new condition, and replacement with a randomly-selected unit from the supply of heterogeneous spares. The objective is to minimize the total expected discounted maintenance costs over an infinite time horizon. We formulate the problem using a mixed observability Markov decision process (MOMDP) model in

which the system's age is observable but its quality must be inferred. We show, under suitable conditions, the monotonicity of the optimal value function in the belief about the system quality and establish conditions under which finite preventive maintenance thresholds exist. A detailed computational study reveals that the optimal policy encourages exploration when the system's quality is uncertain and is more exploitive when the quality is highly certain. The study also demonstrates that substantial cost savings are achieved by utilizing our MOMDP-based method as compared to more naive methods of accounting for heterogeneous spares.

2.0 STAFFING AND PRICING IN CO-SOURCED CALL CENTERS

2.1 INTRODUCTION

Customer contact centers, and particularly call centers, play a central role in bridging the gap between manufacturing and service organizations and their customers. Effectively managed call centers serve as a means of direct communication with customers and foster loyalty with customers. However, call center operating costs can be substantial, with 60–80% of these costs stemming from staffing agents to service incoming calls [2]. Increasingly, organizations are electing to co-source a portion or all of their customer support functions. The practice of outsourcing a portion of these functions is often referred to as *co-sourcing*. That is, rather than handling all service requests with in-house agents, a portion of service capacity can be delegated to an external service provider (or contractor). Organizations choose to outsource their operations for a number of reasons, including (but not limited to): cost reduction, handling call overflow, offering extended hours or leveraging an external provider's expertise. However, cost reduction is arguably the key factor in co-sourcing decisions [44]. Service organizations must weigh the benefits (economic and otherwise) of co-sourcing against the potential costs of forfeiting control of their largest source of direct customer support. For those that decide the benefits of co-sourcing outweigh the risks, the question remains: How much service capacity should be outsourced? This question is complicated by the difficulties associated with managing a call center, namely that call center managers must determine staffing levels, rosters and schedules in order to satisfactorily service incoming calls in a timely manner. These tasks are challenging due to the inherently stochastic and dynamic nature of incoming calls.

We consider the problem of optimally staffing a call center that can exercise the option to co-source a portion of its service capacity to an external contractor. Simultaneously, we consider the related problem faced by the contractor: How should contract prices be set so as to maximize the revenue generated by rendering service to the call center? Specifically, with uncertain knowledge of the call center's demand and operating costs, the contractor seeks to set per-period, per-agent holding and activation prices that maximize the expected total revenues accrued over a finite contract horizon. For its part, the call center responds to the contractor's pricing strategy by setting in-house and on-call staffing levels that minimize their expected total staffing costs over the duration of the contract. However, the call center must contend with uncertain call arrival rates. Additionally, it must satisfy two common quality-of-service (QoS) requirements: (1) the probability that an arbitrary customer abandons their call must not exceed a fixed threshold; and (2) the expected time a customer waits before their call is answered must not exceed a fixed time threshold [4, 31, 38, 61]. Once the contract prices are fixed, the call center decides how many in-house agents to staff (in each period) and how many external agents to place on call for each period. We may view these on-call agents as reserved capacity. Subsequently, as needed, they reactively activate agents during each period in order to satisfy their QoS constraints, paying an additional per-agent cost for each activated agent. We model the interplay between the call center and contractor as a leader-follower game in which the contractor plays the part of the leader, and the call center represents the follower. We show that the corresponding bilevel programming problem can be reformulated as a quadratically-constrained linear program whose solution yields the contractor's optimal per-agent holding and activation prices. The call center's optimal staffing problem is formulated as a two-stage stochastic integer program (SIP) with recourse and is shown to be highly tractable. The value of this modeling framework is illustrated through numerical examples using real call center data. Before proceeding, we next discuss existing relevant research.

The body of literature related to the operation and management of call centers is vast and growing. The cogent survey by Gans et al. [30] spans a wide range of topics including modeling, analysis, forecasting, staffing, rostering, long-term planning, and call routing. More recently, Akşin et al. [2] surveyed the broad literature, paying special attention to the impacts and challenges surrounding recent technological advances and research on psychological aspects of call centers. Other surveys, focused on tractable queueing models [55], examined optimization problems related to call centers [59], and surveyed issues related to multi-skill call centers [56, 4]. In what follows, we review call center contributions that are most pertinent to our work here, namely those concerned with staffing under uncertain arrival rates and call center outsourcing and co-sourcing.

Significant effort has been devoted to developing forecasting, staffing, and scheduling models that incorporate arrival rate uncertainty [45, 93, 47, 106, 28, 80, 91, 43, 65, 81]. Chen and Henderson [20] demonstrated that estimation error, arrival rate nonstationarity and random arrival rates may lead to highly uncertain future demand. Gurvich et al. [38] considered a system with multiple customer classes and agent types, probabilistic QoS constraints, and demand volume uncertainty. Their aim was to obtain the minimal staffing level needed to meet the QoS constraints with a chosen probability. They devised a two-step solution procedure in which the first step provides an approximation of the optimal staffing level and a staffing frontier. Subsequently, deterministic problems whose arrival rates are on the staffing frontier are solved, and their solutions are used to generate a solution to the original chance-constrained staffing problem. Gans et al. [31] proposed an integrated forecasting and staffing model when future staffing levels can be adjusted. They developed a parametric forecast model, discretized via Gaussian quadrature, that is stable under a small number of samples. This stability makes it particularly well-suited for use within a stochastic program. Accordingly, they modeled their staffing problem as a stochastic program with recourse and demonstrated its efficacy in satisfying long-run QoS targets with low expected cost. Recently, Bodur and Luedtke [16] proposed an integrated staffing and scheduling model for multiclass service systems under arrival rate uncertainty. Their problem was formulated as a stochastic integer program (SIP), and a new mixed-integer rounding (MIR) technique was used to improve Bender's cuts. Their MIR method is applicable to any SIP with integer first-stage decision variables. Other recent work exploits heavy-traffic approximations to account for arrival rate uncertainty [7, 8, 61]. Bassamboo et al. [9] considered the problem of staffing with constraints on the fraction of abandoned calls. They proposed a fluid limit approach that allows the original staffing problem to be treated as a newsvendor network problem. Extending the quality and efficiency-driven limiting regime, Maman [60] considered a service system whose arrival process is described by a Poisson mixture model.

The literature related to call center outsourcing/co-sourcing is comparatively sparse and fairly recent. Some papers, primarily focused on the case of total outsourcing, consider contracts with various forms of information asymmetry [1, 40, 79, 78]. Akşin et al. [3] considered the problem of contract design and choice, where the contractor sets prices for two different types of contracts and the call center can select between them. Specifically, pay-per-capacity and pay-for-job contracts were considered. In pay-per-capacity contracts, the contractor handles a fixed call volume, and all overflow is handled by the call center. Pay-for-job contracts function in the opposite manner; a base level of calls are handled in-house, and excess calls are routed to the contractor. They showed that, in the presence of uncertain demand, neither type of contract is universally preferred under a multi-period decision horizon. Only a few papers have studied operational decisions within a co-sourcing setting. Van den Schrieck et al. [100] examined the problem of staffing a vendor that offers co-sourcing services. They investigated the issue of bursty arrivals and proposed two staffing methods that utilize peakedness as a measure of burstiness. The first method is an extension of square-root staffing, and the second makes use of the Hayward approximation principles. Gans and Zhou [32] considered a multi-class model with two types of customers; high-value customers must be handled by in-house agents, while low-value customers can be routed to an external vendor. Four call routing schemes of varied complexity and information technology infrastructure requirements were considered. The simpler schemes were shown to perform well when outsourcing requirements are significant. Milner and Olsen [68] considered a similar problem but investigated the impact of different types of service-level agreements. Koçağa [53] proposed a joint model of staffing and outsourcing when there is a cost associated with each call outsourced. Specifically, they sought to minimize the expected long-run average cost when there are costs associated with customer abandonment, outsourcing calls, and staffing costs. A modified square-root staffing rule and threshold outsourcing rule were proposed and shown to be asymptotically optimal (as the mean arrival rate goes to infinity). Moreover, the level of uncertainty in the arrival rate is of the same order as the inherent system fluctuation under a fixed arrival rate.

Our work here differs from the existing call center staffing models in several important ways. First, in most existing models, it is assumed that the call center pays for each *call* that is routed to an external service provider. By contrast, we assume that external agents function identically to dedicated in-house agents. This modeling feature obviates the need for call routing schemes and serves to simplify the process of aligning the motives of the call center and contractor. Second, to the best of our knowledge, our model is the first to consider the joint problem of staffing and pricing, in the presence of co-sourcing, within a queueing framework. We carry out our analysis by assuming the call center functions as an Erlang-A queueing system but do not employ fluid or diffusion approximations; rather, we rely only on steady-state results for the Erlang-A system. Third, we consider information asymmetry between the call center and the contractor. That is, the contractor must set its prices while possessing only partial information about the demand experienced at the call center and its operating costs. Despite the complexities inherent in these models, we are able to establish insightful results. We show that the feasible region of the staffing problem is an integral polyhedron and provide a closed-form expression for the optimal staffing levels. Hence, the staffing problem is tractable even for very large instances. After establishing simple, but useful, bounds for the contractor's pricing problem, we leverage results from the staffing problem to reformulate the optimization problem as a quadratically-constrained linear program. Finally, we provide a numerical study that illustrates the tractability of the pricing problem and its amenability to including an alternative, profit-based objective.

The remainder of this chapter is organized as follows. In Section 2.2, we describe the mathematical model and describe a joint staffing and pricing problem formulated as a two-stage stochastic program. Section 2.3 discusses bounds on the pricing problem and a reformulation of the program, as well as a closed-form solution to the staffing problem. In Section 2.4, we demonstrate the advantages of using our model via numerical examples that utilize real call center data. Finally, we provide some concluding remarks and directions for future work in Section 2.5.

2.2 MODEL FORMULATION

Consider a call center external provider (termed the *contractor*) who seeks to set prices in order to maximize expected revenue (or profits) over a fixed, finite-horizon contract. The contract horizon is partitioned into non-overlapping time intervals, and we denote this planning horizon by $\mathcal{T} := \{1, \ldots, T\}$, where $t \in \mathcal{T}$ represents the tth period. The contractor has a call center client with its own service needs. Contracts between call centers and outsource providers typically take one of two forms: volume- or capacity-based [32, 3]. In capacity-based contracts, capacity is reserved by the call center *a priori* and paid for whether it is utilized or not. In volume-based contracts, fees are assessed only when the contractor's resources are utilized. By contrast, the model we present here allows for a hybrid structure that exhibits features of both volume- and capacity-based contracts. Specifically, the contracts consist of two, per-period, per-agent cost rates: an "on-call" (or holding) cost rate, denoted by c_o ($c_o \geq 0$), and an "activation" cost rate, denoted by c_a ($c_a \geq 0$). Before formulating the contractor's decision problem as a bilevel optimization model, we first present a model to prescribe how the call center should staff its system, assuming a fixed contract with cost vector $c = (c_o, c_a)$.

2.2.1 Call Center Staffing Problem

We consider an inbound single-class, single-pool call center that receives calls during each of the T periods in the planning horizon. During period $t \in \mathcal{T}$, the call center is modeled as an $M/M/s+$

M queueing system. Specifically, for each period t, calls arrive according to a Poisson process with an uncertain arrival rate λ^t, service times are independent and identically distributed (i.i.d.) exponential random variables with rate parameter μ ($\mu > 0$), there are s servers (agents) available to respond to customer requests, and unserved calls abandon the system after an i.i.d. exponentially-distributed patience time with rate parameter α ($\alpha > 0$). Thus, we treat the call center as an Erlang-A model extended to include uncertainty in arrival rate forecasts. The Erlang-A model is frequently employed to represent call centers with abandonment [30].

In what follows, all random variables are defined on a common measurable space $(\Omega, \mathscr{F})$, and each arrival rate λ^t is assumed to be a discrete random variable with finite support $\Lambda^t := \{\lambda_1^t, \ldots, \lambda_{K(t)}^t\}$ where $K(t) \in \mathbb{N}$ and $\lambda_k^t \geq 0$. We do not impose the assumption that the arrival rates $\{\lambda^t : t \in \mathcal{T}\}$ are mutually independent; however, it is assumed that their unconditional probability mass functions are known *a priori*, i.e. before any staffing decisions are made. For each $t \in \mathcal{T}$, define the probabilities

$$p_k^t := \mathbb{P}\left(\lambda^t = k\right), \quad k \in \Lambda^t$$

where $(\Omega, \mathscr{F}, \mathbb{P})$ is a complete probability space.

The call center seeks to minimize the expected total staffing costs over the planning horizon while meeting quality-of-service (QoS) requirements. For the purposes of this model, we assume the QoS requirements are specified as upper bounds on functions that are nonincreasing in the number of servers s, and nondecreasing in the arrival rate λ^t. For example, two common constraints are ensuring that the steady-state expected delay (time in queue) experienced by an arbitrary arrival, and the probability of abandonment, do not exceed fixed thresholds, i.e.,

$$\mathbb{P}(Ab) \leq \psi_1 \quad \text{and} \quad \mathbb{E}(D) \leq \psi_2, \tag{2.1}$$

where D is the steady-state delay, $\mathbb{P}(Ab)$ is the probability of abandonment, and ψ_1 and ψ_2 are specified thresholds. Define the standard lower incomplete Gamma function

$$\gamma(x, y) := \int_0^y t^{x-1} e^{-t} dt, \quad x > 0,\ y \geq 0,$$

the function $\mathcal{E}$ as the reciprocal of the well-known Erlang-B formula (with deterministic arrival rate λ)

$$\mathcal{E}\left(\frac{\lambda}{\mu}, s\right) = \frac{\sum_{j=0}^{s} \frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j}{\frac{1}{s!}\left(\frac{\lambda}{\mu}\right)^s},$$

and an auxiliary function $J(\cdot)$ by

$$J(s,\lambda,\mu,\alpha) = \frac{\exp\left(\frac{\lambda}{\mu}\right)}{\alpha} \cdot \left(\frac{\alpha}{\lambda}\right)^{\frac{s\mu}{\alpha}} \cdot \gamma\left(\frac{s\mu}{\alpha}, \frac{\lambda}{\alpha}\right).$$

Applying well-known results for the $M/M/s+M$ queue [61], we have

$$\mathbb{P}(Ab|s,\lambda,\mu,\alpha) = \frac{1+(\lambda-n\mu)J'}{\mathcal{E}'+\lambda J'},$$

and

$$\mathbb{E}(D|s,\lambda,\mu,\alpha) = \alpha^{-1}\,\mathbb{P}(Ab|s,\lambda,\mu,\alpha), \tag{2.2}$$

where $\mathcal{E}' = \mathcal{E}(\lambda/\mu, s-1)$ and $J' = J(s,\lambda,\mu,\alpha)$. Therefore, the nonlinear constraints (2.1) are equivalent to

$$s \geq \min\left\{n \in \mathbb{N} : \mathbb{P}(Ab|n,\lambda,\mu,\alpha) \leq \min\{\psi_1, \alpha\cdot\psi_2\}\right\}. \tag{2.3}$$

Define s_k as the minimum number of servers needed to satisfy (2.3) for arrival rate $\lambda = k$. To address the problem of determining the number of in-house and external agents to use, we formulate the problem as a two-stage SIP. For each $t \in \mathcal{T}$, let x^t be the number of scheduled in-house agents and b^t the number of co-sourced agents placed on call. Note that x^t and b^t are first-stage decisions made prior to the realization of the arrival rates λ^t. For convenience, we denote the vector of first-stage decisions by $\boldsymbol{x} = (x^1,\ldots,x^T,b^1,\ldots,b^T)$. The number of *activated* on-call agents, given the arrival rate realization λ^t, will be denoted by the recourse variables $y^t_{\lambda^t}$. The call center's problem of minimizing the expected total staffing costs, subject to QoS constraints, is formulated as the following SIP:

$$\begin{aligned} \min \quad & \sum_{t\in\mathcal{T}} (c_h x^t + c_o b^t) + \mathbb{E}(Q(\boldsymbol{x},\boldsymbol{\lambda})) && \qquad (2.4a)\\ \text{s.t.} \quad & x^t \in \mathbb{Z}_+,\ b^t \in \mathbb{Z}_+, && t \in \mathcal{T} \qquad (2.4b) \end{aligned}$$

where c_h is the call center's in-house per-period, per-agent cost rate, and for each $\omega \in \Omega$,

$$\begin{aligned} Q(\boldsymbol{x},\boldsymbol{\lambda}(\omega)) = \min \quad & \sum_{t\in\mathcal{T}} c_a y^t_{\lambda^t(\omega)} && \qquad (2.5a)\\ \text{s.t.} \quad & x^t + y^t_{\lambda^t(\omega)} \geq s_{\lambda^t(\omega)}, && t \in \mathcal{T} \qquad (2.5b)\\ & y^t_{\lambda^t(\omega)} \leq b^t, && t \in \mathcal{T} \qquad (2.5c)\\ & y^t_{\lambda^t(\omega)} \in \mathbb{Z}_+, && t \in \mathcal{T}. \qquad (2.5d) \end{aligned}$$

The objective function (2.4a) represents the cost of setting in-house staffing and on-call levels plus the expected cost of activated servers. Constraints (2.5b) guarantee that the QoS requirements are met for each realization, and constraints (2.5c) ensure that the number of activated, co-sourced agents does not exceed the number reserved. Equivalently, formulation (2.4) can be expressed in its extensive form

$$\min \quad \sum_{t \in \mathcal{T}} (c_h x^t + c_o b^t) + \sum_{t \in \mathcal{T}} \sum_{k \in \Lambda^t} c_a p_k^t y_k^t \tag{2.6a}$$

$$\text{s.t.} \quad x^t + y_k^t \geq s_k, \qquad t \in \mathcal{T},\ k \in \Lambda^t \tag{2.6b}$$

$$y_k^t \leq b^t, \qquad t \in \mathcal{T},\ k \in \Lambda^t \tag{2.6c}$$

$$x^t \in \mathbb{Z}_+,\ b^t \in \mathbb{Z}_+,\ y_k^t \in \mathbb{Z}_+, \qquad t \in \mathcal{T},\ k \in \Lambda^t \tag{2.6d}$$

The objective function (2.6a) represents the expected total staffing cost over the contract horizon. Constraints (2.6b) ensure that the QoS constraints are met for all realizations, while constraints (2.6c) guarantee that the number of activated, co-sourced agents does not exceed the number reserved. It is important to note that, although the planning horizon is comprised of multiple time periods, it lacks dependence between recourse decisions and future time periods. This lack of dependence allows for the formulation of a two-stage stochastic program, in lieu of a multi-stage formulation.

Next, we examine the optimal staffing levels for model (2.6). We begin by analyzing the case of a single-period problem in which formulation (2.6) simplifies to

$$\min \quad z(x, b, y) = c_h x + c_o b + \sum_{k \in \Lambda} c_a p_k y_k \tag{2.7a}$$

$$\text{s.t.} \quad x + y_k \geq s_k, \qquad k \in \Lambda, \tag{2.7b}$$

$$y_k \leq b, \qquad k \in \Lambda, \tag{2.7c}$$

$$x \in \mathbb{Z}_+,\ b \in \mathbb{Z}_+,\ y_k \in \mathbb{Z}_+, \qquad k \in \Lambda. \tag{2.7d}$$

We assume that $\Lambda = \{\lambda_1, \ldots, \lambda_K\}$ is completely ordered so that $\lambda_i < \lambda_{i+1}$, $i = 1, \ldots, K-1$. Note that this also implies $s_{\lambda_i} \leq s_{\lambda_{i+1}}$, as the QoS measures are assumed to be nondecreasing in the arrival rate. We also let $s_k \equiv s_{\lambda_k}$ and $p_k \equiv p_{\lambda_k}$ for notational convenience. While we seek the optimal first-stage and recourse decisions that jointly minimize $z(x, b, y)$, one can first consider the optimal capacity and recourse decisions for a fixed in-house staffing level. Lemma 2.1 characterizes the optimal decisions b and y, given some fixed staffing level.

Lemma 2.1 *Given first-stage staffing levels $x \leq s_K$, the optimal single-period numbers of on-call and activated agents are respectively given by $b^*(x) = s_K - x$ and $y_k^*(x) = \max\{s_k - x, 0\}$.*

Lemma 2.1 states that, given a staffing level, the optimal co-sourcing decisions are the smallest decisions that satisfy constraints (2.7b) and (2.7c). Thus $z^*(x)$, the optimal objective value as a function of the in-house staffing level, is given by

$$z^*(x) = c_h x + c_o(s_K - x) + \sum_{k \leq K} c_a \, p_k \, \max\{s_k - x, 0\}.$$

Thus, for $n \in \{1, \ldots, K-1\}$ and $x \in [s_n, s_{n+1})$, the marginal cost of increasing the staffing level by one in-house agent is

$$\Delta(x) := z^*(x+1) - z^*(x) = c_h - \left(c_o + c_a \sum_{k>n} p_k \right). \tag{2.8}$$

Lemma 2.2 *The optimal single-period, in-house staffing level x^* that minimizes $z^*(x)$ is given by*

$$x^* = \begin{cases} \max\left\{ s_k \in \mathbb{Z}_+ : c_o + c_a \sum_{i>k} p_i \geq c_h \right\}, & c_h < c_o + c_a, \\ 0, & c_h \geq c_o + c_a. \end{cases} \tag{2.9}$$

Proof. When $c_h \geq c_o + c_a$, it is clear that the marginal cost of increasing the in-house staffing level, $\Delta(x)$, is nonnegative for $x \in \{0, \ldots, s_K\}$; therefore, it is never beneficial to increase x. For the case when $c_h < c_o + c_a$, $\Delta(x) \leq 0$ for $x \in [s_n, s_{n+1})$, as long as

$$c_o + c_a \sum_{n<k\leq K} p_k > c_h.$$

If $c_o + c_a \sum_{k>n} p_k < c_h$, the marginal cost is positive for all $n' \geq n$. As the marginal cost remains constant on the intervals $[s_n, s_{n+1})$, this sign change in Δ must occur where $x = s_k$ for some $k \in \{1, \ldots, K\}$. ∎

Lemmas 2.1 and 2.2 characterize the optimal solution when there is only a single time period. We note here that, if outsourcing is cheaper than staffing with in-house agents, co-sourcing is never optimal in the single-period problem; however, co-sourcing can be optimal when the sum of the holding and activation costs exceeds the in-house cost. While the latter result might appear counterintuitive on first glance, it should be noted that the activation cost is not incurred unless the arrival rate warrants the addition of an on-call agent. Proposition 2.1 extends this result to the multi-period problem given by formulation (2.6).

Proposition 2.1 *The optimal first stage solutions x^{t*} to formulation (2.6) are*

$$x^{t*} = \begin{cases} \max\left\{ s_k^t : \left(c_a \sum_{j>k, j\in\Lambda^t} p_j^t + c_o \right) \geq c_h \right\}, & c_h < c_o + c_a, \\ 0, & c_h \geq c_o + c_a. \end{cases} \tag{2.10}$$

Proof. Because there are no decision variables that link the constraints between periods, the problem can be decomposed into T subproblems, each of which takes the form of (2.7) with $\Lambda = \Lambda^t$ for all t. The result then follows by Lemma 2.2. ∎

2.2.2 Contractor's Pricing Problem

It is assumed that the contractor does not precisely know the call center's in-house cost rate or forecasted arrival rates (i.e., the contractor's information is both incomplete and imperfect). Therefore, the contractor faces the problem of setting contract prices that maximize its expected revenue (or profits) under information asymmetry and the inherent stochasticity of the arrival rates. This problem can be viewed as a leader-follower game in which the contractor acts as the leader, and the call center acts as the follower.

To formulate the problem, we assume that for each period t the arrival rate forecast of λ^t originates from a distribution function F_{λ^t} that is parameterized by some vector $\boldsymbol{\theta}^t$ (that is, $F_{\lambda^t}(\cdot; \boldsymbol{\theta}^t)$). Moreover, we assume that the contractor knows the parametric family $\{F_{\lambda^t}(\cdot; \theta)\}$, but not the precise value of $\boldsymbol{\theta}^t$. Specifically, it is assumed that

$$\left(c_h, \boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \ldots, \boldsymbol{\theta}^T\right) = (c_h, \boldsymbol{\theta})$$

is a random vector with known distribution function G and support $\Xi = \Xi_c \times \Xi_\theta$, where Ξ_c and Ξ_θ are the supports for c_h and $\boldsymbol{\theta}$, respectively.

The contractor's random revenue, R, is the sum of holding costs and conditional expected activation costs given by

$$R(c; (c_h, \boldsymbol{\theta})) = \sum_{t\in\mathcal{T}} b^t(c; (c_h, \boldsymbol{\theta}))c_o + \sum_{k\in\Lambda(\boldsymbol{\theta}^t)} p_k^t(\boldsymbol{\theta}) y_k^t(c; (c_h, \boldsymbol{\theta})), \tag{2.11}$$

where $c = (c_o, c_a)$ are the price levels set by the contractor, $b^t(c; (c_h, \boldsymbol{\theta}))$ and $y_k^t(c; (c_h, \boldsymbol{\theta}))$ are solutions to (2.6) with holding cost c_h and, for each $t \in \mathcal{T}$, λ^t is distributed by $F_{\lambda^t}(\cdot; \boldsymbol{\theta}^t)$, and $p_k^t(\boldsymbol{\theta})$ and $\Lambda(\boldsymbol{\theta}^t)$ are the mass function and support of λ^t induced by $F_{\lambda^t}(\cdot;\boldsymbol{\theta}^t)$, respectively. Thus, in this formulation, the contractor seeks to maximize the expected revenue,

$$\mathbb{E}_G(R(c)) = \int_{\Xi} \left(\sum_{t \in \mathcal{T}} b^t(c;\xi) c_o + \sum_{k \in \Lambda(\xi_2^t)} p_k^t(\xi_2) y_k^t(c;\xi) c_a \right) dG(\xi), \tag{2.12}$$

by selecting the holding and activation costs c_o and c_a where $\xi = (\xi_1, \xi_2)$ for $\xi_1 \in \Xi_c$ and $\xi_2 \in \Xi_\theta$.

Within this framework, the contractor first sets the price levels c_o and c_a, and subsequently, the call center sets their optimal staffing levels using formulation (2.6). It is assumed that, if the follower can choose between multiple decisions, each of which minimize their expected costs, then the decision that most benefits the leader will be chosen. It is well known that leader-follower games can be formulated as bilevel programming problems [24]. Therefore, the optimal contract pricing problem can be stated as follows:

$$\max \quad \mathbb{E}_G\left(R(c)\right) \tag{2.13a}$$

$$\text{s.t.} \quad c_o \geq 0,\ c_a \geq 0, \tag{2.13b}$$

where, for each $\xi = (\xi_1, \xi_2) \in \Xi$ and $c \geq 0$, $y_k^t(c;\xi)$ and $b^t(c;\xi)$ solve

$$\min \sum_{t \in \mathcal{T}} c_o b^t(c;\xi) + \xi_1\, x^t(c;\xi) + \sum_{k \in \Lambda(\xi_2^t)} p_k^t(\xi_2) c_a y_k^t(c;\xi) \tag{2.14a}$$

$$\text{s.t.} \quad -x^t(c;\xi) - y_k^t(c;\xi) \leq -s_k, \qquad t \in \mathcal{T},\ k \in \Lambda(\xi_2^t), \tag{2.14b}$$

$$y_k^t(c;\xi) - b^t(c;\xi) \leq 0, \qquad t \in \mathcal{T},\ k \in \Lambda^t(\xi_2^t), \tag{2.14c}$$

$$x^t(c;\xi) \in \mathbb{Z}_+,\ b^t(c;\xi) \in \mathbb{Z}_+,\ y_k^t(c;\xi) \in \mathbb{Z}_+, \qquad t \in \mathcal{T},\ k \in \Lambda^t(\xi_2^t). \tag{2.14d}$$

2.2.3 Profit-Based Objective

When the contractor's operation is substantially larger than that of the call center, it is reasonable to treat the contractor's pool of agents as infinite, thereby justifying a revenue-based objective. However, in reality, the contractor's resources are also constrained; consequently, the marginal labor costs associated with a particular call center contract can be substantial. Furthermore, the contractor may not have sufficient available capacity to satisfy the required activations of the call center; hence, we also consider a certain type of service-level agreement (SLA). Specifically, we assume that for each agent the contractor is unable to provide (per period), it must pay a fixed cost r ($r > 0$).

If the number of agents to be activated in a time period $t \in \mathcal{T}$ is y^t, and the number of servers that are available by the contractor is $\tilde{y}^t$ ($\tilde{y}^t < y^t$), then the shortfall of servers S^t is given by $S^t = y^t - \tilde{y}^t$. Thus, in any time period, the contractor owes the call center a fee, due to the SLA, of rS^t, where S^t is a random variable with finite support $\{0, \dots, b^t\}$. Additionally, the contractor can increase its period t staffing level by an amount z^t at a per-agent, per-period cost of c_z. Thus, the contractor's profit Π is given by its revenue R less its costs C, where R is defined in (2.11) and C is given by

$$C(c; (c_h, \theta)) = \sum_{t \in \mathcal{T}} c_z z^t + rS^t. \tag{2.15}$$

Hence,

$$\mathbb{E}_G(\Pi(c)) = \mathbb{E}_G(R(c) - C(c)) = \mathbb{E}_G(R(c)) - \sum_{t \in \mathcal{T}} c_z z^t - r\mathbb{E}_G(S^t | z^t, y^t(c; \xi)), \tag{2.16}$$

where the exact form of $\mathbb{E}_G(S^t | z^t, y^t(c; \xi))$ depends on the contractor's particular staffing, scheduling, and server demand considerations. We do not impose any assumptions on the functional form of $\mathbb{E}_G(S^t | z^t, y^t(c; \xi))$. Pragmatically, it is only required that the expectation of S^t, conditioned on $\xi \in \Xi$, be computationally tractable.

2.3 REFORMULATING THE PRICING PROBLEM

Solving the bilevel problem (2.13) is nontrivial using either the revenue- or profit-based objective, so we first set out to obtain lower and upper bounds on the contractor's expected revenues. These bounds help the contractor determine whether or not it may be worthwhile to take on certain call center clients. For the remainder of the chapter, we assume that c_h and $\boldsymbol{\theta}$ are independent random elements with distributions G_c and G_θ, respectively. Proposition 2.2 provides lower and upper bounds on the leader's optimal objective function value.

Proposition 2.2 *Let z_L^* denote the optimal objective value of the contract pricing problem. Then,*

$$\kappa \, \bar{c}_h \le z_L^* \le \kappa \, \mathbb{E}_{G_c}(c_h) \tag{2.17}$$

where $\kappa = \mathbb{E}_{G_\theta}\left(\sum_{t \in \mathcal{T}} \max\left\{s_k : k \in \Lambda(\boldsymbol{\theta}^t)\right\}\right)$ and $\bar{c}_h := \min\{c_h : c_h \in \Xi_c\}$.

Proof. We will first establish the upper bound of (2.17). Let $R(c; \xi)$ be the value of the leader's objective function evaluated at $c = (c_o, c_a)$, $\xi = (\xi_1, \xi_2) \in \Xi$, and $z_F^*(c; \xi)$ be the follower's optimal objective function value for the same arguments. By comparing the objective functions of the leader and follower, it is seen that

$$R(c;\xi) = z_F^*(c;\xi) - \sum_{t\in\mathcal{T}} \xi_1 x^{t*}(c;\xi) \le z_F^*(c;\xi), \tag{2.18}$$

where $x^{t*}(c;\xi)$ are the optimal in-house decisions corresponding to $z_F^*(c;\xi)$. Note that for all c_o and c_a, the decision to staff entirely in-house is feasible. Therefore,

$$z_F^*(c;\xi) \le \xi_1 \sum_{t\in\mathcal{T}} \max\{s_k : k \in \Lambda(\xi_2^t)\}, \tag{2.19}$$

where the right-hand side of (2.19) corresponds to the cost of staffing entirely in-house. Combining (2.18) and (2.19) and taking the expectation of each side of the resulting inequality yields

$$\mathbb{E}_G(R(c)) \le \mathbb{E}_G\left(c_h \sum_{t\in\mathcal{T}} \max\left\{s_k : k \in \Lambda(\boldsymbol{\theta}^t)\right\}\right) = \kappa\, \mathbb{E}_{G_c}(c_h), \tag{2.20}$$

where the equality holds due to the independence of c_h and $\boldsymbol{\theta}$. As (2.20) holds for any $c \ge 0$, it also holds for $c^* \in \operatorname{argmax}\{c \ge 0 : \mathbb{E}_G(R(c))\}$. Therefore,

$$z_L^* = \mathbb{E}_G(R(c^*)) \le \kappa\, \mathbb{E}_{G_c}(c_h).$$

Next, to prove the lower bound, note that if the leader sets the cost at $c = (\bar{c_h}, 0)$, then it is optimal for the follower to outsource entirely for any scenario $\xi \in \Xi$ (as established in Proposition 2.1). Hence,

$$\sum_{t\in\mathcal{T}} \xi_1 x^{t*}((\bar{c}_h, 0);\xi) = 0, \quad \xi = (\xi_1, \xi_2) \in \Xi,$$

which implies

$$R((\bar{c}_h, 0);\xi) = z_F^*((\bar{c}_h, 0);\xi) = \bar{c}_h \sum_{t\in\mathcal{T}} \max\{s_k : k \in \Lambda(\xi_2^t)\}. \tag{2.21}$$

Taking the expectation of both sides of (2.21) shows that

$$\mathbb{E}_G[R((\bar{c}_h, 0))] = \kappa\, \bar{c}_h \le \max_{c\ge 0} \mathbb{E}_G[R(c)] = z_L^*,$$

where the equality follows directly, and the inequality holds since $\bar{c}_h$ is a feasible solution. ■

Corollary 2.1 *If $\Xi_c = \{c_h\}$, the lower and upper bounds of (2.17) are tight and achieved at the point $(c_o^*, c_a^*) = (c_h, 0)$.*

Corollary 2.1 asserts that, if the contractor knows with certainty c_h, then it is optimal to set the on-call cost rate so that it matches the in-house rate, and to set the activation cost rate to zero. For this case, the follower chooses between two equal-cost strategies: (a) use in-house agents exclusively, or (b) use outsourced agents exclusively. This setting makes it clear that some assumption must be made regarding the call center's behavior when faced with multiple optimal solutions. Our model assumes that the follower chooses a cost-minimizing strategy that most benefits the leader (namely option (b)). This assumption is realistic in cases where a company decides that outsourcing their call center simplifies their internal operations, or that outsourcing reduces their overhead (in turn leading to indirect cost savings). Additionally, we note that while Proposition 2.2 is given in terms of the revenue-based objective, it readily extends to the profit-based objective by taking the lower bound to be $\kappa\bar{c}_h - \mathbb{E}_G(C(\bar{c}_h, 0))$ and leaving the upper bound unchanged, or taking it to be $\kappa\,\mathbb{E}_{G_c}(c_h) - \min_c \mathbb{E}_G(C(c))$, noting that computation of the latter may be nontrivial.

In order to solve the optimal contract pricing problem when c_h is not constant, we must solve formulation (2.13) – a quadratic, mixed-integer bilevel program in which each evaluation of the objective function requires the solution of a Lebesgue integral. If the support Ξ is at most countable, it is clear that (2.12) reduces to a weighted sum; otherwise, it can be numerically approximated by sampling methods. In either case, we can consider the follower's problem, given a particular scenario $\xi \in \Xi$. Proposition 2.3 establishes that the feasible region of the subproblem, for each scenario ξ, is an integral polyhedron.

Proposition 2.3 *For any scenario ξ, the set $P(\xi) := \{\boldsymbol{x}(\xi) : A(\xi)\,\boldsymbol{x}(\xi) \leq d(\xi)\}$ is an integral polyhedron, where $\boldsymbol{x}(\xi)$ is the vector of the subset of follower's decisions corresponding to scenario $\xi \in \Xi$ and $A(\xi)\,\boldsymbol{x}(\xi) \leq d(\xi)$ corresponds to the constraint sets (2.14b) and (2.14c).*

Proof. For notational convenience, we suppress the dependence on ξ. First, we note that P is nonempty, as the decision to staff entirely in-house is feasible for $x^t \geq \max_{k\in\Lambda^t}\{s_k\}$. We note that the right-hand side vector d is integral, as it is composed of the integer threshold values s_k and zeroes. Therefore, it suffices to show that the matrix A is totally unimodular. Let the columns of A be ordered so that the first T columns correspond to the coefficients of the decision variables x^t, the next $N := \sum_{t\in\mathcal{T}} |\Lambda^t|$ columns correspond to the coefficients of each y_k^t, and the final T columns correspond to the coefficients of b^t. Next, for any subset of the columns $J \subset \{1, \ldots, 2T + N\}$ let J_1 and J_2 partition J such that $J_1 := J \cap \{1, \ldots, T\}$ and $J_2 := J \cap \{T + 1, \ldots, 2T + N\}$. We note that for any row i corresponding to a constraint given by (2.14b) that

$$\sum_{j \in J_1} a_{ij} \in \{-1, 0\} \quad \text{and} \quad \sum_{j \in J_2} a_{ij} \in \{-1, 0\},$$

as there is at most one non-zero entry in the columns of J_1 or J_2 in row i. Therefore,

$$\left(\sum_{j \in J_1} a_{ij} - \sum_{j \in J_2} a_{ij} \right) \in \{-1, 0, 1\},$$

from which it can be discerned that

$$\left| \sum_{j \in J_1} a_{ij} - \sum_{j \in J_2} a_{ij} \right| \leq 1. \tag{2.22}$$

Next, for any row i corresponding to a constraint given by (2.14c), we have that $a_{ij} = 0$ for all $j \in J_1$ and that there are no more than two non-zero entries in the columns of J_2 within row i. However, these non-zero entries cannot have the same sign; therefore,

$$\sum_{j \in J_1} a_{ij} = 0 \quad \text{and} \quad \sum_{j \in J_2} a_{ij} \in \{-1, 0\},$$

again implying inequality (2.22). Then by Ghouila-Houri [36], the matrix A is totally unimodular; hence, it follows that P is integral (see Theorem 2 of [41]). ■

Proposition 2.3 reveals that the linear relaxation of the follower's subproblem has an integral optimal solution; therefore, the optimal solution to the linear programming (LP) dual of the relaxed follower's subproblem yields an optimal objective value that is equal to the follower's original subproblem. Define $u(\xi_2)$ and $v(\xi_2)$ as the vectors of dual variables corresponding to constraint sets (2.14b) and (2.14c), respectively. Then the dual of the follower's subproblem, for fixed leader's decisions (c_o, c_a), is

$$\begin{aligned}
\max \quad & \sum_{t \in \mathcal{T}} \sum_{k \in \Lambda(\xi_2^t)} s_k u_k^t(\xi_2) && && \text{(2.23a)} \\
\text{s.t.} \quad & \sum_{k \in \Lambda(\xi_2^t)} u_k^t(\xi_2) \leq \xi_1, && t \in \mathcal{T}, && \text{(2.23b)} \\
& \sum_{k \in \Lambda(\xi_2^t)} v_k^t(\xi_2) \leq c_o, && t \in \mathcal{T}, && \text{(2.23c)} \\
& u_k^t(\xi_2) - v_k^t(\xi_2) \leq p_k^t(\xi_2) c_a, && t \in \mathcal{T},\ k \in \Lambda(\xi_2^t), && \text{(2.23d)} \\
& u_k^t(\xi_2) \geq 0,\ v_k^t(\xi_2) \geq 0, && t \in \mathcal{T},\ k \in \Lambda(\xi_2^t). && \text{(2.23e)}
\end{aligned}$$

Additionally, if Ξ is finite, we can represent Ξ by $\{(c_h, \boldsymbol{\theta})_1, (c_h, \boldsymbol{\theta})_2, \ldots, (c_h, \boldsymbol{\theta})_N\}$, for $N = |\Xi|$. By utilizing (2.23), Proposition 2.3, and strong duality, we arrive at the main result of this section, namely Theorem 2.1. In what follows, let $\mathcal{N} = \{1, 2, \ldots, N\}$ and $q_i = \mathbb{P}\big((c_h, \boldsymbol{\theta}) = (c_h, \boldsymbol{\theta})_i\big)$.

Theorem 2.1 *The values c_o^* and c_a^* that solve the quadratically-constrained linear program*

$$\max \quad \sum_{i \in \mathcal{N}} q_i \left(\sum_{t \in \mathcal{T}} -c_h(i) x^t(i) + \sum_{k \in \Lambda^t(i)} s_k(i) u_k^t(i) \right) \tag{2.24a}$$

$$\text{s.t.} \quad x^t(i) + y_k^t(i) \geq s_k, \qquad t \in \mathcal{T},\ i \in \mathcal{N},\ k \in \Lambda^t(i), \tag{2.24b}$$

$$y_k^t(i) \leq b^t(i), \qquad t \in \mathcal{T},\ i \in \mathcal{N},\ k \in \Lambda^t(i), \tag{2.24c}$$

$$\sum_{k \in \Lambda^t(i)} u_k^t(i) \leq c_h(i), \qquad t \in \mathcal{T},\ i \in \mathcal{N}, \tag{2.24d}$$

$$\sum_{k \in \Lambda^t(i)} v_k^t(i) \leq c_o, \qquad t \in \mathcal{T},\ i \in \mathcal{N}, \tag{2.24e}$$

$$u_k^t(i) - v_k^t(i) \leq p_k^t(i) c_a, \qquad t \in \mathcal{T},\ i \in \mathcal{N},\ k \in \Lambda^t(i), \tag{2.24f}$$

$$\begin{aligned} &\sum_{t \in \mathcal{T}} c_o b^t(i) + c_h(i) x^t(i) \\ &\qquad + \sum_{t \in \mathcal{T}} \sum_{k \in \Lambda^t(i)} p_k(i) c_a\, y_k^t(i) - s_k u_k^t(i) = 0, \qquad i \in \mathcal{N}, \end{aligned} \tag{2.24g}$$

$$c_o, c_a, x^t(i), y_k^t(i), b^t(i), u_k^t(i), v_k^t(i) \geq 0, \qquad t \in \mathcal{T},\ i \in \mathcal{N},\ k \in \Lambda^t(i), \tag{2.24h}$$

are the decisions that maximize the contractor's revenue in formulation (2.13) – (2.14)*.*

Proof. The constraints (2.24b) and (2.24c) guarantee that the follower's subproblems are each primal feasible, and constraints (2.24d) – (2.24f) ensure dual feasibility. By strong duality, at optimality, constraints (2.24g) must hold, or equivalently, for each $i \in \mathcal{N}$

$$\sum_{t \in \mathcal{T}} c_o b^t(i) + \sum_{k \in \Lambda^t(i)} p_k(i) c_a y_k^t(i) = \sum_{t \in \mathcal{T}} -c_h(i) x^t(i) + \sum_{k \in \Lambda^t(i)} s_k^t(i) u_k^t(i). \tag{2.25}$$

It is seen that the expectation of the left-hand side of (2.25), with respect to the distribution $\{q_i : i \in \mathcal{N}\}$, is simply the leader's objective function; therefore, the quadratic objective function (2.13) can be replaced with the expected value of the right-hand side of (2.25), and the result is proved. ∎

Theorem 2.1 shows that solving the nonlinear, mixed-integer, bilevel formulation (2.13) – (2.14) can be reduced to solving a quadratically-constrained, linear program with N quadratic equality constraints which are nonconvex. Global optimization techniques can be used to solve this problem directly [12]. However, because general purpose global optimization solvers do not exploit the particular constraint structure of a formulation, they are, in general, too computationally expensive to utilize on large-scale problems. In our case, the number of decision variables in the formulations for our pricing problem grow proportional to the product of the cardinality of the sets $\mathcal{T}, \mathcal{N}$, and $\Lambda^t(i)$. Therefore, even for a small call center (needing 20 or fewer servers per period), considering half-day shifts over a week-long horizon, and sampling only $N = 100$ parameter realizations, the number of decision variables in the formulation given by Theorem 2.1 is over 85,000, making it prohibitively large for standard global optimization techniques. For these reasons, in our numerical studies, we exploit the result of Proposition 2.1 to solve the problem (as formulated in (2.13)) directly.

2.4 NUMERICAL STUDY

In this section, we provide a numerical study to illustrate the usefulness and tractability of the staffing and contract pricing models. These illustrations make use of publicly available data originating from a telephone call center of an anonymous bank in Israel [37]. These data contain the archives of all calls placed to the call center in 1999. The bank's call center is typically staffed 159 hours per week, which we partition into 318 30-minute periods. From the arrival times, arrival counts were tallied for each time period in each week. The arrival rates for the subsequent week were then forecasted using the procedure described in Gans et al. [31]. Specifically, (on a square-root scale) the arrival rate process is treated as a hidden, Gaussian AR(1) process. The Gaussian forecasts were then truncated to ensure nonnegativity and contain all values within four standard deviations of their respective means. Figure 1 depicts the square-root scaled arrival rate forecasts for the first day in the forecasted horizon (whiskers depict the interquartile range). It should be noted that this call center is relatively small and routinely requires fewer than five agents to meet its demand.

While the supports of the forecasted arrival rates are continuous, this is not problematic, as (assuming the arrival rates are bounded) there exists a discretization leading to an exact reformulation. In particular, for period t we assume that the arrival rates are bounded above by four standard deviations above their mean, we call this value $u(t)$, i.e., $u(t) = \mathbb{E}(\lambda^t) + 4\sqrt{\mathrm{Var}(\lambda^t)}$. Hence, we consider the support to be $\Lambda^t = [0, u(t)]$. Thus, for a fixed set of QoS constraints, the number

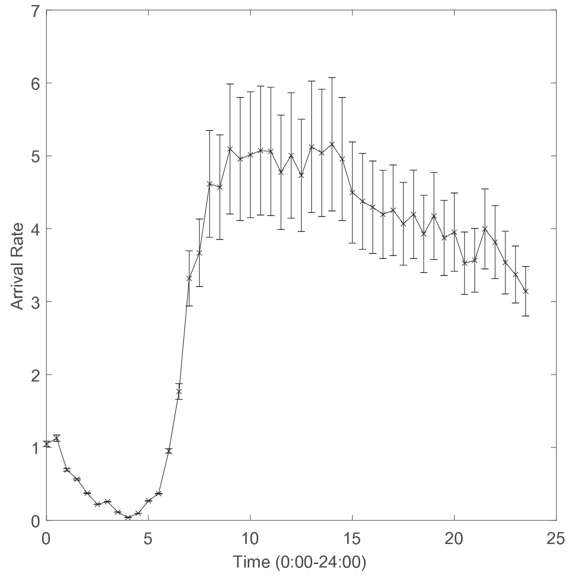


Figure 1: Square-root scale arrival rate forecasts.

of servers needed to satisfy the constraints in time t is contained in the set $\Delta^t = \{1, \ldots, s_{u(t)}\}$, where $s_{u(t)}$ is the minimum number of servers needed to satisfy the QoS requirements for arrival rate $\lambda^t = u(t)$. Next, we determine the probability of needing exactly k servers for each $k \in \Delta^t$. We do this numerically, proceeding backwards from $u(t)$. We find the minimum $\lambda \in \Lambda^t$, such that $s_{u(t)} - 1$ servers are insufficient to satisfy the QoS constraints, call this value $u_{-1}(t)$. Then, the probability that exactly $s_{u(t)}$ servers are needed is given by $\mathbb{P}(\lambda^t \in [u_{-1}(t), u(t)])$. We then proceed to determine the minimum $\lambda \in \Lambda^t$ such that $s_{u(t)} - 2$ servers are insufficient to satisfy the QoS constraints, call this value $u_{-2}(t)$, then the probability that exactly $s_{u(t)} - 1$ servers are needed is given by $\mathbb{P}(\lambda^t \in [u_{-2}(t), u_{-1}(t)])$, etc. Generally, using the notation of Section 2.2.1, for each $k \in \Delta^t$, the probability that k servers are needed (p_k^t) is given by

$$p_k^t = \mathbb{P}(\lambda^t \in \{\lambda \in \Lambda^t : s_\lambda = k\}).$$

The customer patience parameter (α), was approximated based on (2.2), the theoretical relationship between steady-state expected delay and steady-state probability of abandonment, using

$$\alpha \approx \frac{\text{Fraction of calls abandoned}}{\text{Average wait time}}.$$

Lastly, the service rate (μ) was taken to be the reciprocal of the average service time. All problem instances were coded within the MATLAB R2016a computing environment and executed on a personal computer with a 3.50 GHz processor and 8 GB of RAM.

2.4.1 Staffing Level Problem

In order to demonstrate the usefulness of the staffing level model with recourse, we compare its performance to that of other well-known staffing rules. The staffing rules only differ by their prescription of in-house staffing levels; that is, each rule prescribes an in-house staffing level and then subsequently optimally utilizes recourse given that fixed in-house staffing level (optimal second-stage decisions, but possibly suboptimal first-stage decisions). For each staffing rule the in-house staffing level is determined as follows:

- **Worst-case**: This rule staffs in-house agents to meet the QoS requirements for the worst-case (largest) arrival rate realization;
- **Safety Staffing**: Motivated by heavy traffic approximations, this rule staffs in-house agents according to the (expected) offered load, as well as a safety level. Specifically, for service grade parameter β ($\beta \in \mathbb{R}$) and service rate μ, the staffing level in period t is

$$s^t = \frac{\mathbb{E}(\lambda^t)}{\mu} + \beta\sqrt{\frac{\mathbb{E}(\lambda^t)}{\mu}};$$

- **Expected Value**: This rule staffs in-house agents to meet the QoS requirements for the mean arrival rate realization;
- **Stochastic**: This rule uses our optimal in-house staffing levels, $\{x^t : t \in \mathcal{T}\}$, obtained by solving problem (2.6).

For each staffing rule, after the in-house staffing level is determined, the number of on-call and activated servers is determined (optimally) in accordance with Lemma 2.1. It should be noted, however, that because the *Worst-case* staffing rule staffs to meet the largest arrival rate realization, it never requires the use of recourse and only utilizes in-house agents.

In the first numerical example, we illustrate regimes in which our approach results in significant cost savings as compared to the alternative staffing rules. To begin, we estimate the arrival rate distributions, service rate, and patience parameter from the bank's call center data. The mean service and patience times were estimated to be 185.18 and 121.27 seconds, respectively. The remaining model parameters were selected as follows: $\psi_1 = 0.05$, $\psi_2 > 15$, $c_h = 6$, $\beta = 1.75$, and c_a and c_0 were varied between 1 and 6 such that $c_a + c_0 = 7$. The QoS thresholds were selected so that the threshold on the abandonment probability imposed the dominating constraint, cost rates were selected so that co-sourcing is optimal, and β was optimized for ψ_1, the probability of abandonment [33]. Figure 2 depicts the total expected weekly staffing cost under the staffing rules.

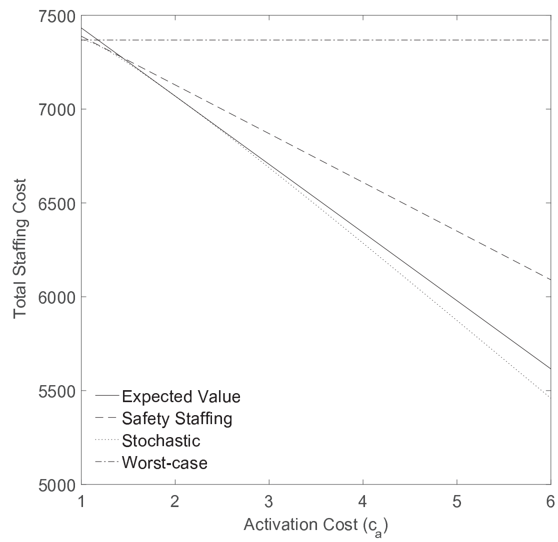


Figure 2: Comparison of one-week staffing costs for varied c_0 and c_a.

With the exception of the *Worst-case* staffing rule, the staffing costs intuitively decrease as the holding cost decreases and the activation cost increases. In the case of *Safety Staffing*, we see that the cost is relatively high across all parameter values. This high cost is due to the fact that the *Safety Staffing* rule frequently overstaffs (when compared to the *Stochastic* staffing rule) in-house agents; therefore, it is not able to fully exploit the cost savings that are possible from maintaining a lower average staffing level and utilizing recourse decisions to staff for higher-than-expected arrival rates as needed. Additionally, as guaranteed by our problem formulation, the *Stochastic* staffing rule has the minimum expected cost across all values of c_a and c_0. However, the relative cost savings moving from the *Expected Value* staffing rule to the *Stochastic* staffing rule is minimal (<2% in all cases). Generally, in the setting where the number of agents required is very small, and the arrivals are relatively stable (low-variance forecasts), our staffing rule provides only minimal benefit over the simpler *Expected Value* staffing rule. However, as the mean and variance of arrival rates increase, our optimization framework provides significant savings. Figure 3 depicts the same setting but with the arrival rate distributions scaled so that their means are twice as large and their variances are five times larger.

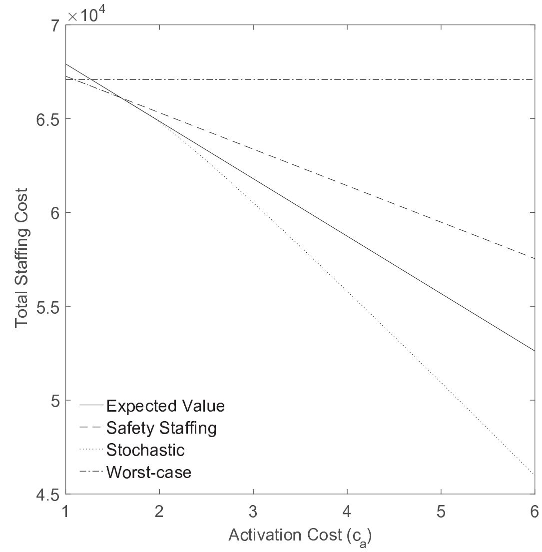
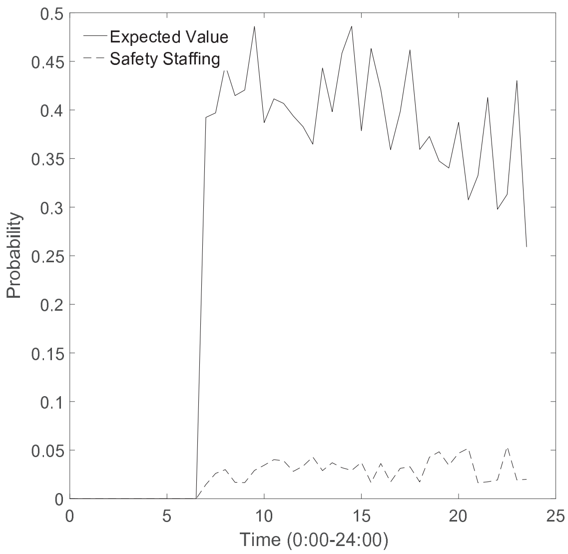


Figure 3: Comparison of one-week staffing costs for varied c_0 and c_a (scaled parameters).

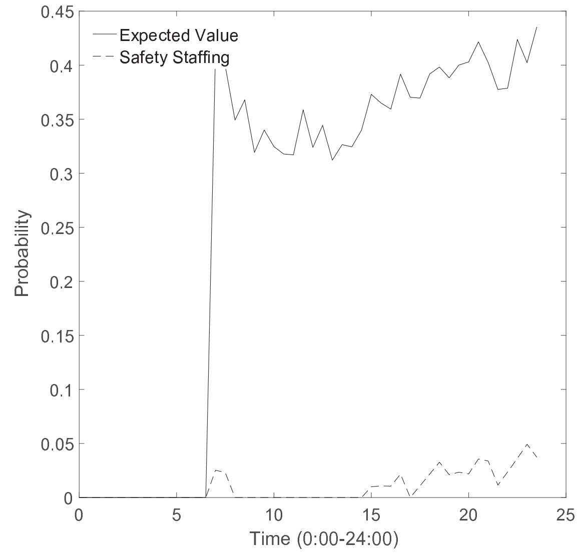
Additionally, the service rate is decreased to one fifth its initial value. As a consequence, the minimum number of required agents is generally much larger, and harder to predict (due to the higher variance). This provides a setting in which the recourse decisions are more meaningful and consequently the *Stochastic* staffing rule can provide savings of nearly 15%. It should also be noted that, as c_a increases (and c_0 correspondingly decreases), the *Stochastic* staffing rule provides increasing benefit.

In our next illustration, we investigate the value of recourse and the trade off between cost and performance. We consider again the bank call center and the scaled version. For this experiment, however, we fix the costs to $c_a = 6$ and $c_0 = 1$ and vary ψ_1 from 0.05 to 0.25 in increments of 0.01. In each case, we fix the value of β so that it corresponds with the probability of abandonment ψ_1. Figure 4 depicts the probability of violating the QoS constraints for the *Expected Value* and *Safety Staffing* rules if recourse is not utilized. The *Safety Staffing* rule provides a high probability of satisfying the QoS constraints, but the *Expected Value* rule fails to meet the QoS requirements

30-40% of the time. These graphs illustrate the benefits of recourse and flexible staffing on call center performance. Specifically, *Safety Staffing* without recourse is not cost effective, and *Expected Value* staffing without recourse performs poorly.



(a) Probability of constraint violation ($\psi_1 = 0.05$).



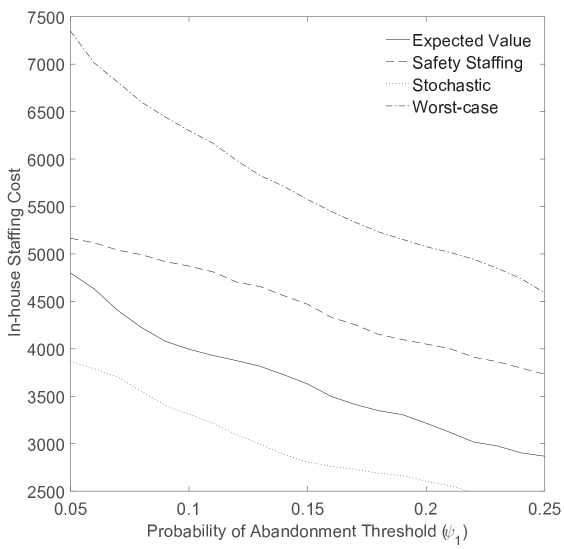
(b) Probability of constraint violation ($\psi_1 = 0.05$).

Figure 4: Probability of violating QoS constraints for unscaled (a) and scaled (b) call centers.

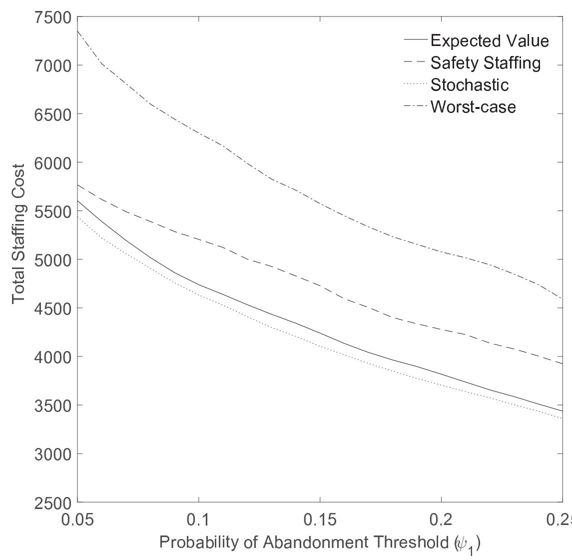
Figures 5 and 6, depict the staffing costs for the unscaled and scaled call centers respectively. In each figure, the in-house and total staffing costs are plotted side-by-side. In all cases, the *Stochastic* rule has lower in-house staffing costs (and correspondingly staffing levels) than all other methods intuitively suggesting that if recourse is available, typical staffing methods will lead to chronic overstaffing. Additionally, we see that the savings due to the *Stochastic* rule are minimal in the unscaled call center, but around 15% in the scaled case. Lastly, we note that the percent savings are reasonably uniform over all values of ψ_1.

2.4.2 Joint Staffing-Pricing Problem

We next consider the contractor's problem of setting revenue-maximizing cost rates c_o and c_a. In what follows, let $U(a,b)$ denote a continuous uniform random variable on $[a,b]$ and $N(c,d)$ denote a normally distributed random variable with mean c and variance d. For each $t \in \mathcal{T}$, we assume that $\lambda^t \sim N(\mu_t, \sigma_t^2)$, where μ_t and σ_t^2 are, respectively, the mean and variance of λ^t estimated from

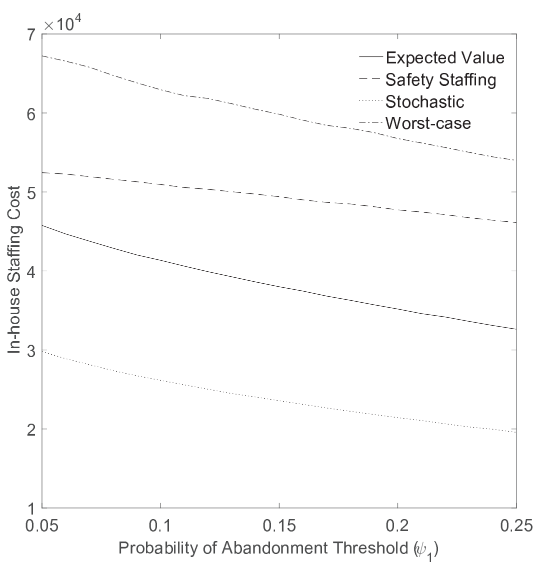


(a) In-house staffing cost.

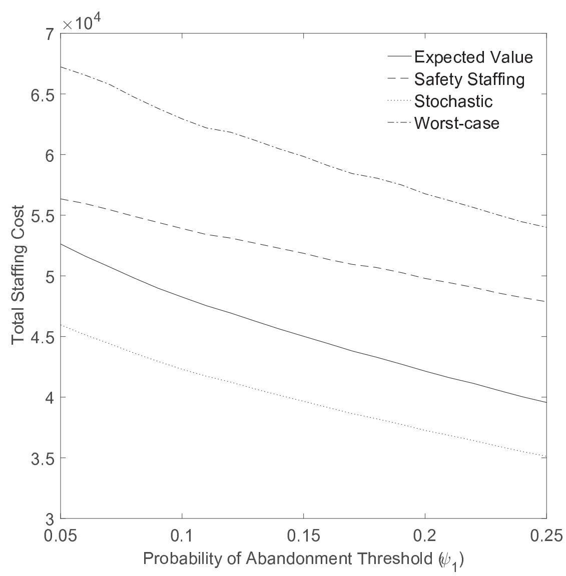


(b) Total staffing cost.

Figure 5: Comparison of one-week staffing cost for varied ψ_1 on the unscaled call center (ψ_1).



(a) In-house Staffing Cost.



(b) Total staffing cost.

Figure 6: Comparison of one-week staffing cost for varied ψ_1 on the scaled call center (ψ_1).

the call center data. The contractor's uncertainty about the call center's cost rate and arrival rate forecasts is captured by letting $c_h \sim U(4, 8)$ and $\lambda^t \sim N(\theta^t, \sigma_t^2)$, where $\theta^t \sim U(0.8\mu_t, 1.2\mu_t)$ for each time period t. The average patience and service times were likewise estimated from the data.

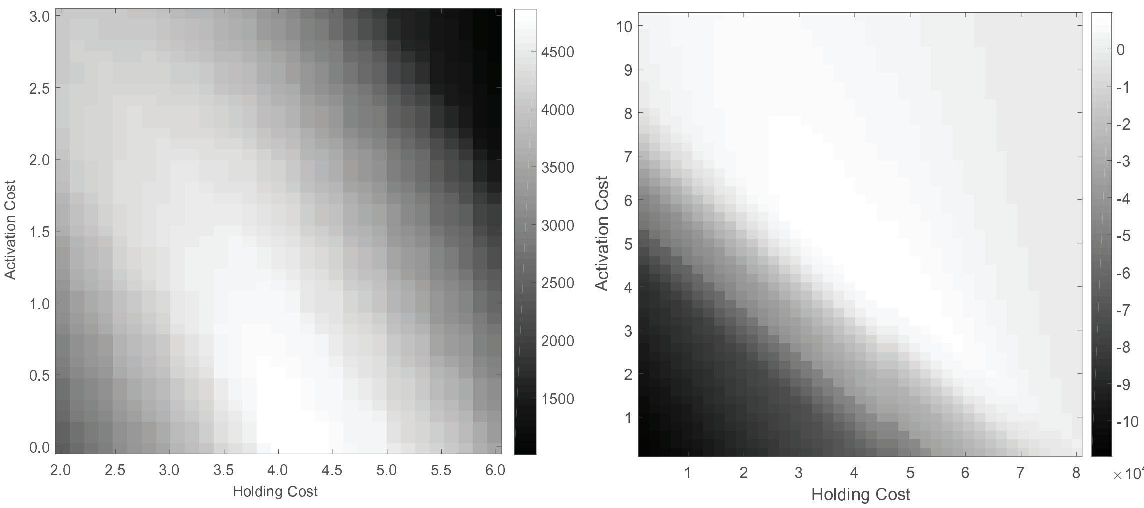
Using formulation (2.13), when costs c_0 and c_a are fixed, the lower-level problem can be solved by determining the call-centers optimal staffing levels according to Proposition 2.1. Utilizing this fact, we determine the revenue-maximizing cost rates by solving formulation (2.13) on a discrete set of values for c_o and c_a. We begin by selecting an appropriate range of values for each. By Proposition 2.1, if $c_o > \max\{c_h : c_h \in \Xi_c\}$, then the contractor's revenue must be zero; hence, the feasible range for c_o is $[0, 8]$. While a similar upper bound can be derived for c_a, it may not be tight. Based on the results of a numerical experiment, it was found that an appropriate feasible range for c_a is $[0, 10]$. Starting from 0, we increased the feasible set of values in increments of 0.01 up to the upper bound of the range so that $c_o \in \{0.00, 0.01, \ldots, 8.00\}$ and $c_a \in \{0.00, 0.01, \ldots, 5.00\}$. We sequentially sampled from the joint distribution of $(c_h, \boldsymbol{\theta})$, solving formulation (2.13) after each sample. That is, for the first problem, a single sample was drawn, and for each discrete value of c_0 and c_a the optimal call center staffing levels were determined. Then the expected cost to the call center, and the corresponding expected revenue for the contractor, were calculated. Lastly, the pair of costs that maximize the revenue over that sample were determined. This same procedure was then performed with two samples, then three samples, and so on. One benefit of this solution procedure is that it can be easily implemented to solve large instances in a parallel fashion. That is, given multiple computing resources, each machine can approximate the value of the contractor's objective function using any number of samples. Once the machines are stopped, the results can be combined without requiring any intermediate communication. In addition to the revenue-based objective function, we considered a profit-based objective function with the cost function

$$\mathbb{E}(C(c)) = 0.5b^t(c; \xi) + r\mathbb{E}_G(S^t|c),$$

where $r = 1.5$ and $S^t|c, \xi \sim \text{Bin}(y^t(c; \xi), 0.05)$.

Figure 7 depicts the contractor's estimated objective function values after drawing 1000 random samples from the distribution of $(c_h, \boldsymbol{\theta})$. By contrasting the revenue- and profit-based objective functions, it is evident that the contractor's costs play a substantial role in the optimal choice of cost rates.

Figures 8 and 9, respectively, depict the optimal holding and activation costs after each new sample is taken. It can be seen that the revenue-based objective function tends to place more value



(a) Revenue-based objective function values.

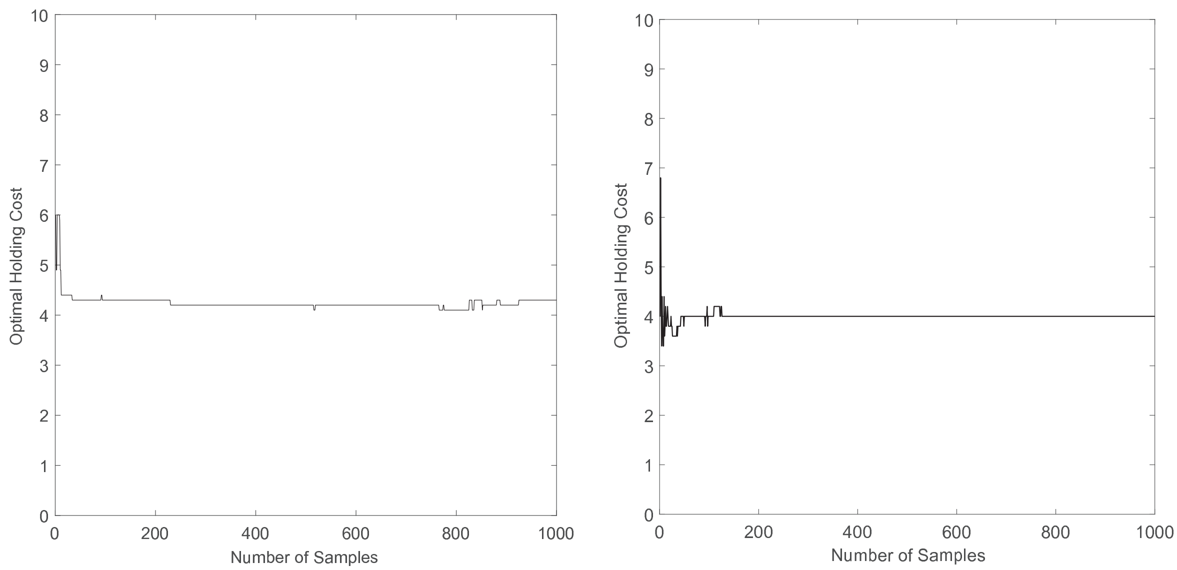
(b) Profit-based objective function values.

Figure 7: Leader's objective function values (lighter shading indicates favorable performance).

on the holding cost than the activation cost and, in this particular instance, the optimal holding and activation costs were $c_o = 4.3$ and $c_a = 0$, respectively. On the other hand, the profit-based objective yields an optimal holding cost of $c_o = 4.0$ and activation cost of $c_a = 4.2$. The profit-based optimal costs guarantee that co-sourcing is optimal for all possible realizations of c_h, whereas the revenue-based optimal costs would provide no revenue for $c_h < 4.3$. Additionally, we note that the nonzero activation cost reduces the overall revenue in the profit-based objective, but helps the contractor to balance their costs by decreasing the overall utilization of their agents and increasing the revenue whenever the agents are actually utilized. Lastly, we note that the optimal solutions converged rapidly for both objective functions, but especially for the profit-based objective with neither cost rate (c_0 or c_a) changing after 160 samples were taken.

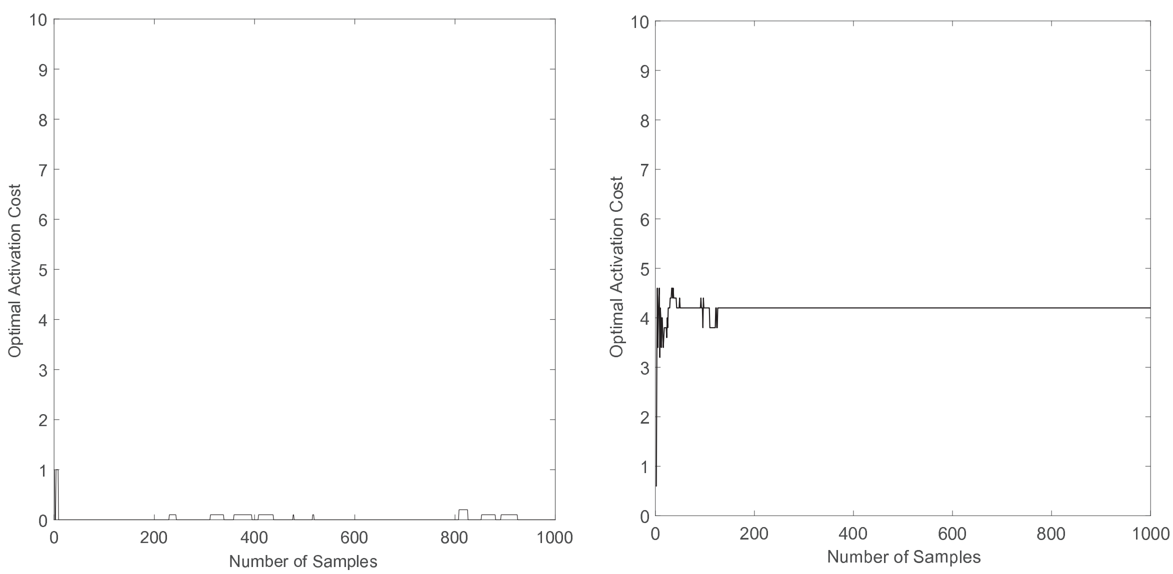
2.5 CONCLUSIONS

We have considered the joint problem of contract pricing and staffing within the context of call center co-sourcing. A leader-follower model was developed and formulated as a bilevel program in which the lower-level is a stochastic integer program with recourse. The model assumes asymmetric



(a) Optimal holding cost for revenue objective. (b) Optimal holding cost for profit objective.

Figure 8: Effect of sample size on the optimal holding cost.



(a) Optimal activation cost for revenue objective. (b) Optimal activation cost for profit objective.

Figure 9: Effect of sample size on the optimal activation cost.

information between the contractor and call center, as well as uncertainty in the call center's arrival rate forecasts. By exploiting the structure of the lower-level problem, namely its totally unimodular constraint matrix, we demonstrated that the bilevel program can be reduced to a single-level linear program with quadratic equality constraints. Additionally, we showed that the staffing problem can be solved efficiently.

Our numerical study demonstrates that the use of co-sourcing and responsive staffing can lead to a substantial cost reduction while maintaining the highest level of service quality for the chosen performance parameters. For appropriately priced contracts, co-sourced staffing proves to be less costly than square-root safety staffing while satisfying QoS requirements more reliably. From a contract pricing standpoint, we showed that determining optimal, or near-optimal, prices is computationally tractable for two-week time horizons. The contractor's uncertainty about the call center's demand and operating costs was captured through a joint probability distribution, and the pricing problem was shown to be stable, even when considering small sample sizes from that distribution.

While our work here provides a foundation for co-sourced contracts and staffing, it also suggests several avenues for future inquiry. One extension is to generalize the queueing model of the call center to include multiple customer classes and multi-skilled call center agents. Doing so necessitates the inclusion of routing considerations (e.g., high-priority classes must be handled in-house while low-priority classes can be handled either in-house or by the contractor). Furthermore, we assumed that in-house and contractor agents operate in an identical manner; however, this assumption is impractical. It will be instructive in future work to consider a multi-skill model that accounts for potential heterogeneity between these two agent pools, while incorporating service-level agreements in the contract. Other potential research directions include the consideration of more general QoS requirements (e.g., satisfying long-run performance goals rather than per-period goals), exploring the impact of alternative contractor objective functions, as well as fluid and diffusion approximations of the call center queueing model.

Finally, the model we presented here focused only a small subset of operational considerations that are crucial to call center decision making. Although forecasting, shift scheduling, and rostering are outside of the scope of this work, their inclusion would improve the realism and practicality of our model. In particular, our framework relies on arrival rate forecasts and the ability to reliably estimate the arrival rate at an arbitrary point in time. Therefore, it will be instructive to explore the impact of estimation error on staffing and pricing decisions.

3.0 MULTI-SYSTEM REPLACEMENT IN A SHARED ENVIRONMENT

3.1 INTRODUCTION

The emergence of low-cost, advanced sensing technologies and real-time condition-monitoring systems have led to increased interest in advanced maintenance planning strategies. Condition-based maintenance (CBM) techniques utilize up-to-date condition information to make well-informed maintenance decisions to achieve important objectives (e.g., minimize maintenance costs, maximize revenue, or maximize system availability). For example, modern wind turbine systems use advanced sensors to measure particle contamination levels in lubricating fluids, shaft torque, electrical discharge, vibrations, acoustic emissions, torsional vibration, and many other signals of degradation [109]. CBM provides an opportunity to exploit degradation measurements, or signals of degradation, for system prognosis and intelligent maintenance decision making that increases system uptime while reducing maintenance and operational costs.

Many large-scale systems that degrade stochastically over time are difficult to analyze in the presence of dependencies, including stochastic, economic, and structural dependencies. Stochastic dependencies are prevalent when integrated components, or systems, do not degrade (or fail) independently. For example, within a single wind farm, wind turbines are exposed to common, local weather conditions and, therefore, operate in a shared ambient environment. This exposure to similar environmental conditions may lead to dependencies in the degradation sample paths of individual wind turbines. Economic dependencies can be viewed as any financial linkages between maintenance actions, e.g., exploitation of shared downtime or rented equipment. Finally, structural dependencies result when maintenance activities performed on one component, or subsystem, require maintenance activities to occur on another component or subsystem. For instance, these types of dependencies exist when a multi-component system is enclosed in a single machine, and disassembly is required to repair or replace failed components.

In this chapter, we consider the problem of determining optimal replacement strategies for multiple, stochastically degrading systems that exhibit both stochastic and economic dependencies. Within this context, we refer to multiple machines with similar characteristics operating in close proximity to one another. The systems operate in a shared, exogenous environment that evolves randomly over time and modulates the rates of degradation of each of the systems. The systems, which are stochastically heterogenous, degrade monotonically until the cumulative level of degradation reaches a threshold, or the system is replaced. The costs of maintaining this system with replacements include a substantial fixed cost that is incurred when any maintenance action is taken. Replacements may occur preventively (before failure) or reactively (in response to a failure). Preventive replacements are less costly than reactive replacements. Finally, a downtime cost is incurred whenever the system is taken off-line to perform replacements. Our objective is obtain cost-minimizing replacement policies that account for preventive and reactive replacement decisions, as well as downtime and fixed setup costs. To this end, we formulate a continuous-time, infinite-horizon discounted Markov decision process (MDP) model, establish important properties of the cost function, reformulate the model using well-devised approximation techniques, and customize an approximate dynamic programming (ADP) algorithm to obtain high-quality policies.

For well over five decades, maintenance optimization models have emerged in the applied probability and operations research communities. Many extensive surveys highlight some of the most prevalent models for both single- and multi-component systems [5, 63, 72, 74, 85, 87, 89, 98, 102]. Most classical models consider single-component systems that operate in a static environment. Furthermore, the majority of these models employ failure-based (as opposed to condition-based) decision making. Recent emphasis on condition-based maintenance strategies has led to a stream of research on degradation-based reliability of single-component systems. These strategies seek to model degradation (or signals of degradation) as a stochastic process evolving in continuous time. Some representative models consider degradation as a Brownian motion process [21, 34, 111], while others assume that degradation is modulated by Markov or semi-Markov processes [48, 49, 50, 51]. Additionally, there exists a significant body of literature on non-condition-based policies for multi-component systems, and surveys of this literature can be found in [23, 27, 69]. Recently, Ko and Byon [52] use asymptotic theory to analytically derive the cost minimizing policy for a large-scale system with finite condition states and independent and identically degrading components.

Over the past two decades, a modest body of literature has emerged for multi-component or multi-system CBM. Marseguerra et al. [62] formulated a joint optimization model that seeks to maximize availability and net profit within a large-scale system with stochastic dependencies. Optimal thresholds, beyond which preventive maintenance should be performed, are computed by Monte Carlo simulation embedded within a genetic algorithm. Castanier et al. [18] considered a discrete-time, series system with shared setup cost for inspection or replacement. The model's decision variables are thresholds on inspection and preventive, corrective, and opportunistic maintenance. These thresholds are determined analytically for a two-component system that degrades in a static environment. Bouvard et al. [17] studied a large-scale system with shared setup cost within the context of commercial heavy vehicle maintenance. Rolling horizon procedures were developed that incorporate component information, and dynamic, analytical maintenance intervals were obtained. Tian and Liao [94] investigated policies similar to those in [18] for the general multi-component case with multiple identical and independent components. Dual threshold policies, for which once any component's failure risk exceeds the first threshold, all components with failure risk over the second threshold are replaced, were obtained numerically. Zhu et al. [113] considered a multi-component system with non-identical, independently degrading components and large shared setup costs for maintenance activities. They examined the case where the degradation paths are described by a random coefficient model and developed a nested enumeration algorithm to simultaneously obtain the optimal maintenance interval and optimal preventive maintenance control limits.

The model we present here differs from existing replacement models in that we consider large-scale systems with condition-based maintenance, stochastic dependency (through a shared modulating environment), economic dependency (through a shared setup cost), non-identical systems, degradation modulated by an exogenous continuous-time stochastic process, and continuous degradation sample paths. These model features present challenges for: (i) exact analysis of the the structure of the value function and optimal policy; and (ii) numerical computation of optimal policies for particular problem instances. Those points notwithstanding, our model is significant, as the gap between theory and practice of maintenance models is appreciable, and these realistic features represent a step towards bridging that gap [25, 85]. In the general case of multiple systems, we establish monotonicity of the value function in the cumulative degradation level for each environment state. Then, under mild conditions, it is shown that this monotonicity extends to the entire state space. Additionally, this framework allows for analysis of the special single-system case for which we show that optimal replacement policies are monotone on the entire state space. This result partially resolves the conjecture of Ulukus et al. [97] when reactive replacements do not occur immediately. Subsequently, we exploit the monotonicity results of the value function to devise a tractable, approximate model with state- and action-space transformations and a customized, basis-function approximation of the action-value function. The novel state space transformation maps abstract degradation levels to probabilities that are easily interpreted. In addition to improved interpretability, the action-value function can be more easily approximated using this new state space. We believe this type of transformation is sufficiently general to find applicability in a wide range of maintenance optimization problems. Finally, we provide a detailed computational study to demonstrate the efficacy of the approximate model in producing near-optimal policies. Specifically, we obtain policies that are nearly indistinguishable from the optimal policy in small-scale instances, as well as policies that significantly outperform heuristics in large-scale instances. To our knowledge, these techniques and algorithms are novel within the maintenance optimization literature.

The remainder of the chapter is organized as follows. In Section 3.2, we describe the environment process and its relationship to the degradation of the systems, and formulate a mathematical model of the sequential decision process. Section 3.3 discusses attributes of the value function and the optimal decision rule. In Section 3.4, we reformulate the problem using an approximate dynamic programming (ADP) model and demonstrate the usefulness of this reformulation through numerical examples in Section 3.5.

3.2 DEGRADATION MODEL AND PROBLEM FORMULATION

Consider a collection of n ($n < \infty$) systems operating in a shared, exogenous environment. The systems are assumed to begin operation in an as-good-as-new condition. The degradation rate of each system is governed by the randomly evolving environment, which occupies one of finitely many states at any point in time. Over time, each system accumulates degradation until it reaches its own fixed, deterministic threshold, above which it is considered to be failed. For the ith system, we denote this failure threshold by ξ_i ($0 < \xi_i < \infty$), $i = 1, \ldots, n$. In what follows, all random variables are defined on a common, complete probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

Let $Z(t)$ denote the state of the environment at time t and $\mathcal{Z} \equiv \{Z(t) : t \geq 0\}$ is the environment process defined on the finite state space $S = \{1, \ldots, \ell\}$. For this model, it is assumed that $\mathcal{Z}$ evolves as an S-valued, irreducible, continuous-time Markov chain (CTMC). The environment state $j \in S$ can be understood as an abstract classification of the exogenous factors that impact the degradation of each system. Specifically, whenever $Z(t) = j \in S$, the ith system degrades linearly at a rate r_i^j. Without loss of generality, it is assumed that, for each system, the degradation rates are positive, finite and monotone increasing in the environment state, i.e., $0 < r_i^1 < r_i^2 < \cdots < r_i^\ell < \infty$, for $i = 1, \ldots, n$. Denote by $X_i(t)$ the cumulative degradation of system i at time t given by

$$X_i(t) = X_i(0) + \int_0^t r_i^{Z(u)} du, \quad t \geq 0, \tag{3.1}$$

where $X_i(0)$ denotes the initial degradation level of system i. Assuming $X_i(0) = 0$, and noting that for each i

$$\int_0^t r_i^{Z(u)} du < \infty,$$

$X_i(t)$ is well defined for each $t \geq 0$. Moreover, as noted in [97], the non-negativity of the degradation rates, $\{r_i^j\}$, ensures that the sample paths of the degradation process $\mathcal{X} \equiv \{X_i(t) : t \geq 0\}$ are piecewise linear and monotone increasing in t.

Now, we introduce a Markov decision process (MDP) model to formulate the problem of optimally replacing multiple systems in a shared environment. The objective is to minimize the sum of the expected total discounted setup, replacement, and downtime costs over an infinite time horizon. This model can be viewed as an extension of the MDP model presented in [97] for a single system, with a minor variation in the costs and system dynamics. The state of the process is an $(n+1)$-dimensional vector of the form $(\boldsymbol{x}, j)$ – a realization of the joint process $(\mathcal{X}, \mathcal{Z})$ in which $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ is the vector of the systems' cumulative degradation levels, and $j \in S$ denotes the current state of the environment. Without loss of generality, we can scale the degradation rates appropriately and assume that $\xi_i = \xi$ for all i; therefore, the state space of the MDP model is the set $\Gamma \equiv [0, \xi]^n \times S$. The set of feasible actions (or action space) is $\mathcal{A} = \{0, 1\}^n$ where, for $\boldsymbol{a} = (a_1, \ldots, a_n) \in \mathcal{A}$, $a_i = 0$ corresponds to taking no action on the ith system, and $a_i = 1$ corresponds to replacing system i. System replacements can be done preventively (before failure) at a fixed cost c_p, or reactively (after a failure) at a fixed cost c_r. It is reasonable to assume that $c_p < c_r$. In addition to the system replacement costs, a fixed cost c_0 is assessed if any maintenance is performed. In practice, this cost may account for equipment costs, crew wages, or travel expenses associated with maintenance. Finally, a cost c_d for the expected per period lost productivity is assessed for each failed system that is inoperable, and we assume that all costs are non-negative and bounded from above, i.e., $c_0, c_p, c_r, c_d \in [0, \infty)$.

Because the environment process $\mathcal{Z}$ evolves as a CTMC on the finite state space S, we employ the common strategy of uniformization (cf. Puterman [77]) and convert to a discrete-time Markov chain (DTMC). Denote by $\mathcal{Q} = [q_{jk}]$ the infinitesimal generator matrix of $\mathcal{Z}$, and let q_j be the total transition rate out of state j, i.e.,

$$q_j \equiv -q_{jj} = \sum_{k:k\neq j} q_{jk}, \quad j \in S.$$

Define a uniformization rate $q \geq \max\{q_j : j \in S\}$. By uniformizing the environment process, the system is inspected at exponentially-distributed intervals of time. We denote by T_m the length of the mth inter-inspection period, which is exponentially-distributed with rate q for each $m = 1, 2, \ldots$. Denote the (random) degradation level, (random) environment state, and action taken at the mth inspection time to be $\boldsymbol{X}_m = [X^i_m]$, Z_m, and $\boldsymbol{A}_m = [A^i_m]$, respectively. Similarly, denote the (random) state at the mth period to be $\boldsymbol{S}_m = (\boldsymbol{X}_m, Z_m)$. For $j, k \in S$, define the transition probabilities of the discretized environment process,

$$p_{jk} = \mathbb{P}(Z_{m+1} = k | Z_m = j),$$

the probability that the environment transitions from state j to k during one inter-inspection period given by

$$p_{jk} = \begin{cases} q_{jk}/q, & k \neq j, \\ 1 - q_j/q, & k = j, \end{cases}$$

and let $\boldsymbol{P} = [p_{jk}]$ be the transition probability matrix of the uniformized chain. Each transition epoch of the uniformized process is then treated as an inspection time. Therefore, the minimum inspection rate is dictated by the environment process, and the inspection times are linked to the environment process through the embedded discrete-time Markov chain $\{Z_n : n \geq 0\}$.

Next, let Y^j_i be the (random) one-step accumulated degradation of system $i \in \{1, 2, \ldots, n\}$ while the environment is in state $j \in S$. The exponential length of the inter-inspection period and the constant rate of degradation jointly imply that Y^j_i is exponentially distributed with rate parameter q/r^j_i. That is,

$$F^j_i(y) := \mathbb{P}\left(Y^j_i \leq y\right) = \mathbb{P}\left(r^j_i T_m \leq y\right) = 1 - \exp\left(-qy/r^j_i\right).$$

For notational convenience, define the indicator function

$$\mathbb{I}_i(\boldsymbol{x}) = \begin{cases} 1, & x_i \geq \xi, \\ 0, & x_i < \xi. \end{cases}$$

That is, $\mathbb{I}_i(\boldsymbol{x})$ indicates whether or not system i is failed in the degradation vector $\boldsymbol{x}$. Therefore, the expected one-step cost associated with state-action pair $(\boldsymbol{s}, \boldsymbol{a})$ is

$$c(\boldsymbol{s}, \boldsymbol{a}) \equiv c_j(\boldsymbol{x}, \boldsymbol{a}) = \begin{cases} c_0 + \sum_{i=1}^{n} \left[a_i(1 - \mathbb{I}_i(\boldsymbol{x}))c_p + a_i\mathbb{I}_i(\boldsymbol{x})c_r + (1 - a_i)\mathbb{I}_i(\boldsymbol{x})c_d\right], & \boldsymbol{a} \neq \boldsymbol{0}, \\ c_d \sum_{i=1}^{n} \mathbb{I}_i(\boldsymbol{x}), & \boldsymbol{a} = \boldsymbol{0}. \end{cases} \tag{3.2}$$

For many applications, the downtime cost c_d may depend on the length of the inter-inspection period. Within our framework, this dependence is reflected by the fact that c_d is a function of the uniformization constant q; however, for notational convenience, we suppress this dependence on q, as it is reasonable to assume that $c_d(q)$ is monotone nonincreasing in q. We discuss the implications of this assumption at the end of Section 3.3. Inspections are assumed to be costless and instantaneous. If the replacement action is taken, the degradation of that system ceases, and the replacement occurs during the current period and ends just prior to the start of the next period. These assumptions ensure that failures do not occur once preventive replacement is decided upon and that the system begins the next period in as-good-as-new condition. All one-step costs are incurred at the beginning of the period and discounted at rate α $(0 < \alpha < 1)$, where

$$\alpha = \frac{q}{\theta + q}$$

for some continuous-time discount rate θ $(0 < \theta < \infty)$. The objective is to minimize the expected total discounted setup, replacement and downtime costs over an infinite planning horizon. The optimal expected total discounted cost, starting in state $\boldsymbol{s} = (\boldsymbol{x}, j)$ and denoted by $V_j(\boldsymbol{x})$, is given as a solution to the Bellman optimality equations

$$V_j(\boldsymbol{x}) = \min_{\boldsymbol{a} \in \mathcal{A}} \left\{ c_j(\boldsymbol{x}, \boldsymbol{a}) + \alpha \sum_{k=1}^{\ell} \left(\int_0^\infty V_k\left(\boldsymbol{x}'\right) q e^{-qt} dt \right) p_{jk} \right\}, \quad j \in S, \tag{3.3}$$

where $\boldsymbol{x}' = [x_i']$ with

$$x_i' = \begin{cases} \min\{\xi,\, x_i + t\, r_i^j\}, & a_i = 0, \\ 0, & a_i = 1. \end{cases}$$

It should be noted that this formulation does not force reactive replacements immediately upon failure. This added flexibility allows the decision maker to pool replacements when appropriate, thereby sharing the cost (c_0) amongst multiple repairs.

In order to solve this problem by conventional MDP solution techniques, the state space must first be discretized so that numerical methods can be applied (e.g., value iteration or policy iteration [11, 42]). It is clear from this formulation that as n, the number of systems, grows large, the problem will suffer from the curse of dimensionality. This occurs because the state space is an n-dimensional hypercube, and the number of permissible actions is 2^n. For example, with 50 systems and one environment state, the number of feasible actions is approximately 10^{15}, and discretizing the state space into $1,000$ states per component yields 10^{100} states. For these reasons, we propose handling this problem by approximate dynamic programming (ADP) methods. In particular, we customize a state-action-reward-state-action (SARSA) algorithm with eligibility traces and basis function approximation [76, 82]. Before doing so, we first provide some useful structural results that help characterize an optimal policy.

3.3 STRUCTURAL RESULTS

In this section, we characterize the attributes of the cost function and optimal policy of the MDP model presented in Section 3.2. We first consider the general, multiple system case before presenting results for the special case of a single system.

3.3.1 Results for the General Case

The first result for n ($n > 1$) systems establishes the existence of a stationary optimal policy, as well as the convergence of the standard value iteration algorithm.

Lemma 3.1 *There exists an optimal, non-randomized stationary replacement policy, and the value iteration algorithm converges to the optimal value.*

Proof. First note that the state space, $\Gamma = [0, \xi]^n \times S$, is Borel-measurable and the action space, $\mathcal{A} = \{0, 1\}^n$, is finite. Additionally, the immediate costs are strictly positive and bounded, and the problem is discounted. Therefore, the result follows immediately from Corollary 9.17.1 of Bertsekas and Shreve [13]. ∎

For the state space Γ, define the binary relation $(\leq)$ as the standard component-wise inequality. That is, for any two vectors $\boldsymbol{s}, \boldsymbol{s}' \in \Gamma \subset \mathbb{R}^{n+1}$, $\boldsymbol{s} \leq \boldsymbol{s}'$ if and only if $s_i \leq s_i'$ for each $i = 1, \ldots, n+1$. It can be verified that $\leq$ is a partial order on Γ and $\mathcal{A}$; therefore, $(\Gamma, \leq)$ and $(\mathcal{A}, \leq)$ are partially-ordered sets (posets). Additionally, as formalized in Lemma 3.2, $(\Gamma, \leq)$ is a lattice (see Birkhoff [15] for a detailed discussion of lattices).

Lemma 3.2 *The partially ordered sets $(\Gamma, \leq)$ and $(\mathcal{A}, \leq)$ are lattices.*

Proof. For any pair of states $\boldsymbol{s}, \boldsymbol{s}' \in \Gamma$, we have that

$$\bar{\boldsymbol{s}} := (\max\{s_1, s_1'\}, \ldots, \max\{s_{n+1}, s_{n+1}'\}) \in \Gamma,$$

and is an upper bound for $\boldsymbol{s}$ and $\boldsymbol{s}'$. Letting $\bar{\boldsymbol{s}}' = (\bar{s}_1', \ldots, \bar{s}_{n+1}')$ be an arbitrary upper bound on $\boldsymbol{s}, \boldsymbol{s}'$, it is seen that $\bar{s}_i' \geq s_i$ and $\bar{s}_i' \geq s_i'$. Hence, $\bar{s}_i' \geq \max\{s_i, s_i'\}$ for each $i = 1, \ldots, n+1$. Therefore, $\bar{\boldsymbol{s}}' \geq \bar{\boldsymbol{s}}$, which implies $\boldsymbol{s} \vee \boldsymbol{s}' = \bar{\boldsymbol{s}} \in \Gamma$. Similarly, by component-wise minimization, $\boldsymbol{s} \wedge \boldsymbol{s}' \in \Gamma$. The proof that $(\mathcal{A}, \leq)$ is a lattice proceeds in an identical fashion. ∎

The notion of submodularity plays an important role in deriving our main results; therefore, we formally define it next.

Definition 3.1 *(Submodularity). A function $f : A \times B \to \mathbb{R}$ is said to be submodular on $A \times B$ if*

$$f(a_2, b_2) - f(a_2, b_1) \leq f(a_1, b_2) - f(a_1, b_1) \tag{3.4}$$

for any $a_1, a_2 \in A$ and $b_1, b_2 \in B$ such that $a_1 \leq a_2$ and $b_1 \leq b_2$.

Submodularity is useful for characterizing structural properties in optimization problems. Here, we restate an important result due to Topkis [95] as Theorem 3.1.

Theorem 3.1 *(Topkis [95]). Let $f : A \times B \to \mathbb{R}$ be a submodular function on $A \times B$, and let $(A, \leq)$ and $(B, \leq)$ be lattices. Then $g^*(b) = \max\{a' \in \operatorname{argmin}_{a \in A} f(a, b)\}$ is nondecreasing in b.*

For the results that follow, define the Q-function,

$$Q(\boldsymbol{s}, \boldsymbol{a}) \equiv Q_j(\boldsymbol{x}, \boldsymbol{a}) := c(\boldsymbol{s}, \boldsymbol{a}) + \alpha \mathbb{E}\left(V(\boldsymbol{S}_1) | \boldsymbol{S}_0 = \boldsymbol{s}, \boldsymbol{A}_0 = \boldsymbol{a}\right). \tag{3.5}$$

By Theorem 3.1, if it can be shown that Q is submodular on $\Gamma \times \mathcal{A}$, then the optimal decision rule,

$$d^*(\boldsymbol{s}) := \max\left\{\boldsymbol{a}' \in \operatorname{argmin}_{\boldsymbol{a}\in\mathcal{A}} Q(\boldsymbol{s}, \boldsymbol{a})\right\}, \tag{3.6}$$

is monotone in Γ. This result is formalized under some specific conditions in Theorem 3.4. However, before proceeding to the main results, Lemma 3.3 provides basic insights into the structure of the expected one-step cost function $c(\boldsymbol{s}, \boldsymbol{a})$; it asserts that the one-step cost does not decrease as the cumulative degradation level increases. Additionally, under a particular cost structure, the one-step cost function is submodular.

Lemma 3.3 *The expected one-step cost function $c(\boldsymbol{s}, \boldsymbol{a})$ is*

(a) monotone nondecreasing in $\boldsymbol{s}$, and

(b) submodular on $\Gamma \times \mathcal{A}$, if $c_r - c_p \leq c_d$.

Proof. We note that if $\boldsymbol{a} \neq \boldsymbol{0}$, then $c_j(\boldsymbol{x}, \boldsymbol{a})$ can be expressed as

$$c_j(\boldsymbol{x}, \boldsymbol{a}) = c_0 + \sum_{i=1}^{n} [a_i c_p + \mathbb{I}_i(\boldsymbol{x}) a_i (c_r - c_p) + \mathbb{I}_i(\boldsymbol{x})(1 - a_i) c_d].$$

Because it is assumed that $c_r > c_p$, Lemma 3.3(a) follows immediately. For Lemma 3.3(b), we seek to show that for $\boldsymbol{a} < \boldsymbol{a}'$ and $\boldsymbol{x} < \boldsymbol{x}'$,

$$c_j(\boldsymbol{x}', \boldsymbol{a}') + c_j(\boldsymbol{x}, \boldsymbol{a}) \leq c_j(\boldsymbol{x}', \boldsymbol{a}) + c_j(\boldsymbol{x}, \boldsymbol{a}'). \tag{3.7}$$

We begin with the case where $\boldsymbol{a} = \boldsymbol{0} < \boldsymbol{a}'$. The left-hand side (l.h.s.) of (3.7) is given by

$$\begin{aligned} c_j(\boldsymbol{x}', \boldsymbol{a}') + c_j(\boldsymbol{x}, \boldsymbol{a}) &= c_0 + \sum_{i=1}^{n} \left[\boldsymbol{a}'_i c_p + \boldsymbol{a}'_i \mathbb{I}_i(\boldsymbol{x}')(c_r - c_p) + \mathbb{I}_i(\boldsymbol{x}')(1 - \boldsymbol{a}'_i) c_d\right] + c_d \sum_{i=1}^{n} \mathbb{I}_i(\boldsymbol{x}) \\ &= c_0 + \sum_{i=1}^{n} \left[\boldsymbol{a}'_i c_p + \boldsymbol{a}'_i \mathbb{I}_i(\boldsymbol{x}')(c_r - c_p) + \left(\mathbb{I}_i(\boldsymbol{x}')(1 - \boldsymbol{a}'_i) + \mathbb{I}_i(\boldsymbol{x})\right) c_d\right]. \end{aligned} \tag{3.8}$$

Similarly, the right-hand side (r.h.s.) of (3.7) is given by

$$c_j(\boldsymbol{x}', \boldsymbol{a}) + c_j(\boldsymbol{x}, \boldsymbol{a}') = c_0 + \sum_{i=1}^{n} \left[\boldsymbol{a}'_i c_p + \boldsymbol{a}'_i \mathbb{I}_i(\boldsymbol{x})(c_r - c_p) + \left(\mathbb{I}_i(\boldsymbol{x})(1 - \boldsymbol{a}'_i) + \mathbb{I}_i(\boldsymbol{x}')\right) c_d\right]. \tag{3.9}$$

After some algebraic manipulation, it can be shown that the quantity in (3.8) is no greater than that in (3.9) if, and only if,

$$c_d \sum_{i=1}^{n} \boldsymbol{a}'_i (\mathbb{I}_i(\boldsymbol{x}') - \mathbb{I}_i(\boldsymbol{x})) \geq (c_r - c_p) \sum_{i=1}^{n} \boldsymbol{a}'_i (\mathbb{I}_i(\boldsymbol{x}') - \mathbb{I}_i(\boldsymbol{x})).$$

By supposition, $c_r - c_p \le c_d$; therefore, inequality (3.7) holds. Similarly, in case $\mathbf{0} < \boldsymbol{a} < \boldsymbol{a}'$, it can be shown that inequality (3.7) holds if, and only if,

$$c_d \sum_{i=1}^{n} \left[\mathbb{I}_i(\boldsymbol{x})(\boldsymbol{a}_i - \boldsymbol{a}_i') + \mathbb{I}_i(\boldsymbol{x}')(\boldsymbol{a}_i' - \boldsymbol{a}_i) \right] \ge (c_r - c_p) \sum_{i=1}^{n} \left[\mathbb{I}_i(\boldsymbol{x})(\boldsymbol{a}_i - \boldsymbol{a}_i') + \mathbb{I}_i(\boldsymbol{x}')(\boldsymbol{a}_i' - \boldsymbol{a}_i) \right].$$

For each $i = 1, \ldots, n$,

$$\mathbb{I}_i(\boldsymbol{x})(\boldsymbol{a}_i - \boldsymbol{a}_i') + \mathbb{I}_i(\boldsymbol{x}')(\boldsymbol{a}_i' - \boldsymbol{a}_i) = (\mathbb{I}_i(\boldsymbol{x}') - \mathbb{I}_i(\boldsymbol{x}))(\boldsymbol{a}_i' - \boldsymbol{a}_i) \ge 0.$$

Therefore, by the condition, $c_r - c_p \le c_d$, inequality (3.7) holds. ∎

We pause here to note that the condition of Lemma 3.3(b), $c_r - c_p \le c_d$, may not hold in practice. In particular, for a large uniformization rate q, it is unlikely that this inequality is valid. Fortunately, a useful property still emerges, namely that the value function is monotone nondecreasing in the cumulative degradation level, even if the condition is relaxed. We formalize this result in Theorem 3.2.

Theorem 3.2 *For each $j \in S$, the value function $V_j(\boldsymbol{x})$ is monotone nondecreasing in the degradation level $\boldsymbol{x} \in \mathcal{X}$.*

Proof. For $(\boldsymbol{x}, j) \in \Gamma$, denote the mth iterate of the value iteration algorithm by $v_j^m(\boldsymbol{x}) \equiv v^m(\boldsymbol{x}, j)$. We prove the theorem by induction on m. Take $v_j^0(\boldsymbol{x}) = 0$ for all $(\boldsymbol{x}, j) \in \Gamma$. Therefore,

$$\begin{aligned} v_j^1(\boldsymbol{x}) &= \min_{\boldsymbol{a} \in \mathcal{A}} \left\{ c_j(\boldsymbol{x}, \boldsymbol{a}) + \alpha \mathbb{E}\left(v^0(\boldsymbol{X}_1, Z_1) | \boldsymbol{X}_0 = \boldsymbol{x}, Z_0 = j, \boldsymbol{A}_0 = \boldsymbol{a} \right) \right\} \\ &= \min_{\boldsymbol{a} \in \mathcal{A}} \left\{ c_j(\boldsymbol{x}, \boldsymbol{a}) + 0 \right\} \\ &= c_d \sum_{i=1}^{n} \mathbb{I}_i(\boldsymbol{x}), \end{aligned}$$

which is monotone nondecreasing in $\boldsymbol{x}$. For the induction hypothesis, assume $v_j^m(\boldsymbol{x})$ is monotone nondecreasing in $\boldsymbol{x} \in \mathcal{X}$ for each $j \in S$. Let $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{X}$ such that $\boldsymbol{x}_1 \le \boldsymbol{x}_2$, then for each $\boldsymbol{a} \in \mathcal{A}$

$$\begin{aligned} v_j^{m+1}(\boldsymbol{x}_1, \boldsymbol{a}) &= c_j(\boldsymbol{x}_1, \boldsymbol{a}) + \alpha \sum_{k=1}^{\ell} \left(\int_0^\infty v_k^m(\boldsymbol{x}_1') q e^{-qt} dt \right) p_{jk} \\ &\le c_j(\boldsymbol{x}_2, \boldsymbol{a}) + \alpha \sum_{k=1}^{\ell} \left(\int_0^\infty v_k^m(\boldsymbol{x}_2') q e^{-qt} dt \right) p_{jk} \\ &= v_j^{m+1}(\boldsymbol{x}_2, \boldsymbol{a}), \end{aligned} \tag{3.10}$$

where the inequality holds due to Lemma 3.3(a) and by noting that, for a fixed $\boldsymbol{a}$, $\boldsymbol{x}_1' \le \boldsymbol{x}_2'$. Minimizing both sides of (3.10) over $\boldsymbol{a} \in \mathcal{A}$ shows that $v_j^{m+1}(\boldsymbol{x}_1) \le v_j^{m+1}(\boldsymbol{x}_2)$. Finally, Lemma 3.1 implies that $v_j^m(\boldsymbol{x}) \to V_j(\boldsymbol{x})$, as $m \to \infty$, and the proof is complete. ∎

Theorem 3.2 asserts that the expected cost-to-go increases as the cumulative degradation level of the system increases. Informally, this corresponds to the intuitive idea that starting in a 'bad' state is, in fact, more costly than starting in a 'good' state. Under some reasonable conditions we can prove stronger structural results.

For the remainder of this section, we impose mild conditions on the uniformized environment process and the degradation rate vectors $\boldsymbol{r}_i$. For completeness, we review the notion of the increasing failure rate (IFR) property of a transition probability matrix.

Definition 3.2 *Let $\boldsymbol{P} = [p_{ij}]$ be the transition probability matrix of a DTMC with state space $S = \{1, \ldots, \ell\}$. Then $\boldsymbol{P}$ is said to be IFR if*

$$\eta_m(i) := \sum_{j=m}^{\ell} p_{ij} \tag{3.11}$$

is nondecreasing in $i \in S$ for all $m = 1, \ldots, \ell$.

Now, let us define the tail distribution function

$$q(\boldsymbol{s}'|\boldsymbol{s}, \boldsymbol{a}) := \mathbb{P}\left(\boldsymbol{S}_{m+1} \geq \boldsymbol{s}'|\boldsymbol{S}_m = \boldsymbol{s}, \boldsymbol{A}_m = \boldsymbol{a}\right). \tag{3.12}$$

The quantity in (3.12) represents the probability that the MDP transitions to a state at least as large as $\boldsymbol{s}'$, given a starting state $\boldsymbol{s}$ and action $\boldsymbol{a}$. For notational convenience, also define the following sets:

$$\mathcal{I}(\boldsymbol{x}') := \{i \in \{1, \ldots, n\} : x_i' = 0\}, \tag{3.13}$$

$$\mathcal{J}(\boldsymbol{a}) := \{i \in \{1, \ldots, n\} : a_i = 1\}, \tag{3.14}$$

$$\mathcal{K}(\boldsymbol{x}, \boldsymbol{x}') := \{i \in \{1, \ldots, n\} : x_i < x_i'\}. \tag{3.15}$$

These sets are used to describe the form of $q(\cdot|\cdot)$, namely,

$$q((\boldsymbol{x}', k')|(\boldsymbol{x}, k), \boldsymbol{a}) = \begin{cases} \exp\left[-q \cdot \max\limits_{v \in \mathcal{K}(\boldsymbol{x}, \boldsymbol{x}')} \left\{\dfrac{x_v' - x_v}{r_v^k}\right\}\right] \displaystyle\sum_{m=k'}^{\ell} p_{km}, & \mathcal{J} \subseteq \mathcal{I}, \mathcal{K} \neq \emptyset, \\ \displaystyle\sum_{m=k'}^{\ell} p_{km}, & \mathcal{J} \subseteq \mathcal{I}, \mathcal{K} = \emptyset, \\ 0, & \mathcal{J} \nsubseteq \mathcal{I}. \end{cases} \tag{3.16}$$

The tail distribution is useful for establishing monotonicity of the value function (3.3) and the corresponding optimal decisions. We begin by describing some useful properties of the tail distribution in Lemma 3.4.

Lemma 3.4 *If $\boldsymbol{P}$ is IFR, and the degradation rates r_i^j are monotone nondecreasing in $j \in S$ for each $i = 1, \ldots, n$, then the tail distribution function $q(\boldsymbol{s}'|\boldsymbol{s}, \boldsymbol{a})$ is*

(a) monotone nondecreasing in $\boldsymbol{s} \in \Gamma$, and

(b) submodular on $\Gamma \times \mathcal{A}$.

Proof. Let $\tilde{\boldsymbol{s}} = (\tilde{\boldsymbol{x}}, \tilde{j}) \in \Gamma$, $\boldsymbol{a} \in A$, and $\boldsymbol{s} = (\boldsymbol{x}, j) \leq \boldsymbol{s}' = (\boldsymbol{x}', j') \in \Gamma$. In the case where $\mathcal{J} \nsubseteq \mathcal{I}$, we have $q(\tilde{\boldsymbol{s}}|\boldsymbol{s}, \boldsymbol{a}) = q(\tilde{\boldsymbol{s}}|\boldsymbol{s}', \boldsymbol{a}) = 0$. Because $\mathcal{K}(\boldsymbol{x}, \boldsymbol{x}')$ is decreasing in $\boldsymbol{x}$, there are three subcases for $\mathcal{J} \subseteq \mathcal{I}$: (i) $\mathcal{K}(\boldsymbol{x}, \tilde{\boldsymbol{x}}) \neq \emptyset$ and $\mathcal{K}(\boldsymbol{x}', \tilde{\boldsymbol{x}}) = \emptyset$, (ii) $\mathcal{K}(\boldsymbol{x}, \tilde{\boldsymbol{x}}) \neq \emptyset$ and $\mathcal{K}(\boldsymbol{x}', \tilde{\boldsymbol{x}}) \neq \emptyset$, and (iii) $\mathcal{K}(\boldsymbol{x}, \tilde{\boldsymbol{x}}) = \emptyset$ and $\mathcal{K}(\boldsymbol{x}', \tilde{\boldsymbol{x}}) = \emptyset$. For subcase (i),

$$q(\tilde{\boldsymbol{s}}|\boldsymbol{s}', \boldsymbol{a}) = \sum_{m=\tilde{k}}^{\ell} p_{j'm} \geq \exp\left(-q \cdot \max_{v \in \mathcal{K}(\boldsymbol{x}, \tilde{\boldsymbol{x}})} \left\{\frac{\tilde{x}_v - x_v}{r_v^j}\right\}\right) \sum_{m=\tilde{k}}^{\ell} p_{jm} = q(\tilde{\boldsymbol{s}}|\boldsymbol{s}, \boldsymbol{a}).$$

For subcase (ii),

$$q(\tilde{\boldsymbol{s}}|\boldsymbol{s}', \boldsymbol{a}) = \sum_{m=\tilde{k}}^{\ell} p_{j'm} \geq \sum_{m=\tilde{k}}^{\ell} p_{jm} = q(\tilde{\boldsymbol{s}}|\boldsymbol{s}, \boldsymbol{a}),$$

because $\boldsymbol{P}$ is IFR. Lastly, for subcase (iii), note by monotonicity of $\boldsymbol{r_i}$, for each $i \in \{1, 2, \ldots, n\}$,

$$\frac{\tilde{x}_v - x'_v}{r_v^{j'}} \leq \frac{\tilde{x}_v - x_v}{r_v^j}.$$

Additionally, by $\mathcal{K}(\boldsymbol{x}', \tilde{\boldsymbol{x}}) \subseteq \mathcal{K}(\boldsymbol{x}, \tilde{\boldsymbol{x}})$,

$$\max_{v \in \mathcal{K}(\boldsymbol{x}', \tilde{\boldsymbol{x}})} \left\{\frac{\tilde{x}_v - x'_v}{r_v^{j'}}\right\} \leq \max_{v \in \mathcal{K}(\boldsymbol{x}, \tilde{\boldsymbol{x}})} \left\{\frac{\tilde{x}_v - x_v}{r_v^j}\right\}.$$

Therefore, by the monotonicity of $\exp(\cdot)$ and the increasing failure rate of $\boldsymbol{P}$ we have

$$\begin{aligned} q(\tilde{\boldsymbol{s}}|\boldsymbol{s}', \boldsymbol{a}) &= \exp\left(-q \cdot \max_{v \in \mathcal{K}(\boldsymbol{x}', \tilde{\boldsymbol{x}})} \left\{\frac{\tilde{x}_v - x'_v}{r_v^{j'}}\right\}\right) \sum_{m=\tilde{k}}^{\ell} p_{j'm} \\ &\geq \exp\left(-q \cdot \max_{v \in \mathcal{K}(\boldsymbol{x}, \tilde{\boldsymbol{x}})} \left\{\frac{\tilde{x}_v - x_v}{r_v^j}\right\}\right) \sum_{m=\tilde{k}}^{\ell} p_{jm} = q(\tilde{\boldsymbol{s}}|\boldsymbol{s}, \boldsymbol{a}). \end{aligned}$$

Thus the tail function is monotone in $\boldsymbol{s}$. For Lemma 3.4(b), we seek to show

$$q(\tilde{\boldsymbol{s}}|\boldsymbol{s}', \boldsymbol{a}') + q(\tilde{\boldsymbol{s}}|\boldsymbol{s}, \boldsymbol{a}) \leq q(\tilde{\boldsymbol{s}}|\boldsymbol{s}, \boldsymbol{a}') + q(\tilde{\boldsymbol{s}}|\boldsymbol{s}', \boldsymbol{a}) \tag{3.17}$$

for $\boldsymbol{a} \leq \boldsymbol{a}' \in A$. Noting $\mathcal{J}(\boldsymbol{a}) \subseteq \mathcal{J}(\boldsymbol{a}')$, we have three cases to consider: (i) $\mathcal{J}(\boldsymbol{a}), \mathcal{J}(\boldsymbol{a}') \nsubseteq \mathcal{I}$, (ii) $\mathcal{J}(\boldsymbol{a}) \subseteq \mathcal{I}$, $\mathcal{J}(\boldsymbol{a}') \nsubseteq \mathcal{I}$, and (iii) $\mathcal{J}(\boldsymbol{a}), \mathcal{J}(\boldsymbol{a}') \subseteq \mathcal{I}$. For case (i), it is clear that both sides of (3.17) equal 0. For case (ii), $q(\tilde{\boldsymbol{s}}|\boldsymbol{s}, \boldsymbol{a}') = q(\tilde{\boldsymbol{s}}|\boldsymbol{s}', \boldsymbol{a}') = 0$, reducing inequality (3.17) to

$$q(\tilde{\boldsymbol{s}}|\boldsymbol{s}, \boldsymbol{a}) \leq q(\tilde{\boldsymbol{s}}|\boldsymbol{s}', \boldsymbol{a}),$$

which holds by Lemma 3.4(a). For case (iii), we note $q(\tilde{\boldsymbol{s}}|\boldsymbol{s},\boldsymbol{a}) = q(\tilde{\boldsymbol{s}}|\boldsymbol{s},\boldsymbol{a}')$ and $q(\tilde{\boldsymbol{s}}|\boldsymbol{s}',\boldsymbol{a}) = q(\tilde{\boldsymbol{s}}|\boldsymbol{s}',\boldsymbol{a}')$. Hence, inequality (3.17) holds in equality. Thus, the tail function is submodular. ■

Next, we review some concepts from stochastic ordering that are needed for the remaining results. The first is that of upper orthant ordering of random vectors. For two n-dimensional, random vectors $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$, we say that $\boldsymbol{Y}_1$ is less than $\boldsymbol{Y}_2$ in the upper orthant order, denoted $\boldsymbol{Y}_1 \leq_{uo} \boldsymbol{Y}_2$, if for all $y \in \mathbb{R}^n$,

$$\mathbb{P}(\boldsymbol{Y}_1 \geq y) \leq \mathbb{P}(\boldsymbol{Y}_2 \geq y).$$

A set U is said to be an upper set if $u_2 \in U$ whenever $u_2 \geq u_1$ and $u_1 \in U$. We say that $\boldsymbol{Y}_1$ is less than $\boldsymbol{Y}_2$ in the usual stochastic order, denoted $\boldsymbol{Y}_1 \leq_{st} \boldsymbol{Y}_2$, if for all upper sets $U \subseteq \mathbb{R}^n$,

$$\mathbb{P}(\boldsymbol{Y}_1 \in U) \leq \mathbb{P}(\boldsymbol{Y}_2 \in U).$$

It is important to note that monotonicity of the tail function, established in Lemma 3.4(a), is equivalent to the condition

$$[\boldsymbol{S}_{m+1}|\boldsymbol{S}_m = \boldsymbol{s}, \boldsymbol{A}_m = \boldsymbol{a}] \leq_{uo} [\boldsymbol{S}_{m+1}|\boldsymbol{S}_m = \boldsymbol{s}', \boldsymbol{A}_m = \boldsymbol{a}], \quad \text{for any } \boldsymbol{s} \leq \boldsymbol{s}'.$$

If $n = 1$, the upper orthant and the usual stochastic order are equivalent. If $n > 1$, the usual stochastic order is stronger than the upper orthant order. These concepts are useful in that they allow us to extend part Lemma 3.4(a) to the usual stochastic order as seen in Lemma 3.5.

Lemma 3.5 *Under the conditions stated in Lemma 3.4, for any $a \in \mathcal{A}$ and $\boldsymbol{s} \leq \boldsymbol{s}' \in \Gamma$,*

$$[\boldsymbol{S}_{m+1}|\boldsymbol{S}_m = \boldsymbol{s}, \boldsymbol{A}_m = \boldsymbol{a}] \leq_{st} [\boldsymbol{S}_{m+1}|\boldsymbol{S}_m = \boldsymbol{s}', \boldsymbol{A}_m = \boldsymbol{a}].$$

Proof. Let $\boldsymbol{S}_{m+1} = (\boldsymbol{X}_{m+1}, \boldsymbol{Z}_{m+1})$. By independence of $\boldsymbol{X}_{m+1}$ and $\boldsymbol{Z}_{m+1}$ it suffices to show:

$$[\boldsymbol{X}_{m+1}|\boldsymbol{S}_m = \boldsymbol{s}, \boldsymbol{A}_m = \boldsymbol{a}] \leq_{st} [\boldsymbol{X}_{m+1}|\boldsymbol{S}_m = \boldsymbol{s}', \boldsymbol{A}_m = \boldsymbol{a}], \tag{3.18}$$

and

$$[\boldsymbol{Z}_{m+1}|\boldsymbol{S}_m = \boldsymbol{s}, \boldsymbol{A}_m = \boldsymbol{a}] \leq_{st} [\boldsymbol{Z}_{m+1}|\boldsymbol{S}_m = \boldsymbol{s}', \boldsymbol{A}_m = \boldsymbol{a}]. \tag{3.19}$$

Inequality (3.19) follows from the IFR assumption of $\boldsymbol{P}$. For $t \geq 0$, define the functions $\psi_1(t)$ and $\psi_2(t)$ by

$$\psi_1(t) = [(1-a_1)(x_1 + r_1^j t), (1-a_2)(x_2 + r_2^j t), \ldots, (1-a_n)(x_n + r_n^j t)],$$

and

$$\psi_2(t) = [(1-a_1)(x_1' + r_1^{j'} t), (1-a_2)(x_2' + r_2^{j'} t), \ldots, (1-a_n)(x_n' + r_n^{j'} t)].$$

By the monotonicity of the degradation rates r_i^j, and the fact that, $\boldsymbol{x} \leq \boldsymbol{x}'$, we see that $\psi_1(t) \leq \psi_2(t)$ for all $t \geq 0$. Letting $T \sim \text{Exp}(q)$, we have

$$\psi_1(T) \overset{\mathrm{d}}{=} [\boldsymbol{X}_{m+1} | \boldsymbol{S}_m = \boldsymbol{s}, \boldsymbol{A}_m = \boldsymbol{a}] \quad \text{and} \quad \psi_2(T) \overset{\mathrm{d}}{=} [\boldsymbol{X}_{m+1} | \boldsymbol{S}_m = \boldsymbol{s}', \boldsymbol{A}_m = \boldsymbol{a}],$$

where $\overset{\mathrm{d}}{=}$ denotes equality in distribution. The result $\mathbb{P}(\psi_1(T) \leq \psi_2(T)) = 1$ implies inequality (3.18) holds, and the proof is complete. ■

Next, in Proposition 3.1, we state without proof a well-known result about the comparability of expectations for random vectors under the usual multivariate stochastic order.

Proposition 3.1 *(Shaked and Shanthikumar [83]). Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two n-dimensional, random vectors such that $\boldsymbol{X} \leq_{st} \boldsymbol{Y}$. For any nondecreasing function ψ on $\mathbb{R}^n$*

$$\mathbb{E}[\psi(\boldsymbol{X})] \leq \mathbb{E}[\psi(\boldsymbol{Y})].$$

We are now prepared to state our main result concerning the value function. Namely, under appropriate conditions the value function $V(\boldsymbol{s})$ is monotone in $\boldsymbol{s} \in \Gamma$.

Theorem 3.3 *If $\boldsymbol{P}$ is IFR and the degradation rates r_i^j are monotone nondecreasing in $j \in S$ for each $i = 1, \ldots, n$, then the value function $V(\boldsymbol{s})$ is monotone nondecreasing in $\boldsymbol{s} \in \Gamma$.*

Proof. For $(\boldsymbol{x}, j) \in \Gamma$, denote the mth iterate of the value iteration algorithm by $v_j^m(\boldsymbol{x}) \equiv v^m(\boldsymbol{s})$. We prove the theorem by induction on m. Take $v_j^0(\boldsymbol{x}) = 0$ for all $(\boldsymbol{x}, j) \in \Gamma$. Therefore, as in the proof of Theorem 3.2, $v_j^1(\boldsymbol{x}) = c_d \sum_{i=1}^n \mathbb{I}_i(\boldsymbol{x})$, which is monotone nondecreasing in $\boldsymbol{s}$. For the induction hypothesis, assume $v^m(\boldsymbol{s})$ is monotone nondecreasing in $\boldsymbol{s} \in \Gamma$. By Lemma 3.5 and Proposition 3.1, for $\boldsymbol{s} \leq \boldsymbol{s}' \in \Gamma$,

$$\mathbb{E}\left(v^m(\boldsymbol{S}_1) | \boldsymbol{S}_0 = \boldsymbol{s}, \boldsymbol{A}_0 = \boldsymbol{a}\right) \leq \mathbb{E}\left(v^m(\boldsymbol{S}_1) | \boldsymbol{S}_0 = \boldsymbol{s}', \boldsymbol{A}_0 = \boldsymbol{a}\right).$$

By Lemma 3.3, $c(\boldsymbol{s}, \boldsymbol{a})$ is monotone in $\boldsymbol{s}$; hence,

$$c(\boldsymbol{s}, \boldsymbol{a}) + \alpha\mathbb{E}\left(v^m(\boldsymbol{S}_1) | \boldsymbol{S}_0 = \boldsymbol{s}, \boldsymbol{A}_0 = \boldsymbol{a}\right) \leq c(\boldsymbol{s}', \boldsymbol{a}) + \alpha\mathbb{E}\left(v^m(\boldsymbol{S}_1) | \boldsymbol{S}_0 = \boldsymbol{s}', \boldsymbol{A}_0 = \boldsymbol{a}\right). \tag{3.20}$$

By minimizing over $\boldsymbol{a} \in \mathcal{A}$ on both sides of inequality (3.20), we obtain $v^{m+1}(\boldsymbol{s}) \leq v^{m+1}(\boldsymbol{s}')$. Thus, by Lemma 3.1, the result is proved. ■

Theorem 3.3 asserts that, if the uniformized environment process is IFR, and the degradation rates are monotone, then the values of the value function are ordered over the entire state space.

3.3.2 Special Case: A Single System

In general, the optimal policy is not necessarily monotone over the entire state space. Indeed, in Section 3.5, we provide two numerical examples illustrating that, even under the conditions of Lemma 3.4, optimal policies need not be monotone. In this subsection, we restrict our attention to the special case of a single system and establish the monotonicity of optimal policies for this case. Before proceeding to the main result, Lemma 3.6 is first needed.

Lemma 3.6 *Suppose there is only a single system (i.e., $n = 1$). If $\boldsymbol{P}$ is IFR, and the degradation rates r^j are monotone nondecreasing in $j \in S$, then $\mathbb{E}(V(\boldsymbol{S}_1)|\boldsymbol{S}_0 = \boldsymbol{s}, A_0 = a)$ is submodular on $\Gamma \times \mathcal{A}$.*

Proof. By Lemma 3.4(b), the tail distribution function $q(\cdot|\cdot)$ is submodular on $\Gamma \times \mathcal{A}$. Define $F(\boldsymbol{s}|\boldsymbol{s}_0, a) = \mathbb{P}(\boldsymbol{S}_1 \leq \boldsymbol{s}|\boldsymbol{S}_0 = \boldsymbol{s}_0, A_0 = a)$. Then for any $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \Gamma$ and $a_1, a_2 \in \mathcal{A}$ such that $\boldsymbol{s}_1 \leq \boldsymbol{s}_2$ and $a_1 \leq a_2$,

$$\int_{\boldsymbol{s} \geq \boldsymbol{s}'} dF(\boldsymbol{s}|\boldsymbol{s}_2, a_2) + \int_{\boldsymbol{s} \geq \boldsymbol{s}'} dF(\boldsymbol{s}|\boldsymbol{s}_1, a_1) \leq \int_{\boldsymbol{s} \geq \boldsymbol{s}'} dF(\boldsymbol{s}|\boldsymbol{s}_1, a_2) + \int_{\boldsymbol{s} \geq \boldsymbol{s}'} dF(\boldsymbol{s}|\boldsymbol{s}_2, a_1)$$

for each $\boldsymbol{s}' \in \Gamma$. Thus,

$$\int_{\boldsymbol{s} \geq \boldsymbol{s}'} \left[dF(\boldsymbol{s}|\boldsymbol{s}_2, a_2) + dF(\boldsymbol{s}|\boldsymbol{s}_1, a_1)\right] \leq \int_{\boldsymbol{s} \geq \boldsymbol{s}'} \left[dF(\boldsymbol{s}|\boldsymbol{s}_1, a_2) + dF(\boldsymbol{s}|\boldsymbol{s}_2, a_1)\right].$$

Let us define $dF_1(\boldsymbol{s}) = dF(\boldsymbol{s}|\boldsymbol{s}_2, a_2) + dF(\boldsymbol{s}|\boldsymbol{s}_1, a_1)$ and $dF_2(\boldsymbol{s}) = dF(\boldsymbol{s}|\boldsymbol{s}_1, a_2) + dF(\boldsymbol{s}|\boldsymbol{s}_2, a_1)$. Noting that $V(\boldsymbol{s})$ is monotone nondecreasing (by Theorem 3.3), and that Proposition 3.1 extends to finite measures, we obtain

$$\int_{\boldsymbol{s} \in \Gamma} V(\boldsymbol{s}) dF_1(\boldsymbol{s}) \leq \int_{\boldsymbol{s} \in \Gamma} V(\boldsymbol{s}) dF_2(\boldsymbol{s}). \tag{3.21}$$

By expanding the terms on both sides of inequality (3.21), the result is obtained. ■

We are now prepared to state the main result for a single system. Theorem 3.4 asserts that, under suitable conditions, the optimal decision rule is monotone nondecreasing over the entire state space Γ.

Theorem 3.4 *For the case of $n = 1$, if $c_r - c_p \leq c_d$, $\boldsymbol{P}$ is IFR, and the degradation rates r^j are monotone nondecreasing in $j \in S$, the optimal decision rule $d^*(\boldsymbol{s})$ is monotone in $\boldsymbol{s} \in \Gamma$.*

Proof. By Lemma 3.3(b), the expected one-step cost function $c(\boldsymbol{s}, \boldsymbol{a})$ is submodular on $\Gamma \times \mathcal{A}$. Similarly, by Lemma 3.6, we have that $\mathbb{E}(V(\boldsymbol{S}_1)|\boldsymbol{S}_0 = \boldsymbol{s}, A_0 = a)$ is submodular. Thus,

$$Q(\boldsymbol{s}, a) = c(\boldsymbol{s}, a) + \alpha\mathbb{E}(V(\boldsymbol{S}_1)|\boldsymbol{S}_0 = \boldsymbol{s}, A_0 = a)$$

is a non-negative, linear combination of submodular functions and is, therefore, submodular. Hence, by Theorem 3.1, the optimal decision rule $d^*(\boldsymbol{s})$ is monotone in $\boldsymbol{s} \in \Gamma$. ∎

Theorem 3.4 ensures that, under suitable conditions, threshold policies are optimal, and these thresholds are monotone over the environment states. Ulukus et al. [97] examined a similar model and proved the optimality of threshold policies for a single-system (or component) setting and conjectured (without proof) that the thresholds are monotone over the environment states. Theorem 3.4 is significant in that it provides at least a partial resolution to this conjecture. Specifically, setting $c_0 = 0$ and $c_d \gg c_r$ in our framework yields a model that is identical to the one studied in [97], except that reactive replacements do not occur immediately following a failure; they occur at the start of the subsequent period. However, the assumption that c_d is very large induces an optimal policy that forces reactive replacements; hence, our model is not more restrictive than the one studied in [97].

The actions satisfying the optimality equation (3.3) for each state (corresponding to an optimal policy) cannot be directly computed by standard numerical methods (e.g. using the value or policy iteration algorithms [77]). Section 3.4 presents an approximation scheme to address the numerical problems associated with obtaining replacement policies.

3.4 AN APPROXIMATE FORMULATION

In this section, we address dual manifestations of Bellman's curses of dimensionality, namely a continuous state space and a combinatorially large action space [10, 76]. To obtain high-quality policies, we customize and employ multiple reinforcement learning techniques: Q-function approximation, reduction to a subset of the action space, and an on-policy approximate dynamic programming (ADP) algorithm, namely the state action reward state action, or SARSA(γ), algorithm. For a complete description of the algorithm, the reader is referred to [82, 92].

Before discussing the details of our approximation scheme, we first introduce a transformation of the state description. In lieu of considering the current cumulative degradation level directly, we consider the state to be a vector of probabilities on the hypercube $[0,1]^n$. In particular, define the state by $\tilde{\boldsymbol{x}} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)$, where

$$\tilde{x}_i = \mathbb{P}\left(X^i_{m+1} \geq \xi | X^i_m = x_i, Z_m = j\right) = \exp\left(\frac{-q(\xi - x_i)}{r^j_i}\right), \tag{3.22}$$

Z_m is the current environment state, X_m^i is the current degradation level of system i, and X_{m+1}^i is the degradation level at the time of the next inspection. That is, for each system i, the corresponding system in the transformed state vector, denoted by $\tilde{x}_i$, is the probability that the system will fail before the next decision epoch. We pause here to remark on two aspects of this transformation: (i) no information is lost in the transformation (given fixed model parameters, the mapping is a bijection); and (ii) the new state implicitly contains information about the underlying degradation process. The first point is important, as we do not fundamentally change the problem. The second point is of practical significance for our function approximation. We utilize a linear basis to approximate the Q-function, and direct inclusion of this information improves the performance of simple basis functions. Intuitively, the degradation level itself is abstract and uninformative, whereas probabilities require no additional knowledge of the underlying degradation process to interpret; hence, this state space is more appropriate. For convenience, we denote the modified state as $\tilde{\boldsymbol{s}}$ and state space as $\tilde{\Gamma} = [0,1]^n \times S$.

The first issue we address is reduction of the action space $\mathcal{A}$. It is clear that for a large number of systems, n, the action space $\mathcal{A} = \{0,1\}^n$ is too large to perform any practical computational operations. For example, for $n = 20$ the cardinality of the action space is over 10^6 and for $n = 30$ the cardinality is over one billion. Unfortunately, it is very difficult to characterize the form of an optimal policy in the multiple system case, so instead we consider a reduced action space based on the transformed state vector $\tilde{\boldsymbol{x}}$ that may, or may not, contain the optimal actions. We define the new action space as

$$\tilde{\mathcal{A}} = \{0, 1, \ldots, n\},$$

where $a = 0$ corresponds to taking no action, and $a = k$ corresponds to replacing all systems i such that $\tilde{x}_i \geq \tilde{x}_k$, for $k = 1, 2 \ldots, n$. The action $a = k > 0$ means that any system whose probability of failing in the next period is at least as high as that of system k is replaced. For example, if $\tilde{\boldsymbol{x}} = (0.3, 0.7, 0.25, 0.6)$ and action $a = 4$ is selected, then all systems whose probability of failing within the next period is at least 0.6 are replaced. In this case, systems 2 and 4 are replaced. It should be noted that this action space can be interpreted as the most general set of threshold policies in the transformed state. That is, for every state $\tilde{\boldsymbol{s}} \in \tilde{\Gamma}$ the optimal action in $\tilde{\mathcal{A}}$ is equivalent to replacing all systems above some optimal threshold $u(\tilde{\boldsymbol{s}})$. Most importantly, this set helps facilitate computational tractability, as it is linear in the number of systems.

The next issue we address is the continuous nature of $\mathcal{X}$. In order to overcome this difficulty, we employ function approximation techniques. First, we recall the definition of the Q-function, or action-value function [76] from equation (3.5). For each $j \in S$, the function $Q_j(\cdot)$ maps state-action pairs to expected total discounted costs, i.e.,

$$Q_j(\boldsymbol{x}, \boldsymbol{a}) = c(\boldsymbol{s}, \boldsymbol{a}) + \alpha \sum_{k=1}^{\ell} \left(\int_0^{\infty} V_k \left(\boldsymbol{x}' \right) q e^{-qt} dt \right) p_{kj}.$$

It is seen that, by the relationship between (3.3) and (3.5), optimal actions can be determined by minimizing the Q-function over the original action space $\mathcal{A}$. We seek to approximate the Q-function as a weighted (linear) sum of basis functions that take the transformed state variables as input [35]. In particular, we approximate $Q_j(\boldsymbol{x}, \boldsymbol{a})$ by $\tilde{Q}_j(\tilde{\boldsymbol{x}}, a)$ defined as

$$\tilde{Q}_j(\tilde{\boldsymbol{x}}, a) := \sum_{k=1}^{K} \lambda_k^j \phi_k^j(\tilde{\boldsymbol{x}}, a), \tag{3.23}$$

where λ_k^j are real-valued coefficients and $\phi_k^j(\cdot)$ are basis functions. By Theorem 3.2, it is known that the value function is monotone in the state space; hence, we choose to represent the Q-function as a linear combination of monotone functions and constant functions. In particular, we utilize constant functions along with functions of the form $\cos(f(\tilde{\boldsymbol{x}}, a))$, where $f : \tilde{\Gamma} \to [0, 1]$. This choice of basis functions is practical, as they are bounded and monotone on $\tilde{\Gamma}$. Additionally, the non-constant basis functions are sigmoidal, enabling flexible modeling of nonlinearity in the value function, and appropriate choices for f allow for dependencies between systems to be captured. As described in [54], these basis functions can be viewed as adapted first-order Fourier basis functions, and we choose the particular form

$$\begin{aligned}\tilde{Q}_j(\tilde{\boldsymbol{x}}, a) = \lambda_0^j + \lambda_1^j \mathbb{I}\left(a > 0\right) + \sum_{i=1}^{n} \lambda_{i,1}^j \mathbb{I}\left(\tilde{x}_i \geq \tilde{x}_a\right) + \sum_{i=1}^{n} \lambda_{i,1}^j \mathbb{I}\left(\tilde{x}_i = 1, a > 0\right) \\ + \sum_{i=1}^{n} \sum_{k=i}^{n} \lambda_{i,k,1}^j \phi_{i,k,1}^j\left(\tilde{x}, a\right) + \sum_{i=1}^{n} \sum_{k=i+1}^{n} \lambda_{i,k,2}^j \phi_{i,k,2}^j\left(\tilde{x}, a\right),\end{aligned} \tag{3.24}$$

where

$$\phi_{i,k,m}^j = \begin{cases} \cos\left(\frac{\pi}{2} \cdot \left(\mathbb{I}\left(\tilde{x}_i < \tilde{x}_a\right) \tilde{x}_i + \mathbb{I}\left(\tilde{x}_k < \tilde{x}_a\right) \tilde{x}_k\right)\right), & m = 1, \\ \cos\left(\frac{\pi}{2} \cdot \mathbb{I}\left(\min\left\{\tilde{x}_i, \tilde{x}_k\right\} < \tilde{x}_a\right) \cdot \left(\tilde{x}_i + \tilde{x}_k\right)\right), & m = 2. \end{cases} \tag{3.25}$$

The first term in (3.24) is a constant, the second term accounts for the impact of the setup cost, the third term accounts for preventive maintenance cost, the fourth for reactive maintenance, and the final two terms account for the impact of the degradation level (given the current action taken). It can be seen in (3.25) that the approximation $\tilde{Q}_j(\cdot)$ considers only pairwise dependencies between systems. This assumption can be relaxed, but the computation time of any iterative learning process would naturally be adversely impacted. For the model in (3.24), the coefficients λ_k^j can be obtained using an iterative, on-policy learning algorithm, such as the SARSA(γ) algorithm, as described in Sutton and Barto [92].

3.5 NUMERICAL EXAMPLES

In this section, we illustrate the efficacy of the proposed approximation framework in obtaining high-quality solutions. Both small- and large-scale examples are considered. Small problem instances are illustrated because exact solutions can be obtained using standard MDP machinery; hence, approximate solutions can be compared to these exact solutions. Larger problem instances are used to assess how well our approximation scheme performs on problems that are otherwise intractable.

Because our aim is to compare distinct policies, we first simulate the evolution of the environment process over a large number of sample paths. The simulation run length is given by the number of decision epochs N (recall that the time between decision epochs is exponentially distributed with rate parameter q). Along each sample path, the total discounted cost is computed for each policy of interest, and these values are compared. It should be noted that, because the expected one-step costs are bounded, and the cost function is discounted, we can determine *a priori* the simulation run length needed to ensure that the total discounted cost is accurate to a fixed constant. In particular, to guarantee the finite approximation is within ε ($\varepsilon > 0$) of the true total discounted cost, the length of the sample paths N must satisfy

$$N \geq \frac{\ln((1-\alpha)\varepsilon/C)}{\ln(\alpha)} - 1,$$

where C is any valid upper bound on the expected one-step costs, and α is the one-step discount rate. For all numerical examples, N was chosen to correspond to $\varepsilon = 0.01$ and C was taken to be $c_0 + nc_d + nc_r$. All problem instances were coded within the MATLAB R2016a computing environment and executed on a personal computer with a 3.50 GHz processor and 8 GB of RAM.

3.5.1 Single-System Problems

First, we present single-system examples to demonstrate the accuracy of our approach in the simplest case. We modify the cost structure of Section 3.2 to fit the single-system case. Rather than having four cost parameters, we simply have the preventive and reactive replacement costs, c_p and c_r respectively, and we force replacement of the system whenever it is found to be failed.

In the first example, we fix the parameter values and solve the resulting problem repeatedly using the approximate formulation. We compute multiple learned solutions over a single problem to explore the impact of a stochastic learning algorithm on the consistency of solutions. For the first instance, the discount rate is $\alpha = 0.99$ and the cost vector is $\boldsymbol{c} = (c_p, c_r) = (3, 10)$. The failure threshold is set to $\xi = 1$, and the environment $\mathscr{Z}$ has state space $S = \{1, 2, 3, 4\}$ with infinitesimal generator matrix

$$\mathcal{Q} = \begin{pmatrix} -5 & 5 & 0 & 0 \\ 2.5 & -5 & 2.5 & 0 \\ 0 & 2.5 & -5 & 2.5 \\ 0 & 0 & 5 & -5 \end{pmatrix}.$$

The inspection rate is $q = 10$ and degradation rates are $\boldsymbol{r} = (r_1, r_2, r_3, r_4) = (2.5, 3, 3.5, 4)$.

The degradation interval $[0, \xi]$ was discretized into $10,000$ states, and the optimal policy was obtained numerically using the value iteration algorithm. For each fixed environment state $j \in S$, the optimal policy was determined to be a threshold policy. This is intuitive, as the problem can viewed as the general problem with $c_0 = 0$ and c_d chosen sufficiently large so as to force reactive replacements. Therefore, Theorem 3.4 ensures that a monotone threshold policy is optimal. In particular, the thresholds were found to be $\boldsymbol{u} = (u_1, u_2, u_3, u_4) = (0.524, 0.469, 0.430, 0.387)$, where for state $(x, j) \in [0, 1] \times S$ the optimal action $a_j^*(x)$ is given by

$$a_j^*(x) = \begin{cases} 0, & x \le u_j, \\ 1, & x > u_j. \end{cases}$$

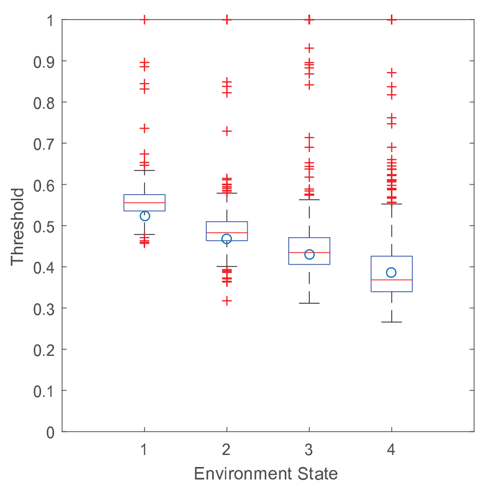
To adapt to the modified cost structure, we simplified the basis functions for our approximation as follows:

$$\phi_j^i(\tilde{x}, a) = \begin{cases} 1, & i = 1, \\ \mathbb{I}(a = 0)\cos(\pi\tilde{x}), & i = 2, \\ \mathbb{I}(a = 1,\, \tilde{x} < 1), & i = 3, \\ \mathbb{I}(\tilde{x} \ge 1), & i = 4, \end{cases}$$

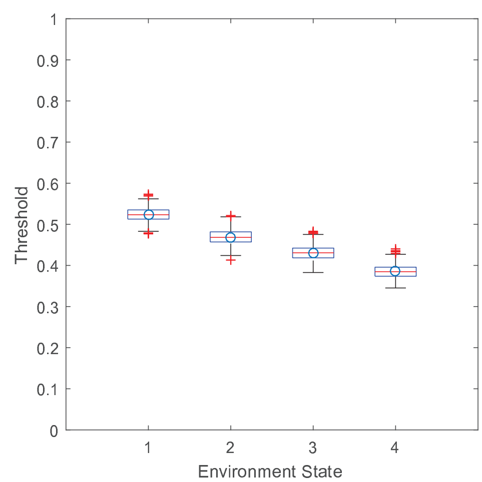
where $\tilde{x}$ is the transformed state as described by (3.22). Therefore, the expected total discounted cost-to-go, starting from state $\boldsymbol{s} = (x, j) \in \Gamma$ and taking action $a \in \mathcal{A}$, is approximated by

$$\tilde{Q}_j(\tilde{x}, a) = \sum_{i=1}^{4} \lambda_j^i \phi_j^i(\tilde{x}, a).$$

In order to test the quality of this approximation, 500 sample paths were simulated, each with $2,000,000$ decision epochs. On each of these sample paths, the approximation model was trained using a SARSA(γ) algorithm with $\gamma = 0.9$. Exploration is encouraged by using an ϵ-greedy action-selection approach, where the best myopic action is chosen with probability $1 - \epsilon$, and a random action is chosen with probability ϵ. For these examples, ϵ is initialized to 0.5 and linearly decreased in increments of 2.5×10^{-7} to 0.0001 by the final iteration.



(a) After 400,000 simulation iterations.



(b) After 2,000,000 simulation iterations.

Figure 10: Boxplots of thresholds from learned policies (optimal thresholds given by circles).

Figure 10 shows box-and-whisker plots of the thresholds derived from the learned policies. In Figure 10(a), we see the thresholds from the policies if learning is terminated at 400,000 iterations. This plot indicates that the bias from parameter initialization is still significant in the learned policies. Generally, the thresholds are still too high, which matches intuition as the initial parameter values would induce a policy that never replaces components. By $2,000,000$ iterations, this upward bias is drastically diminished, and the variance has been greatly reduced. In fact, the mean absolute deviations from the true thresholds are under 0.015 for all environment states. The thresholds from value iteration and the mean learned thresholds are given in Table 1.

Table 1: Comparison of exact and learned thresholds.

Technique	Environment State			
	$j = 1$	$j = 2$	$j = 3$	$j = 4$
Value iteration threshold	0.5238	0.4688	0.4301	0.3865
Mean learned threshold	0.5240	0.4701	0.4303	0.3850

While these policies appear to be accurate, it is important to gauge the impact of small policy variations on the expected total discounted cost. To this end, we first simulate 5,000 sufficiently long sample paths per policy (2,500,000 sample paths in total) as described at the beginning of this section. Each simulation run is initialized in state $(x, j) = (0, 1)$. For learned policy $m \in \{1, \ldots, 500\}$ and sample path $k \in \{1, \ldots, 5,000\}$, we then follow both the optimal policy and the learned policy over each of these sample paths to determine their respective total discounted costs, call them $\tilde{V}_k^m$ and $\hat{V}_k^m$, respectively. The sample-wise values are then averaged to yield approximate expected total discounted costs

$$\tilde{V}^m = \frac{1}{5000} \sum_{k=1}^{5000} \tilde{V}_k^m \quad \text{and} \quad \hat{V}^m = \frac{1}{5000} \sum_{k=1}^{5000} \hat{V}_k^m.$$

The results are summarized in the box-and-whisker plots of Figure 11. While the optimal policy does perform better on average (over the sample paths), the percent difference in the means is nominal at approximately

$$\frac{|\tilde{V} - \hat{V}|}{\tilde{V}} = 4.887 \times 10^{-4},$$

where $\tilde{V} = \sum_m \tilde{V}^m/500$ and $\hat{V} = \sum_m \hat{V}^m/500$. In fact, the approximate policy outperforms the optimal policy in 49.16% of the total sample paths. Therefore, the learned policies are not only superficially similar to the optimal policies, but almost identical in performance as well. Lastly, it should be noted that obtaining the exact solution took 80.5269 seconds (averaged over 100 function calls) and the average time for the approximate model to generate the simulations and train was only 25.8465 (averaged over the 500 sample paths). Thus, even in the single-system case, this solution method is able to generate stable, accurate policies faster than value iteration.

In the next example, 200 single-system problems are randomly generated, and one solution to the approximate formulation is computed for each problem. Our aim is to vary the problem

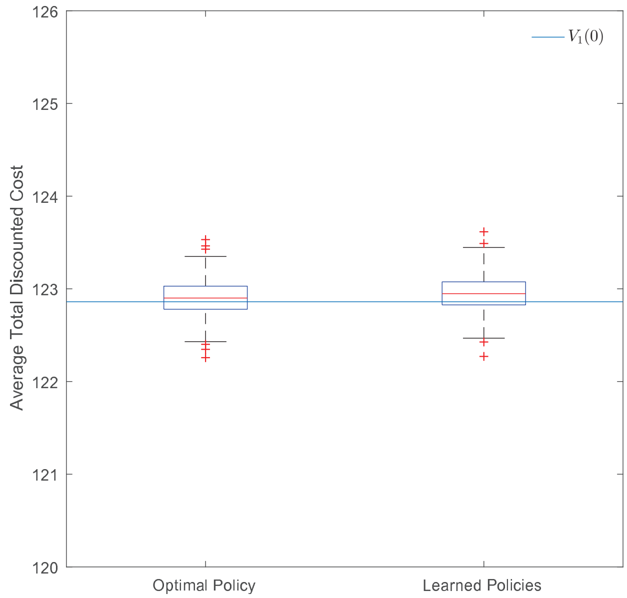


Figure 11: Boxplots comparing cost using the optimal policy versus the learned policies.

parameters over a wide range to assess the robustness of the approximate formulation. In what follows, $U(a,b)$ denotes a continuous uniform random variable on (a,b). Fixing $\alpha = 0.99$, and the number of environment states at $\ell = 4$, we randomly generated $M = 200$ problem instances. (These two parameters are fixed because they significantly affect the learning rate, and the training size will be fixed over the problem set.) For problem $m \in \{1, \ldots, M\}$, the cost vector is denoted $\boldsymbol{c}^m = (c_p^m, c_r^m)$, where $c_p^m \sim U(0,5)$ and $c_r^m \sim c_p^m + U(0,10)$. As before, the failure threshold is fixed at $\xi^m = 1$, corresponding to rates being normalized. The environment process $\mathcal{Z}^m$ has state space $S^m = \{1,2,3,4\}$ with infinitesimal generator matrix $Q^m = [q_{ij}^m]$, where $q_{ij}^m \sim U(0,10)$ for $j \neq i$ and $q_{ii}^m = -\sum_j q_{ij}^m$. The inspection rate is given as $q^m = 2 \cdot \max_i\{-q_{ii}^m\}$, and the degradation rates are $\boldsymbol{r}^m = (r_1^m, r_2^m, r_3^m, r_4^m)$, where $r_1^m \sim U(0,4)$ and $r_j^m \sim r_{j-1}^m + U(0,4)$ for $j = 2,3,4$.

For each problem, the degradation interval $[0,\xi]$ was discretized into $10,000$ states and the optimal policy was obtained numerically using the value iteration algorithm. In each case, a long sample path was simulated with 2,000,000 decision epochs, and the approximate model was trained using a SARSA(γ) algorithm with $\gamma = 0.9$. Exploration was again encouraged by using an ϵ-greedy action-selection approach, where ϵ was initialized to 0.5 and linearly decreased to 0.0001 by the final iteration. Next, to evaluate the quality of the solutions, 5,000 sufficiently long sample paths were

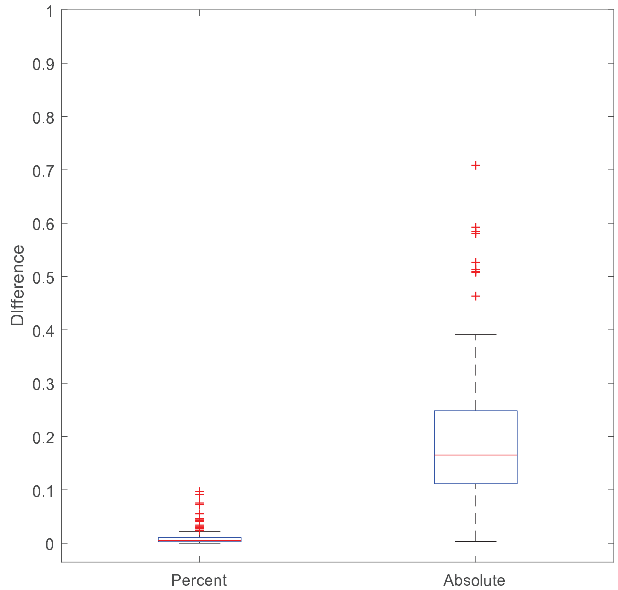


Figure 12: Boxplots of percent and absolute differences.

simulated per problem using the initial state $(x, j) = (0, 1)$. For each problem m and sample path $k_m \in \{1, \ldots, 5,000\}$, the learned policy was followed over the path to determine the total discounted cost, denoted v_k^m. For problem m, we denote the value function, at state $(0, 1) \in \Gamma$ by V^m, and assume that these are given by the final values from the value iteration algorithm. Similarly, for problem m, we denote the approximate value function, at state $(0, 1) \in \Gamma$ by $\widehat{V}^m = \sum_k v_k^m/5000$. Lastly, we define $\delta^m = |V^m - \widehat{V}^m|$ and $\Delta^m = \delta^m/V^m$ to be, respectively, the absolute and percent differences between the actual and approximate value functions. From the data, we find that there is one significant outlier, where $\Delta^m = 0.38$. It is seen that at this data point, the actual value function is very small at $V^m = 1.0649$. We remove this outlier from the data set and present the data in Figure 12. For the percent differences, the sample mean of Δ^m is 0.0099 and the standard deviation is 0.0143. In fact, for approximately 97.5% of the cases $\Delta^m \leq 0.05$. For the absolute differences, we find that the sample mean of δ^m is 0.1908 and the standard deviation is 0.1170. In fact, for over 95% of the cases $\delta^m \leq 0.4$.

3.5.2 Multiple-System Problems

Next, we present multiple-system examples to illustrate solution quality and the tractability of our approach. We consider three problems with $n = 2$, $n = 3$ and $n = 50$, respectively. For cases $n = 2$ and $n = 3$, we obtain "exact" solutions by applying the value iteration algorithm on discretized versions of the problems. Our approximate solutions, obtained by using the methods developed in Section 3.4, are compared to these solutions. Additionally, we describe a set of heuristic policies and compare our solutions to those as well. When $n = 50$, exact solutions cannot be obtained, so our approximation is compared only to heuristic policies.

Example. For the first example, $n = 2$ systems, the discount rate is $\alpha = 0.99$ and the cost vector is $\boldsymbol{c} = (c_0, c_p, c_r, c_d) = (20, 3, 10, 8)$. The failure thresholds are $\xi_i = 1$, for all i, and the environment $\mathcal{Z}$ has state space $S = \{1, 2, 3\}$ with infinitesimal generator matrix

$$\mathcal{Q} = \begin{pmatrix} -5 & 5 & 0 \\ 2.5 & -5 & 2.5 \\ 0 & 5 & -5 \end{pmatrix}.$$

The inspection rate is $q = 10$ and degradation rates were randomly generated as $\boldsymbol{r}_i = (r_i^1, r_i^2, r_i^3) = U(1,2) \cdot (0.5, 1.0, 1.5)$, for $i = 1, 2$, where $U(1,2)$ is a continuous uniform random variable on $(1, 2)$. The particular realization of $\boldsymbol{r}$ is

$$\boldsymbol{r} = \begin{pmatrix} 0.9567 & 1.9134 & 2.8701 \\ 0.5635 & 1.1270 & 1.6905 \end{pmatrix}.$$

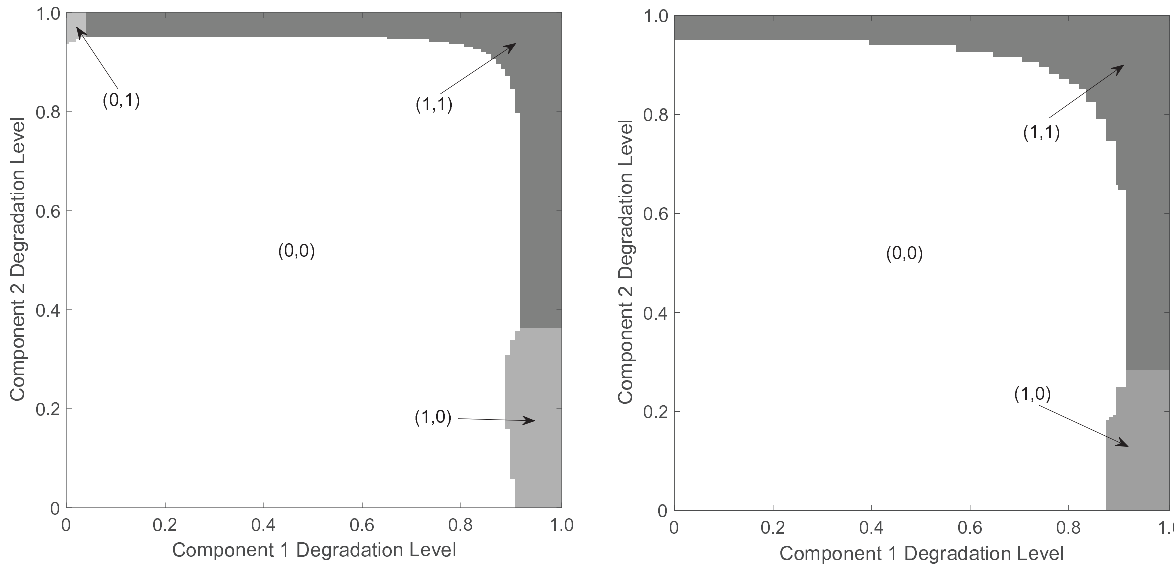
To compute an exact solution, the degradation interval $[0, \xi]$ was uniformly discretized into 1,000 states for each system, yielding a discretized problem with 3,000,000 total states. The optimal policy was obtained numerically using the value iteration algorithm. For each $\boldsymbol{s}$ in the discretized set of states, let $v^m(\boldsymbol{s})$ denote the mth iterate of the value iteration algorithm. The algorithm terminates when the maximum norm of the difference between subsequent value function iterates is below 0.01, i.e.,

$$||v_{m+1} - v_m||_\infty = \max_{\boldsymbol{s}} \{|v_{m+1}(\boldsymbol{s}) - v_m(\boldsymbol{s})|\} \leq 0.01,$$

or after 604,800 seconds (7 days). The algorithm terminated after 401,029.97 seconds (approximately 4.6 days).

In order to train our approximate model, a long sample path was simulated with 4,000,000 decision epochs, and the approximate model was trained using a SARSA(γ) algorithm with $\gamma = 0.9$.

Exploration was again encouraged by using an ϵ-greedy action-selection approach, where ϵ was initialized to 0.5 and linearly decreased to 0.0001 by the final iteration. The training took 369.54 seconds, three orders of magnitude faster than value iteration.



(a) Depiction of optimal policy.

(b) Depiction of learned policy.

Figure 13: Comparison between optimal and learned policy in environment state 1.

In Figure 13, the optimal and learned policies are depicted when the environment state is fixed at 1. We observe that the optimal policy does not have a monotone structure. While the absence of a monotone structure is counterintuitive, the overall structure is not. When one system is very heavily degraded, but the other system is nearly as-good-as-new, it is optimal to replace the heavily degraded system. However, due to economic dependencies, the region where replacing both systems is optimal dominates the replacement regions. This result should be expected, as the shared replacement cost is high and the difference between the preventive and reactive replacement costs is substantial. Additionally, it is optimal to wait longer to perform maintenance actions on system 2 as a consequence of the degradation rates. In particular, we note that $r_2^1 < r_1^1$. Contrasting the optimal and learned policies, we note that while similar, the learned policy is slightly more aggressive and replaces more frequently.

For a cost comparison, we also consider a set of threshold-based replacement policies. To understand these policies, we begin with the policy that ignores the dependency of the systems and decouples the problem. For this policy, we consider each system separately and solve the single-system problem described in Section 3.5.1. We then use the computed thresholds to determine when to replace individual systems independently of the others. We refer to this as a 1-policy. We then generalize this policy to a k-policy, where we replace all systems above their independent threshold whenever there are at least k systems above their threshold. In the case where $n = 50$, two other similarly structured policies were considered, but failed to perform as well: (i) wait until at least k systems are failed and then replace all failed systems; and (ii) wait until at least k systems exceed their threshold and then replace all systems.

As in the single-system cases, we assess the quality of our solutions by following each policy over 5,000 sufficiently long sample paths simulated using the initial state $(x, j) = (\mathbf{0}, 1)$. The results of this comparison are summarized in the box-and-whisker plots of Figure 14. Beginning with

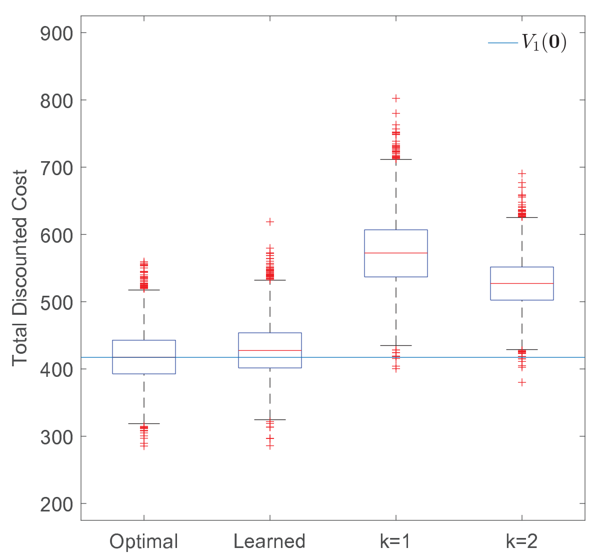


Figure 14: Boxplots comparing cost under different policies.

the heuristic policies, we note that they are both significantly outperformed by the learned policy. On average, the learned policy outperforms the 1-policy by 32.79% and the 2-policy by 22.28%. Moreover, the learned policy outperforms the 1-policy on all sample paths and the 2-policy on all but three sample paths. This performance gap demonstrates the importance of considering structural and economic dependence within our model. Comparing the learned and optimal policies, we note the average total discounted cost for the learned policy was 431.32 compared to 418.74 for the optimal policy (approximately 3% difference). The learned policy even outperformed the optimal policy on 1,041 sample paths.

Example. For this example, $n = 3$ systems, the discount rate is $\alpha = 0.99$ and the cost vector is $\boldsymbol{c} = (c_0, c_p, c_r, c_d) = (20, 3, 10, 8)$. The failure thresholds are $\xi_i = 1$, for all i, and the environment $\mathcal{Z}$ has state space $S = \{1, 2\}$ with infinitesimal generator matrix

$$\mathcal{Q} = \begin{pmatrix} -5 & 5 \\ 5 & -5 \end{pmatrix}.$$

The inspection rate is $q = 10$ and degradation rates were randomly generated as $\boldsymbol{r}_i = (r_i^1, r_i^2) = U(1, 2) \cdot (0.5, 1.0)$. The particular realization of $\boldsymbol{r}$ is

$$\boldsymbol{r} = \begin{pmatrix} 0.9074 & 1.8147 \\ 0.9529 & 1.9058 \\ 0.5635 & 1.1270 \end{pmatrix}.$$

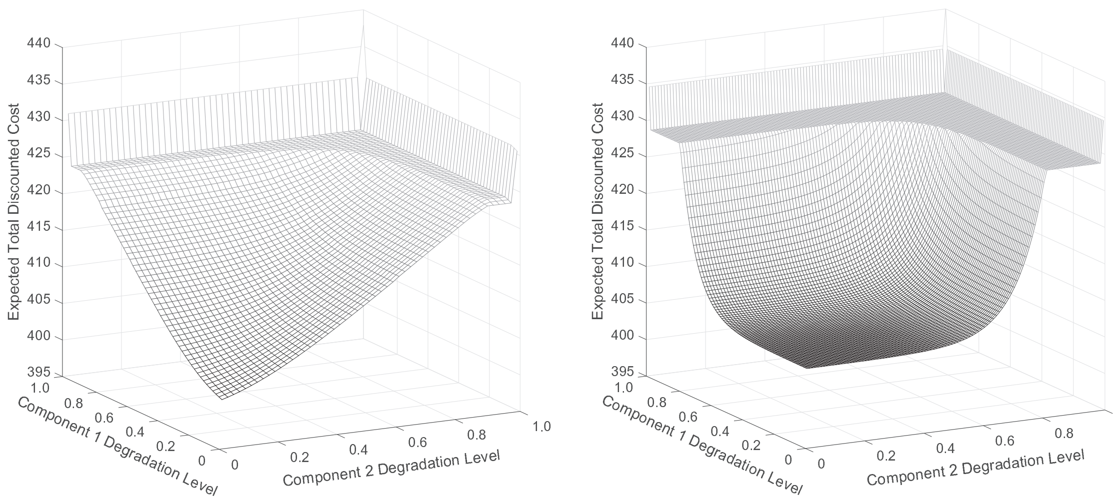
To compute an exact solution, the degradation interval $[0, \xi]$ was uniformly discretized into 125 states, for each system, yielding a discretized problem with 3,906,250 total states. The optimal policy was obtained numerically using the value iteration algorithm. For each $\boldsymbol{s}$ in the discretized set of states, let $v^m(\boldsymbol{s})$ denote the mth iterate of the value iteration algorithm. The algorithm was set to terminate when the maximum norm of the difference between subsequent value function iterates fell below 0.01, i.e.,

$$||v_{m+1} - v_m||_\infty = \max_{\boldsymbol{s}} \{|v_{m+1}(\boldsymbol{s}) - v_m(\boldsymbol{s})|\} \leq 0.01,$$

or after 604,800 seconds (7 days). The algorithm terminated after reaching 604,800 seconds and the maximum norm of the difference between the final two value function iterates was 0.0253.

In order to train our approximate model, a long sample path was simulated with 4,000,000 decision epochs, and the approximate model was trained using a SARSA(γ) algorithm with $\gamma = 0.9$. Exploration was again encouraged by using an ϵ-greedy action-selection approach, where ϵ was initialized to 0.5 and linearly decreased to 0.0001 by the final iteration. The training took 482.48 seconds to complete.

Figure 15 depicts the optimal and learned value functions when the environment is fixed at state 2 and the degradation level of system 3 is fixed at 0.40. At low degradation levels, in regions where it is optimal to wait to replace, the magnitude of the gradient of the value function approximation



(a) Optimal value function.

(b) Learned value function.

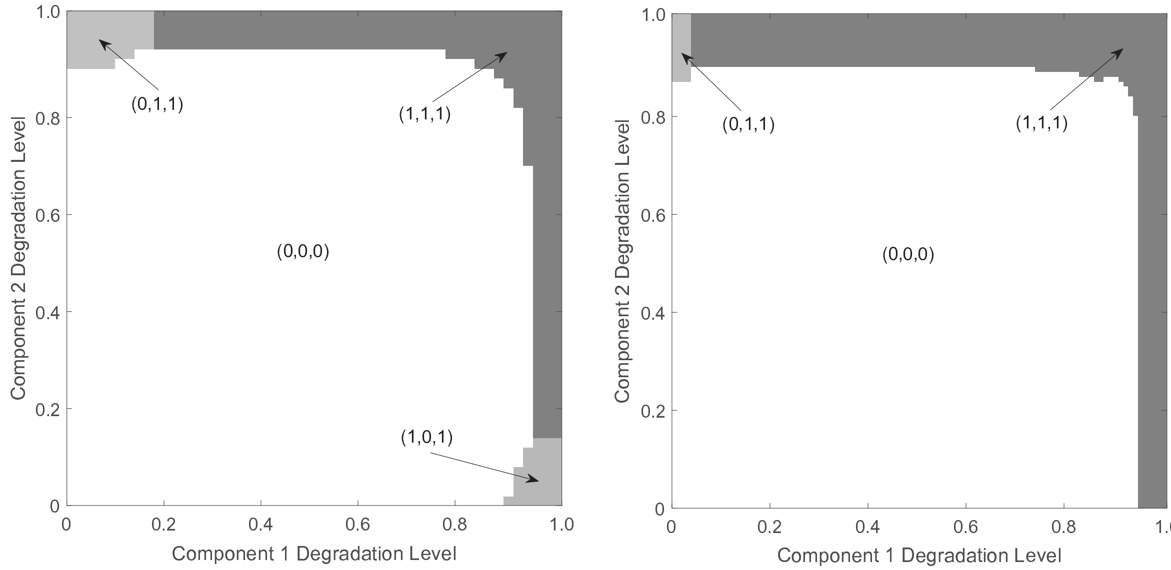
Figure 15: Comparison between optimal and learned value function in environment state 2.

tends to be near zero. Otherwise, the shape and values of the function approximation tend to be very similar to the true optimal value function. These results are typical and were observed across all test cases.

Figure 16 depicts the optimal and learned policies when the environment state is fixed at 1 and the degradation level of system 3 is fixed at 0.40. Again we observe that the optimal policy does not have a monotone structure. In contrast to the two-system case, we see that the policy is reasonably symmetric. This symmetry stems from the similarity between the degradation rates of the first two systems. In Figure 17, we compare the policies at environment state 2, noting that the regions where replacing two systems have largely vanished, but the symmetry is still observed. In both environment states, over a majority of the degradation states, the learned policy calls for replacing at least as many systems as the optimal policy.

Again, we evaluate the quality of the solutions by following each policy over 5,000 sufficiently long sample paths simulated using the initial state $(x, j) = (\mathbf{0}, 1)$. The results of this comparison are summarized in the box-and-whisker plot in Figure 18.

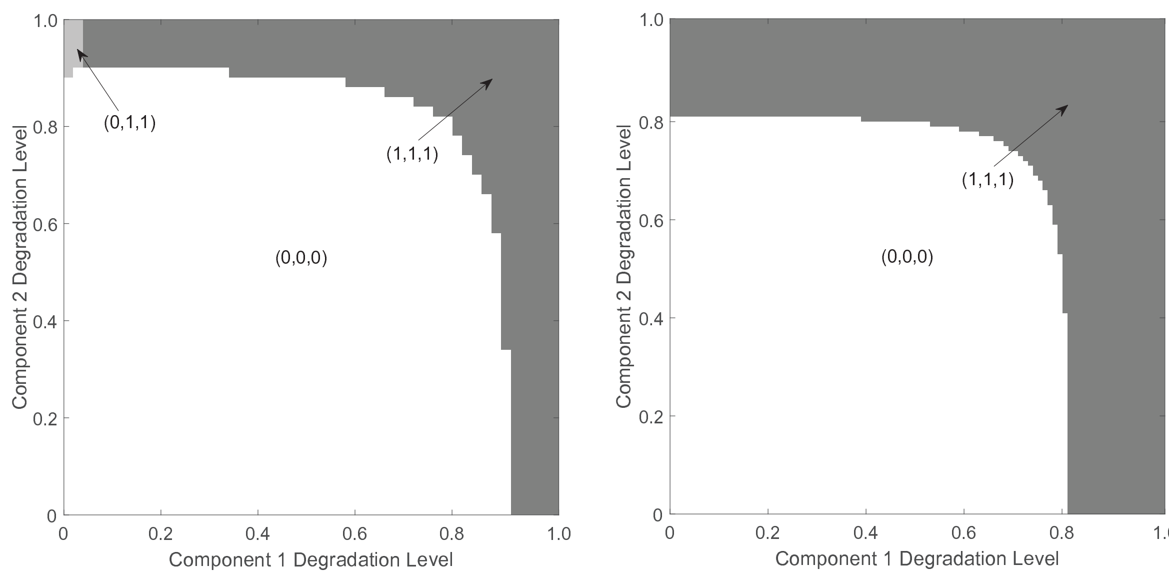
The learned policy again outperforms the heuristic policies, in this case, on average it outperforms the 1-policy by 51.89%, the 2-policy by 42.66%, and the 3-policy by 84.61%. The learned policy outperforms all of the heuristic policies on every sample path. Comparing the learned and



(a) Depiction of optimal policy.

(b) Depiction of learned policy.

Figure 16: Comparison between optimal and learned policy in environment state 1.



(a) Depiction of optimal policy.

(b) Depiction of learned policy.

Figure 17: Comparison between optimal and learned policy in environment state 2.

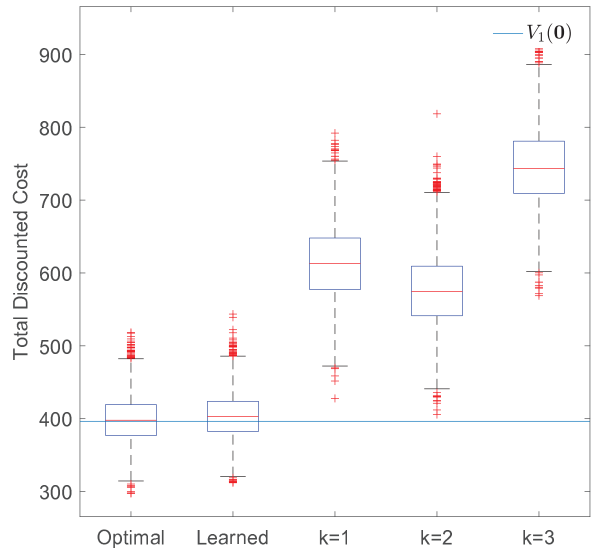


Figure 18: Boxplots comparing cost under different policies.

optimal policies, we note the average total discounted cost for the learned policy was 403.76 compared to 398.91 for the optimal policy (approximately 1.22% difference). The learned policy even outperformed the optimal policy on 1,854 sample paths (approximately 37%). Showing that for small $n > 1$ our approach is able to quickly determine near optimal policies.

Example. For this final example, we consider a larger problem for which finding an optimal solution by value iteration is computationally intractable. Here, there are $n = 50$ systems, the discount rate is $\alpha = 0.99$ and the cost vector is $\boldsymbol{c} = (c_0, c_p, c_r, c_d) = (100, 3, 10, 8)$. The failure thresholds are $\xi_i = 1$, for all i, and the environment $\mathcal{Z}$ has state space $S = \{1, 2, 3, 4\}$ with infinitesimal generator matrix

$$\mathcal{Q} = \begin{pmatrix} -5 & 5 & 0 & 0 \\ 2.5 & -5 & 2.5 & 0 \\ 0 & 2.5 & -5 & 2.5 \\ 0 & 0 & 5 & -5 \end{pmatrix}.$$

The inspection rate is $q = 10$ and degradation rates were randomly generated as $\boldsymbol{r}_i = (r_i^1, r_i^2, r_i^3, r_i^4) = U(1, 2) \cdot (2.5, 3, 3.5, 4)$, where $U(1, 2)$ is a continuous uniform random variable on $(1, 2)$.

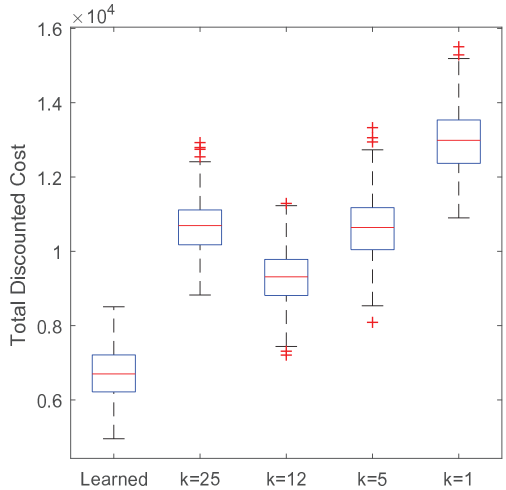


Figure 19: Boxplots comparing cost under different policies.

For this problem, the learned policy was trained for $60,000$ seconds, allowing for approximately $12,000,000$ decision epochs. Once the training was complete, the performance of the learned policy and the k-policies were compared over $5,000$ sample paths, beginning from state $(\boldsymbol{x}, j) = (\mathbf{0}, j)$. The results of this comparison are summarized in Figure 19.

Looking first at the k-policies, it appears that the expected total discounted cost is convex in k. This behavior occurs because, for large k, the downtime costs dominate the overall cost, yet for small k the setup costs are most significant. Even when the setup and downtime costs are most balanced ($k = 12$), the heuristic policy fails to outperform the learned policy. In fact, the learned policy outperforms all other policies sample-wise over all $5,000$ sample-paths. Specifically, the learned policy leads to an average cost savings of 38.05% when compared to the 12-policy and 92.61% when all dependency is ignored ($k = 1$).

4.0 MAINTAINING SYSTEMS WITH HETEROGENEOUS SPARE PARTS

4.1 INTRODUCTION

Operations and maintenance activities constitute a significant proportion of the operational budgets of many organizations [26]. Consequently, companies are employing increasingly sophisticated methods to improve system reliability, ensure safety, and reduce costs. A common assumption in maintenance models is that available spares (components or sub-systems) originate from a homogeneous population in which each spare has identical degradation characteristics. However, in reality, replacement parts may exhibit significant unit-to-unit variability in their quality characteristics. For instance, micro-electro-mechanical systems (MEMS) are known to suffer from a variety of local defects including particulate, ionic, organic, and isolated defects (e.g., voids and stringers) that can cause substantial unit-to-unit variability [39]. More generally, multiple device qualities can stem from manufacturing processes that are still in early stages of development and, therefore, highly variable [108]. Irrespective of the source of heterogeneity, ignoring this variability can reduce overall system reliability and lead to economic losses [19].

In this chapter, we consider the problem of optimally maintaining a stochastically degrading, single-unit system with heterogeneous spare parts. Specifically, the spare parts originate from Y distinct and heterogeneous subpopulations, each of which has its own time-to-failure distribution. Failures are not self-announcing; therefore, the system is inspected periodically to determine its status (functioning or failed). The system continues in operation until it is either preventively or correctively maintained. The available maintenance options include perfect repair, which restores the system to an as-good-as-new condition, or replacement of the system with a randomly-selected unit from the lot of heterogeneous spares. The primary advantage of this framework is that over time, as a system remains in operation, the subpopulation from which it originates becomes more apparent. Consequently, we are able to make better-informed maintenance decisions, thereby reducing long-run maintenance costs. In contrast to other models that consider spare part heterogeneity, we are able to update our beliefs, even in the absence of detailed condition monitoring information, based on the system's age alone. Assuming an intuitive cost structure that includes inspection costs, preventive maintenance costs, and corrective maintenance costs, our objective is to obtain a cost-minimizing policy that accounts for preventive and corrective maintenance decisions. To this end, we formulate an infinite-horizon, discounted MOMDP model and establish important properties of the cost function and optimal policy.

Within the applied probability and operations research communities, maintenance optimization models have been extensively studied over the last five decades. Many existing surveys highlight some of the most prevalent models for both single- and multi-component systems [6, 26, 64, 73, 75, 86, 88, 90, 99, 103]. Particularly relevant to our work here is the survey of Valdez-Flores and Feldman [99], which reviews the maintenance optimization literature of single-unit systems and includes models for inspection, minimal repair, shock, and replacement. Over the past two decades, reliability and maintenance models have increasingly addressed the problem of population heterogeneity. A common method for handling heterogeneous populations is to eliminate low-quality subpopulation(s) before they enter field service through burn-in procedures. Burn-in procedures are tests engineered to stress and detect devices that are likely to incur early failures (infant mortality). For a general background on burn-in procedures and models, the reader is referred to [46, 57] and references contained therein. Mi [66, 67] explores joint problems of burn-in, maintenance, and repair when a system exhibits a bathtub-shaped failure rate function and characterize the corresponding optimal burn-in times and maintenance policies. While early burn-in models focused on lifetime-based burn-in (discarding units that have failed before the end of the burn-in period), recent burn-in models are concerned with degradation-based burn-in (discarding units that have a particular degradation level at the end of the burn-in period). For example, Ye et al. [110] consider a joint burn-in and maintenance problem when the degradation processes of two different subpopulations are modeled as Weiner processes with distinct drift parameters. Optimal burn-in and replacement policies are characterized for age- and block-replacement and shown to be effective when compared to traditional lifetime-based burn-in approaches. Xiang et al. [107] investigate the more general case of joint burn-in and preventive replacement when there are n subpopulations subject to stochastic degradation. Their framework is then extended to the case of burn-in under accelerated conditions (e.g., elevated voltage, humidity, and temperature) with condition-based maintenance (CBM).

In addition to burn-in methods, another fruitful research area related to heterogeneous populations is the modeling of degradation in the presence of unit-to-unit variability. A common strategy is to augment a standard degradation model by allowing for some (or all) of the model's parameters to be random (i.e., to differ between systems within the population). This strategy has been employed when the degradation is modeled as a Brownian motion process [14, 71, 105], a gamma process [58, 96], or when it assumes an exponential functional form [29]. For each of these model types, the primary goal is to provide some measure of remaining useful life. Although these frameworks find wide applicability in CBM applications, they typically require that the system deteriorates with time, the deterioration level is observable at any time, and the device fails when the degradation level reaches a particular threshold [104]. Some recent papers consider CBM strategies with heterogeneous populations (cf. [22, 29, 101]). In lieu of a true burn-in procedure or CBM, Zhang et al. [112] propose a joint inspection-replacement policy. Each system is allowed to enter service, but an early inspection is conducted to determine its health state, at which point the unit is either replaced (if it appears to be defective) or a preventive replacement time is determined based on the health state. They show that this inspection-replacement policy outperforms a joint burn-in and age-based-replacement policy.

The model we present here is distinguished from the existing literature by considering unit-to-unit variability in repairable systems. Within this paradigm, knowledge about the system's quality can be learned and leveraged without a well-defined notion of degradation or failure threshold, or even an observable degradation signal of any kind. Specifically, as the system continues in operation and undergoes repairs over time, we are able to glean valuable information about its population of origin based on its age alone. By way of Bayesian inference procedures, we update our understanding of the system's quality and make better-informed decisions that lead to demonstrable cost savings. Our main contributions include a novel MOMDP framework for accounting for spare part heterogeneity; providing conditions under which the optimal value function is monotone in the belief space; providing conditions under which the optimal policy calls for preventive maintenance; characterizing the optimal policy; and executing a computational study that demonstrates the utility of our proposed framework and provides additional insights into the optimal policy as a exploration/exploitation type policy.

The remainder of this chapter is organized as follows. In Section 4.2, we present the maintenance problem when unit-to-unit variability is significant and formulate a mathematical model of the corresponding sequential decision-making process. Section 4.3 discusses attributes of the value function and the optimal policy. Finally, in Section 4.4, we provide a detailed computational study that illustrates the importance of accounting for heterogeneity, demonstrates the effectiveness of our modeling framework, and highlights some interesting properties of the optimal maintenance policy.

4.2 MODEL FORMULATION

Consider a single-unit, repairable system that begins operation in an as-good-as-new condition, that is, with an operating age of 0. The system continues functioning until it fails, i.e., when its operating age exceeds a probabilistically determined time-to-failure, or until a maintenance action is taken preventively. In what follows, all random variables are defined on a common, complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

It is assumed that the system originates from a lot of spare systems that, despite being visually indistinguishable, have distinct time-to-failure distributions. The lot of spare systems is composed of Y $(Y > 1)$ distinct qualities; for convenience, let $\mathcal{Y} := \{1, 2, \ldots, Y\}$. To ensure that the proportion of each type of system remains constant, we assume that the lot of spare systems is infinitely large. Additionally, we assume that the proportion of systems is known *a priori* and given by the vector $\boldsymbol{\rho} := (\rho_1, \rho_2, \ldots, \rho_Y)$, where ρ_y denotes the proportion of systems of quality $y \in \mathcal{Y}$ and

$$\sum_{y \in \mathcal{Y}} \rho_y = 1.$$

When a system is placed into service, it is able to operate until its (random) time-to-failure T. We denote the distribution function (d.f.) of T by $F(t) := \mathbb{P}(T \leq t)$, $t \geq 0$, and define $\bar{F}(t)$ as the complementary d.f., or survival function $\bar{F}(t) = 1 - F(t)$. Additionally, we assume that each quality has its own (known) failure distribution. We let Q denote the (random) quality of the system; hence, $\rho_y = \mathbb{P}(Q = y)$. The conditional time-to-failure distribution, given $Q = y$, is denoted by $F_y(t) := \mathbb{P}(T \leq t | Q = y)$ and its survival function will be denoted by $\bar{F}_y$. Hence, whenever a new system enters service, the time-to-failure distribution is given by the mixture distribution

$$F(t) = \sum_{y \in \mathcal{Y}} \rho_y F_y(t), \quad t \geq 0.$$

The system is inspected periodically with fixed period τ $(\tau > 0)$, that is, the system is inspected at the times in the set $\{\tau, 2\tau, 3\tau, \ldots\}$. At each inspection epoch, the system is observed to be in one of two states in the set $\mathcal{O} = \{0, 1\}$, where state 0 means the system is failed, and state 1 means the system is functioning. If the system is found to be in state 1, three actions are feasible: do nothing, repair, or replace. We define the set of feasible actions by $\mathcal{A} = \{0, 1, 2\}$, where 0, 1, and 2 denote do nothing, repair, and replace the system, respectively. On the other hand, if the system is found to be in state 0, only repair or replacement are permitted. Whenever a system is repaired, the same system reenters service with a virtual age of 0, and its quality remains unaltered. If a system is replaced, a new system is randomly selected from the lot of spare systems and enters service.

Due to the indistinguishable nature of the systems, the system quality is not known with certainty. Therefore, we define a vector $\boldsymbol{b} = (b_1, b_2, \ldots, b_Y)$, where b_y denotes the current belief, or probability, that $Q = y$. In other words, $\boldsymbol{b}$ is a probability distribution on the support $\mathcal{Y}$. When a new system first enters service, $\boldsymbol{b} = \boldsymbol{\rho}$, and the vector $\boldsymbol{b}$ is updated over time as the system operates. Thus, at each inspection epoch, the state of the system is described by the tuple $(x, \boldsymbol{b}) = (x, b_1, \ldots, b_Y)$, where $x\tau$ is the current age of the system. For convenience, we write $x = \infty$ when the system is failed. Hence, the complete state space is given by $\mathcal{S} = \mathcal{X} \times \mathcal{Y}$, where $\mathcal{X} = \mathbb{N} \cup \{0, \infty\}$.

Denote the (random) virtual age, quality, observation, and action taken at the nth inspection time by X_n, Y_n, O_n, and A_n, respectively. At each inspection, the system is in state $S_n = (X_n, Y_n) \in \mathcal{S}$, action $A_n \in \mathcal{A}$ is taken, the state transitions to (X_{n+1}, Y_{n+1}), and the observation $O_n \in \mathcal{O}$ is received (corresponding to the new state). A cost $C(S_n, A_n)$ is incurred and the process repeats indefinitely. The cost function, C, accounts for any pertinent maintenance and downtime costs associated with taking action A_n while being in state S_n; for the moment, we leave the cost function unspecified. We seek to minimize the total expected discounted costs over an infinite time horizon. Because the state is factorable into a fully observable (age) and partially observable (quality) component, we formulate our sequential decision problem using a mixed-observability Markov decision process (MOMDP) model [70].

More formally, the MOMDP is specified by the tuple $(\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{O}, P_{\mathcal{X}}, P_{\mathcal{Y}}, Z, C, \alpha)$, where $\alpha \in (0, 1)$ is the per-period discount factor and $P_{\mathcal{X}}$, $P_{\mathcal{Y}}$, and Z are functions describing the transition dynamics of the system. Specifically,

$$P_{\mathcal{X}}(x, y, a, x') = \mathbb{P}(X_{n+1} = x' \mid X_n = x, Y_n = y, A_n = a)$$

is the probability that the virtual age transitions to x' when action a is taken starting in state (x, y),

$$P_{\mathcal{Y}}(x, y, a, x', y') = \mathbb{P}(Y_{n+1} = y' \mid X_n = x, Y_n = y, A_n = a, X_{n+1} = x')$$

gives the probability that the partially observable (quality) state transitions to y' when action a is taken from state (x, y) and the fully observable state transitions to state x', and

$$Z(x', y', a, o) = \mathbb{P}(O_n = o | X_{n+1} = x', Y_{n+1} = y', A_n = a)$$

gives the probability that we receive observation o if the system transitions to state (x', y') after taking action a. Next, we define several functions that play critical roles in describing the problem's structure. Let $\bar{g}(t, y; \tau) = \mathbb{P}(T_y \geq t + \tau | T_y \geq t)$, then

$$\bar{g}(t, y; \tau) = \frac{\mathbb{P}(T_y \geq t + \tau, T_y \geq \tau)}{\mathbb{P}(T_y \geq t)} = \frac{\bar{F}_y(t + \tau)}{\bar{F}_y(t)}.$$

Because the inter-inspection time τ is fixed, we can ignore the dependence of $\bar{g}$ on τ and simply write $\bar{g}(t, y)$. Additionally, we define $g(t, y) = 1 - \bar{g}(t, y)$, $\bar{G}(t, \boldsymbol{b}) = \mathbb{P}(T \geq t + \tau | T \geq t, Q \sim \boldsymbol{b})$, and $G(t, \boldsymbol{b}) = 1 - \bar{G}(t, \boldsymbol{b})$. By the law of total probability, we see that

$$\bar{G}(t, \boldsymbol{b}) = \sum_{y \in \mathcal{Y}} \mathbb{P}(T_y \geq t + \tau | T_y \geq t) \mathbb{P}(Q = y | Q \sim \boldsymbol{b}) = \sum_{y \in \mathcal{Y}} \bar{g}(t, y) b_y.$$

If no maintenance action is taken on a system in state (x, y) with $x < \infty$, it is clear that the virtual age will transition to $(x+1)\tau$ if $T_y > (x+1)\tau$; otherwise, it will transition to ∞. Therefore, for $x < \infty$,

$$P_{\mathcal{X}}(x, y, 0, x') = \begin{cases} \bar{g}(x\tau, y), & x' = x + 1, \\ 1 - \bar{g}(x\tau, y), & x' = \infty, \\ 0, & \text{otherwise.} \end{cases}$$

However, if the system is repaired or replaced, then the age is returned to 0; hence,

$$P_{\mathcal{X}}(x, y, 1, x') = P_{\mathcal{X}}(x, y, 2, x') = \begin{cases} 1, & x' = 0, \\ 0, & \text{otherwise.} \end{cases}$$

If the system is allowed to continue operating, or it is repaired, then its quality remains unaltered.

However, if it is replaced, the new quality is again distributed according to $\boldsymbol{\rho}$. Consequently,

$$P_{\mathcal{Y}}(x, y, 0, x', y') = P_{\mathcal{Y}}(x, y, 1, x', y') = \begin{cases} 1, & y = y', \\ 0, & y \neq y', \end{cases}$$

and

$$P_{\mathcal{Y}}(x, y, 2, x', y') = \rho_{y'}.$$

Lastly, we note that $O_n = 1$ if and only if $X_{n+1} < \infty$; hence, $Z(x', y', a, 1) = \mathbb{I}(x' < \infty) = 1 - Z(x', y', a, 0)$. Upon taking action a and receiving observation o, we update the belief vector using Bayes' theorem as follows:

$$\begin{aligned} b'(y'|x, \boldsymbol{b}, x', a, o) &= \mathbb{P}(Q_{n+1} = y'|X_n = x, Q_n \sim \boldsymbol{b}, X_{n+1} = x', A_n = a, O_n = o) \\ &= \eta Z(x', y', a, o) \sum_{y \in \mathcal{Y}} P_{\mathcal{Y}}(x, y, a, x', y') P_{\mathcal{X}}(x, y, a, x') b_y, \end{aligned}$$

where, for any fixed $(x, \boldsymbol{b}, x', a, o)$, η is a normalizing constant. Of particular interest, for $x \in \mathbb{N}$, we have

$$\bar{B}_y(x, \boldsymbol{b}) = b'(y|x, \boldsymbol{b}, x+1, 0, 1) = \left(\bar{G}(x\tau, \boldsymbol{b})\right)^{-1} \bar{g}(x\tau, y) b_y, \tag{4.1}$$

and

$$B_y(x, \boldsymbol{b}) = b'(y|x, \boldsymbol{b}, \infty, 0, 0) = \left(G(x\tau, \boldsymbol{b})\right)^{-1} g(x\tau, y) b_y,$$

for the cases when a working component is allowed to continue operating and either survives or fails, respectively. We then define $\bar{B}(x, \boldsymbol{b}) := (\bar{B}_1(x, \boldsymbol{b}), \ldots, \bar{B}_Y(x, \boldsymbol{b}))$ and $B(x, \boldsymbol{b}) := (B_1(x, \boldsymbol{b}), \ldots, B_Y(x, \boldsymbol{b}))$ as the updated belief vectors when the system survives or fails, respectively.

We assume that at each inspection, a fixed inspection cost c_I is incurred regardless of the action taken. If the system is repaired, a fixed cost c_R is incurred, but if it is replaced, a cost of $c_P > c_R$ is incurred. Lastly, if the maintenance is corrective, i.e., $x = \infty$, an additional penalty of c_F is incurred. This additional cost, c_F, reflects the fact that corrective maintenance may include additional costs such as lost production or overtime labor. The optimal total expected discounted cost, starting in belief state $(x, \boldsymbol{b})$ and denoted by $V(x, \boldsymbol{b})$, is given as a solution to the Bellman optimality equations

$$V(x,\boldsymbol{b}) = \begin{cases} \min\begin{cases} c_I + c_F + c_P + \alpha V(0,\boldsymbol{\rho}), \\ c_I + c_F + c_R + \alpha V(0,\boldsymbol{b}), \end{cases} & x = \infty, \\ \\ \min\begin{cases} c_I + c_P + \alpha V(0,\boldsymbol{\rho}), \\ c_I + c_R + \alpha V(0,\boldsymbol{b}), \\ c_I + \alpha[\bar{G}(x\tau,\boldsymbol{b})V(x+1,\bar{B}(x+1,\boldsymbol{b})) \\ \qquad\qquad +G(x\tau,\boldsymbol{b})V(\infty,B(x+1,\boldsymbol{b}))], \end{cases} & x < \infty. \end{cases} \tag{4.2}$$

The optimal action (or decision) in state $(x,\boldsymbol{b})$ is denoted by $d^*(x,\boldsymbol{b})$ so that $d^*(x,\boldsymbol{b}) = 0$ if it is optimal to do nothing, $d^*(x,\boldsymbol{b}) = 1$ if it is optimal to repair, and $d^*(x,\boldsymbol{b}) = 2$ if it is optimal to replace.

4.3 STRUCTURAL RESULTS

In this section, we examine the attributes of the value function and optimal policy of the MOMDP model formulated in Section 4.2. We begin by presenting several stochastic orders that are used throughout our exposition.

Definition 4.1 (Usual Stochastic Order) *For two random variables X and Y, we say X is smaller than Y in the usual stochastic order (denoted $X \leq_{st} Y$) if*

$$\mathbb{P}(X > x) \leq \mathbb{P}(Y > x)$$

for all $x \in (-\infty,\infty)$.

It should be noted that, given a nondecreasing function ϕ, and assuming existence of the expectations, $X \leq_{st} Y$ implies $\mathbb{E}[\phi(X)] \leq \mathbb{E}[\phi(Y)]$.

Definition 4.2 (Hazard Rate Order) *For two random variables X and Y, we say X is smaller than Y in the hazard rate order (denoted $X \leq_{hr} Y$) if*

$$\mathbb{P}(X > x)\mathbb{P}(Y > y) \geq \mathbb{P}(X > y)\mathbb{P}(Y > x)$$

for all real numbers x and y with $x \leq y$.

Definition 4.3 (Likelihood Ratio Order) *For two random variables X and Y, we say X is smaller than Y in the likelihood ratio order (denoted $X \leq_{lr} Y$) if*

$$\mathbb{P}(X \in A)\mathbb{P}(Y \in B) \geq \mathbb{P}(X \in B)\mathbb{P}(Y \in A)$$

for all measurable sets A and B such that $A \subseteq B$.

Without proof, we next state a well-known result that relates these three types of stochastic orders. For additional details, the reader is referred to Shaked and Shanthikumar [84].

Theorem 4.1 *Suppose X and Y are two continuous or discrete random variables.*

1. *If $X \leq_{lr} Y$, then $X \leq_{hr} Y$.*
2. *If $X \leq_{hr} Y$, then $X \leq_{st} Y$.*

Theorem 4.1 formalizes the notion that the likelihood ratio order is stronger than the hazard rate order, which is stronger than the usual stochastic order. For convenience, if X and Y are discrete random variables with respective probability mass functions p and q, and if X and Y are ordered in a particular sense, we also say that p and q are ordered in the same sense. Furthermore, we take the converse of the statement to be true, e.g., $X \leq_{hr} Y$, if, and only if, $p \leq_{hr} q$. Next, we define the n-simplex, i.e., the space of probability mass functions on n outcomes.

Definition 4.4 (n-Simplex) *The standard n-simplex Δ^n is the simplex formed from the n standard unit vectors. That is,*

$$\Delta^n = \left\{ (x_1, \ldots, x_n) \in \mathbb{R}^n : \sum_{i=1}^{n} x_i = 1,\ x_i \geq 0 \text{ for all } i \right\}.$$

We also refer to Δ^n as the n-dimensional probability simplex.

Using this definition the belief state space can then be defined as $\Gamma = \mathcal{X} \times \Delta^Y$.

The results of this section require that various conditions be met. We next present these conditions and their interpretations before proceeding to the main structural results.

Condition 1 *The function $\bar{g}(t, y)$ is jointly monotone nonincreasing in t and y.*

Requiring $\bar{g}(t, y)$ to be monotone in y is necessary for meaningfully comparing the performance of systems of different qualities. In particular, when $i < j$, at any decision epoch, a system of quality i is more likely to survive until the next decision epoch than a system of quality j. In this sense, a system from lot 1 may be regarded as being of the highest quality, and a system from lot Y of the lowest quality. This condition is equivalent to requiring that $T_y \geq_{hr} T_{y+1}$ by the equivalence of

(1.B.4) and (1.B.7) in [84]. Requiring $\bar{g}(t, y)$ to be monotone in t means that for a fixed quality, as a system ages, its remaining life (stochastically) decreases. This is equivalent to saying that T_y is increasing failure rate (IFR).

Condition 2 *For each $y \in \mathcal{Y}$,*

$$\lim_{x\to\infty} \bar{g}(x\tau, y) = 0.$$

Condition 2 is a type of short-tail condition. In particular, it says that regardless of the system quality, as the system ages, it will eventually become so degraded that it will have zero probability of surviving the inter-inspection period.

Condition 3 *For each $y \neq 1$,*

$$\lim_{x\to\infty} \frac{\bar{g}(x\tau, y)}{\bar{g}(x\tau, 1)} = 0.$$

Condition 3 can be interpreted to mean that (in an asymptotic sense) lower-quality systems degrade more rapidly than the highest quality system.

We are now prepared to state our first result. Under Condition 1, if we observe that the system survives an inter-inspection period, then the updated belief state is smaller, in the likelihood ratio, than the initial belief state. Similarly, if we observe that the system fails during an inter-inspection period, then the updated belief state is larger in the likelihood ratio than the initial belief state.

Proposition 4.1 *Under Condition 1, for $x \in \mathcal{X}$ and $\boldsymbol{b} \in \Delta^Y$*

$$\bar{B}(x, \boldsymbol{b}) \leq_{lr} \boldsymbol{b} \leq_{lr} B(x, \boldsymbol{b}).$$

Proof. First, we show that $\bar{B}(x, \boldsymbol{b}) \leq_{lr} \boldsymbol{b}$. To this end, we must show $b_y/\bar{B}_y(x, \boldsymbol{b})$ is nondecreasing in $y \in \mathcal{Y}$. By equation (4.1), we have

$$\frac{b_y}{\bar{B}_y(x, \boldsymbol{b})} = \frac{\bar{G}(x\tau, \boldsymbol{b})b_y}{\bar{g}(x\tau, y)b_y} = \frac{\bar{G}(x\tau, \boldsymbol{b})}{\bar{g}(x\tau, y)}.$$

The proof is completed by noting that Condition 1 implies $\bar{g}(t, y)$ is monotone nonincreasing in $y \in \mathcal{Y}$. The proof for $\bar{B}(x, \boldsymbol{b}) \leq_{lr} \boldsymbol{b}$ follows similarly by noting that $g(t, y) = 1 - \bar{g}(t, y)$ is monotone nondecreasing in $y \in \mathcal{Y}$. ∎

We next show that under Condition 1, for any fixed belief state, the probability of the system surviving through the inter-inspection period decreases as the virtual age increases; consequently, the probability of failure increases.

Proposition 4.2 *Under Condition 1, for* $\boldsymbol{b} \in \Delta^Y$ *and* $x \in \mathcal{X} \setminus \{\infty\}$,

$$\bar{G}(x\tau, \boldsymbol{b}) \geq \bar{G}((x+1)\tau, \boldsymbol{b}) \quad \textit{and} \quad G(x\tau, \boldsymbol{b}) \leq G((x+1)\tau, \boldsymbol{b}).$$

Proof. We will first show that $\bar{G}(x\tau, \boldsymbol{b}) \geq \bar{G}((x+1)\tau, \boldsymbol{b})$, noting that $G(x\tau, \boldsymbol{b}) \leq G((x+1)\tau, \boldsymbol{b})$ follows since $G(t, \boldsymbol{b}) = 1 - \bar{G}(t, \boldsymbol{b})$. By Condition 1, for each $x \in \mathcal{X} \setminus \{\infty\}$,

$$h_1(y) := \bar{g}(x\tau, y) \geq \bar{g}((x+1)\tau, y) := h_2(y)$$

for all $y \in \mathcal{Y}$. Because the functions h_1 and h_2 are ordered for all y,

$$h_1(Q) \geq_{st} h_2(Q). \tag{4.3}$$

Consequently, by taking expectations on across both sides of inequality (4.3),

$$\bar{G}(x\tau, \boldsymbol{b}) = \mathbb{E}(h_1(Q)) \geq \mathbb{E}(h_2(Q)) = \bar{G}((x+1)\tau, \boldsymbol{b}),$$

where $Q \sim \boldsymbol{b}$. ∎

The next result asserts that, under Condition 1, if two states are ordered such that the belief state is ordered in the sense of the usual stochastic order, and the virtual ages are ordered, then the probabilities of system failure are ordered in the same direction. That is, systems with older virtual ages and stochastically larger belief states are more likely to fail.

Proposition 4.3 *Under Conditions 1 and 2, if* $\boldsymbol{b}_1 \leq_{st} \boldsymbol{b}_2$ *and* $x_1 \leq x_2$ *then*

$$\bar{G}(x_1\tau, \boldsymbol{b}_1) \geq \bar{G}(x_2\tau, \boldsymbol{b}_2) \quad \textit{and} \quad G(x_1\tau, \boldsymbol{b}_1) \leq G(x_2\tau, \boldsymbol{b}_2).$$

Proof. We again proceed by first showing $\bar{G}(x_1\tau, \boldsymbol{b}_1) \geq \bar{G}(x_2\tau, \boldsymbol{b}_2)$, noting this implies $G(x_1\tau, \boldsymbol{b}_1) \leq G(x_2\tau, \boldsymbol{b}_2)$. By Condition 1, we have that $\bar{g}(x_1\tau, y)$ is nonincreasing in $y \in \mathcal{Y}$. Thus, if $Q_i \sim \boldsymbol{b}_i$, $i = 1, 2$, then

$$\bar{G}(x_1\tau, \boldsymbol{b}_1) = \mathbb{E}(\bar{g}(x_1\tau, Q_1)) \geq \mathbb{E}(\bar{g}(x_1\tau, Q_2)) = \bar{G}(x_2\tau, \boldsymbol{b}_2)$$

where the inequality holds by $\boldsymbol{b}_1 \leq_{st} \boldsymbol{b}_2$. The result follows by repeated application of Proposition 4.2. ∎

Next, assume that we have two distributions over the quality of the system in operation that are ordered in the likelihood ratio sense. Under the same observation, i.e., the system is found to be functioning or failed at a particular time, the distributions remain ordered after being updated. This is formalized in Proposition 4.4.

Proposition 4.4 *Under Condition 1, if* $\boldsymbol{b}_1 \leq_{lr} \boldsymbol{b}_2$ *and* $x \in \mathcal{X}$ *then*

$$\bar{B}(x, \boldsymbol{b}_1) \leq_{lr} \bar{B}(x, \boldsymbol{b}_2) \quad \textit{and} \quad B(x, \boldsymbol{b}_1) \leq_{lr} B(x, \boldsymbol{b}_2).$$

Proof. Again, we proceed by only showing that $\bar{B}(x, \boldsymbol{b}_1) \leq_{lr} \bar{B}(x, \boldsymbol{b}_2)$. For $i = 1, 2$, let $\boldsymbol{b}_y^i$ be the yth component of $\boldsymbol{b}_i$. Then,

$$\frac{\bar{B}_y(x, \boldsymbol{b}_2)}{\bar{B}_y(x, \boldsymbol{b}_1)} = \frac{[\bar{G}(x\tau, \boldsymbol{b}_2)]^{-1}}{[\bar{G}(x\tau, \boldsymbol{b}_1)]^{-1}} \frac{\bar{g}(x\tau, y)}{\bar{g}(x\tau, y)} \frac{\boldsymbol{b}_y^2}{\boldsymbol{b}_y^1} = \frac{\bar{G}(x\tau, \boldsymbol{b}_1)}{\bar{G}(x\tau, \boldsymbol{b}_2)} \frac{\boldsymbol{b}_y^2}{\boldsymbol{b}_y^1},$$

but $\boldsymbol{b}_y^2/\boldsymbol{b}_y^1$ is increasing in y by assumption; hence, $\bar{B}_y(x, \boldsymbol{b}_2)/\bar{B}_y(x, \boldsymbol{b}_1)$ is increasing. ∎

We are now prepared to state our first main result. Theorem 4.2 asserts that, under Condition 1, for a fixed virtual age, the value function is monotone in the belief state. Additionally, for any fixed belief state, the value function is largest when the system is in the failed state.

Theorem 4.2 *Under Conditions 1, if* $x \in \mathcal{X}$ *and* $\boldsymbol{b}_1 \leq_{lr} \boldsymbol{b}_2$, *then*

$$V(x, \boldsymbol{b}_1) \leq V(x, \boldsymbol{b}_2)$$

and if $\boldsymbol{b} \in \Delta^Y$ *then*

$$c_F + V(x, \boldsymbol{b}) \leq V(\infty, \boldsymbol{b}).$$

Proof. We proceed by induction on the iterates of the value iteration algorithm. Let v^k be the approximation of the value function V at the kth iteration and assume $v^0(x, \boldsymbol{b}) = 0$ for all $(x, \boldsymbol{b}) \in \mathcal{X} \times \Delta^Y$. Then,

$$v^1(x, \boldsymbol{b}) = \begin{cases} c_I + c_F + c_R, & x = \infty, \\ c_I, & x < \infty; \end{cases}$$

hence, $v^1(x, \boldsymbol{b})$ is constant in $\boldsymbol{b}$ and $v^1(x, \boldsymbol{b}) + c_F < v^1(\infty, \boldsymbol{b})$. Now, assume that $v^k(x, \boldsymbol{b})$ satisfies the induction hypothesis, i.e., if $x \in \mathcal{X}$ and $\boldsymbol{b}_1 \leq_{lr} \boldsymbol{b}_2$ then $v^k(x, \boldsymbol{b}_1) \leq v^k(x, \boldsymbol{b}_2)$, and if $b \in \Delta^Y$ then $c_F + v^k(x, \boldsymbol{b}) \leq v^k(x, \boldsymbol{b})$. For notational convenience, we let $C_P^k = c_I + c_P + \alpha v^k(0, \boldsymbol{\rho})$, $C_R^k(\boldsymbol{b}) = c_I + c_R + \alpha v^k(0, \boldsymbol{b})$, and $C_{DN}^k(x, \boldsymbol{b}) = c_I + \alpha(\bar{G}(x, \boldsymbol{b}) v^k(x+1, \bar{B}(x+1, \boldsymbol{b})) + G(x, \boldsymbol{b}) v^k(\infty, B(x+1, \boldsymbol{b}))$. Then, when $x = \infty$, we have

$$v^{k+1}(\infty, \boldsymbol{b}) = \min\{c_F + C_P^k,\, c_F + C_R^k(\boldsymbol{b})\}.$$

By the induction hypothesis, $C_R^k(\boldsymbol{b})$ is monotone nondecreasing in $\boldsymbol{b}$, and since C_P^k is constant in $\boldsymbol{b}$, it is clear that $v^{k+1}(\infty, \boldsymbol{b})$ is also monotone nondecreasing in $\boldsymbol{b}$.

Next, for $x < \infty$,

$$v^{k+1}(x, \boldsymbol{b}) = \min\{C_P^k,\, C_R^k(\boldsymbol{b}),\, C_{DN}^k(x, \boldsymbol{b})\} \le \min\{C_P^k,\, C_R^k(\boldsymbol{b})\};$$

hence, $v^{k+1}(x, \boldsymbol{b}) + c_F \le c_F + \min\{C_P^k, C_R^k(\boldsymbol{b})\} = v^{k+1}(\infty, \boldsymbol{b})$. To complete the proof, we need to show that $C_{DN}^k(x, \boldsymbol{b})$ is monotone nondecreasing in $\boldsymbol{b}$. Let, $\boldsymbol{b}_1 \le_{lr} \boldsymbol{b}_2$, and, for $i = 1, 2$, let D_i be a random variable such that

$$\mathbb{P}(D_i = q) = \begin{cases} \bar{G}(x, \boldsymbol{b}_i), & q = 0, \\ G(x, \boldsymbol{b}_i), & q = 1, \end{cases}$$

and let $h_i^k(q|x)$ be a function such that

$$h_i^k(q|x) = \begin{cases} v^k(x+1, \bar{B}(x+1, \boldsymbol{b}_i)), & q = 0, \\ v^k(\infty, B(x+1, \boldsymbol{b}_i)), & q = 1. \end{cases}$$

Because $\boldsymbol{b}_1 \le_{lr} \boldsymbol{b}_2$ implies $\boldsymbol{b}_1 \le_{st} \boldsymbol{b}_2$, Proposition 4.2 shows that $G(x, \boldsymbol{b}_1) \le G(x, \boldsymbol{b}_2)$; consequently, $D_1 \le_{st} D_2$. Additionally, by Proposition 4.4, $\bar{B}(x+1, \boldsymbol{b}_1) \le_{lr} \bar{B}(x+1, \boldsymbol{b}_2)$ and $B(x+1, \boldsymbol{b}_1) \le_{lr} B(x+1, \boldsymbol{b}_2)$; therefore, by the induction hypothesis, $h_1^k(q|x) \le h_2^k(q|x)$ for $q = 0, 1$. We can see then that

$$C_{DN}^k(x, \boldsymbol{b}_1) = c_I + \alpha\, \mathbb{E}(h_1^k(D_1|x)) \le c_I + \alpha\, \mathbb{E}(h_2^k(D_1|x)).$$

Next, we note

$$\begin{aligned} h_2^k(0|x) = v^k(x+1, \bar{B}(x+1, \boldsymbol{b}_2)) \le V^k(x+1, \boldsymbol{b}_2) &\le v^k(x+1, B(x+1, \boldsymbol{b}_2)) \\ &\le v^k(\infty, B(x+1, \boldsymbol{b}_2)) \\ &= h_2^k(1|x), \end{aligned}$$

where the first two inequalities follow from the induction hypothesis and Proposition 4.1, and the last inequality follows from the induction hypothesis. Thus, $h_2^k(q|x)$ is nondecreasing in q, so by $D_1 \le_{st} D_2$

$$C_{DN}^k(x, \boldsymbol{b}_1) \le c_I + \alpha\, \mathbb{E}(h_2^k(D_1|x)) \le c_I + \alpha\, \mathbb{E}(h_2^k(D_2|x)) = C_{DN}^k(x, \boldsymbol{b}_2),$$

and the proof is complete. ■

Our next result (Theorem 4.3) establishes that the belief state space can be partitioned into two regions: a region where either doing nothing or repair is optimal, and a region where either doing nothing or replacement is optimal. Moreover, these regions are formed by partitioning Δ^Y and are related to the likelihood ratio ordering.

Theorem 4.3 *Under Condition 1, if* $d^*(x, \boldsymbol{b}) = 2$*, then* $d^*(x', \boldsymbol{b}') \in \{0, 2\}$ *for all* $x' \in \mathbb{N}$ *and* $\boldsymbol{b}' \geq_{lr} \boldsymbol{b}$*. Additionally, if* $d^*(x, \boldsymbol{b}) = 1$*, then* $d^*(x', \boldsymbol{b}') \in \{0, 1\}$ *for all* $x' \in \mathbb{N}$ *and* $\boldsymbol{b}' \leq_{lr} \boldsymbol{b}$*.*

Proof. If $d^*(x, \boldsymbol{b}) = 2$, then $V(x, \boldsymbol{b}) = c_I + c_P + \alpha V(0, \boldsymbol{\rho})$; hence,

$$c_I + c_P + \alpha V(0, \boldsymbol{\rho}) < c_I + c_R + \alpha V(0, \boldsymbol{b}) \tag{4.4}$$

$$\leq c_I + c_R + \alpha V(0, \boldsymbol{b}'), \tag{4.5}$$

where the first inequality follows directly from (4.2), and the second inequality follows from Theorem 4.2. Therefore, repair is not optimal in state $(x', \boldsymbol{b}')$. The proof of the second statement follows by noting that $d^*(x, \boldsymbol{b}) = 1$ and $\boldsymbol{b}' \leq_{lr} \boldsymbol{b}$ results in a reversal of inequalities (4.4) and (4.5). ■

To further understand the structure of the optimal value function and policy, we consider their behavior when the quality is known with certainty. Let the set of vectors $\{\boldsymbol{e}_y : 1 \leq y \leq Y\}$ be the standard basis for Y-dimensional Euclidean space, where $\boldsymbol{e}_y$ denotes the vector with a one in the yth coordinate and zeros elsewhere. Our next result states that, when the quality is known with certainty, the value function is monotone nondecreasing in the virtual age.

Proposition 4.5 *Under Condition 1, for each* $y \in \mathcal{Y}$*, and* $x \in \mathcal{X} \setminus \{\infty\}$*,*

$$V(x, \boldsymbol{e}_y) \leq V(x + 1, \boldsymbol{e}_y).$$

Proof. Again, by induction, if $v^0(x, \boldsymbol{b}) = 0$ for all $(x, \boldsymbol{b}) \in \mathcal{X} \times \Delta^Y$, then again

$$v^1(x, \boldsymbol{b}) = \begin{cases} c_I + c_R + c_F, & x = \infty, \\ c_I, & x < \infty; \end{cases}$$

hence, the base case holds. Additionally, for each $x \in \mathcal{X} \setminus \{\infty\}$ and $y \in \mathcal{Y}$, we note that $v^1(x, \boldsymbol{e}_y) \leq v^1(\infty, \boldsymbol{e}_y)$. Now, assume that for each x, y that $v^k(x, \boldsymbol{e}_y) \leq v^k(x + 1, \boldsymbol{e}_y) \leq v^k(\infty, \boldsymbol{e}_y)$, which also holds in the base case, then by the proof of Theorem 4.2 we know this property is conserved under each iteration of the value iteration algorithm and holds for $k + 1$. Then, noting that

$\bar{G}(x, \boldsymbol{e}_y) = \bar{g}(x, y)$, $G(x, \boldsymbol{e}_y) = g(x, y)$, and $\bar{B}(x, \boldsymbol{e}_y) = B(x, \boldsymbol{e}_y) = \boldsymbol{e}_y$, we see

$$\begin{aligned}
v^{k+1}(x, \boldsymbol{e}_y) &= \min \begin{cases} c_I + \alpha v^k(0, \boldsymbol{\rho}) \\ c_I + \alpha v^k(0, \boldsymbol{e}_y) \\ c_I + \alpha \left(\bar{g}(x, y) v^k(x+1, \boldsymbol{e}_y) + g(x, y) v^k(\infty, \boldsymbol{e}_y) \right) \end{cases} \\
&\leq \min \begin{cases} c_I + \alpha v^k(0, \boldsymbol{\rho}) \\ c_I + \alpha v^k(0, \boldsymbol{e}_y) \\ c_I + \alpha \left(\bar{g}(x, y) v^k(x+2, \boldsymbol{e}_y) + g(x, y) v^k(\infty, \boldsymbol{e}_y) \right) \end{cases} \\
&\leq \min \begin{cases} c_I + \alpha v^k(0, \boldsymbol{\rho}) \\ c_I + \alpha v^k(0, \boldsymbol{e}_y) \\ c_I + \alpha \left(\bar{g}(x+1, y) v^k(x+2, \boldsymbol{e}_y) + g(x+1, y) v^k(\infty, \boldsymbol{e}_y) \right) \end{cases} \\
&= v^{k+1}(x+1, \boldsymbol{e}_y),
\end{aligned}$$

where the first inequality follows directly from the induction hypothesis, and the second inequality follows by noting that $v^k(x+2, \boldsymbol{e}_y) \leq v^k(\infty, \boldsymbol{e}_y)$ and $g(x, y) \leq g(x+1, y)$ (using a stochastic ordering argument similar to that in the proof of Theorem 4.2). ∎

Additionally, for each $\boldsymbol{e}_y$, there exists a virtual age above which it is optimal to preventively maintain and below which it is optimal to do nothing. This result is formalized in Proposition 4.6.

Proposition 4.6 *If $d^*(x, \boldsymbol{e}_y) > 0$, then*

$$V(x+1, \boldsymbol{e}_y) = V(x, \boldsymbol{e}_y) = V(\infty, \boldsymbol{e}_y) - c_F,$$

and

$$d^*(x+1, \boldsymbol{e}_y) = d^*(x, \boldsymbol{e}_y).$$

Proof. If $d^*(x, \boldsymbol{e}_y) > 0$, then

$$\begin{aligned}
V(x, \boldsymbol{e}_y) &= \min\{c_I + c_P + \alpha V(0, \boldsymbol{\rho}), c_I + c_R + \alpha V(0, \boldsymbol{e}_y)\} \\
&\leq c_I + \alpha \left(\bar{g}(x, y) V(x+1, \boldsymbol{e}_y) + g(x, y) V(\infty, \boldsymbol{e}_y) \right) \qquad (4.6) \\
&\leq c_I + \alpha \left(\bar{g}(x, y) V(x+2, \boldsymbol{e}_y) + g(x, y) V(\infty, \boldsymbol{e}_y) \right) \qquad (4.7) \\
&\leq c_I + \alpha \left(\bar{g}(x+1, y) V(x+2, \boldsymbol{e}_y) + g(x+1, y) V(\infty, \boldsymbol{e}_y) \right), \qquad (4.8)
\end{aligned}$$

where inequality (4.6) follows by supposition that the optimal action is preventive maintenance, (4.7) follows by Proposition 4.5, and (4.8) follows by Condition 1. The result then follows immediately by noting that

$$V(x+1,\boldsymbol{e}_y) = \min\begin{cases}\min\{c_I + c_P + \alpha V(0,\boldsymbol{\rho}), c_I + c_R + \alpha V(0,\boldsymbol{e}_y)\} \\ c_I + \alpha\left(\bar{g}(x+1,y)V(x+2,\boldsymbol{e}_y) + g(x+1,y)V(\infty,\boldsymbol{e}_y)\right)\end{cases}$$

$$= \min\{c_I + c_P + \alpha V(0,\boldsymbol{\rho}), c_I + c_R + \alpha V(\infty,\boldsymbol{e}_y)\}$$

$$= V(x,\boldsymbol{e}_y).$$

■

Proposition 4.7, further characterizes the optimal policy and value function when the quality is known with certainty.

Proposition 4.7 *Suppose that Conditions 1 and 2 hold. For each $y \in \mathcal{Y}$,*

1. *if $d^*(x,\boldsymbol{e}_y) \in \{0,2\}$, then $d^*(x,\boldsymbol{e}_y) = 0$ for all $x < x_y^*$ and $d^*(x,\boldsymbol{e}_y) = 2$ for all $x \geq x_y^*$, where*

$$x_y^* = \operatorname{argmin}\left\{x \in \mathbb{N} : \bar{g}(x\tau,y) < \left(\frac{1-\alpha}{\alpha c_F}\right)\left(\frac{c_I + c_F}{1-\alpha} - (c_I + c_F + c_P) - V(0,\boldsymbol{\rho})\right)\right\},$$

2. *if $d^*(x,\boldsymbol{e}_y) \in \{0,1\}$, then $d^*(x,\boldsymbol{e}_y) = 0$ for all $x < x_y^*$ and $d^*(x,\boldsymbol{e}_y) = 1$ for all $x \geq x_y^*$, where*

$$x_y^* = \operatorname{argmin}\left\{x \in \mathbb{N} : \bar{g}(x\tau,y) < \left(\frac{1-\alpha}{\alpha c_F}\right)\left(\frac{c_I + c_F}{1-\alpha} - (c_I + c_F + c_R) - V(0,\boldsymbol{e}_y)\right)\right\},$$

3. *if $d^*(x,\boldsymbol{e}_y) \in \{0,1\}$, then $V(x,\boldsymbol{e}_y) = V(x)$ where $V(x)$ satisfies the following one-dimensional Bellman equations:*

$$V(x) = \min\begin{cases} c_I + c_P + \alpha V(0), \\ c_I + \alpha\left[\bar{g}(x\tau,y)V(x+1) + g(x\tau,y)\left(c_I + c_F + c_R + \alpha V(0)\right)\right].\end{cases}$$

Proof. We first prove Part 3. By the assumption that $d^*(x,\boldsymbol{e}_y) \in \{0,1\}$, and (4.2), we have

$$V(x,\boldsymbol{e}_y) = \min\begin{cases} c_I + c_R + \alpha V(0,\boldsymbol{e}_y), \\ c_I + \alpha\left[\bar{G}(x\tau,\boldsymbol{e}_y)V(x+1,\bar{B}(x+1,\boldsymbol{e}_y)) + G(x\tau,\boldsymbol{e}_y)V(\infty,B(x+1,\boldsymbol{e}_y))\right].\end{cases} \tag{4.9}$$

It can be seen that $\boldsymbol{e}_y = \bar{B}(x, \boldsymbol{e}_y) = B(x, \boldsymbol{e}_y)$, $\bar{G}(x\tau, \boldsymbol{e}_y) = \bar{g}(x\tau, y)$ and $G(x\tau, \boldsymbol{e}_y) = g(x\tau, y)$. Hence, $V(\infty, B(x+1, \boldsymbol{e}_y)) = V(\infty, \boldsymbol{e}_y) = c_I + c_F + c_R + \alpha V(0, \boldsymbol{e}_y)$ and (4.9) can be rewritten as

$$V(x, \boldsymbol{e}_y) = \min \begin{cases} c_I + c_R + \alpha V(0, \boldsymbol{e}_y), \\ c_I + \alpha \left[\bar{g}(x\tau, y)V(x+1, \boldsymbol{e}_y) + g(x\tau, y)(c_I + c_F + c_R + \alpha V(0, \boldsymbol{e}_y))\right], \end{cases} \tag{4.10}$$

completing the proof of Part 3. For Part 2, we see from (4.10) that $d^*(x, \boldsymbol{e}_y) = 1$ if, and only if,

$$c_I + c_R + \alpha V(0, \boldsymbol{e}_y) < c_I + \alpha \left[\bar{g}(x\tau, y)V(x+1, \boldsymbol{e}_y) + g(x\tau, y)(c_I + c_F + c_R + \alpha V(0, \boldsymbol{e}_y))\right]. \tag{4.11}$$

We note that if $d^*(x, \boldsymbol{e}_y) = 1$, then by Proposition 4.6 that $d^*(x+1, \boldsymbol{e}_y) = 1$, and, consequently, $V(x, \boldsymbol{e}_y) = V(x+1, \boldsymbol{e}_y) = c_I + c_R + \alpha V(0, \boldsymbol{e}_y)$. Hence, inequality (4.11) is equivalent to

$$c_I + c_R + \alpha V(0, \boldsymbol{e}_y) < c_I + \alpha \left[\bar{g}(x\tau, y)(c_I + c_R + \alpha V(0, \boldsymbol{e}_y)) + g(x\tau, y)(c_I + c_F + c_R + \alpha V(0, \boldsymbol{e}_y))\right]. \tag{4.12}$$

By noting that $g(x\tau, y) + \bar{g}(x\tau, y) = 1$, we rearrange the inequality to see that $d^*(x, \boldsymbol{e}_y) = 1$ if, and only if,

$$\bar{g}(x\tau, y) < \left(\frac{1-\alpha}{\alpha c_F}\right)\left(\frac{c_I + c_F}{1-\alpha} - (c_I + c_F + c_R) - V(0, \boldsymbol{e}_y)\right). \tag{4.13}$$

The proof is completed by noting that $\bar{g}(x\tau, y)$ is monotone nonincreasing in x, and converges to 0 (by Conditions 1 and 2, respectively). The proof of Part 1 is similar to that of Part 2. ■

Proposition 4.7, Part 3 is useful for computing the optimal policy at extreme points of Δ^Y where repair is optimal. Additionally, as shown in Corollary 4.1, Proposition 4.7, Part 1 is useful for bounding the age replacement threshold of the extreme points of Δ^Y for which repair is not optimal.

Corollary 4.1 *Under Conditions 1-2, for each $y \in \mathcal{Y}$, if $d^*(x, \boldsymbol{e}_y) \in \{0, 2\}$, and if there exist real numbers $\underline{V}$ and $\bar{V}$ such that $\underline{V} \le V(0, \boldsymbol{\rho}) \le \bar{V}$, then $\underline{x}_y \le x^*_y \le \bar{x}_y$, where*

$$\underline{x}_y = \operatorname{argmin}\left\{x \in \mathbb{N} : \bar{g}(x\tau, y) < \left(\frac{1-\alpha}{\alpha c_F}\right)\left(\frac{c_I + c_F}{1-\alpha} - (c_I + c_F + c_P) - \underline{V}\right)\right\}, \text{ and}$$

$$\bar{x}_y = \operatorname{argmin}\left\{x \in \mathbb{N} : \bar{g}(x\tau, y) < \left(\frac{1-\alpha}{\alpha c_F}\right)\left(\frac{c_I + c_F}{1-\alpha} - (c_I + c_F + c_P) - \bar{V}\right)\right\}.$$

There are many ways to bound $V(0, \boldsymbol{\rho})$, but some simple bounds are given by $V(0, \boldsymbol{e}_1) \leq V(0, \boldsymbol{\rho}) \leq \underline{V}(0, \boldsymbol{\rho})$ where $V(0, \boldsymbol{e}_1)$ can be computed as described in Proposition 4.7, Part 3 and $\underline{V}(0, \boldsymbol{\rho})$ is determined by solving the following one-dimensional Bellman equations:

$$\underline{V}(x, \boldsymbol{\rho}) = \min \begin{cases} c_I + c_P + \alpha \underline{V}(0, \boldsymbol{\rho}), \\ c_I + \alpha \left(\bar{G}(x\tau, \boldsymbol{\rho}) \underline{V}(x+1, \boldsymbol{\rho}) + \bar{G}(x\tau, \boldsymbol{\rho})(c_I + c_F + c_P + \alpha \underline{V}(0, \boldsymbol{\rho})) \right). \end{cases} \tag{4.14}$$

The solution to equation (4.14) gives the total expected discounted cost of an optimal replacement-only policy when the belief about the active system's quality is never updated.

In order to prove our final main result concerning the structure of the optimal policy, we need to prove several useful lemmas. The first lemma establishes limits for the Bayesian update functions B and $\bar{B}$.

Lemma 4.1 *For all $\boldsymbol{b} \in \Delta^Y$, under Condition 2,*

$$\lim_{x \to \infty} B(x, \boldsymbol{b}) = \boldsymbol{b},$$

and, under Condition 3,

$$\lim_{x \to \infty} \bar{B}(x, \boldsymbol{b}) = \boldsymbol{e}_1.$$

Proof. Define $\lim_{x \to \infty} B_y(x, \boldsymbol{b}) = \ell_y$, then if ℓ_y exists for ever y, $B(x, \boldsymbol{b}) \to (\ell_1, \ldots, \ell_Y)$. Now, by definition,

$$B_y(x, \boldsymbol{b}) = \frac{g(x, y)}{G(x, \boldsymbol{b})} b_y,$$

where clearly, by Condition 2, we have that $g(x, y) \to 1$ and $G(x, \boldsymbol{b}) = \sum_m g(x, m) b_m \to \sum_m b_m = 1$. Therefore, $B_y(x, \boldsymbol{b}) \to b_y$ and $B(x, \boldsymbol{b}) \to \boldsymbol{b}$. Similarly, we consider $\bar{B}_y(x, \boldsymbol{b})$, where, after some algebraic manipulation,

$$\begin{aligned} \bar{B}_y(x, \boldsymbol{b}) &= \frac{\bar{g}(x, y) b_y}{\sum_m \bar{g}(x, m) b_m} \\ &= \frac{b_y}{\sum_m \frac{\bar{g}(x,m)}{\bar{g}(x,y)} b_m} \\ &= \frac{b_y}{b_y + \sum_{m \neq y} \frac{\bar{g}(x,m)}{\bar{g}(x,y)} b_m}. \end{aligned} \tag{4.15}$$

For $y = 1$, by Condition 3, the expression in (4.15) converges to 1. ∎

Lemmas 4.2 and 4.3 state that, regardless of the time-to-failure distribution, $\bar{G}$, G, $\bar{B}$, and B are continuous in the belief about the system quality.

Lemma 4.2 *For each fixed $x \in \mathcal{X} \setminus \{\infty\}$, $\bar{G}(x, \boldsymbol{b})$ and $G(x, \boldsymbol{b})$ are continuous in $\boldsymbol{b} \in \Delta^Y$.*

Proof. In what follows, the norm $||\cdot||$ will denote the Euclidean norm. For $\epsilon > 0$, consider $\delta(\epsilon) = \epsilon / Y$. Then, for any $\boldsymbol{b}, \boldsymbol{b}' \in \Delta^Y$ such that $||\boldsymbol{b} - \boldsymbol{b}'|| < \delta(\epsilon)$ we seek to show that $\left|\bar{G}(x, \boldsymbol{b}) - \bar{G}(x, \boldsymbol{b}')\right| < \epsilon$. First, we note that $||\boldsymbol{b} - \boldsymbol{b}'|| < \delta(\epsilon)$ implies $\left|b_y - b'_y\right| < \delta(\epsilon)$ for all $y \in \mathcal{Y}$.

Now, by the non-negativity of $\bar{g}(x, y)$ and the triangle inequality, we know that

$$\left|\bar{G}(x, \boldsymbol{b}) - \bar{G}(x, \boldsymbol{b}')\right| = \left|\sum_y \bar{g}(x, y)(b_y - b'_y)\right| \leq \sum_y \bar{g}(x, y) \left|b_y - b'_y\right|,$$

but by the bound on $\left|b_y - b'_y\right|$, we have that

$$\left|\bar{G}(x, \boldsymbol{b}) - \bar{G}(x, \boldsymbol{b}')\right| < \frac{\epsilon}{Y} \sum_y \bar{g}(x, y) \leq \frac{\epsilon}{Y} \sum_y 1 = \epsilon.$$

Therefore, $\bar{G}(x, \boldsymbol{b})$ is continuous in $\boldsymbol{b}$ and because $G(x, \boldsymbol{b}) = 1 - \bar{G}(x, \boldsymbol{b})$, we conclude that $G(x, \boldsymbol{b})$ is also continuous in $\boldsymbol{b}$. ■

Lemma 4.3 *For each fixed $x \in \mathcal{X} \setminus \{\infty\}$, $\bar{B}(x, \boldsymbol{b})$ and $B(x, \boldsymbol{b})$ are continuous in $\boldsymbol{b} \in \Delta^Y$.*

Proof. The function $\bar{B}(x, \cdot)$ is continuous if, and only if, $\bar{B}_y(x, \cdot) : \mathbb{R}^Y \to \mathbb{R}$ is continuous for all $y \in \mathcal{Y}$. By Lemma 4.2, $\bar{B}_y(x, \boldsymbol{b}) = \left(\bar{G}(x, \boldsymbol{b})\right)^{-1} \bar{g}(x, y) b_y$ is a product of continuous functions and is, therefore, continuous. Hence, by the continuity of its components, $\bar{B}(x, \cdot)$ is continuous. The proof that B is continuous is similar. ■

Our next result, states that for each fixed virtual age, the optimal value function, V, is continuous in the belief about the system quality.

Proposition 4.8 *Under Conditions 2-3, for each fixed $x \in \mathcal{X} \setminus \{\infty\}$, $V(x, \boldsymbol{b})$ is continuous in $\boldsymbol{b} \in \Delta^Y$.*

Proof. By induction, if $v^0(x, \boldsymbol{b}) = 0$ for all $(x, \boldsymbol{b}) \in \mathcal{X} \times \Delta^Y$, then

$$v^1(x, \boldsymbol{b}) = \begin{cases} c_I + c_R + c_F, & x = \infty, \\ c_I, & x < \infty, \end{cases}$$

so the base case holds. Now, assume $v^k(x, \boldsymbol{b})$ is continuous in $\boldsymbol{b}$ and note that

$$v^{k+1}(\infty, \boldsymbol{b}) = c_I + c_F + \min\{c_P + \alpha v^k(0, \boldsymbol{\rho}),\, c_R + \alpha v^k(0, \boldsymbol{b})\}.$$

Because the minimum of continuous functions is again continuous, we conclude that $v^{k+1}(\infty, \boldsymbol{b})$ is continuous in $\boldsymbol{b}$. Now, for finite $x \in \mathcal{X}$,

$$v^{k+1} = \min\{v^{k+1}(\infty, \boldsymbol{b}) - c_F,\, C^k_{DN}(x, \boldsymbol{b})\},$$

where $C^k_{DN}(x, \boldsymbol{b}) = c_I + \alpha(\bar{G}(x, \boldsymbol{b})v^k(x+1, \bar{B}(x+1, \boldsymbol{b})) + G(x, \boldsymbol{b})v^k(\infty, B(x+1, \boldsymbol{b}))$. Therefore, we proceed to show that $C^k_{DN}(x, \boldsymbol{b})$ is continuous to complete the proof. By Lemma 4.3, we know that $\bar{B}$ is continuous; hence,

$$\lim_{\boldsymbol{b}_n \to \boldsymbol{b}} \bar{B}(x+1, \boldsymbol{b}_n) = \bar{B}(x+1, \boldsymbol{b}).$$

Then, by the induction hypothesis, we see that

$$\lim_{\boldsymbol{b}_n \to \boldsymbol{b}} v^k(x+1, \bar{B}(x+1, \boldsymbol{b}_n)) = v^k(x+1, \lim_{\boldsymbol{b}_n \to \boldsymbol{b}} \bar{B}(x+1, \boldsymbol{b}_n)) = v^k(x+1, \bar{B}(x+1, \boldsymbol{b}));$$

thus, $v^k(x+1, \bar{B}(x+1, \boldsymbol{b}_n))$ is continuous. Similarly, $v^k(\infty, B(x+1, \boldsymbol{b}_n))$ is also continuous. Lastly, we note that G and $\bar{G}$ are continuous by Lemma 4.2. Because $C^k_{DN}(x, \boldsymbol{b})$ is the composition of continuous functions, it is also continuous. ■

Our next two results characterize the asymptotic behavior of the value function as the virtual age increases. First, Lemma 4.4 gives an expression for the value of doing nothing as the virtual age goes to infinity.

Lemma 4.4 *Under Conditions 2-3,*

$$\lim_{x \to \infty} V_{DN}(x, \boldsymbol{b}) = c_I + \alpha V(\infty, \boldsymbol{b}),$$

where $V_{DN}(x, \boldsymbol{b}) = c_I + \alpha \left(\bar{G}(x, \boldsymbol{b})V(x+1, \bar{B}(x+1, \boldsymbol{b})) + G(x, \boldsymbol{b})V(\infty, B(x+1, \boldsymbol{b}))\right)$.

Proof. We first show that, for all $x < \infty$, $V(x, \boldsymbol{b})$ is finitely bounded. By (4.2), we note that $V(0, \boldsymbol{b}) \leq c_I + c_R + \alpha V(0, \boldsymbol{b})$. Rearranging terms, observe that $V(0, \boldsymbol{b}) \leq (c_I + c_R)/(1-\alpha)$. For any $x < \infty$, $V(x, \boldsymbol{b}) \leq c_I + c_R + \alpha V(0, \boldsymbol{b}) \leq c_I + c_R + \alpha(c_I + c_R)/(1-\alpha)$. By Condition 2 and this finiteness, we have that

$$\lim_{x \to \infty} \bar{G}(x, \boldsymbol{b})V(x+1, \bar{B}(x+1, \boldsymbol{b})) = 0.$$

By Condition 2, Lemma 4.1, and Proposition 4.8,

$$\lim_{x \to \infty} G(x, \boldsymbol{b})V(\infty, B(x+1, \boldsymbol{b})) = V(\infty, \lim_{x \to \infty} B(x+1, \boldsymbol{b})) = V(\infty, \boldsymbol{b}).$$

Therefore, $V_{DN}(x, \boldsymbol{b}) \to c_I + \alpha V(\infty, \boldsymbol{b})$. ■

A natural consequence of Lemma 4.4 is that, for each fixed system belief, the value function is convergent. This is stated, without proof, in Corollary 4.2.

Corollary 4.2 *Under Conditions 2-3,*

$$\lim_{x\to\infty} V(x, \boldsymbol{b}) = c_I + \min\{c_P + \alpha V(0, \boldsymbol{\rho}), c_R + \alpha V(0, \boldsymbol{b}), \alpha V(\infty, \boldsymbol{b})\}.$$

Finally, Theorem 4.4 asserts that for each fixed system belief, $\boldsymbol{b} \in \Delta^Y$, under the appropriate conditions, there exists a threshold in the virtual age beyond which preventive maintenance is optimal. Moreover, this threshold is guaranteed to be finite.

Theorem 4.4 *Under Conditions 1-3, if $c_R < \alpha c_F$, then for each $\boldsymbol{b} \in \Delta^Y$, there exists an $x(\boldsymbol{b}) < \infty$ such that $d^*(x, \boldsymbol{b}) = d^*(x(\boldsymbol{b}), \boldsymbol{b}) > 0$ for all $x > x(\boldsymbol{b})$.*

Proof. By Theorem 4.2,

$$\alpha V(\infty, \boldsymbol{b}) \geq \alpha(c_F + V(x, \boldsymbol{b})),$$

for all $x \in \mathcal{X}$. Therefore,

$$\begin{aligned} \alpha V(\infty, \boldsymbol{b}) \geq \alpha(c_F + V(0, \boldsymbol{b})) &= \alpha c_F + \alpha V(0, \boldsymbol{b}) && (4.16)\\ &> c_R + \alpha V(0, \boldsymbol{b}) && (4.17)\\ &\geq \min\{c_P + \alpha V(0, \boldsymbol{\rho}), c_R + \alpha V(0, \boldsymbol{b})\}, && (4.18) \end{aligned}$$

where inequality (4.17) follows from the assumption that $c_R < \alpha c_F$, and inequality (4.18) by definition of the minimum function. For all $\delta, \epsilon > 0$, there exist two finite, possibly distinct, integers $x_{DN}(\delta, \boldsymbol{b})$ and $x(\epsilon, \boldsymbol{b})$ such that for all $x > x_{DN}(\delta, \boldsymbol{b})$,

$$|V_{DN}(x, \boldsymbol{b}) - (c_I + \alpha V(\infty, \boldsymbol{b}))| < \delta, \tag{4.19}$$

and for all $x > x(\epsilon, \boldsymbol{b})$

$$|V(x, \boldsymbol{b}) - (c_I + \min\{c_P + \alpha V(0, \boldsymbol{\rho}), c_R + \alpha V(0, \boldsymbol{b})\})| < \epsilon, \tag{4.20}$$

where (4.19) follows from Lemma 4.4, and (4.20) follows from Corollary 4.2 and the strict inequality (4.17).

Define

$$d(\boldsymbol{b}) = c_I + \alpha V(\infty, \boldsymbol{b}) - c_I + \min\{c_P + \alpha V(0, \boldsymbol{\rho}), c_R + \alpha V(0, \boldsymbol{b})\},$$

then for any $x > \max\{x_{DN}(d(\boldsymbol{b})/2, \boldsymbol{b}), x(d(\boldsymbol{b})/2, \boldsymbol{b})\}$, we have

$$
\begin{aligned}
V_{DN}(x, \boldsymbol{b}) - V(x, \boldsymbol{b}) &= V_{DN}(x, \boldsymbol{b}) - V(x\boldsymbol{b}) + d(\boldsymbol{b}) - d(\boldsymbol{b}) && (4.21)\\
&= V_{DN}(x, \boldsymbol{b}) - V(x\boldsymbol{b}) + d(\boldsymbol{b}) - [c_I + \alpha V(\infty, \boldsymbol{b}) && (4.22)\\
&\qquad - c_I + \min\{c_P + \alpha V(0, \boldsymbol{\rho}), c_R + \alpha V(0, \boldsymbol{b})\}] \\
&= (V_{DN}(x, \boldsymbol{b}) - c_I + \alpha V(\infty, \boldsymbol{b}) && (4.23)\\
&\quad + [c_I + \min\{c_P + \alpha V(0, \boldsymbol{\rho}), c_R + \alpha V(0, \boldsymbol{b})\} - V(x, \boldsymbol{b})] + d(\boldsymbol{b}) \\
&> -\frac{d(\boldsymbol{b})}{2} - \frac{d(\boldsymbol{b})}{2} + d(\boldsymbol{b}) && (4.24)\\
&= 0, && (4.25)
\end{aligned}
$$

where inequality (4.24) follows by (4.19), (4.20), and the fact that (4.16)-(4.18) imply $d > 0$. Thus, for all $x > \max\{x_{DN}(d/2, \boldsymbol{b}), x(d/2, \boldsymbol{b})\}$, doing nothing is strictly suboptimal. The result follows by defining $x(\boldsymbol{b}) = \max\{x_{DN}(d(\boldsymbol{b})/2, \boldsymbol{b}), x(d(\boldsymbol{b})/2, \boldsymbol{b})\}$. ■

We are unable to confirm that the value function is jointly monotone nondecreasing; however, if such monotonicity holds, we can establish stronger structural properties. Proposition 4.9 asserts that if the value function is jointly monotone, then the optimal decisions are also monotone in the system's virtual age for each fixed belief vector.

Proposition 4.9 *If $V(x, \boldsymbol{b})$ is jointly monotone nondecreasing, then for each $\boldsymbol{b} \in \Delta^Y$, $x \leq x'$ also implies $d^*(x, \boldsymbol{b}) \leq d^*(x', \boldsymbol{b})$.*

Proof. By contradiction, assume that for some $(x, \boldsymbol{b}) \in \mathcal{X} \times \Delta^Y$ and $x' > x$ that $d^*(x, \boldsymbol{b}) > d^*(x', \boldsymbol{b})$. It follows by Theorem 4.3 that $V(x, \boldsymbol{b}) = \min\{c_I + c_R + \alpha V(0, \boldsymbol{b}), c_I + c_R + \alpha V(0, \boldsymbol{\rho})\}$ and $V(x', \boldsymbol{b}) = \min\{c_I + c_R + \alpha V(0, \boldsymbol{b}), c_I + c_R + \alpha V(0, \boldsymbol{\rho}), V_D N x', \boldsymbol{b}\} \leq \min\{c_I + c_R + \alpha V(0, \boldsymbol{b}), c_I + c_R + \alpha V(0, \boldsymbol{\rho})\} = V(x, \boldsymbol{b})$. Therefore, $V(x, \boldsymbol{b}) = V(x', \boldsymbol{b})$ and $d^*(x, \boldsymbol{b}) = d^*(x', \boldsymbol{b})$, which is a contradiction. ■

Lastly, Proposition 4.10 states that if it is optimal to replace in a particular state, then it is optimal to replace for all larger states (with the belief vector being ordered in the likelihood ratio sense). That is, within the region for which replacement is the optimal maintenance action, the preventive age replacement thresholds are monotone nondecreasing in the belief vector.

Proposition 4.10 *If $V(x, \boldsymbol{b})$ is jointly monotone nondecreasing, then if $d^*(x, \boldsymbol{b}) = 2$, then $d^*(x_2, \boldsymbol{b}_2) = 2$ for all $x_2 \geq x_1$ and $\boldsymbol{b}_2 \geq_{lr} \boldsymbol{b}_1$.*

Proof. If $d^*(x_1, \boldsymbol{b}_1) = 2$ then by (4.2) we have that $V(x_1, \boldsymbol{b}_1) = c_I + c_P + \alpha V(0, \boldsymbol{\rho})$, and by joint monotonicity we see that $V(x_1, \boldsymbol{b}_1) \leq V(x_2, \boldsymbol{b}_2)$. Lastly, we note that $V(x_2, \boldsymbol{b}_2) \leq c_I + c_P + \alpha V(0, \boldsymbol{\rho})$. Combining these facts we see that

$$c_I + c_P + \alpha V(0, \boldsymbol{\rho}) \leq V(x_2, \boldsymbol{b}_2) \leq c_I + c_P + \alpha V(0, \boldsymbol{\rho}). \tag{4.26}$$

Therefore, equality holds throughout (4.26), thus completing the proof. ∎

4.4 NUMERICAL EXAMPLES

In this section, we illustrate our maintenance optimization framework on synthetic problem instances. We consider problems that are specifically tailored to illustrate particular properties, in addition to a large bed of randomly-parameterized problem instances. Examined are the qualitative properties of the optimal value function and resulting optimal policy. Additionally, we compare the cost of following the optimal *MOMDP* policy to several other policies.

Throughout all of our numerical examples, it is assumed that the time-to-failure, $(T|Q = y)$, follows a Weibull distribution with common shape parameter $k > 1$ and scale parameter λ_y. Additionally, it is assumed that $\lambda_1 > \lambda_2 > \cdots > \lambda_Y$. These distributional assumptions are made due to the prevalence of the Weibull distribution in modeling the time-to-failure, particularly within the context of maintenance optimization models. Additionally, under these assumptions, it is straightforward to verify that Conditions 1-3 are met.

All two-quality problem instances are coded within the MATLAB R2016a computing environment, and the three-quality problem instance is coded within the Java SE Runtime Environment 8. All codes are executed on a personal computer with a 3.50 GHz processor and 8GB of RAM.

4.4.1 Randomly-Generated Problem Instances

Here, 200 two-quality problems ($Y = 2$) are randomly generated with the aim of varying the problem parameters over a wide range values to assess the robustness of our *MOMDP* policy. In what follows, $U(a, b)$ denotes a continuous uniform random variable on (a, b). Fixing the number of system qualities at $Y = 2$, we randomly generate $M = 200$ problem instances. For problem $m \in \{1, \ldots, M\}$: the discount factor is denoted $\alpha^{(m)}$, where $\alpha \sim U(0.8, 0.9999)$; the cost vector is denoted $\boldsymbol{c}^{(m)} = (c_I^{(m)}, c_F^{(m)}, c_R^{(m)}, c_P^{(m)})$, where $c_I^{(m)} = 1$, $c_F^{(m)} \sim U(4,8)$, $c_R^{(m)} \sim U(1,2)$, and $c_P^{(m)} \sim c_R^{(m)} + U(1,4)$; the time-to-failure distribution shape parameter is denoted $k^{(m)}$ and scale parameter vector is denoted $\boldsymbol{\lambda}^{(m)} = (\lambda_1^{(m)}, \lambda_2^{(m)})$, where $k^{(m)} \sim U(1.1,3)$, $\lambda_2^{(m)} \sim U(1,10)$, and $\lambda_1^{(m)} \sim \lambda_2^{(m)} + U(1,10)$; the initial quality distribution is denoted $\boldsymbol{\rho}^{(m)} = (\rho_1^{(m)}, \rho_2^{(m)})$, where $\rho_1^{(m)} \sim U(0.1, 0.9)$ and $\rho_2^{(m)} = 1 - \rho_1^{(m)}$; and the inter-inspection period is denoted $\tau^{(m)}$, where $\tau^{(m)} \sim U(0.2, 1.5)$.

Because there are only two qualities, the belief state can be written as $\boldsymbol{b} = (b, 1-b)$; hence, the belief space can be simplified to the interval $[0,1]$. In order to compute the *MOMDP* policy, we discretize the interval $[0,1]$ into 1,000 states and truncate $\mathcal{X}$ to be large enough to have negligible impact on the optimal value function. When a value iteration step required the value function iterate be evaluated at a non-grid point, it is approximated using simple linear interpolation between the two nearest points. The optimal value function and policy are then obtained numerically using the value iteration algorithm.

In addition to our *MOMDP* policy, we consider three other policies: *Oracle*, *Heuristic*, and *Naive*. The *Oracle* policy is endowed with additional information in that it is given perfect information about the system quality. It then takes actions prescribed by the *MOMDP* policy but with the belief state fixed to the appropriate extreme point in Δ^Y. The *Oracle* policy provides a performance bound on the total expected discounted maintenance costs, as the additional information guarantees that, in expectation, it will outperform the *MOMDP* model. The *Heuristic* policy is determined by decoupling the problem across the belief states. Specifically, if the belief state is fixed, the problem of determining the optimal policy simplifies to solving a one-dimensional set of Bellman equations in which the time-to-failure distribution is determined by the current mixture distribution. As in the *MOMDP* policy, we begin by discretizing the interval and then solving a one-dimensional problem for each of the 1,000 belief states. This approach reduces the computational burden and storage requirements, and because each of the MDP models is monotone, we can utilize a monotone value iteration algorithm to further enhance the computational savings. The *Heuristic* policy is implemented by updating the belief about the quality of the system in the same manner as the *MOMDP* policy, but at each inspection epoch, it utilizes the one-dimensional policy corresponding to the current belief state to determine which action is taken. Finally, the *Naive* policy fixes the belief state at the initial distribution and, similar to the *Heuristic* policy, solves a one-dimensional MDP to determine whether or not to preventively maintain the system.

In order to compare the costs of the four policies, we use a simulation model. For a given simulation run, we simulate the system's time-to-failure each time it enters service, and when a system is replaced, we randomly draw a new system using the initial distribution $\boldsymbol{\rho}^{(m)}$. For each m, the simulation run length is given by the number of decision epochs $N^{(m)}$. Along each sample path, the total discounted cost is computed for each policy of interest, and these values are compared. It should be noted that, because the expected one-step costs are bounded, and the cost function is discounted, we can determine *a priori* the simulation run length needed to ensure that the total discounted cost is accurate to a fixed constant. More precisely, to guarantee that the finite approximation is within ϵ ($\epsilon > 0$) of the true total discounted cost, the number of decision epochs $N^{(m)}$ must satisfy

$$N^{(m)} \geq \frac{\ln\left[(1-\alpha^{(m)})\epsilon/C^{(m)}\right]}{\ln(\alpha^{(m)})} - 1, \quad m = 1, \ldots, M,$$

where $C^{(m)}$ is any valid upper bound on the expected one-step costs. For all numerical examples, $N^{(m)}$ is chosen to correspond to $\epsilon = 0.01$ and $C^{(m)}$ is taken to be $c_I^{(m)} + c_F^{(m)} + c_P^{(m)}$.

For each problem instance $m \in \{1, \ldots, M\}$, 500 sample paths are simulated and the cost of following each policy is computed. Under a particular policy and problem instance m, we denote the average total discounted maintenance cost (averaged over the 500 sample paths) by $\bar{v}^{(m)}_{policy}$, e.g., $\bar{v}^{(m)}_{Naive}$. In problem instance $m = 12$, the parameter values are as follows:

$$\begin{aligned}
\alpha^{(12)} &= 0.9904 \\
\boldsymbol{c}^{(12)} &= (1, 4.2283, 1.9530, 3.9396) \\
k^{(12)} &= 1.9516 \\
\boldsymbol{\lambda}^{(12)} &= (14.3213, 8.1914) \\
\boldsymbol{\rho}^{(12)} &= (0.4907, 0.5093) \\
\tau^{(12)} &= 0.4942
\end{aligned}$$

Problem instance $m = 12$ is noteworthy in that it exhibits the greatest discrepancy between the *MOMDP* and *Naive* policies; the *MOMDP* policy achieved a 39.47% average cost savings, i.e.,

$$\frac{\bar{v}^{(m)}_{Naive} - \bar{v}^{(m)}_{MOMDP}}{\bar{v}^{(m)}_{Naive}} \times 100\% = 39.47\%.$$

For each problem instance, we establish baseline performance by comparing each policy to the *Oracle* policy. We assess this difference by comparing the average increase in cost realized by using each policy. This increase, for a particular policy in problem instance m, is denoted by $\hat{v}^{(m)}_{policy}$. For example, when comparing the *MOMDP* policy against the *Oracle* policy in problem m, we compute

$$\hat{v}^{(m)}_{MOMDP} = \frac{\bar{v}^{(m)}_{MOMDP} - \bar{v}^{(m)}_{Oracle}}{\bar{v}^{(m)}_{Oracle}} \times 100\%.$$

These percentage increases are then averaged over all 500 problem instances to obtain the average cost increase for a given policy, denoted by $\hat{v}_{policy}$. Table 2 summarizes the average cost increase for each policy, demonstrating significant savings achieved by utilizing the *MOMDP* policy. It is noteworthy that our model yields a nearly 20% improvement over the *Naive* policy, on average.

Table 2: Summary of policy comparison results.

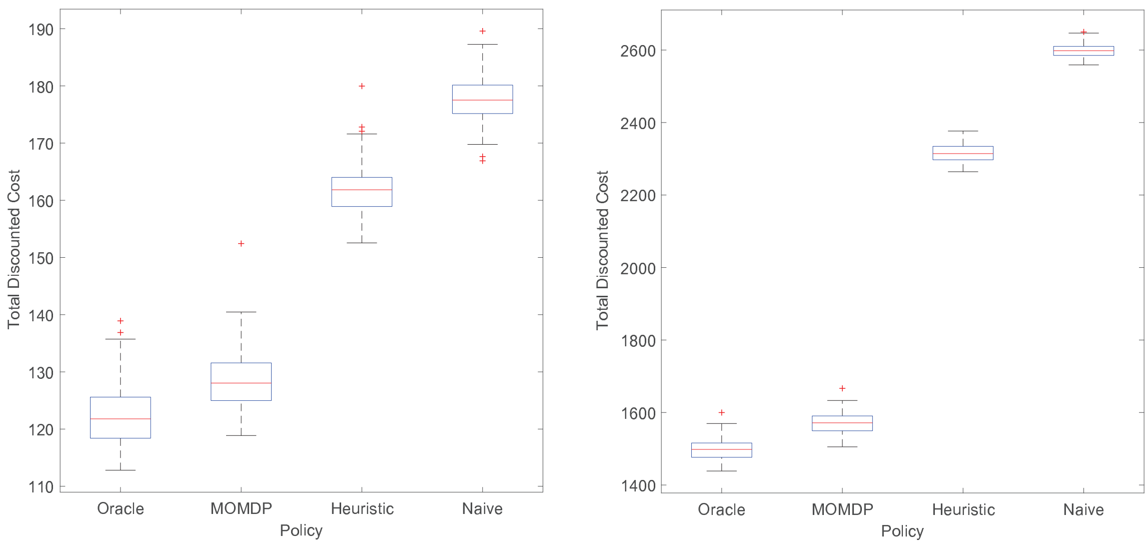
$\hat{v}_{Oracle}$	$\hat{v}_{MOMDP}$	$\hat{v}_{Heuristic}$	$\hat{v}_{Naive}$
–	7.70%	20.90%	26.39%

Figure 20 depicts the cost comparison between two particular problem instances, $m = 10$ and $m = 12$. In problem instance $m = 10$, the parameter values are as follows: $\alpha^{(10)} = 0.9989$, $\boldsymbol{c}^{(10)} = (1, 7.5643, 1.7253, 4.3246)$, $k^{(10)} = 1.4123$, $\boldsymbol{\lambda}^{(10)} = (10.5496, 5.8213)$, $\boldsymbol{\rho}^{(10)} = (0.3754, 0.6246)$, and $\tau^{(10)} = 0.9942$. In these problem instances, the difference between the performance of the *MOMDP* policy and *Oracle* policy are negligible, but there is a large performance gap between the *MOMDP* policy and the *Naive* and *Heuristic* policies. The most striking commonality between these two problem instances is that the discount factor $\alpha^{(m)}$ is large in both cases ($\alpha^{(10)}$, $\alpha^{(12)} > 0.99$). This finding indicates that our framework is likely to outperform these policies under an average cost criterion (as opposed to a discounted cost criterion).

4.4.2 A Specific Two-quality Problem

For the example considered in this section, the number of system qualities is again $Y = 2$. The discount rate is $\alpha = 0.99$ and the cost vector is $\boldsymbol{c} = (c_I, c_F, c_R, c_P) = (1, 2, 3, 4)$, the time-to-failure shape parameter is $k = 2$, the scale parameter vector is $\boldsymbol{\lambda} = (\lambda_1, \lambda_2) = (12, 6)$, the initial distribution is $\boldsymbol{\rho} = (\rho_1, \rho_2) = (0.7, 0.3)$, and the inter-inspection period is $\tau = 0.2$.

To compute the *MOMDP* policy, the belief space, $[0, 1]$, is uniformly discretized into 1,000 states and $\mathcal{X}$ is truncated to be $\{0, 1, \ldots, 200, \infty\}$. The optimal value function and policy are then obtained numerically using the value iteration algorithm. When a step in the value iteration algorithm requires a value function iterate whose belief state is outside this discretization, it is



(a) Boxplots of total discounted cost for $m = 10$.

(b) Boxplots of total discounted cost for $m = 12$.

Figure 20: Boxplots comparing policy costs for problem instances $m = 10$ and $m = 12$.

approximated by linear interpolation. For each $(x, \boldsymbol{b})$ in the discretized and truncated set of states, let $v_k(x, \boldsymbol{b})$ denote the kth iterate of the value iteration algorithm. The algorithm terminates when the maximum norm of the difference between subsequent value function iterates is below 10^{-6}, that is,

$$||v_{k+1} - v_k||_\infty = \max_{x,\boldsymbol{b}}\{|v_{k+1}(x, \boldsymbol{b}) - v_k(x, \boldsymbol{b})|\} \leq 10^{-6}.$$

In the case of only two qualities, the belief space is completely ordered; consequently, as seen in Figure 21, the value function exhibits monotonicity across the entire state space. In Figure 22, the *MOMDP* and *Heuristic* policies are depicted. For each fixed belief, b_1, both policies are of threshold type in age. We note that for the *MOMDP* policy, it is guaranteed to be of threshold-type by Proposition 4.9. Interestingly, in the *MOMDP* policy, we see that near the interface where the repair and replacement regions meet, the age threshold is increasing in both regions. This behavior is somewhat counter intuitive as the time-to-failure is stochastically smaller (in the hazard rate sense) near this interface than it is when b_1 is nearer to 1. This behavior can be understood as a natural exploration that occurs in the *MOMDP* policy. By allowing the system to function longer near this interface, the decision maker obtains failure data that is less likely to be right-censored. This additional data can be used to increase the liklihood that a high quality system is repaired and a low quality system is replaced. Moving away from this interface (by either increasing or decreasing b_1), we see that actions become more exploitative, i.e., quickly replace systems that are likely to be low quality and allow systems that are likely to be high quality to function for longer before preventively repairing. Additionally, we see that near where $b_1 = 1$, the *MOMDP* policy and the *Heuristic* policy are nearly identical. This observation is not surprising, as Proposition 4.7 guarantees that they should be exactly the same when $b_1 = 1$.

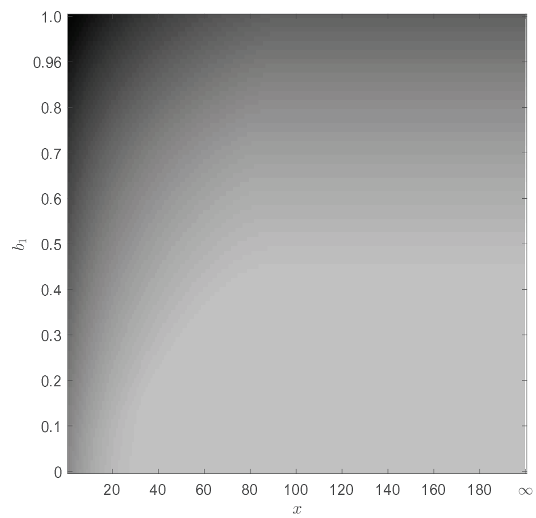
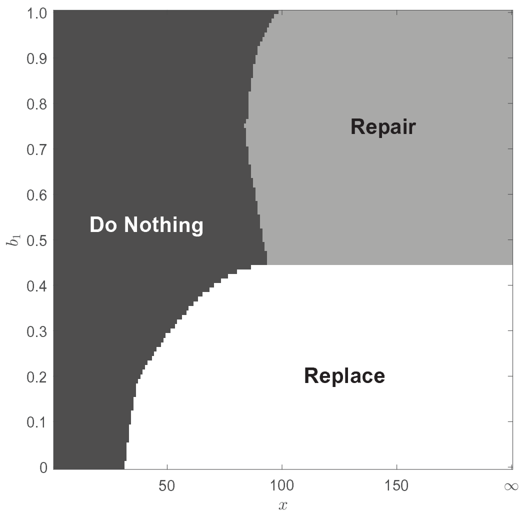


Figure 21: Depiction of the optimal value function (dark colors indicate lower costs).

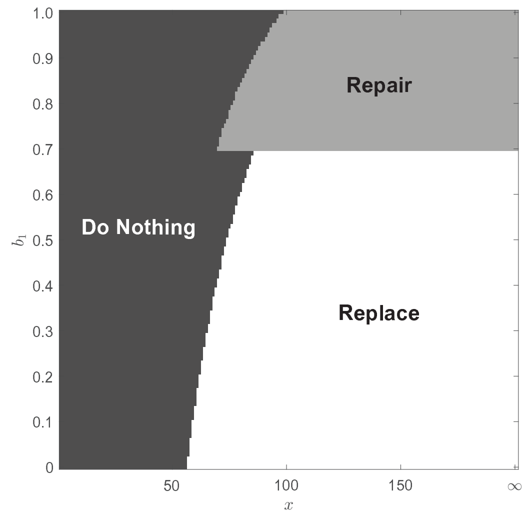
4.4.3 A Specific Three-quality Problem

For the example considered in this section, the number of system qualities is $Y = 3$. The discount rate is $\alpha = 0.99$ and the cost vector is $\boldsymbol{c} = (c_I, c_F, c_R, c_P) = (1, 2, 3, 4)$, the time-to-failure shape parameter is $k = 2$, the scale parameter vector is $\boldsymbol{\lambda} = (\lambda_1, \lambda_2) = (12, 10, 6)$, the initial distribution is $\boldsymbol{\rho} = (\rho_1, \rho_2, \rho_3) = (0.5, 0.2, 0.3)$, and the inter-inspection period is $\tau = 1$.

To compute the *MOMDP* policy, each dimension of the belief space is uniformly discretized into 500 states and $\mathcal{X}$ is truncated to be $\{0, 1, \ldots, 50, \infty\}$. The optimal value function and policy are then obtained numerically using the value iteration algorithm. When a step in the value iteration algorithm requires a value function evaluation at a belief state outside this discretization, it is approximated by bilinear interpolation (with edge cases approximated by linear or barycentric



(a) Depiction of the *MOMDP* policy.



(b) Depiction of the *Heuristic* policy.

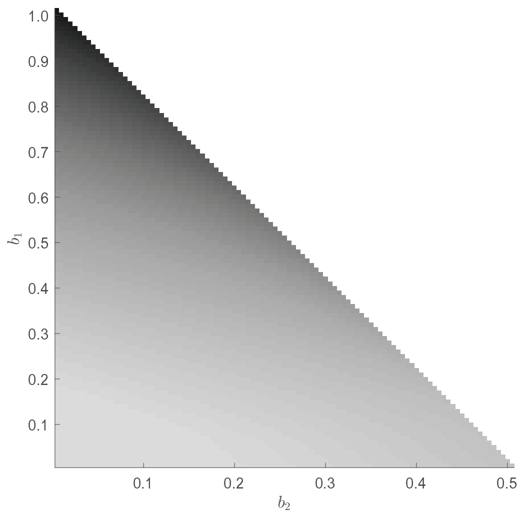
Figure 22: Comparison between the *MOMDP* and *Heuristic* policies.

interpolation). For each $(x, \boldsymbol{b})$ in the discretized and truncated set of states, let $v_k(x, \boldsymbol{b})$ denote the kth iterate of the value iteration algorithm. The algorithm terminates when the maximum norm of the difference between subsequent value function iterates is below 10^{-6}, that is,

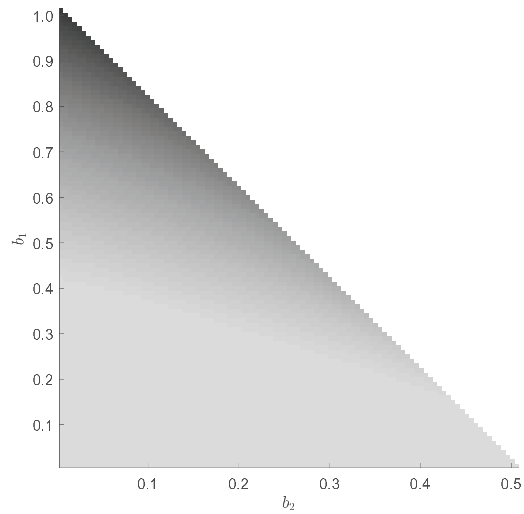
$$||v_{k+1} - v_k||_\infty = \max_{x,\boldsymbol{b}}\{|v_{k+1}(x, \boldsymbol{b}) - v_k(x, \boldsymbol{b})|\} \leq 10^{-6}.$$

Figure 23 depicts a portion of the optimal value function evaluated at virtual age $x = 5$ and $x = 25$. It should be noted that in each image, the origin represents the belief state $\boldsymbol{b} = \boldsymbol{e}_3 = (0, 0, 1)$, and can, therefore, be thought of as the worst belief. For this reason, we see that at $x = 5$ and $x = 25$ starting from this belief state has the highest total expected discounted cost. Similarly, starting from belief state $(1, 0, 0)$ has the lowest cost. It can also be observed that the value function exhibits monotonicity for each fixed x (as guaranteed by Theorem 4.2), but also across the x's, i.e., $V(5, \boldsymbol{b}) < V(25, \boldsymbol{b})$ for each $\boldsymbol{b}$.

Figure 24 provides a graphical depiction of the *MOMDP* policy. By Theorem 4.3, the belief state space can be partitioned into two regions: one in which doing nothing or repair is optimal, and another in which doing nothing or replacement is optimal (see Figure 24(a)). It is not coincidental



(a) The optimal value function at $x = 5$.

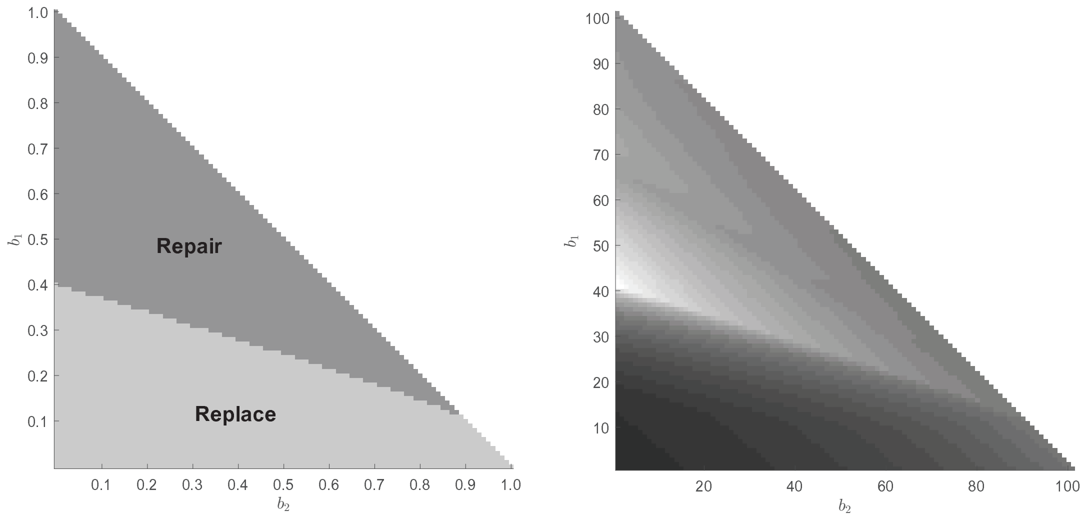


(b) The optimal value function at $x = 25$.

Figure 23: Depiction of the optimal value function (dark colors indicate lower costs).

that these regions are separated in the belief state space by a straight line; rather, it is a further consequence of Theorem 4.3 resulting from the partitioning of Δ^3 being related to the likelihood ratio ordering. In particular, these regions are divided in such a way that if $\boldsymbol{b} \leq_{lr} \boldsymbol{b}'$ and $\boldsymbol{b}'$ is in the repair region, then $\boldsymbol{b}$ is also in the repair region. Similarly, if $\boldsymbol{b} \leq_{lr} \boldsymbol{b}'$ and $\boldsymbol{b}$ is in the replacement region, then $\boldsymbol{b}'$ is also in the replacement region.

By the joint monotonicity of the value function, and Proposition 4.9, the optimal policy is guaranteed to be a threshold-type policy, for each fixed belief state. Figure 24(b) shows the age threshold for each belief state above which it is optimal to perform preventive maintenance. Unsurprisingly, the thresholds are ordered at the corner points of the plot, i.e., the threshold at $\boldsymbol{e}_3$ is the smallest and at $\boldsymbol{e}_1$ is the largest. However, in contrast to the two-quality case, the largest threshold is not when the belief state is $\boldsymbol{e}_1$, but rather at $\boldsymbol{b} = (0.4, 0, 0.6)$. Again, a ridge is formed along the interface between the repair and replacement regions where exploration is encouraged in the form of large thresholds. Additionally, we see that the thresholds are monotone in the belief state within the replacement region, but not within the repair region. By the joint monotonicity of the value function, the monotone age thresholds in the replacement region are guaranteed by Proposition 4.10.



(a) The repair and replacement regions. (b) The preventive maintenance thresholds.

Figure 24: Depiction of the *MOMDP* policy (dark colors indicate lower thresholds).

4.5 CONCLUSIONS AND FUTURE WORK

In this work, we have considered the problem of optimally maintaining a stochastically degrading, single-unit system with heterogeneous spare parts of varying quality. To address this problem, we presented an MOMDP model and investigated its properties. Under intuitive conditions on the time-to-failure distributions, we have established monotonicity properties of the optimal value function and presented a comprehensive characterization of the optimal policy. Additionally, by way of a detailed computational study, we highlighted the cost savings that can be achieved by properly accounting for spare part heterogeneity. These numerical illustrations also revealed that the optimal policy implicitly accounts for the tradeoff between receiving high-quality, uncensored data (which improves long-term decision making) and reducing short-term maintenance costs.

The model we presented herein can be improved in a few important ways. First, our model assumes that the proportion of parts of each quality is fixed and known, and that the number of qualities and their respective time-to-failure distributions are known. Relaxing these assumptions would allow for additional model flexibility and the investigation of tradeoffs between parameter learning and maintenance decisions that exploit the current belief about the parameters. Another promising direction for future research is to relax the assumption that the inter-inspection period

τ is a predetermined model parameter. Two problems related to this relaxation are worthy of further consideration: (i) determining the optimal fixed value of τ; and (ii) allowing the subsequent inter-inspection length to be set at each inspection epoch. In the latter problem, shorter inter-inspection intervals would provide higher-quality information but with an increase in cost. Due to this tradeoff, we suspect the optimal policy may be difficult to fully characterize.

BIBLIOGRAPHY

[1] M. Akan, B. Ata, and M. A. Lariviere. Asymmetric information and economies of scale in service contracting. *Manufacturing and Service Operations Management*, 13(1):58–72, 2011.

[2] O. Z. Akşin, M. Armony, and V. Mehrotra. The modern call center: A multi-disciplinary perspective on operations management research. *Production and Operations Management*, 16(6):665–688, 2007.

[3] O. Z. Akşin, F. de Véricourt, and F. Karaesmen. Call center outsourcing contract analysis and choice. *Management Science*, 54(2):354–368, 2008.

[4] O. Z. Akşin, F. Karaesmen, and E. L. Örmeci. A review of workforce cross-training in call centers from an operations management perspective. In D. Nembhard, editor, *Workforce Cross Training Handbook*. CRC Press, 2007.

[5] R. Barlow and F. Proschan. *Mathematical Theory of Reliability*. John Wiley & Sons, Inc., New York, NY, 1965.

[6] R. E. Barlow and F. Proschan. *Mathematical Theory of Reliability*. John Wiley & Sons, Inc., New York, NY, 1965.

[7] A. Bassamboo, J. M. Harrison, and A. Zeevi. Dynamic routing and admission control in high-volume service systems: Asymptotic analysis via multi-scale fluid limits. *Queueing Systems*, 51(3):249–285, 2005.

[8] A. Bassamboo, J. M. Harrison, and A. Zeevi. Design and control of a large call center: Asymptotic analysis of an LP-based method. *Operations Research*, 54(3):419–435, 2006.

[9] A. Bassamboo, R. S. Randhawa, and A. Zeevi. Capacity sizing under parameter uncertainty: Safety staffing principles revisited. *Management Science*, 56(10):1668–1686, 2010.

[10] R. Bellman. *Dynamic Programming*. Princeton University Press, Princeton, NJ, 1957.

[11] R. Bellman. A Markovian decision process. *Indiana University Mathematics Journal*, 6(4):679–684, 1957.

[12] P. Belotti, J. Lee, L. Liberti, F. Margot, and A. Wächter. Branching and bounds tightening techniques for non-convex minlp. *Optimization Methods and Software*, 24:597–634, 2009.

[13] D. Bertsekas and S. Shreve. *Stochastic Optimal Control: The Discrete-Time Case*. Athena Scientific, Belmont, MA, 1996.

[14] L. Bian and N. Gebraeel. Computing and updating the first-passage time distribution for randomly evolving degradation signals. *IIE Transactions*, 44(11):974–987, 2012.

[15] G. Birkhoff. *Lattice Theory, Third Edition*. American Mathematical Society, Providence, RI, 1967.

[16] M. Bodur and J. R. Luedtke. Mixed-integer rounding enhanced Benders decomposition for multiclass service-system staffing and scheduling with arrival rate uncertainty. *Management Science*, 63(7):2073–2091, 2017.

[17] K. Bouvard, S. Artus, C. Berenguer, and V. Cocquempot. Condition-based dynamic maintenance operataions planning and grouping: Application to commercial heavy vehicles. *Reliability Engineering and System Safety*, 96(6):601–610, 2011.

[18] B. Castanier, A. Grall, and C. Berenguer. A condition-based maintenance policy with non-periodic inspections for a two-unit series system. *Reliability Engineering and System Safety*, 87(1):109–120, 2005.

[19] J. H. Cha and M. Finkelstein. Stochastic analysis of preventive maintenance in heterogeneous populations. *Operations Research Letters*, 40(5):416–421, 2012.

[20] B. P. Chen and S. G. Henderson. Two issues in setting call centre staffing levels. *Annals of Operations Research*, 108(1):175–192, 2001.

[21] N. Chen and K. Tsui. Condition monitoring and remaining useful life prediction using degradation signals: Revisited. *IIE Transactions*, 45(9):939–952, 2013.

[22] N. Chen, Z. Ye, Y. Xiang, and L. Zhang. Condition-based maintenance using the inverse Gaussian degradation model. *European Journal of Operational Research*, 243(1):190–199, 2015.

[23] D. Cho and M. Parlar. A survey of maintenance models for multi-unit systems. *European Journal of Operational Research*, 51(1):1–23, 1991.

[24] B. Colson, P. Marcotte, and G. Savard. An overview of bilevel optimization. *Annals of Operations Research*, 153(1):235–256, 2007.

[25] R. Dekker. Applications of maintenance optimization models: A review and analysis. *Reliability Engineering and System Safety*, 51:229–240, 1996.

[26] R. Dekker. Applications of maintenance optimization models: A review and analysis. *Reliability Engineering & System Safety*, 51(3):229–240, 1996.

[27] R. Dekker, R. Wildeman, and F. van der Duyn Schouten. A review of multi-component maintenance models with economic dependence. *Mathematical Methods of Operations Research*, 45:411–435, 1997.

[28] A. Deslauriers, P. L'Ecuyer, J. Pichitlamken, A. Ingolfsson, and A. N. Avramidis. Markov chain models of a telephone call center with call blending. *Computers and Operations Research*, 34(6):1616–1645, 2007.

[29] A. H. Elwany, N. Z. Gebraeel, and L. M. Maillart. Structured replacement policies for components with complex degradation processes and dedicated sensors. *Operations Research*, 59(3):684–695, 2011.

[30] N. Gans, G. Koole, and A. Mandelbaum. Telephone call centers: Tutorial, review, and research prospects. *Manufacturing and Service Operations Management*, 5(2):79–141, 2003.

[31] N. Gans, H. Shen, Y.-P. Zhou, N. Korolev, A. McCord, and H. Ristock. Parametric forecasting and stochastic programming models for call-center workforce scheduling. *Manufacturing and Service Operations Management*, 17(4):571–588, 2015.

[32] N. Gans and Y.-P. Zhou. Call-routing schemes for call-center outsourcing. *Manufacturing and Service Operations Management*, 9(1):33–50, 2007.

[33] O. Garnett, A. Mandelbaum, and M. Reiman. Designing a call center with impatient customers. *Manufacturing and Service Operations Management*, 4(3):208–227, 2002.

[34] N. Gebraeel, M. Lawley, R. Li, and J. Ryan. Residual-life distributions from component degradation signals: A Bayesian approach. *IIE Transactions*, 37(6):543–557, 2005.

[35] A. Geramifard, T. Walsh, S. Tellex, G. Chowdhary, N. Roy, and J. How. A tutorial on linear function approximators for dynamic programming and reinforcement learning. *Foundations and Trends in Machine Learning*, 6(4):375–451, 2013.

[36] A. Ghouila-Houri. Caractérisation des matrices totalement unimodulaires. *Comptes Rendus de l'Académie des Sciences*, 254:1192–1194, 1962.

[37] I. Guedj and A. Mandelbaum. "Anonymous bank" call-center data. Accessed July 13, 2017: http://iew3.technion.ac.il/serveng/callcenterdata/index.html, 2000.

[38] I. Gurvich, J. Luedtke, and T. Tezcan. Staffing call centers with uncertain demand forecasts: A chance-constrained optimization approach. *Management Science*, 56(7):1093–1115, 2010.

[39] A. L. Hartzell, M. G. da Silva, and H. R. Shea. *MEMS Reliability*. MEMS Reference Shelf. Springer, Boston, MA, 2011.

[40] S. Hasija, E. J. Pinker, and R. A. Shumsky. Call center outsourcing contracts under information asymmetry. *Management Science*, 54(4):793–807, 2008.

[41] A. J. Hoffman and J. B. Kruskal. Integral boundary points of convex polyhedra. In H. W. Kuhn and A. W. Tucker, editors, *Linear Inequalities and Related Systems*, pages 223–246. Princeton University Press, Princeton, NJ, 1956.

[42] R. Howard. *Dynamic Programming and Markov Processes*. The MIT Press, Cambridge, MA, 1960.

[43] A. Ingolfsson, F. Campello, X. Wu, and E. Cabral. Combining integer programming and the randomization method to schedule employees. *European Journal of Operational Research*, 202(1):153–163, 2010.

[44] International Customer Management Institute. *2006 Contact Center Outsourcing Report*. ICMI Press, Annapolis, MD, 2006.

[45] O. B. Jennings, A. Mandelbaum, W. A. Massey, and W. Whitt. Server staffing to meet time-varying demand. *Management Science*, 42(10):1383–1394, 1996.

[46] F. Jensen and N. E. Petersen. *Burn-In: An Engineering Approach to the Design and Analysis of Burn-in Procedures*. John Wiley & Sons Ltd., New York, NY, 1982.

[47] G. Jongbloed and G. Koole. Managing uncertainty in call centres using Poisson mixtures. *Applied Stochastic Models in Business and Industry*, 17(4):307–318, 2001.

[48] J. Kharoufeh and S. Cox. Stochastic models for degradation-based reliability. *IIE Transactions*, 37(6):533–542, 2005.

[49] J. Kharoufeh, D. Finkelstein, and D. Mixon. Availability of periodically inspected systems with Markovian wear and shocks. *Journal of Applied Probability*, 43(2):303–317, 2006.

[50] J. Kharoufeh and D. Mixon. On a Markov-modulated shock and wear process. *Naval Research Logistics*, 56(6):563–576, 2009.

[51] J. Kharoufeh, C. Solo, and M. Ulukus. Semi-markov models for degradation-based reliability. *IIE Transactions*, 42(8):599–612, 2010.

[52] Y. Ko and E. Byon. Condition-based joint maintenance optimization for a large-scale system with homogeneous units. *IIE Transactions*, (DOI: 10.1080/0740817X.2016.1241457).

[53] Y. L. Koçağa, M. Armony, and A. R. Ward. Staffing call centers with uncertain arrival rates and co-sourcing. *Production and Operations Management*, 24(7):1101–1117, 2015.

[54] G. Konidaris, S. Osentoski, and P. Thomas. Value function approximation in reinforcement learning using the Fourier basis. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, San Francisco, CA, 2011. The AAAI Press.

[55] G. Koole and A. Mandelbaum. Queueing models of call centers: An introduction. *Annals of Operations Research*, 113(1):41–59, 2002.

[56] G. Koole and A. Pot. An overview of routing and staffing algorithms in multi-skill customer contact centers. Technical report, Vrije Universiteit, Amsterdam, The Netherlands, 2006.

[57] W. Kuo and Y. Kuo. Facing the headaches of early failures: A state-of-the-art review of burn-in decisions. *Proceedings of the IEEE*, 71(11):1257–1266, 1983.

[58] J. Lawless and M. Crowder. Covariates and random effects in a Gamma process model with applications to degradation and failure. *Lifetime Data Analysis*, 10(3):213–227, 2004.

[59] P. L'Ecuyer. Modeling and optimization problems in contact centers. In *Third International Conference on the Quantitative Evaluation of Systems (QEST'06)*, pages 145–156, Sept 2006.

[60] S. Maman. *Uncertainty in the Demand for Service: The Case of Call Centers and Emergency Departments*. Master's thesis, Technion–Israel Institute of Technology, Haifa, Israel, 2009.

[61] A. Mandelbaum and S. Zeltyn. Staffing many-server queues with impatient customers: Constraint satisfaction in call centers. *Operations Research*, 57(5):1189–1205, 2009.

[62] M. Marseguerra, E. Zio, and L. Podofillini. Condition-based maintenance optimization by means of genetic algorithms and Monte Carlo simulation. *Reliability Engineering and System Safety*, 77(2):151–165, 2002.

[63] J. McCall. Maintenance policies for stochastically failing equipment: A survey. *Management Science*, 11(5):493–524, 1965.

[64] J. J. McCall. Maintenance policies for stochastically failing equipment: A survey. *Management Science*, 11(5):493–524, 1965.

[65] V. Mehrotra, O. Ozlük, and R. Saltzman. Intelligent procedures for intra-day updating of call center agent schedules. *Production and Operations Management*, 19(3):353–367, 2010.

[66] J. Mi. Burn-in and maintenance policies. *Advances in Applied Probability*, 26(1):207–221, 1994.

[67] J. Mi. Minimizing some cost functions related to both burn-in and field use. *Operations Research*, 44(3):497–500, 1996.

[68] J. M. Milner and T. L. Olsen. Service-level agreements in call centers: Perils and prescriptions. *Management Science*, 54(2):238–252, 2008.

[69] R. Nicolai and R. Dekker. Optimal maintenance of multi-component systems: A review. In *Complex System Maintenance Handbook*, pages 263–286, London, 2008. Springer London.

[70] S. C. W. Ong, S. W. Png, D. Hsu, and W. S. Lee. Planning under uncertainty for robotic tasks with mixed observability. *The International Journal of Robotics Research*, 29(8):1053–1068, 2010.

[71] C.-Y. Peng and S.-T. Tseng. Mis-specification analysis of linear degradation models. *IEEE Transactions on Reliability*, 58(3):444–455, 2009.

[72] H. Pham and H. Wang. Imperfect maintenance. *European Journal of Operational Research*, 94(3):425–438, 1996.

[73] H. Pham and H. Wang. Imperfect maintenance. *European Journal of Operational Research*, 94(3):425–438, 1996.

[74] W. Pierskalla and J. Voelker. A survey of maintenance models: The control and surveillance of deteriorating systems. *Naval Research Logistics Quarterly*, 23(3):353–388, 1976.

[75] W. P. Pierskalla and J. A. Voelker. A survey of maintenance models: The control and surveillance of deteriorating systems. *Naval Research Logistics Quarterly*, 23(3):353–388, 1976.

[76] W. Powell. *Approximate Dynamic Programming: Solving the Curses of Dimensionality.* Wiley, Hoboken, NJ, 2007.

[77] M. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley, Hoboken, NJ, 2nd edition, 2005.

[78] Z. J. Ren, F. Zhang, and T. Feng. Service outsourcing: Capacity, quality and correlated costs. *Social Science Research Network*, Online article: https://ssrn.com/abstract=1362036, 2016.

[79] Z. J. Ren and Y.-P. Zhou. Call center outsourcing: Coordinating staffing level and service quality. *Management Science*, 54(2):369–383, 2008.

[80] T. R. Robbins and T. P. Harrison. A simulation based scheduling model for call centers with uncertain arrival rates. In *Proceedings of the 2008 Winter Simulation Conference (WSC 2008)*, pages 2884–2890, Miami, FL, 2008. IEEE.

[81] T. R. Robbins and T. P. Harrison. A stochastic programming model for scheduling call centers with global Service Level Agreements. *European Journal of Operational Research*, 207(3):1608–1619, 2010.

[82] G. Rummery and M. Niranjan. On-line Q-learning using connectionist systems. Technical report, Cambridge University Engineering Department, 1994.

[83] M. Shaked and J. Shanthikumar. *Stochastic Orders*. Springer, New York, NY, 2007.

[84] M. Shaked and J. G. Shanthikumar. *Stochastic Orders*. Springer, New York, NY, 2007.

[85] A. Sharma, G. Yadava, and S. Deshmukh. A literature review and future perspectives on maintenance optimization. *Journal of Quality in Maintenance Engineering*, 17(1):5–25, 2011.

[86] A. Sharma, G. Yadava, and S. Deshmukh. A literature review and future perspectives on maintenance optimization. *Journal of Quality in Maintenance Engineering*, 17(1):5–25, 2011.

[87] Y. Sherif and M. Smith. Optimal maintenance models for systems subject to failure – A review. *Naval Research Logistics Quarterly*, 28(1):47–74, 1981.

[88] Y. Sherif and M. Smith. Optimal maintenance models for systems subject to failure – A review. *Naval Research Logistics Quarterly*, 28(1):47–74, 1981.

[89] N. Singpurwalla. Survival in dynamic environments. *Statistical Science*, 10(1):86–103, 1995.

[90] N. D. Singpurwalla. Survival in dynamic environments. *Statistical Science*, 10(1):86–103, 1995.

[91] S. G. Steckley, S. G. Henderson, and V. Mehrotra. Forecast errors in service systems. *Probability in the Engineering and Informational Sciences*, 23(2):305–332, 2009.

[92] R. Sutton and A. Barto. *Reinforcement Learning: An Introduction*. The MIT Press, Cambridge, MA, 1998.

[93] G. M. Thompson. Labor staffing and scheduling models for controlling service levels. *Naval Research Logistics*, 44(8):719–740, 1997.

[94] Z. Tian and H. Liao. Condition-based maintenance optimization for multi-component systems using proportional hazards model. *Reliability Engineering and System Safety*, 96(5):581–589, 2011.

[95] D. Topkis. *Supermodularity and Complementarity.* Princeton University Press, Princeton, NJ, 1998.

[96] C.-C. Tsai, S.-T. Tseng, and N. Balakrishnan. Optimal design for degradation tests based on Gamma processes with random effects. *IEEE Transactions on Reliability*, 61(2):604–613, 2012.

[97] M. Ulukus, J. Kharoufeh, and L. Maillart. Optimal replacement policies under environment-driven degradation. *Probability in the Engineering and Informational Sciences*, 26(3):405–424, 2012.

[98] C. Valdez-Flores and R. Feldman. A survey of preventive maintenance models for stochastically deteriorating single-unit systems. *Naval Research Logistics*, 36(4):419–446, 1989.

[99] C. Valdez-Flores and R. M. Feldman. A survey of preventive maintenance models for stochastically deteriorating single-unit systems. *Naval Research Logistics*, 36(4):419–446, 1989.

[100] J.-C. Van den Schrieck, O. Z. Akşin, and P. Chevalier. Peakedness-based staffing for call center outsourcing. *Production and Operations Management*, 23(3):504–524, 2014.

[101] C. van Oosterom, H. Peng, and G. van Houtum. Maintenance optimization for a Markovian deteriorating system with population heterogeneity. *IISE Transactions*, 49(1):96–109, 2017.

[102] H. Wang. A survey of maintenance policies of deteriorating systems. *European Journal of Operational Research*, 139(3):469–489, 2002.

[103] H. Wang. A survey of maintenance policies of deteriorating systems. *European Journal of Operational Research*, 139(3):469–489, 2002.

[104] W. Wang. A model to determine the optimal critical level and the monitoring intervals in condition-based maintenance. *International Journal of Production Research*, 38(6):1425–1436, 2000.

[105] X. Wang. Wiener processes with random effects for degradation data. *Journal of Multivariate Analysis*, 101(2):340–351, 2010.

[106] W. Whitt. Staffing a call center with uncertain arrival rate and absenteeism. *Production and Operations Management*, 15(1):88–102, 2006.

[107] Y. Xiang, D. W. Coit, and Q. Feng. n Subpopulations experiencing stochastic degradation: reliability modeling, burn-in, and preventive replacement optimization. *IIE Transactions*, 45(4):391–408, 2013.

[108] Y. Xiang, D. W. Coit, and Q. Feng. Accelerated burn-in and condition-based maintenance for n-subpopulations subject to stochastic degradation. *IIE Transactions*, 46(10):1093–1106, 2014.

[109] W. Yang, P. J. Tavner, C. J. Crabtree, Y. Feng, and Y. Qiu. Wind turbine condition monitoring: Technical and commercial challenges. *Wind Energy*, 17(5):673–693, 2012.

[110] Z. Ye, Y. Shen, and M. Xie. Degradation-based burn-in with preventive maintenance. *European Journal of Operational Research*, 221(2):360–367, 2012.

[111] Z. Ye, Y. Wang, K. Tsui, and M. Pecht. Degradation data analysis using Wiener processes with measurement errors. *IEEE Transactions on Reliability*, 62(4):772–780, 2013.

[112] M. Zhang, Z. Ye, and M. Xie. A condition-based maintenance strategy for heterogeneous populations. *Computers & Industrial Engineering*, 77:103–114, 2014.

[113] Q. Zhu, H. Peng, and G. van Houtum. A condition-based maintenance policy for multi-component systems with a high maintenance setup cost. *OR Spectrum*, 37:1007–1035, 2015.