Data-Driven Robust Optimization in Healthcare Applications

by

Austin Bren

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2018 by the
Graduate Supervisory Committee:

Soroush Saghafian, Co-Chair
Pitu Mirchandani, Co-Chair
Teresa Wu
Rong Pan

ARIZONA STATE UNIVERSITY

May 2018

ABSTRACT

Healthcare operations have enjoyed reduced costs, improved patient safety, and innovation in healthcare policy over a huge variety of applications by tackling problems via the creation and optimization of descriptive mathematical models to guide decision-making. Despite these accomplishments, models are stylized representations of real-world applications, reliant on accurate estimations from historical data to justify their underlying assumptions. To protect against unreliable estimations which can adversely affect the decisions generated from applications dependent on fully-realized models, techniques that are robust against misspecications are utilized while still making use of incoming data for learning. Hence, new robust techniques are applied that (1) allow for the decision-maker to express a spectrum of pessimism against model uncertainties while (2) still utilizing incoming data for learning. Two main applications are investigated with respect to these goals, the first being a percentile optimization technique with respect to a multi-class queueing system for application in hospital Emergency Departments. The second studies the use of robust forecasting techniques in improving developing countries' vaccine supply chains via (1) an innovative outside of cold chain policy and (2) a district-managed approach to inventory control. Both of these research application areas utilize data-driven approaches that feature learning and pessimism-controlled robustness.

DEDICATION

I dedicate this work to my wife, Kristen. Your love and support have been a constant source of my strength, which has enabled this accomplishment.

ACKNOWLEDGMENTS

This dissertation has only been possible through an incredible, encouraging support network. I want to especially thank Dr. Soroush Saghafian, for having patience in times of slow progress, believing in my potential, pressing onward through setbacks, and pushing my work to a much higher level than I could have ever achieved on my own. I want to thank Dr. Pitu Mirchandani, for providing lots of encouragement, providing advice, and for acting as the co-chair of my committee. I also want to thank my committee members, Dr. Rong Pan, Dr. Teresa Wu, for providing guidance on my research.

I want to thank Dr. Michael Clough, who has been an excellent source of advice, engaging conversation, and valuable experience in all matters. Finally, I want to thank my family, who have been so supportive in my pursuit of education, and have believed in me throughout this entire journey.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 Overview

Healthcare operations have enjoyed significant improvements over a huge variety of applications by tackling problems via (1) the creation of descriptive mathematical models and (2) the optimization of said models to guide decision-making. This approach has yielded great successes in the reduction of costs, improvement of patient safety, and innovations in healthcare policy. Despite these accomplishments, models are stylized representations of real-world applications, and as such, must rely on estimations from historical data to justify their underlying assumptions. In settings with scarce or unreliable data, such as those experienced at the inception of a new process, specifying model parameters or underlying stochastic elements can be highly unreliable. Hence, models are subject to inevitable misspecifications which can adversely affect the decisions generated from applications dependent on precise estimation, resulting in excess expenses, higher patient risks, or an unnecessarily strained system.

To combat these issues, many studies pursue techniques that are robust against model misspecifications. These robust procedures often utilize minimax objectives on specialized sets of parameters or models to guide their decision-making. By optimizing against the worst-case scenario within the specified set via engaging in a game against an antagonistic agent, this technique effectively protects their decisions from poorly estimated models. However, naively constructing such robust methods can be ineffective in real applications due to (1) overly-conservative policies and (2) ignoring potential learning from incoming streams of data. To gain a large amount of robustness against model ambiguity, a decision-maker might choose to create expansive ambiguity sets to encompass a large variety of potential situations. However, if these robust sets are too large, the resulting policies will be optimizing over highly unrealistic scenarios with respect to the real application and hence do not generate useful decision-making or insights. Furthermore, robust models for dynamic processes that feature incoming data streams that do not express learning via this information miss out on the significant potential to modify their decision schemes in the presence of new, better model estimates.

Due to the unintended negative consequences of state of the art robust techniques, this research will apply techniques that (1) allow for the decision-maker to express a spectrum of pessimism while (2) still utilizing incoming data for learning. Specifically, we explore two different applications with respect to these goals. The first application is a percentile optimization technique with respect to a multi-class queueing system for application in hospital Emergency Departments (EDs). The second investigates the use of robust forecasting techniques in improving developing countries' vaccine supply chains via (1) a non-traditional "outside of cold chain" policy and (2) a district-managed supply chain network. Both of these research application areas utilize data-driven approaches that feature learning and pessimism-controlled robustness.

1.2 Dissertation Outline

The dissertation is divided into five chapters: Chapter 1 provides the overview and the outline of the dissertation. Chapter 2 studies our application in robust multi-class queueing systems. Such models typically experience ambiguity in real-world settings in the form of unknown parameters, hence we incorporate robustness in the control policies by applying a novel data-driven percentile optimization technique that allows for (1) expressing a controller's optimism level toward ambiguity, and (2) utilizing incoming data in order to learn the true system parameters. Our contributions include showing that the optimal policy under the percentile optimization objective is related to a closed-form priority-based policy. We also identify connections between the optimal percentile optimization and $c\mu$-like policies, which in turn enables us to establish effective but easy-to-use heuristics for implementation in complex systems. Using real-world data collected from a leading U.S. hospital, we also apply our approach to a hospital ED setting, and demonstrate the benefits of using our framework for improving current patient flow policies.

In Chapter 3, we examine the potential of utilizing thermostability in developing countries' vaccine supply chains. Providing immunizations to many developing countries with limited infrastructure is complicated by maintaining the cold chain. For the purposes of increasing vaccination coverage and the potential for the introduction of new vaccines, we utilize the thermostable properties of vaccines in the last mile of delivery which enhances the flexibility and capacity of the supply chain. To combat the inherent ambiguity arising from uncertain demand while maximizing vaccination coverage and keeping wastage costs under control, we develop a robust constrained newsvendor model using distributionally robust optimization procedures.

In Chapter 4, we consider managing the vaccine supply chain at the district level via a "push" mechanism from vaccine depots to IHCs as opposed to traditional "pull" strategies from the IHCs. The last mile of the vaccine supply chain is well-known to suffer from poor data quality, limited transportation capacity, and a lack of managerial oversight. To help tackle these issues, we consider district led immunization delivery system, where vehicles at the district depot routinely supply vaccines to IHCs in its service area, where demand rates at each IHC are either fully observed or are assessed via a Bayesian approach. We develop and test effective policies and heuristics based on lower bounds that can become tight under high population densities.

Finally, Chapter 5 summarizes our contributions stemming from Chapters 2-4 and concludes the thesis.

Chapter 2

DATA-DRIVEN PERCENTILE OPTIMIZATION FOR MULTI-CLASS QUEUEING SYSTEMS WITH MODEL AMBIGUITY

2.1 Introduction

Multi-class queueing systems require dynamic control in environments where servers must process multiple types of jobs that vary with respect to holding costs, service rates, and other defining characteristics. These types of queueing systems are widely used to model call centers, hospitals, manufacturing lines, and service operations, where elements in the queue can be classified based on differing levels of urgency, processing time, or other attributes. For example, in a hospital Emergency Department (ED), patients are classified through a triage system, which differentiates them based on their severity, medical complexity, or other conditions (see, e.g., Saghafian *et al.* (2012), Saghafian *et al.* (2014), and the references therein). Hence, a natural way to analyze ED patient flow is via a multi-class queueing system which separates patients based on their attributes. [1]

In such systems, when all parameters are known, many well-established policies like the $c\mu$ rule have been shown to be optimal for optimizing the system's performance (see, e.g., Van Mieghem (1995) and Buyukkoc *et al.* (1985)). However, the assumption that all the model parameters are perfectly known is often unrealistic, especially in settings with little supporting data, inaugural system launch, or various other sources of ambiguity. A manager with incorrect parameter specifications may enforce policies that perform poorly, or may not have confidence in using a policy that is obtained

[1] See, e.g., Saghafian *et al.* (2015) for a recent review of various models used to optimize patient flow and improve ED operations.

from a model with parameters that s/he does not fully trust. In an effort to combat such mistrust, we consider a form of model ambiguity caused by the ambiguity in parameters termed *parameter ambiguity*, and develop strategies that directly take these into account.

Traditionally, robust optimization protects against parameter ambiguity by utilizing a minimax objective on an ambiguity set of parameters which are assumed to contain the true system parameters. However, this type of robustness (a) can result in overly pessimistic policies and (b) ignores the significant potential to learn about the true system parameters from data acquired both before and after system launch. Even when this pessimism is reduced by choosing tighter ambiguity sets, the policies generated are not capable of learning from incoming data. To avoid these deficiencies, we model parameter ambiguity via a Partially Observable Markov Decision Process (POMDP), an extension of Markov Decision Processes (MDPs), which allows for (a) imperfect state knowledge, and (b) learning in a Bayesian manner. A POMDP supports the distribution of the underlying system parameters, known as the belief space, and updates this distribution to reflect received observations. This is ideal from a learning perspective; however, in a POMDP, the decision-maker is assumed to have an initial prior belief which is often a subjective value, guided by scarce data, error-prone expert opinion, intuition, or instinct. For these reasons, Bayesian critics distrust such learning mechanisms, citing the unreliability of the prior specification in real-world applications [2] .

To incorporate robustness to such a prior belief (hence gaining robustness to parameter ambiguity), we integrate our POMDP model with a *percentile optimization* approach. Percentile optimization is traditionally used to avoid overly conservative

[2] Though we mainly focus on a queueing model, our approach can be used for the general class of Bayesian decision-making problems where the decision-maker faces ambiguity with respect to parameters that shape his/her prior (see Corollaries A.1 and A.2 in Online Appendix A.2).

policies by offering a certain level of performance over a percentage of the ambiguity set (see, e.g., Delage and Mannor (2010) and Nemirovski and Shapiro (2006)). We extend percentile optimization in order to incorporate robustness to the belief about the model parameters rather than relying on a robustness generated directly from the parameters themselves. In this way, we investigate strategies where the controller learns the main model parameters (e.g., unknown service rates) while simultaneously controlling the underlying system for superior performance, which contrasts with robust techniques that only focus on parameter ambiguity sets. Thus, our framework allows generating policies that are robust to parameter ambiguities (considering a manager's pessimism level), while simultaneously learning about the true model from data/observation of the system's performance in a Bayesian manner.

Our main contributions stem from extending the robust percentile optimization approach for integration with POMDPs. We find that the percentile optimization objective reduces to the minimax and minimin objectives when the optimism level is set to its lowest and highest values, respectively and show that the optimal policies under these objectives are myopic $c\mu$ priority policies. Understanding the non-robust problem (which assumes a specified initial belief) proves to be essential in finding robust policies where the belief is subject to ambiguity. We find that optimal robust policies can be formed using specific non-robust policies via a geometric structure known as the *convex floating body.* Therefore, to solve the robust percentile problem, we first solve the non-robust problem that has a known initial belief. As the rate of observations increases, we find that a priority-based policy that acts as an extension of the well-known $c\mu$ rule becomes asymptotically optimal to the non-robust problem. This policy, which we term E$c\mu$, is myopic and prioritizes the class with the largest expected $c\mu$ value. The proposed E$c\mu$ policy utilizes incoming data for learning (unlike the traditional $c\mu$ rule), and is extremely simple to implement.

Due to its foundation in POMDPs, the robust framework we consider is computationally ambitious and necessitates finding tractable methods for implementation. Using the analytical insights gained from the connection between non-robust and robust policies, constraints via the convex floating body, and the relation of E$c\mu$ to the non-robust objective, we develop a heuristic for the robust problem that (a) is highly scalable to large problem instances, and (b) shows strong performance in extensive simulation experiments. We also develop analytical bounds to the non-robust problem based on queueing systems with fully known parameters. These bounds are (a) tight under a variety of conditions, and (b) can be used to more effectively compute optimal robust policies. Furthermore, since the bounds are based on non-learning policies, they can be computed in an efficient manner.

Finally, we demonstrate the benefits of our approach in a real-world setting by utilizing data that we have collected from a leading U.S. hospital, and by establishing the advantages of using our framework in improving the current ED patient flow policies. Our percentile optimization framework is the first study in the literature to yield data-driven policies for use in EDs that hedge against parameter ambiguity. We find that highly congested EDs are well-suited to our percentile optimization framework, especially in geographical areas with uncertain/unstable patient population characteristics. Additionally, our approach explicitly avoids overly conservative policies that focus only on the "worst-case" scenarios. As a result, we find that percentile optimization performs well over a large spectrum of optimism/pessimism. In particular, our simulations calibrated with hospital data suggest that, by using our approach, an ED manager can significantly improve performance regardless of his/her disposition.

The rest of the chapter is organized as follows. In Section 2.2, we provide a literature review of the related studies. Section 2.3 introduces the non-robust continuous-time formulation of our problem, which is uniformized into a discrete-time problem in Section 2.3.1, and lays the foundation for the percentile framework developed in Section 2.3.2. We provide the majority of our analytical insights in Sections 2.4 and 2.5, where we establish optimal policies for the non-robust and robust formulations, and identify upper/lower bound results. Section 2.6 introduces a heuristic to the robust problem that is rooted in the analytical insights generated from Section 2.4. In Section 2.7, we present various numerical experiments, discuss the application of our work for improving patient flow in EDs, and use real-world data obtained from a leading U.S. hospital to evaluate the potential benefits of our approach. Finally, in Section 2.8, we present our concluding remarks.

2.2 Literature Review

The literature surrounding multi-class queueing systems aims to analyze complex structures and discover their optimal control policies such as the $c\mu$ policy and its variations (see, e.g., Buyukkoc *et al.* (1985), Van Mieghem (1995), Saghafian and Veatch (2016), and the references therein). A common tool used to analyze and control such systems is Markov Decision Processes (MDPs). However their use is limited to the unrealistic case where the decision-maker is assumed to completely know all the parameters of the model (e.g. service rates). Most notably, this includes a perfect knowledge assumption of the transition matrices that guide a system's state transitions. This assumption can be problematic in various practical applications in which service rates (or other parameters) are not perfectly known. Mannor *et al.* (2007) and Nilim and El Ghaoui (2005) found that small changes in such parameters can result in significant differences in decision-making strategies. However, a synthesis of most studies on dynamic control in queueing systems indicates the use of tools that heavily rely on a full knowledge about the system's parameters. This is despite the fact that in practice such parameters are typically unknown and often hard to estimate.

Robust methods applied to queueing models are largely involved with reducing the computational burden of characterizing queueing metrics and policies. Su (2006) studies a fluid approximation of a multi-class queueing model's holding cost under a robust paradigm established by Bertsimas and Sim (2004a) and Bertsimas and Sim (2004b). Bertsimas *et al.* (2011) focuses on finding bounds for performance measures through a method rooted in robust optimization, and studies the performance of this method on tandem and multi-class single server queueing networks. Jain *et al.* (2010) finds that a queueing network with control over traffic intensities has a simple threshold type policy under a robust objective. For more recent studies on robust techniques used in queueing systems we refer to Pedarsani *et al.* (2014), Bandi and Bertsimas (2012), Bandi *et al.* (2015), and the references therein. This stream of research is mainly aimed at increasing tractability by focusing on "worst-case" (i.e., fully pessimistic) scenarios, and establishing related performance metrics. Unlike this stream, our goal is to provide policies that (a) are more optimistic (i.e., less conservative), and (b) incorporate learning from online system-run data/observations.

Adding robustness when facing parameter ambiguity is a topic of significant interest to a variety of fields including economics, operations research/management, computer science, and decision theory among others. Typically, robustness in MDPs is added using a "minimax" objective, since this often results in tractable analyses as shown in Nilim and El Ghaoui (2005), Iyengar (2005), and the references therein. Other studies such as Chen and Farias (2013) deal with ambiguities by considering policies that offer guarantees on expected performance. Still other methods of incorporating robustness include regret minimization (Lim *et al.* (2012)), relative entropy (Bagnell *et al.* (2001)), and martingale-based approaches (Hansen and Sargent (2007)) that provide less conservative, and hence, potentially more realistic alternatives to minimax techniques. In particular, Delage and Mannor (2010) identify a robust approach applied to MDPs called percentile optimization that effectively avoids over-conservatism (see also Nemirovski and Shapiro (2006) and Wiesemann *et al.* (2013) for related studies). Instead of finding policies that are tailored to work well in worst-case scenarios, the percentile optimization method finds policies that maximize performance with respect to a level of belief about the true parameters for a given level of optimism. [3]

Chow *et al.* (2017) also utilize this type of robustness to develop risk-constrained policies for MDPs. However, a significant deficit in current percentile optimization approaches is the lack of ability to *learn* about the true parameters over time. Delage and Mannor (2007) work to fill this gap via a similar formulation to our approach, and find second-order approximations to MDPs that experience transition parameter uncertainty. However, the Dirichlet-type uncertainty assumed in transition parameters does not fit our queueing problem, and in our work, we extend the percentile optimization approach with respect to ambiguity in the initial belief. Thus, system data/observations can be used for learning the true operational model, and as we will show, this ability to learn itself adds a strong layer of robustness for controlling queueing systems (e.g., hospital patient flows) that face parameter ambiguity. Learning to overcome ambiguities are also discussed in Bassamboo and Zeevi (2009), which models a call center application using a data-driven technique. However, their work (a) does not include any notion of robustness, and (b) focuses on near-optimal policies with performance bounds. Our work differs in modeling approach by our joint focus on learning and robustness, and in methodology by our contributions in characterizing the exact optimal policies.

[3] The percentile objective originally arose in single-period contexts (see, e.g., Charnes and Cooper (1959) and Prékopa (1995)).

Data-driven parameter learning has been incorporated in POMDPs: Ross *et al.* (2011) explores a finite-horizon POMDP model that updates a posterior of its parameter belief in a Bayesian manner, and Thrun (1999) investigates a POMDP in continuous action and state spaces that relies on particle filtering techniques to determine the belief state. Unlike learning mechanisms, robust methods are almost non-existent in POMDP frameworks. Osogami (2015) shows that traditional minimax approaches with convex ambiguity sets can be extended to POMDPs while still retaining its structural features (such as convexity). In a new approach, Saghafian (2018) extends POMDPs to a new class termed Ambiguous POMDPs (APOMDPs) which incorporates ambiguity in transition and observation probabilities in a robust fashion. The robustness in Saghafian (2018) is achieved by considering α-maximin (α-MEU) preferences, and by incorporating the decision-maker's temperament toward model ambiguity. Different from the APOMDP approach of Saghafian (2018), we utilize a percentile optimization objective to hedge against ambiguities.

2.3 The Multi-Class Queueing Control Problem with Parameter Ambiguity

We begin by considering a continuous time multi-class queueing control problem with preemption, where a single [4] server is responsible for serving n classes of customers over an infinite time horizon. Unlike the traditional version of this model, we assume the controller does not know the main parameters of the system, and hence, is faced with parameter ambiguity. We focus on the case where the ambiguity is on service rates. To this end, we start by excluding dynamic arrivals to the system, and instead

[4] For analytical tractability, we restrict our attention to single-server scenarios. Cases with multiple servers may interfere with some of our main analytical results, notably the relation to multi-armed bandit problems and the optimality of $c\mu$-like policies. In Section 2.7, we investigate the robustness of the insights we gain via simulation experiments.

consider a *clearing system* [5] version of the problem. We relax this assumption in Sections 2.7 and A.1.4 by allowing for dynamic arrivals, and find that many of our major results are transferable from the clearing system. Our general approach can also be used for systems where arrival rates or other parameters are ambiguous by modifying the underlying dynamic program to include these components along with their learning mechanisms. However, this appears to increases the problem's complexity without providing additional insights.

With $\mathcal{N} = \{1, \ldots, n\}$ denoting the set of customer classes, we assume each customer of class $i \in \mathcal{N}$ accrues a cost $\hat{c}_i > 0$ for each unit of time spent in the system. Let $\hat{\mathbf{c}} = (\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_n)$ be the cost vector, $\alpha \in (0, \infty)$ the discount rate, and $\mathbf{X}(t) = (X_1(t), X_2(t), \ldots, X_n(t))$ the vector of the number of customers in the system, where $X_i(t)$ is number of class i customers in the system at time t. In line with many robust approaches, we begin by outlining an ambiguity set (i.e. a "cloud" of models) that is assumed to include the true model. To this end, and for tractability, we assume service times for each class are i.i.d. exponential [6] random variables with unknown rates for each class. The true service rate for each class $i \in \mathcal{N}$ is chosen by Nature at time $t = 0$, and lies within ambiguity set [7] $\mathcal{M}_i = \{\hat{\mu}_{i,1}, \ldots, \hat{\mu}_{i,m_i}\}$. We further assume that service times for different classes are independent. For future notational convenience, we let $\mathcal{J}_i = \{1, \ldots, m_i\}$. Throughout the chapter, we assume $m_i \in \mathbb{N}$, and $\hat{\mu}_{i,j} \neq \hat{\mu}_{i,k}$ for each $i \in \mathcal{N}$ and distinct $j, k \in \mathcal{J}_i$. Though the ambiguity sets $\mathcal{M}_i$ are discrete, the continuous case can be approximated arbitrarily closely by

[5] Clearing systems are typically used to model busy periods by focusing on the customers/jobs already in the system. The goal is then to clear the system with the minimum cost.

[6] In Section 2.7 we relax the exponential distribution assumption. For instance, our data shows that service times in EDs are close to log-normal. As we will show, our main insights and heuristic control procedures remain effective even when the service times are not exponential.

[7] The general nature of our ambiguity sets enhances the flexibility of our framework. For ambiguity sets reminiscent of other robust literature, we may choose to build each $\mathcal{M}_i$ to surround some nominal value estimated from historical data. This is in fact the strategy we use in our ED application of Section 2.7.1.

increasing the number of potential service rates m_i to make the mesh size of $\mathcal{M}_i$ close to zero.

Over time, the controller can learn the true service rates by observing the process history which includes all previous service durations, control actions, and observations of service completions. For Markovian systems with incomplete information, it has been shown in Bertsekas (1995) that the Bayesian belief on the unknown parameters with respect to the observed process history is a *sufficient statistic.* We let $\mathcal{B}$ be the set of all such sufficient statistics, i.e., the set of possible belief distributions on the system's service parameters. Letting $m = \sum_{i \in \mathcal{N}} m_i$, each $\mathbf{b} \in \mathcal{B}$ is an m-dimensional vector of the form $\mathbf{b} = (b_{1,1}, b_{1,2}, \ldots, b_{1,m_1}, b_{2,1}, \ldots, b_{n,m_n})$ with the condition that each $b_{i,j} \geq 0$ and that $\sum_{j=1}^{m_i} b_{i,j} = 1$ for each $i \in \mathcal{N}$. In this setting, if $\hat{\mu}_i^* \in \mathcal{M}_i$ is the true (unknown) service rate for class $i \in \mathcal{N}$, $\mathbb{P}\left(\hat{\mu}_{i,j} = \hat{\mu}_i^* | \mathbf{b}\right) = b_{i,j}$. We further assume that the observation made after serving one class does not affect the belief about another. This is aligned with the assumption that service time of one class is independent of that of another class.

To find policies that optimally prescribe which customer class the server should serve at any time, given (a) the available information summarized in the current belief about the service rates, and (b) the number of customers in each queue, it is known that one can restrict attention to policies that are deterministic, stationary, and Markovian (see, e.g., Sondik (1971), Smallwood and Sondik (1973), and Bertsekas (1995)). Consequently, an admissible non-anticipative policy π maps the current belief and queue length information (information state) to the set of actions: $\pi : \mathbb{Z}_+^n \times \mathcal{B} \rightarrow \mathcal{N} \bigcup \{0\}$, with the additional condition that π can serve only customer classes that have non-empty queues, and serves the fictitious class "0" when the server is idled (e.g., when all the queues are empty). Our model described above is schematically illustrated in Figure 2.1.

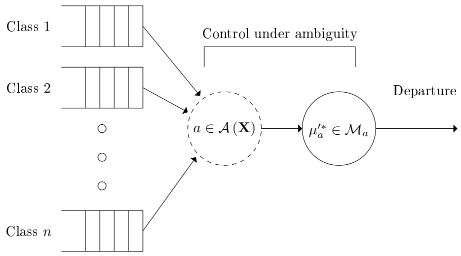


Figure 2.1: The server serves a class a customer with an unknown rate $\hat{\mu}_a^*$ belonging to ambiguity set $\mathcal{M}_a$.

We let Π be the set of all admissible policies, and $\mathbf{X}^\pi(t) = (X_1^\pi(t), X_2^\pi(t), \ldots, X_n^\pi(t)) \in \mathbb{Z}_+^n$ be the number of customers in the system under policy $\pi \in \Pi$ at time t. In Appendix A.2, Lemma A.17 shows that idling the server when at least one customer class queue is non-empty is always suboptimal; hence, we consider only non-idling policies in our analysis. For a given policy π, the expected discounted *true cost* the system experiences is

$$\mathrm{E}_\pi \left[\int_{t=0}^{\infty} e^{-\alpha t} \hat{\mathbf{c}} \mathbf{X}^\pi \left(t\right)^{\mathrm{T}} dt | \mathbf{X}(0) \right],$$

given the true transition parameters chosen by Nature at time $t = 0$, where the notation "T" represents transpose, and E_π is expectation with respect to the probability measure induced by π. However, since the controller does not know the true transition matrix (as service rates are unknown), we are interested in the expected cost with respect to the controller's belief:

$$\mathrm{J}^\pi \left(\mathbf{X}(0), \mathbf{b}(0)\right) = \mathrm{E}_{\pi, \mathbf{b}(0)} \left[\int_{t=0}^{\infty} e^{-\alpha t} \hat{\mathbf{c}} \mathbf{X}^\pi \left(t\right)^{\mathrm{T}} dt | \mathbf{X}\left(0\right) \right], \tag{2.1}$$

where $\mathrm{E}_{\pi, \mathbf{b}(0)}$ denotes expectation with respect to both the initial belief $\mathbf{b}(0)$ and π. We refer to $\mathrm{J}^\pi(\mathbf{X}(0), \mathbf{b}(0))$ as the *non-robust cost* under policy π, since it assumes a perfectly assigned $\mathbf{b}(0)$ (which is inevitably hard to quantify for any decision-maker who is faced with model ambiguity). The optimal non-robust cost is then given by $\mathrm{J}\left(\mathbf{X}(0), \mathbf{b}(0)\right) = \inf_{\pi \in \Pi} \mathrm{J}^\pi \left(\mathbf{X}(0), \mathbf{b}(0)\right)$. In what follows, we first use uniformization to work with the discrete-time model of the non-robust scenario, where the initial belief is given. We then adopt percentile optimization to enable the decision-maker/controller to reduce his/her reliance on $\mathbf{b}(0)$, and thereby make robust decisions.

2.3.1 A Discrete-Time Non-Robust Framework

The continuous-time Markov chain $\{\mathbf{X}^{\pi}(t) : t \geq 0\}$ can be converted to a discrete-time equivalent using the well-known uniformization technique (Lippman (1975)). Following this method, we first select a *uniformized* exponentially distributed random variable ξ with a rate $\psi > \max_{i \in \mathcal{N}, j \in \mathcal{J}_i} \hat{\mu}_{i,j}$ which serves as our rate of observations made as follows. If the server completes service to a customer of class i a uniformized unit of time (i.e., at the end of each period), an observation indicating the "successful" service to class i is recorded. Otherwise, if no service completion is observed within this time, an observation is recorded indicating an "incomplete" service to class i. We note that this *uniformization rate* ψ may be arbitrarily large so as to approximate continuous observations.

We let σ be the Bayesian learning operator such that $\sigma(\mathbf{b}, a, \theta)$ is an m-dimensional vector representing the updated belief after taking action a and receiving observation θ, when the prior belief is $\mathbf{b}$. Since there are only two outcomes for observations for any given action, we let "$+$" signify an observed service completion ("success") during the uniformized time period, and "$-$" represent an incomplete service ("failure") in that period. In this setting, we use a discrete-time dynamic program with uniformized parameters $\mu_{i,j} = \hat{\mu}_{i,j}/\psi$. For notational convenience, we let $\mathrm{E}[\mu_i|\mathbf{b}] = \sum_{j=1}^{m_i} \mu_{i,j} b_{i,j}$ be the expected service transition probability of class $i \in \mathcal{N}$ given belief $\mathbf{b}$. In this way, the Bayesian learning operator updates belief $\mathbf{b}$ with components $b_{i,j}$ to belief

$\bar{\mathbf{b}} = \sigma(\mathbf{b}, a, \theta)$ with components $\bar{b}_{i,j} = \sigma(\mathbf{b}, a, \theta)_{i,j}$, where $a, i \in \mathcal{N}, j \in \mathcal{J}_i$, and

$$\sigma(\mathbf{b}, a, +)_{i,j} = \begin{cases} \frac{\mu_{a,j} b_{a,j}}{\sum_{k=1}^{m_a} \mu_{a,k} b_{a,k}} = \frac{\mu_{a,j} b_{a,j}}{\mathrm{E}[\mu_a|\mathbf{b}]} & : i = a \\ b_{i,j} & : i \neq a \end{cases} \tag{2.2}$$

for a successful service observation, and

$$\sigma(\mathbf{b}, a, -)_{i,j} = \begin{cases} \frac{(1-\mu_{a,j}) b_{a,j}}{\sum_{k=1}^{m_a} (1-\mu_{a,k}) b_{a,k}} = \frac{(1-\mu_{a,j}) b_{a,j}}{(1-\mathrm{E}[\mu_a|\mathbf{b}])} & : i = a \\ b_{i,j} & : i \neq a \end{cases} \tag{2.3}$$

for a failed service observation. Equations (2.2) and (2.3) are established due to the fact that under realized parameter $\mu_{a,j}$, the probability of successful service in a given period is $\mu_{a,j}$ and probability of incomplete service is $(1-\mu_{a,j})$. With this, and defining a discrete-time discounting factor $\beta = \frac{\psi}{\psi+\alpha}$ and instantaneous cost $\mathbf{c}\mathbf{X}^{\mathrm{T}} = \frac{\hat{\mathbf{c}}\mathbf{X}^{\mathrm{T}}}{\psi+\alpha}$, where $\mathbf{X}$ is an n-dimensional vector representing queue lengths, we can identify the non-robust optimal policy and the associated cost via the dynamic program

$$\begin{aligned} \mathrm{V}_{t+1}(\mathbf{X}, \mathbf{b}) = \mathbf{c}\mathbf{X}^{\mathrm{T}} + \beta \Bigg[\min_{a \in \mathcal{A}(\mathbf{X})} \Big\{ & \mathrm{E}[\mu_a|\mathbf{b}] \, \mathrm{V}_t(\mathbf{X} - \mathbf{e}_a, \sigma(\mathbf{b}, a, +)) \\ & + \; (1 - \mathrm{E}[\mu_a|\mathbf{b}]) \, \mathrm{V}_t(\mathbf{X}, \sigma(\mathbf{b}, a, -)) \Big\} \Bigg], \end{aligned} \tag{2.4}$$

with the terminal condition $\mathrm{V}_0(\mathbf{X}, \mathbf{b}) = \mathbf{c}\mathbf{X}^{\mathrm{T}}$. In this setting, taking the limit as $t \to \infty$, we define $\mathrm{V}(\mathbf{X}, \mathbf{b}) = \lim_{t\to\infty} \mathrm{V}_t(\mathbf{X}, \mathbf{b})$, and note that $\mathrm{V}(\mathbf{X}, \mathbf{b}) = \inf_{\pi\in\Pi} \mathrm{J}^\pi(\mathbf{X}, \mathbf{b})$ (see Lemma A.11 in Online Appendix A.2 for a rigorous treatment), where $\mathrm{J}^\pi(\mathbf{X}, \mathbf{b})$ is defined in (2.1). To account for evaluating non-optimal policies, we let $\mathrm{V}^\pi_{t+1}(\mathbf{X}, \mathbf{b})$ be a value function similar to that of the dynamic program (2.4) with minimization operator replaced by serving the class prescribed by policy π. Likewise, we let $\mathrm{V}^\pi(\mathbf{X}, \mathbf{b}) = \lim_{t\to\infty} \mathrm{V}^\pi_t(\mathbf{X}, \mathbf{b})$ be the infinite-horizon dynamic program value function under policy π.

2.3.2 Gaining Robustness via Percentile Optimization

Since the controller is facing ambiguity with respect to the true model, s/he may distrust his/her initial prior on the cloud of models, $\mathbf{b}(0)$. The specification of $\mathbf{b}(0)$ is subject to model sensitivities, especially in applications in which there is little or highly variable data to perfectly quantify it. Often, the selection of a prior is a process that requires sussing out probabilities and parameter values from experts in the field, which can be a highly subjective and inaccurate task [8] .

In traditional robust optimization, one would choose a policy assuming that Nature, being an antagonistic character, picks the worst-case initial belief vector $\mathbf{b}(0)$ for a chosen policy. Hence, the traditional *minimax* robust objective can be defined by first considering the worst-case cost under a policy $\pi \in \Pi$:

$$\mathrm{R}^{\pi}\left(\mathbf{X}\right) = \max_{\mathbf{b}\in\mathcal{B}} \mathrm{V}^{\pi}\left(\mathbf{X}, \mathbf{b}\right).$$

The cost under the minimax robust objective is then $\mathrm{R}\left(\mathbf{X}\right) = \inf_{\pi\in\Pi} \mathrm{R}^{\pi}\left(\mathbf{X}\right)$. In this setting, the controller assumes that Nature will pick the transition parameters that result in the maximum cost for any given policy, and chooses a policy that minimizes the cost of this worst-case outcome.

In sharp contrast to this type of robustness, which typically yields overly pessimistic control policies, is the overly optimistic *minimin* objective defined by:

$$\mathrm{N}^{\pi}\left(\mathbf{X}\right) = \min_{\mathbf{b}\in\mathcal{B}} \mathrm{V}^{\pi}\left(\mathbf{X}, \mathbf{b}\right),$$

and $\mathrm{N}\left(\mathbf{X}\right) = \inf_{\pi\in\Pi} \mathrm{N}^{\pi}\left(\mathbf{X}\right)$, under which the controller chooses a policy assuming Nature picks the transition parameters resulting in the best-case cost for any given policy. In what follows, we first show that both minimax and minimin optimal policies

[8] This is indeed a general criticism to Bayesianism and goes well beyond the queueing setting of this thesis.

are within the well-known class of $c\mu$ policies. Thus, they (a) are fully myopic, and (b) have very simple forms.

Proposition 2.1 (**Minimax/Minimin $c\mu$ Optimal Policies**). *At any state* $(\mathbf{X}, \mathbf{b})$*, optimal policies to the minimax and minimin objectives serve classes* $\arg\max_{a\in\mathcal{A}(\mathbf{X})} \left(\min_{j\in\mathcal{J}_a} c_a\mu_{a,j}\right)$ *and* $\arg\max_{a\in\mathcal{A}(\mathbf{X})} \left(\max_{j\in\mathcal{J}_a} c_a\mu_{a,j}\right)$*, respectively.*

Proposition 2.1 establishes that optimal policies under both minimax and minimin objectives are myopic priority disciplines (known as the $c\mu$ rule) with respect to the smallest and largest transition rates within the ambiguity set for each class, respectively. However, it should be noted that such policies (a) ignore the potential for learning from the system behavior, and (b) only consider the potentially unrealistic extreme best and worst-case scenarios and can perform poorly in real-world applications. To address this deficit, we next investigate how the percentile optimization approach provides a balancing alternative between these two extreme strategies, while incorporating learning about the hidden probabilities associated with the true transition parameters (i.e., service rates).

To this end, for a given $\epsilon \in [0, 1]$, we define the percentile optimization program:

$$\mathrm{Y}^{\pi}(\mathbf{X}, \epsilon) = \inf_{y_\epsilon \in [\mathrm{N}^{\pi}(\mathbf{X}), \mathrm{R}^{\pi}(\mathbf{X})]} y_\epsilon \tag{2.5}$$

$$s.t.\ \mathbb{P}_{\mathbf{B}}\left(\mathrm{V}^{\pi}\left(\mathbf{X}, \mathbf{B}\right) \le y_\epsilon\right) \ge 1 - \epsilon, \tag{2.6}$$

and let $\mathrm{Y}\left(\mathbf{X}, \epsilon\right) = \inf_{\pi_{h\in\Pi}} \mathrm{Y}^{\pi}\left(\mathbf{X}, \epsilon\right)$ represent the optimal percentile objective. In (2.5), we impose that $\mathrm{N}^{\pi}\left(\mathbf{X}\right) \le y_\epsilon \le \mathrm{R}^{\pi}\left(\mathbf{X}\right)$ so that the value of the objective is within the most optimistic and pessimistic values attainable for any given belief in accordance with the policy, hence enforcing "realizable" expected costs. The probability operator, $\mathbb{P}_{\mathbf{B}}$, in (2.6) is defined with respect to a specified probability density function over the prior belief space [9] , where $\mathbf{B}$ is a random variable whose realization is $\mathbf{b}$. The

[9] One may criticize the use of the percentile objective due to the potential ambiguity of $\mathbb{P}_{\mathbf{B}}$;

percentile optimization program (2.5)-(2.6) allows us to find a *chance-constrained* policy: it emphasizes policy performance over a portion of the belief space. We thus term the policy that is the solution under the optimal percentile objective as $(1-\epsilon)\%$ *chance-constrained policy.* Intuitively, the smaller the ϵ, the more protection from poor parameter settings since the proportion of the belief space that performs worse than y_ϵ becomes smaller.

It is important to note that the percentile objective acts as a bridge between non-robust and robust objectives; expressing a manager's optimism level is a core ambition of this type of robustness. For instance, the chance-constrained policy reduces to the minimax and minimin policies when ϵ is 0 and 1, respectively.

Proposition 2.2 (**Percentile/Minimax/Minimin Relationship**). *The percentile objective, minimax, and minimin policies share the following relation:*

(i) If $\epsilon = 0$ and $\mathbb{P}_{\mathbf{B}}(\mathbf{B} = \mathbf{b}) > 0$ for all $\mathbf{b} \in \mathcal{B}$, then the optimal policy and cost under both minimax and percentile objectives are the same.

(ii) If $\epsilon = 1$, then the optimal policy and cost under the minimin and percentile objectives are the same.

The additional condition $\mathbb{P}_{\mathbf{B}}(\mathbf{B} = \mathbf{b}) > 0$ for all $\mathbf{b} \in \mathcal{B}$ in part (i) is necessary, since $\mathbb{P}_{\mathbf{B}}$ with zeros allows percentile objective to "ignore" certain portions of the belief space while still satisfying constraint (2.6). For example, if $\mathbb{P}_{\mathbf{B}}$ is the degenerate distribution with respect to a point $\mathbf{b}$, $\mathrm{Y}(\mathbf{X}, 0) = \mathrm{V}(\mathbf{X}, \mathbf{b})$.

however, it should be noted that this is a second-order distribution, and perturbations in $\mathbb{P}_{\mathbf{B}}$ result in very similar convex floating bodies, which is the geometric structure investigated in Section 2.4 that generates our optimal robust policies.

2.4 Structure of Optimal Policies under the Percentile Objective

Analyzing program (2.5)-(2.6) is inherently complex both analytically and computationally. However, we find that the solution to this program is linked to solving the non-robust problem. Hence, we first consider the solution of the dynamic program (2.4), identify important characteristics of these solutions over the belief space, establish the link between non-robust and robust policies, and finally work to characterize optimal percentile policies. In Section 2.6, we develop an easy-to-use heuristic based on these insights to facilitate tractable solutions.

As the observation rate increases, tending toward continuous observations, the non-robust problem can be transferred to a multi-armed bandit (MAB) problem by noting that (a) under any action, only the belief about transition parameters and number of customers in the served class (the "arms" of the MAB) change, and (b) the "discounted cost" can be reinterpreted as "discounted savings" of the MAB due to our clearing system environment (for further discussion, see Lemma A.3 in Online Appendix A.2). MAB problems are typically solved by indexing policies related to the expected savings in cost experienced through exclusively serving one class over time.

To take advantage of the above-mentioned connection, we term the myopic policy that serves the class $a \in \mathcal{A}(\mathbf{X})$ with largest value of $c_a \mathrm{E}\left[\mu_a | \mathbf{b}\right]$ the "E$c\mu$" policy. Thus, we denote $\pi^{c\mu}$ that serves $\arg\max_{a \in \mathcal{A}(\mathbf{X}(t))} c_a \mathrm{E}\left[\mu_a | \mathbf{b}(t)\right]$ as the E$c\mu$ policy. This policy can be viewed as an extension of the traditional $c\mu$ policy (often seen in the literature surrounding control of multi-class queueing systems) for queueing systems with ambiguous parameters. [10] The expectation operator in this policy dynamically combines all the possible $c\mu$ values for each class based on the belief at time t. In

[10] Argon and Ziya (2009) demonstrate the optimality of a similar policy in an average-cost non-learning queueing environment when service rates are known, but customer class is not fully observed.

the following theorem, we show that the E$c\mu$ policy is asymptotically optimal for the non-robust problem as the observation rate increases.

THEOREM 2.1 (E$c\mu$ **Asymptotic Optimality**). *The* E$c\mu$ *policy* $\pi^{c\mu}$ *is asymptotically optimal for the non-robust problem:* $\lim_{\psi\to\infty} \mathrm{V}^{\pi^{c\mu}}(\mathbf{X},\mathbf{b}) = \lim_{\psi\to\infty} \mathrm{V}(\mathbf{X},\mathbf{b})$ *for* $\mathbf{X} \in \mathbb{Z}_+^n$ *and* $\mathbf{b} \in \mathcal{B}$.

Theorem 2.1 is surprising in its simplicity since problems based on POMDP formulations typically do not yield closed-form results. In contrast to the usual complexities, the asymptotic optimality of the E$c\mu$ policy implies that the only information necessary to make decisions is the expected transition rates among non-empty queues. Therefore, queue lengths are essentially irrelevant to the decision-maker. Rather, the E$c\mu$ policy features a momentum property; if the current action a prescribed by the policy yields enough successes so that $c_a\mathrm{E}[\mu_a|\mathbf{b}]$ does not fall below the threshold defined by $c_{\hat{a}}\mathrm{E}[\mu_{\hat{a}}|\mathbf{b}]$ of the next highest available class $\hat{a}$, the E$c\mu$ policy will continue to serve class a regardless of the state of other classes. In turn, this means that the policy will not attempt to serve a class with smaller $c_{\hat{a}}\mathrm{E}[\mu_{\hat{a}}|\mathbf{b}]$ until other classes with larger values have experienced a sufficient number of service failures, or have cleared their queue. This property may run counter-intuitive to the exploration-minded individual; even if a class has the potential to be endowed with a very large $c_a\mu_{a,j}$ value (under the realization of system parameters), this potential is only rated on the basis of its contribution to the expected service rate.

Another important property of the E$c\mu$ policy is that under mild conditions, $\mathrm{V}^{\pi^{c\mu}}(\mathbf{X},\mathbf{b})$ is piecewise-linear over the belief space (excluding beliefs near edges and faces of $\mathcal{B}$). [11]

[11] An infinite horizon POMDP value function is not always guaranteed to be piecewise-linear (see, e.g. White and Harrington (1980)).

Proposition 2.3 (**Piecewise-Linearity of the Approximate Non-Robust Value Function**). *Let $\mathcal{B}'$ be any closed subset of $\mathcal{B}$ such that for any $\mathbf{b} \in \mathcal{B}', b_{i,j} > 0$ for all $i \in \mathcal{N}, j \in \mathcal{J}_i$. If $\min_{j \in \mathcal{J}_i} c_i \mu_{i,j} \neq \min_{j \in \mathcal{J}_k} c_k \mu_{k,j}$ for any distinct pair $i, k \in \mathcal{N}$, then $\mathrm{V}^{\pi^{c\mu}}(\mathbf{X}, \mathbf{b})$ is piecewise-linear on $\mathcal{B}'$.*

This result is related to two facts: (*i*) for any given initial prior $\mathbf{b} \in \mathcal{B}'$ (and $\mathbf{X} \in \mathbb{Z}_+^n$), the E$c\mu$ policy is unique, unless $\mathbf{b}$ lies on the break-points of the piecewise-linear function $\mathrm{V}^{\pi^{c\mu}}(\mathbf{X}, \mathbf{b})$ (see Lemma A.7 and 2.3 in Online Appendix A.2), and (*ii*) policies can be evaluated as linear functions of the belief in any POMDP. Therefore, with respect to closed, non-zero portions of the belief space, the value function $\mathrm{V}^{\pi^{c\mu}}(\mathbf{X}, \mathbf{b})$ is differentiable (except at breakpoints). As we will show in Theorem 2.2, the differentiability of the value function strongly enhances the relationship between optimal policies of the non-robust problem and those under the robust percentile optimization program (2.5)-(2.6). Thus, in identifying an asymptotically optimal policy that exhibits this property enables us to solve the robust percentile optimization program in an efficient way. This is an important insight to our search for robust chance-constrained policies especially since, as Zhang (2010) states, there are no known general conditions over which a POMDP value function is differentiable on its entire belief space.

To the purpose of finding robust chance-constrained policies, we introduce the following set of policies. Fix the initial $\mathbf{X}$, and let $\mathcal{K}_{\mathbf{b}} = \left\{\pi_{\mathbf{b}}^1, \pi_{\mathbf{b}}^2, \ldots, \pi_{\mathbf{b}}^k\right\}$ be any finite set of optimal policies to the non-robust problem when the initial prior is $\mathbf{b}$, and $\mathbf{p} = (p_1, p_2, \ldots, p_k)$ be an associated distribution such that $\sum_{i=1}^k p_i = 1$. We define a policy $\pi_{\mathcal{K}_{\mathbf{b}}}^{\mathbf{p}}$ to be a *randomized policy*, if at time 0, an element of $\mathcal{K}_{\mathbf{b}}$, $\pi_{\mathbf{b}}^i$, is chosen with probability p_i, which will dictate all current and future decisions. [12]

[12] For these randomized policies, we disallow policies that are not picked at time zero for the purpose of targeting specific contours of the value function.

Interestingly, similar to other non-learning robust problems (see, e.g., Bertsimas and Thiele (2006)), we find that there exists a randomized policy that forms an optimal solution to the robust percentile problem. This means that there exists an optimal robust policy that randomizes between optimal non-robust policies obtained for a single belief point $\mathbf{b} \in \mathcal{B}$. Furthermore, we shed light on conditions (associated with the differentiability of $\mathrm{V}(\mathbf{X}, \mathbf{b})$ with respect to the belief space) such that a *deterministic* non-robust policy is optimal even for the robust percentile problem.

THEOREM 2.2 (**Chance-Constrained Policy**). *For any given* $\epsilon \geq 0$*, there exists a* $\mathbf{b}^* \in \mathcal{B}$ *and a distribution* $\mathbf{p}^*$ *forming a randomized policy* $\pi^{\mathbf{p}^*}_{\mathcal{K}_{\mathbf{b}^*}}$ *that is optimal under the percentile optimization program* (2.5)-(2.6) [13] *:* $\mathrm{Y}^{\pi^{\mathbf{p}^*}_{\mathcal{K}_{\mathbf{b}^*}}}(\mathbf{X}, \epsilon) = \mathrm{Y}(\mathbf{X}, \epsilon) = \mathrm{V}(\mathbf{X}, \mathbf{b}^*)$*. Furthermore, if* $\mathrm{V}^{\pi_{\mathbf{b}}}(\mathbf{X}, \mathbf{b})$ *is differentiable at* $\mathbf{b}^*$*, then* $\mathcal{K}_{\mathbf{b}^*}$ *consists of a single policy, and hence,* $\pi^{\mathbf{p}^*}_{\mathcal{K}_{\mathbf{b}^*}}$ *is deterministic.*

The above result significantly reduces the complexity of the search for optimal robust policies. Importantly, it implies that we can combine policies associated with the function $\mathrm{V}(\mathbf{X}, \mathbf{b}^*)$ to find chance-constrained policies. In this way, we no longer need to look at the general space of policies, but rather can focus on the class of non-robust optimal policies. Moreover, Proposition 2.3 shows that the differentiability condition of Theorem 2.2 can be met by a surface that converges to the value function. If $\mathbf{b}^*$ lies on a linear segment of the value function that is not a breakpoint, $\mathcal{K}_{\mathbf{b}^*}$ can be composed of a single policy yielding a deterministic chance-constrained policy. Hence, under this assumption, one need not be concerned with finding $\mathbf{p}^*$.

However, Theorem 2.2 leaves us with an important question: what belief, $\mathbf{b}^*$, should be used to form the chance-constrained policy $\pi^{\mathbf{p}^*}_{\mathcal{K}_{\mathbf{b}^*}}$ for a given percentile problem? If such a $\mathbf{b}^*$ is characterized, then the solution to the percentile problem

[13] For notational convenience, we suppress the dependency of $\mathbf{p}^*$ and $\mathbf{b}^*$ on ϵ.

can easily be found by a randomization of non-robust policies associated with $\mathbf{b}^*$. The answer to this question turns out to be closely related to the geometrical concept of the *convex floating body* first discussed by Dupin (1822), and later used in robust literature to generate ambiguity sets that guarantee performance for policies evaluated within these sets (see, e.g., Lagoa *et al.* (2005) and Bertsimas *et al.* (2013)). However, we utilize the convex floating body in order to characterize $\mathbf{b}^*$, which generates a policy satisfying the chance-constrained objective.

Definition 2.1 (**Convex Floating Body**). *Let $\mathcal{W}_\epsilon = \{(\mathbf{w}, w) \in \mathbb{R}^m \times \mathbb{R} : \mathbb{P}_{\mathbf{B}}\left(\mathbf{B}\mathbf{w}^{\mathrm{T}} \geq w\right) \leq \epsilon\}$ be the set of all half spaces that "cut off" ϵ or less volume of the belief space $\mathcal{B}$ with respect to $\mathbb{P}_{\mathbf{B}}$. An ϵ-based convex floating body on $\mathcal{B}$ is $\mathcal{L}_\epsilon = \bigcap_{\{\mathbf{w},w\}\in\mathcal{W}_\epsilon} \left\{\mathbf{b} \in \mathcal{B} : \mathbf{b}\mathbf{w}^{\mathrm{T}} \leq w\right\}$. We let $\delta\mathcal{L}_\epsilon$ be the boundary of $\mathcal{L}_\epsilon$ [14] .*

Based on the above definition, a convex floating body is the region left from hyperplanes "cutting off" a specified volume (ϵ) from an object. For every $\mathbf{b} \in \delta\mathcal{L}_\epsilon$, there exists a hyperplane that divides $\mathcal{B}$ into two pieces, one which has volume less than or equal to ϵ. Figure 2.2 illustrates the convex floating body of a sphere with uniform density, which is either the empty set or another sphere. We study convex floating bodies with respect to the density measure $\mathbb{P}_{\mathbf{B}}$ on the belief space of our priors to characterize $\mathbf{b}^*$, and thereby find optimal chance-constrained policies as discussed in Theorem 2.2.

For the purposes of characterizing $\mathbf{b}^*$, it is important that $\mathcal{L}_\epsilon$ is non-empty. Fortunately, Fresen (2013) states that when $\mathbb{P}_{\mathbf{B}}$ is a log-concave probability distribution, $\mathcal{L}_\epsilon$ exists so long as $\epsilon \leq e^{-1}$. Hence, for many robust applications which tend toward pessimism (where ϵ is small), under common distributions, the convex floating body

[14] We note that if $\mathcal{L}_\epsilon$ is nonempty, $\delta\mathcal{L}_\epsilon$ always exists since closed, convex, and compact sets are equal to the convex hull of their boundary.

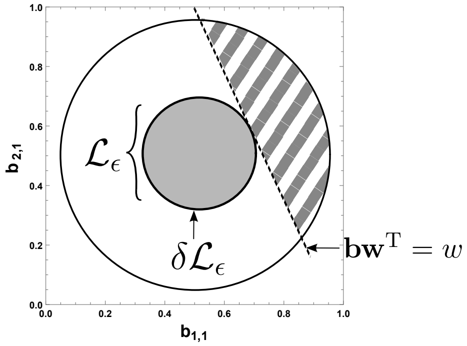


Figure 2.2: A convex floating body $\mathcal{L}_\epsilon$ when $\mathbb{P}_{\mathbf{B}}$ has uniform density within the circle and is zero elsewhere. It is generated from the intersection of halfspaces $(\mathbf{w}, w) \in \mathcal{W}_\epsilon$, and the striped area must contain less than or equal to ϵ volume. $(n = 2, m_1, m_2 = 2)$

is guaranteed to exist [15] . If $\mathcal{L}_\epsilon$ is nonempty, we find that $\mathbf{b}^*$ (defined in Theorem 2.2) is found at the largest value of the non-robust problem on the boundary of the convex floating body.

Proposition 2.4 (**Characterizing** $\mathcal{K}_{\mathbf{b}^*}$). *For nonempty* $\mathcal{L}_\epsilon$*,*
$\mathbf{b}^* = \mathrm{argmax}_{\mathbf{b}\in\delta\mathcal{L}_\epsilon} \mathrm{V}\left(\mathbf{X}, \mathbf{b}\right)$*, where* $\mathbf{b}^*$ *satisfies* $\mathrm{Y}\left(\mathbf{X}, \epsilon\right) = \mathrm{V}\left(\mathbf{X}, \mathbf{b}^*\right).$

Interestingly, Proposition 2.4 relates percentile optimization to a minimax objective: one can search for a *worst-case* belief within a specified set. Since $\mathrm{V}\left(\mathbf{X}, \mathbf{b}\right)$ is concave in $\mathbf{b}$ (by the convexity results of Sondik (1971) and Smallwood and Sondik (1973)), if $\delta\mathcal{L}_\epsilon$ is easily characterized, we can apply gradient-based optimization to solve the problem rather than evaluating the entire surface which is computationally intractable. Although Theorem 2.2 states that $\mathcal{K}_{\mathbf{b}^*}$ is a singleton when the value function is differentiable at $\mathbf{b}^*$, the differentiability is not always guaranteed. To this end, in the proof of Proposition 2.4 (see Online Appendix A.2), we characterize $\mathbf{p}^*$. We find that the distribution $\mathbf{p}^*$ such that the contour $\{\mathbf{b} \in \mathcal{B} | \mathrm{V}^{\pi^{\mathbf{p}^*}_{\mathcal{K}_{\mathbf{b}^*}}}\left(\mathbf{X}, \mathbf{b}\right) = \mathrm{V}\left(\mathbf{X}, \mathbf{b}^*\right)\}$ is a subgradient hyperplane to $\mathcal{L}_\epsilon$.

[15] For additional discussion and examples of convex floating bodies, see Online Appendix A.1.6.

In general, since non-robust policies are only partially characterized (they converge to E$c\mu$ policies asymptotically), it is important to connect the E$c\mu$ policies to the percentile optimization objective. The following corollary is similar to Proposition 2.4 and shows that there exists a finite randomization of E$c\mu$ policies that are asymptotically optimal to the percentile objective as $\psi \to \infty$.

Corollary 2.1 (**Robust** E$c\mu$ **Optimality**). *If $\mathcal{L}_\epsilon$ is nonempty, then there exists a policy π that is a finite randomization of* E$c\mu$ *policies such that* $\mathrm{Y}^\pi(\mathbf{X},\epsilon) - \mathrm{Y}(\mathbf{X},\epsilon) \leq \mathrm{V}^{\pi^{c\mu}}(\mathbf{X},\hat{\mathbf{b}}) - \mathrm{V}(\mathbf{X},\mathbf{b}^*)$, *where* $\hat{\mathbf{b}} = \arg\max_{\mathbf{b}\in\delta\mathcal{L}_\epsilon} \mathrm{V}^{\pi^{c\mu}}(\mathbf{X},\mathbf{b})$ *and* $\mathbf{b}^*$ *is defined in Theorem 2.2.*

This corollary holds despite the fact that $\mathrm{V}^{\pi^{c\mu}}(\mathbf{X},\mathbf{b})$ is not guaranteed to be concave in $\mathbf{b}$. In fact, if it is concave in $\mathbf{b}$, the randomized policy π can be directly built from non-robust policies. However, if $\mathrm{V}^{\pi^{c\mu}}(\mathbf{X},\mathbf{b})$ is not concave in $\mathbf{b}$, we can still form the appropriate randomized policy satisfying Corollary 2.1 via a randomization of policies that satisfy minimax solutions within the set of E$c\mu$ policies on the boundary of the convex floating body, namely $\min_{\mathbf{b}^1\in\mathcal{B}} \max_{\mathbf{b}^2\in\delta\mathcal{L}_\epsilon} \mathrm{V}^{\pi^{c\mu}_{\mathbf{b}^1}}(\mathbf{X},\mathbf{b}^2)$.

With respect to optimal solutions to the percentile objective, additional results can further confine $\mathcal{K}_{\mathbf{b}^*}$ (of Theorem 2.2) by noting that $\mathbf{b}^*$ must lie near the extreme belief state with worst-case transition parameters. We denote this "worst-case" belief state by $\mathbf{b}_0$, and note that it is composed of components

$$b^0_{i,j} = \begin{cases} 1 & : \text{ if } \mu_{i,j} = \min_{k\in\mathcal{J}_i} \mu_{i,k}, \\ 0 & : \text{ otherwise.} \end{cases} \tag{2.7}$$

It can be shown (see the proof of Proposition 2.5) that for any policy, $\mathbf{b}_0$ is the worst-case (most expensive) belief state for the system. To further characterize $\mathbf{b}^*$, we define the concept of *visibility* (adopted from geometry literature but repurposed for our needs).

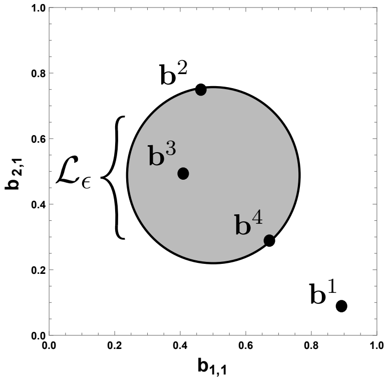


Figure 2.3: Belief points $\mathbf{b}^2$ and $\mathbf{b}^3$ are *not* visible from reference belief $\mathbf{b}^1$, whereas $\mathbf{b}^4$ is visible from reference belief $\mathbf{b}^1$ $(n = 2, m_1, m_2 = 2)$.

DEFINITION 2.2 (**Visibility**). *A belief point* $\mathbf{b} \in \mathcal{L}_\epsilon$ *is said to be visible from a reference belief* $\mathbf{b}^1 \in \mathcal{B}$ *if* $\{\mathbf{b}^2 \in \mathcal{B} : \mathbf{b}^2 = \eta\mathbf{b} + (1-\eta)\mathbf{b}^1, \eta \in [0,1]\} \bigcap \mathcal{L}_\epsilon = \mathbf{b}$.

As demonstrated in Figure 2.3, a belief $\mathbf{b}$ in the convex floating body is visible from a reference belief $\mathbf{b}^1$ if, on the line segment connecting these points, only $\mathbf{b}$ lies within the convex floating body. This implies that if the reference belief point $\mathbf{b}^1$ is distinct from $\mathbf{b}$, and $\mathbf{b}$ is visible from $\mathbf{b}^1$, then $\mathbf{b}$ must lie in the boundary ($\mathbf{b} \in \delta\mathcal{L}_\epsilon$). However, not every point on $\delta\mathcal{L}_\epsilon$ is visible from a reference point $\mathbf{b}^1$. In the following Proposition, we show that the belief $\mathbf{b}^*$ (introduced in Theorem 2.2) must be visible from the worst-case belief state $\mathbf{b}_0$.

PROPOSITION 2.5 (**Visibility of $\mathbf{b}^*$**). *If* $\mathcal{L}_\epsilon$ *is nonempty, then there exists a* $\mathbf{b}^*$ *visible from the worst-case belief* $\mathbf{b}_0$*.*

Proposition 2.5 significantly helps us find $\mathbf{b}^*$ (of Theorem 2.2): we only need to search part of $\delta\mathcal{L}_\epsilon$ which is visible from $\mathbf{b}_0$. Proposition 2.5 also can facilitate establishing effective heuristics which circumvent the calculation of the non-robust problem. For instance, Figure 2.4 demonstrates the implications of Proposition 2.5 for a uniform type $\mathbb{P}_{\mathbf{B}}$: $\mathbf{b}^*$ lies somewhere on the dashed line.

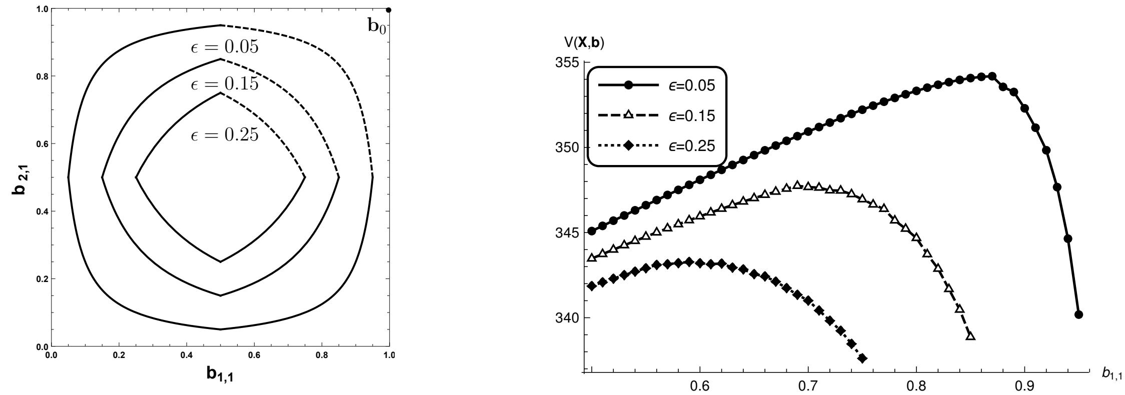


Figure 2.4: On the left, convex floating bodies $\mathcal{L}_\epsilon$ for $\epsilon = 0.05, 0.15, 0.25$ with $n = 2, m_1, m_2 = 2$, and uniform $\mathbb{P}_\mathbf{B}$. To be visible from $\mathbf{b}_0$, belief $\mathbf{b}^*$ associated with Proposition 2.4 must lie on the dashed lines assuming $\mu_{1,1} < \mu_{1,2}$ and $\mu_{2,1} < \mu_{2,2}$ (Proposition 2.5). On the right, $\mathrm{V}((10,10), \mathbf{b})$ is evaluated on these boundaries when $\mu_{1,1} = 0.1, \mu_{1,2} = 0.2, \mu_{2,1} = 0.05, \mu_{2,2} = 0.25$. Belief $\mathbf{b}^*$ lies at the peak of these curves.

2.5 Asymptotically Tight Bounds

Although we have characterized the optimal policies of the non-robust and percentile problems, evaluating the non-robust value function $\mathrm{V}(\mathbf{X}, \mathbf{b})$ is still a computationally complex problem (see, e.g., Littman *et al.* (1998), Mundhenk *et al.* (2000), and Papadimitriou and Tsitsiklis (1987) for an in-depth discussion regarding the complexity of POMDP programs). If the value function $\mathrm{V}(\mathbf{X}, \mathbf{b})$ and the convex floating body's boundary $\delta\mathcal{L}_\epsilon$ are known, the solution to the percentile optimization is easily characterizable (Theorem 2.2, Proposition 2.4, and Proposition 2.5). Therefore, we provide computationally tractable bounds to the non-robust problem that can be evaluated in closed-form to facilitate the computability of chance-constrained policies.

The bounds we form are based on the performance of (a) queues under no model ambiguity with fixed rate parameters equal to $\mathrm{E}[\mu_i|\mathbf{b}]$, and (b) following a particular server allocation priority rule based on the initial parameter belief. These imply that our bounds rely only on the valuation of fixed priority-based policies that do not change with dynamic observations, significantly reducing the computational complexity of the problem.

For a given belief $\hat{\mathbf{b}} \in \mathcal{B}$, consider a counterpart system identical to our original setting with the exception of the ambiguity sets being $\hat{\mathcal{M}}_i = \left\{ \mathrm{E}[\mu_i|\hat{\mathbf{b}}] \right\}$ (analogous to the original ambiguity sets $\mathcal{M}_i$). That is, the counterpart queueing system has fully known service rates that are calculated based on taking an expectation of service rates in $\mathcal{M}_i$ over belief $\hat{\mathbf{b}}$. Obviously, the optimal policy for this system is the traditional $c\mu$ rule, since all of its parameters are fully known. Let $\pi_{\hat{\mathbf{b}}}$ denote this $c\mu$ rule and $\bar{\mathrm{V}}^{\pi_{\hat{\mathbf{b}}}}(\mathbf{X}, \hat{\mathbf{b}})$ be the associated infinite-horizon cost of the counterpart system under $\pi_{\hat{\mathbf{b}}}$. It is important to emphasize that $\pi_{\hat{\mathbf{b}}}$ exhaustively serves class $\arg\max_{a \in \mathcal{A}(\mathbf{X})} c_a \mathrm{E}[\mu_a|\hat{\mathbf{b}}]$ until no customer of that class remains in the system, and acts only as a function of the queue state, not of belief, even when $\pi_{\hat{\mathbf{b}}}$ is implemented in the original system. When $\pi_{\hat{\mathbf{b}}}$ is implemented in the original system, we denote the infinite-horizon cost by $\mathrm{V}^{\pi_{\hat{\mathbf{b}}}}(\mathbf{X}, \hat{\mathbf{b}})$. Using the counterpart system's cost and its associated policy, we can bound the non-robust cost (which is needed to calculate the robust cost; see Theorem 2.2 and Proposition 2.4) using the following proposition.

Proposition 2.6 (**Asymptotically Tight Bounds**). *For any state* $(\mathbf{X}, \hat{\mathbf{b}})$*, the non-robust cost* $\mathrm{V}(\mathbf{X}, \hat{\mathbf{b}})$ *is bounded as* $\bar{\mathrm{V}}^{\pi_{\hat{\mathbf{b}}}}(\mathbf{X}, \hat{\mathbf{b}}) \leq \mathrm{V}(\mathbf{X}, \hat{\mathbf{b}}) \leq \mathrm{V}^{\pi_{\hat{\mathbf{b}}}}(\mathbf{X}, \hat{\mathbf{b}})$. *Furthermore:*

(i) The gap between the upper and lower bound costs decrease to zero as queue length X_i *increases to infinity, where* $i = \arg\max_{a \in \mathcal{A}(\mathbf{X})} c_a \mathrm{E}[\mu_a|\hat{\mathbf{b}}]$.

(ii) The gap between the upper and lower bound costs monotonically decrease to zero as $\mathrm{Var}[\mu_i|\hat{\mathbf{b}}]$ *decrease to zero (for all* $i \in \mathcal{N}$*).*

Both the upper and lower bounds of Proposition 2.6 are easily calculable (see

Online Appendix A.2). Furthermore, under the conditions above, these bounds become arbitrarily close approximations, which adds computational tractability to the problem as well as analytical insight to the relationship between our non-robust and traditional $c\mu$ policies. In particular, part (*ii*) of Proposition 2.6 supports the intuition that gathering more data on unknown service parameters can provide more accurate bound information. Part (*i*) of Proposition 2.6 provides conditions under which the myopic, non-learning policy's cost converges to that of the optimal policy.

Remark 1. Since the percentile objective relies on the computation of the non-robust problem, the bound results can be easily applied to the percentile formulation as well. For instance, one can refine the search for $\mathrm{argmax}_{\mathbf{b}\in\delta\mathcal{L}_\epsilon} \mathrm{V}\left(\mathbf{X},\mathbf{b}\right)$ as in Proposition 2.4: if the upper bound for a $\mathbf{b}\in\delta\mathcal{L}_\epsilon$ is less than the lower bound for $\mathbf{b}'\in\delta\mathcal{L}_\epsilon$, $\mathbf{b}$ must not be the belief point $\mathbf{b}^*$. Since most infinite-horizon POMDPs are calculated by finite-horizon approximations, a second application of the bounds is to use them as the terminal cost used in the finite-horizon dynamic program. That is, when evaluating the finite-horizon approximation, one can replace $\mathrm{V}_0(\mathbf{X},\mathbf{b})$ by lower and upper bounds $\bar{\mathrm{V}}^{\pi_{\hat{\mathbf{b}}}}(\mathbf{X},\hat{\mathbf{b}})$ and $\mathrm{V}^{\pi_{\mathbf{b}}}\left(\mathbf{X},\mathbf{b}\right)$, respectively. This can provide very tight bounds on the POMDP, since after a certain number of "learning periods," where the POMDP is explicitly evaluated, the controller might have collected enough information to have enough confidence in the true transition parameters.

2.6 An Analytically-Rooted Heuristic Policy

Chance-constrained policies are inherently difficult to calculate, even given the analytical results established in the previous section. To circumvent complexity arising from (a) the PSPACE-hard problem of evaluating a POMDP over a belief space with high dimensionality, and (b) finding the shape of the convex floating body which requires high-dimensional polytope approximations, we now introduce an effective heuristic policy. This heuristic policy operates by simply choosing the E$c\mu$ policy associated with the belief point on the convex floating body's boundary $\delta\mathcal{L}_\epsilon$ that minimizes the distance from $\mathbf{b}_0$ (the worst-case parameter settings for each class characterized in (2.7)). This is typically an easy-to-perform task, especially in the cases of uniform and spherical type distributions on the belief space, allowing for managers to benefit from our approach without requiring demanding computations. Moreover, as we will show in Section 2.7, this heuristic performs extremely well both on randomly generated data and on real-world data that we have collected from a leading U.S. hospital.

We term the E$c\mu$ policy with expectation taken based on belief point $\arg\min_{\mathbf{b}\in\delta\mathcal{L}_\epsilon} \|\mathbf{b}_0 - \mathbf{b}\|$, where $\|\cdot\|$ is the l^2-norm, as the $(1-\epsilon)\%$ E$c\mu$ *heuristic policy.* This heuristic policy takes advantage of three main structural results of the chance-constrained policy (that we established in the previous section), while providing a much simpler version of it:

(1) It assumes that the true optimal policies of the non-robust problem are E$c\mu$, a fact supported by Theorem 2.1 which shows the asymptotic relationship of the optimal policies to E$c\mu$.

(2) It locates belief $\arg\min_{\mathbf{b}\in\delta\mathcal{L}_\epsilon} \|\mathbf{b}_0 - \mathbf{b}\|$ to be near $\mathbf{b}^*$ (of Theorem 2.2) based on Proposition 2.4. The worst-case (most expensive) belief state is $\mathbf{b}_0$, and through the proof of Proposition 2.5 (see Online Appendix A.2) the value function is non-increasing in λ with respect to belief $\lambda\mathbf{b} + (1-\lambda)\mathbf{b}_0$ for $\lambda \in [0,1]$. Thus, $\arg\max_{\mathbf{b}\in\delta\mathcal{L}_\epsilon} \mathrm{V}(\mathbf{X},\mathbf{b})$ is expected to be near $\mathbf{b}_0$. [16]

(3) It takes advantage of the fact that $\arg\min_{\mathbf{b}\in\delta\mathcal{L}_\epsilon} \|\mathbf{b}_0 - \mathbf{b}\|$ satisfies Proposition 2.5 (since this belief is visible from $\mathbf{b}_0$).

[16] This does not imply that $\arg\max_{\mathbf{b}\in\delta\mathcal{L}_\epsilon} \mathrm{V}(\mathbf{X},\mathbf{b}) = \arg\min_{\mathbf{b}\in\delta\mathcal{L}_\epsilon} \|\mathbf{b}_0 - \mathbf{b}\|$. $\mathrm{V}(\mathbf{X},\mathbf{b})$ is only assured to be non-increasing on line segments connected to $\mathbf{b}_0$.

2.7 Numerical Experiments

We now perform various numerical experiments in order to (a) identify the advantages of chance-constrained policies in a variety of environments under model ambiguity, (b) demonstrate the sensitivities of the underlying queueing models, (c) study the effectiveness of the proposed E$c\mu$ heuristic in mimicking the optimal chance-constrained policies, and (d) demonstrate the implications of our results in real-world applications. To pursue these goals, we present our analyses in five parts: we (a) establish the sensitivities in initial prior selection, (b) investigate how our policies perform over a large parameter suite but in a relatively small queueing system, (c) evaluate our proposed heuristic alongside percentile, minimax, and minimin policies in a larger system, (d) demonstrate the gap between the E$c\mu$ and optimal (non-robust) policies, and (e) apply the E$c\mu$ heuristic to a hospital Emergency Department (ED) setting using real-world data, and discuss its significant implications on improving the current patient flow policies.

To help establish the necessity of our robust percentile formulation, it is first important to establish the sensitivities of the non-robust value function under small perturbations in belief. To this end, we evaluate the expected cost under a variety of parameter settings when $n = 2, m_1 = 2$, and $m_2 = 2$ with respect to a "central prior" $\bar{\mathbf{b}} = (0.5, 0.5, 0.5, 0.5)$, that assumes a uniform distribution on parameters, a slightly pessimistic $\bar{\mathbf{b}}^p = (0.6, 0.4, 0.6, 0.4)$, and a slightly optimistic prior $\bar{\mathbf{b}}^o = (0.4, 0.6, 0.4, 0.6)$. Table 2.1 displays the results from comparing the percentage difference between non-robust value functions evaluated at these priors (for various parameter configurations) via the expression

$$\frac{|\mathrm{V}(\mathbf{X}, \mathbf{b}) - \mathrm{V}(\mathbf{X}, \hat{\mathbf{b}})|}{\left(\mathrm{V}(\mathbf{X}, \mathbf{b}) + \mathrm{V}(\mathbf{X}, \hat{\mathbf{b}})\right)/2}\%$$

Percentage Differences							Percentage Differences						
$\mathbf{X}$	$(\mu_{1,1},\mu_{1,2})$	$(\mu_{2,1},\mu_{2,2})$	(c_1,c_2)	$\bar{\mathbf{b}}$ vs $\bar{\mathbf{b}}^p$	$\bar{\mathbf{b}}$ vs $\bar{\mathbf{b}}^o$	$\bar{\mathbf{b}}^p$ vs $\bar{\mathbf{b}}^o$	$\mathbf{X}$	$(\mu_{1,1},\mu_{1,2})$	$(\mu_{2,1},\mu_{2,2})$	(c_1,c_2)	$\bar{\mathbf{b}}$ vs $\bar{\mathbf{b}}^p$	$\bar{\mathbf{b}}$ vs $\bar{\mathbf{b}}^o$	$\bar{\mathbf{b}}^p$ vs $\bar{\mathbf{b}}^o$
$(5,5)$	$(0.1,0.2)$	$(0.15,0.3)$	$(0.1,0.1)$	5.24	5.52	10.76	$(5,5)$	$(0.05,0.15)$	$(0.1,0.2)$	$(0.15,0.2)$	5.31	5.8	11.1
$(10,10)$				4.25	4.98	9.22	$(10,10)$				4.15	4.47	8.62
$(15,15)$				3.62	4.29	7.9	$(15,15)$				3.39	3.37	6.75
$(5,5)$	$(0.05,0.15)$	$(0.1,0.2)$	$(0.1,0.1)$	5.21	6.1	11.3	$(5,5)$	$(0.1,0.2)$	$(0.15,0.3)$	$(0.2,0.15)$	5.33	5.8	11.12
$(10,10)$				4.11	4.63	8.74	$(10,10)$				4.7	4.77	9.46
$(15,15)$				3.19	3.7	6.89	$(15,15)$				4.02	3.89	7.91
$(5,5)$	$(0.05,0.1)$	$(0.06,0.08)$	$(0.1,0.1)$	3.03	3.23	6.26	$(5,5)$	$(0.05,0.15)$	$(0.1,0.2)$	$(0.2,0.15)$	6.32	6.0	12.32
$(10,10)$				2.24	2.23	4.47	$(10,10)$				4.75	4.6	9.35
$(15,15)$				1.58	1.63	3.21	$(15,15)$				3.64	3.81	7.45
$(5,5)$	$(0.1,0.2)$	$(0.1,0.2)$	$(0.15,0.3)$	5.46	5.27	10.72	$(5,5)$	$(0.05,0.1)$	$(0.06,0.08)$	$(0.2,0.15)$	3.44	3.95	7.39
$(10,10)$				4.39	4.89	9.27	$(10,10)$				2.66	2.63	5.3
$(15,15)$				3.97	4.05	8.02	$(15,15)$				2.04	1.96	4.0
Average %											**3.72**	**3.93**	**7.64**

Table 2.1: Percentage gaps for $\bar{\mathbf{b}} = (0.5, 0.5, 0.5, 0.5)$, $\bar{\mathbf{b}}^p = (0.6, 0.4, 0.6, 0.4)$, and $\bar{\mathbf{b}}^o = (0.4, 0.6, 0.4, 0.6)$ where $n = 2, m_1 = 2, m_2 = 2$.

for two distinct priors $\mathbf{b}, \hat{\mathbf{b}} \in \mathcal{B}$.

Even with relatively small perturbations to the selection of the prior, as can be seen from Table 2.1, differences in value function are substantial. Thus, we make the following:

Observation 2.1 (**Sensitivity to Prior Specification**). *The expected cost of the non-robust problem is sensitive to the choice of prior.*

Can slight perturbations in the prior also cause significant differences in policies obtained from the non-robust framework? To answer this, we again turn to our $n = 2, m_1 = 2, m_2 = 2$ environment and investigate the differences non-robust policies E$c\mu$ experience as their prior changes from $\bar{\mathbf{b}}, \bar{\mathbf{b}}^p$, and $\bar{\mathbf{b}}^o$. We run simulations in which the true parameter settings are selected according to $\bar{\mathbf{b}}$. To identify differences between the policies at different initial priors, we track the cumulative number of attempts to serve class 1 by time t under each policy, and depict the results in

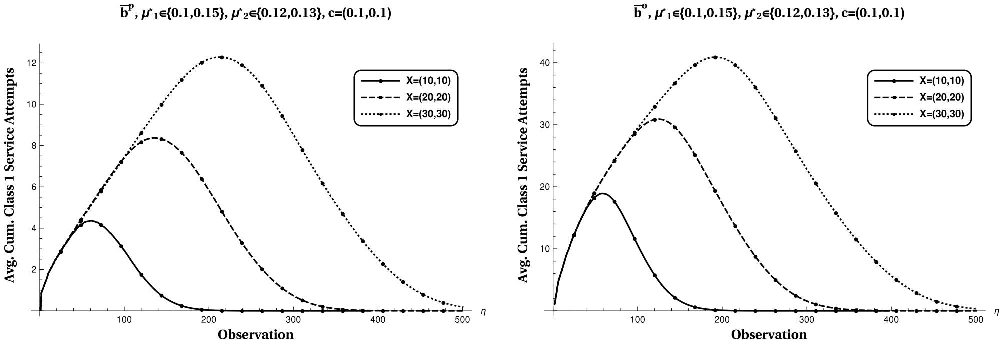


Figure 2.5: Comparison of two non-robust policies under slight perturbations of the initial prior $\bar{\mathbf{b}}$ ($\mu_{1,1} = 0.1, \mu_{1,2} = 0.15, \mu_{2,1} = 0.12, \mu_{2,2} = 0.13$).

Figure 2.5.

Figure 2.5 shows that policies experience an extended period of time in which they disagree on the class to serve. This is especially evident when a large number of customers are in the system, indicating that policies only begin to converge after having finished the service of the class. Furthermore, as discussed earlier, the E$c\mu$ policy experiences momentum toward serving a customer after a successful service. Therefore, as can be seen from Figure 2.5, two policies with slightly different starting beliefs (i.e., initial priors) may experience very different action profiles.

Thus, not only does the value function experience sensitivity among different selections of the prior, but these differences also correspond to policy changes. Thus, we make the following:

Observation 2.2 (**Sensitivity in Policy**). *The policies generated from the non-robust problem are sensitive to the choice of prior.*

If the duration of time where the optimal policy experiences learning is relatively small, the choice of the initial prior becomes inconsequential, since the difference between initial priors will be quickly "washed out" by the incoming data. To test

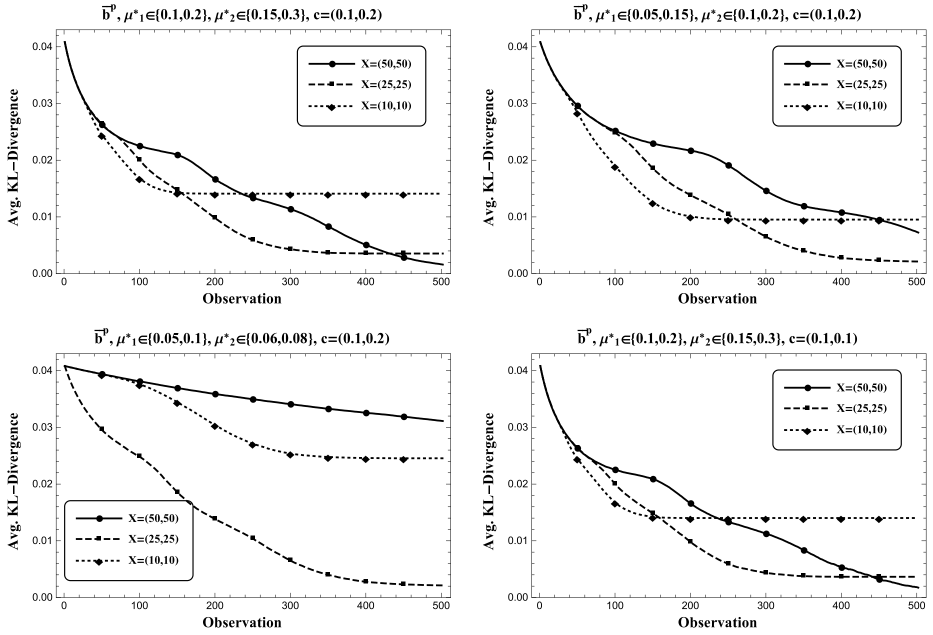


Figure 2.6: Comparison of the average KL-divergence between two policies' beliefs when the true prior is $\bar{\mathbf{b}}$.

whether or not the differences between initial priors is long-lasting, in Figure 2.6 we compare the KL-divergence [17] between two beliefs after each observation under their associated policies.

From Figure 2.6, it is evident that the beliefs converge to one another given that there are enough customers to serve. In general the learning is faster for smaller queue states until all of the customers have been served since the policies are in effect closer to one another. However, even in the smaller queue settings, the learning rate is not fast enough to disregard the choice of the initial prior. Thus, we make the following:

Observation 2.3 (**Slow Convergence in Belief**). *The differences between the beliefs about the correct model with differing initial priors is long-lasting.*

To better understand the relative performance of our robust percentile policies, we

[17] For two belief points, $\mathbf{b}, \hat{\mathbf{b}} \in \mathcal{B}$ with all positive components in the setting where $n = 2, m_1 = 2$, and $m_2 = 2$, the KL-divergence is $D_{KL}(\mathbf{b}||\hat{\mathbf{b}}) = \sum_{i=1}^{2} \sum_{j=1}^{2} b_{1,i}b_{2,j} \log \frac{b_{1,i}b_{2,j}}{\hat{b}_{1,i}\hat{b}_{2,j}}$.

start by considering a large parameter suite including over $1,000$ parameter settings in an $n = 2, m_1 = 2$, and $m_2 = 2$ setting with four different $\mathbb{P}_{\mathbf{B}}$ distributions at their 95% chance-constrained policy. We name these $\mathbb{P}_{\mathbf{B}}$ distributions f_1, f_2, f_3 and f_4 respectively: f_1, f_2, and f_3 are truncated multivariate normal distributions with means $\boldsymbol{\mu}_1 = (0.5, 0.5)$, $\boldsymbol{\mu}_2 = (0.4, 0.4)$, $\boldsymbol{\mu}_3 = (0.6, 0.6)$ and covariance matrices $\boldsymbol{\Sigma}_1 = \left(\begin{smallmatrix} 1.5 & 0.0 \\ 0.0 & 1.5 \end{smallmatrix}\right)$, $\boldsymbol{\Sigma}_2 = \left(\begin{smallmatrix} 0.5 & 0.0 \\ 0.0 & 0.5 \end{smallmatrix}\right)$, and $\boldsymbol{\Sigma}_3 = \left(\begin{smallmatrix} 0.5 & 0.0 \\ 0.0 & 0.5 \end{smallmatrix}\right)$ respectively. Finally, f_4 is the uniform distribution.

We include two non-learning robust policies (minimin and minimax) as benchmarks for the performance of our robust percentile policies and compare the policies by evaluating their total cost when each model (i.e., parameter configuration) is equally likely. That is, we assume that the true (but unknown) prior of our system is $\bar{\mathbf{b}} = (0.5, 0.5, 0.5, 0.5)$, and we evaluate the total cost under 95% chance-constrained, minimax, and minimin policies. Furthermore, we assume $c_1 = c_2$. In every problem instance, we assume $\mu_{2,1} < \mu_{1,1}$ and $\mu_{1,2} < \mu_{2,2}$ so that the policy is not uniform throughout the belief space, which provides incentive for gaining additional knowledge. Further detail on this parameter suite is presented in Online Appendix A.1.1.

We next compare our proposed policies with other non-learning robust policies (minimax and minimin). In Table 2.2, we present the results of this comparison expressed by the average (among all models) optimality gap percentage under various policies. The optimality gap percentage for policy π at $\mathbf{b}$ is defined as

$$\frac{\mathrm{V}^{\pi}(\mathbf{X}, \mathbf{b}) - \mathrm{V}(\mathbf{X}, \mathbf{b})}{\mathrm{V}(\mathbf{X}, \mathbf{b})}\%.$$

From Table 2.2, we observe that on average, our proposed chance-constrained policies perform much better than the other non-learning policies. Since there is equal chance of every parameter configuration, non-learning policies serve the wrong class for a realized set of parameters 50% of the time, which results in poor performance.

Optimality Gap (%)						
X	Minimax	Minimin	95% Chance Constrained f_1	95% Chance Constrained f_2	95% Chance Constrained f_3	95% Chance Constrained f_4
$(2, 2)$	3.17	15.51	1.84	2.07	2.2	1.97
$(2, 5)$	2.52	13.58	0.85	0.81	0.86	0.86
$(2, 10)$	1.35	8.21	0.52	0.51	0.54	0.49
$(5, 2)$	4.48	8.73	2.65	2.74	2.33	2.21
$(5, 5)$	5.01	10.3	0.85	0.81	0.75	0.79
$(5, 10)$	3.37	7.56	0.61	0.58	0.48	0.57
$(10, 2)$	4.14	4.05	1.76	1.93	1.32	1.35
$(10, 5)$	5.49	5.79	0.53	0.51	0.54	0.55
$(10, 10)$	4.34	5.15	0.33	0.33	0.35	0.35
Ave.	3.76	8.76	1.10	1.14	1.04	1.02

Table 2.2: Performance of various robust policies over the test suite ($n = 2, m_1 = 2, m_2 = 2$).

Comparing the chance-constraint policies under f_1, f_2, f_3, and f_4 in Table 2.2 reveals yet another interesting insight: they exhibit similar performance. The reason behind this is three-fold: (a) as a property of Proposition 2.5, since we used 95% chance-constrained policies, each $\mathbf{b}^*$ tends to be near $\mathbf{b}_0$, (b) even though the distributions f_1, f_2, f_3, and f_4 are different (e.g., they have differing covariance structures and are centered at different beliefs), their convex floating bodies are quite similar, and (c) the chance-constrained policies we propose exhibit learning. Hence, we can make the following:

Observation 2.4 (**Sensitivity**). *The performance of chance-constrained policies is not sensitive to the choice of* $\mathbb{P}_{\mathbf{B}}$.

In Section 2.6, we introduced the E$c\mu$ heuristic as an easy-to-implement policy that mimics the performance of robust optimal chance-constrained policies. To

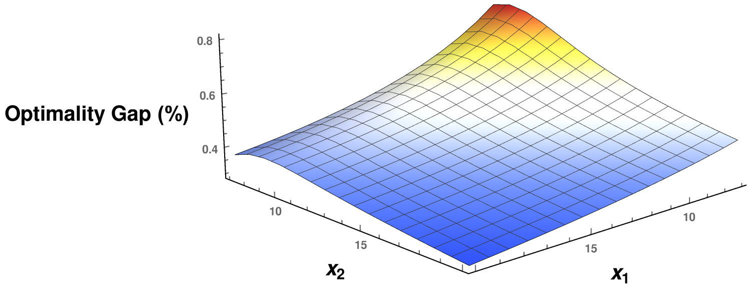


Figure 2.7: The optimality gap (%) of E$c\mu$ policy when evaluated on the central prior $\bar{\mathbf{b}}$ ($\mu_{1,1} = 0.6, \mu_{1,2} = 0.7, \mu_{2,1} = 0.5, \mu_{2,2} = 0.8$).

demonstrate the validity of the first assumption underlying this heuristic – that the optimal policies of the non-robust problem are E$c\mu$ – in Figure 2.7 we depict the percent optimality gap of the E$c\mu$ heuristic policy by comparing its cost to that of the optimal non-robust policies in a situation where ψ is small. Since we know that E$c\mu$ becomes optimal as ψ becomes large (Theorem 2.1), this poses a "worst-case" scenario for the performance of the E$c\mu$ policies. From Figure 2.7, we can make the following:

Observation 2.5 (**Near Optimality of** E$c\mu$). *Even when ψ is small, the* E$c\mu$ *performance is close to the non-robust optimal policy, especially when the system is highly congested.*

Observation 2.5 confirms that the myopic E$c\mu$ policy provides us with a good approximation of the optimal POMDP value function (as we would expect given its asymptotic relationship to the chance-constrained policy; see Theorem 2.1). However, using such a rule to find the explicit surface of the POMDP value function is computationally challenging, even though the E$c\mu$ policy is simple. This is because policy evaluation (even when a policy is known) in POMDPs is PSPACE complete (see, e.g., Mundhenk *et al.* (2000)). Hence, the ideal task of searching for the max of the convex floating body as in Proposition 2.4, even with the help of Proposition 2.5, is highly difficult even in moderate problem instances where $n > 3$ and $m > 6$. Furthermore, often times the shape of $\mathcal{L}_\epsilon$ is difficult to determine explicitly as is the case even in the simple uniform distributions in more than two dimensions, which further complicates our search. Hence, for implementation in real applications, we turn to our robust heuristic policy.

To gain deeper insights into the performance of our heuristic, we simulate systems with $m_1 = m_2 = m_3 = 3$ with uniform $\mathbb{P}_{\mathbf{B}}$ in the largest inscribed sphere of the belief space. To also evaluate the robustness of our proposed heuristic vis-a-vis the optimal percentile policy as well as minimin and minimax policies, we use CVar(q), which is the average cost within the most costly q% of our simulated runs. Therefore, if $\mathcal{S} = \{s_1, \ldots, s_r\}$ is the set of the costs from a simulation of r runs ordered from most costly to least costly, then

$$\text{CVar}\,(q) = \frac{\sum_{i=1}^{\lceil (1-q)(r-1)+1 \rceil} s_i}{\lceil (1-q)(r-1)+1 \rceil}.$$

This statistic may roughly be seen as a function that increases in pessimism, since we use fewer low cost data points in the expectation as q increases. [18]

Using a 95% chance-constrained policy, the E$c\mu$ heuristic, minimin, and minimax policies, Figure 2.8 illustrates performance over $20,000$ simulation runs. [19] The leftmost subfigures display the raw CVar values. However, we direct our attention to the rightmost figures, which display the percentage gap (of CVars) between the four selected polices and "best" policy at a given q. From Figure 2.8, we observe the following:

[18] For instance, one would expect the minimax policy to perform well in comparison to other policies at CVar(1).

[19] The associated confidence intervals are tight, so we only show the averages.

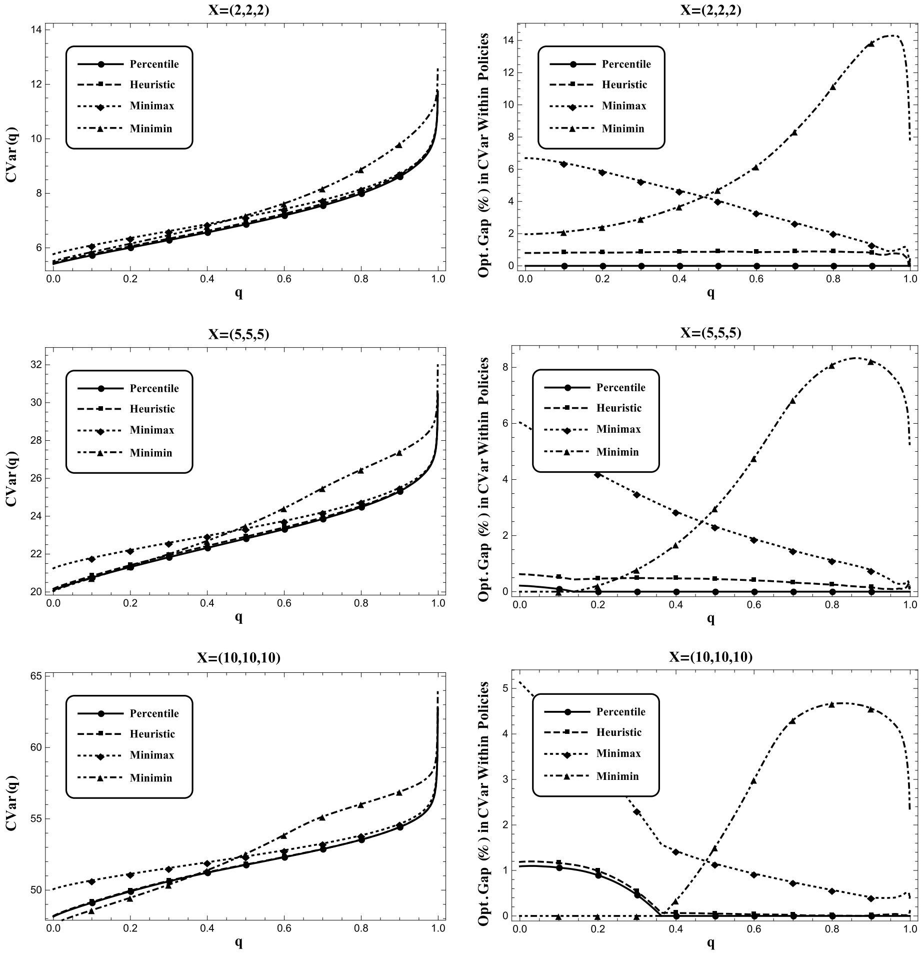


Figure 2.8: Comparison of policies with respect to CVar ($20,000$ simulated runs and a uniform $\mathbb{P}_{\mathbf{B}}$ on the largest inscribed sphere of the belief space).

Observation 2.6 (**Heuristic Performance**). *The* E$c\mu$ *heuristic performs nearly identically to the chance-constrained policy, with a diminishing difference as the system becomes more congested.*

We note that percentile optimization is not concerned about the "worst-case" scenarios, and rather optimizes based on a proportion of the belief space. Hence, being a statistic concerned with the tail performance of the distribution of costs, CVar (as compared to the expected cost) provides us with a more accurate representation of the value of robustness that percentile optimization offers. Furthermore, Figure 2.8 demonstrates that the proposed heuristic captures the essence of the chance-constrained policy in that it lies near the optimal policy, mirroring its performance in each simulated run. Overall, our goal to provide an alternative to the over-conservatism and over-optimism of the minimax and minimin policies seems to be met by our percentile optimization technique, which is consistent with established robust optimization literature (see, e.g., Bertsimas and Sim (2004b)). Moreover, though our policies are generated from a fixed pessimism level (i.e., 95% chance-constrained), they perform well throughout the spectrum of optimism/pessimism in the CVar statistic.

Even in cases where the chance-constrained policy is inferior to other policies with regard to the CVar statistic (e.g., the fourth row of Figure 2.8 with $\mathbf{X} = (10, 10, 10)$, where the minimin policy is seen to perform best with regard to CVar(0)), we can see that fixed priority policies (e.g., those obtained under the minimin objective) miss out on the advantages of robustness that the chance-constrained policy offers throughout the optimism spectrum. Furthermore, percentile optimization is flexible: by modifying ϵ, we can change our policy's focus to be more or less optimistic to the point of becoming a minimax and minimin policy itself (Proposition 2.2). A similar advantage is also gained in the APOMDP framework of Saghafian (2018), where α-maximin expected utility (α-MEU) preferences are used.

2.7.1 Real-World Application: ED Patient Prioritization

In most hospital Emergency Departments (EDs) in the U.S., patients upon arrival are sorted by means of an urgency-based triage system into one of (typically) five classes known as Emergency Severity Index (ESI) levels. These ESI levels classify patients in descending order of urgency so that a patient of ESI 1, being in dire condition, is immediately treated, whereas patients of levels 4 and 5 are sent to a "fast track" area to be treated. Therefore, the classes served by the main section of the ED (the majority of arrivals) are those with ESI levels 2 and 3 (see, e.g., Saghafian *et al.* (2012), Saghafian *et al.* (2014), and the references therein). We denote ESI 2 and 3 patients by "Urgent" and "Non-Urgent" patients, respectively.

As patients wait to receive treatment their condition may worsen over time and lead to adverse medical events. Sprivulis *et al.* (2006) and Plunkett *et al.* (2011) show that higher patient mortality is associated with longer waiting times prior to seeing a physician. Other research (e.g., an extremely large study on data of nearly 14 million patients by Guttmann *et al.* (2011)) indicates that the Risk of Adverse Events (ROAE) for patients increases with higher waiting times leading to higher mortality and hospital admission rates. Therefore, with the objective of increasing patient safety, we consider the goal of minimizing average ROAE for ED patients, and investigate optimal prioritization policies. To do so, we assume adverse events occur based on a Poisson process with a higher rate for urgent patients, and note that ROAEs in this setting play the role of holding cost parameters in our multi-class queueing model introduced earlier. The same approach is used in Saghafian *et al.* (2014), where the benefits of further stratifying these levels in terms of a patient's *complexity* is discussed. Simple patients are those that experience only a single interaction with the physician, and thus are more quickly treated by the ED than complex patients, whose treatment necessitates several interactions with the physician interspersed with various tests (CT scans, MRI, etc.).

Figure 2.9 (left) illustrates a schematic flow of patients as a multi-class queueing system. To analyze the multi-class queueing system of Figure 2.9 (right) in a traditional way, one needs to obtain point estimates of various parameters (e.g., service/treatment rates for each class), a task which is subject to inevitable errors. [20]

[20] Even after using a large data set that we have collected from a leading U.S. hospital, which

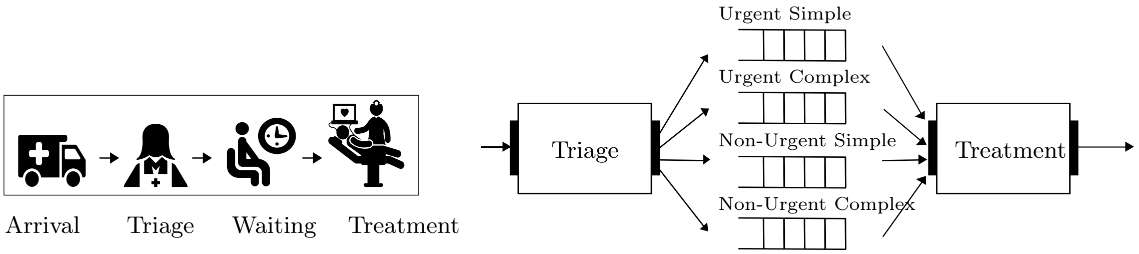


Figure 2.9: Patient flow in hospital Emergency Departments (left: the overall flow; right: the multi-class flow).

Furthermore, triaged urgency and complexity levels are subject to misclassifications, which further confuses the true parameter settings of the system. Although misclassifications can be included in the analysis when all of the parameters of the system are known, the misclassification probabilities themselves are also hard to quantify. These create parameter ambiguity, and one needs to use robust analyses to hedge against them. However, current ED patient prioritization policies are based on analyses that ignore such ambiguities.

To demonstrate the benefits of our percentile optimization approach, we now focus on two questions: how should EDs prioritize their patients given that they are faced with parameter ambiguity? and how much benefit can they get by taking ambiguities into consideration? To answer these questions, we first model the ED from a broad perspective with non-stationary Poisson process arrivals and known service rates for all four classes: Urgent Simple (US), Urgent Complex (UC), Non-Urgent Simple (NS), and Non-Urgent Complex (NC) patients. In this way, we model the ED as a single "super-server" (i.e., with a pooled capacity that we estimate from our data set so as to match the input-output process of the ED as a whole). This allows us to gain

includes data about more than 18,000 patient visits, we see that our point estimates are not reliable due to various reasons including the large variation among patient characteristics as well as the need to estimate parameters for each patient class separately.

insights into the questions we raised above by noting that the ED queueing model of Figure 2.9 (right) is essentially a special case of our general model depicted in Figure 2.1 with $n = 4$.

Patient arrivals in an ED fluctuate throughout a given day, so we model these arrivals with a non-stationary Poisson process with hourly rates shown in Figure A.6 in Online Appendix A.1 which depicts the actual time-dependent arrival rates to the ED based on our data set. Furthermore, since patient LOS in our data has a lognormal distribution, we fit lognormal service distributions to match the LOS of patients for each class of patients. Next, we design our "cloud of models" by perturbing the fitted rate parameters such that for each class i with fitted rate $\hat{\mu}_{i,3}$, we incorporate four additional possible rate parameters so $\hat{\mu}_{i,1} < \hat{\mu}_{i,2} < \hat{\mu}_{i,3} < \hat{\mu}_{i,4} < \hat{\mu}_{i,5}$. Because patients become fairly stable upon seeing a physician, we focus on adverse events in the waiting area of EDs, and assume ROAE drops to zero once the treatment stage begins. Our model is non-preemptive, which is a reflection of physicians' behavior in EDs: upon initiating treatment to a patient, they rarely pause treatment to serve a different patient. Since there is a possibility that the ROAE for simple patients differs from that of complex patients, we also consider a variety of such "cost" structures in our study.

Though this model allows for dynamic arrivals (unlike our model introduced in Section 2.3), we can still incorporate chance-constrained policies through the use of our heuristic, and compare its performance to the complexity-based prioritization policy that serves classes US, UC, NS, and NC in descending priority (demonstrated to be optimal for EDs in Saghafian *et al.* (2014) when ambiguity is ignored), minimax, and minimin policies. To do so, we simply modify the Bayesian belief to also incorporate arrival data. We simulate these policies, and track the non-discounted ROAE by assuming that $\mathbb{P}_{\mathbf{B}}$ is uniform. The result of 20,000 simulated days expressed in terms of the CVar statistic is reported in Figure 2.10 (see Online Appendix A.1.7 for four additional ROAE settings and in-depth discussions).

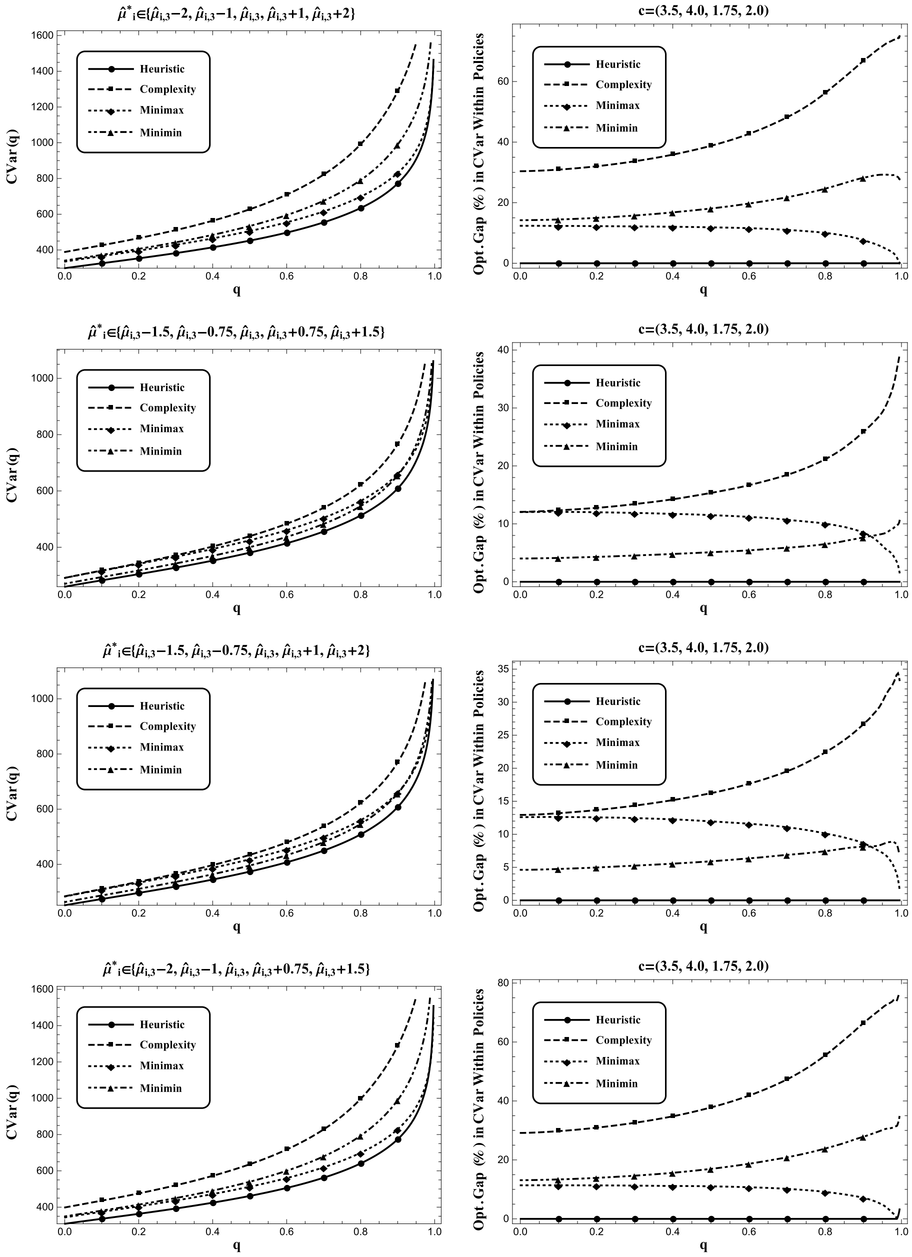


Figure 2.10: $20,000$ simulated days in the ED for the complexity-based prioritization, 95% E$c\mu$ heuristic, minimin, and minimax policies, when $\mathbb{P}_{\mathbf{B}}$ is uniform, and the cloud of models perturbs the fitted service rate $\hat{\mu}_{i,3}$ in terms of two-hour time increments with $\mathbf{c} = (3.5, 4.0, 1.75, 2.0)$. (Triage levels US, UC, NS, and NC are denoted 1,2,3, and 4, respectively.)

A widely discussed topic in the literature surrounding EDs is the "overcrowding" issue (see e.g. Derlet and Richards (2000), Derlet *et al.* (2001), and Trzeciak and Rivers (2003)) that stems from high arrival rates and limited resources (such as capacity, physicians, equipment, etc). Overcrowding in EDs results in high ROAE that endangers patients. The third row of Figure 2.10 demonstrates how policies perform in overcrowded EDs by considering an ambiguity set with smaller service rates (in comparison to the other ambiguity sets). We note that percentile optimization, in comparison with other policies, is especially suited for studying patient prioritization in overcrowded EDs. This is because under heavy congestion, chance-constrained policies learn faster, since more classes are available to serve at any given time. Furthermore, as we show in Corollary A.3 in Online Appendix A.2, the E$c\mu$ policy becomes asymptotically optimal when arrivals occur during intense bursts followed by lull periods. Since hospital EDs typically experience long periods of heavy traffic in the afternoon followed by little traffic after midnight (see the actual arrival pattern depicted in Figure A.6 in Online Appendix A.1), this further establishes our approach in hospital ED applications. Using these results, we can make the following:

Observation 2.7 (**High Traffic**). *Our percentile optimization approach performs well for prioritizing patients in EDs, especially in highly congested ones (e.g. those in busy research hospitals).*

Also, Figure 2.10 shows that, once again, the chance-constrained policies nearly dominate the entire spectrum of the CVar statistic since they explicitly incorporate both learning and robustness. Hence, even though our stylized environment is less detailed than those ED flow models in studies such as Huang *et al.* (2015), Saghafian *et al.* (2012), Saghafian *et al.* (2014), Saghafian *et al.* (2015), and the references there in (which feature patient feedback), these experiments indicate a performance advantage over complexity-based prioritization, which suggests implementation regardless of optimism/pessimism levels. Hence, to establish the potential benefits percentile optimization can offer to EDs over the current status quo, we make the following:

OBSERVATION 2.8 (**Improved System Performance**). *Percentile optimization can improve the performance of EDs regardless of a manager's disposition.*

In systems with high traffic, learning may occur at an advanced rate, since it has available customers from each class a majority of the time the system is online. Hence, while static priority policies continue to serve the "wrong" classes (due to the underlying parameter ambiguity), the chance-constrained policy quickly identifies the optimal $c\mu$ priority using the observed values. This enhances the quality the robustness percentile optimization offers, especially since one is typically more concerned with overcrowded/busy systems (EDs with low traffic have short patient LOS naturally, and are not in significant need for optimization).

Furthermore, our "clearing" system is a model often used to study queues undergoing overcrowded situations. Therefore, a more congested ED is a better fit to our original model, and in considering dynamic arrivals, we can reconfirm all the previous insights generated in the "clearing" environment. This further confirms the results of Section A.1.4 (within Online Appendix A.1), where we show that most of the main insights gained from the "clearing" system holds for systems with dynamic arrivals.

Finally, we note that in communities with unstable patient population characteristics, where ED service rates or misclassification probabilities are more ambiguous, ED managers can incorporate percentile optimization to effectively hedge against such ambiguities. Moreover, percentile optimization is well-suited to high levels of ambiguity. In our simulations, this is captured through modifying our cloud of models to incorporate larger differences in the fitted parameters (see the first row of Figure 2.10

and compare it with the second row). Hence, when patient population characteristics are unstable, percentile optimization stands out as a method that protects from negative consequences of focusing only on extreme outcomes, while simultaneously learning from incoming data. This results in the following:

OBSERVATION 2.9 (**Uncertain Population Characteristics**). *Percentile optimization can significantly help EDs that are placed in geographical areas with unstable or unknown patient population characteristics to better prioritize their patients.*

2.8 Conclusion

Multi-class queues are versatile structures widely used in operations management that see a large variety of applications in both service and manufacturing sectors. In such environments, often exact parameter specification is rife with estimation errors that (if ignored) can cause system managers to implement wrong policies. We identify and implement a novel data-driven percentile optimization framework for use in POMDPs. Our method layers chance-constrained optimization on a non-robust learning model, effectively enabling learning of the true system state parameters, and allowing the manager to set an optimism level indicating the extent of protection against poor parameter scenarios s/he desires. We characterize the optimal policies to both the non-robust and percentile problems and find that chance-constrained policies can be established via the non-robust problem.

Since percentile optimization problems are typically computationally difficult, we introduce an analytically-rooted heuristic that can be used to effectively incorporate robustness in managing large and complex service or manufacturing systems. To further improve computational tractability, we find asymptotically tight bounds to the non-robust problem, which can be used to efficiently solve the percentile optimization problem.

Finally, we demonstrate the efficacy of our methods numerically in both stylized and realistic environments. Using real-world data collected from a leading hospital, we observe that our approach provides promising results in improving current patient flow policies, especially for overcrowded EDs, or those facing unknown patient population characteristics. Since ED managers typically do not fully know the service rate parameters, traditional patient flow policies based on queueing models that assume full service rate knowledge subject patients to higher risk than chance-constrained policies. Our work is the first to take into account the inevitable ambiguities in ED operations, and sheds light on the dire consequences of ignoring such ambiguities.

Chapter 3

OUTSIDE OF COLD CHAIN: IMPROVING VACCINE DELIVERY IN DEVELOPING COUNTRIES VIA ROBUST FORECASTING AND INVENTORY MANAGEMENT

3.1 Introduction

With the goal of reducing child mortality, the Expanded Program on Immunization (EPI) developed by the World Health Organization (WHO) currently outlines routine immunization guidelines with the purpose of increasing coverage. In developing countries, vaccines are sent from a centrally located depot to regional and district depots until they reach integrated health centers (IHCs) where they are administered via fixed and outreach sessions. However, to ensure potency, vaccines are recommended to lie in a temperature-controlled environment throughout the supply chain, which invokes the title "cold chain." Hence, to provide vaccination coverage, IHC workers with access to refrigeration retrieve vaccines in regular intervals, whereas other locations with limited or no access to refrigeration, must engage in regular outreach or mobile immunization sessions (see, e.g. Haidari *et al.* (2013)). Since the cold chain necessitates refrigeration, cold-packs, and other temperature-regulating devices, many developing countries struggle to maintain inventory flow, especially in areas with poor transportation and inadequate storage infrastructure. These complications are further exacerbated when there is little or unreliable data to forecast vaccine demands; such uncertainty induces poor inventory policies, which can lead to excessive waste or a loss in coverage. This "last mile" of the vaccine supply chain, which includes the transportation to and management of inventory for IHCs, is the source of many problems in the cold chain and is the focus of the chapter.

According to the WHO's strategic immunization plan, some of the major challenges toward effectively distributing vaccinations include limited cold chain capacities, service delivery points, planning/leadership deficiencies, poor data management, and unreliable forecasting techniques (World Health Organization (2014)). The first of these challenges is documented in numerous studies. For example, in a district of Cameroon, Akoh *et al.* (2016) shows that 16.7% of health facilities had non-functional refrigerators, 54.8% had no access to effective transportation, and 11.9% had experienced vaccine stockouts in the three months prior to the survey. In 2013, 96% of Nigeria's health facilities were found to have no working refrigerators and 43% of the existing cold chain equipment was found to be non-functional (Nigeria's Ministry of Health (2013)). The critical condition of Nigeria's cold chain is corroborated by Ophori *et al.* (2014), who found that despite large quantities of cold chain equipment, much of it remains non-functional and beyond the point of repair. According to Ethiopia's multiyear plan, in addition to lack of regular maintenance, 35% of cold chain equipment (such as refrigerators/freezers) is nonfunctioning, and 83% of equipment is at least 10 years old (Ethiopia's Federal Ministry of Health (2010)). Since such infrastructural inadequacies negatively impact vaccine access, we seek to enable immunization activities, even when refrigeration is limited or unavailable.

In addition to the infrastructural inadequacies, as stated in the WHO's strategic plan, one of the reasons for subpar outreach efforts is that many vaccine supply chains lack data-informed decision-making which can result in poor inventory management practices. These logistical issues can lead to costly missed opportunities or waste due to the perishable nature of vaccines. As Zaffran *et al.* (2013) state, most vaccine supply chains are ill-equipped to handle the increasing demands leading to reduced coverage (via stockouts) and costly waste if logistics are not adequately supported.

India: Immunization Schedule

Antigen	**VVM**	**cm^3**	**Cost**	**Birth**	**6 Wk**	**10 Wk**	**14 Wk**	**9 Mo**	**16 Mo**	**>16 Mo**
BCG	30	3.3	0.075	✓						
HepB	30	4.4	0.175	✓						
OPV	2	2.0	0.18	✓	✓	✓	✓		✓	
Penta	7	7.8	1.35		✓	✓	✓			
JapEnc	14	6.9	0.41					✓	✓	
Measles	14	2.6	0.237					✓	✓	
DPT	14	3.0	0.2						✓	✓
TT	30	2.1	0.11							✓

Table 3.1: The immunization schedule for India, where the VVM rating, timetable, cost per dose, and volume per dose are obtained from WHO (2016) and the WHO's immunization forecast tool.

The waste due to logistical supply chain failures such as expiry, exposure to extreme temperatures, and breakage may have been somewhat de-emphasized in the past due to low vaccine prices (e.g., the cost per dose of diphtheria, pertussis, and tetanus (DPT) vaccine is only about $0.20, see Table 3.1). However, that dismissal loses relative credit with new sophisticated vaccines such as the pneumococcal vaccine costing as much as $7 per dose. Hence, the fact that many countries see 50% wastage rates is simply unsustainable (World Health Organization (2005)). Shen *et al.* (2014) also identifies the necessity of incorporating and generating data to guide the decisions surrounding the supply chain, citing it as one of the key elements for a reliable network. As such, there is a dire need for information systems that can support better management of vaccine chain coordination when little data are available.

Despite these difficulties, vaccine logistics policies have been formed without consideration to the actual stability of the vaccines under suboptimal conditions. As Galazka *et al.* (1998) aptly states, “However, this approach has led to the gradual emergence of a dogmatic view of the cold chain, preventing health workers from taking full advantage of the actual heat stabilities of different vaccines.” To take advantage of the heat stability properties of different vaccines, a relatively new WHO approved approach termed Outside of Cold Chain (OCC) has been developed which allows for managing vaccine supply chains outside of the typical temperature ranges (see, e.g., Villadiego (2008) and the references within). OCC procedures allow for transportation and storage in cool or ambient temperatures while still guaranteeing the potency of the vaccine. This can help to reduce the dependency on refrigeration capacity, decrease the need for heavy ice packs used to cool vaccine carriers and cold-boxes, and extend the cold-life of storage containers by permitting a wider range of temperature environments in the cold chain.

To make such an endeavor possible, it is necessary to be able to ensure the potency of vaccines in the presence of less-than-ideal temperature conditions. Vaccine vial monitors (VVMs), which are temperature sensitive monitors placed on individual vials of vaccines, are present on nearly all vaccinations in current production, and are available in a variety of temperature sensitivity levels, ranging from the highly sensitive VVM2 to the resilient VVM30 (see Table 3.1 for the sensitivities of a variety of vaccines). VVMs enable a vaccine to be used even after exposure to temperatures so long as the VVM has not expired.

OCC practices have helped to enable vaccination campaigns, especially in areas where transportation and storage capacities are limited. For example, Juan-Giner *et al.* (2014) demonstrated the feasibility of using a tetanus toxoid vaccine in an experiment conducted in Chad. The vaccine remained potent even though it was exposed to a maximum of 30 days of exposure to ambient temperatures ($< 40^\circ$ C).

Ren *et al.* (2009) studied the potency of vaccines for measles (which is heat-sensitive) and hepatitis B (which is freeze sensitive) which were kept outside cold chain in China. The OCC techniques have even been studied with vaccines most sensitive to temperature. Specifically, the oral polio vaccine (OPV) is notoriously heat sensitive, yet Halm *et al.* (2010) and Zipursky *et al.* (2011) investigate the effects of OPV kept outside of cold chain in a vaccination campaign in Mali and Chad, respectively, and find that with proper care, OCC techniques can still be effective in immunization activities. These studies imply that with careful management, the cold chain may be more flexible than the conventional 2° - 8° C temperature range specification in the last mile of delivery.

We develop a new inventory control methodology that takes advantage of OCC procedures in order to help provide routine vaccination services for communities with limited infrastructure while directly taking into account the major related complications arising from the uncertainty of demands and the potential for increased deterioration due to exposure. Therefore, we model the last mile of the supply chain in a robust framework similar to a multi-product newsvendor problem (MPNP) which will help determine level of inventory necessary to accommodate immunization sessions. We aim to provide answers to the questions:

- How can vaccine availability be improved without large infrastructure investment?
- To what degree can improved demand forecasts reduce the strain on the supply chain?
- How should vaccines be distributed in environments where demand is highly uncertain?

We design an algorithm for determining optimal policies in the case of general forecasts and also deliver insights for the special case of Normal forecasts via bounds and other analytical results, which reveal the relative priorities for ordering different kinds of immunizations when capacity constraints become tight. Additionally, we identify the benefits of our robust approach via a numerical case study in Section 3.6 that uses real population data from an IHC in Bihar, India. We find that a robust policy-maker can reduce costs due to wastage while still achieving high immunization coverage, especially in the presence of large levels of ambiguity. Furthermore, we show that utilizing the robust OCC approach can reduce the required transportation and refrigeration capacities for providing high levels of immunization coverage and discuss the relative importance of transportation and refrigeration capacities.

The remainder of this chapter is organized as follows: we present a literature review composed of the current state of vaccine supply chains and related work on MPNPs in Section 3.2. Then we formulate the problem in Section 3.3, and proceed to solve the single-period formulation of our problem in Section 3.4 whose results are used in Section 3.5 to develop techniques for handling a multi-period context. Finally, we provide a numerical analysis in Section 3.6 that tests our approach in simulated settings.

3.2 Literature Review

Much research has been conducted for the purpose of improving vaccine supply chains in developing countries via large simulation models, studied under a variety of settings to illuminate the potential effects of policy changes on cost and vaccine availability. Studies such as Assi *et al.* (2013), which considers the implications of removing the regional level of Niger's vaccine supply chain, or Brown *et al.* (2014), which proposes alternative inventory transportation strategies, examine the effects of changes to the supply chain network itself. Other studies investigate the impact of decision-making within the network, such as Dhamodharan and Proano (2012) or Assi *et al.* (2011), which make recommendations for vaccine vial sizing, or Haidari *et al.* (2013), which weighs the relative importance of increasing storage capacity vs transportation capacity. Finally, others consider the forecasting and management of inventory levels such as Mueller *et al.* (2016), which weighs the benefits of commercial forecasting systems throughout the vaccine supply chain, or Rajgopal *et al.* (2011) which compares several strategies for managing routine immunizations at an IHC in a setting with unlimited transportation and storage capacity. However, these studies are based on simulations that feature fully known stochastic features and rely on established inventory polices to capture the behavior of the network as a whole. Since in reality, demand distributions are not fully known to policy-makers, we approach the management of inventory in the last mile of the vaccine supply chain via a more detailed model that allows us to consider ambiguity in the distribution of demand and incorporate transportation/storage capacity constraints.

The study of robust techniques which protect against distributional ambiguity in inventory control problems has a rich trove of literature, mainly focusing on newsvendor variations and related problems. The first studies in this area, such as the pioneering work of Scarf (1959) followed by Gallego and Moon (1993) consider moment-based approaches, where only mean and variance information is used to generate robust policies. More recent work focuses on data-driven approaches such as Wang *et al.* (2016), which considers ambiguity sets based on the likelihood function, or Xin *et al.* (2013) which studies the time-consistency properties of multi-period newsvendor problems. We refer to Gabrel *et al.* (2014) for an in-depth review of recent work in robust inventory control problems. In our work, since vaccine supply chains currently face limitations in transportation and storage capacity while making ordering decisions with respect to several vaccine types, we focus on robust solutions that jointly protect against distributional ambiguities of the entire suite of vaccines in a setting with capacity constraints, where demand evolves according to an autoregressive process.

Since decisions are based on maximizing coverage levels while simultaneously reducing the costs resulting from both wastage and holding excess inventory subject to both transportation and refrigeration constraints, the resulting problem is closely related to the MPNP with budget constraints. This type of newsvendor variant, first studied by Hadley (1963), is a single-period inventory problem with several demand streams and constraints on order quantities. Many studies utilize Lagrange-multiplier techniques to solve single constraint problems. Erlebacher (2000) develops optimal order solutions in the special cases of similar cost/demand structures as well as the case of uniformly distributed demands. Abdel-Malek *et al.* (2004) gives special attention to the case of uniform and exponential demands, as well as a general algorithm for establishing optimal, or near-optimal solutions. Moon and Silver (2000) investigate a dynamic programming method used in both distribution-known and distribution-free approaches which can guarantee integer-valued orders. Zhang *et al.* (2009) utilizes properties of optimal solutions to develop highly efficient algorithms for both continuous and discrete demand cases. Using these ideas, Zhang (2012) extends this work for multiple constraints. Vairaktarakis (2000) considers three related robust formulations of the problem based on worst-case demand realizations and finds efficient solutions to these objectives. Other studies such as Ben-Daya and Raouf (1993), Lau and Lau (1995), and Lau and Lau (1997) study cases with multiple constraints and design effective solution methodologies and heuristics. For other related studies, see Khouja (1999), Abdel-Malek *et al.* (2004), Abdel-Malek and Montanari (2005), Zhang and Hua (2008) and the references within.

3.3 Problem Description

In regular periods (bi-monthly or monthly, see, e.g. Assi *et al.* (2013) and Haidari *et al.* (2013)), n different types of vaccines are collected to be administered via fixed and/or outreach sessions. IHC workers travel to the district level to pick up vaccines to satisfy demand (in terms of doses), and are not required to place orders in advance for future periods (i.e., orders are typically placed with zero lead-time). Vaccination schedules imply that a given period's demand has dependence with past demands since they direct the timetable for the immunization of each child. For example, in many African countries, the diphtheria-tetanus-pertussis (DTP) vaccine is scheduled to be administered in 3 doses for infants at 6, 10, and 14 weeks of age. Similarly, Oral Polio Vaccine (OPV) is scheduled to be administered to newborns at 6, 10, and 14 weeks of age. To capture these dependencies, we assume that vaccine demand is forecasted according to a p-th order vector autoregressive process (VAR(p)). Thus, the forecast vector $\hat{\mathbf{V}}^t = (\hat{V}_1^t, \ldots, \hat{V}_n^t)'$ for the n vaccine types at period t is given by

$$\hat{\mathbf{V}}^t = \mathbf{a}_0 + A_1 \hat{\mathbf{V}}^{t-1} + \ldots + A_p \hat{\mathbf{V}}^{t-p} + \boldsymbol{\epsilon}^t, \tag{3.1}$$

where $\boldsymbol{\epsilon}^t \sim N(\mathbf{0}, \Omega)$ is a zero mean multivariate error term with covariance matrix Ω made up of components $\Omega_{i,j}$ for $i, j \in \mathcal{N} = \{1, 2, \ldots, n\}$. In (3.1), "$\,'\,$" denotes the transpose operator, $p \in \mathbb{Z}^+$, $\mathbf{a}_0 = (a_1^0, \ldots, a_n^0)'$ is an n-dimensional vector, and A_i (for $i = 1, \ldots, p$) is an $n \times n$ matrices with components $a_{j,l}^i$ for $j, l \in \mathcal{N}$. For tractability, we assume that the VAR(p) process given by (3.1) is time-stationary (in long run). We let $\hat{f}_i^t$ and $\hat{F}_i^t$ denote the marginal probability density and cumulative distribution function for $\hat{V}_i^t$ (i.e., demand forecast for vaccine i at time t), and let $\hat{f}^t$ denote the joint density function for $\hat{\mathbf{V}}^t$ given all relevant previous realizations $\hat{\mathbf{v}}^{t-1}, \ldots, \hat{\mathbf{v}}^{t-p}$. We assume that $\hat{f}^t$ is such that $\mathbb{P}(\hat{V}_i^t < 0)$ is negligible for all $i \in \mathcal{N}$ since $\hat{\mathbf{V}}^t$ represents a forecast for demand.

As described in the Section 3.2, determining the demand process for vaccines is challenging for multiple reasons. Often the last mile of the vaccine supply chain is subject to limited, or poor data quality. Furthermore, if new vaccines are introduced or additional outreach sessions are implemented, forecasting the true demand for these new inclusions can be difficult simply due to uncertainty in the target populations. Hence, the forecasted demand distribution $\hat{f}^t$ is not completely reliable, and can be subject to error. To make decisions that are robust to this error, we assume that the true distribution for demand $\mathbf{V}^t$, f^t, is chosen by Nature and lies in an ambiguity set surrounding the VAR(p) forecasted distribution, $\hat{f}^t$. This ambiguity set is defined as

$$\mathbb{D}(\hat{f}^t, \eta) = \left\{ f \in \mathcal{P}_{\hat{f}^t} : \sum_{i=1}^{n} \int_{\forall v_i : \hat{f}_i^t(v_i) > 0} f_i^t(v_i) \ln \frac{f_i^t(v_i)}{\hat{f}_i^t(v_i)} dv_i \leq \eta \right\}, \tag{3.2}$$

where $\mathcal{P}_{\hat{f}^t}$ denotes the set of all distributions on $\mathbb{R}^n$ that are absolutely continuous to $\hat{f}^t$ and f_i^t denote the marginals of f^t for $i \in \mathcal{N}$. Hence, $\mathbb{D}(\hat{f}^t, \eta)$ includes all densities whose total KL-divergence from the marginals of $\hat{f}^t$ do not exceed η. Importantly, we focus on ambiguity sets surrounding the marginals of $\hat{f}^t$ rather than its joint density since, in MPNPs, it is well known that the correlation structure of the density chosen by Nature at time t does not affect the expect cost of any ordering decision. Hence, by focusing on $\mathbb{D}(\hat{f}^t, \eta)$, we utilize an ambiguity set that encompasses the key features that can affect the policy-maker's decisions. $\mathbf{V}^{t-j}$ are generated from previous forecasts of demand in accordance with (3.1), hence, to ensure starting conditions, we assume that at least p periods of demand have been initially recorded to facilitate ordering decisions in the early periods.

In accordance with the proposed partial OCC strategy, all vaccines are typically transported to the IHC via a vaccine carrier with cold packs. This ensures the necessary temperature requirements for heat-sensitive vaccines, and extends the time to expiration for heat-resistant vaccines (see, e.g., Chen and Kristensen (2009)). Upon reaching the IHC, highly heat sensitive vaccines are placed in refrigeration, and less sensitive vaccines are stored within the vaccine carrier for the duration of a period. [1] Thus, the decision-maker must order vaccines with respect to capacity constraints on both the refrigerator and vaccine carrier. To model these capacity constraints, we assume that each dose of a vaccine of type i has volume $w_i > 0$, and let $r_i = w_i$ if the vaccine is to be placed in refrigeration and $r_i = 0$ otherwise. Letting $\mathbf{w} = (w_1, \ldots, w_n)'$ and $\mathbf{r} = (r_1, \ldots, r_n)'$ refer to vectors of these parameters, and $b_c > 0$ and $b_r > 0$ be the total capacity available to the vaccine carrier and refrigerator, respectively, we define

$$\mathcal{X}(\mathbf{s}^t) = \left\{ \mathbf{x} \in \mathbb{R}^n_+ \middle| x_i \geq s^t_i, \mathbf{w}'(\mathbf{x} - \mathbf{s}) \leq b_c, \mathbf{r}'\mathbf{x} \leq b_r \right\} \tag{3.3}$$

as the set of feasible "order-up-to" quantities of vaccines. Here, order-up-to quantities must satisfy refrigerator and vaccine carrier capacities given the level of vaccines in refrigeration prior to ordering, which we denote by the non-negative vector $\mathbf{s}^t = (s^t_1, \ldots, s^t_n)'$. If vaccines are ordered to the level $\mathbf{x}^{t-1}$ in period $t-1$, the held inventory $\mathbf{s}^t$ is the realization of $\mathbf{S}^t = (\mathbf{S}^{t-1} + \mathbf{x}^{t-1} - \mathbf{V}^{t-1})^+$, where the operator "+" used on any vector returns a new vector with each component being the maximum of the original component and zero. Nature, as a robust agent, aims to maximize costs by selecting densities from $\mathbb{D}(\hat{f}^t, \eta)$ at each period t given the decision-maker's current level of vaccines and forecast given by (3.1), which depends on the past p estimates. Hence, we let an admissible policy for Nature be the mapping $\xi : \mathbb{R}^n_+ \times \mathbb{R}^p_+ \to \mathbb{D}(\hat{f}, \eta)$ and define Ξ as the set of all such policies. Here, $\hat{f}$ is determined via (3.1) so that $\mathbf{V}^t_\xi = (V^t_{1,\xi}, V^t_{2,\xi}, \ldots, V^t_{n,\xi})'$ denotes the random variable of demand at time t. The decision-maker orders vaccines according to the current level of inventory and the

[1] Our study also encompasses vaccine management in areas with no refrigeration, such as areas targeted for outreach sessions, IHCs with no available electricity or access to refrigerators, or mobile strategies used to target nomadic populations.

demand density chosen by Nature. As such, we define an admissible ordering policy as a mapping $\zeta : \mathbb{R}_+^n \times \Xi \to \mathcal{X}(\mathbf{S})$, and define $\mathcal{Z}$ as the set of all such admissible ordering policies. Furthermore, we let $\mathbf{X}_\zeta^t = (X_{1,\zeta}^t, \ldots, X_{n,\zeta}^t)'$ denote the order quantity vector at time t under policy ζ, where $X_{i,\zeta}^t$ is the order quantity at time t for type i vaccine.

To differentiate between vaccine classes, we define $\mathcal{O}_c = \{i \in \mathcal{N} | r_i = 0\}$ as the set of OCC vaccines that will be stored in the vaccine carrier, and similarly $\mathcal{O}_r = \{i \in \mathcal{N} | r_i > 0\}$ as the set of vaccines to be refrigerated upon arrival at the IHC. Naturally, OCC vaccines will experience degradation due to the additional heat exposure. Thus, each vaccine within $\mathcal{O}_c$ left at the end of the period is wasted (and hence, $s_i^t = 0$ for all $i \in \mathcal{O}_c$), incurring a cost to the system. Refrigerated vaccines that do not experience such accelerated degradation still induce system costs due to their natural expiration date and the valuable/limited space they occupy in the cold chain. To capture these costs, we consider a holding/overage cost, and denote it by $\mathbf{h} = (h_1, \ldots, h_n)'$.

Missing vaccination opportunities due to lack of available vaccines is highly costly in most immunization chains (especially those in Africa), since it reduces coverage as well as confidence in immunization sessions. In the case that there is a relative weighting to the importance of administrating a given vaccine, we let $\mathbf{u} = (u_1, \ldots, u_n)'$ be a positive vector denoting the cost of missed opportunity (i.e., underage cost) for each unit of each type vaccine. [2] In most vaccine policies, missing a vaccine of type i is considered to be more deleterious than incurring a waste (see, e.g., World Health Organization (2005)). Hence, we expect that $u_i > h_i$, though we do not impose this condition.

It is the aim of the decision-maker to manage vaccine levels so as to minimize the

[2] This cost can be set to be the vector of all ones in the case that there is no relative weighting between the cost of such missed opportunities.

overall underage and overage costs in a robust way (i.e., considering potential forecast errors). Therefore, defining

$$H_i(x_i, v_i) = u_i(v_i - x_i)^+ + h_i(x_i - v_i)^+, \tag{3.4}$$

We consider the objective to be finding the policy that minimizes the worst-case infinite-horizon discounted cost:

$$\arg\inf_{\zeta \in \mathcal{Z}} \sup_{\xi \in \Xi} \limsup_{T \to \infty} \mathrm{E}\left[\sum_{t=1}^{T} \sum_{i=1}^{n} \beta^t H_i(X_{i,\zeta}^t, V_{i,\xi}^t)\right], \tag{3.5}$$

where $\beta \in (0, 1)$ is a discounting factor. This robust objective allows us to take into account the ambiguities surrounding demand for vaccines. The level of ambiguity within each component is reflected in η; a policy-maker with high levels of ambiguity would choose η to be very large whereas a policy-maker with less ambiguity would choose η to be small, effectively shrinking the ambiguity set to be near the estimate for each demand model. In this way, Nature chooses the density for demand via ξ by considering all models that come within a proximity measured via the KL-divergence to the non-robust nominal model.

3.4 Single-Period Properties

To identify strategies for solving (3.5), we first carefully examine its single period properties. To this end, in Section 3.4.1, we start by studying the single-period version of (3.5) with a general underlying nominal density. Using the results we obtain, in Section 3.4.2, we then tackle the single-period version of (3.5) for the special case in which the nominal density is multivariate normal. In Section 3.5, we then use insights gained to solve (3.5) in its general format (i.e., multi-period version). Finally, in Section 3.6 we perform various numerical experiments using a real-world case study of immunization in Bihar, India, (as well as synthetic data) to gain more insights.

3.4.1 Robust Single-Period Problem

Suppressing the notation for period, t, and letting $\hat{f}$ denote a general positive nominal density, the single-period version of (3.5) takes the form

$$\mathop{\arg\inf}_{\mathbf{x}\in\mathcal{X}(\mathbf{s})} \sup_{f\in\mathbb{D}(\hat{f},\eta)} \mathrm{E}_f\left[\sum_{i=1}^{n} H_i(x_i, \mathrm{V}_i)\right]. \tag{3.6}$$

To solve (3.6), we define

$$G(\mathbf{x},\alpha) = \alpha\sum_{i=1}^{n} \ln \mathrm{E}_{\hat{f}}\left[e^{H_i(x_i,V_i)/\alpha}\right] + \alpha\eta, \tag{3.7}$$

and present the following proposition which identifies an equivalent problem to (3.6).

Proposition 3.1 (**Robust Objective Equivalence**). *If there exists $\alpha > 0$ such that*
$\sum_{i=1}^{n} \mathrm{E}_{\hat{f}}\left[e^{H_i(x_i,V_i)/\alpha}\right] < \infty$ *for all $\mathbf{x}\in\mathcal{X}(\mathbf{s})$,* (3.5) *is equivalent to the objective:*

$$\mathop{\text{minimize}}_{\mathbf{x}\in\mathcal{X}(\mathbf{s}),\alpha\geq 0} \; G(\mathbf{x},\alpha). \tag{3.8}$$

Objective (3.8) *is jointly convex in $\mathbf{x}$ and α, and if $\mathbf{x}^*$ and α^* solve* (3.8)*, the maximizing (i.e., worst-case) density in* (3.6) *is*

$$f(\mathbf{v}) = \prod_{i=1}^{n} \hat{f}_i(v_i)\frac{e^{H_i(x_i^*,v_i)/\alpha^*}}{\mathrm{E}_{\hat{f}}\left[e^{H_i(x_i^*,V_i)/\alpha^*}\right]}. \tag{3.9}$$

Proposition 3.1 is established via the class of robust optimization problems investigated by Hu and Hong (2012). However, our work differs by focusing on the KL-divergence from the marginal densities. Since (3.8) is a convex optimization problem, it results in a simpler framework for solving Problem (3.5).

Inspecting (3.7), and noting that η is only present in the term $\alpha\eta$, it is intuitive for α to be monotonically decreasing in η. This means that a policy-maker with large levels of ambiguity will have a smaller α^*. This fact is established in the following lemma by showing that the log term of (3.7) is monotone in α.

Lemma 3.1 (**Monotonicity of α^* in η**). *$\alpha^*(\eta) = \arg\min_{\alpha \geq 0} G(\mathbf{x}^*, \alpha)$ is decreasing in η.*

Due to this monotone relation between α and η, insights regarding α via the simpler, equivalent objective (3.6), correspond directly to the level of ambiguity expressed in η. Hence, by studying the impact of α, we can reveal the effect of a policy-maker's ambiguity level without directly searching for η.

Next, we note that if we can characterize $\mathbf{x}^*$ as a function of α which we denote by $\mathbf{x}^{*\alpha} = (x_1^{*\alpha}, x_2^{*\alpha}, \ldots, x_n^{*\alpha})$, solving (3.8) reduces to a simple univariate search for an optimizing α. With this goal, we define the following functions which we term *marginal benefit functions*:

$$\pi_i^\alpha(x_i) = \frac{G_{x_i}(x_i, \alpha)}{w_i}. \tag{3.10}$$

In (3.10) $G_{x_i} = \partial G / \partial x_i$ (for $i \in \mathcal{N}$) which is a function of only x_i and α (and not x_j for $j \neq i$), since G is completely additively separable. Though π_i^α are interpreted slightly differently between OCC and refrigerated vaccines as seen in the forthcoming Theorem 3.1, marginal benefit functions generally act as the rate of cost change relative to the rate of volume consumption for a given vaccine. Furthermore, since it is the derivative of a convex function, π_i^α is monotone, and has a lower bound equal to $-u_i/w_i$ (see Lemma B.2 in Online Appendix B.3). Hence, we define an associated inverse function to π_i^α:

$$\pi_i^{\alpha,-1}(q) = \begin{cases} \inf\{x \geq 0 | q \leq \pi_i^\alpha(x)\} & q \geq -u_i/w_i \\ 0 & q < -u_i/w_i. \end{cases} \tag{3.11}$$

The case with $q < -u_i/w_i$ in (3.11) is incorporated for notational convenience since the function is minimized at $-u_i/w_i$. Naturally, if capacity is unlimited ($b_c = b_r = \infty$), the optimal x_i as a function of α, $x_i^{*\alpha}$, corresponds either to the zero of π_i^α or to s_i

if $\pi_i^\alpha(s_i) > 0$. We denote this special unconstrained order-up-to level as $\hat{x}_i^\alpha$, and note that it acts as the "desired order-up-to" value that can be used to identify optimality conditions for $x_i^{*\alpha}$.

In the following theorem, we utilize the properties of the marginal benefit functions to identify optimality conditions. This, in turn, will grant (a) an efficient way of finding solutions to (3.6), and (b) analytical insights into the optimal order-up-to quantities.

Theorem 3.1 (**Robust Optimality Conditions**). *Let*

$$\nu_c = \max\Big\{\nu \leq 0 | \sum_{i\in\mathcal{N}} \max\{\pi_i^{\alpha,-1}(\nu), s_i\} - s_i \leq b_c\Big\},$$

$$\nu_r = \max\Big\{\nu \leq 0 | \sum_{i\in\mathcal{N}} \max\{\pi_i^{\alpha,-1}(\nu), s_i\} r_i \leq b_r\Big\},$$

and $$\nu_b = \max\Big\{\nu \leq 0 | \sum_{i\in\mathcal{O}_c} \max\{\pi_i^{\alpha,-1}(\nu), s_i\} - s_i + \sum_{i\in\mathcal{O}_r} \max\{\pi_i^{\alpha,-1}(\nu_r), s_i\} - s_i \leq b_c\Big\}.$$

(i) If $\sum_{i=1}^n (\hat{x}_i^\alpha - s_i)w_i \leq b_c$ and $\sum_{i=1}^n \hat{x}_i^\alpha r_i \leq b_r$, then $x_i^{*\alpha} = \hat{x}_i^\alpha$ for all $i \in \mathcal{N}$.

(ii) If $\sum_{i=1}^n (\hat{x}_i^\alpha - s_i)w_i \leq b_c$ and $\sum_{i=1}^n \hat{x}_i^\alpha r_i > b_r$, then $x_i^{*\alpha} = \hat{x}_i^\alpha$ for all $i \in \mathcal{O}_c$ and $x_i^{*\alpha} = \max\{\pi_i^{\alpha,-1}(\nu_r), s_i\}$ for all $i \in \mathcal{O}_r$.

(iii) If $\sum_{i=1}^n (\hat{x}_i^\alpha - s_i)w_i > b_c$ and $\sum_{i=1}^n \hat{x}_i^\alpha r_i \leq b_r$, then $x_i^{*\alpha} = \max\{\pi_i^{\alpha,-1}(\nu_c), s_i\}$ for all $i \in \mathcal{N}$.

(iv) If $\sum_{i=1}^n (\hat{x}_i^\alpha - s_i)w_i > b_c$ and $\sum_{i=1}^n \hat{x}_i^\alpha r_i > b_r$, then $x_i^{*\alpha} = \max\{\pi_i^{\alpha,-1}(\nu_b), \pi_i^{\alpha,-1}(\nu_c), s_i\}$ for all $i \in \mathcal{O}_c$, $x_i^{*\alpha} = \max\big\{\min\{\pi_i^{\alpha,-1}(\nu_c), \pi_i^{\alpha,-1}(\nu_r)\big\}, s_i\}$ for all $i \in \mathcal{O}_r$.

Noting that the refrigeration constraint does not affect vaccines in $\mathcal{O}_c$, Theorem 3.1 demonstrates the intuitive notion that optimal order quantities aim for the zero of the marginal benefit function. If this quantity can be met for all vaccines without violating constraints, the optimal order is $\hat{x}_i^\alpha$. Otherwise, if an optimal order for a vaccine $i \in \mathcal{O}_c$ is less than its infinite-capacity quantity ($x_i^{*\alpha} < \hat{x}_i^{\alpha}$), then $x_j^{*\alpha} < \hat{x}_j^{\alpha}$ for all other non-zero orders, where by "zero-order", we refer to the event that $x_i^{*\alpha} = s_i$, since this is the case where no new vaccines are ordered. In a similar vein, if an order for a vaccine $i \in \mathcal{O}_r$ has $x_i^{*\alpha} < \hat{x}_i^{\alpha}$, then $x_j^{*\alpha} < \hat{x}_j^{\alpha}$ for all other non-zero orders $j \in \mathcal{O}_r$.

Therefore, if unconstrained order quantities cannot be met, the limiting constraint is tight. Moreover, the marginal benefit function under optimal order quantities within each class of vaccines ($\mathcal{O}_c$ and $\mathcal{O}_r$) are equal when constraints are tight, which is intuitive since they share a common constraint. Similarly, when the limiting constraint is the vaccine carrier (and not the refrigeration), all non-zero order quantities have identical marginal benefit values. Importantly, Theorem 3.1 also allows us to construct the optimal order quantities of the robust problem (3.6) by performing a search for specific values of π_i^{α}, which is the basis of the Single-Period Algorithm presented in Table 1. This algorithm only involves finding (1) $\hat{x}_i^{\alpha}$ for each vaccine class, and (2) at most three target values for the modified robust marginal benefit function ratios π^{α}. Notably, if π_i^{α} is known, each of these steps can be easily accomplished via binary search algorithms allowing for the easy calculation of robust order quantities. The Single-Period Algorithm of Table 2 can also be used as a myopic solution for the multi-period problem, but we provide further characterization for the multi-period setting in Section 3.5.

3.4.2 Robust Single-Period Problem with Normal Demand

Since each period's forecast is multivariate normal in a VAR(p), to further characterize the solution to (3.5), we next generate insights by investigating the single-period problem in the special case where the nominal forecast $\hat{f}$ is multivariate normal. Again, assuming that the probability of negative demand is negligible and letting μ_i and σ_i denote the mean and standard deviation parameters of $\hat{f}_i$ (the marginal

Table 2: Single-Period Algorithm

Initialize. Calculate $\hat{x}_i^\alpha$ for $i \in \mathcal{N}$ and determine case (i)-(iv) via Theorem 3.1.

Case (i). Let $x_i^{*\alpha} = \hat{x}_i^\alpha$ for $i \in \mathcal{N}$.

Case (ii). Let $x_i^{*\alpha} = \hat{x}_i^\alpha$ for $i \in \mathcal{O}_c$. Find

$$q_r^* = \max \left\{ q \leq 0 \middle| \textstyle\sum_{i \in \mathcal{O}_r} \max\{\pi_i^{\alpha,-1}(q), s_i\} w_i \leq b_r \right\},$$

and let $x_i^{*\alpha} = \max\{\pi_i^{\alpha,-1}(q_r^*), s_i\}$ for all $x \in \mathcal{O}_r$.

Case (iii). Find

$$q_{cr}^* = \max \left\{ q \leq 0 \middle| \textstyle\sum_{i=1}^n (\pi_i^{\alpha,-1}(q) - s_i)^+ w_i \leq b_c \right\},$$

and let $x_i^{*\alpha} = \max\{\pi_i^{\alpha,-1}(q_{cr}^*), s_i\}$ for all $i \in \mathcal{N}$.

Case (iv). If $b_c - b_r + \sum_{i=1}^n s_i r_i > 0$, find the quantities

$$q_c^* = \max \left\{ q \leq 0 \middle| \textstyle\sum_{i \in \mathcal{O}_c} \pi_i^{\alpha,-1}(q)^+ w_i \leq b_c - b_r + \sum_{i=1}^n s_i r_i \right\},$$

$$q_r^* = \max \left\{ q \leq 0 \middle| \textstyle\sum_{i \in \mathcal{O}_r} \max\{\pi_i^{\alpha,-1}(q), s_i\} w_i \leq b_r \right\}.$$

· If $\sum_{i \in \mathcal{O}_c} \hat{x}_i^\alpha w_i \leq b_c - b_r + \sum_{i=1}^n s_i r_i$, let $x_i^{*\alpha} = \max\{\pi_i^{\alpha,-1}(q_r^*), s_i\}$ for all $x \in \mathcal{O}_r$ and $x_i^{*\alpha} = \hat{x}_i^\alpha$ for all $x \in \mathcal{O}_c$.

· Else if $\sum_{i \in \mathcal{O}_c} \hat{x}_i^\alpha w_i > b_c - b_r + \sum_{i=1}^n s_i r_i$ and $q_c^* \geq q_r^*$,

let $x_i^{*\alpha} = \max\{\pi_i^{\alpha,-1}(q_r^*), s_i\}$ for all $x \in \mathcal{O}_r$ and $x_i^{*\alpha} = \max\{\pi_i^{\alpha,-1}(q_c^*), s_i\}$ for all $x \in \mathcal{O}_c$.

· Else find $x_i^{*\alpha}$ according to Case (iii) for all $i \in \mathcal{N}$.

of $\hat{f}$ with respect to $\hat{V}_i$), the partial derivatives of $G(\mathbf{x}, \alpha)$ introduced in (3.7) (and hence π_i^α introduced in (3.10)) can be expressed in closed form (see, e.g. Propositions B.1 and B.2 in Online Appendix B.3). Fortunately, these show that the condition $\sum_{i=1}^n \mathrm{E}_{\hat{f}} \left[e^{H_i(x_i, V_i)/\alpha} \right] < \infty$ for all $\mathbf{x} \in \mathcal{X}(\mathbf{s})$ required in Proposition 3.1 is obviously satisfied for any positive finite α, which allows us to utilize (3.8).

Using these closed form expressions, we first investigate bounds by considering the case where there is infinite carrier and refrigeration capacity ($b_c = b_r = \infty$). These bounds are closely related to the non-robust problem (i.e., when $\eta = 0$). Hence, we define

$$\hat{x}_i = \max \left\{ \hat{F}_i^{-1} \left(\frac{u_i}{u_i + h_i} \right), s_i \right\},$$

which is either the well-known newsvendor critical fractile based on the forecasted demand for the vaccine type i, or the current level of vaccine in storage with respect to the nominal forecast.

PROPOSITION 3.2 (**Infinite-Capacity Robust Bounds**). *For all $i \in \mathcal{N}$, suppose $\hat{f}_i \sim N(\mu_i, \sigma_i^2)$. Then:*

1. *If $h_i \leq u_i$,*

$$\max\left\{s_i, \mu_i + \frac{(u_i - h_i)\sigma_i^2}{2\alpha}, \hat{x}_i\right\} \leq \hat{x}_i^\alpha$$
$$\leq \max\left\{s_i, \min\left\{\mu_i + \frac{(u_i - h_i)\sigma_i^2}{2\alpha} + \frac{\alpha \ln(u_i/h_i)}{u_i + h_i}, \hat{x}_i + \frac{u_i\sigma_i^2}{\alpha}\right\}\right\}$$

2. *If $h_i \geq u_i$,*

$$\max\left\{s_i, \mu_i + \frac{(u_i - h_i)\sigma_i^2}{2\alpha} + \frac{\alpha \ln(u_i/h_i)}{u_i + h_i}, \hat{x}_i - \frac{h_i\sigma_i^2}{\alpha}\right\} \leq \hat{x}_i^\alpha$$
$$\leq \max\left\{s_i, \min\left\{\mu_i + \frac{(u_i - h_i)\sigma_i^2}{2\alpha}, \hat{x}_i\right\}\right\}$$

Proposition 3.2 implies that in the uncapacitated robust problem, the optimal order quantities are bounded by an interplay between α, the difference between h_i and u_i, the variance of the nominal model, and the non-robust ordering quantity $\hat{x}_i$. The bounds converge in many asymptotic cases. For example, when a policy-maker places equal weight on a missed opportunity and an overage (i.e., when u_i approaches h_i), the bounds converge to the point $\max\{\mu_i, s_i\}$. This is intuitive, since Nature should see no clear advantage to choosing distributions that induce higher as opposed to lower demands when overage and underage costs are equal with a symmetric nominal distribution. However, this behavior is not expressed in practice, since policy-makers typically prefer to see a low outage rate compared to that of overage if they have the capacity to do so. In fact, a typical vaccine policy-maker would settle cost of missed opportunity moderately greater than that of overage. This is inclination is more dramatically demonstrated in the case when $u_i \to \infty$. For this case, the bounds show that the order quantity goes to infinity as expected due to an increasing fear of stock outages, which is much more aligned with most IHCs' goal of near-zero outages in developed countries where capacity is ample. As noted earlier, however, in developing countries, the limited capacity disallows reaching this goal.

We also note that when α becomes small, $\frac{\alpha \ln(u_i/h_i)}{u_i+h_i}$ terms in the bounds given in Proposition 3.2 go to zero, and since α must decrease as η increases (Lemma 3.1), the bounds converge as ambiguity increases. Since these bounds either go to infinity or zero depending on the underage/overage costs, a policy-maker with high levels of ambiguity and no capacity constraints will engage in extreme ordering behavior. On the other hand, as α becomes large or σ_i becomes small, the bounds converge to the original newsvendor order quantities which indicates that as the level of ambiguity decreases, the ordering decisions tend toward the non-robust optimal case.

Interestingly, the term $\frac{\alpha \ln(u_i/h_i)}{u_i+h_i}$ in upper and lower bounds is not affected by σ_i. However, as α decreases, these bounds deviate by at least a factor of the variability in the nominal distribution. Hence, once again, from the perspective of a policy-maker who is not facing tight capacity restrictions, higher variability and higher ambiguity aversion results in more extreme ordering decisions. Since the robust problem with capacity restrictions necessitate ordering at most as many vaccines as those in the uncapacitated robust problem, $x_i^{*\alpha}$ with $b_c = b_r = \infty$ serve as an upper bound to the optimal capacitated order quantities.

In certain cases, tighter upper/lower bounds than those found in Proposition 3.2 can be obtained. Due to the closed form of π_i^α and the symmetry with respect to underage and overage costs found therein, if both $\frac{u_i\sigma_i}{\alpha}$ and $\frac{h_i\sigma_i}{\alpha}$ lie beneath or above 0.84 (which occurs whenever the level of ambiguity is high *or* low), the following upper/lower bounds can be established.

Corollary 3.1 (**Cost/Ambiguity-Dependent Bounds**). *If* $\frac{u_i \sigma_i}{\alpha} \geq \frac{h_i \sigma_i}{\alpha} \geq 0.84$, *or if* $\frac{u_i \sigma_i}{\alpha} \leq \frac{h_i \sigma_i}{\alpha} \leq 0.84$, $\hat{x}_i^\alpha \leq \max\{s_i, \mu_i + \frac{(u_i - h_i)\sigma_i^2}{\alpha}\}$. *Otherwise, if* $\frac{h_i \sigma_i}{\alpha} \geq \frac{u_i \sigma_i}{\alpha} \geq 0.84$, *or if* $\frac{h_i \sigma_i}{\alpha} \leq \frac{u_i \sigma_i}{\alpha} \leq 0.84$, $\hat{x}_i^\alpha \geq \max\{s_i, \mu_i + \frac{(u_i - h_i)\sigma_i^2}{\alpha}\}$.

Via these tightened bounds, Corollary 3.1 further reinforces that higher levels of ambiguity result in more extreme ordering behavior in the case of infinite capacity. Moreover, tying these results with Proposition 3.2 can help a policy-maker to plan for appropriate refrigeration and carrier capacities at different ambiguity levels since the infinite-capacity case acts as target for order quantities in the case with limited capacity.

In these cases where refrigeration and/or carrier capacity is finite, by investigating π_i^α in light of Theorem 3.1, it is easy to show that order quantities are monotonically increasing in μ_i (see, e.g., Proposition B.3 in Online Appendix B.3). An equally intuitive, but less direct result maintains that the optimal order quantity $x_i^{*\alpha}$ also experiences monotone behavior in σ_i.

Proposition 3.3 (**Monotonic Ordering in** σ_i). *For* $i \in \mathcal{N}$, *if* $u_i \geq h_i$ *and* $x_i^{*\alpha} \geq \mu_i$, $x_i^{*\alpha}$ *is nondecreasing in* σ_i. *Otherwise, if* $u_i \leq h_i$, $x_i^{*\alpha}$ *is nonincreasing in* σ_i.

Proposition 3.3 implies an important relationship between variance and the effect of ambiguity on order quantities. It can be shown that α must increase in systems with larger σ_i (see, e.g., Proposition B.4 in Online Appendix B.3). Hence, comparing the relative order quantities of vaccine type i in the case where $u_i \geq h_i$, policy-makers with larger ambiguity levels and lower σ_i can actually order *fewer* type i vaccines than policy-makers with lower levels of ambiguity and higher σ_i. Hence, variability has the effect of increasing the flexibility of Nature's ambiguity set, resulting in more larger order quantities, even when a policy-maker's ambiguity levels are relatively small.

Importantly, these monotone results (in μ_i and σ_i) consider only a fixed α. Inspecting the bounds of Proposition 3.2, the infinite-capacity order quantities move in a linear fashion with respect to changes in the mean and variance. Now, increases in μ_i when capacity constraints are tight, and increases in σ_i induce an exponential increase in cost in (3.7), hence the α term in (3.8) must also increase to compensate and reduce costs. Proposition 3.2 suggests that larger α result in less extreme ordering behavior, hence, optimal orders to (3.8) change only by a linear factor of mean and variance if this intuition is correct. In this way, a policy-maker can be much less concerned about the sensitivity surrounding mean and variance estimates in their initial forecasts since parameter shifts only perturb ordering behavior on the same magnitude as the original non-robust case.

Therefore, to reinforce these insights and investigate ordering behavior as the level of ambiguity changes, we identify a monotone relation between α and optimal order quantities in the following proposition. Since α affects all vaccines simultaneously, simple monotone properties relating to optimal order quantities are not established in the same manner as Proposition 3.3 (and Proposition B.3 in Online Appendix B.3). However, since optimal order quantities become independent decisions between vaccine classes in the infinite capacity case, we can find monotone properties of $\hat{x}_i^\alpha$ with respect to α, which in turn imply important behaviors of $x_i^{*\alpha}$.

Proposition 3.4 (**Monotonic Ordering in** α). *For all $i \in \mathcal{N}$, if $u_i \geq h_i$ $\hat{x}_i^\alpha$ is nonincreasing in α. Otherwise, if $u_i \leq h_i$, $\hat{x}_i^\alpha$ is nondecreasing in α.*

In addition to the fact that it implies reduced concern about policy sensitivity to parameters (as discussed above), Proposition 3.4 yields similar policy implications to that of Proposition 3.3. It demonstrates that as the level of ambiguity η increases, the target order quantities become more extreme. Tying this result with those found in Proposition 3.2 and Corollary 3.1, we can further improve the infinite-capacity bounds. If $\overline{x}^\alpha$ and $\underline{x}^\alpha$ represent the tightest upper/lower bounds implied by Proposition 3.2 and Corollary 3.1, Proposition 3.4 implies that $\min_{0<\alpha\leq\hat{\alpha}} \overline{x}^\alpha$ and $\max_{\alpha\geq\hat{\alpha}} \underline{x}^\alpha$ provide an upper and lower bound to the infinite capacity case with $\hat{\alpha}$ that are at least as tight as the original bounds when $u_i \geq h_i$. Otherwise, if $u_i \leq h_i$, $\min_{\alpha\geq\hat{\alpha}} \overline{x}^\alpha$ and $\max_{0<\alpha\leq\hat{\alpha}} \underline{x}^\alpha$ represent the tightened bounds. Hence, a policy-maker can refer to monotone bounds in α for the purpose of studying changes in ordering behaviors and better determining appropriate levels of ambiguity aversion by observing the infinite-capacity case (which act as ordering targets), as shown in Section 3.6.

Importantly, Proposition 3.4 paired with Theorem 3.1 allows us to gain insights into the behavior of the optimal policy with respect to changes in η in the capacitated case. From the following corollary, we have conditions under which tight capacity constraints will remain tight as the level of ambiguity increases (i.e., as η increases).

COROLLARY 3.2 (**Full Capacity**). *If $u_i \geq h_i$ for all $i \in \mathcal{N}$ and $\sum_{i=1}^n (x_i^{*\alpha} - s_i)w_i = b_c$ then $\sum_{i=1}^n (x_i^{*\hat{\alpha}} - s_i)w_i = b_c$ for all $\hat{\alpha} \leq \alpha$.*

This result reinforces the notion that robust decision-makers who fear outages (i.e. $u_i \leq h_i$) and reach their carrier capacity constraint continue to do so as the level of ambiguity (i.e. η) increases. The reason this does not carry through to the refrigeration constraint is that the decision-maker may need to alter the composition of orders to include a greater quantity of $\mathcal{O}_c$ vaccines in favor of refrigerated vaccines, and hence forfeit remaining capacity in refrigeration.

Since η affects all vaccine demands simultaneously, Proposition 3.4 only shows a monotonic correspondence to more extreme order quantities in the infinite-capacity case or where capacity is non-restrictive. However, as long as $u_i > h_i$, it can be shown that $\pi_i^\alpha(x_i)$ approaches the lower bound $-u_i/w_i$ (established in Lemma B.2) when α is sufficiently small. This implies that the target values for the robust marginal benefit functions in the robust algorithm decrease when η is large. This induces extreme ordering behavior, even in the case with capacity restrictions.

PROPOSITION 3.5 (**Orders as $\eta \to \infty$**). *Define* $\gamma_c = \min_{i \in \mathcal{O}_c} -\frac{u_i}{w_i}\mathbb{1}\{u_i > h_i\}$ *and* $\gamma_r = \min_{i \in \mathcal{O}_r} -\frac{u_i}{w_i}\mathbb{1}\{u_i > h_i\}$. *If* $u_i \neq h_i$ *for all* $i \in \mathcal{N}$, *then optimal order quantities have the following asymptotic properties:*

(i) *If* $u_i < h_i$, $\lim_{\eta \to \infty} x_i^* - s_i = 0$.

(ii) *If* $\gamma_c < 0$ *and* $\gamma_c \leq \gamma_r$, $\lim_{\eta \to \infty} x_i^* = s_i$ *for all* $i \in \mathcal{N}$ *that have* $-u_i/w_i > \gamma_c$, *and* $\lim_{\eta \to \infty} b_c - \sum_{i=1}^{n}(x_i^* - s_i)w_i = 0$.

(iii) *If* $\gamma_r < \gamma_c$, $\lim_{\eta \to \infty} x_i^* = s_i$ *for all* $i \in \mathcal{O}_r$ *with* $-u_i/w_i > \gamma_r$ *and* $\lim_{\eta \to \infty} x_i^* = s_i$ *for all* $i \in \mathcal{O}_c$ *that have* $-u_i/w_i > \gamma_c$. *Furthermore,*

$$\lim_{\eta \to \infty} \min \left\{ b_r - \sum_{i=1}^{n} x_i^* r_i, b_c - \sum_{i=1}^{n} (x_i^* - s_i) w_i \right\} = 0,$$

and if $\gamma_c < 0$, $\lim_{\eta \to \infty} b_c - \sum_{i=1}^{n}(x_i^* - s_i)w_i = 0$.

Proposition 3.5 shows that the optimal ordering level for all vaccines with $u_i < h_i$ go to zero-orders as η, the level of ambiguity, becomes large. This implies that for very expensive vaccines that also have tight expiration dates, a policy-maker with extremely high levels of ambiguity should order nothing. However, in the case where $u_i > h_i$, each vaccine will only have a non-zero ordering level if it has $-u_i/w_i = \gamma_c$ when $i \in \mathcal{O}_c$, or $-u_i/w_i = \gamma_r$ if $i \in \mathcal{O}_r$. This implies that the determining factor for non-zero ordering quantities when η is large is the ratio $-u_i/w_i$. If these ratios are unique, the optimal ordering decision becomes to fill the carrier capacity with a single vaccine type if $\gamma_c < 0$ and $\gamma_c \leq \gamma_r$, and at most two types of vaccines if $\gamma_r < \gamma_c$. In the case where all vaccines have equal underage costs (i.e., $u_i = 1$ with $h_i < 1$ for all

$i \in \mathcal{N}$), if an OCC vaccine has smallest volume per dose (i.e., w_i), the carrier is filled with only this vaccine. Otherwise, if a refrigerated vaccine has smallest volume per dose, the carrier is filled with this vaccine and the OCC vaccine with smallest w_i.

3.5 Robust Multi-Period Problem

Despite solving the single-period problem, the multi-period problem (3.5) is highly complex. This is because the problem lies in a continuous space and is highly dimensional, requiring consideration for $\hat{\mathbf{v}}^{t-1}, \ldots, \hat{\mathbf{v}}^{t-p}$ in order to forecast each demand epoch. Furthermore, Nature's policy is no longer easily determined due to the fact that leftover inventory can potentially be carried over. Hence, in general, the problem (3.5) is out of reach.

However, under certain conditions we can still utilize the single-period problem to gain insights into the optimal solutions of the multi-period problem. One obvious case is where even refrigerated vaccines have tight expiration dates, and hence, cannot be carried over to future periods. More generally, consider the case where the action space is expanded to enable the decision-maker to modify $\mathbf{s}^t$ to any $\mathbf{0}$. When this action is enabled, we say that the decision-maker can "return" vaccines. In the case that all leftover vaccines in refrigeration at the IHC can be sent back to the district depot, this action implies vaccines are returned to be stored at the district level.

Assuming that the vaccines are returnable, and the carrier's capacity is not binding, we provide the following result which casts the multi-period objective as a series of single-period problems.

PROPOSITION 3.6 (**Multi-Period Objective Equivalence**). *If vaccines are returnable and $b_c = \infty$, the optimal ordering quantities to the multi-period objective* (3.5) *at each period t can be determined via*

$$\underset{\mathbf{x}\in\mathcal{X}(\mathbf{0}),\alpha\geq 0}{\text{minimize}} \;\; \alpha\sum_{i=1}^{n}\ln \mathrm{E}_{\hat{f}_t}\left[e^{H_i(x_i,V_i)/\alpha}\right] + \alpha\eta. \tag{3.12}$$

The result also holds if vaccines are not returnable and $b_c < \infty$, but $\mathcal{O}_c = \mathcal{N}$.

Proposition 3.6 allows us to consider two important cases. The first case, when returns are possible with $b_c = \infty$, arises when the carrier's capacity is not a limiting factor and either (1) policy-makers assign low costs to vaccine waste resulting from shifting refrigeration inventories, or (2) when transportation to the upstream of the vaccine supply chain is possible. This can occur when deliveries are accomplished in round trips by the district depot (see, e.g., Brown *et al.* (2014)), or when mobile strategies are utilized to reach remote communities such as those in Nigeria, Nicer-acgua, and Kenya (see, e.g., Ryman *et al.* (2008), and the references therein). It also is possible whenever there is adequate freezer and carrier capacity so that carriers can be outfitted with cold packs for the return trip. Obviously, infinite-capacity carriers are a fictitious construct; however, when deliveries are accomplished in vehicles that transport a large volume of vaccines at once, vaccine capacity is no longer a limiting constraint on ordering. Furthermore, even when transportation capacity is limited, in settings where deliveries can be made with higher frequency, even a finite-capacity carrier can render the constraint induced by b_c non-binding.

The second case, when $\mathcal{N} = \mathcal{O}_c$, describes a purely OCC process. This situation arises whenever an IHC is planning of outreach sessions or when refrigeration simply is not available at the IHC, like many IHCs in Sub-Saharan Africa where electricity is scarce and equipment is subject to failure (see, e.g., Ophori *et al.* (2014), Haidari *et al.* (2013)). When refrigeration is not available, our approach enables the policy-maker to maximize coverage while minimizing the wastage costs resulting from leftover vaccines from a given demand period. When planning for outreach sessions, it is important to establish appropriate levels of inventory since transporting vaccines for use in such sessions necessarily exposes vaccines to temperatures and other factors that can accelerate their degradation. Hence, our approach helps determine inventory levels that consider maximizing the coverage for sessions, while still minimizing costs of exposure due to vaccines leaving refrigeration.

If the district depot delivers vaccines directly to IHCs, another related scenario may arise. As described in Proposition 3.6, since vaccines are delivered to multiple IHCs, trucks must be endowed with large transportation capacity. Hence, from the perspective of the IHC, transportation capacity can be viewed as infinite. However, in this case, storage capabilities at the IHC are binding due to the refrigeration capacity, as well as the OCC storage in vaccine carriers available at the IHC. Hence, if transportation capacity is infinite, yet the storage capacity for carriers at the IHC is limited because of either (1) the number of carriers available, or (2) the number of carriers that can be kept in the cool chain (with respect to limit on cool packs which are to be cycled through refrigeration), we find that, a single-period strategy can again be employed so long as vaccines are returnable.

Corollary 3.3 (**District-Delivered Policy**). *If vaccines are returnable and delivered to an IHC via the district depot, the optimal order-up-to quantities to the multi-period objective* (3.5) *at each period* $t = 1, 2, \ldots$ *can be determined via*

$$\underset{\mathbf{x}\in\hat{\mathcal{X}},\alpha\geq 0}{\text{minimize}} \ \alpha \sum_{i=1}^{n} \ln \mathrm{E}_{\hat{f}_t}\left[e^{H_i(x_i,V_i)/\alpha}\right] + \alpha\eta, \tag{3.13}$$

where $\hat{\mathcal{X}}$ *is given by* $\hat{\mathcal{X}} = \left\{ \mathbf{x} \in \mathbb{R}^n_+ \middle| \sum_{i\in\mathcal{O}_c} x_i w_i \leq b_c, \sum_{i\in\mathcal{O}_r} x_i w_i \leq b_r \right\}.$

In (3.13), Corollary 3.3 shows that optimal order quantities in this case is determined by two fully myopic singly-constrained problems, and hence can be solved via two instances of the Single-Period Algorithm. This is due to the fact that in this case, any quantity of OCC and non-OCC vaccines can be transported to the IHC, though storage capacity is limited by refrigeration and vaccine carrier space, respectively. Therefore OCC and non-OCC vaccines do not share any capacity constraints which accounts for the modified action space $\hat{\mathcal{X}}$, since non-OCC vaccines no longer are transported in the carrier used to store OCC vaccines, however, Nature's demand distribution must still be selected jointly in (3.13).

In general, the multi-period problem appears to be out of reach. However, when we consider the system with backordering where the remaining inventory transitions to $s_i^t = \min\{x_i^{t-1} - v_i^{t-1}, 0\}$ for OCC vaccines and $s_i^t = x_i^{t-1} - v_i^{t-1}$ for refrigerated vaccines, instead of the lost-sales model, we can still characterize optimal policies as modified base-stock.

Proposition 3.7 (**Base-Stock Optimality**). *When demands are backordered, for the infinite-horizon problem, there exists a stationary base-stock level* $\mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{R}^n$ *such that the optimal action to the multi-period objective* (3.5) *from an initial inventory position* $\mathbf{s}$ *and previous* p *forecasts is* $x_i = s_i$ *if* $s_i \geq y_i$*, and* $x_i = y_i$ *if* $(\max\{y_1, s_1\}, \ldots, \max\{y_n, s_n\}) \in \mathcal{X}(\mathbf{s})$.

Systems with backordering are well-known to be approximations of lost-sales models when the level of coverage is high, hence under conditions with enough capacity to ensure low levels of outages, a policy-maker can expect optimal policies to be very near a modified base-stock policy. This enhances the intuition that ordering targets exist such that, with sufficient capacity, a policy-makers can order up to these levels regardless their current inventory state $\mathbf{s}^t$.

3.6 Numerical Experiments

To generate further insights, we first consider a case study, and examine the multi-period problem using real-world data from an IHC in India. We then turn our attention to the single-period problem and perform various experiments under a variety of parameter configurations (i.e., synthetic data).

3.6.1 Multi-Period Problem. Immunization in Bihar (India)

Lim *et al.* (2016) use partial data of villages near the Tetia Bambar IHC in Bihar, India to develop a representative population map for the purposes of determining outreach center locations. Figure 3.1 shows this map with village locations, populations outside 5km, three routine outreach immunizations, and center locations as determined by Lim *et al.* (2016). We utilize this map to study our vaccine management problem in a multi-period context in two main scenarios: the first scenario considers only the planning of the outreach immunization, and the second scenario plans inventory levels for the entire IHC.

Table 3.1 shows relevant immunization data that we use including the level of thermostability by VVM, the volume and price per dose, and the immunization schedule for children in India. We begin by generating nominal parameters for our underlying VAR(p) model by simulating a system that features i.i.d. newborn arrivals in each period with a perfect adherence to the schedule (see Online Appendix B.1).

We first consider inventory control policies over the outreach sessions to the center with population 10,749, (see Figure 3.1). In this case, the problem acts as a fully OCC problem since vaccines are distributed at the location and are returned to the IHC within the day. Since schedules for vaccines BCG/HepB and JE/Measles schedules are identical, we assume that the demand for each type of vaccine is also identical which allows us to treat these as bundled vaccines. Then, we study the problem with $\mathbf{u} = (2, 1, 2, 1, 1, 1)'$, $\mathbf{h} = \lambda(0.25, 0.11, 0.647, 0.2, 1.35, 0.18)'$, $\mathbf{w} = (7.7, 2.1, 9.5, 3, 7.8, 2)'$ in accordance with Table 3.1, where vaccines are indexed in the order BCG/HepB, TT, JE/Measles, DTP, Penta, and OPV respectively. We

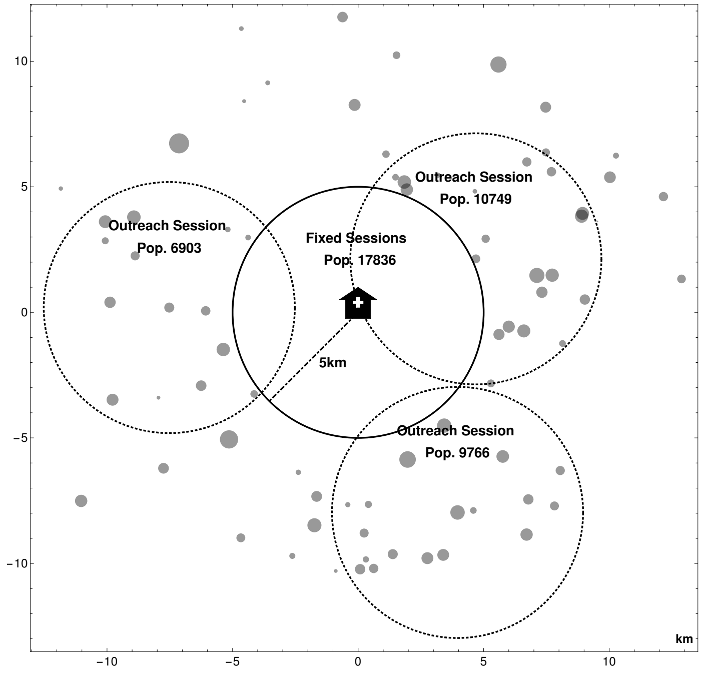
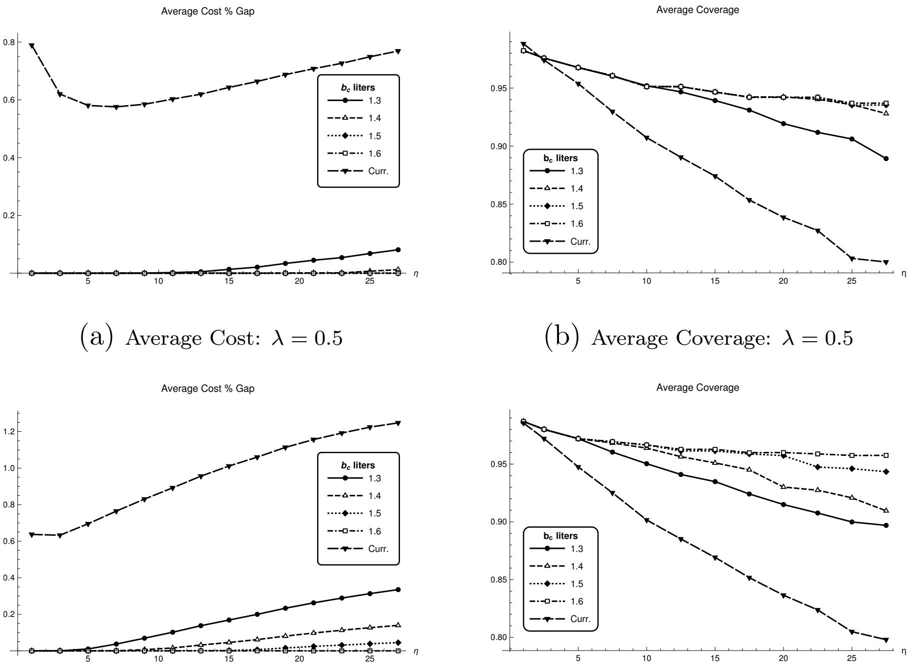


Figure 3.1: Tetia Bambar IHC in Bihar, India with 3 outreach sessions. Circle diameters represent population of villages outside 5km from the IHC. Total population = 57,734. Map generated from data provided in Lim *et al.* (2016).

further assume that each outreach session occurs on bimonthly intervals per recommendations from WHO (2010). Here, we let $\lambda > 0$ represent a scaling factor for our holding costs to maintain the relative cost difference while investigating the effects of different holding cost levels. By changing λ, we generate insights into the effect of a policy-maker's aversion to outages on immunization coverage.

Figure 3.2 shows the average cost and coverage levels with respect to varying levels of ambiguity aversion, capacity levels, and holding costs compared to a traditional outreach policy. The traditional order quantities are established via WHO (2010) which simply prescribes ordering 1.33 times the average vaccine demand. This traditional order policy requires approximately 1.3 liters of carrier capacity which is fairly typical of vaccine carriers. To ensure that our robust policies do not gain advantage



(a) Average Cost: $\lambda = 0.5$

(b) Average Coverage: $\lambda = 0.5$

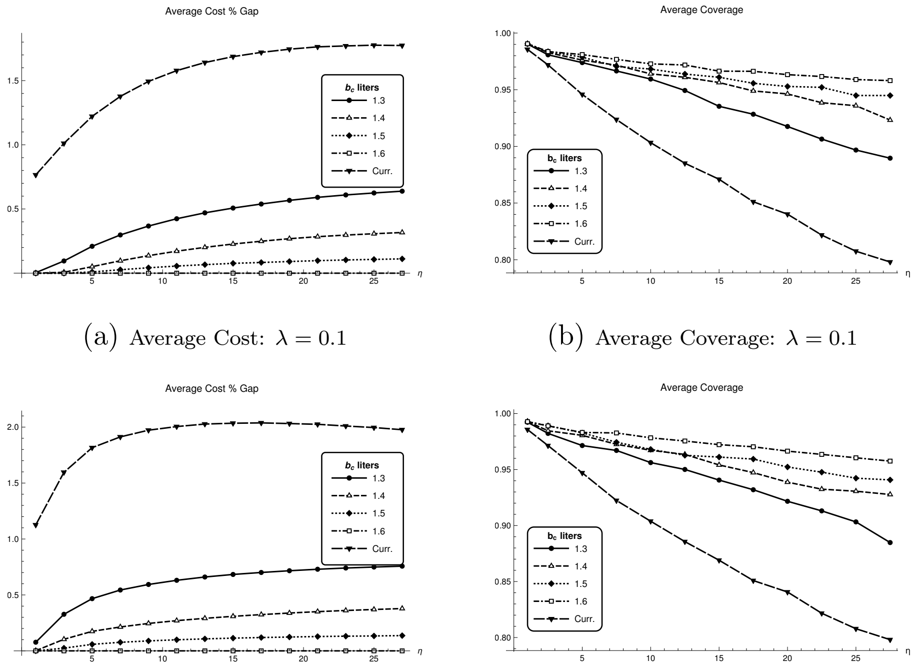
(c) Average Cost: $\lambda = 0.25$

(d) Average Coverage: $\lambda = 0.25$

Figure 3.2: Average discounted cost % gap between policies (left column) and coverage (right column) in outreach sessions with various levels of ambiguity aversion when $\lambda = 0.5, 0.25, 0.1$, and 0.05 and $\beta = 0.95$. The current policy orders 1.33 of each vaccine's expected demand (WHO (2010)).

simply via increased carrier capacity, we vary carrier capacity levels at $1.3, 1.4, 1.5$, and 1.6 liters.

The right column of Figure 3.2 shows that when η is small, the average coverage under each policy is also large since in such systems there is only minor demand ambiguity. However, even when such ambiguity is small, the left column shows that the traditional policy pays a heavy price in holding costs since this policy does not forecast based on the schedule, but rather assumes i.i.d. demand in each period. These large holding costs showcase the trade-off that traditional vaccine policies make in return for simplicity and demonstrates the potential gains that can be obtained by



(a) Average Cost: $\lambda = 0.1$

(b) Average Coverage: $\lambda = 0.1$

(c) Average Cost: $\lambda = 0.05$

(d) Average Coverage: $\lambda = 0.05$

Figure 3.3: Average discounted cost % gap between policies (left column) and coverage (right column) in outreach sessions with various levels of ambiguity aversion when $\lambda = 0.5, 0.25, 0.1$, and 0.05 and $\beta = 0.95$. The current policy orders 1.33 of each vaccine's expected demand (WHO (2010)).

considering the natural autoregressive nature of vaccine demand. When less waste due to overage can be obtained while still maintaining acceptable coverage levels, the strain on the cold-chain's capacity is naturally reduced via smaller inventory levels. This can be expressed via reduced strain on both cold storage and transportation capacity at each level of the supply chain. Therefore, we make the following:

OBSERVATION 3.1 (**Outreach Overage Costs**). *Robust inventory policies can reduce vaccine waste while still maintaining high coverage levels even when the level of ambiguity is small.*

As the level of distrust for the model increases and η becomes large, the disparity between robust optimal policies and traditional policies becomes more pronounced. Even when the robust policy is constrained to the carrier capacity of 1.3 liters, the robust optimal policy reduces the costs and underages by nearly a factor of 2 throughout the spectrum of η. When the capacity is increased, these gains are even more apparent in both cost and underages. This is especially true when λ is small since in this case the robust policy can leverage the extra capacity to better protect against outages; when η is small and λ is large, as in Figures 4.1a and 4.1b, these improvements are negligible since the high overage costs lead to orders near, or below the capacity constraint. However, when η is large and λ is small, increasing capacity yields moderate cost/underage improvements, though these improvements see diminishing returns as seen in Figures 3.2c-3.3d. Hence, we make the following:

Observation 3.2 (**Increasing Outreach Capacity**). *Outreach systems with vaccines that have low overage costs (e.g., BCG, OPV, and TT) experience the greatest impact from increases in carrier capacity.*

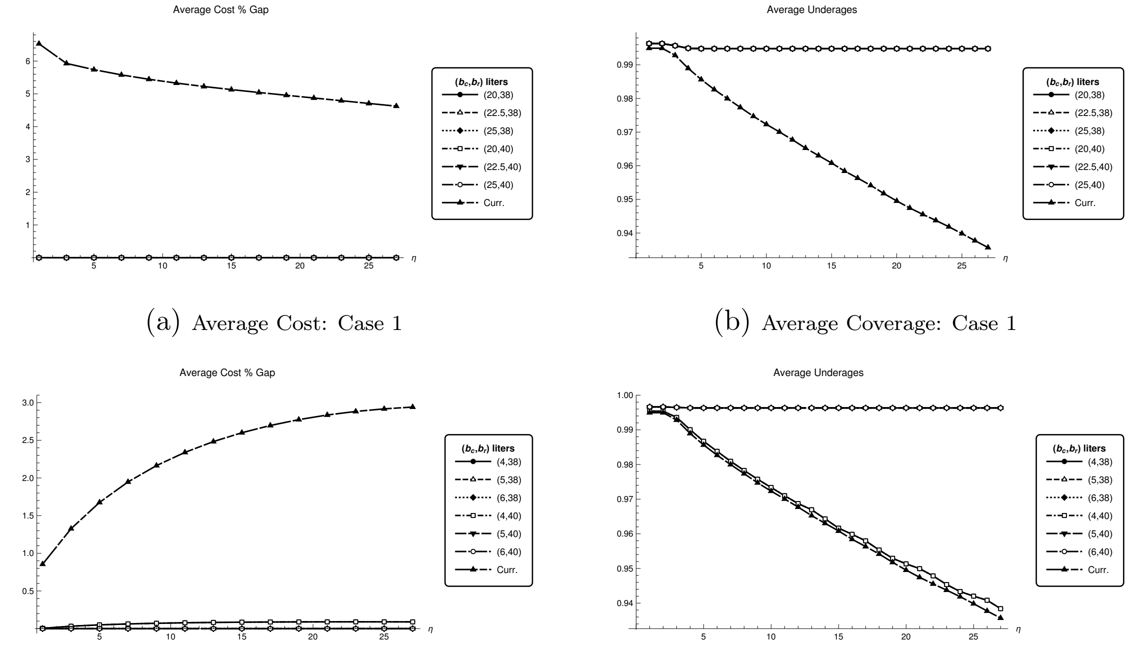
Next, we consider the scenario where some vaccines are stored in refrigeration, and some in cool storage within the vaccine carrier at the IHC. Since BCG/HepB and TT are the least heat sensitive vaccines, we let these be our OCC vaccines, whereas JE/Measles, DTP, Penta, and OPV are to remain in refrigeration. Since OCC vaccines experience higher exposure to heat, we let $\mathbf{u} = (2, 1, 2, 1, 1, 1)'$, $\mathbf{h} = (0.25\lambda_1, 0.11\lambda_1, 0.647\lambda_2, 0.2\lambda_2, 1.35\lambda_2, 0.18\lambda_2)'$, $\mathbf{w} = (7.7, 2.1, 9.5, 3, 7.8, 2)'$, where $\lambda_1 > \lambda_2$. We assume that vaccines are stocked in monthly intervals, which is the most common restocking interval for IHCs. Furthermore, since traditional IHC policies dictate 6 weeks order-up-to levels, we assume there is enough refrigeration and transportation capacity to make to achieve 6 weeks of stock with near 100% certainty for OCC vaccines (see Appendix B.1 for further details).

Since determining optimal order quantities when refrigerators are utilized in the general case requires solving a highly dimensional dynamic program with continuous state space, we leverage Proposition 3.6 and Corollary 3.3 to observe the performance of our policies in two settings. In Case 1, we order vaccines and determine Nature's worst-case demand densities under the assumptions of Proposition 3.6. In this way, both the policy-maker and Nature behave myopically, and so long as the constraint on transportation capacity is not tight, this behavior is optimal. In Case 2, we order vaccines according to the assumptions of Corollary 3.3, where vaccines are delivered to the IHC, and b_c and b_r act only as storage capacity constraints for OCC and non-OCC vaccines respectively.

Figure 3.4 shows the results of simulations with $\lambda_1 = 0.25$ and $\lambda_2 = 0.05$ in both cases. In Case 1, though there appears to be large improvements with respect to coverage and cost in our policies to the traditional approach, there is little differentiation between the robust policies when transportation and storage capacities are manipulated. The reason for this is simple: in settings where the traditional policy has enough storage and transportation to satisfy 6 weeks demand, there is more than enough capacity for robust policies to order their desired quantities of vaccines, even when η is moderately large. Thus, in this experiment, the order quantities are almost never restricted by capacity constraints even when we reduce capacity by $5 - 10\%$, hence, there is very little difference between the settings with larger capacities. Therefore, we make the following:

Observation 3.3 (**IHC Capacity**). *Our approach can reduce the current requirements for storage and transportation capacity at the IHC level by* $5 - 10\%$.

Furthermore, it is clear from Figures 3.4a and 3.4b that, even though the traditional policy achieves high coverage when η is small, it does so at the expense of



(a) Average Cost: Case 1

(b) Average Coverage: Case 1

(c) Average Cost: Case 2

(d) Average Coverage: Case 2

Figure 3.4: Average cost % gap (left column) and coverage (right column) in Case 1 (associated with Proposition 3.6) and Case 2 (associated with Corollary 3.3) at IHCs with various levels of ambiguity aversion when $\lambda_1 = 0.25$ and $\lambda_2 = 0.05$. The current policy orders 1.5 of each vaccine's expected infinite horizon demand (WHO (2010)).

high overage costs. Moreover, these order quantities do not adequately protect coverage levels when η becomes large, yet the robust policy is capable of simultaneously attaining high coverage levels with low overage costs, even with high η.

Similar behavior to Case 1 can be observed in Case 2 in Figures 3.4c and 3.4d, where the robust policy demonstrates high performance in reducing costs as opposed to the traditional policy. However, when $b_c = 4$, even though the costs are dramatically reduced, vaccine coverage achieves similar levels to the traditional policy. This indicates that when capacity constraints are tight, Nature greatly limits the capability of attaining high coverage, yet this also implies that small increases in capacity can have dramatic effects on improving coverage, especially when η is high. Hence, we make the following:

Observation 3.4 (**Case 2 Coverage**). *When vaccines are delivered to the IHC in accordance with Corollary 3.3, increases in OCC capacity can greatly increase coverage, especially when demand is highly uncertain.*

3.6.2 Single Period Problem

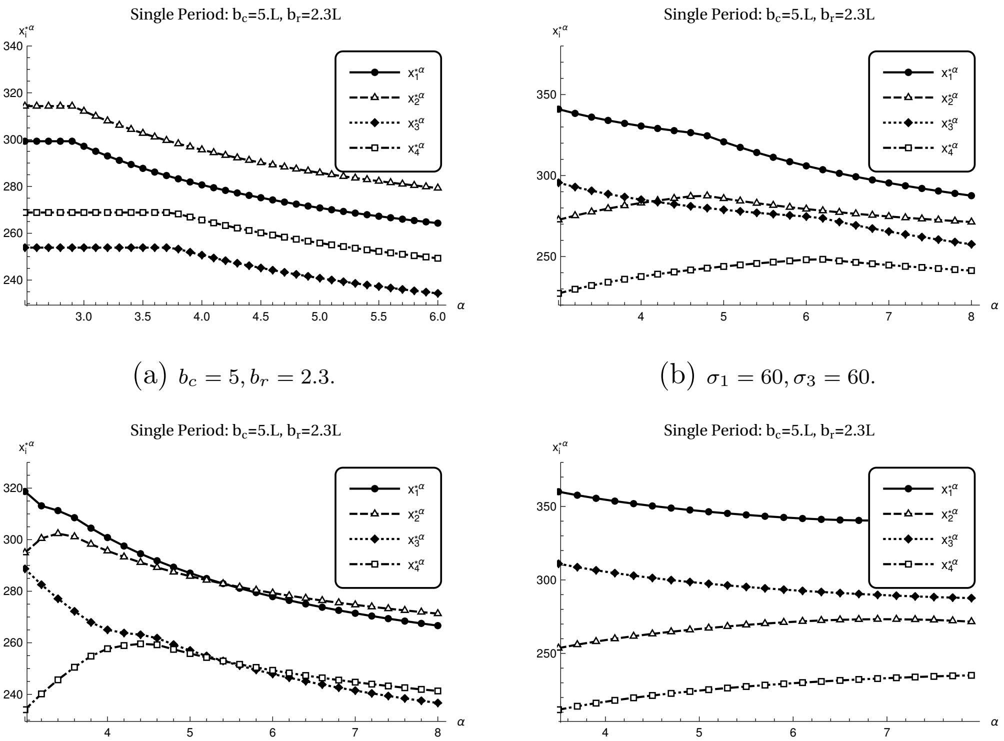
To gain deeper insight into the single period problem, we now examine order behavior under a variety of parameter configurations. For a related investigation on the performance of our analytical bounds, we refer to Online Appendix B.2. Throughout our following analysis, we utilize the BCG vaccine as a working example, whose nominal parameter settings consist of $\mu_i = 220$, $\sigma_i = 40$, $w_i = 0.0044$ (as estimated in Section 3.6.1), and whose underage/overage costs have been normalized to $u_i = 1$ and $h_i = 0.75$ respectively. Throughout our analysis, we consider the case where $u_i > h_i$ since (1) this case is much more closely aligned with real-world vaccine policy-maker's disposition, and (2) the case where $h_i > u_i$ is highly symmetric to that of $u_i > h_i$, hence insights into these cases follow in a similar manner.

In the general single-period problem, we investigate a system with $n = 4$ where $\mathcal{O}_c = \{1, 2\}$ and $\mathcal{O}_r = \{3, 4\}$ in order to understand the effects of parameter shifts when capacity constraints become restrictive. In order to generate insights while avoiding redundancy, we focus on cases with $s_i = 0, h_i \leq u_i$, and a mean fixed to $\boldsymbol{\mu} = (230, 245, 200, 215)'$ to help differentiate ordering behaviors. In cases with non-zero inventory, s_i simply become lower bounds order quantities, and scenarios with $h_i \geq u_i$ yield highly symmetric results to that of $u_i \geq h_i$. Therefore, we study problem instances that deviate from a central setting where $\sigma_i = 40, u_i = 1, h_i = 0.75$ and $w_i = 0.0044$, for all $i \in \mathcal{N}$. This environment features uniform characteristics so that the effect of shifting parameters can be more easily observed.

Figure 3.5 show order quantities in settings that feature binding constraints on both refrigeration and carrier capacity. Since each of the vaccines feature identical features with the exception of their mean, the marginal benefit functions feature identical costs when $x_i^{*\alpha} - \mu_i = x_j^{*\alpha} - \mu_j$, and hence, the difference between their optimal order quantities is exactly the difference between their means until the capacity constraints become tight. As expected via Corollary 3.2, once the capacity becomes a binding constraint, it remains binding as $\alpha \to 0$. Furthermore, due to their identical features, optimal order quantities stay constant with small enough α as shown in Figure 3.5a, hence, as Proposition 3.5 suggests, once a policy-maker achieves a certain level of ambiguity, all optimal orders will remain the same for higher levels of ambiguity.

Variability plays an important role in ordering levels, since generally speaking, higher variability in the nominal distribution permits Nature to choose from a more hazardous pool of densities. Figure 3.5b demonstrates ordering in a setting identical to Figure 3.5a except that σ_1 and σ_3 are increased to 60. As expected from Propositions 3.3 and 3.5, the priority to fill vaccines shifts from x_2 and x_4 to filling x_1 and x_3 as $\alpha \to 0$ since the potential costs from these types become more severe when the variance is higher. Importantly, these increases in variance effect order quantities for non-OCC vaccines more than OCC vaccines since refrigeration and carrier constraints act jointly on non-OCC vaccines, leading to decreased order quantities for vaccines with smaller variance than their OCC counterpart which can be observed in Figure 3.5b, where x_4 begins decreasing prior to x_2 as $\alpha \to 0$. Therefore, increases in variability increases the priority for a vaccine, which can lead a policy-maker to reduce other vaccine orders, especially for non-OCC vaccines.

Turning our attention to ordering behavior with respect to overage/underage Fig-



(a) $b_c = 5, b_r = 2.3$.

(b) $\sigma_1 = 60, \sigma_3 = 60$.

(c) $u_1 = 1.1, u_3 = 1.1$.

(d) $h_1 = 0.1, h_3 = 0.1$.

Figure 3.5: Order quantities for vaccines with $\boldsymbol{\mu} = (230, 245, 200, 215)'$, $u_i = 1$, $h_i = 0.75$, $s_i = 0$, $w_i = 0.0044$ in the base case, where $\mathcal{O}_c = \{1, 2\}$ and $\mathcal{O}_r = \{3, 4\}$.

ures 3.5c and 3.5d show the effect of modifying h_i/u_i in refrigerated vaccines and OCC vaccines. As shown in Figure 3.5d, even with a major decreases in overage costs for vaccines 1 and 3, order quantities are not highly sensitive to α since, for moderate α values, the cost due to overages is relatively minor in comparison to underage costs as a result of the capacity constraints which directly limits the likelihood of overages. Hence, when capacity constraints become tight, a policy-maker can be relatively unconcerned about overage costs as opposed to their treatment of other parameters. Alternatively, Figure 3.5c shows a setting with a relatively modest increase in underage costs for vaccines 1 and 3, which demonstrates that even small increases in underage costs results in large increases in relative order quantities, especially when capacity constraints become tight. Again, this increase is due to the fact that underage is the primary contributor to costs when capacity constraints become binding. Based on these, we make the following:

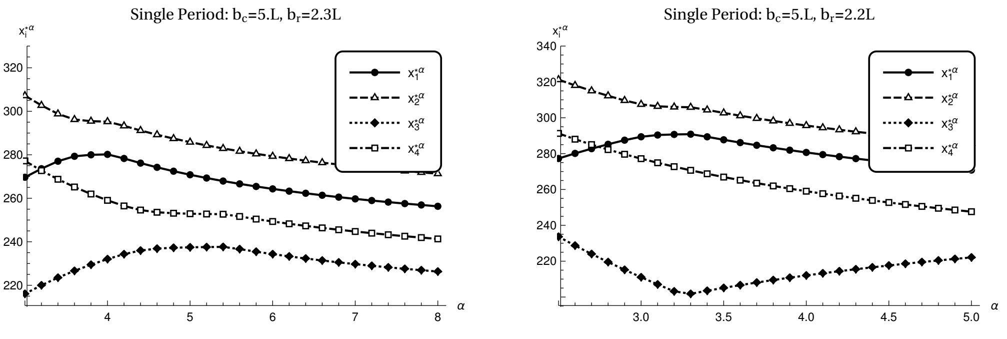
Observation 3.5 (**Effect of Underage Cost**). *When capacity constraints are tight, a policy-maker should place an especially high priority on ordering vaccines with large underage cost (e.g., OPV in high-risk areas).*

Therefore, if some vaccines can be viewed as more critical than others (i.e., those that have a higher chance of reducing mortality), a policy-maker should prioritize these vaccines, even at the cost of reducing the order sizes of less critical vaccines.

The final major factor that affects order quantities is the volume of the vaccine. In Figure 3.6a volume parameters w_1 and w_3 are increased from 4.4ml to 5ml, which discourages such orders for when the capacity constraints become tight. As Figure 3.6b shows, vaccine volume can have dramatic effects, especially when both capacity constraints become active. These behaviors indicate that it can be better to fill vaccines that are more volume-efficient than to fill vaccines that require more space, especially in highly uncertain environments with limited capacities. Thus, we make the following:

Observation 3.6 (**Effect of Vaccine Volume**). *When capacity is limited, ordering vaccines with low volume requirements (e.g., TT, OPV, and Measles) is heavily incentivized, whereas ordering for vaccines with high volume requirements (e.g., Penta and JapEnc) is greatly reduced.*

Since KL-divergence permits a highly diverse ambiguity set, it is interesting to identify the shape of the robust density chosen by Nature. Figure 3.7 demonstrates a few infinite capacity settings where $x_i^{*\alpha} = \hat{x}^{\alpha}$ and $\mu_i = 10, \sigma_i = 1, u_i = 1$. Figures 3.7a and 3.7b show the density shape when $\alpha = 30$ and $\alpha = 20$ respectively with $h_i =$



(a) $w_1 = .005, w_3 = .005, b_r = 2.3.$

(b) $w_1 = .005, w_3 = .005, b_r = 2.2.$

Figure 3.6: Order quantities for vaccines with $\boldsymbol{\mu} = (230, 245, 200, 215)'$, $u_i = 1$, $h_i = 0.75$, $s_i = 0$, $w_i = 0.0044$ in the base case, where $\mathcal{O}_c = \{1, 2\}$ and $\mathcal{O}_r = \{3, 4\}$.

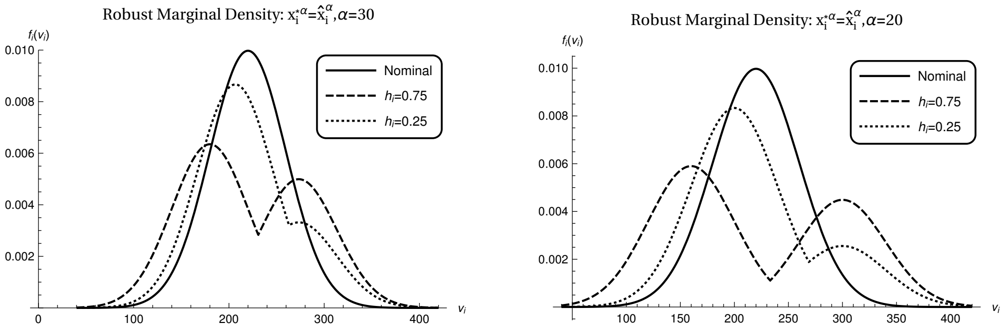
0.75 and 0.25. Interestingly, their shape resembles mixture distributions with two highly distinct phases, with the smaller peak tending toward the case of underages and the larger peak tending toward the case of overages. This implies that the larger order quantities protect against the underage costs by reducing the density for larger realizations. As α decreases, the spread of the robust density increases, while simultaneously further dividing the two peaks of the density. Hence, a robust policy-maker who fears underage (i.e. $u_i \geq h_i$) will actually expect larger overages rather than larger underages when they experience no capacity restrictions. Surprisingly, this contradicts the intuition that the worst-case scenario experiences high levels of demand.

3.7 Conclusion

Maintaining the cold chain for vaccine distribution in developing countries is subject to unreliable population data, limited refrigeration capacities, and aging infrastructure. To gain insights into policies that can improve vaccine delivery in such countries, we develop an approach that takes advantage of the thermostable properties of vaccines. We take advantage of natural autoregressive characteristics implied by vaccine schedules to tackle the high levels of ambiguity experienced in demand forecasts, we apply a robust optimization technique that utilizes ambiguity sets composed of densities constrained by KL-divergence from the nominal forecast.

We provide a variety of analytical findings including (1) an efficient algorithm for determining optimal order quantities, (2) easily calculable bounds, and (3) policy behavior with respect to the level of ambiguity, variability, and capacity restrictions. Furthermore, via a case study based on real-world data, we find that both outreach and fixed immunization strategies can benefit from our approach by reducing waste while ensuring high coverage levels in the presence of forecast ambiguities. Furthermore, since we rely on OCC strategies, our approach can reduce the capacity necessary for adequate coverage, which is important in areas with limited/unreliable refrigeration space, especially as the quantity and selection of vaccines included in immunization schedules increases.

Since our research mainly considers inventory control systems at the last mile of the vaccine supply chain, future studies could investigate decisions at higher levels of the supply chain to help regional and district depots store appropriate levels of inventory while facing uncertain populations. Developing countries suffer from a lack of data-informed decision-making and managerial oversight, hence policy-makers are in need of practical considerations to the propagation of upstream data to help ensure high levels of coverage in the presence of demand ambiguity.



(a) $\alpha = 0.8, x_i^{*\alpha} = \hat{x}_i^{\alpha}$.

(b) $\alpha = 0.8, x_i^{*\alpha} = \hat{x}_i^{\alpha}$.

Figure 3.7: Robust marginal densities f_i for the infinite-capacity case with $\alpha = 20$ and 30, $\mu_i = 220, \sigma_i = 40, u_i = 1$, and $h_i = 0.75$ and 0.25.

Chapter 4

DISTRICT-MANAGED VACCINE SUPPLY NETWORKS WITH DEMAND UNCERTAINTY

4.1 Introduction

Providing immunizations to populations in developing countries is well-known to be one of the most effective actions to combat high mortality rates, especially for children. As such, organizations such as UNICEF, the WHO, and GAVI partner with developing countries in order to enable high vaccination coverage. However, despite these efforts, the vaccine supply chain for developing countries is currently at capacity and is further being strained due to increases in vaccine demands, the introduction of additional vaccines, insufficient storage and transportation capacity, aging equipment, and the deficient utilization of data for forecasting and informed decision-making. As such, to improve vaccination coverage, it is necessary to not only invest in the modernization and expansion of vaccine supply chains, but also to investigate management practices that can ensure high levels of coverage for the purpose of reaching populations in need of immunization.

For this reason, we propose a new management system from the perspective of the district level of the supply chain; instead of the traditional approach which relies on Integrated Health Centers (IHCs) fetching vaccines from the district supply depot, we investigate an approach where vaccines are delivered to the IHC by the district level. The district depot, which supplies vaccines to a number of IHCs, can more effectively manage its own inventory levels if it also assumes responsibility for monitoring the inventory levels of the IHCs. However, in developing countries with limited capability for information sharing, communication between levels of the supply chain is typically poor leading to (1) incomplete knowledge of inventory levels throughout at the IHC level and (2) uncertainty in the true demand for vaccines. We aim to address these issues via (1) improved demand forecasting, (2) reduced capacity strain, (3) proper/safe disposal of vaccine wastes, (4) improved upstream information flow in the supply chain, (5) reduced strain on IHC workers, and (6) additional monitoring at the IHC level of the supply chain.

4.2 Literature Review

In developing countries, vaccines are typically transported from a centrally located national vaccine store to regional stores, then to district depots. IHC workers then fetch vaccines from the district depots and distribute immunizations via outreach and fixed immunization sessions. These supply chains feature a wide range of challenges including aging infrastructure, the necessity of maintaining a cool environment for the product, and limited storage/transportation capacity at each level which hinders the ability to provide full coverage to their population (see, e.g. Nigeria's Ministry of Health (2013), Ophori *et al.* (2014), and Adair-Rohani *et al.* (2013)). To help reduce costs and increase coverage, many studies investigate the potential for modifying current practices and supply chain structures. Among these, in a study on Benin's vaccine supply chain, Brown *et al.* (2014) shows that logistics costs can be reduced by delivering vaccines to IHCs in groups via trucks from higher levels of the supply chain as opposed to a strategy that can serve only a single IHC at a time (usually via motorcycles). In our work, we also consider such a delivery strategy, focusing specifically on a network of IHCs where inventory levels are not observable until delivery and the demand parameter may be learned from incoming data in a Bayesian context.

Along with infrastructural challenges, the WHO's strategic immunization plan also indicates a lack of leadership, ineffective data management, and poor demand forecasting as areas in dire need of improvement (World Health Organization (2014)). As noted by a GAVI report on Nigeria, remote locations are particularly susceptible to issues arising from lack of effective communication and inventory management decisions at the district and IHC levels of the supply chain (GAVI (2014)). Shen *et al.* (2014) and Zaffran *et al.* (2013) state that establishing data-informed inventory management solutions to vaccine supply chains is essential to providing coverage, especially as increases in demand due to rising populations and the introduction of new vaccines strain the system. Yet, maintaining rigorous data records is already a challenging task for IHC workers, and as many audits show, reliable data often does not propagate back up the supply chain to inform inventory policy (see, e.g. Wagenaar *et al.* (2015), Bosch-Capblanch *et al.* (2009), Chilundo *et al.* (2004), Lim *et al.* (2008), Murray *et al.* (2003) and the many references within). These studies have found the quality of data lacking, and often prescribe additional supervision to improve record-keeping practices. Therefore, any policy approach that requires frequent, intricate data records while simultaneously determining inventory decisions based off of this data from the IHC level would be (1) extremely difficult to implement (due to the higher levels of effort and training), and (2) subject to many human recording errors.

In our goal to protect against such parameter uncertainty while maintaining implementable policies, the problem we tackle is a combination of periodic inventory control within a network under demand uncertainty. In this vein, there is a wealth of literature that uses robust optimization to aid in determining effective policies in the presence of demand ambiguities. Among these, Bertsimas and Thiele (2006) incorporate robustness against parameter uncertainty in a tractable manner by applying bounds to the forecast errors via a "Budgets of Uncertainty" via Bertsimas and Sim

(2004a). Other strategies such as those seen in Adida and Perakis (2006), Bienstock and ÖZbay (2008), See and Sim (2010), Klabjan *et al.* (2013) and the references within tackle uncertain demands in inventory control problems via fluid models, data-driven approaches, and by utilizing moment constraints.

Due to the difficulty of transferring information in developing countries, we consider an environment where demands are unobservable until the decision to order inventory. We first examine the problem with a single IHC from both a fully observed and Bayesian perspective, where uncertainty is expressed in rate of the Poisson process that guides demand. This resembles traditional Bayesian inventory control approaches like Scarf (1959) and Azoury (1985); however, we focus on delivery timings with unobserved demands between excursions.

Since our problem takes place in a network where deliveries must be made to several IHCs with varying characteristics, our problem takes the form of a specialized Inventory Routing Problem (IRP). This class of problems, first studied by Bell *et al.* (1983), minimize transportation and inventory costs involved with the distribution of products to a network of demand sites. With the exception of few recent papers such as Grønhaug *et al.* (2010), Engineer *et al.* (2012), and Uggen *et al.* (2013), as Coelho *et al.* (2013) note in their literature review, most IRP studies assume fixed, deterministic consumption rates. In our problem, we not only assume stochastic consumption, but also consider consumption under parameter uncertainty via a Bayesian approach.

IRPs are well-known to be difficult to solve exactly (see, e.g. Coelho *et al.* (2013), Andersson *et al.* (2010) and the references within), and as such many realistic implementations rely on heuristics or solutions based on simplified problem instances. The literature is filled with a large variety of such strategies including those based on metaheuristics, column generation, and clustering strategies. Following the IRP literature, we provide integer programming (IP) formulations with simplified policy spaces that can be solved in small to medium problem instances. We also employ a cluster-type heuristic algorithm which can be used on larger problem instances. Unlike most of the IRP literature, we identify an easily calculable lower bound converges asymptotically when populations become large.

4.3 Problem Description

As opposed to the traditional approach where IHCs fetch their vaccines from the district level, we investigate the case where the district level delivers to IHCs directly. We consider a network where the district depot delivers to n IHCs at periodic time intervals. Since vaccine inventory can be expressed in terms of the total amount of inventory necessary to immunize a single child rather than determining the levels of each type of vaccine individually (see, e.g., UNICEF, WHO (2012)) and since IHCs utilize order-up-to policies (see, e.g., Rajgopal *et al.* (2011)), we assume that order-up-to levels have been determined for each IHC, so that whenever the district delivers vaccines to IHC $i \in \{1, \ldots, n\} = \mathcal{N}$, they ensure q_i FIC units of vaccines are in stock at the beginning of the period. In developing countries, storage capacity at the IHC level is often highly limited, hence in these settings, q_i can simply be the available refrigeration volume for vaccines. Following the literature like Brown *et al.* (2014), we assume that demand for each IHC $i \in \mathcal{N}$ occurs according to independent Poisson processes with mean λ_i per period.

Supplying the network of IHCs requires the use of vehicles and other resources due to fuel, maintenance, and labor. Therefore, we assign a fixed cost k per route in addition to a travel cost $d_{i,j} > 0$ whenever an excursion travels from node i to j for $i, j \in \mathcal{N} \bigcup \{0\}$. Here, index 0 corresponds to the district depot and we assume that the network experiences symmetric costs so that $d_{i,j} = d_{j,i}$ for all $i, j \in \mathcal{N} \bigcup \{0\}$. Furthermore, due to (1) the time spent unloading and recording inventory levels and

(2) limited transportation capacity, we assume that the district can fill at most $m \leq n$ IHCs in a given excursion. We note that this is a simplification of the real system; in reality, if few vaccines are distributed in an excursion due to low demand realizations or order-up-to levels, there is the potential to serve additional IHCs in the same trip. However, in setting m constant for all excursions, we establish a baseline that can protect against adverse scenarios where all of the transportation capacity is necessary to serve the IHCs in a route. Hence such policies guarantee a service level even when they are modified to incorporate additional visits based on the inventory levels in a given trip. [1]

The decision-maker seeks to minimize the costs of underages and transportation, thus we denote the cost of a missed opportunity due to inventory outages be set to υ. Since communication is limited between levels of the supply chain, we assume that previous realizations of demands are not observed until the district delivers vaccines to the IHC. Letting $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_n) \in \mathbb{Z}_+^n$ be the vector of the number of periods since each IHC was last visited, a policy can be expressed as a mapping that takes the current period and number of periods since last delivery to each IHC and returns the IHCs that will be visited on the given period. Hence, we let a policy be a function $\pi : \mathbb{Z}_+ \times \mathbb{Z}_+^n \to \mathbb{B}^n$ so that $\pi(t, \boldsymbol{\tau}) = \mathbf{a} = (a_1, \ldots, a_n)$, where $a_i = 1$ if IHC i is visited on period t under policy π and $a_i^\pi = 0$ otherwise and let Π be the set of all policies.

To form optimal routes based on the IHCs visited in a given period, $\mathbf{a}$, we let $\mathcal{O}(\mathbf{a}) = \{i \in \mathcal{N} | a_i > 0\}$, and define $\psi : \mathbb{B}^n \to \mathbb{R}$ as the optimal cost to the constrained multiple Traveling Salesman Problem (mTSP) with fixed cost k per salesman for nodes $i \in \mathcal{N}$ with $a_i > 0$ which can be solved via the following Mixed Integer Program

[1] The restriction to visiting m IHCs can be extended to reflect the transportation capacity demand rate at each facility, as discussed in Online Appendix C.2.

(MIP).

$$
\begin{aligned}
\psi(\mathbf{a}) = \min \quad & kr + \sum_{0 \le i \le n} \sum_{\substack{0 \le j \le n \\ j \ne i}} d_{i,j} u_{i,j} && (4.1) \\
\text{s.t.} \quad & \sum_{j \in \mathcal{O}(\mathbf{a})} (u_{0,j} + u_{j,0}) = 2r, \\
& \sum_{\substack{0 \le i \le n \\ i \ne j}} u_{i,j} = 1, & j \in \mathcal{O}(\mathbf{a}) \\
& \sum_{\substack{0 \le i \le n \\ j \ne i}} u_{i,j} = 1, & i \in \mathcal{O}(\mathbf{a}) \\
& b_i - b_j + m u_{i,j} \le m - 1, & i, j \in \mathcal{O}(\mathbf{a}), i \ne j, \\
& r \in \mathbb{N}, u_{i,j} \in \mathbb{B}, b_i \in \mathcal{N}
\end{aligned}
$$

Here, $u_{i,j}$ are the usual TSP variables that are 1 if the edge i to j is used in a tour and 0 otherwise, b_i represent the node potentials in the network, and r gives number of tours scheduled (see, e.g. Miller *et al.* (1960) and Bektas (2006)).

Now, letting $\boldsymbol{\tau}_t^\pi = (\tau_{1,t}^\pi, \ldots, \tau_{n,t}^\pi)$ be the number of periods since each IHC was last visited and $\mathbf{a}_t^\pi = (a_{1,t}^\pi, \ldots, a_{n,t}^\pi)$ be the action under policy π at time t, and letting X_t denote the demand at time t, the cost under policy π at time t can be expressed

$$
Z_t^\pi = \begin{cases}
\psi(\mathbf{a}_t^\pi) + \upsilon \sum_{i=1}^n \left(X_{i,t} - \left(q_i - \mathbb{1}\{\tau_{i,t}^\pi \ne 1\} \sum_{\hat{t}=t-\tau_{i,t}^\pi}^{t-1} X_{i,\hat{t}} \right)^+ \right)^+ & t < T, \\
\psi(\mathbf{1}_n) + \upsilon \sum_{i=1}^n \left(X_{i,t} - \left(q_i - \mathbb{1}\{\tau_{i,t}^\pi \ne 1\} \sum_{\hat{t}=t-\tau_{i,t}^\pi}^{t-1} X_{i,\hat{t}} \right)^+ \right)^+ & t = T, \\
0 & \text{Otherwise,}
\end{cases}
\tag{4.2}
$$

Hence, given an initial $\boldsymbol{\tau}_1$ (which is not affected by π) our goal is to find the policy that minimizes the objective:

$$
\operatorname*{arg\,min}_{\pi \in \Pi} \mathrm{E} \left[\frac{\sum_{t=1}^T Z_t^\pi}{T} \right]. \tag{4.3}
$$

We refer to the case when $T \to \infty$ as the infinite horizon average cost case.

However, in developing countries, populations served by IHCs are not well understood, hence the rate of demand can be uncertain. Therefore, instead of assuming a fully observed parameters λ_i, we approach the problem from a Bayesian perspective. Assuming the initial prior for each IHC's λ_i is gamma with parameters $\alpha_{i,0}$ and $\beta_{i,0}$, a sufficient statistic from the observations up to time t is given by $\alpha_{i,0} + \sum_{i=1}^{t} X_i$ and $t + \beta_{i,0}$. Letting $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)$ and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_n)$, the policies for the Bayesian case are mappings $\pi : \mathbb{Z}_+ \times \mathbb{Z}_+^n \times \mathbb{R}_+^n \times \mathbb{R}_+^n \to \mathbb{Z}_+$ which take the prior parameters along with $\boldsymbol{\tau}$ and the current period t, and return the IHCs that are served in a given period. Hence, in the Bayesian case, $\pi(t, \boldsymbol{\tau}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \mathbf{a}$, which represents the IHCs that are visited at time t under policy π. In a nearly identical manner to the case where λ is fully observed, we let $\hat{\Pi}$ denote the set of such policies and express the cost under policy π as

$$\hat{Z}_t^\pi = \begin{cases} \psi(\mathbf{a}_t^\pi) + \upsilon \sum_{i=1}^n \left(X_{i,t} - \left(q_i - \mathbb{1}\{\tau_{i,t}^\pi \neq 1\} \sum_{\hat{t}=t-\tau_{i,t}^\pi}^{t-1} X_{i,\hat{t}} \right)^+ \right)^+ & t < T, \\ \psi(\mathbf{1}_n) + \upsilon \sum_{i=1}^n \left(X_{i,t} - \left(q_i - \mathbb{1}\{\tau_{i,t}^\pi \neq 1\} \sum_{\hat{t}=t-\tau_{i,t}^\pi}^{t-1} X_{i,\hat{t}} \right)^+ \right)^+ & t = T, \\ 0 & \text{Otherwise,} \end{cases} \tag{4.4}$$

Then, given $\boldsymbol{\tau}_1$ and initial priors $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$, the Bayesian objective can be expressed:

$$\min_{\pi \in \hat{\Pi}} \mathrm{E}_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0} \left[\sum_{t=1}^{T} \hat{Z}_t^\pi \right], \tag{4.5}$$

where prior parameters are updated at time t in accordance with $\alpha_{i,t} = \alpha_{i,0} + \sum_{i=1}^{t} X_i$ and $t + \beta_{i,0}$ if $a_{i,t}^\pi = 1$ since demand observations only occur when IHCs are visited.

4.4 Fully-Observed Case

Both (4.3) and (4.5) can be solved via dynamic programs. First focusing on (4.3), we define the cost function

$$c_i(\lambda) = \begin{cases} \upsilon\left(\lambda(1 - F(q_i - 1, \lambda)) - q_i(1 - F(q_i, \lambda))\right) & \lambda > 0, \\ 0 & \lambda = 0, \end{cases} \tag{4.6}$$

for each $i \in \mathcal{N}$, where F is the Poisson CDF with parameter λ which is the expected number of underages resulting from a single period of Poisson demand with mean λ when the order-up-to level is given by q_i. Then, we can express (4.3) in terms of a dynamic program which operates backwards in time:

$$\mathrm{V}_t(\boldsymbol{\tau}) = \begin{cases} \min_{\mathbf{a} \in \mathbb{B}^n} \psi(\mathbf{a}) + \sum_{i=1}^n c_i(\lambda_i \tau_i) - c_i(\lambda_i(1 - \tau_i) & \\ \qquad\qquad + \mathrm{V}_{t-1}(((1 - a_1)\tau_1 + 1, ..., (1 - a_n)\tau_n + 1)) & t > 0, \\ \psi(\mathbf{1}_n) + \sum_{i=1}^n c_i(\lambda_i \tau_i) - c_i(\lambda_i(\tau_i - 1)) & t = 0, \end{cases} \tag{4.7}$$

Here, $\boldsymbol{\tau}$ still consists of the number of periods since the last delivery to each IHC, and act as the system state at time t. If the decision-maker chooses action $a_i = 1$, then cost $c_i(\lambda_i \tau_i)$ is collected and the next period's τ_i transitions to 1. Otherwise, if $a_i = 0$, the system delays delivery, and τ_i transitions to $\tau_i + 1$. If the initial state is $\boldsymbol{\tau}$, $\sum_{t=1}^T Z_t^\pi = \mathrm{V}_t(\boldsymbol{\tau})$, hence, the infinite horizon case of (4.7) can be calculated as the limit of the dynamic program $\lim_{T \to \infty} \mathrm{V}_T(\boldsymbol{\tau})/T$.

The problem (4.7) is highly computationally complex due to its n-dimensional state and action spaces, where the action space is further demanding as a result of the service constraints on each trip. Even given an optimal action $\mathbf{a}$, simply determining $\psi(\mathbf{a})$ is well-known to be an NP-Hard problem, and the naive MIP formulation of (4.7) features far too many constraints and variables to be tractable. However, using the structural results of the case when $n = 1$, we can determine a lower bound to

(4.7).

When $n = 1$, the (4.7) takes on a highly simplistic form since the timing of routes is only based on a single IHC's inventory level and ψ only takes on values $2d_{1,0}$ in the case that $a_1 = 1$ and 0 otherwise. Furthermore, since no parameter learning occurs, intuitively, there is no advantage to altering the number of periods between deliveries over different periods. Since such a policy maintains an identical delivery pattern, we say π is a fixed-τ policy if it designates τ periods between every delivery. As our intuition suggests, fixed-τ policies are optimal to the fully observed objective. This allows us to easily determine the optimal policy, as well as evaluate (4.7) via the following proposition:

Proposition 4.1 (**Optimal Fully-Observed Policy**). *In the case where* $n = 1$ *and* $k + 2d_{1,0} < vq_1$*, letting* $\hat{\tau} = \operatorname{argmin}_{\tau \geq 1} \frac{1}{\tau}(k + 2d_{1,0} + c_1(\tau\lambda_1))$*,*

(i) *The optimal policy to the infinite horizon average cost case of* (4.3) *is fixed-*τ *with ordering period* $\hat{\tau}$ *with average cost* $\frac{1}{\hat{\tau}}(k + 2d_{1,0} + c_1(\hat{\tau}\lambda_1))$*.*

(ii) *In the finite-horizon problem, if* $T + 1 \bmod \hat{\tau} = 0$*, the fixed-*$\tau$ *policy with* $\hat{\tau}$ *is optimal with cost* $\mathrm{V}_T(1) = \frac{T+1}{\hat{\tau}}(k + 2d_{1,0} + c_1(\hat{\tau}\lambda_1))$*.*

(iii) *In any finite-horizon problem with* T*,* $\mathrm{V}_T(1) \leq \frac{T+1}{\hat{\tau}}(k + 2d_{1,0} + c_1(\hat{\tau}\lambda_1))$*.*

This reinforces the notion that the fully-observed case minimizes the cost by minimizing the average cost between deliveries. When IHCs must be served on an individual basis due to remote locations or transportation capacity restrictions, a simple routine delivery policy should be implemented if the demand rate is established. In finite horizon settings, when $T+1 \bmod \hat{\tau} = 0$, an average cost of $\frac{1}{\tau}(k + 2d_{1,0} + c_1(\tau\lambda_1))$ per period can be achieved by delivering every $\hat{\tau}$ periods, whereas when $T + 1 \bmod \hat{\tau} \neq 0$ the terminating conditions of (4.7) imply that at least some delivery must violate the fixed-τ policy associated with $\hat{\tau}$. However, the performance of the fixed-τ policy (with the exception of the final delivery) in these cases is very near optimal performance, since the average cost for all periods except those associated with the final delivery can be shown to result in lower cost than all other policies.

We specify that $k + 2d_{1,0} < \upsilon q_1$ since otherwise, the cost of delivering the vaccines will outweigh any mitigated underage costs which induces an optimal strategy of zero deliveries. In practice, since the leading metric of vaccine supply chains is immunization coverage, the cost of missed opportunity is large in comparison to transportation costs, hence this condition is naturally satisfied in any realistic setting. Hence, so long as the costs of delivery are not prohibitive, a manager that has fully characterized demand should strive to engage in equally spaced deliveries. Furthermore, since c_1 is an increasing function of λ_1, which induces a smaller $\hat{\tau}$, supply chains will naturally experience higher costs and more deliveries when demands increase.

Using the structural results of the single-IHC case, we can determine an easily calculable lower bound to (4.7).

Proposition 4.2 (**Fully-Observed Lower Bounds**). *If $k + 2d_{i,0} < \upsilon q_i$ for all $i \in \mathcal{N}$, the following lower bounds to* (4.7) *(and hence* (4.3)*) hold:*

$$\sum_{i=1}^{n} \min_{\tau \geq 1} \frac{1}{\tau} \left(\frac{k + 2d_{0,i}}{m} + c_i(\tau\lambda_i) \right) \leq \frac{\mathrm{V}_T(\mathbf{1}_n)}{T+1} \leq \lim_{T \to \infty} \frac{\mathrm{V}_T(\mathbf{1}_n)}{T}. \tag{4.8}$$

In addition to revealing a lower bound for both finite and infinite horizon cases, Proposition 4.2 enables a sufficient condition for optimality of (4.7) when the distance costs go to zero. Referring to routes that feature many IHC visits as "dense" and letting a "fully dense" route describe a route that visits exactly m IHCs, when $d_{i,j} = 0$, if a policy exists that consists of only fully dense routes, where each IHC is visited every $\arg\min_{\tau \geq 1} \frac{1}{\tau} \left(\frac{k}{m} + c_i(\tau\lambda_i) \right)$ periods, this policy experiences a cost identical to the lower bound of Proposition 4.2 and hence is optimal to (4.7). Even for cases where such a policy is not possible, and distance costs are small, (4.8) suggests that optimal policies to (4.7) strive to (1) enable fully dense trips and (2) target trips near $\arg\min_{\tau\geq 1}\frac{1}{\tau}\left(\frac{k+2d_{i,0}}{m}+c_i(\tau\lambda_i)\right)$ for each IHC. Hence, these lower bounds naturally improve when n increases, m decreases, and when IHCs with similar λ_i and q_i are tightly clustered, since these conditions become easier to satisfy. Thus, a manager is incentivized to serve IHCs with similar characteristics if they lie near one another in regular intervals in accordance with $\hat{\tau}$ of Proposition 4.1.

Though the lower bounds (4.8) can be near the optimal cost in the above settings, they can be improved under some assumptions concerning the routes in the optimal policy. If it is assumed that each route includes at least $\hat{m}$ IHCs, letting

$$\hat{\psi}(i,\hat{m}) = \min_{\mathbf{a}\in\mathbb{B}^n}\left\{\psi(\mathbf{a})\middle| r=1, a_i>0, \sum_{j\in\mathcal{N}} a_j \geq \hat{m}\right\} \tag{4.9}$$

denote the transportation cost of the smallest route of $\hat{m}$ IHCs that includes IHC i, the lower bounds (4.8) can be improved to

$$\sum_{i=1}^{n}\min_{\tau\geq 1}\frac{1}{\tau}\left(\frac{\hat{\psi}(i,\hat{m})}{m}+c_i(\tau\lambda_i)\right) \leq \frac{\mathrm{V}_T(\mathbf{1}_n)}{T+1} \leq \lim_{T\to\infty}\frac{\mathrm{V}_T(\mathbf{1}_n)}{T} \tag{4.10}$$

by substituting the transportation cost $k+2d_{i,0}$ with $\psi(i,\hat{m})$. Since inventory routing problems usually are composed of dense routes, and since the capacity restrictions via m further incentivize route density in our problem, (4.10) provides a means of obtaining more realistic bounds in cases where n is small, m is large, and the network is more sparse, which can worsen the guaranteed lower bounds (4.8).

4.5 Bayesian Case

We turn our attention to the case where the demand rates are not fully observed, (4.5), which can also be expressed via a dynamic program. Similar to the fully-observed case, we define cost functions for the Bayesian case

$$\hat{c}_i(\tau,\alpha,\beta) = \begin{cases} \frac{v}{\beta}\left((q_i+\tau\alpha)h(q_i,\tau\alpha,\beta)+(q_i\beta-\alpha\tau)(-1+H(q_i,\tau\alpha,\beta))\right), & \tau>0, \\ 0, & \tau=0. \end{cases} \tag{4.11}$$

Here, H and h are distribution functions of the negative binomial distribution

$$h(y,\alpha,\beta) = \left(\frac{1}{\beta+1}\right)^y\left(1-\frac{1}{\beta+1}\right)^\alpha\binom{y+\alpha-1}{\alpha-1}. \tag{4.12}$$

Cost function $\hat{c}_i$ gives the expected cost due to missed opportunities when the time since last service is τ and prior parameters for demand are the pair α and β for all $i\in\mathcal{N}$. The Bayesian objective (4.5) can be solved via the dynamic program

$$\begin{aligned}\hat{\mathrm{V}}_t(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\tau}) = \min_{\mathbf{a}\in\mathbb{B}^n}\psi(\mathbf{a}) &+ \sum_{i=1}^n \hat{c}_i(\tau_i,\alpha_i,\beta_i)-\hat{c}_i(\tau_i-1,\alpha_i,\beta_i) \\ &+ \mathrm{E}_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\tau}}\big[\hat{\mathrm{V}}_{t-1}((\alpha_1+\sum_{\hat{t}=t-\tau_1}^{t}X_{1,t},...,\alpha_n+\sum_{\hat{t}=t-\tau_n}^{t}X_{n,t}), \\ &\qquad (\beta_1+a_1\tau_1,...,\beta_n+a_n\tau_n), \\ &\qquad ((1-a_1)\tau_1+1,...,(1-a_n)\tau_n+1))\big]\end{aligned} \tag{4.13}$$

if $t>0$, and terminating state

$$\hat{\mathrm{V}}_t(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\tau}) = \psi(\mathbf{1}_n)+\sum_{i=1}^n \hat{c}_i(\tau_i,\alpha_i,\beta_i)-\hat{c}_i(\tau_i-1,\alpha_i,\beta_i) \tag{4.14}$$

if $t=0$. However, given the high level of complexity in the fully-observed problem (4.7), the Bayesian program (4.13) is hopelessly complex to solve in general, even over small time horizons and small n. Fortunately, (4.13) admits a lower bound, similar to (4.8).

To accomplish this, we again turn to the case when $n=1$, where $\hat{\mathrm{V}}_t(1)$ can be expressed

$$\hat{\mathrm{V}}_t(\alpha,\beta) = \min_{\tau\geq 1}\begin{cases} k + 2d_{1,0} + \hat{c}_1(\tau,\alpha,\beta) & \\ + \sum_{y=0}^{\infty}\hat{\mathrm{V}}_{t-\tau}(\alpha+y,\beta+\tau)h(y,\tau\alpha,\beta) & t-\tau > 0 \\ k + 2d_{1,0} + \hat{c}_1(\tau,\alpha,\beta) & t-\tau \leq 0. \end{cases} \tag{4.15}$$

Unlike (4.13), actions between deliveries are by definition delayed actions, hence these waiting periods can be immediately woven into the dynamic program and we drop τ from the state space (see Lemma C.1 in Online Appendix C.3). Inspecting (4.15) reveals that $\beta - t$ remains constant throughout the dynamic program. This implies that $\hat{\mathrm{V}}_t(\alpha,\beta)$ can be viewed as a two-dimensional dynamic program by linking β and t. Thus, to solve for period t, it suffices to solve for the $t-1$ periods of single-dimensional components of the dynamic program.

Since $\hat{\mathrm{V}}_t$ is the Bayesian counterpart to V_t, the two dynamic programs feature many connections. Therefore, letting $\hat{\mathrm{V}}_t^{\pi}(\alpha,\beta)$ to be the evaluation of policy π on (4.15), we make the following proposition.

Proposition 4.3 (**Learning Objective Characterization**). *The following relations between fully observed and learning objectives hold:*

(*i*) *For any fixed-τ policy π, $\frac{t+1}{\hat{\tau}}(k + 2d_{1,0} + c_1(\hat{\tau}\alpha/\beta)) \leq \hat{\mathrm{V}}_t^{\pi}(\alpha,\beta)$,*

(*ii*) $\lim_{\beta\to\infty}\hat{\mathrm{V}}_t(\lambda_1\beta,\beta) = \mathrm{V}_t(\lambda_1)$,

(*iii*) $\min_{\tau\in\mathbb{Z}^+}\mathrm{E}_g\left[\frac{t+1}{\tau}(k + 2d_{1,0} + c_1(\tau\lambda))\right] \leq$
$\hat{\mathrm{V}}_t(\alpha,\beta) \leq \min_{\tau\in\mathbb{Z}^+}\frac{t+1}{\tau}(k + 2d_{1,0} + \hat{c}_1(\tau,\alpha,\beta))$.

Proposition 4.3 implies that the non-learning and Bayesian objectives are highly related. Result (*i*) implies that fully observed demand parameters are preferable to any parameter uncertainty when the decision-maker is restricted to fixed-τ policies.

This is because when the Bayesian objective is restricted to non-learning policies, the convex properties of $\hat{c}$ in α allows us to invoke Jensen's inequality. However, in general we do not have $\frac{t+1}{\hat{\tau}}(k + 2d_{1,0} + c_1(\hat{\tau}\alpha/\beta)) \leq \hat{\mathrm{V}}_t(\alpha, \beta)$ since $\hat{\mathrm{V}}_t(\alpha, \beta)$ is non-convex in α and β.

Since β acts roughly as the level of information, (*ii*) shows that as the level of information increases, the Bayesian case converges to the fully observed case, hence $\hat{\mathrm{V}}_t(\alpha, \beta) \approx \frac{t+1}{\hat{\tau}}(k + 2d_{1,0} + c_1(\hat{\tau}\alpha/\beta))$ for large β. This intuitive result reinforces the fact that information is the link between objectives and further shows that as T becomes large, policies in the Bayesian case become fixed-τ.

Further linking non-learning and Bayesian cases, (*iii*) shows that the Bayesian objective can be bounded from below via the non-robust objective, and bounded above via $\hat{c}$. Defining the probability density function

$$\hat{g}(\lambda, \alpha, \beta, \tau) = \frac{\lambda^{\alpha-1}\left(\frac{\beta}{\beta+1}\right)^{\alpha\tau}(\beta+\tau)^{\alpha}e^{-\lambda(\beta+\tau)}\,{}_1F_1\left(\alpha\tau; \alpha; \frac{\lambda(\beta+\tau)}{\beta+1}\right)}{\Gamma(\alpha)},$$

where ${}_1F_1$ is the hypergeometric function, we can use these bounds to generate sufficient conditions for optimality criteria via the following corollary.

Corollary 4.1 (**Sufficient Optimality Condition**). *If there exists $0 < \tau < t$ such that*

$$\frac{t+1}{\tau}(k + 2d_{1,0} + \hat{c}_1(\tau, \alpha, \beta)) \leq \begin{cases} \mathrm{E}_{\hat{g}}\left[\frac{t+1}{\tau}(k + 2d_{1,0} + c_1(\hat{\tau}\lambda))\right] & \hat{\tau} < t \\ k + 2d_{1,0} + \hat{c}_1(\hat{\tau}, \alpha, \beta) & \hat{\tau} \geq t \end{cases} \qquad (4.16)$$

for every $\hat{\tau} \neq \tau$, then the optimal delivery waiting period for $\hat{\mathrm{V}}_t(\alpha, \beta)$ is τ.

By substituting the upper bound as the valuation for policies under action τ, and substituting the lower bound in the cost-to-go for policies under all other actions $\hat{\tau}$, if action τ is still preferable to the decision-maker, τ must represent an optimal action.

Since Corollary 4.1 relies on fixed-τ policies, the conditions necessary to declare τ an optimal action are only satisfied for sufficiently small T since the potential for learning policies to reduce costs can be made arbitrarily large with T. However, the following proposition shows that optimal actions in the Bayesian case can be easily bounded.

Proposition 4.4 (**Delivery Bounds**). *If $k + 2d_{1,0} \leq \hat{c}(2\tau, \alpha, \beta) - 2\hat{c}(\tau, \alpha, \beta)$, then the optimal delivery waiting period to $\hat{\mathrm{V}}_t(\alpha, \beta)$ is less than 2τ.*

The proposition is established intuitively; if two deliveries can be accomplished with smaller expected value than a single delivery over the same number of periods, there is no incentive for the decision-maker to deliver with any larger gap since the information state can remain the same after two orders. Naturally this implies that a manager can expect to engage in faster deliveries as the expected demand rate, α/β, increases.

The results on the single-IHC case helps to form a lower bound to (4.13).

Proposition 4.5 (**Bayesian Lower Bounds**). *Letting $\hat{\mathrm{V}}'_{T,i}(\alpha_i, \beta_i)$ refer to* (4.15) *with fixed and distance costs k/m and $2d_{0,i}/m$ and underage cost function $\hat{c}_i$,* (4.13) *has the lower bound*

$$\sum_{i=1}^{n} \hat{\mathrm{V}}'_{T,i}(\alpha_i, \beta_i) \leq \hat{\mathrm{V}}_T(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{1}_n). \tag{4.17}$$

Like the lower bound (4.8) in the fully observed case, (4.17) results from the fact that a system only capable of serving one IHC per trip with fixed and travel costs of k/m and $2d_{i,0}/m$ results in a lower cost than a system with fixed and travel costs of k and $d_{i,j}$ that can serve m IHCs each trip since the former is essentially a less restricted version of the latter. Unlike Proposition 4.2, even when distance costs go to zero, the lower bound (4.17) are only realizable if there exists fully dense trips that satisfy the waiting periods implied by the dynamic program $\hat{\mathrm{V}}'_{t,i}(\alpha_i, \beta_i)$ at each period. Since the prior parameters naturally deviate based on observations, such a scenario is impossible to guarantee, hence, in general, only trivial settings can result in achievable lower bounds. However, the lower bounds still provide a baseline from which to compare the performance of suboptimal policies that can be calculated by evaluating n instances of (4.15).

Similar to (4.10), these lower bounds can be improved by assuming that the optimal policy has route density such that every route visits at least $\hat{m}$ IHCs. Then, letting $\hat{\mathrm{V}}''_{T,i}(\alpha_i, \beta_i)$ denote a modified version of $\hat{\mathrm{V}}'_{T,i}(\alpha_i, \beta_i)$ where term $(k + 2d_{i,0})/m$ is replaced with with $(k + \hat{\psi}(i, \hat{m}))/m$, we gain the improved bounds

$$\sum_{i=1}^{n} \hat{\mathrm{V}}''_{T}(\alpha_i, \beta_i) \leq \hat{\mathrm{V}}_T(\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{1}_n). \tag{4.18}$$

4.6 Fixed-τ Policies Reduction

Since both (4.3) and (4.5) become highly intractable, we focus on a reduced policy space which only use variations on fixed-τ policies. Then, the problems can be expressed as more easily solvable MIPs than the original dynamic programming formulations (4.7) and (4.13).

Therefore, since the fully observed single-IHC problem is solved via fixed-τ policies and the lower bounds (4.8) suggest that such policies also have a strong connection to (4.7), we consider the infinite-horizon average case where a manager is restricted to utilizing only fixed-τ policies. Thus, defining $\bar{\tau}$ as the maximal allowable number of period between deliveries for any IHC in the network and letting $T_{\bar{\tau}}$ be the least common multiple of $\{1, ..., \bar{\tau}\}$, the infinite-horizon fully observed problem restricted to using only fixed-τ policies can be solved via the MIP (4.19).

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{n}\sum_{\tau=1}^{\bar\tau}\sum_{j=0}^{\tau-1} y_{i,\tau,j} c_i(\tau\lambda_i)/\tau + \frac{1}{T_{\bar\tau}}\bigg(\sum_{t=1}^{T_{\bar\tau}}\big(kr_t + \sum_{i=0}^{n}\sum_{j=0}^{n} d_{i,j}u_{i,j,t}\big)\bigg) && (4.19)\\
\text{s.t.} \quad & \sum_{\tau=1}^{\bar\tau}\sum_{j=0}^{\tau-1} y_{i,\tau,j} = 1, && i = 1, ..., n\\
& \sum_{j=1}^{n}(u_{0,j,t} + u_{j,0,t}) = 2r_t, && t = 1, ..., T_{\bar\tau}\\
& \sum_{i \neq j} u_{i,j,t} = \sum_{\tau=1}^{\bar\tau} y_{j,\tau,t \bmod \tau}, && j = 1, ..., n, t = 1, ..., T_{\bar\tau},\\
& \sum_{j \neq i} u_{i,j,t} = \sum_{\tau=1}^{\bar\tau} y_{i,\tau,t \bmod \tau}, && i = 1, ..., n, t = 1, ..., T_{\bar\tau},\\
& b_{i,t} - b_{j,t} + mu_{i,j,t} \leq m - 1, && i, j \in \mathcal{N}, i \neq j, t = 1, ..., T_{\bar\tau}\\
& r_t \in \mathbb{N}, y_{i,\tau,j} \in \mathbb{B}, u_{i,j,t} \in \mathbb{B}, b_{i,t} \in \mathcal{N}
\end{aligned}
$$

Program (4.19) takes advantage of the fact that fixed-τ policies feature a natural periodicity every $T_{\bar\tau}$ periods, hence it suffices to consider only these $T_{\bar\tau}$ periods to solve the infinite horizon problem. Similar to (4.1), in (4.19), $u_{i,j,t} = 1$ when the edge from i to j is used at time t and is 0 otherwise, $y_{i,\tau,j} = 1$ indicates that IHC i is served every τ periods with first service occurring on period $j + 1$, and r_t represents the total number of trips made on period t. These decision variables, along with the fact that (4.19) only requires $T_{\bar\tau}$ periods, account for why the MIP formulation is much simpler than the dynamic program (4.7); policies for each IHC no longer deviate throughout time, but all the information for a policy at an IHC is held within $y_{i,\tau,j}$, which describes the time between individual services paired with its starting point.

Interestingly, when the travel costs go to zero (i.e., $d_{i,j} = 0$), so long as there exists fully dense trips that facilitate deliveries every $\arg\min_{\tau\geq 1} \frac{1}{\tau}\left(\frac{k}{m} + c_i(\tau\lambda_i)\right)$ periods for each IHC, (4.19) will still achieve the lower bound of Proposition 4.2, and hence will result in an optimal policy. Even when travel costs are non-zero, (4.19) can still approach this lower bound. If IHCs can be grouped into disjoint subsets $\mathcal{O}_1, \mathcal{O}_2, \ldots,$ so that $d_{i,j}$ is small pairwise within each subset, then $d_{i,0} \approx d_{j,0}$ for each $i, j \in \mathcal{O}_l$. Hence, treating each subset as an instance of (4.19), if there exists fully utilized trips that deliver every $\arg\min_{\tau \geq 1} \frac{1}{\tau}\left(\frac{k+2d_{i,0}}{m} + c_i(\tau\lambda_i)\right)$ periods for each IHC $i \in \mathcal{O}_l$ for each disjoint subset, (4.19) will approximately yield the lower bounds of Proposition 4.2. Even when such a policy is not achievable for all IHCs, if most can be served in this manner, performance can still be near the lower bounds, hence, in networks that feature natural clustering, (4.19) can result in a close upper bound to (4.7).

Though (4.19) is a simplification of (4.7), it can still be a challenging problem when $\bar{\tau}$ and n is large. This is because even when all fixed-τ policies have been determined (i.e., all $y_{i,\tau,j}$ are known), the routing constraints (see, e.g., Miller *et al.* (1960)) form a series of $T_{\bar{\tau}}$ mTSPs which are known to be even more challenging than the traditional TSP. Furthermore, since $T_{\bar{\tau}}$ is the least common multiple of all whole numbers up to $\bar{\tau}$, $T_{\bar{\tau}}$ grows exponentially, hence the number of mTSPs that must be embedded into (4.19) increases exponentially in $\bar{\tau}$. Each embedded mTSP sees quadratic increases in decision variables in n, thus the problem can obviously still become highly intractable. District depots usually supply fewer than $20 - 25$ IHCs, and if each period represents the time interval of a week, $\bar{\tau}$ should be no larger than 6 (see, e.g., Assi *et al.* (2013), Brown *et al.* (2014)). Though these smaller settings are not out of reach, they are still time consuming, requiring many hours of computational effort. Thus, in both smaller settings and in cases where n and $\bar{\tau}$ are large, we provide a scalable heuristic in Section 4.7 based on the structural results of (4.19).

A similar strategy can be employed to incorporate learning via a tractable Bayesian approach by modifying the policy space to feature fixed-τ policies. In a T period problem, consider the restricted policy space where (1) all IHCs are required to be visited on periods $T_1, T_2, \dots$, where $T_i < T_{i+1}$ and (2) each IHC must be served according to a fixed-τ strategy between periods T_i and T_{i+1}. Hence, like (4.19), we aim to reduce the complexity of (4.13) by enforcing fixed-τ strategies, but allow the decision-maker to dynamically update their strategies between mandatory delivery periods. The optimal policy for periods $t \in \{1, \dots, T_1\}$ can be solved via the MIP (4.20).

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{n} \sum_{\tau=1}^{\bar{\tau}} \sum_{j=0}^{\tau-1} y_{i,\tau,j} \Bigg(\hat{c}_i(j+1, \alpha_i, \beta_i) + \tau \lfloor \tfrac{T_1-(j+1)}{\tau} \rfloor \hat{c}_i(\tau, \alpha_i, \beta_i) && (4.20) \\
& + \hat{c}_i \left(T_1 - \tau \lfloor \tfrac{T_1-(j+1)}{\tau} \rfloor - (j+1), \alpha_i, \beta_i \right) \Bigg) \\
& + \psi(\mathbf{1}_n) + \sum_{t=1}^{T_1-1} \big(k r_t + \sum_{i=0}^{n} \sum_{j=0}^{n} d_{i,j} u_{i,j,t} \big) \\
\text{s.t.} \quad & \sum_{\tau=1}^{\bar{\tau}} \sum_{j=0}^{\tau-1} y_{i,\tau,j} = 1, && i = 1, ..., n \\
& \sum_{j=1}^{n} (u_{0,j,t} + u_{j,0,t}) = 2 r_t, && t = 1, ..., T_1 - 1 \\
& \sum_{i \neq j} u_{i,j,t} = \sum_{\tau=1}^{\bar{\tau}} y_{j,\tau,t \bmod \tau}, && j = 1, ..., n, t = 1, ..., T_1 - 1, \\
& \sum_{j \neq i} u_{i,j,t} = \sum_{\tau=1}^{\bar{\tau}} y_{i,\tau,t \bmod \tau}, && i = 1, ..., n, t = 1, ..., T_1 - 1, \\
& b_{i,t} - b_{j,t} + m u_{i,j,t} \leq m - 1, && i, j \in \mathcal{N}, i \neq j, t = 1, ..., T_1 - 1 \\
& r_t \in \mathbb{N}, y_{i,\tau,j} \in \mathbb{B}, u_{i,j,t} \in \mathbb{B}, b_{i,t} \in \mathcal{N}
\end{aligned}
$$

MIP (4.20) can restrict its attention to the first T_1 periods since the condition that all IHCs experience deliveries at T_1 ensures an identical system state at T_1 under all feasible policies in the restricted policy space. Due to the fact that only fixed-τ strategies are employed, (4.20) is highly similar to (4.19): $y_{i,\tau,j} = 1$ if deliveries to IHC i are made every τ periods starting with period $j+1$, $u_{i,j,t}$ indicates if the edge from i to j is used on period t, and r_t represents the number of trips made at period t. However, there are distinct differences in the objective since the Bayesian problem is not infinite horizon with periodicity. Instead, underage costs are calculated in three portions: the underage cost from periods $1, ..., j+1$ is accounted for by the term $\hat{c}_i(j+1, \alpha_i, \beta_i)$, the underage costs with τ periods between trips corresponds to the $\hat{c}_i(\tau, \alpha_i, \beta_i)$ term, and underage cost from periods leading up to the required delivery at period T_1 is given by the $\hat{c}_i(T_1 - \frac{T_1 - T_1 \bmod \tau}{\tau} - (j+1), \alpha_i, \beta_i)$ term. Also, we note that travel costs accrue identically to (4.19) up to period $T_1 - 1$, at which point travel costs on period T_1 are accounted for via $\psi(\mathbf{1}_n)$.

Since it shares a comparable complexity due to the same number of decision variables and constraints as found in (4.19), (4.20) grants a computationally feasible space to express policies that exhibit parameter learning. However, in the case that n or T_1 become large, we provide a heuristic in Section 4.7 similar to those provided for the fully observed MIP (4.19). The simplifications used to attain tractability in the form of the MIP (4.20) increase the costs to (4.13) due to increased policy restrictions, yet, if T_i are chosen carefully, this approach can still be shown to yield high performance as compared to the lower bounds of Proposition (4.5). If each $T_i \bmod T_{\bar{\tau}} = 0$, the mandatory services at T_i will naturally occur (i.e., the the fixed-τ policy delivers at time T_i) for IHCs that have $y_{i,\tau,\tau-1} = 1$. Hence, the additional cost resulting from service periods $T_1, T_2, \ldots$ can be made small since this requirement is not highly restrictive when T_i is chosen appropriately. Obviously, the other main simplification is the restriction to fixed-τ policies. However, this restriction can also be managed by careful selection of T_i: if there is a large amount of uncertainty in parameters, T_i can be chosen to be smaller, resulting in a faster response to new information. Otherwise, if there is little parameter uncertainty T_i can be made larger, which results in fewer costs due to required service periods, with the trade-off of having a slower response to new information.

4.7 Class-Based Heuristic

Since solving the network case can become computationally challenging, even with the simplifications used in Section 4.6, we develop an easy-to-implement, scalable heuristics based on fixed-τ strategies. To accomplish this, our heuristics endeavor to mimic the MIP reductions (4.19) and (4.20) by arranging IHCs in groups that satisfy fully dense routes, aiming for deliveries that occur every $\arg\min_{\tau\geq 1}\frac{1}{\tau}\big(\frac{k+2d_{i,0}}{m}+c_i(\tau\lambda_i)\big)$ periods in the fully observed case and $\arg\min_{\tau\geq 1}\frac{1}{\tau}\big(\frac{k+2d_{i,0}}{m}+\hat{c}_i(\tau,\alpha_i,\beta_i)\big)$ in the Bayesian case for each IHC.

Therefore, to classify IHCs in accordance with their single-IHC order period, we define $\theta(i)=\arg\min_{1\leq\tau\leq\bar{\tau}}\frac{1}{\tau}\big(\frac{k+2d_{i,0}}{m}+c_i(\tau\lambda_i)\big)$ and $\hat{\theta}(i)=\arg\min_{1\leq\tau\leq\bar{\tau}}\frac{1}{\tau}\big(\frac{k+2d_{i,0}}{m}+\hat{c}_i(\tau,\alpha_i,\beta_i)\big)$ return the "class" of IHC i in the fully observed and Bayesian cases respectively. Then, letting $\mathcal{M}\subseteq\mathcal{N}$ and $\mathcal{M}(i)=\{j\in\mathcal{M}|\theta(j)=i\}$ refer to the set of IHCs in a particular class within $\mathcal{M}$, we make the following observation: if i divides j and $\lfloor\frac{|\mathcal{M}(j)|i}{j}\rfloor>1$, IHCs of class j can be substituted for class i while still satisfying service every j periods by serving $\lfloor\frac{|\mathcal{M}(j)|i}{j}\rfloor$ IHCs of class j every i periods. This can be exploited to obtain a quantity of class i such that it is divisible by m which can be used to implement fully dense trips to these IHCs. For example, when $m=2$, $|\mathcal{M}(2)|=1$ and $|\mathcal{M}(6)|=3$, if we visit a different class 6 IHC along with the class 2 IHC every 2 periods, each IHC can be visited according to its class in a fully dense route.

When class j are used to generate class i in this way, we refer to this as reducing class j into class i. Thus, we define $\texttt{red}:\mathbb{Z}^3_+\to\mathbb{Z}^2_+$ as the reducing function from j into i,

$$\texttt{red}(i,j,a)=\begin{cases}(\lfloor\frac{ai}{j}\rfloor, a-\lfloor\frac{ai}{j}\rfloor\frac{j}{i}) & j \bmod i = 0\\ (0,a) & \text{Otherwise,}\end{cases}$$

Algorithm 1 Moderate IHC Arrangement

```
function SOFT_REDUCE($\boldsymbol{a}, i$)
    $\boldsymbol{b} \leftarrow \boldsymbol{a}$
    $b_i \leftarrow b_i \bmod m$
    if $b_i = 0$ or $i = \bar{\tau}$ then return $\boldsymbol{b}$
    for $j = i + 1, j++,$ while $j \leq \bar{\tau}$ do
        if $b_j \bmod m \neq 0$ then
            if $b_i + \texttt{red}(i, j, b_j)_1 \geq m$ then
                $b_j = (\texttt{red}(i, j, b_j)_1 - (m - b_i)))\frac{j}{i} + \texttt{red}(i, j, b_j)_2 \bmod m$
                $b_i = 0$ return $\boldsymbol{b}$
            else
                $b_i = b_i + \texttt{red}(i, j, b_j)_1$
                $b_j = \texttt{red}(i, j, a_j)_2$
    return $(b_1, ..., b_i, a_{i+1}, ...a_{\bar{\tau}})$
```

where we refer to $\texttt{red}(i, j)_1$ and $\texttt{red}(i, j)_2$ as the first and second elements of the associated 2-vector, which are simply the quotient and remainder resulting from ai/j when i divides j and $(0, a)$ otherwise.

Since IHCs with low class experience frequent visits, they encourage dense routes to an even greater degree than higher classes. Thus, our heuristic prioritizes determining dense routes for lower classes via reductions. To reflect this characteristic in our heuristic, we design Algorithms 1 and 2, which take a targeted class and the number of each class of customers and return IHC groupings that yield dense routes. Algorithm 1 attempts to find fully utilized trips for class i via reductions of other classes. It prioritizes reducing smaller classes into i before resorting to reductions of larger classes, yet it only reduces class j into class i if the remaining members of class j becomes a multiple of m. In this way, the Algorithm 1 does not break up potentially fully utilized excursions composed of class j. Algorithm 2 is identical to the Algorithm 1 with the exception that it will break up classes in order to find fully dense excursions for class i.

Algorithms 1 and 2 only return the quantity of each class which yield dense routes with a focus on reducing IHCs of a particular class. To assign the routes themselves,

Algorithm 2 Strict IHC Arrangement

```
function HARD_REDUCE($\boldsymbol{a}, i$)
    $\boldsymbol{b} \leftarrow \boldsymbol{a}$
    $b_i \leftarrow b_i \bmod m$
    if $b_i = 0$ or $i = \bar{\tau}$ then return $\boldsymbol{b}$
    for $j = i + 1, j++,$ while $j \leq \bar{\tau}$ do
        if $b_i + \texttt{red}(i, j, b_j)_1 \geq m$ then
            $b_j = (\texttt{red}(i, j, b_j)_1 - (m - b_i)))\frac{j}{i} + \texttt{red}(i, j, b_j)_2 \bmod m$
            $b_i = 0$ return $\boldsymbol{b}, \boldsymbol{0}$
        else
            $b_i = b_i + \texttt{red}(i, j, b_j)_1$
            $b_j = \texttt{red}(i, j, a_j)_2$
    $\boldsymbol{a} \leftarrow \boldsymbol{b}$
    for $j = i + 1, j++,$ while $j \leq \bar{\tau}$ do
        if $b_i + b_j < m$ then
            $b_i = b_i + b_j$
            $b_j = 0$
        else
            $b_j = b_j - (m - b_i)$
            $b_i = 0$ return $\boldsymbol{b}, \boldsymbol{a} - \boldsymbol{b}$
    return $\boldsymbol{0}, \boldsymbol{0}$
```

we employ a MIP solution which minimizes the transportation cost associated with the groupings produced by the reduction Algorithms 1 and 2, which is detailed in Online Appendix C.1.

The principal advantage of our heuristic (see, e.g., Algorithm 3 in Online Appendix C.1) is that its complexity only grows linearly in $\bar{\tau}$ and avoids the exponential growth of $T_{\bar{\tau}}$ which allows for it to be applied to problems with finer time periods than our fixed-τ MIP formulations permit. Furthermore, though it still requires the solution of TSP-related MIPs, these problems are called at most $\lceil n/m \rceil$ times, and unlike MIPs (4.19) and (4.20), our heuristic uses with smaller subsets of $\mathcal{N}$, which further reduces the associated computational complexity in n. Hence, as stated in Section 4.6, though n is usually less than 20, our heuristic can still provide tractable solutions in settings with larger n.

4.8 Numerical Study

To further investigate the problem under various parameter settings, we present the following numerical study which consists of evaluating our MIP solutions, heuristic, and bounds over a large parameter suite. In order to avoid biases resulting from class/location similarity (which naturally improves our heuristic's performance), we randomly generate IHC and depot locations via a multivariate normal random variable with mean $\mathbf{0}$, and covariance $I\sigma$, where $\sigma > 0$. Then, with a FIC volume of 68.2cm^3 (according to India's immunization schedule from WHO (2016)) we assume that each IHC is equipped with refrigeration capacity of 30 liters (see, e.g., WHO (2000) and WHO (2017)), which corresponds $q_i = 450$ for all $i \in \mathcal{N}$. Then, we vary n, m, k, and σ, as well as $\boldsymbol{\lambda}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}$ in the fully observed and Bayesian cases respectively.

As discussed in Section 4.6, our MIP solutions are still computationally difficult problems, even in the reduced solution space where fixed-τ policies are employed. Hence, in Tables 4.1-4.2 we evaluate the cost percentage gap from the smaller of (1) best MIP solution obtained in 30 minutes of computation time and (2) the solution obtained via our heuristic approach in the fully-observed and Bayesian cases respectively. First turning our attention to the fully-observed case, we consider three settings for $\boldsymbol{\lambda}$: For all $i \in \mathcal{N}$, In Case 1: $\lambda_i = 100 + 25(i \bmod 6)$, in Case 2: $\lambda_i = 100 + 25(i \bmod 3)$, and in Case 3: $\lambda_i = 150 + 25(i \bmod 4)$ per week. This generates classes of IHCs that range from $1-4$, and allows us to investigate settings with low demand rates (Case 2), high demand rates (Case 3), or a mixture of both (Case 1). In the Bayesian analog, we consider cases where the prior parameters are set such that for all $i \in \mathcal{N}$, in Case 1: $\alpha_i/\beta_i = 100 + 25(i \bmod 6)$, in Case 2: $\alpha_i/\beta_i = 100 + 25(i \bmod 3)$, and in Case 3: $\alpha_i/\beta_i = 150 + 25(i \bmod 4)$ per week. To reflect different levels of information, we consider cases where, for all $i \in \mathcal{N}$, IHCs share the same value $\beta_i = 5, 10, 15, 20, 25$ since it is expected that the level of information throughout the network is approximately the same.

Since our heuristic is based on achieving highly dense routes with the aim of accommodating routine trips to each IHC in accordance with their class, it is expected for the performance of our heuristic to become stronger when n increases and m decreases. In these settings, a larger proportion of fully dense trips should be able to be accommodated without compromising travel expenses by visiting IHCs too frequently. Naturally, these travel costs comprise the bulk of the suboptimal behavior in our heuristic; by its design, high underage costs are already mitigated in our heuristic since it guarantees visits to each IHC with respect to its class. Observing averages in terms of n and m in both Table 4.1 and 4.2 reveals that this intuition holds true, since the performance of our heuristic increases as compared to the MIP solution. Therefore, since large n implies high populations size (due to the necessity of many IHCs), and since small m can indicate either low transportation capacity or high levels of demand at each IHC, we make the following:

Observation 4.1 (**Large Population/Low Transport Capacity**). *In settings large populations (n) and/or low transportation capacity (m), a manager can expect high performance from our class-based routing heuristic.*

Additionally, the heuristic's capability for accomodating fully dense trips improves when IHC classes are small and relatively uniform. This is because (1) m IHCs of the same class can always be grouped together to form fully dense trips, and (2) smaller classes can more easily reduce larger classes. Tables 4.1 and 4.2 reflect this inclination: Case 1, which features a wider range of IHC classes and largest percentage gap than Cases 2 and 3 in both Fully-Observed and Bayesian problems. Furthermore, Case

Numerical Suite Performance: % Gap							
Type	**MIP** %	**Heur** %	**MIP Time**	**Heur Time**	**LB**$_1$ %	**LB**$_2$ %	**LB**$_3$ %
$n = 9$	0.98	10.79	29.43	0	36	32	30
12	3.25	3.77	29.87	0.09	31	28	26
15	4.92	2.29	30	0.01	28	26	24
18	5.41	1.36	30	0.02	29	27	25
$m = 3$	3.37	1.25	29.95	0	19	16	13
5	4.68	3.66	30	0.01	30	28	26
7	2.95	8.41	29.58	0.09	43	41	39
$k = 0.25$	1.04	4.69	30	0.02	48	42	38
0.5	1.32	6	28.94	0.01	46	41	37
1	2.56	3.97	30	0.03	41	36	33
5	4.67	5.35	30	0.01	34	32	30
Case 1	2.9	7.2	29.78	0.01	33	31	29
Case 2	6.43	2.68	30	0.01	32	29	27
Case 3	1.26	3.8	29.7	0.1	28	25	23
$d = 0.05$	3.95	4.84	29.74	0.01	27	25	23
0.1	4.41	4.16	30	0.14	29	26	25
0.5	3.23	4.11	29.92	0.01	31	28	26
1	2.83	4.63	30	0.01	34	31	28
1.5	3.96	4.6	29.48	0.02	32	29	27
Ave.	3.63	4.48	29.84	0.04	31	28	26

Table 4.1: The performance of our heuristic and bounds, under different parameters. Timing of the MIP and Heuristic problems are in minutes.

3 has the smallest percentage gap due to the small IHC classes as a result of larger mean demand rates. Interestingly, in the Bayesian case, there does not appear to be large differences between the levels of β, which implies that the heuristic is well-suited for any level of uncertainty in parameters.

Since the heuristic and lower bounds assume fully dense trips with zero costs resulting from IHC-to-IHC transportation, as d_{ij} increases, it is expected that the performance of our heuristic will worsen. Though some small deterioration can be observed in both Table 4.1 and 4.2, this decrease in performance is largely unremarkable, which can be attributed to the fact that the rate at which IHCs are optimally served is quite insensitive to distance. Consider Figure 4.1, which demonstrates that the main determining factor to the cost (in the fully-observed case) is associated with the term c_i or $\hat{c}_i$, and hence, the choice of τ is relatively unaffected by distance costs.

Numerical Suite Performance: % Gap

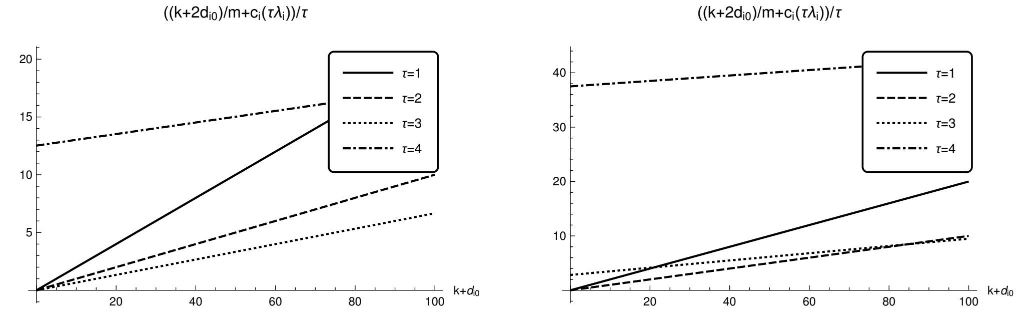
Type	**MIP** %	**Heur** %	**MIP Time**	**Heur Time**	**LB**$_1$ %	**LB**$_2$ %	**LB**$_3$ %
$n = 9$	4.21	10.95	22.19	0	45	40	35
12	1.3	3.01	26.98	0.01	42	38	34
15	2.69	2	26.79	0.02	41	37	34
18	4	0.98	30	0.02	37	35	32
$m = 3$	2.74	1.26	25.7	0	28	24	19
5	2.41	3.94	26.96	0.01	43	39	36
7	4.1	7.87	28.64	0.02	52	49	46
$k = 0.25$	1.67	5.38	25.74	0.01	48	43	39
0.5	2.9	4.57	28.81	0.01	45	41	37
1	3.48	5.45	25.78	0.01	42	38	34
5	4.3	2.21	27.94	0.01	30	28	26
Case 1	2.42	8.82	27.65	0.01	43	39	36
Case 2	3.59	2.64	28.43	0.01	43	39	35
Case 3	3.2	1.96	25.01	0.02	39	34	30
$\beta = 5$	3.9	1.97	30	0.01	38	34	31
10	2.65	4.71	27.07	0.02	42	38	35
15	2.64	5.46	25.35	0.01	42	38	34
20	3.14	4.89	26.94	0.01	41	38	34
25	3.17	4.89	23.6	0.01	43	39	35
$d = 0.05$	2.56	3.74	30	0.01	38	35	32
0.1	3.89	2.46	25.02	0.01	37	34	31
0.5	4.23	5.15	25.7	0.01	45	41	37
1	1.85	6.11	27.57	0.01	43	39	35
1.5	2.95	4.46	24.82	0.02	44	38	34
Ave.	3.1	4.4	27.12	0.01	41	37	34

Table 4.2: The performance of our heuristic and bounds, under different parameters. Timing of the MIP and Heuristic problems are in minutes.

Hence, even though the heuristic does not capure IHC-to-IHC travel costs, so long as these distances are not extremely large, the heuristic still tends to choose an optimal τ. Therefore, we make the following:

Observation 4.2 (**Sensitivity In Transportation Costs**). *Changes in fixed costs (k) and transportation costs (d_{ij}) do not have large impacts on the frequency of IHC visits.*

In developing countries where data is limited, it is unlikely that a policy-maker can determine a known demand form, which calls into question the assumption of a



(a) $\lambda = 125$

(b) $\lambda = 150$

Figure 4.1: Class sensitivity with respect to changes in transportation costs for the fully observed case. The Bayesian case behaves in an almost identical fashion.

Poisson demand process. Therefore, to test the robustness of our approach against misspecifications of the demand distribution, we evaluate the differences in expected underage, $\mathrm{E}\left[(\sum_{t=1}^{\tau} X_{i,t} - q_i)^+\right]$, when the underlying density of demand at each period deviates from Poisson. Since Poisson random variables have variances which scale according to λ, to capture deviations from our Poisson assumption, we test the expected underage when demand occurs according to discretized normal random variables with mean λ_i and standard deviation $\theta_i\sqrt{\lambda_i}$.

Figure 4.2 demonstrates the difference in expected underage associated with Normal and Poisson demand for various τ. As expected, when θ_i approaches 1, the expected underage from Normal demands becomes a good match to the expected underage in the Poisson demand process. However, even in the case where $|\theta_i - 1|$ is large, the expected underage gap remains within a small margin of the Poisson case. Importantly, since these gaps are small enough to still provide a high level of differentiation between IHC classes, our heuristic can still achieve a high level of performance in both fully-observed and Bayesian cases even when demand does not occur according to a Poisson process.

Observation 4.3 (**Sensitivity to Poisson Demand**). *Our heuristic is not highly sensitive to deviations in the distribution of demand.*

Notably, the lower bounds in Tables 4.1 and 4.2 tend to perform poorly despite acting as a basis for a high-performance heuristic. This is due to two main contributing factors: (a) IHC-to-IHC transportation costs (which are ignored in the lower bounds), and (b) a failure to serve all IHCs via fully-dense trips in accordance with their class. To investigate the impacts of these two factors, we consider the optimality percentage gap of the lower bounds when IHCs are shifted away from the district depot as pictured in Figure 4.3a in two settings: Case 1 where IHCs can be served via fully-dense trips with respect to their class, and Case 2 where this is not possible. This allows us to observe the effect of both factors: as the cluster shifts away from the depot, IHC-to-IHC travel costs become small in comparison to the total costs, which reduces Factor (a) of our heuristic as a contributor to suboptimality. Likewise, we can compare the effects of Factor (b) by observing Case 1 against Case 2.

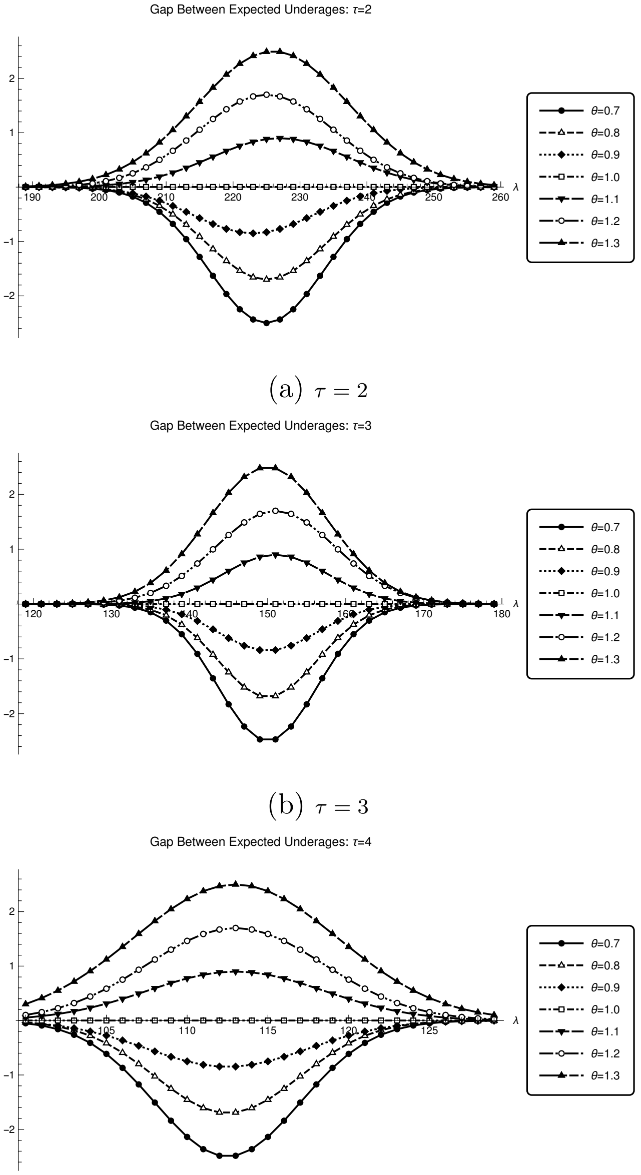
Figure 4.3b shows that, as expected, in Case 1, when the IHC-to-IHC transportation costs become small in comparison to the total costs, the bounds converge to the optimal cost. Furthermore, when IHCs become more dispersed (i.e., σ becomes large), this effect is more pronounced. Notably, the percentage gap can be reduced when n becomes large since this naturally results in a closer spacing between IHCs. In Case 2, we also see reduced percentage gap as the distance from IHCs to depot becomes small. However, since it takes place in an environment where fully-dense trips are not possible, the bounds can never achieve the optimal cost since the lower bounds assume a transportation cost of $(k + 2d_{i,0})/(m\tau)$ for each IHC. This effect is especially obvious in Tables 4.1 and 4.2 in the cases where $\sigma = 0.05$, which despite negligable IHC-to-IHC transportation costs, still experiences a large percentage gap. However, we again note that when n becomes large, the proportion of trips that are fully-dense also naturally becomes large, resulting in smaller percentage gap due to Factor (b).

Observation 4.4 (**Lower Bound Performance**). *When IHCs are grouped closely to one another, the lower bounds (and hence the heuristic) see increased performance.*

4.9 Conclusion

In order to aid developing countries' vaccine supply chains that experience (1) limited transportation capacity, (2) poor data quality, (3) insufficient managerial oversight, and (4) a lack of communication between levels of the supply chain, we identify a new strategy for delivering vaccines directly from the district depot that can directly address these problems while also providing a means of learning demand rates in a Bayesian manner.

Though the general problem is highly challenging in both fully observed and Bayesian cases, by investigating the single-IHC cases, we find easily calculable lower bounds and other features that help to characterize asymptotically optimal policies that transfer to the multiple-IHC case. Though these policies can be solved via a MIP approach, we also identify an easy-to-implement heuristic approach that operates by grouping IHCs into classes based on their demand rate and transportation costs. Finally, we verify our approach numerically via a large parameter suite, which demonstrates the high-performance of our heuristic, especially in settings with (1) numerous IHCs, (2) low transportation capacity, or (3) highly clustered IHCs. This allows for managers to design policies that naturally mitigate vaccine outages while avoiding excessive transportation costs without undergoing heavy computational burdens.

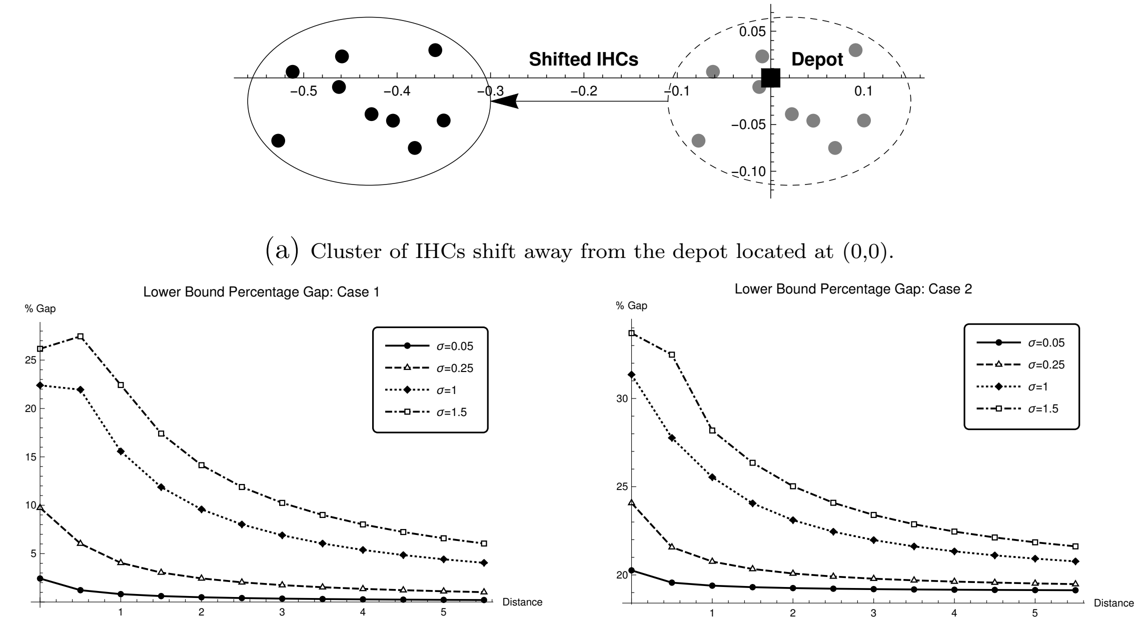


(a) $\tau = 2$

(b) $\tau = 3$

(c) $\tau = 4$

Figure 4.2: Gap between Normal and Poisson expected underages, $\mathrm{E}\left[\left(\sum_{t=1}^{\tau} X_{i,t} - q_i\right)^{+}\right] - c_i(\tau\lambda_i)$, when $q_i = 450$.



(a) Cluster of IHCs shift away from the depot located at (0,0).

(b) Case 1

(c) Case 2

Figure 4.3: Optimal percentage gap of lower bounds as IHC clusters shift away from the district depot. Case 1 has $n = 9$ with 6 Class 2 IHCs and 3 Class 3 IHCs. Case 2 has $n = 10$ with 7 Class 2 IHCs and 3 Class 3 IHCs ($m = 3, k = 1$, and $q_i = 450$ for all $i \in \mathcal{N}$). IHCs are located according to a multivariate normal random variable, with mean in accordance with the distance from the district depot, and covariance $I\sigma$.

Chapter 5

CONCLUSION

5.1 Contributions

Healthcare operations has experienced great improvements over a vast array of problems via the development and optimization of models for the purpose of informing policies that can help reduce costs and increase patient safety. However, models exist only as stylized representations of real-world scenarios, hence they rely on estimations from data and/or content experts. Thus, an estimated model can be an unreliable representation of the real-world, especially in settings with little or highly variable supporting data. This can lead to poor decision-making, resulting in increased costs or reduced patient outcomes.

To help reduce the negative consequences of model misspecifications in healthcare operations, instead of optimizing with respect to a single model, policy-makers can implement robust techniques by instead considering a suite of models, engaging in a minimax game against an antagonistic agent, and hence can safeguard themselves from potentially adverse scenarios. However, this robust methodology can result in (1) overly-conservative policies that (2) ignore learning from incoming data streams. To mitigate these drawbacks, we have studied robust frameworks for two main health care applications that allow decision-makers to engage in policies that protect against these misspecifications in accordance with their pessimism levels while utilizing incoming data to learn about the true underlying model.

In Chapter 2, we investigated a percentile optimization technique for multi-class queueing systems with unknown service rates that can utilize incoming data for learning the true system parameters. We found that the optimal policies to the non-robust parameter-learning problem take the form of an easily expressible policy that can be used to generate optimal policies to the robust problem. Since the general robust problem is highly complex, by further characterizing the robust optimal policies, we identified a high-performance, easy-to-implement heuristic that we applied to a Hospital Emergency departments application, and found that our approach could benefit Emergency departments with high congestion and unstable/unknown populations.

In Chapter 3, we propose implementing a novel inventory policy for the last mile in developing countries supply chains that utilizes the inherent thermostable properties of the vaccines. We model this problem as a MPNP and give initial results concerning its optimal policy and a simple algorithm for finding optimal ordering policies. Since the underlying model requires the estimation of many parameters (such as those guiding the arrivals, proportion of arrivals, and wastage rates), we develop a KL-divergence constrained objective that can help to protect against these ambiguities. We find that our approach can help to reduce the necessary storage capacities at IHCs, identify relative ordering behavior between vaccine types in settings with highly uncertain demand, while highlighting the benefits of increased transportation capacity and informed demand forecasting.

Finally, in Chapter 4 we propose a model that utilizes a new district-managed approach of the vaccine supply chain to transform the traditional pull system to a data-informed push system, effectively integrating the last two levels of the supply chain. We study networks where the demand rate of each IHC are (1) fully known or (2) known only up to a prior via a Bayesian approach and establish effective policies by implementing analytically driven MIP and heuristic solutions. In addition to naturally providing additional managerial oversight, improving data reliability via consolidation, and reducing the load on IHC workers, we find that this approach is effective at maintaining high levels of coverage while still utilizing incoming data for informed decision-making.

BIBLIOGRAPHY

Abdel-Malek, L. and R. Montanari, “An analysis of the multi-product newsboy problem with a budget constraint”, International Journal of Production Economics **97**, 3, 296–307 (2005).

Abdel-Malek, L., R. Montanari and L. C. Morales, “Exact, approximate, and generic iterative models for the multi-product newsboy problem with budget constraint”, International Journal of Production Economics **91**, 2, 189–198 (2004).

Adair-Rohani, H., K. Zukor, S. Bonjour, S. Wilburn, A. Kuesel, R. Hebert and E. Fletcher, “Limited electricity access in health facilities of sub-Saharan Africa: a systematic review of data on electricity access, sources, and reliability”, Global Health: Science and Practice **1**, 2, 249–261 (2013).

Adida, E. and G. Perakis, “A robust optimization approach to dynamic pricing and inventory control with no backorders”, Mathematical Programming **107**, 1, 97–129 (2006).

Akoh, W., J. Ateudjieu, J. Nouetchognou, M. Yakum, F. Nembot, S. Sonkeng, M. Fopa and P. Watcho, “The expanded program on immunization service delivery in the Dschang health district, west region of Cameroon: a cross sectional survey”, BMC Public Health **16**, 1, 801 (2016).

Andersson, H., A. Hoff, M. Christiansen, G. Hasle and A. Løkketangen, “Industrial aspects and literature survey: Combined inventory management and routing”, Computers & Operations Research **37**, 9, 1515–1536 (2010).

Argon, N. and S. Ziya, “Priority assignment under imperfect information on customer type identities”, Manufacturing & Service Operations Management **11**, 4, 674–693 (2009).

Assi, T., S. T. Brown, A. Djibo, B. A. Norman, J. Rajgopal, J. S. Welling, S. Chen, R. R. Bailey, S. Kone, H. Kenea *et al.*, “Impact of changing the measles vaccine vial size on Niger’s vaccine supply chain: a computational model”, BMC Public Health **11**, 1, 1 (2011).

Assi, T., S. T. Brown, S. Kone, B. A. Norman, A. Djibo, D. L. Connor, A. R. Wateska, J. Rajgopal, R. B. Slayton and B. Y. Lee, “Removing the regional level from the Niger vaccine supply chain”, Vaccine **31**, 26, 2828–2834 (2013).

Azoury, K., “Bayes solution to dynamic inventory models under unknown demand distribution”, Management science **31**, 9, 1150–1160 (1985).

Bagnell, J., A. Y. Ng and J. Schneider, “Solving uncertain Markov decision problems”, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, Tech. Rep. CMU-RI-TR-01-25 (2001).

Bandi, C. and D. Bertsimas, “Tractable stochastic analysis in high dimensions via robust optimization”, Mathematical Programming **134**, 1, 23–70 (2012).

Bandi, C., D. Bertsimas and N. Youssef, "Robust queueing theory", Operations Research **63**, 3, 676–700 (2015).

Bassamboo, A. and A. Zeevi, "On a data-driven method for staffing large call centers", Operations Research **57**, 3, 714–726 (2009).

Bektas, T., "The multiple traveling salesman problem: an overview of formulations and solution procedures", Omega **34**, 3, 209–219 (2006).

Bell, W. J., L. M. Dalberto, M. L. Fisher, A. J. Greenfield, R. Jaikumar, P. Kedia, R. G. Mack and P. J. Prutzman, "Improving the distribution of industrial gases with an on-line computerized routing and scheduling optimizer", Interfaces **13**, 6, 4–23 (1983).

Ben-Daya, M. and A. Raouf, "On the constrained multi-item single-period inventory problem", International Journal of Operations & Production Management **13**, 11, 104–112 (1993).

Bertsekas, D., *Dynamic Programming and Optimal Control*, vol. 1 (Athena Scientific Belmont, MA, 1995).

Bertsimas, D., D. Gamarnik and A. A. Rikun, "Performance analysis of queueing networks via robust optimization", Operations Research **59**, 2, 455–466 (2011).

Bertsimas, D., V. Gupta and N. Kallus, "Data-driven robust optimization", arXiv preprint arXiv:1401.0212 (2013).

Bertsimas, D. and M. Sim, "The price of robustness", Operations Research **52**, 1, 35–53 (2004a).

Bertsimas, D. and M. Sim, "The price of robustness", Operations Research **52**, 1, 35–53 (2004b).

Bertsimas, D. and A. Thiele, "A robust optimization approach to inventory theory", Operations Research **54**, 1, 150–168 (2006).

Bienstock, D. and N. ÖZbay, "Computing robust basestock levels", Discrete Optimization **5**, 2, 389–414 (2008).

Bosch-Capblanch, X., O. Ronveaux, V. Doyle, V. Remedios and A. Bchir, "Accuracy and quality of immunization information systems in forty-one low income countries", Tropical Medicine & International Health **14**, 1, 2–10 (2009).

Brown, S. T., B. Schreiber, B. E. Cakouros, A. R. Wateska, H. M. Dicko, D. L. Connor, P. Jaillard, M. Mvundura, B. A. Norman, C. Levin *et al.*, "The benefits of redesigning Benin's vaccine supply chain", Vaccine **32**, 32, 4097–4103 (2014).

Buyukkoc, C., P. Varaiya and J. Walrand, "The $c\mu$ rule revisited", Advances in Applied Probability **17**, 237–238 (1985).

Charnes, A. and W. W. Cooper, “Chance-constrained programming”, Management Science **6**, 1, 73–79 (1959).

Chen, D. and D. Kristensen, “Opportunities and challenges of developing thermostable vaccines”, Expert Review of Vaccines **8**, 5, 547–557 (2009).

Chen, Y. and V. Farias, “Simple policies for dynamic pricing with imperfect forecasts”, Operations Research **61**, 3, 612–624 (2013).

Chilundo, B., J. Sundby and M. Aanestad, “Analysing the quality of routine malaria data in Mozambique”, Malaria Journal **3**, 1, 3 (2004).

Chow, Y., M. Ghavamzadeh, L. Janson and M. Pavone, “Risk-constrained reinforcement learning with percentile risk criteria”, arXiv preprint arXiv:1512.01629 (2017).

Coelho, L. C., J. Cordeau and G. Laporte, “Thirty years of inventory routing”, Transportation Science **48**, 1, 1–19 (2013).

Delage, E. and S. Mannor, “Percentile optimization in uncertain Markov decision processes with application to efficient exploration”, in “Proceedings of the 24th international conference on Machine learning”, pp. 225–232 (ACM, 2007).

Delage, E. and S. Mannor, “Percentile optimization for Markov decision processes with parameter uncertainty”, Operations Research **58**, 203–213 (2010).

Derlet, R. W. and J. R. Richards, “Overcrowding in the nation’s Emergency Departments: Complex causes and disturbing effects”, Annals of Emergency Medicine **35**, 1, 63–68 (2000).

Derlet, R. W., J. R. Richards and R. L. Kravitz, “Frequent overcrowding in U.S. Emergency Departments”, Academic Emergency Medicine **8**, 2, 151–155 (2001).

Dhamodharan, A. and R. Proano, “Determining the optimal vaccine vial size in developing countries: a Monte Carlo simulation approach”, Health care management science **15**, 3, 188–196 (2012).

Dupin, C., *Applications de Géométrie et de Méchanique* (Bachelier, successeur de Mme. Ve. Courcier, libraire, 1822).

Engineer, F. G., K. C. Furman, G. L. Nemhauser, M. Savelsbergh and J. Song, “A branch-price-and-cut algorithm for single-product maritime inventory routing”, Operations Research **60**, 1, 106–122 (2012).

Erlebacher, S. J., “Optimal and heuristic solutions for the multi-item newsvendor problem with a single capacity constraint”, Production and Operations Management **9**, 3, 303–318 (2000).

Ethiopia’s Federal Ministry of Health, “Ethiopia comprehensive multi-year plan 2011-2015”, (2010).

Fresen, D., “A multivariate Gnedenko law of large numbers”, The Annals of Probability **41**, 5, 3051–3080 (2013).

Gabrel, V., C. Murat and A. Thiele, “Recent advances in robust optimization: An overview”, European Journal of Operational Research **235**, 3, 471–483 (2014).

Galazka, A., J. Milstien and M. Zaffran, *Thermostability of Vaccines* (Citeseer, 1998).

Gallego, G. and I. Moon, “The distribution free newsboy problem: review and extensions”, Journal of the Operational Research Society **44**, 8, 825–834 (1993).

GAVI, “GAVI alliance country tailored approach for Nigeria 2014-2018.”, (2014).

Geyer, C., “Lower-truncated poisson and negative binomial distributions”, (2017).

Grønhaug, R., M. Christiansen, G. Desaulniers and J. Desrosiers, “A branch-and-price method for a liquefied natural gas inventory routing problem”, Transportation Science **44**, 3, 400–415 (2010).

Guttmann, A., M. J. Schull, M. J. Vermeulen and T. A. Stukel, “Association between waiting times and short term mortality and hospital admission after departure from Emergency Department: Population based cohort study from Ontario, Canada”, British Medical Journal **342** (2011).

Hadley, G. W., “Analysis of inventory systems”, Tech. rep. (1963).

Haidari, L. A., D. L. Connor, A. R. Wateska, S. T. Brown, L. E. Mueller, B. A. Norman, M. M. Schmitz, P. Paul, J. Rajgopal, J. S. Welling *et al.*, “Augmenting transport versus increasing cold storage to improve vaccine supply chains”, PloS one **8**, 5, e64303 (2013).

Halm, A., I. Yalcouyé, M. Kamissoko, T. Keïta, N. Modjirom, S. Zipursky, U. Kartoglu and O. Ronveaux, “Using oral polio vaccine beyond the cold chain: a feasibility study conducted during the national immunization campaign in Mali”, Vaccine **28**, 19, 3467–3472 (2010).

Hansen, L. and T. Sargent, “Recursive robust estimation and control without commitment”, Journal of Economic Theory **136**, 1, 1–27 (2007).

Hu, Z. and J. Hong, “Kullback-Leibler divergence constrained distributionally robust optimization”, Available on optimization online (2012).

Huang, J., B. Carmeli and A. Mandelbaum, “Control of patient flow in emergency departments, or multiclass queues with deadlines and feedback”, Operations Research **63**, 4, 892–908 (2015).

Iyengar, G. N., “Robust dynamic programming”, Mathematics of Operations Research **30**, 2, 257–280 (2005).

Jain, A., A. Lim and J. G. Shanthikumar, “On the optimality of threshold control in queues with model uncertainty”, Queueing Systems **65**, 2, 157–174 (2010).

Juan-Giner, A., C. Domicent, C. Langendorf, M. Roper, P. Baoundoh, F. Fermon, P. Gakima, S. Zipursky, M. Tamadji and R. Grais, "A cluster randomized non-inferiority field trial on the immunogenicity and safety of tetanus toxoid vaccine kept in controlled temperature chain compared to cold chain", Vaccine **32**, 47, 6220–6226 (2014).

Khouja, M., "The single-period (news-vendor) problem: literature review and suggestions for future research", Omega **27**, 5, 537–553 (1999).

Klabjan, D., D. Simchi-Levi and M. Song, "Robust stochastic lot-sizing by means of histograms", Production and Operations Management **22**, 3, 691–710 (2013).

Lagoa, C. M., X. Li and M. Sznaier, "Probabilistically constrained linear programs and risk-adjusted controller design", SIAM Journal on Optimization **15**, 3, 938–951 (2005).

Lau, H. and A. H. Lau, "The multi-product multi-constraint newsboy problem: Applications, formulation and solution", Journal of Operations Management **13**, 2, 153–162 (1995).

Lau, H. and A. H. Lau, "The newsstand problem: A capacitated multiple-product single-period inventory problem", European Journal of Operational Research **94**, 1, 29–42 (1997).

Lim, A. E. B., J. G. Shanthikumar and G. Vahn, "Robust portfolio choice with learning in the framework of regret: Single-period case", Management Science **58**, 9, 1732–1746 (2012).

Lim, J., E. Claypool, B. A. Norman and J. Rajgopal, "Coverage models to determine outreach vaccination center locations in low and middle income countries", Operations Research for Health Care **9**, 40–48 (2016).

Lim, S., D. Stein, A. Charrow and C. Murray, "Tracking progress towards universal childhood immunisation and the impact of global initiatives: a systematic analysis of three-dose diphtheria, tetanus, and pertussis immunisation coverage", The Lancet **372**, 9655, 2031–2046 (2008).

Lippman, S. A., "Applying a new device in the optimization of exponential queuing systems", Operations Research **23**, 4, 687–710 (1975).

Littman, M. L., J. Goldsmith and M. Mundhenk, "The computational complexity of probabilistic planning", Journal of Artificial Intelligence Research **9**, 1, 1–36 (1998).

Mannor, S., D. Simester, P. Sun and J. N. Tsitsiklis, "Bias and variance approximation in value function estimates", Management Science **53**, 2, 308–322 (2007).

Miller, C. E., A. W. Tucker and R. A. Zemlin, "Integer programming formulation of traveling salesman problems", Journal of the ACM (JACM) **7**, 4, 326–329 (1960).

Moon, I. and E. A. Silver, “The multi-item newsvendor problem with a budget constraint and fixed ordering costs”, Journal of the Operational Research Society **51**, 5, 602–608 (2000).

Mueller, L., L. Haidari, A. Wateska, R. Phillips, M. Schmitz, D. Connor, B. Norman, S. Brown, J. Welling and B. Lee, “The impact of implementing a demand forecasting system into a low-income country’s supply chain”, Vaccine **34**, 32, 3663–3669 (2016).

Mundhenk, M., J. Goldsmith, C. Lusena and E. Allender, “Complexity of finite-horizon Markov decision process problems”, Journal of the ACM (JACM) **47**, 4, 681–720 (2000).

Murray, C., B. Shengelia, N. Gupta, S. Moussavi, A. Tandon and M. Thieren, “Validity of reported vaccination coverage in 45 countries”, The Lancet **362**, 9389, 1022–1027 (2003).

Nemirovski, A. and A. Shapiro, “Convex approximations of chance constrained programs”, SIAM Journal on Optimization **17**, 4, 969–996 (2006).

Nigeria’s Ministry of Health, “National routine immunization strategic plan”, (2013).

Nilim, A. and L. El Ghaoui, “Robust control of Markov decision processes with uncertain transition matrices”, Operations Research **53**, 5, 780–798 (2005).

Ophori, E., M. Tula, A. Azih, R. Okojie and P. Ikpo, “Current trends of immunization in Nigeria: prospect and challenges”, Tropical medicine and health **42**, 2, 67–75 (2014).

Osogami, T., “Robust partially observable Markov decision process.”, in “ICML”, pp. 106–115 (2015).

Papadimitriou, C. H. and J. N. Tsitsiklis, “The complexity of Markov decision processes”, Mathematics of Operations Research **12**, 3, 441–450 (1987).

Pedarsani, R., J. Walrand and Y. Zhong, “Robust scheduling and congestion control for flexible queueing networks”, in “2014 International Conference on Computing, Networking and Communications (ICNC)”, pp. 467–471 (IEEE, 2014).

Plunkett, P. K., D. G. Byrne, T. Breslin, K. Bennett and B. Silke, “Increasing wait times predict increasing mortality for emergency medical admissions”, European Journal of Emergency Medicine **18**, 4, 192–196 (2011).

Prékopa, A., *Stochastic Programming* (Klewer Academic Publishers, Dordrecht, 1995).

Rajgopal, J., D. Connor, T. Assi, B. Norman, S. Chen, R. Bailey, A. Long, A. Wateska, K. Bacon, S. Brown *et al.*, “The optimal number of routine vaccines to order at health clinics in low or middle income countries”, Vaccine **29**, 33, 5512–5518 (2011).

Ren, Q., H. Xiong, Y. Li, R. Xu and C. Zhu, “Evaluation of an outside-the-cold-chain vaccine delivery strategy in remote regions of western China”, Public Health Reports pp. 745–750 (2009).

Ross, S., J. Pineau, B. Chaib-draa and P. Kreitmann, “A Bayesian approach for learning and planning in partially observable Markov decision processes”, Journal of Machine Learning Research **12**, 1729–1770 (2011).

Ryman, T., V. Dietz and L. Cairns, “Too little but not too late: results of a literature review to improve routine immunization programs in developing countries”, BMC Health Services Research **8**, 1, 134 (2008).

Saghafian, S., “Ambiguous Partially Observable Markov Decision Processes: Structural Results and Applications.”, Working Paper, Harvard University (2018).

Saghafian, S., G. Austin and S. J. Traub, “Operations research/management contributions to Emergency Department patient flow optimization: Review and research prospects”, IIE Transactions on Healthcare Systems Engineering **5**, 2 (2015).

Saghafian, S., W. J. Hopp, M. P. Van Oyen, J. S. Desmond and S. L. Kronick, “Patient streaming as a mechanism to improve responsiveness in Emergency Departments”, Operations Research **60**, 5, 1080–1097 (2012).

Saghafian, S., W. J. Hopp, M. P. Van Oyen, J. S. Desmond and S. L. Kronick, “Complexity-augmented triage: A tool for improving patient safety and operational efficiency”, Manufacturing and Service Operations Management **16**, 3, 329–345 (2014).

Saghafian, S. and M. H. Veatch, “A $c\mu$ rule for two-tiered parallel servers”, IEEE Transactions on Automatic Control **61**, 4, 1046–1050 (2016).

Scarf, H., “Bayes solutions of the statistical inventory problem”, The Annals of Mathematical Statistics pp. 490–508 (1959).

See, C. and M. Sim, “Robust approximation to multiperiod inventory management”, Operations Research **58**, 3, 583–594 (2010).

Shen, A., R. Fields and M. McQuestion, “The future of routine immunization in the developing world: challenges and opportunities”, Global Health: Science and Practice **2**, 4, 381–394 (2014).

Smallwood, R. and E. J. Sondik, “The optimal control of partially observable Markov processes over a finite horizon”, Operations Research **21**, 5, 1071–1088 (1973).

Sondik, E. J., *The Optimal Control of Partially Observable Markov Processes*, Ph.D. thesis, Stanford University (1971).

Sprivulis, P. C., J. Da Silva, I. G. Jacobs, A. Frazer and G. A. Jelinek, “The association between hospital overcrowding and mortality among patients admitted via Western Australian Emergency Departments”, Medical Journal of Australia **184**, 5, 208 (2006).

Su, H., *Robust Fluid Control of Multiclass Queueing Networks*, Master's thesis, Massachusetts Institute of Technology (2006).

Thrun, S., "Monte Carlo POMDPs", Advances in Neural Information Processing Systems **12**, 1064–1070 (1999).

Trzeciak, S. and E. P. Rivers, "Emergency Department overcrowding in the United States: an emerging threat to patient safety and public health", Emergency Medicine Journal **20**, 5, 402–405 (2003).

Uggen, K. T., M. Fodstad and V. S. Nørstebø, "Using and extending fix-and-relax to solve maritime inventory routing problems", Top **21**, 2, 355–377 (2013).

UNICEF, WHO, "User guide for the WHO Vaccine Volume Calculator", (2012).

Vairaktarakis, G. L., "Robust multi-item newsboy models with a budget constraint", International Journal of Production Economics **66**, 3, 213–226 (2000).

Van Mieghem, J. A., "Dynamic scheduling with convex delay costs: The generalized $c\mu$ rule", The Annals of Applied Probability pp. 809–833 (1995).

Villadiego, S., "Use of vaccines outside of the cold chain: A literature review", PATH Publications (2008).

Wagenaar, B., S. Gimbel, R. Hoek, J. Pfeiffer, C. Michel, J. Manuel, F. Cuembelo, T. Quembo, P. Afonso, V. Porthé *et al.*, "Effects of a health information system data quality intervention on concordance in Mozambique: time-series analyses from 2009–2012", Population health metrics **13**, 1, 1 (2015).

Wang, Z., P. Glynn and Y. Ye, "Likelihood robust optimization for data-driven problems", Computational Management Science **13**, 2, 241–261 (2016).

White, C. and D. Harrington, "Application of Jensen's inequality to adaptive suboptimal design", Journal of Optimization Theory and Applications **32**, 1, 89–99 (1980).

WHO, "Product Information Sheets: Expanded Programme on Immunization", (2000).

WHO, "Microplanning for immunization service delivery using the Reaching Every District (RED) strategy (WHO/IVB/09.11)", (2010).

WHO, "EPI Fact Sheet: India. Regional Office for South-East Asia", (2016).

WHO, "Vaccine Management Handbook: How to Calculate Vaccine Volumes and Cold Chain Capacity Requirements. Module VMH-E3-01.1.", (2017).

Wiesemann, W., D. Kuhn and B. Rustem, "Robust Markov decision processes", Mathematics of Operations Research **38**, 1, 153–183 (2013).

World Health Organization, "Monitoring vaccine wastage at country level: guidelines for programme managers", (2005).

World Health Organization, “Regional strategic plan for immunization 2014-2020”, (2014).

Xekalaki, E., “Chance mechanisms for the univariate generalized Waring distribution and related characterizations”, in “Statistical distributions in scientific work”, pp. 157–171 (Springer, 1981).

Xin, L., D. Goldberg and A. Shapiro, “Distributionally robust multistage inventory models with moment constraints”, arXiv preprint arXiv:1304.3074 (2013).

Zaffran, M., J. Vandelaer, D. Kristensen, B. Melgaard, P. Yadav, K. Antwi-Agyei and H. Lasher, “The imperative for stronger vaccine supply and logistics systems”, Vaccine **31**, B73–B80 (2013).

Zhang, B., “Multi-tier binary solution method for multi-product newsvendor problem with multiple constraints”, European Journal of Operational Research **218**, 2, 426–434 (2012).

Zhang, B. and Z. Hua, “A unified method for a class of convex separable nonlinear knapsack problems”, European Journal of Operational Research **191**, 1, 1–6 (2008).

Zhang, B., X. Xu and Z. Hua, “A binary solution method for the multi-product newsboy problem with budget constraint”, International Journal of Production Economics **117**, 1, 136–141 (2009).

Zhang, H., “Partially observable Markov decision processes: A geometric technique and analysis”, Operations Research **58**, 1, 214–228 (2010).

Zipursky, S., L. Boualam, D. O. Cheikh, J. Fournier-Caruana, D. Hamid, M. Janssen, U. Kartoglu, G. Waeterloos and O. Ronveaux, “Assessing the potency of oral polio vaccine kept outside of the cold chain during a national immunization campaign in Chad”, Vaccine **29**, 34, 5652–5656 (2011).

APPENDIX A

MULTI-CLASS QUEUEING SYSTEMS WITH MODEL AMBIGUITY

A.1 Numerical Experiments and Extensions

A.1.1 Parameter Suite

We explicitly describe the parameter suite associated with the experiment summarized in Table 2.2. In each parameter configuration, was let $\mathbf{c} = (1, 1)$ and $\beta = 0.99$. Since there is only one optimal policy when $\mu_{1,i} > \mu_{2,j}$, (or $\mu_{1,i} < \mu_{2,j}$) for all $i, j \in \{1, 2\}$, our parameter suite focuses on the case where $\mu_{2,1} < \mu_{1,1} < \mu_{1,2} < \mu_{2,2}$ so that the minimax policy focuses on class 1 and minimin policy focuses on class 2. To limit the study for tractability purposes, we let the parameters vary from 0.1 to 0.9 on a grid with 0.1 increments so that each $\mu_{i,j} = 0.1 * k$ for $i, j \in \{1, 2\}$ and $k \in \{1, 2, \ldots, 9\}$. There are 126 configurations of parameters that satisfy these requirements. Therefore, for each $\mathbf{X}$ specified in Table 2.2, we evaluate each policy (minimax, minimin, and 95% chance-constrained for f_1, f_2, f_3 and f_4) on each parameter configuration at the central prior $\bar{\mathbf{b}}$, resulting in $1,134$ examples for each policy.

A.1.2 Heuristic Performance

To help establish the near-optimality of the E$c\mu$ heuristic, we utilize the parameter suite explicitly described above in Section A.1.1 and evaluate the percentile objective using this heuristic. Table A.1 gives the optimality gap percentage of the percentile objective under the heuristic policy π, with respect to the parameter suite associated with Table 2.2 evaluated as

$$\frac{\mathrm{Y}^{\pi}(\mathbf{X}, \epsilon) - \mathrm{Y}(\mathbf{X}, \epsilon)}{\mathrm{Y}(\mathbf{X}, \epsilon)}\%.$$

The extremely small gaps indicate that our heuristic performs very closely to the optimal percentile policy. Again, this is due to (a) the relationship of the E$c\mu$ policy to the optimal non-robust policies, (b) the fact that we utilize a prior on the boundary of $\mathcal{L}_{\epsilon}$, and (c) the visibility of this prior from the worst-case belief.

	Optimality Gap (%)			
$\mathbf{X}$	95% Chance Constrained f_1	95% Chance Constrained f_2	95% Chance Constrained f_3	95% Chance Constrained f_4
$(2,2)$	0.02	0.03	0.01	0.01
$(2,5)$	0.08	0.12	0.1	0.10
$(2,10)$	0.06	0.08	0.07	0.07
$(5,2)$	0.13	0.18	0.05	0.04
$(5,5)$	0.10	0.11	0.12	0.12
$(5,10)$	0.10	0.11	0.11	0.12
$(10,2)$	0.14	0.17	0.09	0.08
$(10,5)$	0.23	0.27	0.27	0.27
$(10,10)$	0.26	0.27	0.25	0.27
Ave.	0.12	0.15	0.12	0.12

Table A.1: Performance of the heuristic over the test suite ($n = 2, m_1 = 2, m_2 = 2$).

To show that our heuristic policy performs well, even in large instances, we perform simulations for very large problem instances, well-beyond our capability for finding the optimal chance-constrained solution. Using a uniform $\mathbb{P}_{\mathbf{B}}$ with $\epsilon = 0.05$ and $\mathbf{X} = (10, 10, 10, 10, 10, 10)$ as well as $\mathbf{X} = (30, 30, 30, 30, 30, 30)$ we again evaluate over the CVar statistic with respect to the heuristic and minimax policies which can be seen in Figure A.1. In this experiment, we see many of the same performance characteristics as our smaller examples. Notably, the heuristic ourperforms the minimax policy of the spectrum of optimism/pessimism.

A.1.3 Sensitivity: Robust Case

To show that the sensitivities to policies with respect to the selection of the prior are alleviated in our robust framework, we re-examine the experiment associated

Class	$\mu_{i,1}$	$\mu_{i,2}$	$\mu_{i,3}$	$\mu_{i,4}$	$\mu_{i,5}$	$\mu_{i,6}$
1	0.1	0.105	0.11	0.115	0.12	0.125
2	0.085	0.15	0.125	0.14	0.165	0.185
3	0.08	0.11	0.14	0.15	0.165	0.175
4	0.07	0.08	0.1	0.13	0.145	0.155
5	0.065	0.075	0.085	0.09	0.155	0.2
6	0.055	0.065	0.075	0.08	0.115	0.175

Table A.2: Ambiguity set for the service rates for each of our 6 customer classes in our extended example.

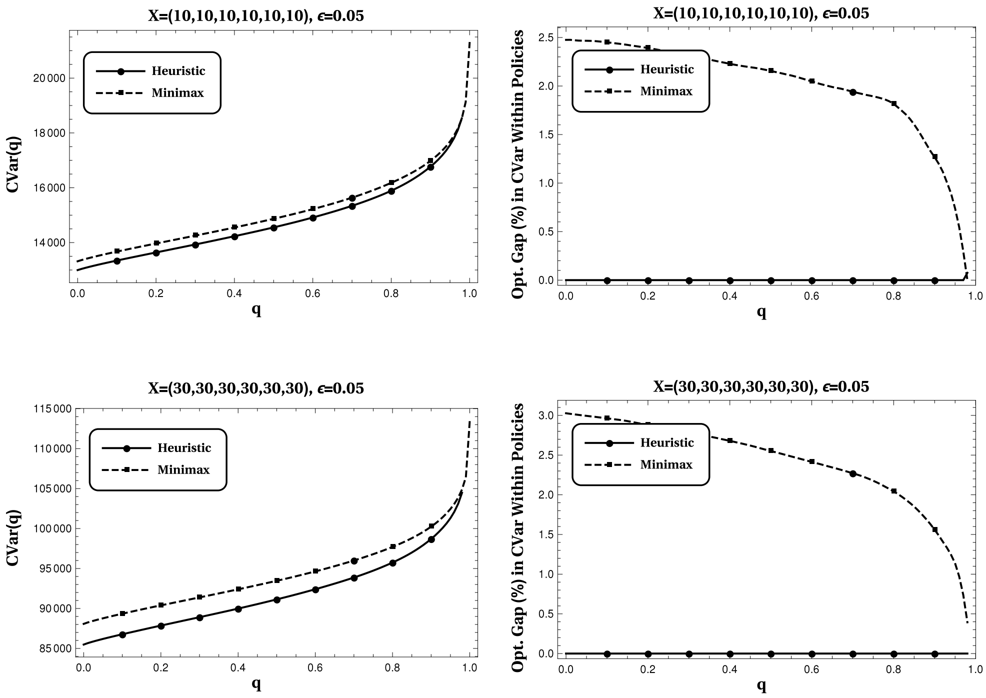


Figure A.1: $10,000$ simulations of the 95% E$c\mu$ heuristic and minimax policies, when $\mathbb{P}_{\mathbf{B}}$ is uniform and $\mathbf{c} = (1.0, 1.0, 1.0, 1.0, 1.0, 1.0)$.

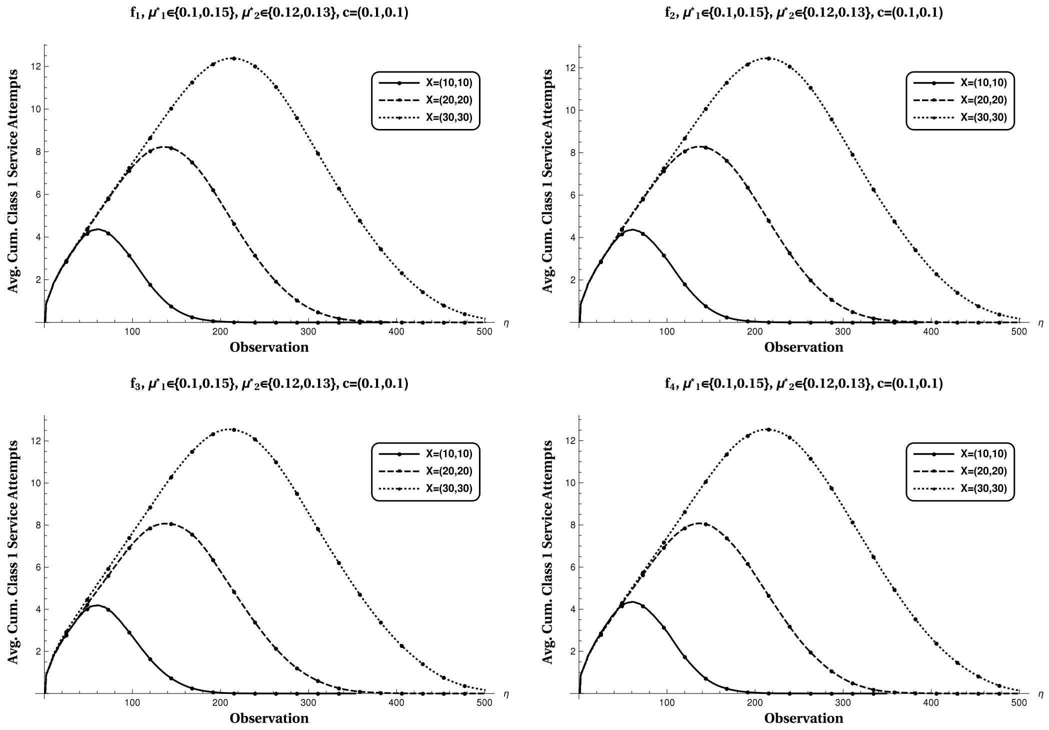


Figure A.2: Comparison of four 95% chance-constrained policies with prior densities f_1, f_2, f_3, and f_4 against $\bar{\mathbf{b}}$ ($\mu_{1,1} = 0.1, \mu_{1,2} = 0.15, \mu_{2,1} = 0.12, \mu_{2,2} = 0.13$).

with Figure 2.5. Keeping the experiment setting identical to Figure 2.5, we evaluate policies over prior distributions f_1, f_2, f_3, and f_4 associated with Table 2.2. Again, we run simulations in which the true parameter settings are selected according to $\bar{\mathbf{b}}$ and keep track of the cumulative number of attempts to serve class 1 by time t under each policy, and depict the results in Figure A.2.

Figure A.2 shows that these policies are nearly identical, which is expected since the belief points that generate these policies lie near one another, a fact that is also reflected in Table 2.2.

A.1.4 Extension: Systems with Dynamic Arrivals

Consider the case where arrivals occur in accordance with independent Poisson processes with associated rates $\hat{\lambda}_i$ for each class $i \in \mathcal{N}$. For the stability of the queue, we assume that $\sum_{i\in\mathcal{N}} \hat{\lambda}_i < \min_{i\in\mathcal{N},j\in\mathcal{J}_i} \hat{\mu}_{i,j}$. We can express the system as a dynamic program similar to (2.4) using uniformized parameters with respect to $\psi > \max_{i\in\mathcal{N},j\in\mathcal{J}_i} \hat{\mu}_{i,j} + \sum_{i\in\mathcal{N}} \hat{\lambda}_i$ so that $\lambda_i = \hat{\lambda}_i/\psi$. We also modify our belief update mechanism σ to $\hat{\sigma}$ so that the updated belief after receiving an observation is based on components

$$\hat{\sigma}\left(\mathbf{b}, a, +\right)_{a,j} = \frac{\mu_{a,j} b_{a,j}}{\sum_{k=1}^{m_a} \mu_{a,k} b_{a,k}} = \frac{\mu_{a,j} b_{a,j}}{\mathrm{E}\left[\mu_a|\mathbf{b}\right]}, \tag{A.1}$$

for "successful" service observations, and

$$\hat{\sigma}\left(\mathbf{b}, a, -\right)_{a,j} = \frac{\left(1 - \mu_{a,j} - \sum_{i\in\mathcal{N}} \lambda_i\right) b_{a,j}}{\sum_{k=1}^{m_a} \left(1 - \mu_{a,k} - \sum_{i\in\mathcal{N}} \lambda_i\right) b_{a,k}} = \frac{\left(1 - \mu_{a,j} - \sum_{i\in\mathcal{N}} \lambda_i\right) b_{a,j}}{\left(1 - \mathrm{E}\left[\mu_a|\mathbf{b}\right] - \sum_{i\in\mathcal{N}} \lambda_i\right)}, \tag{A.2}$$

for "failed" service observations, and $\sigma\left(\mathbf{b}, a, \theta\right)_{i,j} = b_{i,j}$ for $i \neq a$ since parameter belief is independent between customer classes. In this way, the discrete-time equivalent problem when arrivals occur according to Poisson processes is given by

$$\begin{aligned}\hat{\mathrm{V}}\left(\mathbf{X}, \mathbf{b}\right) = \mathbf{c}\mathbf{X}^{\mathrm{T}} + \beta \Bigg[\min_{a\in\mathcal{A}(\mathbf{X})} \Big\{ &\mathrm{E}\left[\mu_a|\mathbf{b}\right] \hat{\mathrm{V}}\left(\mathbf{X} - \mathbf{e}_a, \hat{\sigma}\left(\mathbf{b}, a, +\right)\right) + \sum_{i\in\mathcal{N}} \lambda_i \hat{\mathrm{V}}\left(\mathbf{X} + \mathbf{e}_i, \mathbf{b}\right) \\ &+ \big(1 - \sum_{i\in\mathcal{N}} \lambda_i - \mathrm{E}\left[\mu_a|\mathbf{b}\right]\big) \hat{\mathrm{V}}\left(\mathbf{X}, \hat{\sigma}\left(\mathbf{b}, a, -\right)\right) \Big\} \Bigg],\end{aligned} \tag{A.3}$$

where, in the case of $\mathbf{X} = \mathbf{0}$, the idling action is used. As with the clearing system, we show that this is the discrete-time equivalent to the continuous-time problem in Lemma A.12 in Online Appendix A.2.

To transfer the case of arrivals to the percentile optimization criterion, we let $\hat{\mathrm{R}}^{\pi}\left(\mathbf{X}\right) = \max_{\mathbf{b}\in\mathcal{B}} \hat{\mathrm{V}}^{\pi}\left(\mathbf{X}, \mathbf{b}\right)$ and $\hat{\mathrm{N}}^{\pi}\left(\mathbf{X}\right) = \min_{\mathbf{b}\in\mathcal{B}} \hat{\mathrm{V}}^{\pi}\left(\mathbf{X}, \mathbf{b}\right)$ be the minimax and minimin robust objectives in the case of Poisson arrival streams. Then, we define

$$\hat{\mathrm{Y}}^{\pi}(\mathbf{X}, \epsilon) = \inf_{y_\epsilon \in [\hat{\mathrm{N}}^{\pi}(\mathbf{X}), \hat{\mathrm{R}}^{\pi}(\mathbf{X})]} y_\epsilon$$

$$s.t.\ \mathbb{P}_{\mathbf{B}}\left(\hat{\mathrm{V}}^{\pi}\left(\mathbf{X},\mathbf{B}\right)\leq y_{\epsilon}\right)\geq 1-\epsilon, \tag{A.4}$$

as the percentile objective adapted to Poisson arrivals with $\hat{\mathrm{Y}}\left(\mathbf{X},\epsilon\right)=\inf_{\pi\in\Pi}\mathrm{Y}^{\pi}\left(\mathbf{X},\epsilon\right)$.

Interestingly, many of the results we found for the percentile problem with respect to the clearing system can be transferred to the case of Poisson arrivals due to the preservation of $\mathbf{b}_0$ as a "worst-case" belief and the concavity of $\hat{\mathrm{V}}\left(\mathbf{X},\mathbf{b}\right)$ with respect to its belief state. Letting $\hat{\Pi}_{\mathbf{b}}$ denote the set of non-robust policies associated with $\hat{\mathrm{V}}\left(\mathbf{X},\mathbf{b}\right)$, and appropriately altering $\mathcal{K}_{\mathbf{b}}$ to be composed of such policies, Theorem 2.2 and Propositions 2.1, 2.4, and 2.5 can be interpreted from the perspective of Poisson arrivals with associated non-robust value function $\hat{\mathrm{V}}\left(\mathbf{X},\mathbf{b}\right)$ and percentile objective $\hat{\mathrm{Y}}\left(\mathbf{X},\epsilon\right)$. Hence, we gain Proposition A.1.

Proposition A.1 (**Dynamic Arrivals**). *The result of Theorem 2.2, and Propositions 2.1, 2.4, and 2.5 holds when arrivals occur according to independent Poisson processes.*

Proposition A.1 suggests that, even for systems with dynamic arrivals, we can gain essentially all of the insights necessary for constructing effective policies under the robust percentile objective by studying the non-robust problem of a clearing system. Hence, we can use the same strategies for implementing the E$c\mu$ policy to solve the modified percentile objective as used in the clearing system.

To further connect the E$c\mu$ policy to the case with dynamic arrivals, we also show that under some conditions, the E$c\mu$ policy remains asymptotically optimal to the non-robust problem even when the system is not a clearing one. For instance, consider a queueing system that undergoes intense bursts of arrivals during which the probability of clearing any given class is near 0. If these bursts are followed by long periods of no arrivals, so that the probability of clearing the system between arrival bursts is near 1 under any non-idling policy, the system resembles a series of clearing systems. In these types of arrival processes, the E$c\mu$ policy becomes asymptotically optimal. The proof and details of this are outlined in the proof of Corollary A.3 in Online Appendix A.2. This is an important insight with respect to hospital EDs – the application we study in Section 2.7.1 – since their arrival behavior exhibit these traits; typically EDs experience long periods of heavy traffic during peak hours followed by little traffic after midnight, where the system clears. [1]

A.1.5 Sensitivity to Prior Belief: Arrivals Case

Numerical experiments that lead to Observations 2.2 and 2.3 establish the sensitives inherent within non-robust problem with regard to the specification of belief. We find that the robust problem does not experience these sensitivities by the replicating the numerical experiments with under different $\mathbb{P}_{\mathbf{B}}$ that were used in the experiment associated with Table 2.2.

To show that actions are not sensitive to the selection of $\mathbb{P}_{\mathbf{B}}$, we compare the cumulative actions taken as in experiment 2. Figure A.3 demonstrates that even under significantly different distributions, the difference between policies is on average, only differ by about two actions, which is not a large degree of difference given the number of actions it takes to empty the queue and in comparison to Figure 2.5.

To show that the difference in beliefs over time is not significant, we compare the KL-divergence between the beliefs of which two separate 95% chance constrained robust policies are generated in Figure A.4. Obviously, since their initial beliefs are very similar, and their associated policies are also very similar, the KL-divergence remains very small, especially in comparison to Figure 2.6.

[1] A year of data show that typically a peak arrival rate is seen close to noon, followed by a lull period close to midnight as shown by Figure A.6 in Online Appendix A.1.7.

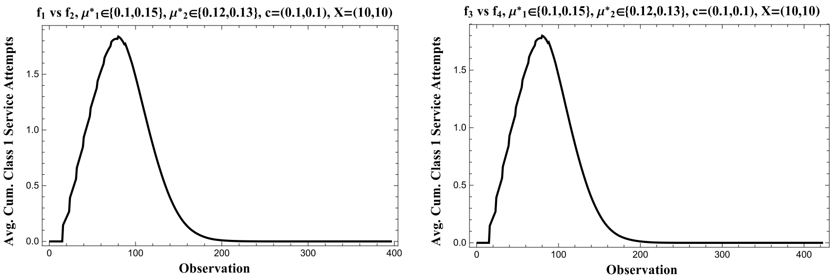


Figure A.3: Comparison of two 95% chance constrained robust policies under f_1, f_2, f_3, and f_4 associated with Table 2.2.

A.1.6 Expressible Convex Floating Bodies

The set $\mathcal{L}_\epsilon$ (and hence $\delta\mathcal{L}_\epsilon$) typically needs to be estimated by a polytope since most distributions result in convex floating bodies with no easy closed-form representation. However, upper and lower bounds to the percentile objective can be found by optimizing over sets (in the sense of Proposition 2.4) that contain or are contained by $\mathcal{L}_\epsilon$ which converge to $\Upsilon(\mathbf{X}, \epsilon)$ as the sets converge to $\mathcal{L}_\epsilon$. The details of this are expressed in the proof of Lemma A.9 in Online Appendix A.2. Additionally, with certain $\mathbb{P}_\mathbf{B}$, the problem of estimating $\mathcal{L}_\epsilon$ may be altogether circumvented. This is specifically the case when $\mathbb{P}_\mathbf{B}$ has the form of a spherical-type distribution defined below.

Definition A.1 (**Spherical Distribution**). *We say $\mathbb{P}_\mathbf{B}$ is a spherical distribution centered at $\mathbf{b}_1$, if, for any $\epsilon \in \mathbb{R}^+$, $\mathcal{L}_\epsilon = \{\mathbf{b}_2 \in \mathcal{B} : \|\mathbf{b}_2 - \mathbf{b}_1\| \leq d\}$ for some $d \in \mathbb{R}^+$, where $\|\cdot\|$ is the l^2-norm.*

In cases with spherical distributions, searching for $\mathbf{b}^*$ is simplified even in large dimensional spaces, since we have the expression for $\delta\mathcal{L}_\epsilon$ and bounds based on the visibility from $\mathbf{b}_0$. Thus, the problem is reduced to searching for the maximum of a concave function on a sphere.

Another distribution that features an easily expressible convex floating body is a

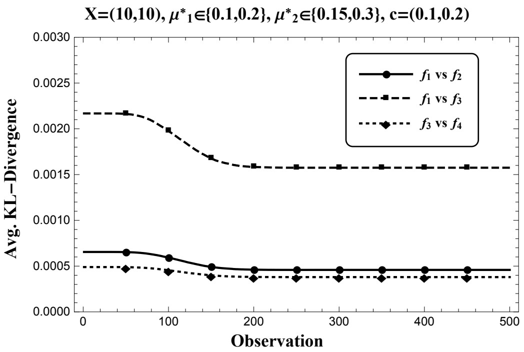


Figure A.4: Comparison of the average KL-divergence between two 95% chance constrained robust policies' beliefs under f_1, f_2, f_3, and f_4 associated with Table 2.2.

special case when $\mathbb{P}_{\mathbf{B}}$ is uniform. If $n = 2, m_1 = 2, m_2 = 2$, and $\mathbb{P}_{\mathbf{B}}$ is uniform, a small modification of a result by Calgar (2010) shows that $\delta\mathcal{L}_\epsilon$ is given by a curve defined in four quadrants as:

$$b_{2,1} = \begin{cases} \frac{b_{1,1}-1+0.5\epsilon}{b_{1,1}-1} & : 0.5 \leq b_{1,1} \leq 1-\epsilon, \quad 0.5 \leq b_{2,1} \leq 1-\epsilon \\ \frac{b_{1,1}-0.5\epsilon}{b_{1,1}} & : \epsilon \leq b_{1,1} \leq 0.5, \qquad 0.5 \leq b_{2,1} \leq 1-\epsilon \\ \frac{0.5\epsilon}{b_{1,1}} & : \epsilon \leq b_{1,1} \leq 0.5, \qquad \epsilon \leq b_{2,1} \leq 0.5 \\ -\frac{0.5\epsilon}{b_{1,1}-1} & : 0.5 \leq b_{1,1} \leq 1-\epsilon, \quad \epsilon \leq b_{2,1} \leq 0.5 \end{cases}$$

for $0 < \epsilon < 0.5$ as shown in Figure 2.4. If we evaluate $\mathrm{V}(\mathbf{X}, \mathbf{b})$ along the quadrant visible to $\mathbf{b}_0$, belief $\mathbf{b}^*$ is revealed as the maximum on this curve. We use this uniform case and various spherical cases in Section 2.7 to show that since our approach includes learning, it provides robustness to the specification of $\mathbb{P}_{\mathbf{B}}$. Therefore, even though exact closed-form representations of $\mathcal{L}_\epsilon$ in general are rare, polytope or spherical approximations are sufficient for the purposes of percentile optimization in our framework.

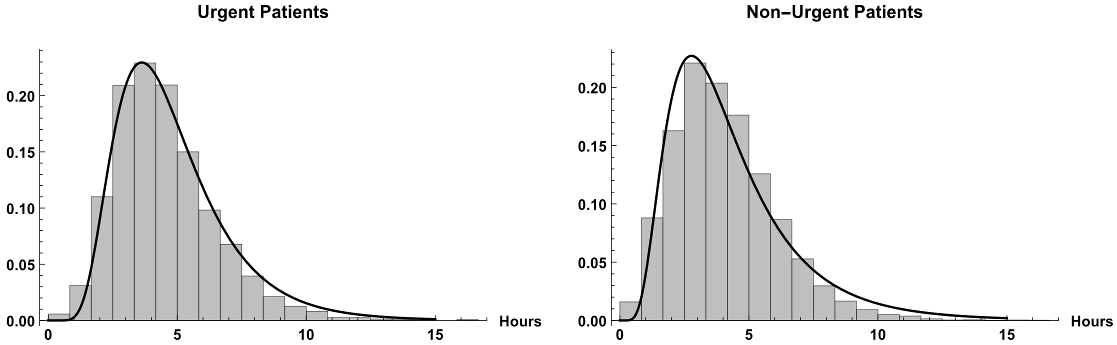


Figure A.5: Normalized histograms of the patients' Length of Stay (LOS) based on a year of data collected from a partner hospital.

A.1.7 The Hospital Emergency Department Setting: Calibration Using Data

We fit a model with fully known parameters such that first of all, the $c\mu$ priority rule (which is optimal with known exponential service rates) follows the ED protocol of prioritizing patients in order of Urgent Simple (US), Urgent Complex (UC), Non-Urgent Simple (NS), and Non-Urgent Complex (NC) classes that is proven by Saghafian *et al.* (2014) to be optimal under fully observed parameters. Secondly, we ensure that this model matches our data collected from a partner hospital seen in Figures A.5 and A.6. Furthermore, in line with what is described in the main body, we assume that the costs (ROAE) do not accumulate for patients who have started service (stabilized patients) and that the service is nonpreemtive.

To accomplish these goals, we model the arrivals as a non-stationary Poisson process with arrival rates changing every two hours according to the data obtained from our partner hospital. We assume that 51% and 49% of patients are simple and complex respectively in urgent and non-urgent patients which are the typical proportions reported in Saghafian *et al.* (2014).

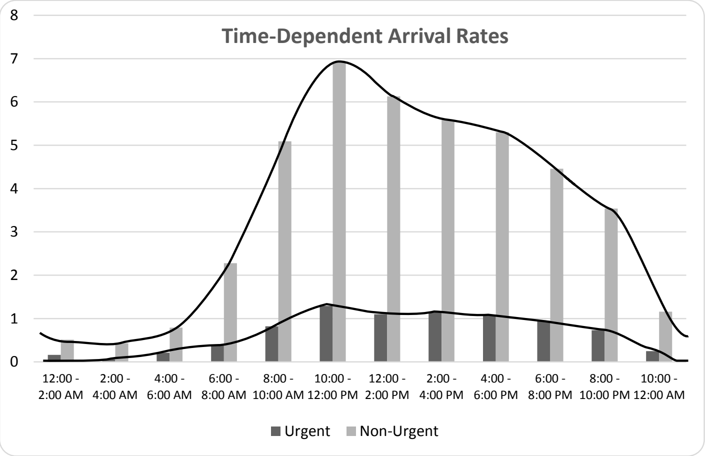


Figure A.6: Per hour arrival rates of urgent and non-urgent classes based on a year of data collected from a partner hospital.

In hospital EDs, patients are subject to misclassifications. Urgent/non-urgent patients experience misclassification errors of $9\% - 15\%$, whereas simple/complex misclassifications occur at a rate near 17% (see, e.g., Saghafian *et al.* (2014), and the references within). In our model, we use error-impacted service rates and treat the each class of patient as their triaged class, as opposed to their "true" class. Hence, the queue containing patients classified as UC may include patients that are actually within US, UC, NS, and NC, and the associated rate parameter of this queue reflects this mixture. Hereafter, we refer to each class as the "error-impacted" class.

Since LOS in urgent and non-urgent patients appears to be lognormally distributed, we assume service time distributions of the ED as a superserver are also lognormal. Complex patients by definition experience multiple visits with physicians and undergo tests which usually require more processing time. Therefore, we assume that the service rate for a complex patient is less than that of a simple patient. Next, we fit the service rates (of the ED as a superserver) so that the mean LOS is the mean LOS seen from our data set. Letting US, UC, NS, NC be classes $1, 2, 3,$

and 4, respectively, we find that lognormal distributions with rates (denoted $\hat{\mu}_{i,3}$), $\hat{\mu}_{1,3} = 6.8, \hat{\mu}_{2,3} = 2.72, \hat{\mu}_{3,3} = 10.2, \hat{\mu}_{4,3} = 4.08$ and scale parameter 0.5 see the same mean LOS as in our data set. We note that these rates are estimated from our data for 2-hour periods of time. To incorporate model ambiguity, we generate *ambiguity sets* by incorporating four additional rate parameters $\hat{\mu}_{i,1}, \hat{\mu}_{i,2}, \hat{\mu}_{i,4}, \hat{\mu}_{i,5}$ to our fitted rates $\hat{\mu}_{i,3}$ so that for each $i \in \mathcal{N}$, $\hat{\mu}_{i,1} < \hat{\mu}_{i,2} < \hat{\mu}_{i,3} < \hat{\mu}_{i,4} < \hat{\mu}_{i,5}$.

To study the effect of incorporating ambiguities, we wish to compare our proposed data-driven percentile optimization to the complexity-based prioritization (as recommended by Saghafian *et al.* (2014)), minimax, and minimin approaches. Since arrivals are included in the ED model, we must begin by modifying our Bayesian updating mechanism. To accomplish this, we use uniformization rate $\psi = 30$ to uniformize parameters $\mu_{i,j}$ corresponding to their continuous time rates. We choose this rate since is fast relative to our estimated service rates, allowing for us to employ our asymptotic results gained from Theorem 2.1. At a given time, let λ_i be the uniformized rate associated with an arrival of class i. Then the updated belief $\sigma(\mathbf{b}, a, \theta)$ after receiving an observation is based on components

$$\sigma(\mathbf{b}, a, +)_{a,j} = \frac{\mu_{a,j} b_{a,j}}{\sum_{k=1}^{m_a} \mu_{a,k} b_{a,k}} = \frac{\mu_{a,j} b_{a,j}}{\mathrm{E}[\mu_a|\mathbf{b}]}, \tag{A.5}$$

for "successful" service observations, and

$$\sigma(\mathbf{b}, a, -)_{a,j} = \frac{\left(1 - \mu_{a,j} - \sum_{i\in\mathcal{N}} \lambda_i\right) b_{a,j}}{\sum_{k=1}^{m_a} \left(1 - \mu_{a,k} - \sum_{i\in\mathcal{N}} \lambda_i\right) b_{a,k}} = \frac{\left(1 - \mu_{a,j} - \sum_{i\in\mathcal{N}} \lambda_i\right) b_{a,j}}{\left(1 - \mathrm{E}[\mu_a|\mathbf{b}] - \sum_{i\in\mathcal{N}} \lambda_i\right)}, \tag{A.6}$$

for "failed" service observations, and $\sigma(\mathbf{b}, a, \theta)_{i,j} = b_{i,j}$ for $i \neq a$ (since parameter belief is independent between customer classes).

We assume $\mathbb{P}_{\mathbf{B}}$ is uniform within $\mathcal{B}$ and wish to use a 95% chance-constrained policy on the system. Since this problem is highly dimensional and the convex floating body (introduced in the main body) is not easily determined, we utilize our E$c\mu$

heuristic. The heuristic states that our 95% chance-constrained policy should be approximately the E$c\mu$ policy with initial belief with components $\{b_{i,1}, b_{i,2}, b_{i,3}, b_{i,4}, b_{i,5}\} = \{0.51, 0.1225, 0.1225, 0.1225, 0.1225\}$ for each $i \in \mathcal{N}$.

To test our percentile approach against other methods, we simulated a day under each method and examined the total cost over this period using the CVar metric with the following ambiguity sets:

$\hat{\mu}_{i,1}$	$\hat{\mu}_{i,2}$	$\hat{\mu}_{i,4}$	$\hat{\mu}_{i,5}$
$\hat{\mu}_{i,3} - 2.0$	$\hat{\mu}_{i,3} - 1.0$	$\hat{\mu}_{i,3} + 1.0$	$\hat{\mu}_{i,3} + 2.0$
$\hat{\mu}_{i,3} - 1.5$	$\hat{\mu}_{i,3} - 0.75$	$\hat{\mu}_{i,3} + 0.75$	$\hat{\mu}_{i,3} + 1.5$
$\hat{\mu}_{i,3} - 2.0$	$\hat{\mu}_{i,3} - 1.0$	$\hat{\mu}_{i,3} + 0.75$	$\hat{\mu}_{i,3} + 1.5$
$\hat{\mu}_{i,3} - 1.5$	$\hat{\mu}_{i,3} - 0.75$	$\hat{\mu}_{i,3} + 1.0$	$\hat{\mu}_{i,3} + 2.0$

Table A.3: Ambiguity sets $\mathcal{M}_i$ considered for the different service rates of the Emergency Department under consideration.

These ambiguity sets represent scenarios of busier and queues (e.g., the third ambiguity set of Table A.3), less busy queues (e.g., the fourth ambiguity set of Table A.3), and ambiguity sets with tighter or looser considerations (e.g., the second and first ambiguity set of Table A.3 respectively).

Since it is possible that the true ROAE for complex patients is different from that of simple patients, so we test our method choosing a variety of adverse event (i.e., cost) configurations. Namely, we assume that simple patients ROAE is less than or equal to their complex counterpart, and perform simulations for when $c_1 = kc_2$ and $c_3 = kc_4$ for $k \in \{1/2, 5/8, 3/4, 7/8, 1\}$. Thus, we perform 20 simulations to analyze each combination of cost setting and ambiguity set. We implement a warm-up period of 6 hours for each run and perform $20,000$ replications for each configuration. The results of these simulations are shown in Figure 2.10 in Section 2.7.1, and in Figures A.7, A.8, A.9, and A.10.

To show that our approach can be used in even larger settings, we conduct an additional experiment that features week-long (7 day) simulations of the ED with a uniform $\mathbb{P}_{\mathbf{B}}$ and the ambiguity shown in Table A.4. The results of these simulations are shown in Figure A.11 and demonstrate that our chance-constrained policies still dominate the spectrum of pessimism/optimism as compared to other prioritization schemes. This implies that our approach offers significant advantages over non-learning policies, even in large problem instances.

Class	$\hat{\mu}_{i,1}$	$\hat{\mu}_{i,2}$	$\hat{\mu}_{i,3}$	$\hat{\mu}_{i,4}$	$\hat{\mu}_{i,5}$	$\hat{\mu}_{i,6}$	$\hat{\mu}_{i,7}$	$\hat{\mu}_{i,8}$	$\hat{\mu}_{i,9}$
US	4.80	5.30	5.80	6.30	6.80	7.30	7.80	8.30	8.80
UC	0.72	1.22	1.72	2.22	2.72	3.22	3.72	4.22	4.72
NS	8.20	8.70	9.20	9.70	10.20	10.70	11.20	11.70	12.20
NC	2.08	2.58	3.08	3.58	4.08	4.58	5.08	5.58	6.08

Table A.4: Ambiguity set for the service rates of each patient class for our extended hospital ED example.

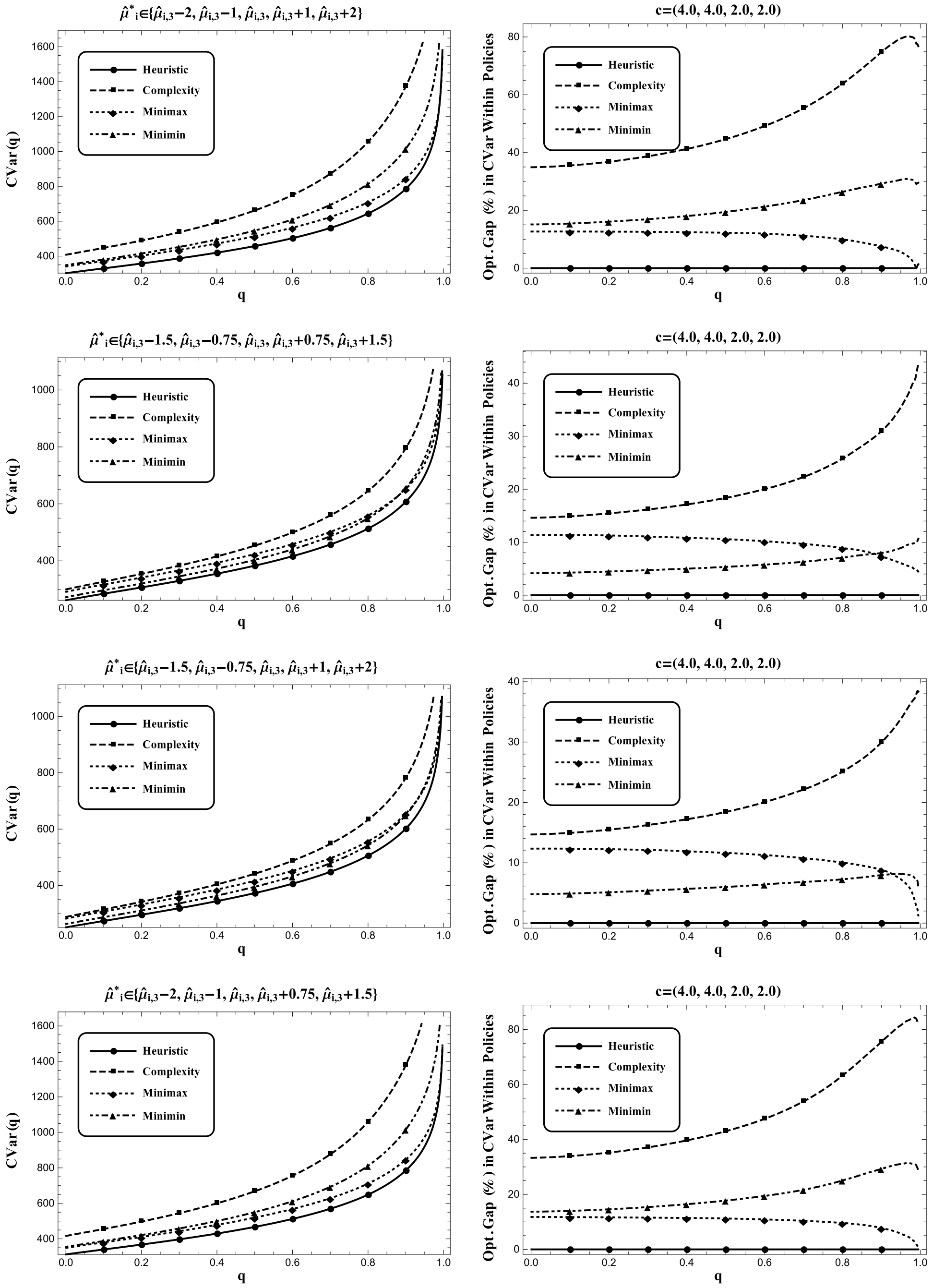


Figure A.7: 20,000 simulated days in the ED for the complexity-based prioritization, 95% E$c\mu$ heuristic, minimin, and minimax policies, when $\mathbb{P}_{\mathbf{B}}$ is uniform, and the cloud of models perturbs the fitted service rate $\hat{\mu}_{i,3}$ in terms of two-hour time increments with $\mathbf{c} = (4.0, 4.0, 2.0, 2.0)$. (Triage levels US, UC, NS, and NC are denoted 1,2,3, and 4, respectively.)

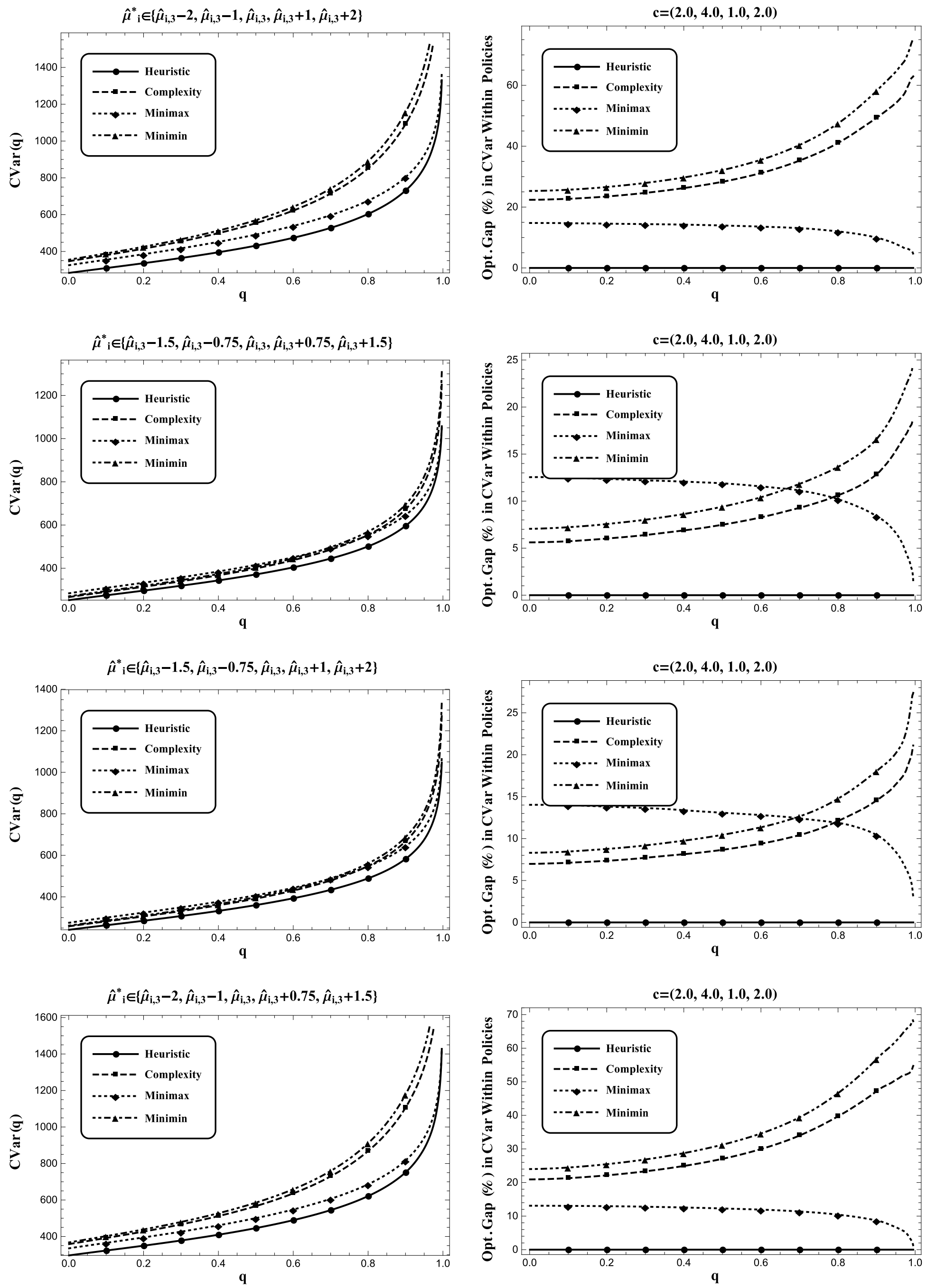


Figure A.8: 20,000 simulated days in the ED for the complexity-based prioritization, 95% E$c\mu$ heuristic, minimin, and minimax policies, when $\mathbb{P}_{\mathbf{B}}$ is uniform, and the cloud of models perturbs the fitted service rate $\hat{\mu}_{i,3}$ in terms of two-hour time increments with $\mathbf{c} = (2.0, 4.0, 1.0, 2.0)$.

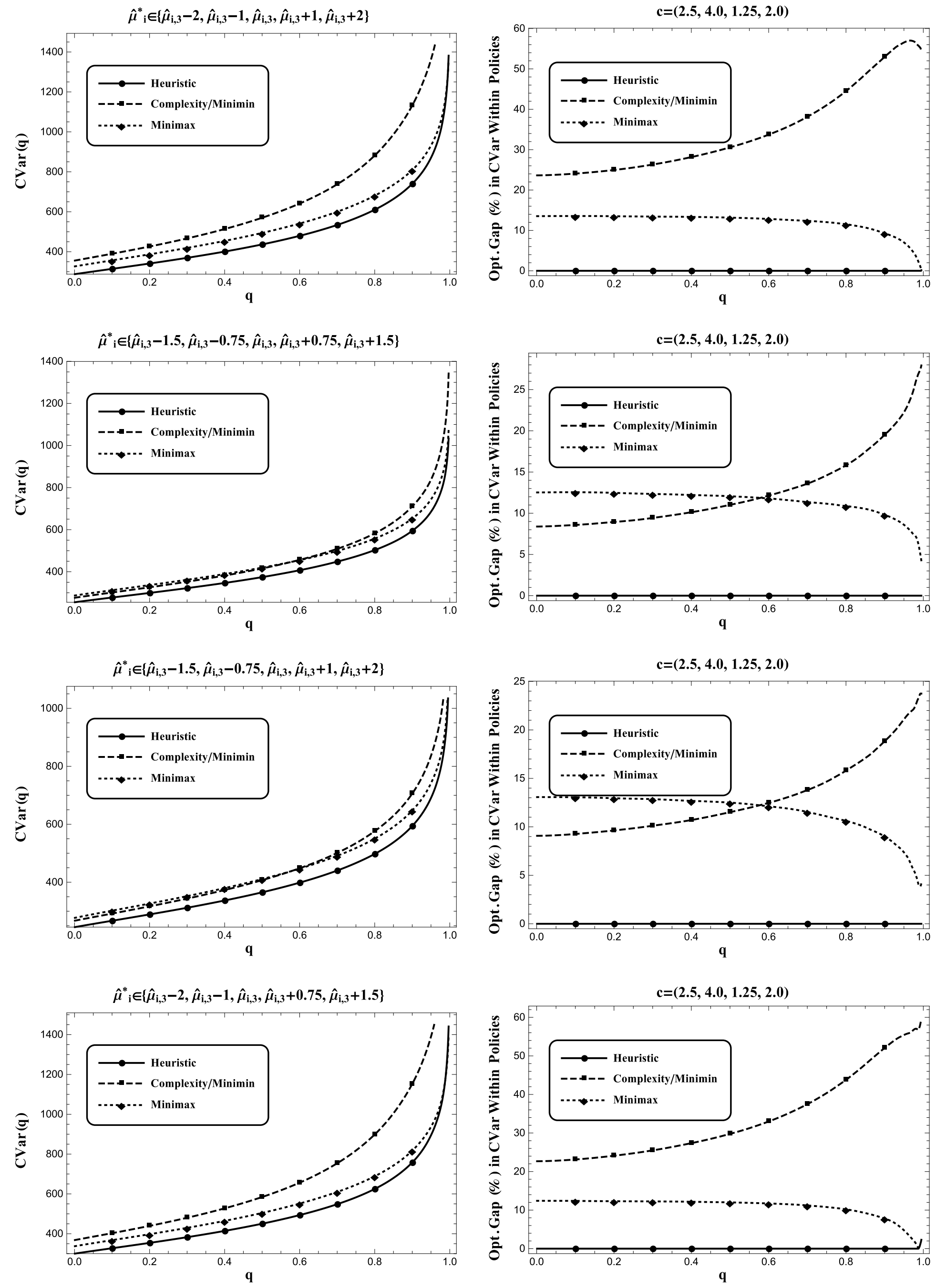


Figure A.9: 20,000 simulated days in the ED for the complexity-based prioritization, 95% E$c\mu$ heuristic, minimin, and minimax policies, when $\mathbb{P}_{\mathbf{B}}$ is uniform, and the cloud of models perturbs the fitted service rate $\hat{\mu}_{i,3}$ in terms of two-hour time increments with $\mathbf{c} = (2.5, 4.0, 1.25, 2.0)$.

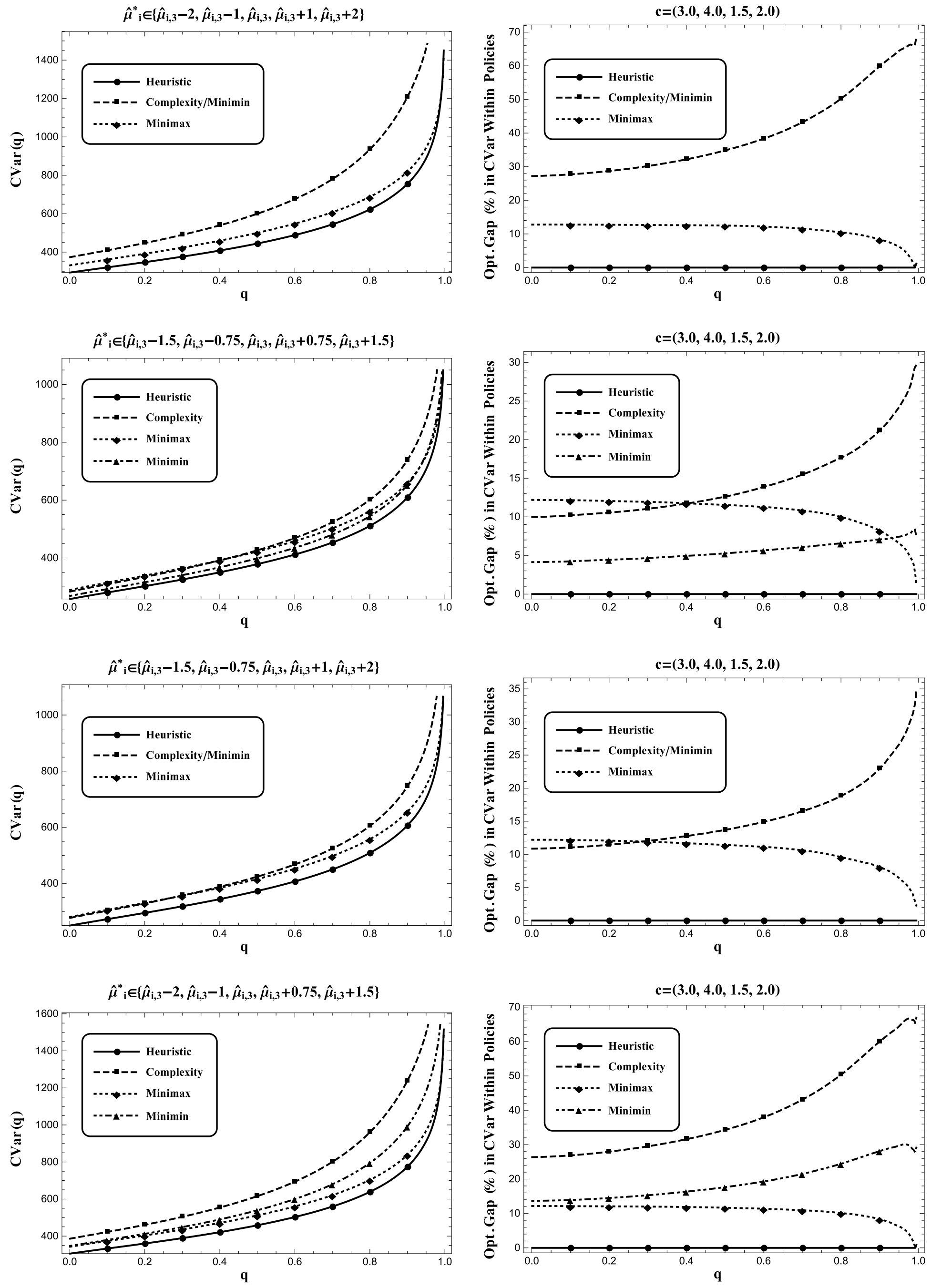


Figure A.10: 20,000 simulated days in the ED for the complexity-based prioritization, 95% E$c\mu$ heuristic, minimin, and minimax policies, when $\mathbb{P}_{\mathbf{B}}$ is uniform, and the cloud of models perturbs the fitted service rate $\hat{\mu}_{i,3}$ in terms of two-hour time increments with $\mathbf{c} = (3.0, 4.0, 1.5, 2.0)$.

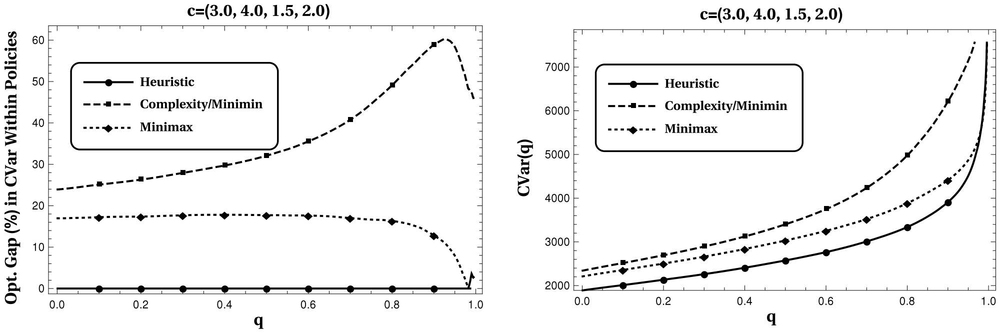


Figure A.11: $1,000$ simulated weeks in the ED for the complexity-based prioritization, 95% E$c\mu$ heuristic, minimin, and minimax policies, when $\mathbb{P}_{\mathbf{B}}$ is uniform, and the cloud of models perturbs the fitted service rate $\hat{\mu}_{i,3}$ in terms of two-hour time increments with $\mathbf{c} = (3.0, 4.0, 1.5, 2.0)$. (Triage levels US, UC, NS, and NC are denoted 1,2,3, and 4, respectively.)

A.2 Proofs of Propositions, Lemmas, and Theorems

Lemma A.1. *For all* $a \in \mathcal{A}(\mathbf{X}), t \in \mathbb{N} \bigcup \{0\}$, *and* $\mathbf{b} \in \mathcal{B}$, $\mathrm{V}_t\left(\mathbf{X}-\mathbf{e}_a, \mathbf{b}\right) < \mathrm{V}_t\left(\mathbf{X}, \mathbf{b}\right).$ *Furthermore, if* π *is a priority policy (i.e. it chooses to prioritize classes of customers),* $\mathrm{V}_t^{\pi}\left(\mathbf{X}-\mathbf{e}_a, \mathbf{b}\right) < \mathrm{V}_t^{\pi}\left(\mathbf{X}, \mathbf{b}\right).$

Proof. We proceed by induction on t. In the base case, when $t = 0$, the assertion is true since $\mathbf{c}\left(\mathbf{X}-\mathbf{e}_a\right)^{\mathrm{T}} < \mathbf{c}\mathbf{X}^{\mathrm{T}}$. For the inductive step, we suppose the assertion holds for t. Suppose the optimal action for $\mathrm{V}_t\left(\mathbf{X}, \mathbf{b}\right)$ is action a'. If $a' = a$ and $X_a - 1 = 0$, note that $\mathrm{V}\left(\mathbf{X}-\mathbf{e}_a, \mathbf{b}\right) = \mathrm{V}\left(\mathbf{X}-\mathbf{e}_a, \sigma\left(\mathbf{b}, a, +\right)\right) = \mathrm{V}\left(\mathbf{X}-\mathbf{e}_a, \sigma\left(\mathbf{b}, a, -\right)\right)$ since there are no members of class a left to serve. This implies any changes in belief concerning this class has no effect on cost. Therefore,

$$\begin{aligned}
&\mathrm{V}_{t+1}\left(\mathbf{X}-\mathbf{e}_a, \mathbf{b}\right)\\
&\leq \mathbf{c}\left(\mathbf{X}-\mathbf{e}_a\right)^{\mathrm{T}} + \beta\Bigg[\Big\{\mathrm{E}\left[\mu_{a'}|\mathbf{b}\right]\mathrm{V}_t\left(\mathbf{X}-\mathbf{e}_a, \sigma\left(\mathbf{b}, a', +\right)\right)\\
&+ \left(1-\mathrm{E}\left[\mu_{a'}|\mathbf{b}\right]\right)\mathrm{V}_t\left(\mathbf{X}-\mathbf{e}_a, \sigma\left(\mathbf{b}, a', -\right)\right)\Big\}\Bigg]\\
&< \mathbf{c}\mathbf{X}^{\mathrm{T}} + \beta\Bigg[\Big\{\mathrm{E}\left[\mu_{a'}|\mathbf{b}\right]\mathrm{V}_t\left(\mathbf{X}-\mathbf{e}_{a'}, \sigma\left(\mathbf{b}, a', +\right)\right)\\
&+ \left(1-\mathrm{E}\left[\mu_{a'}|\mathbf{b}\right]\right)\mathrm{V}_t\left(\mathbf{X}, \sigma\left(\mathbf{b}, a', -\right)\right)\Big\}\Bigg]\\
&= \mathrm{V}_{t+1}\left(\mathbf{X}, \mathbf{b}\right),
\end{aligned}$$

by the inductive hypothesis. Otherwise,

$$\begin{aligned}
&\mathrm{V}_{t+1}\left(\mathbf{X}-\mathbf{e}_a, \mathbf{b}\right)\\
&\leq \mathbf{c}\left(\mathbf{X}-\mathbf{e}_a\right)^{\mathrm{T}} + \beta\Bigg[\Big\{\mathrm{E}\left[\mu_{a'}|\mathbf{b}\right]\mathrm{V}_t\left(\mathbf{X}-\mathbf{e}_a-\mathbf{e}_{a'}, \sigma\left(\mathbf{b}, a', +\right)\right)\\
&+ \left(1-\mathrm{E}\left[\mu_{a}|\mathbf{b}\right]\right)\mathrm{V}_t\left(\mathbf{X}-\mathbf{e}_a, \sigma\left(\mathbf{b}, a', -\right)\right)\Big\}\Bigg]\\
&< \mathbf{c}\mathbf{X}^{\mathrm{T}} + \beta\Bigg[\Big\{\mathrm{E}\left[\mu_{a'}|\mathbf{b}\right]\mathrm{V}_t\left(\mathbf{X}-\mathbf{e}_{a'}, \sigma\left(\mathbf{b}, a', +\right)\right)
\end{aligned}$$

$$+ \left(1 - \mathrm{E}\left[\mu_a | \mathbf{b}\right]\right) \mathrm{V}_t\left(\mathbf{X}, \sigma\left(\mathbf{b}, a', -\right)\right) \Big\}\Big]$$

$$= \mathrm{V}_{t+1}\left(\mathbf{X}, \mathbf{b}\right),$$

by the inductive hypothesis. For the second portion of the proof, it is easy to see that the priority discipline would choose a' until none remain in the class, so the proof remains the same by substituting V_t^π in place of V_t. □

Lemma A.2. *For all $a \in \mathcal{A}(\mathbf{X}), t \in \mathbb{N} \bigcup \{0\}$, and $\mathbf{b} \in \mathcal{B}$, $\mathrm{V}_t\left(\mathbf{X} - \mathbf{e}_a, \sigma\left(\mathbf{b}, a, +\right)\right) < \mathrm{V}_t\left(\mathbf{X}, \sigma\left(\mathbf{b}, a, -\right)\right)$. Furthermore, if π is a priority policy (i.e. it chooses to prioritize classes of customers), $\mathrm{V}_t^\pi\left(\mathbf{X} - \mathbf{e}_a, \sigma\left(\mathbf{b}, a, +\right)\right) < \mathrm{V}_t^\pi\left(\mathbf{X}, \sigma\left(\mathbf{b}, a, -\right)\right)$.*

Proof. We proceed by induction on t. In the base case, when $t = 0$, the assertion is true since $\mathbf{c}\left(\mathbf{X} - \mathbf{e}_a\right)^{\mathrm{T}} < \mathbf{c}\mathbf{X}^{\mathrm{T}}$. For the inductive step, we suppose the assertion holds for t. Suppose the optimal action for $\mathrm{V}_t\left(\mathbf{X}, \sigma\left(\mathbf{b}, a, -\right)\right)$ is action a'. Similar to Lemma A.1, if $a' = a$ and $X_a - 1 = 0$, note that $\mathrm{V}\left(\mathbf{X} - \mathbf{e}_a, \sigma(\mathbf{b}, a, +)\right) = \mathrm{V}\left(\mathbf{X} - \mathbf{e}_a, \sigma(\sigma\left(\mathbf{b}, a, +\right), a, +)\right) = \mathrm{V}\left(\mathbf{X} - \mathbf{e}_a, \sigma(\sigma\left(\mathbf{b}, a, +\right), a, -)\right)$ since there are no members of class a left to serve. Again, this implies any changes in belief concerning this class has no effect on cost.

$$\mathrm{V}_{t+1}\left(\mathbf{X} - \mathbf{e}_a, \sigma\left(\mathbf{b}, a, +\right)\right)$$

$$\leq \mathbf{c}\left(\mathbf{X} - \mathbf{e}_a\right)^{\mathrm{T}} + \beta \Big[\Big\{\mathrm{E}\left[\mu_{a'} | \sigma\left(\mathbf{b}, a, +\right)\right] \mathrm{V}_t\left(\mathbf{X} - \mathbf{e}_a, \sigma\left(\sigma\left(\mathbf{b}, a, +\right), a', +\right)\right)$$

$$+ \left(1 - \mathrm{E}\left[\mu_{a'} | \sigma\left(\mathbf{b}, a', +\right)\right]\right) \mathrm{V}_t\left(\mathbf{X} - \mathbf{e}_a, \sigma\left(\sigma\left(\mathbf{b}, a, +\right), a', -\right)\right) \Big\}\Big]$$

$$< \mathbf{c}\mathbf{X}^{\mathrm{T}} + \beta \Big[\Big\{\mathrm{E}\left[\mu_{a'} | \sigma\left(\mathbf{b}, a, +\right)\right] \mathrm{V}_t\left(\mathbf{X} - \mathbf{e}_{a'}, \sigma\left(\sigma\left(\mathbf{b}, a, +\right), a', +\right)\right)$$

$$+ \left(1 - \mathrm{E}\left[\mu_{a'} | \sigma\left(\mathbf{b}, a, +\right)\right]\right) \mathrm{V}_t\left(\mathbf{X}, \sigma\left(\sigma\left(\mathbf{b}, a, +\right), a', -\right)\right) \Big\}\Big]$$

$$\leq \mathbf{c}\mathbf{X}^{\mathrm{T}} + \beta \Big[\Big\{\mathrm{E}\left[\mu_{a'} | \sigma\left(\mathbf{b}, a, -\right)\right] \mathrm{V}_t\left(\mathbf{X} - \mathbf{e}_{a'}, \sigma\left(\sigma\left(\mathbf{b}, a, -\right), a', +\right)\right)$$

$$+ \left(1 - \mathrm{E}\left[\mu_{a'} | \sigma\left(\mathbf{b}, a, -\right)\right]\right) \mathrm{V}_t\left(\mathbf{X}, \sigma\left(\sigma\left(\mathbf{b}, a, -\right), a', -\right)\right) \Big\}\Big]$$

$$= \mathrm{V}_{t+1}\left(\mathbf{X}, \sigma\left(\mathbf{b}, a, -\right)\right),$$

since $\mathrm{E}\left[\mu_{a'}|\sigma\left(\mathbf{b}, a, -\right)\right] \leq \mathrm{E}\left[\mu_{a'}|\sigma\left(\mathbf{b}, a, +\right)\right]$, the inductive hypothesis, and since $\mathrm{V}_t\left(\mathbf{X} - \mathbf{e}_a, \mathbf{b}\right) < \mathrm{V}_t\left(\mathbf{X}, \mathbf{b}\right)$ by Lemma A.1. Otherwise,

$$\begin{aligned}
&\mathrm{V}_{t+1}\left(\mathbf{X} - \mathbf{e}_a, \sigma\left(\mathbf{b}, a, +\right)\right)\\
&\leq \mathbf{c}\left(\mathbf{X} - \mathbf{e}_a\right)^{\mathrm{T}} + \beta\left[\left\{\mathrm{E}\left[\mu_{a'}|\sigma\left(\mathbf{b}, a, +\right)\right]\mathrm{V}_t\left(\mathbf{X} - \mathbf{e}_a - \mathbf{e}_{a'}, \sigma\left(\sigma\left(\mathbf{b}, a, +\right), a', +\right)\right)\right.\right.\\
&\left.\left.+\ \left(1 - \mathrm{E}\left[\mu_{a'}|\sigma\left(\mathbf{b}, a, +\right)\right]\right)\mathrm{V}_t\left(\mathbf{X} - \mathbf{e}_a, \sigma\left(\sigma\left(\mathbf{b}, a, +\right), a', -\right)\right)\right\}\right]\\
&< \mathbf{c}\mathbf{X}^{\mathrm{T}} + \beta\left[\left\{\mathrm{E}\left[\mu_{a'}|\sigma\left(\mathbf{b}, a, +\right)\right]\mathrm{V}_t\left(\mathbf{X} - \mathbf{e}_{a'}, \sigma\left(\sigma\left(\mathbf{b}, a, +\right), a', +\right)\right)\right.\right.\\
&\left.\left.+\ \left(1 - \mathrm{E}\left[\mu_{a'}|\sigma\left(\mathbf{b}, a, +\right)\right]\right)\mathrm{V}_t\left(\mathbf{X}, \sigma\left(\sigma\left(\mathbf{b}, a, +\right), a', -\right)\right)\right\}\right]\\
&\leq \mathbf{c}\mathbf{X}^{\mathrm{T}} + \beta\left[\left\{\mathrm{E}\left[\mu_{a'}|\sigma\left(\mathbf{b}, a, -\right)\right]\mathrm{V}_t\left(\mathbf{X} - \mathbf{e}_{a'}, \sigma\left(\sigma\left(\mathbf{b}, a, -\right), a', +\right)\right)\right.\right.\\
&\left.\left.+\ \left(1 - \mathrm{E}\left[\mu_{a'}|\sigma\left(\mathbf{b}, a, -\right)\right]\right)\mathrm{V}_t\left(\mathbf{X}, \sigma\left(\sigma\left(\mathbf{b}, a, -\right), a', -\right)\right)\right\}\right]\\
&= \mathrm{V}_{t+1}\left(\mathbf{X}, \sigma\left(\mathbf{b}, a, -\right)\right),
\end{aligned}$$

since $\mathrm{E}\left[\mu_{a'}|\sigma\left(\mathbf{b}, a, -\right)\right] \leq \mathrm{E}\left[\mu_{a'}|\sigma\left(\mathbf{b}, a, +\right)\right]$, by the inductive hypothesis, and since $\mathrm{V}_t\left(\mathbf{X} - \mathbf{e}_a, \mathbf{b}\right) < \mathrm{V}_t\left(\mathbf{X}, \mathbf{b}\right)$ by Lemma A.1. For the second portion of the proof, it is easy to see that the priority discipline would choose a' until none of the class remain in the system, so the proof remains the same by substituting V_t^{π} in place of V_t. □

Proof of Proposition 2.1. To prove that the minimax policy is associated with the $c\mu$ policy that prioritizes $\arg\max_{a\in\mathcal{A}(\mathbf{X})}\min_{j\in\mathcal{J}_a} c_a\mu_{a,j}$, first we show that for *any* priority policy π and belief $\mathbf{b}$, $\lambda \in [0,1]$, $\mathrm{V}^{\pi}(\mathbf{X}, \lambda\mathbf{b}_0 + (1-\lambda)\mathbf{b})$ is nondecreasing as λ increases.

Choosing $\delta > 0$ such that $\lambda - \delta > 0$, we proceed by induction on t. In the base case, when $t = 1$, the assertion is true since $\mathrm{V}_1^{\pi}(\mathbf{X}, (\lambda-\delta)\mathbf{b}_0 + (1-\lambda+\delta)\mathbf{b}) \leq \mathrm{V}_1^{\pi}(\mathbf{X}, \lambda\mathbf{b}_0 + (1-\lambda)\mathbf{b})$ since $\mathrm{E}[\mu_a|\lambda\mathbf{b}_0 + (1-\lambda)\mathbf{b}] \leq \mathrm{E}[\mu_a|(\lambda-\delta)\mathbf{b}_0 + (1-\lambda+\delta)\mathbf{b}]$ for all $a \in \mathcal{A}(\mathbf{X})$. For the inductive step, we suppose the assertion holds for t. Then, suppose that the action chosen by $\pi(\mathbf{X}, \lambda\mathbf{b}_0 + (1-\lambda)\mathbf{b}) = a$. Then,

$$
\begin{aligned}
&\mathrm{V}^{\pi}_{t+1}\left(\mathbf{X}, (\lambda-\delta)\mathbf{b}_0 + (1-\lambda+\delta)\mathbf{b}\right)\\
&= \mathbf{c}\mathbf{X}^{\mathrm{T}} + \beta\Bigg[\Big\{\mathrm{E}\left[\mu_a|(\lambda-\delta)\mathbf{b}_0 + (1-\lambda+\delta)\mathbf{b}\right]\mathrm{V}^{\pi}_{t}\left(\mathbf{X}-\mathbf{e}_a, \sigma\left((\lambda-\delta)\mathbf{b}_0 + (1-\lambda+\delta)\mathbf{b}, a, +\right)\right)\\
&+\ \left(1-\mathrm{E}\left[\mu_a|(\lambda-\delta)\mathbf{b}_0 + (1-\lambda+\delta)\mathbf{b}\right]\right)\mathrm{V}^{\pi}_{t}\left(\mathbf{X}, \sigma\left((\lambda-\delta)\mathbf{b}_0 + (1-\lambda+\delta)\mathbf{b}, a, -\right)\right)\Big\}\Bigg]\\
&\leq \mathbf{c}\mathbf{X}^{\mathrm{T}} + \beta\Bigg[\Big\{\mathrm{E}\left[\mu_a|\lambda\mathbf{b}_0 + (1-\lambda)\mathbf{b}\right]\mathrm{V}^{\pi}_{t}\left(\mathbf{X}-\mathbf{e}_a, \sigma\left(\lambda\mathbf{b}_0 + (1-\lambda)\mathbf{b}, a, +\right)\right)\\
&+\ \left(1-\mathrm{E}\left[\mu_a|\lambda\mathbf{b}_0 + (1-\lambda)\mathbf{b}\right]\right)\mathrm{V}^{\pi}_{t}\left(\mathbf{X}, \sigma\left(\lambda\mathbf{b}_0 + (1-\lambda)\mathbf{b}, a, -\right)\right)\Big\}\Bigg]\\
&= \mathrm{V}^{\pi}_{t+1}\left(\mathbf{X}, \lambda\mathbf{b}_0 + (1-\lambda)\mathbf{b}\right),
\end{aligned}
$$

because

$$\mathrm{V}^{\pi}_{t}\left(\mathbf{X}, \sigma\left((\lambda-\delta)\mathbf{b}_0 + (1-\lambda+\delta)\mathbf{b}, a, \theta\right)\right) \leq \mathrm{V}_{t}\left(\mathbf{X}, \sigma\left(\delta\mathbf{b}_0 + (1-\lambda)\mathbf{b}, a, \theta\right)\right),$$

by the inductive hypothesis, and since

$$\mathrm{V}^{\pi}_{t}\left(\mathbf{X}-\mathbf{e}_a, \sigma\left(\mathbf{b}, a, +\right)\right) \leq \mathrm{V}^{\pi}_{t}\left(\mathbf{X}, \sigma\left(\mathbf{b}, a, -\right)\right),$$

by Lemma A.2. Therefore, noting that $\mathrm{R}^{\pi}\left(\mathbf{X}\right) = \mathrm{V}^{\pi}\left(\mathbf{X}, \mathbf{b}_0\right)$, and since our system is identical to that of Buyukkoc *et al.* (1985) when the belief is composed of only zeros and ones, the $c\mu$ policy that prioritizes $\arg\max_{a\in\mathcal{A}(\mathbf{X})}\min_{j\in\mathcal{J}_a} c_a\mu_{a,j}$ is optimal for the minimax objective.

Similarly, to prove that the minimin policy is associated with the $c\mu$ policy that prioritizes $\arg\max_{a\in\mathcal{A}(\mathbf{X})}\max_{j\in\mathcal{J}_a} c_a\mu_{a,j}$, let $\mathbf{b}_1$ be the belief with components

$$b^1_{i,j} = \begin{cases} 1 & : \text{ if } \mu_{i,j} = \max_{k\in\mathcal{J}_i}\mu_{i,k} \\ 0 & : \text{ otherwise.} \end{cases}$$

First we show that for *any* priority policy π and belief $\mathbf{b}$, $\lambda \in [0,1]$, $\mathrm{V}^{\pi}(\mathbf{X}, \lambda\mathbf{b}_1 + (1-\lambda)\mathbf{b})$ is nonincreasing as λ increases.

Choosing δ such that $\lambda - \delta > 0$, we proceed by induction on t. In the base case, when $t = 1$, the assertion is true since $\mathrm{V}_1^\pi(\mathbf{X}, (\lambda - \delta)\mathbf{b}_1 + (1 - \lambda + \delta)\mathbf{b}) \geq \mathrm{V}_1^\pi(\mathbf{X}, \lambda\mathbf{b}_1 + (1 - \lambda)\mathbf{b})$ since $\mathrm{E}[\mu_a|\lambda\mathbf{b}_1 + (1 - \lambda)\mathbf{b}] \geq \mathrm{E}[\mu_a|(\lambda - \delta)\mathbf{b}_1 + (1 - \lambda + \delta)\mathbf{b}]$ for all $a \in \mathcal{A}(\mathbf{X})$. For the inductive step, we suppose the assertion holds for t. Then, suppose that the action chosen by $\pi(\mathbf{X}, \lambda\mathbf{b}_1 + (1 - \lambda)\mathbf{b}) = a$. Then,

$$\begin{aligned}
&\mathrm{V}_{t+1}^\pi\left(\mathbf{X}, (\lambda - \delta)\mathbf{b}_1 + (1 - \lambda + \delta)\mathbf{b}\right) \\
&= \mathbf{c}\mathbf{X}^{\mathrm{T}} + \beta\Bigg[\Big\{\mathrm{E}\left[\mu_a|(\lambda - \delta)\mathbf{b}_1 + (1 - \lambda + \delta)\mathbf{b}\right]\mathrm{V}_t^\pi\left(\mathbf{X} - \mathbf{e}_a, \sigma\left((\lambda - \delta)\mathbf{b}_1 + (1 - \lambda + \delta)\mathbf{b}, a, +\right)\right) \\
&+ \left(1 - \mathrm{E}\left[\mu_a|(\lambda - \delta)\mathbf{b}_1 + (1 - \lambda + \delta)\mathbf{b}\right]\right)\mathrm{V}_t^\pi\left(\mathbf{X}, \sigma\left((\lambda - \delta)\mathbf{b}_1 + (1 - \lambda + \delta)\mathbf{b}, a, -\right)\right)\Big\}\Bigg] \\
&\geq \mathbf{c}\mathbf{X}^{\mathrm{T}} + \beta\Bigg[\Big\{\mathrm{E}\left[\mu_a|\lambda\mathbf{b}_1 + (1 - \lambda)\mathbf{b}\right]\mathrm{V}_t^\pi\left(\mathbf{X} - \mathbf{e}_a, \sigma\left(\lambda\mathbf{b}_1 + (1 - \lambda)\mathbf{b}, a, +\right)\right) \\
&+ \left(1 - \mathrm{E}\left[\mu_a|\lambda\mathbf{b}_1 + (1 - \lambda)\mathbf{b}\right]\right)\mathrm{V}_t^\pi\left(\mathbf{X}, \sigma\left(\lambda\mathbf{b}_1 + (1 - \lambda)\mathbf{b}, a, -\right)\right)\Big\}\Bigg] \\
&= \mathrm{V}_{t+1}^\pi\left(\mathbf{X}, \lambda\mathbf{b}_1 + (1 - \lambda)\mathbf{b}\right),
\end{aligned}$$

because

$$\mathrm{V}_t^\pi\left(\mathbf{X}, \sigma\left((\lambda - \delta)\mathbf{b}_1 + (1 - \lambda + \delta)\mathbf{b}, a, \theta\right)\right) \geq \mathrm{V}_t\left(\mathbf{X}, \sigma\left(\lambda\mathbf{b}_1 + (1 - \lambda)\mathbf{b}, a, \theta\right)\right)$$

by the inductive hypothesis, and since

$$\mathrm{V}_t^\pi\left(\mathbf{X} - \mathbf{e}_a, \sigma\left(\mathbf{b}, a, +\right)\right) \leq \mathrm{V}_t^\pi\left(\mathbf{X}, \sigma\left(\mathbf{b}, a, -\right)\right),$$

by Lemma A.2. By the same argument as we used for the minimax portion of the proof while noting that $\mathrm{N}^\pi\left(\mathbf{X}\right) = \mathrm{V}^\pi\left(\mathbf{X}, \mathbf{b}_1\right)$, and since our system is identical to that of Buyukkoc *et al.* (1985) when the belief is composed of only zeros and ones, the $c\mu$ policy that prioritizes $\arg\max_{a \in \mathcal{A}(\mathbf{X})} \max_{j \in \mathcal{J}_a} c_a\mu_{a,j}$ is optimal for the minimin objective. □

Proof of Proposition 2.2. For the first half of the proposition, we must relate the robust objective and the percentile objective with ϵ set to zero. Note that,

$$\mathrm{Y}\left(\mathbf{X}, 0\right) = \inf_{\pi \in \Pi}\left\{\inf_{y_\epsilon \in [\mathrm{N}^\pi(\mathbf{X}), \mathrm{R}^\pi(\mathbf{X})]}\left\{y_\epsilon|\mathbb{P}\left(\mathrm{V}^\pi\left(\mathbf{X}, \mathbf{B}\right) \leq y_\epsilon\right) = 1\right\}\right\},$$

which implies that any y_ϵ chosen with respect to some policy π must be greater than or equal to every $\mathrm{V}^\pi(\mathbf{X}, \mathbf{b})$ for $\mathbf{b} \in \mathcal{B}$ since $\mathbb{P}_{\mathbf{B}}(\mathbf{B} = \mathbf{b}) > 0$. The value y_ϵ is being minimized, so for any given policy we must choose $y_\epsilon = \max_{\mathbf{b}\in\mathcal{B}} \{\mathrm{V}^\pi(\mathbf{X}, \mathbf{b})\}$. Substituting this back into the original percentile objective we obtain

$$\inf_{\pi\in\Pi} \left\{ \max_{\mathbf{b}\in\mathcal{B}} \{\mathrm{V}^\pi(\mathbf{X}, \mathbf{b})\} \right\},$$

which is the same as $\mathrm{R}(\mathbf{X})$.

For the second half of the proposition, we note that

$$\mathrm{Y}(\mathbf{X}, 1) = \inf_{\pi\in\Pi} \left\{ \inf_{y_\epsilon \in [\mathrm{N}^\pi(\mathbf{X}), \mathrm{R}^\pi(\mathbf{X})])} \{y_\epsilon | \mathbb{P}_{\mathbf{B}}(\mathrm{V}^\pi(\mathbf{X}, \mathbf{b}) \le y_\epsilon) \ge 0\} \right\}$$

Thus, it is easy to see that setting $y_\epsilon = \mathrm{N}(\mathbf{X})$ paired with the nominal priority policy π satisfies the probability constraint since $\mathbb{P}_{\mathbf{B}}(\cdot \le y_\epsilon) \ge 0$ is satisfied for all real numbers. Furthermore, $\mathrm{N}(\mathbf{X})$ is the smallest value any percentile objective may obtain, and hence, the proof is complete. □

Lemma A.3 (**Non-Robust MAB Formulation**). *The dynamic program* (2.4) *has an equivalent Multi-Armed Bandit (MAB) formulation as $\psi \to \infty$.*

Proof. In order to show that (2.4) can be formulated as a MAB as the observation rate goes to infinity, we simply show that it is analogous to the traditional reward-based MAB dynamic programming formulation in which only the active class generates rewards and experiences state transitions, while all other customer classes remain frozen. We can achieve this by noting that a customer of class $i \in \mathcal{N}$ incurs a total discounted cost $\sum_{t=0}^{\tau} \beta^\tau c_i$, where τ is the period in which the customer is served, thus leaving the system. Therefore, an equivalent representation of our discrete-time objective is given by

$$\inf_{\pi\in\Pi} \left\{ \sum_{t=0}^{\infty} \beta^t \left(\mathbf{c}\mathbf{X}^{\mathrm{T}} - \mathrm{E}\left[\sum_{i=1}^{n} \sum_{k=1}^{\infty} \beta^k W_{i,k}^\pi \middle| \mathbf{X}(0), \mathbf{b}(0) \right] \right) \right\},$$

where $W_{i,k}^{\pi}$ is the random variable taking value $\sum_{t=0}^{\infty}\beta^t c_i = \frac{c_i}{1-\beta}$ if a customer from class i is served at time k under policy π and is zero otherwise. Therefore, we can write,

$$\mathrm{V}\left(\mathbf{X},\mathbf{b}\right) = \frac{\mathbf{c}\mathbf{X}^{\mathrm{T}}}{1-\beta} - \hat{\mathrm{W}}\left(\mathbf{X},\mathbf{b},0\right),$$

where,

$$\begin{aligned}\hat{\mathrm{W}}\left(\mathbf{X},\mathbf{b},i\right) = \frac{c_i}{1-\beta} + \max_{a\in\mathcal{N}}\beta\Big[&\mathbb{1}\left\{X_a>0\right\}\Big[\mathrm{E}[\mu_a|\mathbf{b}]\hat{\mathrm{W}}\left(\mathbf{X}-\mathbf{e}_a,\sigma\left(\mathbf{b},+,a\right),a\right)\\ &+\ \left(1-\mathrm{E}[\mu_a|\mathbf{b}]\right)\hat{\mathrm{W}}\left(\mathbf{X},\sigma\left(\mathbf{b},-,a\right),0\right)\Big]+\mathbb{1}\left\{X_a=0\right\}\hat{\mathrm{W}}\left(\mathbf{X},\mathbf{b},0\right)\Big].\end{aligned}$$

Here, $\mathbb{1}$ is the indicator function and we remind the reader that $c_0 = 0$ since it is the class that enables "idling" policies.

The above system offers rewards immediately after completing service to a customer. Let us consider an alternative system where the reward for the ith customer of class a is given immediately after the service to the $i+1$ customer of class a is initiated. This can be described by the dynamic program:

$$\begin{aligned}\mathrm{W}\left(\mathbf{X},\mathbf{b},\mathbf{Y}\right) = \max_{a\in\mathcal{N}}\mathbb{1}\left\{Y_a=1\right\}&\left(\frac{c_i}{1-\beta}\right)\\ &+\beta\left[\mathbb{1}\left\{X_a>0\right\}\left[\mathrm{E}[\mu_a|\mathbf{b}]\mathrm{W}\left(\mathbf{X}-\mathbf{e}_a,\sigma\left(\mathbf{b},+,a\right),\mathbf{Y}+\mathbb{1}\left\{Y_a=0\right\}\mathbf{e}_a\right)\right.\right.\\ &+\ \left.\left.\left(1-\mathrm{E}[\mu_a|\mathbf{b}]\right)\mathrm{W}\left(\mathbf{X},\sigma\left(\mathbf{b},-,a\right),\mathbf{Y}-\mathbb{1}\left\{Y_a=1\right\}\mathbf{e}_a\right)\right]\right].\\ &+\mathbb{1}\left\{X_a=0\right\}\mathrm{W}\left(\mathbf{X},\mathbf{b},\mathbf{Y}-\mathbb{1}\left\{Y_a=1\right\}\mathbf{e}_a\right)].\end{aligned}$$

This seemingly more complicated dynamic program relies on an extra state $\mathbf{Y}\in\mathbb{Z}_+^n$ with elements $Y_a\in\{0,1\}$. We regard this state as a "primer" indicator. If a customer from class a has completed service and is just waiting for the server to begin service to the following customer of its class, then $Y_a = 1$. Otherwise it is zero. Note that as the observations become continuous, $\hat{\mathrm{W}}$ and W become identical. This is because the policy for W that serves customer a whenever $Y_a = 1$, and otherwise serves the class that $\hat{\mathrm{W}}$ would serve can only differ from each other finitely many (namely $\mathbf{X}\mathbf{1}^{\mathrm{T}}$) times while the probability of service during those periods where $\mathbf{Y}\mathbf{1}^{\mathrm{T}} > 0$ becomes arbitrarily small (we assume the decision-maker begins with $\mathbf{Y} = \mathbf{0}$). Thus, as observations become continuous, the cost difference for policies that differ finitely many times goes to zero.

Since under action a, the only rewards generated are associated with this class and all other classes remain frozen in state and rewards, this dynamic program is a MAB. □

Since we have shown that we can express our problem as a MAB when the observation rate is appropriately fast, we aspire to calculate the Gittins index introduced by Gittins (1979) which provides an optimal policy for the system. The numerator of this index for action a associated with $\mathrm{W}(\mathbf{X}, \mathbf{b}, \mathbf{0})$ is identical to the value obtained by repeatedly serving class a and completely ignoring other classes. This value can be calculated via the dynamic program,

$$\begin{aligned}\mathrm{U}_{t+1}(X_a, \mathbf{b}, a) = \beta \Bigg(\mathrm{E}\left[\mu_a|\mathbf{b}\right] \left(\mathrm{U}_t\left(X_a - 1, \sigma(\mathbf{b}, a, +), a\right) + \frac{c_a}{1-\beta} \right) \\ + \left(1 - \mathrm{E}\left[\mu_a|\mathbf{b}\right]\right) \mathrm{U}_t\left(X_a, \sigma(\mathbf{b}, a, -), a\right) \Bigg),\end{aligned}$$

where $\mathrm{U}_0(X_a, \mathbf{b}, a) = \mathrm{U}_t(0, \mathbf{b}, a) = 0$. We wish to show that these values have an easily calculable closed form. To this end, we define,

$$g(t, X_a)_{a,l} = \begin{cases} \sum_{k=0}^{t-1} \beta^k + \sum_{j=X_a}^{t-1} \left[(-1)^{j+X_a+1} \mu_{a,l}^{j} \binom{j-1}{X_a-1} \left(\sum_{k=j}^{t-1} \binom{k}{j} \beta^k \right) \right] & : 0 < X_a < t \\ \sum_{k=0}^{t-1} \beta^k & : 0 < t \leq X_a \\ 0 & : X_a = 0, \end{cases} \tag{A.7}$$

which will help with this task.

Lemma A.4 (**Closed Form U_t Representation**).
$\mathrm{U}_t(X_a, \mathbf{b}, a) = \frac{c_a \beta}{1-\beta} \sum_{j=1}^{m_a} b_{a,j} \mu_{a,j} g(t, X_a)_{a,j}$.

Proof. We accomplish the proof via induction. Assuming $X_a > 0$, for the base case, when $t = 1$ we have

$$\begin{aligned}
\mathrm{U}_1(X_a, \mathbf{b}, a) &= \beta \left(\mathrm{E}\left[\mu_a|\mathbf{b}\right] \left(\mathrm{U}_0\left(X_a - 1, \sigma(\mathbf{b}, +, a), a\right) + \frac{c_a}{1-\beta} \right) \right.\\
&\quad \left. + \left(1 - \mathrm{E}\left[\mu_a|\mathbf{b}\right]\right) \mathrm{U}_0\left(X_a, \sigma(\mathbf{b}, -, a), a\right) \right)\\
&= \frac{c_a\beta}{1-\beta}\mathrm{E}[\mu_a|\mathbf{b}]\\
&= \frac{c_a\beta}{1-\beta}\sum_{j=1}^{m_a} \mu_{a,j} b_{a,j} g(1, X_a)_{a,j}.
\end{aligned}$$

For the inductive step,

$$\begin{aligned}
\mathrm{U}_{t+1}(X_a, \mathbf{b}, a) &= \beta \left(\mathrm{E}\left[\mu_a|\mathbf{b}\right] \left(\mathrm{U}_t\left(X_a - 1, \sigma(\mathbf{b}, +, a), a\right) + \frac{c_a}{1-\beta} \right) \right.\\
&\quad \left. + \left(1 - \mathrm{E}\left[\mu_a|\mathbf{b}\right]\right) \mathrm{U}_t\left(X_a, \sigma(\mathbf{b}, -, a), a\right) \right)\\
&= \beta \left(\mathrm{E}[\mu_a|\mathbf{b}] \left(\frac{c_a}{1-\beta}\beta \sum_{j=1}^{m_a} \mu_{a,j} \frac{\mu_{a,j} b_{a,j}}{\mathrm{E}[\mu_a|\mathbf{b}]} g(t-1, X_a - 1)_{a,j} + \frac{c_a}{1-\beta} \right) \right.\\
&\quad \left. + (1 - \mathrm{E}[\mu_a|\mathbf{b}]) \frac{c_a}{1-\beta}\beta \sum_{j=1}^{m_a} \mu_{a,j} \frac{(1-\mu_{a,j}) b_{a,j}}{1 - \mathrm{E}[\mu_a|\mathbf{b}]} g(t-1, X_a)_{a,j} \right)\\
&= \beta \left(\frac{c_a}{1-\beta}\mathrm{E}[\mu_a|\mathbf{b}] + \frac{c_a\beta}{1-\beta} \sum_{j=1}^{m_a} \mu_{a,j}^2 b_{a,j} g(t-1, X_a - 1)_{a,j} \right.\\
&\quad \left. + \frac{c_a\beta}{1-\beta} \left(- \sum \mu_{a,j}^2 b_{a,j} g(t-1, X_a)_{a,j} + \sum_{j=1}^{m_a} \mu_{a,j} b_{a,j} g(t-1, X_a)_{a,j} \right) \right)\\
&= \frac{c_a\beta}{1-\beta} \left(\mathrm{E}[\mu_a|\mathbf{b}] + \left(\sum_{j=1}^{m_a} \mu_{a,j} b_{a,j} \left(\mu_{a,j} [g(t-1, X_a - 1)_{a,j} - g(t-1, X_a)_{a,j}] \right.\right.\right.\\
&\quad \left.\left.\left. + g(t-1, X_a)_{a,j} \right)\right)\right) \quad \text{(A.8)}\\
&= \frac{c_a\beta}{1-\beta} \left(\mathrm{E}[\mu_a|\mathbf{b}] + \left(\sum_{j=1}^{m_a} \mu_{a,j} b_{a,j} \frac{(g(t, X_a)_{a,j} - 1)}{\beta} \right) \right)\\
&\qquad \text{(A.9)}\\
&= \frac{c_a\beta}{1-\beta} \sum_{j=1}^{m_a} \mu_{a,j} b_{a,j} g(t, X_a)_{a,j},
\end{aligned}$$

by the identity $\beta\mu_{a,j}\left(g(t-1, X_a - 1)_{a,j} - g(t-1, X_a)_{a,j}\right) + \beta g(t-1, X_a)_{a,j}) + 1 = g(t, X_a)_{a,j}$ used from (A.8) to (A.9) which is proven in Lemma A.5. □

LEMMA A.5. $g(t, X_a)_{a,j} =$

$\beta\mu_{a,j}\left(g(t-1, X_a-1)_{a,j} - g(t-1, X_a)_{a,j}\right) + \beta g(t-1, X_a)_{a,j}) + 1.$

Proof. Ignoring the subscripts for the entirety of the proof, we begin with the base case, where $t = 2$. When $X_a = 1$, we have

$$\beta\mu\left(g(1,0) - g(1,1)\right) + \beta g(1,1)) + 1 = -\mu\beta + \beta + 1 = g(2,1).$$

Otherwise, with $X_a > 1$, we have

$$\beta\mu\left(g(1, X_a-1) - g(1, X_a)\right) + \beta g(1, X_a)) + 1 = \beta + 1 = g(2, X_a).$$

When $t > 2$, in the case where $X_a = 1$, we obtain

$$\begin{aligned}
&\beta\mu\left(g(t-1,0) - g(t-1,1)\right) + \beta g(t-1,1)) + 1 = -\beta\mu g(t-1,1)) + \beta g(t-1,1) + 1 \\
&\quad = 1 + \sum_{k=0}^{t-2}\beta^{k+1} + \sum_{j=1}^{t-2}\left[(-1)^j\mu^j\left(\sum_{k=j}^{t-2}\binom{k}{j}\beta^{k+1}\right)\right] - \sum_{j=0}^{t-2}\left[(-1)^j\mu^{j+1}\left(\sum_{k=j}^{t-2}\binom{k}{j}\beta^{k+1}\right)\right] \\
&\quad = \sum_{k=0}^{t-1}\beta^k + \sum_{j=1}^{t-2}\left[(-1)^j\mu^j\left(\sum_{k=j}^{t-1}\binom{k}{j}\beta^{k+1}\right)\right] - \sum_{j=0}^{t-2}\left[(-1)^j\mu^{j+1}\left(\sum_{k=j}^{t-2}\binom{k}{j}\beta^{k+1}\right)\right].
\end{aligned}$$

If we examine the coefficients to μ^j, they are

$$\begin{aligned}
(-1)^j\left(\sum_{k=j-1}^{t-2}\binom{k}{j-1}\beta^{k+1} + \sum_{k=j}^{t-2}\binom{k}{j}\beta^{k+1}\right) &= (-1)^j\left(\beta^j + \sum_{k=j}^{t-2}\left(\binom{k}{j-1} + \binom{k}{j}\right)\beta^{k+1}\right) \\
&= (-1)^j\sum_{k=j}^{t-1}\binom{k}{j}\beta^k,
\end{aligned}$$

since $\binom{j}{i} + \binom{j}{i+1} = \binom{j+1}{i+1}$. This proves the assertion for $X_a = 1$ if we compare to $g(t,1)$. If $X_a > 1$,

$$\begin{aligned}
&\beta\mu\left(g(t-1, X_a-1) - g(t-1, X_a)\right) + \beta g(t-1, X_a)) + 1 \\
&\quad = \sum_{k=0}^{t-1}\beta^k + \sum_{j=X_a}^{t-2}\left[(-1)^{j+X_a+1}\mu^j\binom{j-1}{X_a-1}\left(\sum_{k=j}^{t-2}\binom{k}{j}\beta^{k+1}\right)\right] \\
&\quad + \sum_{j=X_a-1}^{t-2}\left[(-1)^{j+X_a}\mu^{j+1}\binom{j-1}{X_a-2}\left(\sum_{k=j}^{t-2}\binom{k}{j}\beta^{k+1}\right)\right]
\end{aligned}$$

$$- \sum_{j=X_a}^{t-2} \left[(-1)^{j+X_a+1} \mu^{j+1} \binom{j-1}{X_a-1} \left(\sum_{k=j}^{t-2} \binom{k}{j} \beta^{k+1} \right) \right].$$

If we examine the coefficients of μ^j, they are

$$\begin{aligned}
(-1)^{j+X_a+1}\binom{j-1}{X_a-1}\sum_{k=j}^{t-2}\binom{k}{j}\beta^{k+1} &+ (-1)^{j+X_a-1}\binom{j-2}{X_a-2}\sum_{k=j-1}^{t-2}\binom{k}{j-1}\beta^{k+1} \\
&+ (-1)^{j+X_a-1}\binom{j-2}{X_a-1}\sum_{k=j-1}^{t-2}\binom{k}{j-1}\beta^{k+1} \\
&= (-1)^{j+X_a+1}\left(\binom{j-1}{X_a-1}\sum_{k=j}^{t-2}\binom{k}{j}\beta^{k+1}\right. \\
&\left.+\binom{j-1}{X_a-1}\sum_{k=j-1}^{t-2}\binom{k}{j-1}\beta^{k+1}\right) \\
&= (-1)^{j+X_a-1}\binom{j-1}{X_a-1}\left(\sum_{k=j}^{t-2}\binom{k+1}{j}\beta^{k+1}+\beta^j\right) \\
&= (-1)^{j+X_a-1}\binom{j-1}{X_a-1}\sum_{k=j}^{t-1}\binom{k}{j}\beta^k,
\end{aligned}$$

which concludes our lemma if we examine compare this to $g(t, X_a)$. □

Lemma A.6. *Function $\sum_{j=i}^{t}(-1)^{j+i+1}\binom{j-1}{i-1}\binom{t}{j}x^j$ with $x \in [0,1]$ is non-positive and decreasing in x and t.*

Proof. We begin by proving that the function is negative and decreasing in x by taking the first derivative with respect to x,

$$\begin{aligned}
\frac{d}{dx}\sum_{j=i}^{t}(-1)^{j+i+1}\binom{j-1}{i-1}\binom{t}{j}x^j &= \sum_{j=i}^{t}(-1)^{i+j+1}\binom{j-1}{i-1}\binom{t}{j}jx^{j-1} \\
&= \sum_{j=i}^{t}(-1)^{i+j+1}\binom{j}{i}\binom{t}{j}x^{j-1} \\
&= \binom{t}{i}\sum_{j=i}^{t}(-1)^{i+j+1}\binom{t-i}{j-i}x^{j-1} \\
&= \binom{t}{i}(1-x)^{t-i}x^{i-1}(-1)^{2i+1},
\end{aligned}$$

which only has zeros at 1 and 0. Furthermore, since the x^i term is negative, and for small enough x, dominates the rest of the polynomial, we know that the derivative of the function is always less than 0, which proves that the function is decreasing in x over $[0,1]$. Furthermore, since the function evaluated at $x = 0$ is zero, the function must be non-positive.

To prove the function is decreasing in t, we note that we can express the function as

$$-x^i \binom{t}{i} {}_2F_1\left(1, i-t; 1+i; -x\right),$$

where ${}_2F_1$ is the Gaussian or ordinary hypergeometric function. Now, the derivative of this function with respect to t is

$$\frac{d}{dt} - x^i \binom{t}{i} {}_2F_1\left(1, i-t; 1+i; -x\right) = x^i \binom{t}{k} \Bigg({}_2F_1\left(1, i-t; 1+i; -x\right) \left(-\sum_{j=1}^{t} \frac{1}{j} + \sum_{j=1}^{t-i} \frac{1}{j}\right) + \frac{d}{dt} {}_2F_1\left(1, i-t; 1+i; -x\right) \Bigg).$$

Now, since $-x^i \binom{t}{i} {}_2F_1\left(1, i-t; 1+i; -x\right) < 0$ as shown by the previous portion of the proof, ${}_2F_1\left(1, i-t; 1+i; -x\right) > 0$, hence ${}_2F_1\left(1, i-t; 1+i; -x\right) \left(-\sum_{j=1}^{t} \frac{1}{j} + \sum_{j=1}^{t-i} \frac{1}{j}\right) < 0$. Now we need to show that ${}_2F_1\left(1, i-t-1; 1+i; -x\right) - {}_2F_1\left(1, i-t; 1+i; -x\right)$ is negative to show that the second part of the equation is also negative. Using *Gauss' continued fraction representation* (as seen in equation (3) of Karp and Sitnik (2009)),

$$\begin{aligned} {}_2F_1\left(1, i-t-1; 1+i; -x\right) - {}_2F_1\left(1, i-t; 1+i; -x\right) = (1-1) + \left(\frac{(i-t-1)i - (i-t)i}{1+i} x\right) \\ + \ldots + \frac{(i-t-1+l)(i+l) - (i-t+l)(i+l)}{1+i+2l} x \\ + \frac{(l+1)(1+t+1+l) - (l+1)(1+t+l)}{1+i+2l+1} x + \ldots, \end{aligned}$$

and since

$$\begin{aligned} \frac{(i-t-1+l)(i+l) - (i-t+l)(i+l)}{1+i+2l} x + \frac{(l+1)(1+t+1+l) - (l+1)(1+t+l)}{1+i+2l+1} x \\ = \frac{-i-l}{1+i+2l} x + \frac{l+1}{2+i+2l} x < 0, \end{aligned}$$

the proof is complete. □

Proof of Theorem 2.1. The Gittins index associated with $\mathrm{W}\left(\mathbf{X}, \mathbf{b}, \mathbf{0}\right)$ with respect to action a is given by,

$$\max_{\tau \geq 0} \left\{ \frac{\mathrm{U}_\tau(X_a, \mathbf{b}, a)}{\sum_{t=0}^{\tau} \beta^t} \right\} = \max_{\tau \geq 0} \left\{ \frac{\frac{c_a \beta}{1-\beta} \sum_{j=1}^{m_a} \mu_{a,j} b_{a,j} g(t, X_a)_{a,j}}{\sum_{t=0}^{\tau} \beta^t} \right\}.$$

Now, defining

$$h(\mathbf{X}, a, l, t) = \begin{cases} \frac{c_a \mu_{a,l} \left(\sum_{k=0}^{t-1} \beta^k + \sum_{j=X_a}^{t-1} (-1)^{j+X_a+1} \mu_{a,l}^{j} \binom{j-1}{X_a-1} \left(\sum_{k=j}^{t-1} \binom{k}{j} \beta^k \right) \right)}{\sum_{k=0}^{t} \beta^k} & : X_a < t \\ c_a \mu_{a,l} \frac{\sum_{k=0}^{t-1} \beta^k}{\sum_{k=0}^{t} \beta^k} & : X_a \geq t \end{cases} \tag{A.10}$$

serving the Gittins index rule is equivalent to serving

$$\arg\max_{a \in \mathcal{A}(\mathbf{X})} \sum_{l=1}^{m_a} b_{a,l} h(\mathbf{X}, a, l, t)$$

We first show that the index associated with g is bounded by $\frac{c_a \beta}{1-\beta} \mathrm{E}[\mu_a | \mathbf{b}]$. We examine the difference

$$\begin{aligned} &\mathrm{U}_{\tau+1}(X_a, \mathbf{b}, a) - \mathrm{U}_\tau(X_a, \mathbf{b}, a) \\ &\qquad = \frac{\beta c_a}{1-\beta} \sum_{j=1}^{m_a} \mu_{a,j} b_{a,j} \left(\beta^\tau \sum_{k=X_a}^{\tau} \binom{k-1}{X_a-1} \binom{\tau}{k} (-1)^{k+X_a+1} \mu_{a,j}^k + \beta^\tau \right) \\ &\qquad = \frac{\beta c_a}{1-\beta} \sum_{j=1}^{m_a} \mu_{a,j} b_{a,j} \left(\beta^\tau \left(\sum_{k=X_a}^{\tau} \binom{k-1}{X_a-1} \binom{\tau}{k} (-1)^{k+X_a+1} \mu_{a,j}^k + 1 \right) \right). \end{aligned}$$

Now, as seen in Lemma A.6, we have

$$\sum_{k=X_a}^{\tau} \binom{k-1}{X_a-1} \binom{\tau}{k} (-1)^{k+X_a+1} \mu_{a,j}^k + 1 < 1, \tag{A.11}$$

which implies that an upper bound of the index is found by,

$$\frac{\mathrm{U}_\tau(X_a, \mathbf{b}, a)}{\sum_{t=0}^{\tau} \beta^t} \leq \frac{\frac{c_a \beta}{1-\beta} \mathrm{E}[\mu_a | \mathbf{b}] \sum_{t=0}^{\tau} \beta^t}{\sum_{t=0}^{\tau} \beta^t} = \frac{c_a \beta}{1-\beta} \mathrm{E}[\mu_a | \mathbf{b}].$$

This proves property that the indices stemming from h are bounded by $c_a \mathrm{E}[\mu_a|\mathbf{b}]$. Furthermore, this bound implies that

$$\sum_{j=X_a}^{t-1} (-1)^{j+X_a+1} \mu_{a,l}^{j} \binom{j-1}{X_a-1} \left(\sum_{k=j}^{t-1} \binom{k}{j} \beta^k \right) < \beta^t,$$

since otherwise the bound would be attained or exceeded.

Now, when we write h in terms of the uniformization rate (when $X_a < t$),

$$\frac{\frac{\hat{c}_a}{\psi+\alpha} \left(\frac{\hat{\mu}_{a,l}}{\psi}\right) \left(\sum_{k=0}^{t-1} \left(\frac{\psi}{\psi+\alpha}\right)^k + \sum_{j=X_a}^{t-1} (-1)^{j+X_a+1} \left(\frac{\hat{\mu}_{a,l}}{\psi}\right)^j \binom{j-1}{X_a-1} \left(\sum_{k=j}^{t-1} \binom{k}{j} \left(\frac{\psi}{\psi+\alpha}\right)^k \right) \right)}{\sum_{k=0}^{t} \left(\frac{\psi}{\psi+\alpha}\right)^k}$$

for any $\delta > 0$, there exists ψ large enough so that

$$\left| h(\mathbf{X}, a, l, t) - \frac{\frac{\hat{c}_a}{\psi+\alpha} \left(\frac{\hat{\mu}_{a,l}}{\psi}\right) \left(\sum_{k=0}^{t-1} \left(\frac{\psi}{\psi+\alpha}\right)^k \right)}{\sum_{k=0}^{t} \left(\frac{\psi}{\psi+\alpha}\right)^k} \right| < \delta.$$

This is because the term $\left(\frac{\hat{\mu}_{a,l}}{\psi}\right)^j$ on the right hand of the expression goes to zero faster than the left hand side. This result holds trivially when $X_a \geq t$. Therefore, it holds that when ψ is large, for any action, $a \in \mathcal{A}(\mathbf{X})$,

$$\sum_{l=1}^{m_a} b_{a,l} h(\mathbf{X}, a, l, t) = c_a \mathrm{E}\left[\mu_a|\mathbf{b}\right] \frac{\sum_{k=0}^{t-1} \beta^k}{\sum_{k=0}^{t} \beta^k} - \delta$$

for arbitrarily small δ, which implies the E$c\mu$ policy is optimal as $\psi \to \infty$. Furthermore, since $\mathrm{V}(\mathbf{X}, \mathbf{b}) = \frac{\mathbf{c}\mathbf{X}^{\mathrm{T}}}{1-\beta} - \hat{\mathrm{W}}(\mathbf{X}, \mathbf{b}, 0)$, and $\lim_{\psi\to\infty} \hat{\mathrm{W}}(\mathbf{X}, \mathbf{b}) = \lim_{\psi\to\infty} \mathrm{W}(\mathbf{X}, \mathbf{b})$ as discussed in Lemma A.3, the assertion is proven. □

Lemma A.7 (**Finite Cardinality of $\Pi_{\mathbf{b}}$**). *If $\min_{j\in\mathcal{J}_i} c_i\mu_{i,j} \neq \min_{j\in\mathcal{J}_k} c_k\mu_{k,j}$ for any distinct $i, k \in \mathcal{N}$, and $b_{i,j} > 0$ for each component of $\mathbf{b} \in \mathcal{B}$, then $|\Pi_{\mathbf{b}}| < \infty$.*

Proof. Without loss of generality, let $\min_{j\in\mathcal{J}_i} \mu_{i,j} = \mu_{i,1}$ and assume $c_1\mu_{1,1} < c_2\mu_{2,1} < \ldots < c_n\mu_{n,1}$.

We begin by proving that for any belief $\mathbf{b} \in \mathcal{B}$ with $b_{i,j} > 0$ for all $i \in \mathcal{N}, j \in \mathcal{J}_i$, for any $\delta > 0$, there exists k such that the Bayesian updated belief after experiencing k service incompletions of class a, which we denote $\mathbf{b}'$, has component $b'_{i,1} > 1 - \delta$.

From Equation (2.3), $\sigma(\mathbf{b}, a, -)_{a,j} = \frac{(1-\mu_{a,j})b_{a,j}}{(1-\mathrm{E}[\mu_a|\mathbf{b}])}$, which implies that the belief $\mathbf{b}'$ resulting from observing k service incompletions

$$\hat{\mathbf{b}} = \sigma(\ldots\sigma(\sigma(\mathbf{b}, a, -), a, -)\ldots, a, -)$$

has components that are related in the following manner:

$$\hat{b}_{a,1} = \hat{b}_{a,j}\left(\frac{1-\mu_{a,1}}{1-\mu_{a,j}}\right)^k \frac{b_{a,1}}{b_{a,j}}.$$

Now, since $\left(\frac{1-\mu_{a,1}}{1-\mu_{a,j}}\right)^k > 1$, $\hat{b}_{a,1}$ becomes arbitrarily large with respect to any other component $\hat{b}_{a,j}$ for large enough k, which implies that $\hat{b}_{a,1}$ becomes arbitrarily close to 1 for large enough k.

Therefore, after experiencing k service incompletions to class $a \in \mathcal{N}$, $\mathrm{E}[\mu_a|\hat{\mathbf{b}}]$ becomes arbitrarily close to $\min_{j\in\mathcal{J}_a} c_a\mu_{a,j}$, since terms $\hat{b}_{a,j}\mu_{a,j}$ become near zero for any $j \neq 1$. Since order of observations does not effect the updated belief state, suppose we choose $k_1, k_2, \ldots, k_n$ such that $\hat{\mathbf{b}}$, the updated belief state of $\mathbf{b}$ after experiencing X_i successful service observations and k_i incomplete service observations from each class $i \in \mathcal{N}$, has $\mathrm{E}[\mu_1|\hat{\mathbf{b}}]c_1 < \mathrm{E}[\mu_2|\hat{\mathbf{b}}]c_2 < \ldots < \mathrm{E}[\mu_n|\hat{\mathbf{b}}]c_n$.

Since successful observations only increase $\mathrm{E}[\mu_n|\hat{\mathbf{b}}]c_n$, and since E$c\mu$ policies choose to serve the largest $\mathrm{E}[\mu_i|\hat{\mathbf{b}}]c_i$ under any path, the path will have either successfully served, or will continue to serve class n (without attempting service of other classes) until all class n customers are served after time $t = \sum_{i\in\mathcal{N}}(k_i + X_i)$.

Suppose all members of n are served at time τ. Then, in a similar manner, either all customers of class $n-1$ are served or will continue to be served after $\tau + \sum_{i=1}^{n-1}(k_i + X_i)$. In this way, we see that there are finitely many indexing policies with starting belief

$\mathbf{b}$ and queue state $\mathbf{X}$, since after given periods of time, the indexing policy is either guaranteed to have completed service to a class, or continues to serve that class until completion without attempting service at other classes.

□

Proof of Proposition 2.3. Without loss of generality, let $\min_{j \in \mathcal{J}_i} \mu_{i,j} = \mu_{i,1}$ and assume $c_1\mu_{1,1} < c_2\mu_{2,1} < \ldots < c_n\mu_{n,1}$. The proof follows in a similar manner to the proof of Lemma A.7. As in the proof of Lemma A.7, for any belief $\mathbf{b} \in \mathcal{B}'$, there exists $k_{\mathbf{b}}^1, k_{\mathbf{b}}^2, \ldots, k_{\mathbf{b}}^n$ such that the updated belief state of $\mathbf{b}$ after experiencing X_i successful service observations and $k_{\mathbf{b}}^i$ incomplete service observations from each class $i \in \mathcal{N}$ which we term $\hat{\mathbf{b}}$, has $\mathrm{E}[\mu_1|\hat{\mathbf{b}}]c_1 < \mathrm{E}[\mu_2|\hat{\mathbf{b}}]c_2 < \ldots < \mathrm{E}[\mu_n|\hat{\mathbf{b}}]c_n$.

Now, letting $k_i = \sup_{\mathbf{b} \in \mathcal{B}'} k_{\mathbf{b}}^i$, we know that the updated belief state of $\mathbf{b} \in \mathcal{B}'$ after experiencing X_i successful service observations and k_i incomplete service observations from each class $i \in \mathcal{N}$ which we term $\mathbf{b}^*$, has $\mathrm{E}[\mu_1|\mathbf{b}^*]c_1 < \mathrm{E}[\mu_2|\mathbf{b}^*]c_2 < \ldots < \mathrm{E}[\mu_n|\mathbf{b}^*]c_n$. Furthermore, each k_i exists and is finite since $\mathcal{B}'$ is a closed set.

Using a similar argument to Lemma A.7, every indexing policy with starting belief $\mathbf{b} \in \mathcal{B}'$ and queue state $\mathbf{X}$ chooses to serve the largest index under any path of an indexing policy, so the path will have either successfully served, or will continue to serve class n (without attempting service of other classes) until all class n customers are served after time $t = \sum_{i \in \mathcal{N}}(k_i + X_i)$.

For a given belief $\mathbf{b}$, suppose all members of n are served at time $\tau_{\mathbf{b}}$. Then, in a similar manner, either all customers of class $n-1$ are served or will continue to be served after $\tau + \sum_{i=1}^{n-1}(k_i + X_i)$. In this way, we see that there are finitely many indexing policies within $\mathcal{B}'$, since after given periods of time, an indexing policy of a belief in $\mathcal{B}'$ is either guaranteed to have completed service to a class, or continues to serve that class until completion without attempting service at other classes.

Since every policy takes the form of a hyperplane over $\mathcal{B}'$, $\mathrm{V}^{\pi_{\mathbf{b}}}(\mathbf{X}, \mathbf{b})$ must be piecewise-linear on $\mathcal{B}'$. □

Proof of Theorem 2.2. We begin the proof by showing that the optimal policy must be a subgradient to the value function. Suppose that π is optimal to the percentile objective. Since policies are evaluated as linear functions of $\mathbf{b}$, there exists $(\hat{\mathbf{v}}, v) \in \mathbb{R}^m \times \mathbb{R}$ such that $\mathrm{V}^{\pi}(\mathbf{X}, \mathbf{b}) = \hat{\mathbf{v}}\mathbf{b}^{\mathrm{T}} + v$. Furthermore, since by definition $\mathrm{V}(\mathbf{X}, \mathbf{b})$ is optimal (minimized) at $\mathbf{b}$, $\hat{\mathbf{v}}\mathbf{b}^{\mathrm{T}} + v \geq \mathrm{V}(\mathbf{X}, \mathbf{b})$ so it lies completely within the epigraph of $\mathrm{V}(\mathbf{X}, \mathbf{b})$ for all $\mathbf{b} \in \mathcal{B}$.

Now, $\mathrm{V}(\mathbf{X}, \mathbf{b}) = \min_{(\mathbf{v},v)\in\mathcal{V}} \mathbf{v}\mathbf{b}^{\mathrm{T}} + v$ for some set $\mathcal{V} \subset \{(\mathbf{v}, v) \in \mathbb{R}^m \times \mathbb{R}\}$ since POMDPs are composed of linear hyperplanes of the belief. We consider the extended POMDP value function for all $\mathbf{b} \in \mathbb{R}^m$ defined as $\mathrm{V}(\mathbf{X}, \mathbf{b}) = \min_{(\mathbf{v},v)\in\mathcal{V}} \mathbf{v}\mathbf{b}^{\mathrm{T}} + v$. That is, the extended value function that also considers $\mathbf{b} \notin \mathcal{B}$ defines these points using the hyperplane set of the POMDP.

Suppose that $(\hat{\mathbf{v}}, \hat{v})$ is in the epigraph of $\mathrm{V}(\mathbf{X}, \mathbf{b})$ over $\mathcal{B}$, but not for $\mathbb{R}^m$. That is, there exists some $\hat{\mathbf{b}} \in \mathbb{R}^m$ such that $\hat{\mathbf{v}}\hat{\mathbf{b}}^{\mathrm{T}} + \hat{v} < \mathrm{V}(\mathbf{X}, \hat{\mathbf{b}})$. Then, let let us define the convex set $\mathcal{Q} = \left\{\mathbf{b} \in \mathbb{R}^m : \mathrm{V}(\mathbf{X}, \mathbf{b}) \geq \hat{\mathbf{v}}\mathbf{b}^{\mathrm{T}} + v\right\}$. Then, either $\min_{\mathbf{b}\in\mathcal{Q}} \hat{\mathbf{v}}\mathbf{b}^{\mathrm{T}} + \hat{v}$ or $\max_{\mathbf{b}\in\mathcal{Q}} \hat{\mathbf{v}}\mathbf{b}^{\mathrm{T}} + \hat{v}$ is finite depending on if $\hat{\mathbf{v}}\mathbf{b}^{\mathrm{T}}$ is increasing or decreasing toward $\mathcal{Q}$ from $\mathrm{int}(\mathcal{B})$. Then, let us define $d = \min_{\mathbf{b}\in\mathcal{Q}} \hat{\mathbf{v}}\mathbf{b}^{\mathrm{T}} + \hat{v}$ if it is finite, and $d = \max_{\mathbf{b}\in\mathcal{Q}} \hat{\mathbf{v}}\mathbf{b}^{\mathrm{T}} + \hat{v}$ otherwise.

Consider $\mathcal{S} = \left\{\mathbf{b} \in \mathbb{R} | \hat{\mathbf{v}}\mathbf{b}^{\mathrm{T}} + \hat{v} = d\right\}$. This is a supporting hyperplane to $\mathcal{Q}$. Define $\mathcal{S}_1$ to be the halfspace defined by $\mathcal{S}$ containing $\mathcal{Q}$. Likewise, let $\mathcal{S}_2$ be the opposing halfspace defined by $\mathcal{S}$ which must contain $\mathcal{B}$.

Now, there exists a subgradient to $\mathrm{V}(\mathbf{X}, \mathbf{b})$ defined by $\bar{\mathbf{v}}\mathbf{b}^{\mathrm{T}} + \bar{v}$ where $\bar{\mathbf{v}}\mathbf{b}^{\mathrm{T}} + \bar{v} = d$ for all $\mathbf{b} \in \mathcal{S}$ because $\bar{\mathbf{v}}\mathbf{b}^{\mathrm{T}} + \bar{v} \geq \mathrm{V}(\mathbf{X}, \mathbf{b})$ for all $\mathbf{b} \in \mathcal{S}$. Since it is a subgradient, $\bar{\mathbf{v}}\mathbf{b}^{\mathrm{T}} + \bar{v} \geq \mathrm{V}(\mathbf{X}, \mathbf{b})$ for all $\mathbf{b} \in \mathbb{R}^m$ which implies that $\bar{\mathbf{v}}\mathbf{b}^{\mathrm{T}} + \bar{v} \geq \hat{\mathbf{v}}\mathbf{b}^{\mathrm{T}} + \hat{v}$ for all

$\mathbf{b} \in \mathcal{S}_1$. This implies that $\bar{\mathbf{v}}\mathbf{b}^{\mathrm{T}} + \bar{v} \leq \hat{\mathbf{v}}\mathbf{b}^{\mathrm{T}} + \hat{v}$ for all $\mathbf{b} \in \mathcal{S}_2$. This implies, that if we could find a policy $\pi_{\hat{h}}$ such that $\mathrm{V}^{\pi_{\hat{h}}}(\mathbf{X}, \mathbf{b}) = \bar{\mathbf{v}}\mathbf{b}^{\mathrm{T}} + \bar{v}$, then the associated $\mathrm{Y}^{\pi_{\hat{h}}}(\mathbf{X}, \epsilon)$ would be smaller too.

Now, $\mathrm{V}(\mathbf{X}, \mathbf{b})$ can be composed of stationary, non-randomizing policies hence $\mathcal{V}$ can be composed of policy vectors that are stationary and non-randomizing. Since $\mathrm{V}(\mathbf{X}, \mathbf{b})$ is concave, it is differentiable almost everywhere. According to Clarke (1975), the subdifferential at any point $\mathbf{b}$ can be composed of the gradients of the function within an open ball surrounding $\mathbf{b}$. Since in our extended framework, all $\mathbf{b}$ can be surrounded by such an open ball, it implies that the subdifferential at any point $\mathbf{b} \in \mathbb{R}^m$ is composed of the convex hull of such gradients, or, $(\mathbf{v}, v) \in \mathcal{V}$. Now, the subdifferential is a convex set, which implies that by Carathèodory's theorem, any element of the subdifferential can be composed of a convex combination of finitely many (at most $m-1$ elements) of the convex hull. Now, since randomization policies are evaluated as convex combinations of the stationary policies of $\mathcal{V}$, the proof that there exists a randomization of stationary Markov policies that is optimal to the percentile objective is complete.

Now, as we prove in Lemma A.8, $\mathrm{V}(\mathbf{X}, \mathbf{b})$ is non-decreasing toward $\mathbf{b}_0$. Thus, $\mathcal{Z}_y = \{\mathbf{b} \in \mathcal{B} | \mathrm{V}(\mathbf{X}, \mathbf{b}) \geq y\}$ is a convex set that includes $\mathbf{b}_0$ so long as $y \leq \mathrm{V}(\mathbf{X}, \mathbf{b}_0)$. (We need not consider any larger value for y by definition of the percentile objective). Suppose that π is optimal so that $\mathrm{V}^{\pi}(\mathbf{X}, \mathbf{b}) = \mathbf{v}\mathbf{b}^{\mathrm{T}} + v$. Then suppose that $\mathrm{Y}^{\pi}(\mathbf{X}, \epsilon) = y$. Consider the set $\mathcal{H} = \left\{\mathbf{b} \in \mathcal{B} | \mathbf{v}\mathbf{b}^{\mathrm{T}} + v = y\right\}$. On one half-space, $\mathcal{H}_1 = \left\{\mathbf{b} \in \mathcal{B} | \mathbf{v}\mathbf{b}^{\mathrm{T}} + v \leq y\right\}$ and on the other half-space $\mathcal{H}_2 = \left\{\mathbf{b} \in \mathcal{B} | \mathbf{v}\mathbf{b}^{\mathrm{T}} + v > y\right\}$. For the percentile value assigned to policy to be valid, $\mathbb{P}_{\mathbf{B}}(\mathbf{B} \in \mathcal{H}_2) < \epsilon$. We note that $\mathcal{Z}_y$ must be a subset of $\mathcal{H}_2$.

If $\mathcal{H}$ is a supporting hyperplane to $\mathcal{Z}_y$, then it is a subgradient to $\mathrm{V}(\mathbf{X}, \mathbf{b})$ for some $\mathbf{b} \in \mathcal{H} \bigcap \mathcal{Z}_{\epsilon}$, hence it can be formed from a randomization of stationary policies at a point $\mathbf{b}^*$ where $\mathrm{V}(\mathbf{X}, \mathbf{b}^*) = y$. Now suppose that $\mathcal{H}$ is not a supporting hyperplane to $\mathcal{Z}_y$. Now, there exists a subgradient $(\hat{\mathbf{v}}, \hat{v})$ that forms a supporting hyperplane to $\mathcal{Z}_y$ while still satisfying the probability constraint. That is, $\hat{\mathcal{H}} = \left\{\mathbf{b} \in \mathcal{B} | \hat{\mathbf{v}}\mathbf{b}^{\mathrm{T}} + \hat{v} = y\right\}$ is a supporting hyperplane to $\mathcal{Z}_y$ while for $\hat{\mathcal{H}}_2 = \left\{\mathbf{b} \in \mathcal{B} | \hat{\mathbf{v}}\mathbf{b}^{\mathrm{T}} + \hat{v} > y\right\}$ we have $\mathbb{P}_{\mathbf{B}}(\mathbf{B} \in \hat{\mathcal{H}}_2) < \epsilon$ since $\mathcal{Z}_y \subset \mathcal{S}_2$. Thus, a policy that forms a supporting plane to $\mathcal{Z}_y$ can satisfy the same percentile objective, which by above argument, proves the theorem.

□

COROLLARY A.1 (**Generalized Chance-Constrained Policies**). *Theorem 2.2 holds for any MDP that can express its parameter ambiguity in a learning context as a POMDP.*

Proof. If an MDP with parameter ambiguity can be expressed as a POMDP, we can express its fully observed component as $\mathbf{X}$ and the current belief about its parameters as $\mathbf{b}$, hence we can describe the non-robust problem via a POMDP value function $\mathrm{V}(\mathbf{X}, \mathbf{b})$. Inspecting the proof of Theorem 2.2, we can see that the only properties we use to prove the Theorem are shared by all POMDPs. Specifically, we use the linearity of policies evaluated over the belief space and the convexity of $\mathrm{V}(\mathbf{X}, \mathbf{b})$ with respect to belief. Hence, the Theorem carries over to the other MDPs with parameter ambiguity expressed as a POMDP. □

LEMMA A.8 (**Value Function Nonincreasing on Line Segments**). *For any* $\mathbf{b} \in \mathcal{B}$ *and* $\lambda \in [0, 1]$, $\mathrm{V}(\mathbf{X}, \lambda\mathbf{b}_0 + (1 - \lambda)\mathbf{b})$ *is nonincreasing as* λ *increases.*

Proof. The proof follows immediately as a special case of Lemma A.15. □

Proof of Proposition 2.4. Letting y_ϵ denote the optimal value of the percentile objective, we first prove that $y_\epsilon \geq \mathrm{V}(\mathbf{X}, \hat{\mathbf{b}})$ for all $\hat{\mathbf{b}} \in \mathcal{L}_\epsilon$. Suppose by contradiction that $y_\epsilon < \mathrm{V}(\mathbf{X}, \hat{\mathbf{b}})$ for some $\hat{\mathbf{b}} \in \mathcal{L}_\epsilon$. Then $\mathbb{P}_{\mathbf{B}}\left(\mathrm{V}^\pi\left(\mathbf{X}, \mathbf{B}\right) \geq y_\epsilon\right) \leq \epsilon$. Since V^π is linear in belief, $\mathrm{V}^\pi\left(\mathbf{X}, \mathbf{b}\right) = \mathbf{b}\mathbf{w}^{\mathrm{T}} + w$ for some $(\mathbf{w}, w) \in \mathbb{R}^m \times \mathbb{R}$. Therefore, since $\mathbb{P}_{\mathbf{B}}\left(\mathbf{B}\mathbf{w}^{\mathrm{T}} + w \geq y_\epsilon\right) \leq \epsilon$, $(\mathbf{w}, y_\epsilon - w) \in \mathcal{W}_\epsilon$ as in definition 2.1. However, since $\hat{\mathbf{b}} \in \mathcal{L}_\epsilon$, we have $\hat{\mathbf{b}}\mathbf{w}^{\mathrm{T}} + w \leq y_\epsilon$ because this is true for all elements of $\mathcal{W}_\epsilon$. Therefore, $\mathrm{V}(\mathbf{X}, \hat{\mathbf{b}}) \leq \mathrm{V}^\pi(\mathbf{X}, \hat{\mathbf{b}}) = \hat{\mathbf{b}}\mathbf{w}^{\mathrm{T}} + w \leq y_\epsilon$ which is a contradiction, and thus, the assertion holds.

Next, suppose that there exists $\mathbf{b}' \in \mathcal{L}_\epsilon$ such that $\mathrm{V}\left(\mathbf{X}, \mathbf{b}'\right) = \mathrm{R}\left(\mathbf{X}\right)$. By the above argument, $y_\epsilon \geq \mathrm{R}\left(\mathbf{X}\right)$. Furthermore, since $\mathrm{V}\left(\mathbf{X}, \mathbf{b}\right)$ is non-increasing as $\mathbf{b}$ strays from $\mathbf{b}_0$ on a line segment (as proven in the proof of Lemma A.8, there must exist $\mathbf{b}'' \in \delta\mathcal{L}_\epsilon$ such that $\mathrm{V}\left(\mathbf{X}, \mathbf{b}''\right) = \mathrm{R}\left(\mathbf{X}\right)$. Otherwise, there would exist $\mathbf{b}' \in \mathcal{L}_\epsilon \setminus \delta\mathcal{L}_\epsilon$ such that $\mathrm{V}\left(\mathbf{X}, \mathbf{b}'\right) = \mathrm{R}\left(\mathbf{X}\right)$, while for the $\lambda \in (0, 1)$ that satisfies $\lambda\mathbf{b}_0 + (1 - \lambda)\mathbf{b}' \in \delta\mathcal{L}_\epsilon$, $\mathrm{V}\left(\mathbf{X}, \lambda\mathbf{b}_0 + (1 - \lambda)\mathbf{b}'\right) < \mathrm{R}\left(\mathbf{X}\right)$, which is a contradiction to the proof of Lemma A.8. Letting π be the minimax $c\mu$ policy associated with $\mathrm{R}\left(\mathbf{X}\right)$, by the proof of Proposition 2.1, $\mathrm{V}^\pi\left(\mathbf{X}, \mathbf{b}\right) \leq \mathrm{R}\left(\mathbf{X}\right)$ for all $\mathbf{b} \in \mathcal{B}$. Therefore, $\mathrm{V}\left(\mathbf{X}, \mathbf{b}''\right) = \mathrm{R}\left(\mathbf{X}\right) = \mathrm{V}^\pi\left(\mathbf{X}, \mathbf{b}''\right)$. Thus, $\mathcal{K}_{\mathbf{b}''} = \{\pi\}$ is an optimal chance-constrained policy in this case.

Otherwise, suppose there does not exist $\mathbf{b}' \in \mathcal{L}_\epsilon$ such that $\mathrm{V}\left(\mathbf{X}, \mathbf{b}\right) = \mathrm{R}\left(\mathbf{X}\right)$. Then, for $\mathbf{b}' \in \mathcal{L}_\epsilon$, $\mathrm{V}\left(\mathbf{X}, \mathbf{b}'\right)$ is strictly decreasing as $\mathbf{b}'$ strays from $\mathbf{b}_0$ on a line. This is because, for any $\mathbf{b}' \in \mathcal{L}_\epsilon$, and $\lambda \in (0, 1]$ such that $\lambda\mathbf{b}_0 + (1 - \lambda)\mathbf{b}' \in \mathcal{L}_\epsilon$, $\mathrm{V}\left(\mathbf{X}, \lambda\mathbf{b}_0 + (1 - \lambda)\mathbf{b}'\right) < \mathrm{V}\left(\mathbf{X}, \mathbf{b}_0\right)$, so by the non-increasing result of the proof of Lemma A.8 and the concavity of the value function, $\mathrm{V}\left(\mathbf{X}, \lambda\mathbf{b}_0 + (1 - \lambda)\mathbf{b}'\right)$ is strictly increasing in λ while the belief remains in $\mathcal{L}_\epsilon$.

Next, define

$$Q_y = \left\{\mathbf{b} \in \mathcal{B} | \mathrm{V}\left(\mathbf{X}, \mathbf{b}\right) \geq y\right\},$$

and observe that Q_y is convex, since $\mathrm{V}\left(\mathbf{X}, \mathbf{b}\right)$ is concave. $\mathcal{L}_\epsilon$ is formed through an intersection of half-spaces, and thus, it must also be convex. If $\mathbf{b}_0 \in \mathcal{L}_\epsilon$, then by previous argument, $\mathbf{b}_0 \in \delta\mathcal{L}_\epsilon$ (since it is composed of zeros and ones) and hence

$\mathcal{K}_{\mathbf{b}_0} = \{\pi\}$ can be composed of a single policy.

Otherwise, for large enough y, Q_y and $\mathcal{L}_\epsilon$ are non-intersecting while $Q_y \neq \emptyset$. Consider the y for which $Q_y \bigcap \{\mathcal{L}_\epsilon \setminus \delta\mathcal{L}_\epsilon\} = \emptyset$ and $Q_y \bigcap \mathcal{L}_\epsilon \neq \emptyset$. This y exists since $\mathrm{V}(\mathbf{X}, \mathbf{b})$ is continuously strictly decreasing as $\mathbf{b} \in \mathcal{L}_\epsilon$ strays from $\mathbf{b}_0$ as shown above. By the *hyperplane separation theorem*, there must exist a separating hyperplane defined by the tuple $(\mathbf{v}, v) \in \mathbb{R}^m \times \mathbb{R}$ such that for every $\mathbf{b}' \in \mathcal{L}_\epsilon$, $\mathbf{b}\mathbf{v}^{\mathrm{T}} \leq v$ and for every $\mathbf{b} \in Q_y$, $\mathbf{b}\mathbf{v}^{\mathrm{T}} \geq v$. Now, by the definition of the convex floating body, there does not exist $(\mathbf{w}, w) \in \mathbb{R}^m \times \mathbb{R}$ such that $\mathbb{P}_{\mathbf{B}}\left(\mathbf{B}\mathbf{w}^{\mathrm{T}} \geq w\right) \leq \epsilon$ while for a point $\mathbf{b}' \in \mathcal{L}_\epsilon$, $\mathbf{b}\mathbf{w}^{\mathrm{T}} > w$. Therefore, since $\mathbf{b}'\mathbf{v}^{\mathrm{T}} \leq v$ for all $\mathbf{b}' \in \mathcal{L}_\epsilon$, $\mathbb{P}_{\mathbf{B}}\left(\mathbf{B}\mathbf{v}^{\mathrm{T}} \geq v\right) \leq \epsilon$ because otherwise there would exist some $\mathbf{b}' \in \mathcal{L}_\epsilon$ such that $\mathbf{b}'\mathbf{v}^{\mathrm{T}} > v$.

This separating hyperplane must intersect with some point $\mathbf{b}^* \in \delta\mathcal{L}_\epsilon$ which implies that $\mathbf{b}^* = \arg\max_{\mathbf{b} \in \delta\mathcal{L}_\epsilon} \mathrm{V}(\mathbf{X}, \mathbf{b})$ since if it were not, there would exist a larger y' such that $Q_{y'} \bigcap \{\mathcal{L}_\epsilon \setminus \delta\mathcal{L}_\epsilon\} = \emptyset$ while $Q_{y'} \bigcap \mathcal{L}_\epsilon \neq \emptyset$. Thus, $y_\epsilon \geq \mathrm{V}(\mathbf{X}, \mathbf{b}^*)$. Now, for every indexing policy $\pi_{\mathbf{b}^*}$, $\{\mathbf{b} \in \mathcal{B} | \mathrm{V}^{\pi_{\mathbf{b}^*}}(\mathbf{X}, \mathbf{b}) = \mathrm{V}^{\pi_{\mathbf{b}^*}}(\mathbf{X}, \mathbf{b}^*)\}$ is a supporting hyperplane of Q_y since $\mathrm{V}^{\pi_{\mathbf{b}^*}}(\mathbf{X}, \mathbf{b})$ is linear in $\mathbf{b}$.

Any supporting hyperplane of Q_y can be formed from a convex combination of supporting hyperplanes $\{\mathbf{b} \in \mathcal{B} | \mathrm{V}^{\pi_{\mathbf{b}^*}}(\mathbf{X}, \mathbf{b}) = \mathrm{V}(\mathbf{X}, \mathbf{b}^*)\}$ as proven in our proof of Theorem 2.2. Thus, there exists $\pi^{\mathbf{p}^*}_{\mathcal{K}_{\mathbf{b}^*}}$ such that $\{\mathbf{b} \in \mathcal{B} | \mathrm{V}^{\pi^{\mathbf{p}^*}_{\mathcal{K}_{\mathbf{b}^*}}}(\mathbf{X}, \mathbf{b}) = \mathrm{V}(\mathbf{X}, \mathbf{b}^*)\} = \left\{\mathbf{b} \in \mathcal{B} | \mathbf{b}\mathbf{v}^{\mathrm{T}} = v\right\}$ which implies that $\mathbb{P}_{\mathbf{B}}(\mathrm{V}^{\pi^{\mathbf{p}^*}_{\mathcal{K}_{\mathbf{b}^*}}}(\mathbf{X}, \mathbf{b}) \geq \mathrm{V}(\mathbf{X}, \mathbf{b}^*)) \leq \epsilon$. Therefore, $y_\epsilon = \mathrm{V}(\mathbf{X}, \mathbf{b}^*)$ with optimal policy $\pi^{\mathbf{p}^*}_{\mathcal{K}_{\mathbf{b}^*}}$. □

Proof of Proposition 2.5. Proposition 2.4 asserts that the $\mathbf{b}^*$ lies at the maximum of the convex floating body. Thus, by contradiction, if $\mathbf{b}^* \in \mathcal{L}_\epsilon$ is *not* visible from reference belief $\mathbf{b}_0$, it implies there exists $\lambda \in (0, 1]$ such that $\mathrm{V}(\mathbf{X}, \mathbf{b}^*) \leq \mathrm{V}(\mathbf{X}, \lambda\mathbf{b}_0 + (1-\lambda)\mathbf{b}^*)$ while $\lambda\mathbf{b}_0 + (1-\lambda)\mathbf{b}^* \in \mathcal{L}_\epsilon$. However, by Lemma A.8, this is a contradiction, which implies that $\mathbf{b}^*$ must be visible from $\mathbf{b}_0$. □

Lemma A.9 (**Inner/Outer Approximations to $\mathcal{L}_\epsilon$**). *For any nonempty convex sets $\mathcal{S}_o$ and $\mathcal{S}_i$ such that $\mathcal{S}_i \subseteq \mathcal{L}_\epsilon \subseteq \mathcal{S}_o$, the policies generated from $\mathcal{S}_o$ and $\mathcal{S}_i$ are upper and lower bounds to the optimal policy.*

Proof. The proof is straightforward: If we generate policies assuming that $\mathcal{S}_o$ is indeed $\mathcal{L}_\epsilon$ in the same manner as Proposition 2.5 and Proposition 2.4, we know that the minimax belief associated with $\mathcal{S}_o$, which we denote $\mathbf{b}_o^*$ must have a cost larger than that of $\mathbf{b}^*$ since $\mathcal{L}_\epsilon \subseteq \mathcal{S}_o$. Similarly, if we generate policies assuming that $\mathcal{S}_i$ is indeed $\mathcal{L}_\epsilon$ in the same manner as Proposition 2.5 and Proposition 2.4, we know that the minimax belief associated with $\mathcal{S}_i$, which we denote $\mathbf{b}_i^*$ must have a cost smaller than that of $\mathbf{b}^*$ since $\mathcal{S}_o \subseteq \mathcal{L}_\epsilon$. Obviously, as $\mathcal{S}_o \to \mathcal{L}_\epsilon$ and $\mathcal{S}_i \to \mathcal{L}_\epsilon$, $\mathrm{V}(\mathbf{X}, \mathbf{b}_o^*) \to \mathrm{V}(\mathbf{X}, \mathbf{b}^*)$ and $\mathrm{V}(\mathbf{X}, \mathbf{b}_i^*) \to \mathrm{V}(\mathbf{X}, \mathbf{b}^*)$ since the minimax operation is performed on successively smaller (larger) sets which implies that the policies converge to the percentile objective. □

Corollary A.2 (**Generalized Convex Floating Body**). *For any MDP that can express its parameter ambiguity in a learning context as a POMDP, if there exists an extreme belief point $\mathbf{b}_0$ composed of zeros and ones such that Lemma A.8 holds, then Propositions 2.4, 2.5, and Lemma A.9 also hold.*

Proof. Similar to our proof of Corollary A.1, if we inspect the proofs of Propositions 2.4, 2.5, and Lemma A.9, all they require is the general structural properties of POMDPs as well as Lemma A.8. Hence if the MDP with ambiguity can be modeled as a POMDP and there exists an extreme belief $\mathbf{b}_0$ such that Lemma A.8 holds, our results related to the location of $\mathbf{b}^*$ also transfer to this model. □

Traditional priority policies such as the $c\mu$ rule seen in the literature serve a single class until there are no customers left of that class, then move on to other classes in a similar fashion. Evaluating these policies turns out to be a simple task that has a closed form solution. First we let $\mathcal{S} = \{s_1, s_2, \ldots, s_n\}$ index the priority of classes, where s_1 is the highest priority class, and s_n is the lowest priority class.

We define

$$\begin{aligned}
&\mathrm{D}_t(X_{s_i}, \mathbf{b}, s_i) = \\
&\qquad \mathrm{U}_t(X_{s_i}, \mathbf{b}, s_i) + \frac{(1-\beta)}{c_{s_i}} \mathrm{D}_t(X_{s_{i+1}}, \mathbf{b}, s_{i+1}) \left(\mathrm{U}_t(X_{s_i}, \mathbf{b}, s_i) - \mathrm{U}_t(X_{s_i} - 1, \mathbf{b}, s_i)\right),
\end{aligned}$$

and,

$$\mathrm{D}_t(X_{s_n}, \mathbf{b}, s_n) = \mathrm{U}_t(X_{s_n}, \mathbf{b}, s_n).$$

Lemma A.10. *The priority policy $\pi_{\mathcal{S}}$ associated with $\mathcal{S}$ is evaluated as*

$$\mathrm{V}^{\pi_{\mathcal{S}}}(\mathbf{X}, \mathbf{b}) = \frac{\mathbf{c}\mathbf{X}^{\mathrm{T}}}{1-\beta} - \lim_{t \to \infty} \mathrm{D}_t(X_{s_1}, \mathbf{b}, s_1).$$

Proof of Lemma A.10. Consider the dynamic program,

$$\begin{aligned}
\mathrm{F}_{t+1}(X_a, \mathbf{b}, a) = \beta \Bigg(&\mathrm{E}\left[\mu_a | \mathbf{b}\right] \left(\mathrm{F}_t\left(X_a - 1, \sigma(\mathbf{b}, +, a), a\right) + \frac{c_a}{1-\beta} + d\mathbb{1}\left\{X_a - 1 = 0\right\} \right) \\
&+ \left(1 - \mathrm{E}\left[\mu_a | \mathbf{b}\right]\right) \mathrm{F}_t\left(X_a, \sigma(\mathbf{b}, -, a), a\right) \Bigg),
\end{aligned}$$

with the terminal condition

$$\mathrm{F}_0(X_a, \mathbf{b}, a) = \mathrm{F}_t(0, \mathbf{b}, a) = 0.$$

This dynamic programming value function is nearly identical to U except that a reward of d is given upon the service of the final customer of the class. It is evident that if we replace d by the reward generated by serving the remaining classes to completion according to the priority policy $\mathcal{S}$, F_t as t tends to infinity is the value of priority discipline $\mathcal{S}$. We wish to show that in this way, D is equivalent to F.

It is clear that all the value that comes from $\frac{c_a}{1-\beta}$ in t time periods in F_t is given by $\mathrm{U}_t\left(X_{s_i}, \mathbf{b}, s_i\right)$ as is proven in our proof of Theorem 2.1, which explains the first term of $\mathrm{D}_t(X_{s_i}, \mathbf{b}, s_i)$.

Replacing $\mathrm{D}_{t+1}(X_{s_{i+1}}, \mathbf{b}, s_{i+1})$ to d in the dynamic program $\mathrm{D}_{t+1}(X_{s_i}, \mathbf{b}, s_i)$, it is easy to identify that,

$$\mathrm{U}_t(X_{s_i}, \mathbf{b}, s_i) - \mathrm{U}_t(X_{s_i} - 1, \mathbf{b}, s_i),$$

is the value of the rewards obtained from serving the final person in class s_i. Since this value is in terms of c_a, multiplying by $\frac{1-\beta}{c_a}$ and multiplying by d gives the rewards in terms of d which is what we wanted. Since we are counting total savings in $\mathrm{D}_t(X_{s_i}, \mathbf{b}, s_i)$, the value of the priority policy becomes the total idled system cost minus the savings, or $\frac{\mathbf{cX}^{\mathrm{T}}}{1-\beta} - \lim_{t\to\infty} \mathrm{D}_t(X_{s_1}, \mathbf{b}, s_1)$. This concludes the proof. □

Proof of Proposition 2.6. To prove the lower bound, we begin by noting that the CDF of a geometric distribution denoted $G_p(k)$ is concave in success probability parameter since its second derivative $-(-1+k)k(1-p)^{-2+k}$ is always negative. We show that in any given sample path in the system associated with $\mathrm{V}(\mathbf{X}, \mathbf{b})$ (which we call 'system 1' for the duration of the proof) every customer can be served as fast with a higher probability in a system with known rate parameters $\mathrm{V}'(\mathbf{X}, \mathbf{b})$ (which we call 'system 2').

Since the optimal policy to system 1 is non-idling, it must serve each customer eventually. Suppose that in a given sample path of system 1, a customer of class i is served in a total of k periods. The probability of serving this customer in k or fewer periods in system 1 is simply $\sum_{j\in\mathcal{J}_i} b_{i,j} G_{\mu_{i,j}}(k)$, whereas in system 2, the probability of serving this customer in k or fewer periods is $G_{\mathrm{E}[\mu_i|\mathbf{b}]}(k)$. Since the CDF of a geometric random variable is concave in its parameter, $\sum_{j\in\mathcal{J}_i} b_{i,j} G_{\mu_{i,j}}(k) \leq G_{\mathrm{E}[\mu_i|\mathbf{b}]}(k)$. Therefore, in every sample path of system 1, every customer is served slower than system 2 with known rate parameters $\mathrm{E}[\mu|\mathbf{b}]$ which proves the assertion. The proof of the upper bound is immediate as the result of evaluating a suboptimal policy.

To prove result (i) we begin by showing that the second term of $\mathrm{D}_t(X_{s_i}, \mathbf{b}, s_i)$ and $\mathrm{D}_t(X_{s_i}, \mathrm{E}[\mu_{s_i}|\mathbf{b}], s_i)$ are decreasing as X_i increases for any given t. We note that in the functions

$$\begin{aligned}&\mathrm{D}_t(X_{s_i}, \mathbf{b}, s_i) = \\ &\qquad \mathrm{U}_t(X_{s_i}, \mathbf{b}, s_i) + \mathrm{D}_t(X_{s_{i+1}}, \mathbf{b}, s_{i+1}) \left(\mathrm{U}_t(X_{s_i}, \mathbf{b}, s_i) - \mathrm{U}_t(X_{s_i} - 1, \mathbf{b}, s_i)\right) \frac{1-\beta}{c_{s_i}}\end{aligned}$$

and

$$\begin{aligned}\mathrm{D}_t(X_{s_i}, \mathrm{E}[\mu_{s_i}|\mathbf{b}], s_i) = {} & \mathrm{U}_t(X_{s_i}, \mathrm{E}[\mu_{s_i}|\mathbf{b}], s_i) \\ & + \mathrm{D}_t(X_{s_{i+1}}, \mathrm{E}[\mu_{s_{i+1}}|\mathbf{b}], s_{i+1}) \left(\mathrm{U}_t(X_{s_i}, \mathrm{E}[\mu_{s_i}|\mathbf{b}], s_i)\right. \\ & \left. - \mathrm{U}_t(X_{s_i} - 1, \mathrm{E}[\mu_{s_i}|\mathbf{b}], s_i)\right) \frac{1-\beta}{c_{s_i}},\end{aligned}$$

the second term falls to zero for large X_{s_i}, since the reward resulting from serving the last customer of class s_i is less than $\frac{c_{s_i}\beta^{X_{s_i}}}{1-\beta}$ which decreases to zero as X_i increases.

All that is left is to compare $\mathrm{U}_t(X_{s_i}, \mathbf{b}, s_i)$ and $\mathrm{U}_t(X_{s_i}, \mathrm{E}[\mu|\mathbf{b}], s_i)$, for which the difference becomes arbitrarily small as X_{s_i} becomes large, since the reward generated between the two systems is identical for the first $X_{s_i} - 1$ time periods, i.e., $\mathrm{U}_{X_{s_i}-1}(X_{s_i}, \mathbf{b}, s_i) = \mathrm{U}_{X_{s_i}-1}(X_{s_i}, \mathrm{E}[\mu|\mathbf{b}], s_i)$. Since the reward generated in the periods greater than or equal to X_{s_i} can be no more than $\frac{X_{s_i} c_{s_i} \beta^{X_{s_i}}}{1-\beta}$ which goes to zero as X_i tends to infinity property (i) is proven.

To prove result (ii), we show that $\mathrm{U}_t(X_{s_i}, \mathrm{E}[\mu|\mathbf{b}], a)$ and $\mathrm{U}_t(X_{s_i}, \mathbf{b}, a)$ approach each other as a linear function of variance of $\mathbf{b}$ via a proof similar to that of *Holder's defect formula.* Denoting $d(i, \mu_{a,l})_t = \mu_{a,l} g(t, i)_{a,l}$, notice that $\mathrm{U}_t(X_{s_i}, \mathbf{b}, a) =$ $\mathrm{E}\left[d(X_{s_i}, \mu_a)_t\right]$ and $\mathrm{U}_t(X_{s_i}, \mathrm{E}[\mu|\mathbf{b}], a) = d(X_{s_i}, \mathrm{E}\left[\mu_a|\mathbf{b}\right])_t$. Furthermore, by earlier statement, $\mathrm{E}\left[d(X_{s_i}, \mu_{a,l})_t\right] \leq d(X_{s_i}, \mathrm{E}\left[\mu_a|\mathbf{b}\right])_t$. Applying Taylor's Theorem with the Lagrange form of the remainder to d at point $\mu \in (0, 1)$, to obtain for each $\mu_{a,l} \in \mathcal{M}_a$,

$$d(X_{s_i}, \mu_{a,l})_t = d(X_{s_i}, \mu)_t + d'(X_{s_i}, \mu)_t(\mu_{a,l} - \mathrm{E}[\mu|\mathbf{b}]) + \frac{1}{2} d''(X_{s,i}, w_l)(\mu_{a,l} - \mu)^2,$$

where w_l is a real number between μ and $\mu_{a,l}$. Multiply the above equation by $b_{a,k}$ and sum to obtain after simplification:

$$\sum_{k=1}^{m_a} b_{a,k} d(X_{s_i}, \mu_{a,k}) = d(X_{s_i}, \mu) + \frac{1}{2}\sum_{k=1}^{m_i} b_{a,k}(\mu_{a,k} - \mu)^2 d''(X_{s_i}, w_k).$$

This implies that

$$\mathrm{E}[d(X_{s_i}, \mu_{a,l})_t] - d(X_{s_i}, \mathrm{E}\,[\mu_a|\mathbf{b}])_t = \frac{1}{2}\sum_{k=1}^{m_a} b_{a,k}(\mu_{a,k} - \mathrm{E}\,[\mu_a|\mathbf{b}])^2 d''(X_{s_i}, w_k).$$

Since $d(X_i, \mu)$ is a polynomial (in terms of μ), there exists $q, r \in \mathbb{R}$ such that

$$q < d''(X_{s,i}, w_l) < r,$$

which implies

$$\frac{q}{2}\mathrm{Var}[\mu_a|\mathbf{b}] < \frac{1}{2}\sum_{k=1}^{m_a} b_{a,k}(\mu_{a,k} - \mathrm{E}\,[\mu_a|\mathbf{b}])^2 d''(X_{s_i}, w_{a,k}) < \frac{r}{2}\mathrm{Var}[\mu_a|\mathbf{b}].$$

This in turn implies that there exists $\hat{\mu} \in [q, r]$ such that

$$\frac{\hat{\mu}}{2}\sum_{k=1}^{m_a} b_{a,k}(\mu_{a,k} - \mathrm{E}\,[\mu_a|\mathbf{b}])^2 = \frac{1}{2}\sum_{k=1}^{m_a} b_{a,k}(\mu_{a,k} - \mathrm{E}\,[\mu_a|\mathbf{b}])^2 d''(X_{s_i}, w_k)$$

for all $k \in \mathcal{N}$. Thus,

$$\begin{aligned}\mathrm{E}[d(X_{s_i}, \mu_{a,l})_t] - d(X_{s_i}, \mathrm{E}\,[\mu_a|\mathbf{b}])_t &= \mathrm{U}_t(X_{s_i}, \mathbf{b}, a) - \mathrm{U}_t(X_{s_i}, \mathrm{E}[\mu|\mathbf{b}], a)\\ &= \frac{\hat{\mu}}{2}\sum_{k=1}^{m_i} p_{a,k}(\mu_{a,k} - \mathrm{E}\,[\mu_a|\mathbf{b}])^2,\end{aligned}$$

which implies $\mathrm{U}_t(X_{s_i}, \mathbf{b}, a) - \mathrm{U}_t(X_{s_i}, \mathrm{E}[\mu|\mathbf{b}], a) = q\mathrm{Var}[\mu_a|\mathbf{b}]$ for some $q \in \mathbb{R}$. □

Lemma A.11. $\mathrm{V}\,(\mathbf{X}, \mathbf{b}) = \lim_{t\to\infty} \mathrm{V}_t\,(\mathbf{X}, \mathbf{b}) = \inf_{\pi\in\Pi} \mathrm{J}^\pi(\mathbf{X}, \mathbf{b})$

Proof. Under fully observed transition parameters, it is obvious that the continuous system corresponds to the discrete system in the infinite horizon since when there is no learning, this is a very standard uniformization procedure.

Let $\hat{\mathcal{S}} = \{\mathbf{s} \in \mathbb{Z}^n | s_i \in \mathcal{J}_i\}$ and define operator $\mathbb{P}\left(\mathbf{s}|\mathbf{b}\right) = \prod_{i\in\mathcal{N}} b_{i,s_i}$. Furthermore, let $\mathbf{b}^{\mathbf{s}}$ be the vector of ones and zeros corresponding to $\mathbf{s}$ (that is, $b^{\mathbf{s}}_{i,j} = 1$ if $s_i = j$ and is 0 otherwise). For these "fully known" beliefs, it is easy to show that $\mathrm{V}\left(\mathbf{X}, \mathbf{b}^{\mathbf{s}}\right) = \mathrm{J}^{\pi}(\mathbf{X}, \mathbf{b}^{\mathbf{s}})$ (via the above argument).

Therefore, since

$$\mathrm{J}^{\pi}(\mathbf{X}, \mathbf{b}) = \sum_{\mathbf{s}\in\hat{\mathcal{S}}} \mathbb{P}\left(\mathbf{s}|\mathbf{b}\right) \mathrm{J}^{\pi}(\mathbf{X}, \mathbf{b}^{\mathbf{s}}) = \sum_{\mathbf{s}\in\hat{\mathcal{S}}} \mathbb{P}\left(\mathbf{s}|\mathbf{b}\right) \mathrm{V}^{\pi}(\mathbf{X}, \mathbf{b}^{\mathbf{s}}),$$

it is sufficient to show that $\mathrm{V}^{\pi}\left(\mathbf{X}, \mathbf{b}\right) = \sum_{\mathbf{s}\in\hat{\mathcal{S}}} \mathbb{P}\left(\mathbf{s}|\mathbf{b}\right) \mathrm{V}^{\pi}(\mathbf{X}, \mathbf{b}^{\mathbf{s}})$ to prove the assertion.

Proceeding via induction, the initial step is clear since

$$\mathrm{V}^{\pi}_0\left(\mathbf{X}, \mathbf{b}\right) = \mathbf{c}\mathbf{X}^{\mathrm{T}} = \sum_{\mathbf{s}\in\hat{\mathcal{S}}} \mathbb{P}\left(\mathbf{s}|\mathbf{b}\right) \mathrm{V}^{\pi}_0(\mathbf{X}, \mathbf{b}^{\mathbf{s}}).$$

For the inductive step, note that

$$\mathbb{P}(\mathbf{s}|\sigma(\mathbf{b}, a, +)) = \frac{\mu_{a,\mathbf{s}_a}}{\mathrm{E}\left[\mu_a|\mathbf{b}\right]}\mathbb{P}(\mathbf{s}|\mathbf{b})$$

and similarly

$$\mathbb{P}(\mathbf{s}|\sigma(\mathbf{b}, a, -)) = \frac{(1-\mu_{a,\mathbf{s}_a})}{(1-\mathrm{E}\left[\mu_a|\mathbf{b}\right])}\mathbb{P}(\mathbf{s}|\mathbf{b})$$

so that

$$\begin{aligned}
\mathrm{V}^{\pi}_{t+1}\left(\mathbf{X}, \mathbf{b}\right) &= \mathbf{c}\mathbf{X}^{\mathrm{T}} + \beta\Bigg[\mathrm{E}\left[\mu_a|\mathbf{b}\right]\mathrm{V}^{\pi}_t\left(\mathbf{X} - \mathbf{e}_a, \sigma(\mathbf{b}, a, +)\right) \\
&\quad + (1-\mathrm{E}\left[\mu_a|\mathbf{b}\right])\mathrm{V}^{\pi}_t\left(\mathbf{X}, \sigma(\mathbf{b}, a, -)\right)\Bigg] \\
&= \mathbf{c}\mathbf{X}^{\mathrm{T}} + \beta\Bigg[\mathrm{E}\left[\mu_a|\mathbf{b}\right]\sum_{\mathbf{s}\in\hat{\mathcal{S}}}\frac{\mathbb{P}\left(\mathbf{s}|\mathbf{b}\right)\mu_{a,\mathbf{s}_a}}{\mathrm{E}\left[\mu_a|\mathbf{b}\right]}\mathrm{V}^{\pi}_t(\mathbf{X} - \mathbf{e}_a, \mathbf{b}^{\mathbf{s}}) \\
&\quad + (1-\mathrm{E}\left[\mu_a|\mathbf{b}\right])\sum_{\mathbf{s}\in\hat{\mathcal{S}}}\frac{\mathbb{P}\left(\mathbf{s}|\mathbf{b}\right)(1-\mu_{a,\mathbf{s}_a})}{(1-\mathrm{E}\left[\mu_a|\mathbf{b}\right])}\mathrm{V}^{\pi}_t(\mathbf{X}, \mathbf{b}^{\mathbf{s}})\Bigg] \\
&= \mathbf{c}\mathbf{X}^{\mathrm{T}} + \beta\Bigg[\sum_{\mathbf{s}\in\hat{\mathcal{S}}}\mathbb{P}\left(\mathbf{s}|\mathbf{b}\right)\mu_{a,\mathbf{s}_a}\mathrm{V}^{\pi}_t(\mathbf{X} - \mathbf{e}_a, \mathbf{b}^{\mathbf{s}}) \\
&\quad + \sum_{\mathbf{s}\in\hat{\mathcal{S}}}\mathbb{P}\left(\mathbf{s}|\mathbf{b}\right)(1-\mu_{a,\mathbf{s}_a})\mathrm{V}^{\pi}_t(\mathbf{X}, \mathbf{b}^{\mathbf{s}})\Bigg]
\end{aligned}$$

$$= \sum_{\mathbf{s}\in\hat{\mathcal{S}}} \mathbb{P}\left(\mathbf{s}|\mathbf{b}\right) \Big\{ \mathbf{c}\mathbf{X}^{\mathrm{T}} + \beta \Big[\mu_{a,\mathbf{s}_a} \mathrm{V}_t^{\pi}(\mathbf{X}-\mathbf{e}_a, \mathbf{b}^{\mathbf{s}})$$

$$+ (1-\mu_{a,\mathbf{s}_a}) \mathrm{V}_t^{\pi}(\mathbf{X}, \mathbf{b}^{\mathbf{s}}) \Big] \Big\}$$

$$= \sum_{\mathbf{s}\in\hat{\mathcal{S}}} \mathbb{P}\left(\mathbf{s}|\mathbf{b}\right) \mathrm{V}_{t+1}^{\pi}(\mathbf{X}, \mathbf{b}^{\mathbf{s}})$$

where the action a is chosen by policy π. This concludes the proof. □

LEMMA A.12. $\hat{\mathrm{V}}\left(\mathbf{X},\mathbf{b}\right) = \lim_{t\to\infty} \hat{\mathrm{V}}_t\left(\mathbf{X},\mathbf{b}\right) = \inf_{\pi\in\Pi} \hat{\mathrm{J}}^{\pi}(\mathbf{X},\mathbf{b})$

Proof. We prove the lemma in a similar fashion to the proof of Lemma A.11. Under fully observed transition parameters, it is obvious that the continuous system corresponds to the discrete system in the infinite horizon since when there is no learning, this is a very standard uniformization procedure.

Let $\hat{\mathcal{S}} = \{\mathbf{s} \in \mathbb{Z}^n | s_i \in \mathcal{J}_i\}$ and define operator $\mathbb{P}\left(\mathbf{s}|\mathbf{b}\right) = \prod_{i\in\mathcal{N}} b_{i,s_i}$. Furthermore, let $\mathbf{b}^{\mathbf{s}}$ be the vector of ones and zeros corresponding to $\mathbf{s}$ (that is, $b^{\mathbf{s}}_{i,j} = 1$ if $s_i = j$ and is 0 otherwise). For these "fully known" beliefs, it is easy to show that $\hat{\mathrm{V}}\left(\mathbf{X},\mathbf{b}^{\mathbf{s}}\right) = \hat{\mathrm{J}}^{\pi}(\mathbf{X},\mathbf{b}^{\mathbf{s}})$ (via the above argument).

Therefore, since

$$\hat{\mathrm{J}}^{\pi}(\mathbf{X},\mathbf{b}) = \sum_{\mathbf{s}\in\hat{\mathcal{S}}} \mathbb{P}\left(\mathbf{s}|\mathbf{b}\right) \hat{\mathrm{J}}^{\pi}(\mathbf{X},\mathbf{b}^{\mathbf{s}}) = \sum_{\mathbf{s}\in\hat{\mathcal{S}}} \mathbb{P}\left(\mathbf{s}|\mathbf{b}\right) \hat{\mathrm{V}}^{\pi}(\mathbf{X},\mathbf{b}^{\mathbf{s}}),$$

it is sufficient to show that $\hat{\mathrm{V}}^{\pi}\left(\mathbf{X},\mathbf{b}\right) = \sum_{\mathbf{s}\in\hat{\mathcal{S}}} \mathbb{P}\left(\mathbf{s}|\mathbf{b}\right) \hat{\mathrm{V}}^{\pi}(\mathbf{X},\mathbf{b}^{\mathbf{s}})$ to prove the assertion.

Proceeding via induction, the initial step is clear since

$$\hat{\mathrm{V}}_0^{\pi}\left(\mathbf{X},\mathbf{b}\right) = \mathbf{c}\mathbf{X}^{\mathrm{T}} = \sum_{\mathbf{s}\in\hat{\mathcal{S}}} \mathbb{P}\left(\mathbf{s}|\mathbf{b}\right) \hat{\mathrm{V}}_0^{\pi}(\mathbf{X},\mathbf{b}^{\mathbf{s}}).$$

For the inductive step, note that

$$\mathbb{P}(\mathbf{s}|\sigma(\mathbf{b},a,+)) = \frac{\mu_{a,\mathbf{s}_a}}{\mathrm{E}\left[\mu_a|\mathbf{b}\right]} \mathbb{P}(\mathbf{s}|\mathbf{b})$$

and similarly

$$\mathbb{P}(\mathbf{s}|\sigma(\mathbf{b},a,-)) = \frac{(1-\sum_{i\in\mathcal{N}}\lambda_i - \mu_{a,\mathbf{s}_a})}{(1-\sum_{i\in\mathcal{N}}\lambda_i - \mathrm{E}\left[\mu_a|\mathbf{b}\right])}\mathbb{P}(\mathbf{s}|\mathbf{b})$$

so that

$$\begin{aligned}
\hat{\mathrm{V}}^{\pi}_{t+1}\left(\mathbf{X},\mathbf{b}\right) &= \mathbf{c}\mathbf{X}^{\mathrm{T}} + \beta\Big[\mathrm{E}\left[\mu_a|\mathbf{b}\right]\hat{\mathrm{V}}^{\pi}_{t}\left(\mathbf{X}-\mathbf{e}_a,\hat{\sigma}(\mathbf{b},a,+)\right)\\
&+\sum_{i\in\mathcal{N}}\lambda_i\hat{\mathrm{V}}^{\pi}_{t}\left(\mathbf{X}+\mathbf{e}_i,\mathbf{b}\right)\\
&+(1-\sum_{i\in\mathcal{N}}\lambda_i-\mathrm{E}\left[\mu_a|\mathbf{b}\right])\hat{\mathrm{V}}^{\pi}_{t}\left(\mathbf{X},\hat{\sigma}(\mathbf{b},a,-)\right)\Big]\\
&=\mathbf{c}\mathbf{X}^{\mathrm{T}}+\beta\Big[\mathrm{E}\left[\mu_a|\mathbf{b}\right]\sum_{\mathbf{s}\in\hat{\mathcal{S}}}\frac{\mathbb{P}\left(\mathbf{s}|\mathbf{b}\right)\mu_{a,\mathbf{s}_a}}{\mathrm{E}\left[\mu_a|\mathbf{b}\right]}\hat{\mathrm{V}}^{\pi}_{t}(\mathbf{X}-\mathbf{e}_a,\mathbf{b}^{\mathbf{s}})\\
&+\sum_{i\in\mathcal{N}}\lambda_i\sum_{s\in\hat{\mathcal{S}}}\mathbb{P}(\mathbf{s}|\mathbf{b})\hat{\mathrm{V}}^{\pi}_{t}\left(\mathbf{X}+\mathbf{e}_i,\mathbf{b}^{\mathbf{s}}\right)\\
&+(1-\sum_{i\in\mathcal{N}}\lambda_i-\mathrm{E}\left[\mu_a|\mathbf{b}\right])\sum_{\mathbf{s}\in\hat{\mathcal{S}}}\frac{\mathbb{P}\left(\mathbf{s}|\mathbf{b}\right)(1-\sum_{i\in\mathcal{N}}\lambda_i-\mu_{a,\mathbf{s}_a})}{(1-\sum_{i\in\mathcal{N}}\lambda_i-\mathrm{E}\left[\mu_a|\mathbf{b}\right])}\hat{\mathrm{V}}^{\pi}_{t}(\mathbf{X},\mathbf{b}^{\mathbf{s}})\Big]\\
&=\mathbf{c}\mathbf{X}^{\mathrm{T}}+\beta\Big[\sum_{\mathbf{s}\in\hat{\mathcal{S}}}\mathbb{P}\left(\mathbf{s}|\mathbf{b}\right)\mu_{a,\mathbf{s}_a}\mathrm{V}^{\pi}_{t}(\mathbf{X}-\mathbf{e}_a,\mathbf{b}^{\mathbf{s}})\\
&+\sum_{s\in\hat{\mathcal{S}}}\mathbb{P}(\mathbf{s}|\mathbf{b})\sum_{i\in\mathcal{N}}\lambda_i\hat{\mathrm{V}}^{\pi}_{t}\left(\mathbf{X}+\mathbf{e}_i,\mathbf{b}^{\mathbf{s}}\right)\\
&+\sum_{\mathbf{s}\in\hat{\mathcal{S}}}\mathbb{P}\left(\mathbf{s}|\mathbf{b}\right)(1-\sum_{i\in\mathcal{N}}\lambda_i-\mu_{a,\mathbf{s}_a})\mathrm{V}^{\pi}_{t}(\mathbf{X},\mathbf{b}^{\mathbf{s}})\Big]\\
&=\sum_{\mathbf{s}\in\hat{\mathcal{S}}}\mathbb{P}\left(\mathbf{s}|\mathbf{b}\right)\Big\{\mathbf{c}\mathbf{X}^{\mathrm{T}}+\beta\Big[\mu_{a,\mathbf{s}_a}\mathrm{V}^{\pi}_{t}(\mathbf{X}-\mathbf{e}_a,\mathbf{b}^{\mathbf{s}})\\
&+\sum_{i\in\mathcal{N}}\hat{\mathrm{V}}^{\pi}_{t}\left(\mathbf{X}+\mathbf{e}_i,\mathbf{b}^{\mathbf{s}}\right)+(1-\sum_{i\in\mathcal{N}}\lambda_i-\mu_{a,\mathbf{s}_a})\mathrm{V}^{\pi}_{t}(\mathbf{X},\mathbf{b}^{\mathbf{s}})\Big]\Big\}\\
&=\sum_{\mathbf{s}\in\hat{\mathcal{S}}}\mathbb{P}\left(\mathbf{s}|\mathbf{b}\right)\mathrm{V}^{\pi}_{t+1}(\mathbf{X},\mathbf{b}^{\mathbf{s}})
\end{aligned}$$

where the action a is chosen by policy π. This concludes the proof. □

LEMMA A.13. *In the Poisson arrivals case, for all* $a\in\mathcal{A}(\mathbf{X}), t\in\mathbb{N}\bigcup\{0\}$, *and* $\mathbf{b}\in\mathcal{B}$, $\hat{\mathrm{V}}_t\left(\mathbf{X}-\mathbf{e}_a,\mathbf{b}\right)<\hat{\mathrm{V}}_t\left(\mathbf{X},\mathbf{b}\right)$.

Proof. Consider $\hat{\mathrm{V}}_t^{\pi_{\mathbf{b}}}(\mathbf{X}, \mathbf{b})$ and $\hat{\mathrm{V}}_t^{\pi}(\mathbf{X} - \mathbf{e}_a, \mathbf{b})$ where $\pi_{\mathbf{b}}$ is an optimal stationary policy to the system with starting state $(\mathbf{X}, \mathbf{b})$ and π is a policy that idles whenever the system associated with starting state $(\mathbf{X}, \mathbf{b})$ under policy $\pi_{\mathbf{b}}$ attempts to serve the first customer of class a and is otherwise identical to $\pi_{\mathbf{b}}$. We refer to these as system 1 and system 2 respectively. Note that enacting such a policy is possible because the decision-maker under system 2 knows the probabilities associated with each of the sample paths of system 1 and hence can base a policy on the events in system 1.

Now, the probability of any given parameter configuration is identical for both systems since both have the same starting belief $\mathbf{b}$. Therefore, the transition probabilities (the chance of serving or failing to serve a given customer) under any action is identical for both systems. Hence, the difference in cost between the two systems after $\pi_{\mathbf{b}}$ has served the first customer of class a is identical, but the difference between the two systems for every other period is exactly $\beta^t \mathbf{c}\mathbf{e}_a^{\mathrm{T}}$ since system 1 always holds an identical number of customers to system 2 with the exception of the additional customer of class a that has yet to be successfully served. Hence $\hat{\mathrm{V}}_t^{\pi}(\mathbf{X} - \mathbf{e}_a, \mathbf{b}) < \hat{\mathrm{V}}_t^{\pi_{\mathbf{b}}}(\mathbf{X}, \mathbf{b})$ which implies that $\hat{\mathrm{V}}_t(\mathbf{X} - \mathbf{e}_a, \mathbf{b}) < \hat{\mathrm{V}}_t(\mathbf{X}, \mathbf{b})$. □

LEMMA A.14. *In the Poisson arrivals case, for all $a \in \mathcal{A}(\mathbf{X}), t \in \mathbb{N} \bigcup \{0\}$, and $\mathbf{b} \in \mathcal{B}$, $\hat{\mathrm{V}}_t(\mathbf{X} - \mathbf{e}_a, \hat{\sigma}(\mathbf{b}, a, +)) < \hat{\mathrm{V}}_t(\mathbf{X}, \hat{\sigma}(\mathbf{b}, a, -))$.*

Proof. Similar to Lemma A.13, consider $\hat{\mathrm{V}}_t^{\pi_{\hat{\sigma}(\mathbf{b},a,-)}}(\mathbf{X}, \hat{\sigma}(\mathbf{b}, a, -))$ and $\hat{\mathrm{V}}_t^{\pi}(\mathbf{X} - \mathbf{e}_a, \hat{\sigma}(\mathbf{b}, a, +))$ where $\pi_{\hat{\sigma}(\mathbf{b},a,-)}$ is an optimal stationary policy to the system with starting state $(\mathbf{X}, \hat{\sigma}(\mathbf{b}, a, -))$ and π is a policy that idles whenever the system associated with starting state $(\mathbf{X}, \hat{\sigma}(\mathbf{b}, a, -))$ under policy $\pi_{\hat{\sigma}(\mathbf{b},a,-)}$ attempts to serve a customer of class a and the system with starting state $(\mathbf{X} - \mathbf{e}_a, \hat{\sigma}(\mathbf{b}, a, +))$ has fewer customers of class a and is otherwise identical to $\pi_{\mathbf{b}}$. We refer to these as system 1 and system 2 respectively. Note that enacting such a policy is possible because the decision-maker under system 2 knows the probabilities associated with each of the sample paths of system 1 and hence can base a policy on the events in system 1.

Now, the probability of any given parameter configuration is identical for both systems with the exception of class a customers where system 2 has a higher probability of serving customers since for any $\mu_{a,j}$, the probability $\mathbb{P}_{\hat{\sigma}(\mathbf{b},a,-)}\left(\mu_a^* \geq \mu_{a,j}\right) \leq \mathbb{P}_{\hat{\sigma}(\mathbf{b},a,+)}\left(\mu_a^* \geq \mu_{a,j}\right)$. This implies that every time a customer of class a is successfully served by system 2, that customer may or may not be served by system 1. In this way, on any given sample path, system 1 is guaranteed to have at least as many customers of class a as system 2, and an equal number of customers for every other class. Since the initial system holding costs are $\mathbf{c}\mathbf{X}^{\mathrm{T}}$ and $\mathbf{c}\left(\mathbf{X}-\mathbf{e}_a\right)^{\mathrm{T}}$, the inequality is strict, which proves the lemma.

□

Lemma A.15 (**Value Function Nonincreasing on Line Segments: Arrivals Case**). *For any* $\mathbf{b} \in \mathcal{B}$ *and* $\eta \in [0,1]$, $\hat{\mathrm{V}}(\mathbf{X}, \eta\mathbf{b}_0 + (1-\eta)\mathbf{b})$ *is nonincreasing as* η *increases. Similarly,* $\hat{\mathrm{V}}(\mathbf{X}, \eta\mathbf{b}_1 + (1-\eta)\mathbf{b})$ *is nondecreasing as* η *increases.*

Proof. For the first half of the proof, similar to Lemma A.8, choosing δ such that $\eta - \delta > 0$, we proceed by induction on t. In the base case, when $t = 0$, the assertion is true since $\hat{\mathrm{V}}_0(\mathbf{X}, (\eta-\delta)\mathbf{b}_0 + (1-\eta+\delta)\mathbf{b}) = \mathbf{c}\mathbf{X}^{\mathrm{T}} = \hat{\mathrm{V}}_0(\mathbf{X}, \eta\mathbf{b}_0 + (1-\eta)\mathbf{b})$.

For the inductive step, we suppose the assertion holds for t. Then, suppose that the optimal action for $\hat{\mathrm{V}}_{t+1}(\mathbf{X}, \eta\mathbf{b}_0 + (1-\eta)\mathbf{b})$ is action $a \in \mathcal{A}(\mathbf{X})$. Then,

$$
\begin{aligned}
&\hat{\mathrm{V}}_{t+1}\left(\mathbf{X},(\eta-\delta)\mathbf{b}_0+(1-\eta+\delta)\mathbf{b}\right)\\
&\leq \mathbf{c}\mathbf{X}^{\mathrm{T}}+\beta\Bigg[\bigg\{\mathrm{E}\left[\mu_a|(\eta-\delta)\mathbf{b}_0+(1-\eta+\delta)\mathbf{b}\right]\hat{\mathrm{V}}_t\left(\mathbf{X}-\mathbf{e}_a,\hat{\sigma}\left((\eta-\delta)\mathbf{b}_0+(1-\eta+\delta)\mathbf{b},a,+\right)\right)\\
&+\sum_{i\in\mathcal{N}}\lambda_i\hat{\mathrm{V}}_t\left(\mathbf{X}+\mathbf{e}_i,(\eta-\delta)\mathbf{b}_0+(1-\eta+\delta)\mathbf{b}\right)\\
&+(1-\sum_{i\in\mathcal{N}}\lambda_i-\mathrm{E}\left[\mu_a|(\eta-\delta)\mathbf{b}_0+(1-\eta+\delta)\mathbf{b}\right])\hat{\mathrm{V}}_t\left(\mathbf{X},\hat{\sigma}\left((\eta-\delta)\mathbf{b}_0+(1-\eta+\delta)\mathbf{b},a,-\right)\right)\bigg\}\Bigg]\\
&\leq \mathbf{c}\mathbf{X}^{\mathrm{T}}+\beta\Bigg[\bigg\{\mathrm{E}\left[\mu_a|\eta\mathbf{b}_0+(1+\delta)\mathbf{b}\right]\hat{\mathrm{V}}_t\left(\mathbf{X}-\mathbf{e}_a,\hat{\sigma}\left((\eta-\delta)\mathbf{b}_0+(1-\eta+\delta)\mathbf{b},a,+\right)\right)\\
&+\sum_{i\in\mathcal{N}}\lambda_i\hat{\mathrm{V}}_t\left(\mathbf{X}+\mathbf{e}_i,(\eta-\delta)\mathbf{b}_0+(1-\eta+\delta)\mathbf{b}\right)\\
&+(1-\sum_{i\in\mathcal{N}}\lambda_i-\mathrm{E}\left[\mu_a|\eta\mathbf{b}_0+(1-\eta)\mathbf{b}\right])\hat{\mathrm{V}}_t\left(\mathbf{X},\hat{\sigma}\left((\eta-\delta)\mathbf{b}_0+(1-\eta+\delta)\mathbf{b},a,-\right)\right)\bigg\}\Bigg]\\
&\leq \mathbf{c}\mathbf{X}^{\mathrm{T}}+\beta\Bigg[\bigg\{\mathrm{E}\left[\mu_a|\eta\mathbf{b}_0+(1+\delta)\mathbf{b}\right]\hat{\mathrm{V}}_t\left(\mathbf{X}-\mathbf{e}_a,\hat{\sigma}\left(\eta\mathbf{b}_0+(1-\eta)\mathbf{b},a,+\right)\right)\\
&+\sum_{i\in\mathcal{N}}\lambda_i\hat{\mathrm{V}}_t\left(\mathbf{X}+\mathbf{e}_i,\eta\mathbf{b}_0+(1-\eta)\mathbf{b}\right)\\
&+(1-\sum_{i\in\mathcal{N}}\lambda_i-\mathrm{E}\left[\mu_a|\eta\mathbf{b}_0+(1-\eta)\mathbf{b}\right])\hat{\mathrm{V}}_t\left(\mathbf{X},\hat{\sigma}\left(\eta\mathbf{b}_0+(1-\eta)\mathbf{b},a,-\right)\right)\bigg\}\Bigg]\\
&=\hat{\mathrm{V}}_{t+1}\left(\mathbf{X},\eta\mathbf{b}_0+(1-\eta)\mathbf{b}\right),
\end{aligned}
$$

because for any $\hat{\mathbf{X}}$,

$$
\hat{\mathrm{V}}_t\left(\hat{\mathbf{X}},\hat{\sigma}\left((\eta-\delta)\mathbf{b}_0+(1-\eta+\delta)\mathbf{b},a,\theta\right)\right)\leq\hat{\mathrm{V}}_t\left(\hat{\mathbf{X}},\hat{\sigma}\left(\delta\mathbf{b}_0+(1-\eta)\mathbf{b},a,\theta\right)\right),
$$

by the inductive hypothesis, and since for any $\hat{\mathbf{b}}$,

$$
\hat{\mathrm{V}}_t\left(\hat{\mathbf{X}}-\mathbf{e}_a,\hat{\sigma}\left(\hat{\mathbf{b}},a,+\right)\right)<\hat{\mathrm{V}}_t\left(\hat{\mathbf{X}},\hat{\sigma}\left(\hat{\mathbf{b}},a,-\right)\right),
$$

by Lemma A.14.

Now, we remind the reader from the proof of Proposition 2.1, $\mathbf{b}_1$ is the belief with components

$$
b_{i,j}^1=\begin{cases}1 & : \text{ if } \mu_{i,j}=\max_{k\in\mathcal{J}_i}\mu_{i,k}\\ 0 & : \text{ otherwise.}\end{cases}
$$

Following the same strategy of the first portion of the proof, the base case with $t = 0$ is trivial. For the inductive step, we suppose the assertion holds for t. Then, suppose that the optimal action for $\hat{\mathrm{V}}_{t+1}(\mathbf{X}, \eta\mathbf{b} + (1-\eta)\mathbf{b}_1)$ is action $a \in \mathcal{A}(\mathbf{X})$. Then, with $\delta > 0$ such that $\eta - \delta \geq 0$,

$$\begin{aligned}
&\hat{\mathrm{V}}_{t+1}\left(\mathbf{X}, (\eta-\delta)\mathbf{b} + (1-\eta+\delta)\mathbf{b}_1\right) \\
&\leq \mathbf{c}\mathbf{X}^{\mathrm{T}} + \beta \Bigg[\Bigg\{ \mathrm{E}\left[\mu_a | (\eta-\delta)\mathbf{b} + (1-\eta+\delta)\mathbf{b}_1\right] \hat{\mathrm{V}}_t\left(\mathbf{X} - \mathbf{e}_a, \hat{\sigma}\left((\eta-\delta)\mathbf{b} + (1-\eta+\delta)\mathbf{b}_1, a, +\right)\right) \\
&+ \sum_{i \in \mathcal{N}} \lambda_i \hat{\mathrm{V}}_t\left(\mathbf{X} + \mathbf{e}_i, (\eta-\delta)\mathbf{b} + (1-\eta+\delta)\mathbf{b}_1\right) \\
&+ (1 - \sum_{i \in \mathcal{N}} \lambda_i - \mathrm{E}\left[\mu_a | (\eta-\delta)\mathbf{b} + (1-\eta+\delta)\mathbf{b}_1\right]) \hat{\mathrm{V}}_t\left(\mathbf{X}, \hat{\sigma}\left((\eta-\delta)\mathbf{b} + (1-\eta+\delta)\mathbf{b}_1, a, -\right)\right) \Bigg\}\Bigg] \\
&\leq \mathbf{c}\mathbf{X}^{\mathrm{T}} + \beta \Bigg[\Bigg\{ \mathrm{E}\left[\mu_a | \eta\mathbf{b} + (1+\delta)\mathbf{b}_1\right] \hat{\mathrm{V}}_t\left(\mathbf{X} - \mathbf{e}_a, \hat{\sigma}\left((\eta-\delta)\mathbf{b} + (1-\eta+\delta)\mathbf{b}_1, a, +\right)\right) \\
&+ \sum_{i \in \mathcal{N}} \lambda_i \hat{\mathrm{V}}_t\left(\mathbf{X} + \mathbf{e}_i, (\eta-\delta)\mathbf{b} + (1-\eta+\delta)\mathbf{b}_1\right) \\
&+ (1 - \sum_{i \in \mathcal{N}} \lambda_i - \mathrm{E}\left[\mu_a | \eta\mathbf{b} + (1-\eta)\mathbf{b}_1\right]) \hat{\mathrm{V}}_t\left(\mathbf{X}, \hat{\sigma}\left((\eta-\delta)\mathbf{b} + (1-\eta+\delta)\mathbf{b}_1, a, -\right)\right) \Bigg\}\Bigg] \\
&\leq \mathbf{c}\mathbf{X}^{\mathrm{T}} + \beta \Bigg[\Bigg\{ \mathrm{E}\left[\mu_a | \eta\mathbf{b} + (1+\delta)\mathbf{b}_1\right] \hat{\mathrm{V}}_t\left(\mathbf{X} - \mathbf{e}_a, \hat{\sigma}\left(\eta\mathbf{b} + (1-\eta)\mathbf{b}_1, a, +\right)\right) \\
&+ \sum_{i \in \mathcal{N}} \lambda_i \hat{\mathrm{V}}_t\left(\mathbf{X} + \mathbf{e}_i, \eta\mathbf{b} + (1-\eta)\mathbf{b}_1\right) \\
&+ (1 - \sum_{i \in \mathcal{N}} \lambda_i - \mathrm{E}\left[\mu_a | \eta\mathbf{b} + (1-\eta)\mathbf{b}_1\right]) \hat{\mathrm{V}}_t\left(\mathbf{X}, \hat{\sigma}\left(\eta\mathbf{b} + (1-\eta)\mathbf{b}_1, a, -\right)\right) \Bigg\}\Bigg] \\
&= \hat{\mathrm{V}}_{t+1}\left(\mathbf{X}, \eta\mathbf{b} + (1-\eta)\mathbf{b}_1\right),
\end{aligned}$$

because for any $\hat{\mathbf{X}}$,

$$\hat{\mathrm{V}}_t\left(\hat{\mathbf{X}}, \hat{\sigma}\left((\eta-\delta)\mathbf{b} + (1-\eta+\delta)\mathbf{b}_1, a, \theta\right)\right) \leq \hat{\mathrm{V}}_t\left(\hat{\mathbf{X}}, \hat{\sigma}\left(\delta\mathbf{b} + (1-\eta)\mathbf{b}_1, a, \theta\right)\right),$$

by the inductive hypothesis, and since for any $\hat{\mathbf{b}}$,

$$\hat{\mathrm{V}}_t\left(\hat{\mathbf{X}} - \mathbf{e}_a, \hat{\sigma}\left(\hat{\mathbf{b}}, a, +\right)\right) < \hat{\mathrm{V}}_t\left(\hat{\mathbf{X}}, \hat{\sigma}\left(\hat{\mathbf{b}}, a, -\right)\right),$$

by Lemma A.14.

□

Lemma A.16 (**Minimax/Minimin $c\mu$ Optimal Policies: Arrivals Case**). *At any state* $(\mathbf{X}, \mathbf{b})$*, the optimal policies to the minimax and minimin objectives within the Poisson arrivals case serve classes* $\arg\max_{a \in \mathcal{A}(\mathbf{X})}(\min_{j \in \mathcal{J}_a} c_a \mu_{a,j})$ *and* $\arg\max_{a \in \mathcal{A}(\mathbf{X})}(\max_{j \in \mathcal{J}_a} c_a \mu_{a,j})$*, respectively.*

Proof. Using Lemma A.15, we can prove the lemma using a condensed argument of that used in Proposition 2.1. First, since our system is identical to that of Buyukkoc et al. (1985) when the belief is composed of only ones and zeros, the $c\mu$ policy (we denote π_0) that prioritizes $\arg\max_{a \in \mathcal{A}(\mathbf{X})}(\min_{j \in \mathcal{J}_a} c_a \mu_{a,j})$ is optimal to the system when the belief is $\mathbf{b}_0$ and the policy (we denote π_1) $\arg\max_{a \in \mathcal{A}(\mathbf{X})}(\max_{j \in \mathcal{J}_a} c_a \mu_{a,j})$ is optimal to the system when the belief is $\mathbf{b}_1$ respectively. The value function is concave and formed from the minimum of a set of hyperplanes (as a result of being a POMDP). Now, $\hat{\mathrm{V}}^{\pi_0}(\mathbf{X}, \mathbf{b})$ is one of these hyperplanes, and due to Lemma A.15, $\arg\max_{\mathbf{b} \in \mathcal{B}} \hat{\mathrm{V}}^{\pi_0}(\mathbf{X}, \mathbf{b}) = \mathbf{b}_0$ hence we know that $\hat{\mathrm{R}}(\mathbf{X}) = \hat{\mathrm{V}}^{\pi_0}(\mathbf{X}, \mathbf{b}_0)$. Using this same logic, $\arg\min_{\mathbf{b} \in \mathcal{B}} \hat{\mathrm{V}}^{\pi_1}(\mathbf{X}, \mathbf{b}) = \mathbf{b}_1$ hence we know that $\hat{\mathrm{N}}(\mathbf{X}) = \hat{\mathrm{V}}^{\pi_1}(\mathbf{X}, \mathbf{b}_1)$. □

Proof of Corollary 2.1. Each E$c\mu$ policy forms a hyperplane over the belief space so that an E$c\mu$ policy based on initial state $\hat{\mathbf{b}}$ is evaluated as $\mathrm{V}^{\pi^{c\mu}_{\hat{\mathbf{b}}}}(\mathbf{X}, \mathbf{b}) = \mathbf{v}\mathbf{b}^{\mathrm{T}} + v$ for all $\mathbf{b} \in \mathcal{B}$. Now consider the set of E$c\mu$ policies expressed as a set $\mathcal{V} = \{(\mathbf{v}, v) \in \mathbb{R}^n \times \mathbb{R} | \exists \hat{\mathbf{b}} \in \mathcal{B} \text{ s.t. } \mathrm{V}^{\pi^{c\mu}_{\hat{\mathbf{b}}}}(\mathbf{X}, \mathbf{b}) = \mathbf{v}\mathbf{b}^{\mathrm{T}} + v\}$.

Consider the function $\mathrm{Z}(\mathbf{b}) = \min_{(\mathbf{v},v) \in \mathcal{V}} \mathbf{v}\mathbf{b}^{\mathrm{T}} + v$ defined for $\mathbf{b} \in \mathcal{B}$. Let $\mathrm{Z}^{\bar{\pi}_{\hat{\mathbf{b}}}}(\mathbf{b}) = \sum_{i=1}^{r}(\mathbf{v}_i\mathbf{b}^{\mathrm{T}} + v_i)p_i$ where $\mathbf{v}_i\mathbf{b}^{\mathrm{T}} + v_i = \mathrm{Z}(\hat{\mathbf{b}})$ for all $(\mathbf{v}_i, v_i) \in \mathcal{V}$ for $\{1, 2, \ldots, r\}$ denote the evaluation of a finite randomization of E$c\mu$ policies based on belief point $\hat{\mathbf{b}}$. Note that $\mathrm{Z}^{\bar{\pi}_{\hat{\mathbf{b}}}}(\mathbf{b}) = \mathrm{V}^{\bar{\pi}_{\hat{\mathbf{b}}}}(\mathbf{X}, \mathbf{b})$. For the proof we define Z to avoid confusion since $\bar{\pi}_{\hat{\mathbf{b}}}$ may be composed of policies that are not within $\pi^{c\mu}_{\hat{\mathbf{b}}}$. Instead, $\bar{\pi}_{\hat{\mathbf{b}}}$ is composed of policies that satisfy $\min_{\mathbf{b} \in \mathcal{B}} \mathrm{V}^{\pi^{c\mu}_{\mathbf{b}}}(\mathbf{X}, \hat{\mathbf{b}})$.

$\mathrm{Z}(\mathbf{b})$ is concave since it is the minimum of concave functions. Furthermore, since the traditional $c\mu$ policy is optimal to $\mathrm{V}(\mathbf{X}, \mathbf{b}_0)$ as established in Proposition 2.1, we are assured that $\mathrm{Z}(\mathbf{b})$ is non-decreasing on linear segments from $\mathbf{b}_0$ via the same argument in Proposition 2.1.

Note that the only properties that Theorem 2.2 and Proposition 2.4 relied on was the concavity of $\mathrm{V}(\mathbf{X}, \mathbf{b})$, and the fact that it is non-increasing on lines from $\mathbf{b}_0$, hence if we replace $\mathrm{V}(\mathbf{X}, \mathbf{b})$ with $\mathrm{Z}(\mathbf{b})$ in these proofs, we are guaranteed that there exists $\bar{\pi}_{\hat{\mathbf{b}}}$, a finite randomization of E$c\mu$ policies based on point $\hat{\mathbf{b}} = \arg\max_{\mathbf{b}\in\delta\mathcal{L}_\epsilon} \mathrm{Z}(\mathbf{b})$ such that $\mathbb{P}_{\mathbf{B}}\left(\mathrm{Z}^{\bar{\pi}_{\hat{\mathbf{b}}}}(\mathbf{B}) \geq \mathrm{Z}^{\bar{\pi}_{\hat{\mathbf{b}}}}(\hat{\mathbf{b}})\right) \leq \epsilon$. Hence, $\mathrm{Y}^{\bar{\pi}_{\hat{\mathbf{b}}}}(\mathbf{X}, \epsilon) = \mathrm{Z}^{\bar{\pi}_{\hat{\mathbf{b}}}}(\hat{\mathbf{b}})$.

Now, $\mathrm{Z}^{\bar{\pi}_{\hat{\mathbf{b}}}}(\hat{\mathbf{b}}) \leq \max_{\mathbf{b}\in\delta\mathcal{L}_\epsilon} \mathrm{V}^{\pi^{c\mu}_{\mathbf{b}}}(\mathbf{X}, \mathbf{b})$, so we have that

$$\mathrm{Y}^{\bar{\pi}_{\hat{\mathbf{b}}}}(\mathbf{X}, \epsilon) - \mathrm{Y}(\mathbf{X}, \epsilon) = \mathrm{Z}^{\bar{\pi}_{\hat{\mathbf{b}}}}(\hat{\mathbf{b}}) - \mathrm{V}(\mathbf{X}, \mathbf{b}^*) \leq \max_{\mathbf{b}\in\delta\mathcal{L}_\epsilon} \mathrm{V}^{\pi^{c\mu}_{\mathbf{b}}}(\mathbf{X}, \mathbf{b}) - \mathrm{V}(\mathbf{X}, \mathbf{b}^*)$$

□

Lemma A.17 (**Non-idling under Arrivals**). *There exists an optimal non-idling policy to a system under general arrivals*

Proof of Lemma A.17. The proof follows via a simple sample path argument. □

Consider a system that experiences arrivals in the following manner: If the system at time t has a positive level of customers, for any non-idling policy, let t' denote the first time after t that the policy π clears the system of a given class (i.e., $X_i^\pi(t') = 0$ and $X_i^\pi(t' - \delta) = 1$ for small enough $\delta > 0$). Let t'' be the first time in which, under π, $\mathbf{X}(t'') = \mathbf{0}$. If the probability of a path occurring such that an arrival in the time interval (t', t'') is bounded by p, then we say that the arrival process is bursty at level $(1 - p)$. We show in Corollary A.3, that as $p \to 0$, the E$c\mu$ policy becomes asymptotically optimal to such a system.

It is useful to consider bursty arrival processes since it naturally occurs in systems that become heavily overloaded due to high levels of arrivals for a period of time then experience extended lull periods that allows the system to clear. For example, consider a multiclass queue with arrivals that occur as an interrupted Poisson process (which is a type of Markov modulated Poisson process). In this system, alongside the queue and belief state, another state $I(t)$ denotes whether the system is open ($I(t) = 0$) or closed ($I(t) = 1$). When $I(t) = 0$, arrivals of each class occur according to a Poisson process with rate λ_i. However, when the system is closed ($I(t) = 1$), no arrivals occur. The open system transitions to the closed system in an exponentially distributed amount of time with rate ω, and likewise, the closed system transitions to the open system in an exponentially distributed amount of time with rate γ. We assume that the system is stable, which can always be ensured by large enough ω and small enough γ.

Now, to show that such a system is arbitrarily bursty, (in accordance with our definition), we first show that the probability of serving a class to zero during a busy period goes to zero as ω and λ_i increase. Let us consider the scenario that maximizes this probability, that is, policies that serve a single class with priority. If the initial state for a given class of customers is $X_i(0) = X_i > 0$, and the system is "open", (i.e., $I(0) = 0$), then the initial state of interest is $(I(0), X_i(0)) = (0, X_i)$. Then, under the policy that prioritizes class i, (say π_i), if we let p_1 be the probability that $(I^{\pi_i}(t), X_i^{\pi_i}(t)) = (0, 0)$ before $(I^{\pi_i}(t), X_i^{\pi_i}(t)) = (1, X_i^{\pi_i}(t))$, (that is, p_1 is the probability that the system clears the system of class i before to transitioning to the "closed" system), then, assuming that the true service parameter is given by μ_i,

$$p_1 = \sum_{t=X_i}^{\infty} \left(\frac{\lambda}{\mu_i + \lambda_i + \omega}\right)^{t-1} \left(\frac{\mu_i}{\mu_i + \lambda_i + \omega}\right)^{t} = \frac{(\frac{\lambda_i}{\mu_i+\lambda_i+\omega})^{X_i}(\frac{\mu_i}{\mu_i+\lambda_i+\omega})^{X_i}(\omega + \lambda_i + \mu_i)^3}{\lambda(\omega^2 + 2\omega\lambda + \lambda^2 + 2\omega\mu_i + \lambda\mu_i + \mu_i^2)},$$

since the number of services that must occur to empty the system of class i customers is exactly X_i plus the number of arrivals that occur to the system. Obviously, p_1 goes to zero as ω and λ increase. Now, under any non-idling policy π let us examine the probability of a state $(I(0), \mathbf{X}(0)) = (1, \mathbf{X}(0))$ transitioning to $(1, \mathbf{0})$ before $(0, \mathbf{X}^{\pi}(t))$ where $\mathbf{X}(t) > 0$. That is, let p_2 be the probability of the system clearing under any non-idling policy before it transitions to an open system. Then, p_2 can easily be expressed as

$$p_2 = \left(\frac{\mu_1}{\mu_1 + \gamma}\right)^{X_1} \left(\frac{\mu_2}{\mu_2 + \gamma}\right)^{X_2} \cdots \left(\frac{\mu_n}{\mu_n + \gamma}\right)^{X_n},$$

which goes to 1 as $\gamma \to 0$. Since this can be done for arbitrarily large starting state, and the system can be made stable for large enough ω, the probability of seeing a path under any non-idling policy that experiences an arrival in the time interval (t', t'') is bounded.

As another example of a bursty process, consider the batch process for which arrivals occur according to a Poisson process with parameter λ in batches according to some distribution f. Thus, a batch containing customers $\mathbf{X}$ occur according to f at exponentially distributed time periods with parameter λ. If the current system state is given by $\mathbf{X}(0) = \{X_1, X_2, \ldots, X_n\}$ and we let p_3 be the probability of clearing the system under any non-idling policy before another batch arrival occurs, this can be calculated in a similar manner to p_2 by

$$p_3 = \left(\frac{\mu_1}{\mu_1 + \lambda}\right)^{X_1} \left(\frac{\mu_2}{\mu_2 + \lambda}\right)^{X_2} \cdots \left(\frac{\mu_n}{\mu_n + \lambda}\right)^{X_n}.$$

Naturally, p_3 increases to 1 as $\lambda \to 0$, so if the system is stable (which can be ensured by small enough λ), p must be bounded.

Corollary A.3 (E$c\mu$ **with Bursty Arrivals**). *The* E$c\mu$ *policy is asymptotically optimal to a system with bursty arrivals as* $p \to 0$.

Proof of Corollary A.3. By Lemma A.17, there exists an optimal policy that is non-idling for any system with arrivals. Let π be such a policy. Furthermore, let π_0 be the policy that always idles. Hence $\mathbf{X}^{\pi_0}(t)$ is the cumulative number of customers from each class within the system at time t when no services are completed. Hence, if we define $\mathbf{W}^{\pi}(t) = \{W_1^{\pi}(t), W_2^{\pi}(t), \ldots, W_n^{\pi}(t)\} \in \mathbb{Z}_+^n$ as the number of customers in each class successfully served by policy π at time t, it is obvious that $\mathbf{X}^{\pi}(t) = \mathbf{X}^{\pi_0}(t) - \mathbf{W}^{\pi}(t)$ hence we can write

$$\begin{aligned}
&\operatorname{arginf}_{\pi\in\Pi}\mathrm{E}_{\pi,\mathbf{b}(0)}\left[\int_{t=0}^{\infty} e^{-\alpha t}\hat{\mathbf{c}}\mathbf{X}^{\pi}\left(t\right)^{\mathrm{T}} dt|\mathbf{X}\left(0\right)\right] \\
&\qquad = \operatorname{arginf}_{\pi\in\Pi}\mathrm{E}_{\pi,\mathbf{b}(0)}\left[\int_{t=0}^{\infty} e^{-\alpha t}\hat{\mathbf{c}}(\mathbf{X}^{\pi_0}\left(t\right) - \mathbf{W}^{\pi}\left(t\right))^{\mathrm{T}} dt|\mathbf{X}\left(0\right)\right] \\
&\qquad = \operatorname{argsup}_{\pi\in\Pi}\mathrm{E}_{\pi,\mathbf{b}(0)}\left[\int_{t=0}^{\infty} e^{-\alpha t}\hat{\mathbf{c}}\mathbf{W}^{\pi}\left(t\right)^{\mathrm{T}} dt|\mathbf{X}\left(0\right)\right].
\end{aligned}$$

Now, for any two non-idling policies π_1, π_2 $\mathbf{X}^{\pi_1}(t) = \mathbf{0}$ implies $\mathbf{X}^{\pi_2}(t) = \mathbf{0}$ and likewise $\mathbf{X}^{\pi_2}(t) = \mathbf{0}$ implies $\mathbf{X}^{\pi_1}(t) = \mathbf{0}$ since the cumulative service time remains the same across policies. Therefore, at the time in which non-idling policies experience a zero-state, they have identical queue state as well as observation histories (i.e., belief). Hence, letting t'' be the first time to a zero-state (cleared system) under a non-idling policy,

$$\begin{aligned}
&\operatorname{arginf}_{\pi\in\Pi}\mathrm{E}_{\pi,\mathbf{b}(0)}\left[\int_{t=0}^{\infty} e^{-\alpha t}\hat{\mathbf{c}}\mathbf{X}^{\pi}\left(t\right)^{\mathrm{T}} dt|\mathbf{X}\left(0\right)\right] = \\
&\qquad \operatorname{arginf}_{\pi\in\Pi}\mathrm{E}_{\pi,\mathbf{b}(0)}\left[\int_{t=0}^{t''} e^{-\alpha t}\hat{\mathbf{c}}\mathbf{X}^{\pi}\left(t\right)^{\mathrm{T}} dt|\mathbf{X}\left(0\right)\right]
\end{aligned}$$

which can be expressed (as before) as

$$\begin{aligned}
&\operatorname{arginf}_{\pi\in\Pi}\mathrm{E}_{\pi,\mathbf{b}(0)}\left[\int_{t=0}^{t''} e^{-\alpha t}\hat{\mathbf{c}}(\mathbf{X}^{\pi_0}\left(t\right) - \mathbf{W}^{\pi}\left(t\right))^{\mathrm{T}} dt + \int_{t=t''}^{\infty} e^{-\alpha t}\hat{\mathbf{c}}(\mathbf{X}^{\pi_0}\left(t''\right) - \mathbf{W}^{\pi}\left(t''\right))^{\mathrm{T}} dt|\mathbf{X}\left(0\right)\right] \\
&= \operatorname{argsup}_{\pi\in\Pi}\mathrm{E}_{\pi,\mathbf{b}(0)}\left[\int_{t=0}^{t''} e^{-\alpha t}\hat{\mathbf{c}}\mathbf{W}^{\pi}\left(t\right)^{\mathrm{T}} dt + \int_{t=t''}^{\infty} e^{-\alpha t}\hat{\mathbf{c}}\mathbf{W}^{\pi}\left(t''\right)^{\mathrm{T}} dt|\mathbf{X}\left(0\right)\right].
\end{aligned}$$

Now, suppose that the arrival times for the first arrival of each class i before t'' occur at t_i. If there are no arrivals, let $t_i = \infty$. Consider a system closely related to our original non-robust clearing system and the arrivals systems discussed above.

With starting state

$$\hat{\mathbf{X}}(0) = \{\mathbb{1}\{X_1(0) > 0\} X_1^{\pi_0}(t''), \ldots, \mathbb{1}\{X_n(0) > 0\} X_n^{\pi_0}(t'')\},$$

the only arrivals that can occur in this system are from those classes with $\hat{X}_i(0) = 0$. If arrivals of this class do occur, they occur at the same moment in time, specifically at time t_i so that $\hat{X}_i^{\pi}(t_i) = X_i^{\pi_0}(t'')$ regardless of policy π, and no further arrivals of this class can occur. Such a system is a branching-bandit (or arm-acquiring bandit) version of our original clearing system, hence, the E$c\mu$ policy is asymptotically optimal to the system since the index-rule is still optimal, and the class with highest E$c\mu$ value that is non-empty should be served in this system.

Now, in this branching-bandit system, we know that $\hat{\mathbf{X}}^{\pi}(t'') = 0$, under any non-idling policy π since the cumulative time to serve the customers remains the same as the previous arrivals system. Furthermore, so long as no arrivals occur after the first time a class is cleared of customers, $\hat{\mathbf{W}}^{\pi}(t) = \mathbf{W}^{\pi}(t)$ since the two systems have the same classes available to them at each moment in these cases, and hence can serve customers at the same times. Due to the bursty condition on arrivals, this occurs at least $(1-p)$ of the time, since the probability of clearing a class before the last arrival prior to t'' is bounded by p. Hence, since $\hat{\mathbf{W}}^{\pi_{c\mu}}(t)$ is asymptotically optimal to the branching bandit, and is identical to the bursty arrivals case at least $(1-p)$ of the time, the E$c\mu$ policy is asymptotically optimal to the bursty arrivals case at least $(1-p)$ of the time, hence E$c\mu$ is optimal as $p \to 0$. □

APPENDIX B

ROBUST FORECASTING AND INVENTORY MANAGEMENT

B.1 India Numerical Example

For our numerical examples, we provide the parameters determined by the nominal model, and the ambiguity set we design surrounding them.

In our first example to the outreach center with population 10,749, we assume bimonthly outreach sessions. Since BCG/HepB and JE/Measles share immunization schedules, it is reasonable to assume that the demand for these vaccines are identical (or close to identical) in each period. Hence, we fit a 6-dimensional VAR(p) with lag 2 where each dimension represents BCG/HepB, TT, JE/Measles, DTP, Penta, and OPV for dimensions $1, \ldots, 6$ respectively to a time series with perfect vaccine adherence. As per WHO recommendations (see, e.g. WHO (2010)), the expected number of newborns at each session for our population can be given by $10,749 \times .03 \times 1/24 = 13.44$. Hence, to account for the variability in these births, we let a $\mathrm{N}(13.44, 3)$ random variable dictate the arrivals to the system. Fitting the data to a VAR(2) yields

$$\mathbf{a}_0 = \begin{pmatrix} 11.672 \\ 28.104 \\ 27.092 \\ 11.638 \\ 3.817 \\ 29.16 \end{pmatrix}, A_1 = \begin{pmatrix} -0.073 & 0.023 & -0.048 & -0.002 & -0.019 & 0.044 \\ -0.016 & -0.067 & -0.036 & 0.01 & -0.042 & 0.023 \\ -0.063 & 0.025 & -0.079 & 0.014 & -0.025 & 0.027 \\ -0.008 & 0.026 & 0.042 & -0.059 & 0.356 & -0.007 \\ 0.009 & 0.001 & -0.01 & -0.399 & -0.013 & 0.009 \\ -0.124 & 0.03 & -0.107 & -0.403 & -0.034 & 0.064 \end{pmatrix},$$

$$A_2 = \begin{pmatrix} 0.048 & -0.002 & 0.027 & -0.01 & 0.023 & -0.009 \\ 0.082 & 0.036 & 0.018 & 0.031 & 0.033 & -0.044 \\ -0.073 & 0.015 & 0.019 & 0.039 & -0.050 & 0.030 \\ 0.006 & -0.018 & -0.003 & 0.019 & 0.025 & 0.004 \\ 1.028 & 0.020 & 0.005 & 0.019 & 0.813 & -0.007 \\ 0.985 & 0.039 & 0.012 & 0.008 & 0.792 & 0.046 \end{pmatrix},$$

$$\Omega = \begin{pmatrix} 8.885 & -0.240 & -0.295 & -0.298 & -0.151 & 8.485 \\ -0.240 & 17.673 & -0.866 & -0.134 & -0.045 & -0.847 \\ -0.295 & -0.866 & 17.600 & -0.150 & 0.169 & 8.675 \\ -0.298 & -0.134 & -0.150 & 15.103 & -0.017 & -0.692 \\ -0.151 & -0.045 & 0.169 & -0.017 & 3.406 & 3.306 \\ 8.485 & -0.847 & 8.675 & -0.692 & 3.306 & 20.476 \end{pmatrix}.$$

To fit a model describing vaccine demand to the IHC as a whole to a 6-dimensional VAR(p) with lag 2, we must consider other factors such as wastage rates in our considerations. According to WHO (WHO (2000) and WHO (2017)), about 4.4 liters of cold storage volume is required per 10,000 population for a 6 week's worth of stock when the volume to fully immunize a child is 54.2. This implies that, when we use a wastage factor of 1.4, the average stock used in a month is

$$\frac{4400}{54.2} \times \frac{57734}{10000} \times 98.42 \times \frac{4}{6} = 30752.3$$

or, 30.752 liters. Then, a 6 week stock (as recommended by WHO) necessitates 46.128 liters of storage capacity. Calibrating this to our model and treating inevitable wastes (via the wastage factor) as fictitious demands, we let arrivals occur according to N(156.23, 30.) random variables

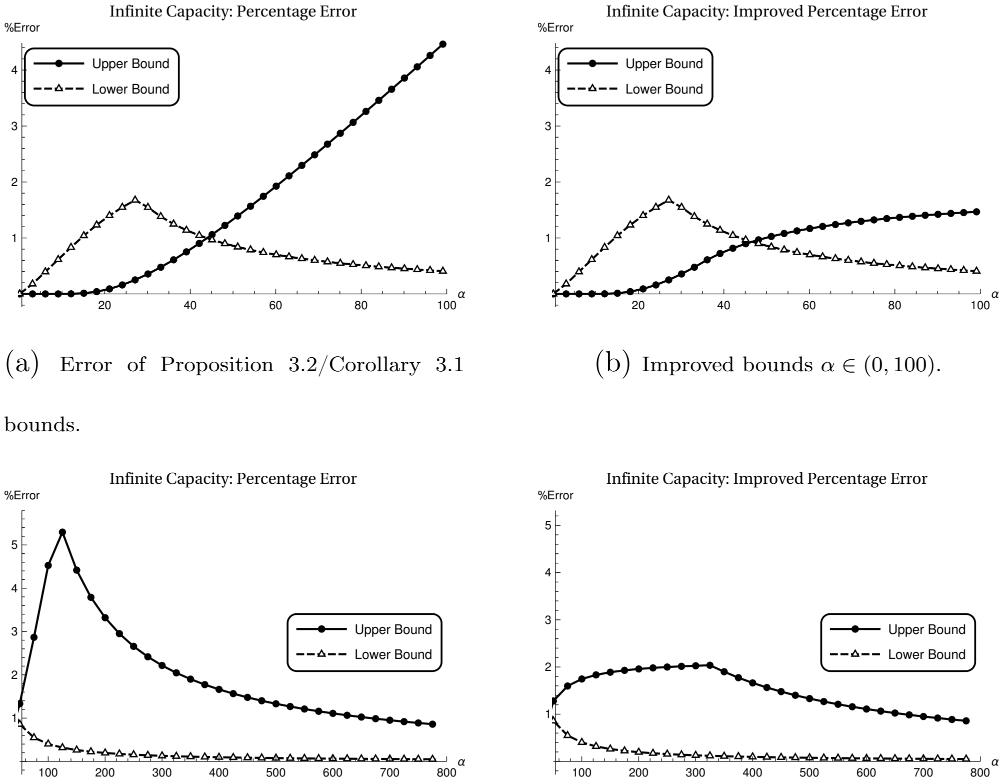
$$\mathbf{a}_0 = \begin{pmatrix} 236.390 \\ 440.357 \\ 433.995 \\ 221.564 \\ 67.704 \\ 529.555 \end{pmatrix}, A_1 = \begin{pmatrix} -0.023 & 0.019 & -0.023 & -0.029 & -0.026 & 0.018 \\ -0.072 & -0.002 & -0.004 & 0.009 & -0.052 & 0.04 \\ -0.039 & -0.006 & 0.003 & 0.009 & -0.047 & 0.034 \\ 0.007 & -0.027 & -0.001 & -0.023 & 0.31 & 0.003 \\ 0.065 & 0.005 & 0.032 & -0.405 & 0.054 & -0.047 \\ 0.081 & 0.015 & 0.023 & -0.425 & 0.043 & -0.052 \end{pmatrix},$$

$$A_2 = \begin{pmatrix} 0.053 & 0.008 & 0.033 & -0.003 & 0.03 & -0.043 \\ 0.045 & 0.006 & -0.044 & 0.033 & 0.021 & -0.016 \\ 0.072 & -0.001 & 0.025 & -0.013 & 0.024 & -0.027 \\ 0.052 & 0.023 & 0.021 & -0.006 & 0.002 & 0.003 \\ 0.999 & 0.003 & 0.005 & -0.004 & 0.77 & 0.024 \\ 1.043 & 0.029 & 0.049 & -0.014 & 0.771 & -0.013 \end{pmatrix},$$

$$\Omega = \begin{pmatrix} 1529.42 & 12.581 & -45.207 & 59.274 & 6.77 & 1522.11 \\ 12.581 & 3061.15 & 63.552 & -34.332 & 25.446 & 60.144 \\ -45.207 & 63.552 & 3119.46 & -53.924 & 17.026 & 1534.83 \\ 59.274 & -34.332 & -53.924 & 2624.7 & 11.727 & 31.727 \\ 6.77 & 25.446 & 17.026 & 11.727 & 583.566 & 603.764 \\ 1522.11 & 60.144 & 1534.83 & 31.727 & 603.764 & 3656.55 \end{pmatrix}.$$

B.2 Bound Performance

We investigate the analytical bounds (see Proposition 3.2, 3.4, and Corollary 3.1) in the infinite capacity case. In the discussion following Proposition 3.4, we observed that the bounds of Proposition 3.2 and Corollary 3.1 could be tightened by exploiting the monotonicity of $\hat{x}_i^\alpha$ in α. To understand the magnitude of this effect, we evaluate the percentage gap, calculated as $|\widetilde{x} - \hat{x}^{*\alpha}|/\hat{x}^{*\alpha}$, between $\hat{x}_i^\alpha$ and the upper/lower bounds with and without these improvements via the expression, where $\widetilde{x}$ represents the upper or lower bound to $\hat{x}^{*\alpha}$. Comparing the non-improved percentage gap shown in Figures B.1a and B.1a against their improved counterparts, shown in Figures B.1b and B.1d, indicates that these improvements can significantly tighten the bounds until $\hat{x}_i + \frac{u_i \sigma_i^2}{\alpha}$ becomes the upper bound at $\alpha \approx 325$. These bounds can provide valuable service to a policy-maker who wishes to determine an appropriate level of ambiguity. Since $\hat{x}_i^\alpha$ acts as a "target" ordering quantity, if α^* is found to imply an extremely large $\hat{x}_i^{*\alpha}$ (via our easily calculable bounds), the level of capacity necessary to carry these quantities can be assessed, allowing for the policy-maker to more easily



(a) Error of Proposition 3.2/Corollary 3.1 bounds.

(b) Improved bounds $\alpha \in (0, 100)$.

(c) Error of Proposition 3.2/Corollary 3.1 bounds.

(d) Improved Bounds $\alpha \in (50, 800)$.

Figure B.1: Percentage error of the infinite capacity bounds via Proposition 3.2 and Corollary 3.1 and improved bounds (via Proposition 3.4) for a BCG vaccine with $\mu_i = 220$, $\sigma_i = 40$, $h_i = 0.75$, $u_i = 1$, and $s_i = 0$ with large and small α.

gauge whether his/her level of optimism is overly pessimistic.

Figure B.1 also demonstrates the asymptotic results implied by the bounds of Proposition 3.2 as α becomes large, or approaches 0. Interestingly, Figure B.1b shows that the lower bounds quickly reach $\hat{x}_i$ as α increases, whereas the upper bounds cling closer to the upper bound for small values of α. This behavior suggests that when the level of ambiguity is large, the upper bound provides a better approximation for $\hat{x}_i^\alpha$ than the lower bound. However, when the ambiguity is small, the lower bound becomes a better approximation for $\hat{x}_i^\alpha$ than the upper bound.

To study the performance of our bounds under larger/smaller variance, Figures B.2a and B.2b show the percentage gap on the improved bounds (via Proposition 3.4) when σ_i is set to 60 and 20. As expected, since Figure B.2a experiences a percentage gap as large as 2.5% as opposed to the less than 1% gap of Figure B.2b, larger variances correspond to higher percentage gap, though even with these large deviations in variance, our bounds remain relatively close throughout the spectrum of α. Additionally, we can see that the α necessary to make $\hat{x}_i + \frac{u_i\sigma_i^2}{\alpha}$ the lower bound is increasing in σ_i, which implies that the upper bound remains a closer bound to $\hat{x}_i^\alpha$ for larger α as σ_i increases.

Changes in cost result in similar behaviors to those experienced in σ_i. Figures B.2c and B.2d consider the case when h_i is 0.85 and 0.5; comparing these with Figure B.1b shows that as h_i/u_i approach 1, bounds are observed to become tighter. Hence, for a policy-maker who places equal weight on missed opportunities between vaccines, requirements for vaccines with large overage costs (i.e., vaccines that are expensive or highly subject to deterioration) can be more accurately estimated than those with low overage costs. Furthermore, α necessary to make $\hat{x}_i + \frac{u_i\sigma_i^2}{\alpha}$ the lower bound increases as h_i/u_i approach 1. This is because the term $\frac{\alpha \ln(u_i/h_i)}{u_i+h_i}$ becomes small as h_i approaches u_i.

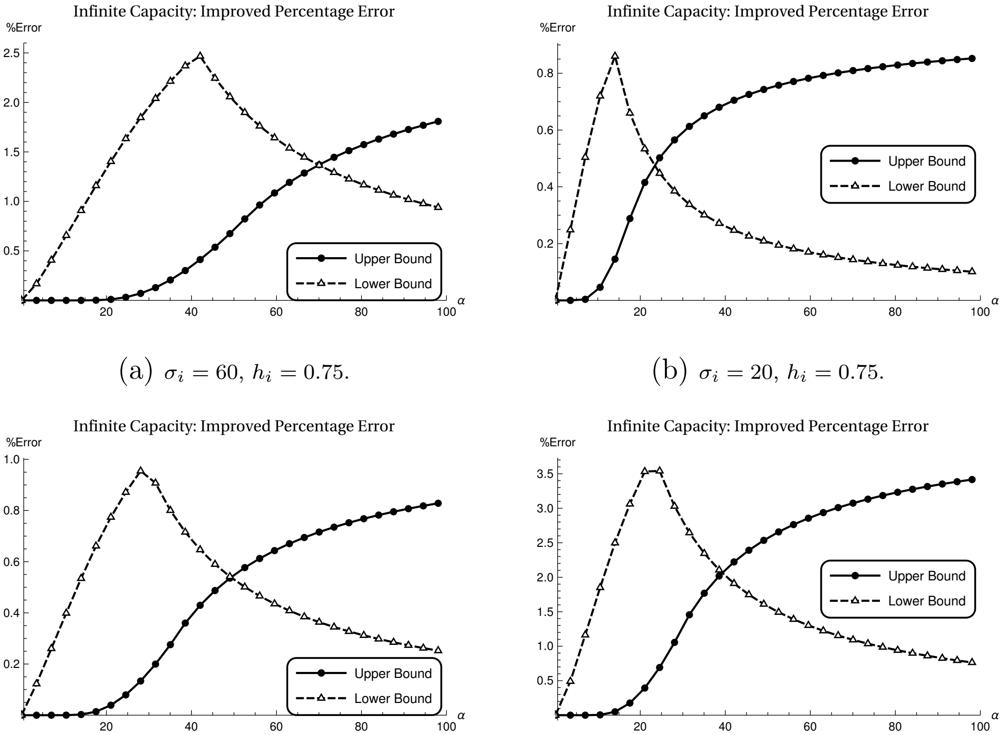
B.3 Proofs of Propositions, Lemmas, and Theorems

Lemma B.1 (π_i **Properties**). *For $i \in \mathcal{N}$, if $\hat{x}_i > s_i$,*

1. *$\pi_i(x_i)$ is non-decreasing.*

2. *$-u_i/w_i \leq \pi_i(x_i)$ for all $x_i \geq 0$.*

3. *$\pi_i^{-1}(q)$ is strictly increasing on $q \geq -u_i/w_i$.*



(a) $\sigma_i = 60$, $h_i = 0.75$.

(b) $\sigma_i = 20$, $h_i = 0.75$.

(c) $\sigma_i = 40$, $h_i = 0.85$.

(d) $\sigma_i = 40$, $h_i = 0.5$.

Figure B.2: Percentage error of the infinite capacity bounds via Propositions 3.2, 3.4, and Corollary 3.1 for a vaccine with $\mu_i = 220$, $u_i = 1$, and $s_i = 0$ with bound improvements from Proposition 3.4.

Proof. Property (i) is easily established since $F_i(x_i)$ is non-decreasing. Due to this fact, for all $x_i \in [0, x_i^*]$,

$$-u_i \leq -u_i + (h_i + u_i)F_i(x_i) \leq -u_i + (h_i + u_i)F_i(x_i^*) = 0.$$

Dividing by w_i gives result (ii): $-u_i/w_i \leq \pi_i(x_i) \leq 0$. Result (iii) follows easily from our definition of F_i^{-1}.

□

Proof of Proposition 3.1: We want to show that the problem

$$\min_{\mathbf{x} \in \mathcal{X}(\mathbf{s})} \mathrm{E}_f \left[\sum_{i=1}^{n} H_i(x_i, s_i) \right]$$

$$\text{s.t.} \ \sum_{i=1}^{n} \int_{\mathcal{V}_i} f_i(v_i) \ln \frac{f_i(v_i)}{\hat{f}_i(v_i)} dv_i \le \eta$$

is equivalent to solving

$$\underset{\mathbf{x}\in\mathcal{X}(\mathbf{s}),\alpha\ge 0}{\text{minimize}} \ \alpha \sum_{i=1}^{n} \ln \mathrm{E}_{\hat{f}} \left[e^{H_i(x_i,V_i)/\alpha} \right] + \alpha\eta.$$

To show this, we define the density

$$g(\mathbf{v}) = \prod_{i=1}^{n} \hat{f}_i(v_i)$$

which is a density that induces independence on the components of $\mathbf{V}$. Let $\mathcal{V} = \{\mathbf{v} \in \mathbb{R}^n | g(\mathbf{v}) > 0\}$ and consider the problem

$$\min_{\mathbf{x}\in\mathcal{X}(\mathbf{s})} \mathrm{E}_f \left[\sum_{i=1}^{n} H_i(x_i, V_i) \right] \tag{B.1}$$

$$\text{s.t.} \ \int_{\mathcal{V}} f(\mathbf{v}) \ln \frac{f(\mathbf{v})}{g(\mathbf{v})} d\mathbf{v} \le \eta.$$

Now, Hu and Hong (2012) show that Problem (B.1) can be solved via

$$\underset{\mathbf{x}\in\mathcal{X}(\mathbf{s}),\alpha\ge 0}{\text{minimize}} \ \alpha \ln \mathrm{E}_g \left[e^{\sum_{i=1}^{n} H_i(x_i,V_i)/\alpha} \right] + \alpha\eta \tag{B.2}$$

so long as there exists $\alpha > 0$ such that $\mathrm{E}_g \left[e^{\sum_{i=1}^{n} H_i(x_i,V_i)/\alpha} \right]$ is finite, which is an equivalent condition to $\sum_{i=1}^{n} \mathrm{E}_{\hat{f}} \left[e^{H_i(x_i,V_i)/\alpha} \right] < \infty$. Now, Problem (B.2) is equivalent to

$$\underset{\mathbf{x}\in\mathcal{X}(\mathbf{s}),\alpha\ge 0}{\text{minimize}} \ \alpha \sum_{i=1}^{n} \ln \mathrm{E}_{\hat{f}} \left[e^{H_i(x_i,V_i)/\alpha} \right] + \alpha\eta$$

since the marginals of g and $\hat{f}$ are identical by definition.

Hu and Hong (2012) also show that we can recast (B.1) as

$$\text{maximize} \ \mathrm{E}_g \left[L(\mathbf{V}) \sum_{i=1}^{n} H_i(x_i, V_i) \right]$$

$$\text{s.t. } \mathrm{E}_g\left[L(\mathbf{V})\ln L(\mathbf{V})\right] \leq \eta.$$

where $L(\mathbf{v}) = f(\mathbf{v})/g(\mathbf{v})$ is the likelihood ratio. Furthermore, they show that the optimal solution to L which we denote L^* takes the form

$$L^*(\mathbf{v}) = \frac{e^{\sum_{i=1}^n H_i(x_i,v_i)/\alpha}}{\mathrm{E}_g\left[e^{\sum_{i=1}^n H_i(x_i,V_i)/\alpha}\right]} = \prod_{i=1}^n \frac{e^{H_i(x_i,v_i)/\alpha}}{\mathrm{E}_g\left[e^{H_i(x_i,V_i)/\alpha}\right]}$$

via the independence obtained by density g. Since L^* can be expressed in product form and $L^*(\mathbf{v}) = f(\mathbf{v})/g(\mathbf{v})$ we obtain

$$\prod_{i=1}^n \frac{e^{H_i(x_i,v_i)/\alpha}}{\mathrm{E}_g\left[e^{H_i(x_i,V_i)/\alpha}\right]} g_i(v_i) = f(\mathbf{v})$$

which implies that L^* induces an independent distribution f. Hence, we can write

$$L^*(\mathbf{v}) = \prod_{i=1}^n \frac{f_i(v_i)}{g_i(v_i)}.$$

With these results, Problem (B.1) can be restated as

$$\min_{\mathbf{x}\in\mathcal{X}(\mathbf{s})} \mathrm{E}_f\left[\sum_{i=1}^n H_i(x_i, V_i)\right] \tag{B.3}$$

$$\text{s.t. } \int_{\mathcal{V}} f(\mathbf{v}) \ln \frac{f_i(v_i)}{g_i(v_i)} d\mathbf{v} \leq \eta.$$

Now, the constraint of (B.3) can be restated as

$$\int_{\mathcal{V}} f(\mathbf{v}) \ln \prod_{i=1}^n \frac{f_i(v_i)}{g_i(v_i)} d\mathbf{v} = \sum_{i=1}^n \int_{\mathcal{V}_i} f_i(v_i) \ln \frac{f_i(v_i)}{g_i(v_i)} dv_i = \sum_{i=1}^n \int_{\mathcal{V}_i} f_i(v_i) \ln \frac{f_i(v_i)}{\hat{f}_i(v_i)} dv_i$$

since the marginals of g and $\hat{f}$ are obviously identical (via the definition of g. This means that Problem (B.1) can be restated as

$$\min_{\mathbf{x}\in\mathcal{X}(\mathbf{s})} \mathrm{E}_f\left[\sum_{i=1}^n H_i(x_i, V_i)\right]$$

$$\text{s.t. } \sum_{i=1}^n \int_{\mathcal{V}_i} f_i(v_i) \ln \frac{f_i(v_i)}{\hat{f}_i(v_i)} dv_i \leq \eta.$$

and as we have shown above, can be solved via

$$\underset{\mathbf{x}\in\mathcal{X}(\mathbf{s}),\alpha\geq 0}{\text{minimize}}\ \alpha\sum_{i=1}^{n}\ln \mathrm{E}_{\hat{f}}\left[e^{H_i(x_i,V_i)/\alpha}\right]+\alpha\eta.$$

so long as $\sum_{i=1}^{n}\mathrm{E}_{\hat{f}}\left[e^{H_i(x_i,V_i)/\alpha}\right]<\infty$. The fact that the objective is convex follows directly from Hu and Hong (2012), who shows that if $H_i(x_i,v_i)$ is convex in x_i for every v_i, the function $\alpha\ln \mathrm{E}_{\hat{f}}[e^{H_i(x_i,V_i)}]$ is convex in α and x_i.

The fact that

$$f(\mathbf{v})=\prod_{i=1}^{n}\hat{f}_i(v_i)\frac{e^{H_i(x_i,v_i)/\alpha}}{\mathrm{E}_g\left[e^{H_i(x_i,V_i)/\alpha}\right]}.$$

follows directly from the fact that $L^*(\boldsymbol{v})g(\boldsymbol{v})=f(\boldsymbol{v})$.

□

Proof of Lemma 3.1. First, we note that by Hu and Hong (2012), the function $\alpha\ln \mathrm{E}_{\hat{f}}[e^{H_i(x_i,V_i)/\alpha}]$ is convex in α since $H_i(x_i,v_i)$ is convex in x_i for every v_i. Hence, if the derivative is non-positive as $\alpha\to\infty$, the derivative must be non-positive for all $\alpha>0$ since convex functions see increasing derivatives. The partial derivative

$$\begin{aligned}\frac{\partial(\alpha\ln \mathrm{E}_{\hat{f}}[e^{H_i(x_i,V_i)/\alpha}])}{\partial\alpha}&=\ln \mathrm{E}_{\hat{f}}[e^{H_i(x_i,V_i)/\alpha}]-\frac{\mathrm{E}_{\hat{f}}[\frac{H_i(x_i,V_i)}{\alpha}e^{H_i(x_i,V_i)/\alpha}]}{\mathrm{E}_{\hat{f}}[e^{H_i(x_i,V_i)/\alpha}]}\\&=\ln \mathrm{E}_{\hat{f}}[e^{H_i(x_i,V_i)\beta}]-\frac{\mathrm{E}_{\hat{f}}[\beta H_i(x_i,V_i)e^{\beta H_i(x_i,V_i)}]}{\mathrm{E}_{\hat{f}}[e^{\beta H_i(x_i,V_i)}]}\end{aligned}$$

when we substitute $\beta=\frac{1}{\alpha}$. Now,

$$\lim_{\beta\to 0}\left[\ln \mathrm{E}_{\hat{f}}[e^{H_i(x_i,V_i)\beta}]-\frac{\mathrm{E}_{\hat{f}}[\beta H_i(x_i,V_i)e^{\beta H_i(x_i,V_i)}]}{\mathrm{E}_{\hat{f}}[e^{\beta H_i(x_i,V_i)}]}\right]=0,$$

which implies that the derivative is non-positive as $\alpha\to\infty$, hence our function is decreasing in α.

□

Lemma B.2 (**Robust Marginal Properties**). *For $i\in\mathcal{N}$,* *1.* $\pi_i^\alpha(x_i)$ *is non-decreasing.*

2. $-u_i/w_i \leq \pi_i^\alpha(x_i)$ *for all* $x_i \geq 0$.

3. $\pi_i^{\alpha,-1}(q)$ *is strictly increasing on* $q \geq -u_i/w_i$.

Proof. Examining the partial derivative of G reveals

$$G_{x_i}(x_i, \alpha) = \frac{\mathrm{E}_{\hat{f}}[e^{H_i(x_i,v_i)/\alpha} \frac{d}{dx_i} H_i(x_i, v_i)]}{\mathrm{E}_{\hat{f}}[e^{H_i(x_i,v_i)/\alpha}]},$$

hence, when x_i is set to zero (assuming a non-negative distribution $\hat{f}$), we gain

$$G_{x_i}(0, \alpha) = \frac{\mathrm{E}_{\hat{f}}[e^{H_i(x_i,v_i)/\alpha}(-u_i)]}{\mathrm{E}_{\hat{f}}[e^{H_i(x_i,v_i)/\alpha}]} < 0. \tag{B.4}$$

Hence, since by Hu and Hong (2012), the function $\alpha \ln \mathrm{E}_{\hat{f}}[e^{H_i(x_i,V_i)/\alpha}]$ is convex in x_i since $H_i(x_i, v_i)$ is convex in x_i for every v_i and the derivatives of convex functions are monotone, we gain Property (i). Further examining the inequality (B.4), we can see that

$$G_{x_i}(0, \alpha) = -u_i$$
$$G_{x_i}(\hat{x}_i^\alpha, \alpha) \geq 0$$

by the definition of $\hat{x}_i^\alpha$. Hence, Property (ii) is established dividing the expressions by w_i.

Property (iii) follows naturally due to the monotonic properties (i) and (ii).

□

Proof of Theorem 3.1. The Lagrangian Dual problem to

$$\underset{\mathbf{x} \in \mathcal{X}(\mathbf{s}), \alpha \geq 0}{\text{minimize}} \; G(\mathbf{x}, \alpha)$$

can be expressed

$$\underset{\mathbf{k}, \lambda_1, \lambda_2 \geq 0}{\text{maximize}} \left\{ \inf_{\mathbf{x} \in \mathbb{R}^n_+, \alpha \geq 0} G(\mathbf{x}, \alpha) + \sum_{i=1}^n (s_i - x_i) k_i \right. \tag{B.5}$$

$$+ \lambda_1 \sum_{i=1}^{n} (k_i(x_i - s_i) - b_k) + \lambda_2 \sum_{i=1}^{n} (r_i x_i - b_r) \Bigg\}$$

Now, the derivative of the Lagrangian Dual problem with respect to x_i set to zero becomes

$$w_i \pi_i^\alpha(x_i) - k_i + \lambda_1 w_i + \lambda_2 r_i = 0. \tag{B.6}$$

For a zero order, $\pi_i^\alpha(x_i) = k_i/w_i - \lambda_1 - \lambda_2$, hence $\pi_i^\alpha(x_i) \geq \pi_j^\alpha(x_j)$ for any non-zero order $j \in \mathcal{O}_c$ if $i \in \mathcal{O}_c$ or $j \in \mathcal{O}_r$ if $i \in \mathcal{O}_r$.

For a non-zero order, $k_i = 0$, hence, setting the partial to zero results in

$$\pi_i^\alpha(x_i) = -\lambda_1 - \lambda_2$$

for all non-zero orders for $i \in \mathcal{O}_r$ and

$$\pi_i^\alpha(x_i) = -\lambda_1$$

for all non-zero order for $i \in \mathcal{O}_c$ since $w_i = 0$ for all $i \in \mathcal{O}_c$.

(i) If $\sum_{i=1}^n (\hat{x}_i^\alpha - s_i) w_i \leq b_c$ and $\sum_{i=1}^n r_i \hat{x}_i^\alpha \leq b_r$, then the problem is unconstrained and $\lambda_1 = \lambda_2 = 0$, hence $x_i^{*\alpha} = \hat{x}_i^\alpha$ for all $i = 1, \ldots n$.

(ii) If $\sum_{i=1}^n (\hat{x}_i^\alpha - s_i) w_i \leq b_c$ and $\sum_{i=1}^n r_i \hat{x}_i^\alpha > b_r$, then $\lambda_1 = 0$ since it is a non-binding constraint, hence $x_i^{*\alpha} = \hat{x}_i^\alpha$ for all $i \in \mathcal{O}_c$. However, refrigeration is a binding constraint, thus $\lambda_2 > 0$ and $x_i^{*\alpha} = \pi_i^{\alpha,-1}(-\lambda_2)$ for all non-zero orders $i \in \mathcal{O}_r$.

(iii) If $\sum_{i=1}^n (\hat{x}_i^\alpha - s_i) k_i > b_c$ and $\sum_{i=1}^n r_i \hat{x}_i \leq b_r$, then $\lambda_2 = 0$ since it is a non-binding constraint. However, the carrier constraint is still active, thus $\lambda_1 > 0$ and $x_i^{*\alpha} = \pi_i^{\alpha,-1}(-\lambda_1)$ for all non-zero orders $i = 1, \ldots, n$.

(iv) If $\sum_{i=1}^n (\hat{x}_i^\alpha - s_i) k_i > b_c$ and $\sum_{i=1}^n r_i \hat{x}_i^\alpha > b_r$, then $\lambda_1 > 0$. This implies that $x_i^{*\alpha} = \pi_i^{\alpha,-1}(-\lambda_1)$ for all non-zero orders $i \in \mathcal{O}_c$ and $x_i^* = \pi_i^{\alpha,-1}(-\lambda_1 - \lambda_2)$ for

all non-zero orders $i \in \mathcal{O}_r$. This directly implies that $\pi_i(x_i^{*\alpha}) \geq \pi_j(x_j^{*\alpha})$ for all non-zero orders $i \in \mathcal{O}_c$ and $j \in \mathcal{O}_r$.

□

Proposition B.1 (**Expected Exponential Cost Function**). *The expectation within* (3.7) *can be expressed as:*

$$\begin{aligned} \mathrm{E}_{\hat{f}}\left[e^{H_i(x_i,V_i)/\alpha}\right] &= e^{\frac{h_i(x_i-\mu_i)}{\alpha}+\frac{\sigma_i^2h_i^2}{2\alpha^2}}\,\Phi\left(\frac{x_i-\mu_i}{\sigma_i}+\frac{\sigma_ih_i}{\alpha}\right) \\ &\quad + e^{\frac{u_i(\mu_i-x_i)}{\alpha}+\frac{\sigma_i^2u_i^2}{2\alpha^2}}\,\Phi\left(\frac{\mu_i-x_i}{\sigma_i}+\frac{\sigma_iu_i}{\alpha}\right). \end{aligned} \tag{B.7}$$

Proof. We first note that

$$\begin{aligned} \mathrm{E}_{\hat{f}}\left[e^{H_i(x_i,V_i)/\alpha}\right] &= \mathrm{E}_{\hat{f}}\left[\exp\left(\left(u_i(V_i-x_i)^+ + h_i(x_i-V_i)^+\right)/\alpha\right)\right] \\ &= \int_{x_i}^{\infty} \exp\left(u_i(V_i-x_i)/\alpha\right)\hat{f}_i(V_i)dV_i \\ &+ \int_{0}^{x_i} \exp\left(h_i(x_i-V_i)/\alpha\right)\hat{f}_i(V_i)dV_i \end{aligned}$$

Now, we know that the moment generating function for a truncated normal distribution with bounds a and b can be expressed as

$$e^{\mu_it+\sigma_i^2t^2/2}\left[\frac{\Phi((b-\mu_i)/\sigma_i-\sigma_it)-\Phi((a-\mu_i)/\sigma_i-\sigma_it)}{\Phi((b-\mu_i)/\sigma_i)-\Phi((a-\mu_i)/\sigma_i)}\right].$$

Furthermore, noting that the expectations are identical to the moment generating functions without the normalization (which is the denominator of the expression), it is easy to see that the expectations can be decomposed into expectations of truncated distributions

$$\begin{aligned} \int_{x_i}^{\infty} \exp\left(u_i(V_i-x_i)/\alpha\right)\hat{f}_i(V_i)dV_i &= e^{-u_ix_i/\alpha}\int_{x_i}^{\infty} e^{u_iV_i/\alpha}\hat{f}_i(V_i)dV_i \\ &= e^{\frac{u_i\mu_i}{\alpha}+\frac{\sigma_i^2u_i^2}{2\alpha^2}-\frac{u_ix_i}{\alpha}}\left(1-\Phi\left(\frac{x_i-\mu_i}{\sigma_i}-\frac{\sigma_iu_i}{\alpha}\right)\right) \end{aligned}$$

$$= e^{\frac{u_i(\mu_i - x_i)}{\alpha} + \frac{\sigma_i^2 u_i^2}{2\alpha^2}} \Phi\left(\frac{\mu_i - x_i}{\sigma_i} + \frac{\sigma_i u_i}{\alpha}\right),$$

where $\frac{u_i}{\alpha}$ is substituted for t, and

$$\begin{aligned}\int_0^{x_i} \exp\left(h_i(x_i - V_i)^+/\alpha\right) \hat{f}_i(V_i) dV_i &= e^{h_i x_i/\alpha} \int_0^{x_i} e^{-h_i V_i/\alpha} \hat{f}_i(V_i) dV_i \\ &= e^{\frac{-h_i \mu_i}{\alpha} + \frac{\sigma_i^2 h_i^2}{2\alpha^2} + \frac{h_i x_i}{\alpha}} \Phi\left(\frac{x_i - \mu_i}{\sigma_i} + \frac{\sigma_i h_i}{\alpha}\right) \\ &= e^{\frac{h_i(x_i - \mu_i)}{\alpha} + \frac{\sigma_i^2 h_i^2}{2\alpha^2}} \Phi\left(\frac{x_i - \mu_i}{\sigma_i} + \frac{\sigma_i h_i}{\alpha}\right),\end{aligned}$$

where $-\frac{h_i}{\alpha}$ is substituted for t assuming that $\Phi\left(-\frac{\mu_i}{\sigma_i} - \frac{\sigma_i u_i}{\alpha}\right)$ is near zero (which identical to the assumption of positive demand).

□

Proposition B.2 (**Cost Function Partials**). *The partial derivative of* (3.7) *with respect to x_i is given by:*

$$G_{x_i}(x_i, \alpha) = \frac{h_i e^{\frac{h_i(x_i - \mu_i)}{\alpha} + \frac{\sigma_i^2 h_i^2}{2\alpha^2}} \Phi\left(\frac{x_i - \mu_i}{\sigma_i} + \frac{\sigma_i h_i}{\alpha}\right) - u_i e^{\frac{u_i(\mu_i - x_i)}{\alpha} + \frac{\sigma_i^2 u_i^2}{2\alpha^2}} \Phi\left(\frac{\mu_i - x_i}{\sigma_i} + \frac{\sigma_i u_i}{\alpha}\right)}{e^{\frac{h_i(x_i - \mu_i)}{\alpha} + \frac{\sigma_i^2 h_i^2}{2\alpha^2}} \Phi\left(\frac{x_i - \mu_i}{\sigma_i} + \frac{\sigma_i h_i}{\alpha}\right) + e^{\frac{u_i(\mu_i - x_i)}{\alpha} + \frac{\sigma_i^2 u_i^2}{2\alpha^2}} \Phi\left(\frac{\mu_i - x_i}{\sigma_i} + \frac{\sigma_i u_i}{\alpha}\right)} \tag{B.8}$$

Proof. With the form of the expected cost known via Proposition B.1, via the chain rule, it can be shown that

$$\begin{aligned}\frac{\partial}{\partial x_i} e^{\frac{h_i(x_i - \mu_i)}{\alpha} + \frac{\sigma_i^2 h_i^2}{2\alpha^2}} \Phi\left(\frac{x_i - \mu_i}{\sigma_i} + \frac{\sigma_i h_i}{\alpha}\right) &= \frac{h_i \exp\left(\frac{h_i^2 \sigma_i^2}{2\alpha^2} + \frac{h_i(x_i - \mu_i)}{\alpha}\right) \Phi\left(\frac{h_i \sigma}{\alpha} + \frac{x_i - \mu_i}{\sigma_i}\right)}{\alpha} \\ &+ \frac{\exp\left(\frac{h_i^2 \sigma^2}{2\alpha^2} - \frac{1}{2}\left(-\frac{h_i \sigma_i}{\alpha} - \frac{x_i - \mu_i}{\sigma_i}\right)^2 + \frac{h_i(x_i - \mu_i)}{\alpha}\right)}{\sqrt{2\pi}\sigma_i} \\ &= \frac{h_i \exp\left(\frac{h_i^2 \sigma_i^2}{2\alpha^2} + \frac{h_i(x_i - \mu_i)}{\alpha}\right) \Phi\left(\frac{h_i \sigma}{\alpha} + \frac{x_i - \mu_i}{\sigma_i}\right)}{\alpha} \\ &+ \frac{\exp\left(-\frac{\mu_i^2}{2\sigma_i^2} - \frac{x_i^2}{2\sigma_i^2} + \frac{\mu_i x_i}{\sigma_i^2}\right)}{\sqrt{2\pi}\sigma_i}\end{aligned}$$

$$= \frac{h_i \exp\left(\frac{h_i^2\sigma_i^2}{2\alpha^2} + \frac{h_i(x_i-\mu_i)}{\alpha}\right)\Phi\left(\frac{h_i\sigma}{\alpha} + \frac{x_i-\mu_i}{\sigma_i}\right)}{\alpha} + \frac{\exp\left(-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}\right)}{\sqrt{2\pi}\sigma_i}$$

with respect to the first term of the expectation, and

$$\begin{aligned}\frac{\partial}{\partial x_i} e^{\frac{u_i(\mu_i-x_i)}{\alpha}+\frac{\sigma_i^2u_i^2}{2\alpha^2}}\Phi\left(\frac{\mu_i-x_i}{\sigma_i}+\frac{\sigma_iu_i}{\alpha}\right) &= -\frac{u_i\exp\left(\frac{\sigma_i^2u_i^2}{2\alpha^2}+\frac{u_i(\mu_i-x_i)}{\alpha}\right)\Phi\left(\frac{\mu_i-x_i}{\sigma_i}+\frac{\sigma_iu_i}{\alpha}\right)}{\alpha} \\ &\quad - \frac{\exp\left(\frac{\sigma_i^2u_i^2}{2\alpha^2} - \frac{1}{2}\left(\frac{\sigma_iu_i}{\alpha} - \frac{x_i-\mu_i}{\sigma_i}\right)^2 + \frac{u_i(\mu_i-x)}{\alpha}\right)}{\sqrt{2\pi}\sigma_i} \\ &= -\frac{u_i\exp\left(\frac{\sigma_i^2u_i^2}{2\alpha^2}+\frac{u_i(\mu_i-x_i)}{\alpha}\right)\Phi\left(\frac{\mu_i-x_i}{\sigma_i}+\frac{\sigma_iu_i}{\alpha}\right)}{\alpha} \\ &\quad - \frac{\exp\left(-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}\right)}{\sqrt{2\pi}\sigma_i}\end{aligned}$$

with respect to the second term of the expectation. Adding these expressions leads to the cancellation of the second exponential terms, giving

$$\frac{1}{\alpha}\left(h_ie^{\frac{h_i(x_i-\mu_i)}{\alpha}+\frac{\sigma_i^2h_i^2}{2\alpha^2}}\Phi\left(\frac{x_i-\mu_i}{\sigma_i}+\frac{\sigma_ih_i}{\alpha}\right) - u_ie^{\frac{u_i(\mu_i-x_i)}{\alpha}+\frac{\sigma_i^2u_i^2}{2\alpha^2}}\Phi\left(\frac{\mu_i-x_i}{\sigma_i}+\frac{\sigma_iu_i}{\alpha}\right)\right).$$

Now, using the closed form results from Proposition B.1, the partial derivative $G_{x_i}(x_i, \alpha)$ is given by

$$\begin{aligned}&\frac{\partial}{\partial x_i}\alpha\sum_{i=1}^{n}\ln \mathrm{E}_{\hat{f}}\left[e^{H_i(x_i,V_i)/\alpha}\right] + \alpha\eta \\ &= \frac{h_ie^{\frac{h_i(x_i-\mu_i)}{\alpha}+\frac{\sigma_i^2h_i^2}{2\alpha^2}}\Phi\left(\frac{x_i-\mu_i}{\sigma_i}+\frac{\sigma_ih_i}{\alpha}\right) - u_ie^{\frac{u_i(\mu_i-x_i)}{\alpha}+\frac{\sigma_i^2u_i^2}{2\alpha^2}}\Phi\left(\frac{\mu_i-x_i}{\sigma_i}+\frac{\sigma_iu_i}{\alpha}\right)}{e^{\frac{h_i(x_i-\mu_i)}{\alpha}+\frac{\sigma_i^2h_i^2}{2\alpha^2}}\Phi\left(\frac{x_i-\mu_i}{\sigma_i}+\frac{\sigma_ih_i}{\alpha}\right) + e^{\frac{u_i(\mu_i-x_i)}{\alpha}+\frac{\sigma_i^2u_i^2}{2\alpha^2}}\Phi\left(\frac{\mu_i-x_i}{\sigma_i}+\frac{\sigma_iu_i}{\alpha}\right)}.\end{aligned}$$

□

Proof of Proposition 3.2. We prove the middle terms of the bounds found in Proposition 3.2 in the following. To prove the third terms, we separate the proofs into

Lemmas B.3 and B.4 respectively. We begin by noting that the $\max\{s_i, \cdot\}$ term is immediate for each case, since we are restricted from ordering less than is currently in stock. Now, in the first case, when $\lambda w_i < 1$, we begin by establishing the lower bound. We investigate the numerator of (B.8) when $x_i = \mu_i + \frac{\sigma_i^2(u_i - h_i)}{2\alpha}$. In this case,

$$\begin{aligned}\Phi\left(\frac{x_i - \mu_i}{\sigma_i} + \frac{\sigma_i h_i}{\alpha}\right) &= \Phi\left(\frac{(u_i + h_i)\sigma_i}{2\alpha}\right)\\ &= \Phi\left(\frac{\mu_i - x_i}{\sigma_i} + \frac{\sigma_i u_i}{\alpha}\right)\end{aligned}$$

hence, the Φ terms of the positive and negative sides of expression (B.8) are equal. Now, the numerator of (B.8) simplifies to

$$\Phi\left(\frac{(1 + \lambda w_i)\sigma_i}{2\alpha}\right)\left(h_i e^{\frac{h_i u_i \sigma_i^2}{2\alpha^2}} - u_i e^{\frac{h_i u_i \sigma_i^2}{2\alpha^2}}\right) < 0.$$

Therefore, it is negative at this point, and by Hu and Hong (2012) the problem is convex (which implies monotonicity of derivatives), we have established a lower bound for the problem.

To establish the upper bound in the case with $h_i < u_i$, we set $x_i = \mu_i + \frac{\sigma_i^2(u_i - h_i)}{2\alpha} + \frac{\alpha \ln(u_i/h_i)}{u_i + h_i}$. Via our previous result, clearly

$$\Phi\left(\frac{x_i - \mu_i}{\sigma_i} + \frac{\sigma_i h_i}{\alpha}\right) > \Phi\left(\frac{\mu_i - x_i}{\sigma_i} + \frac{\sigma_i u_i}{\alpha}\right).$$

Now, substituting x_i in the expression sans Φ terms yields

$$\begin{aligned}&h_i e^{\frac{h_i(x_i - \mu_i)}{\alpha} + \frac{\sigma_i^2 h_i^2}{2\alpha^2}} - u_i e^{\frac{u_i(\mu_i - x_i)}{\alpha} + \frac{\sigma_i^2 u_i^2}{2\alpha^2}}\\ &= h_i e^{\left(\frac{h_i\left(2\alpha^2 \ln(u_i/h_i) + \sigma_i^2 u_i (h_i + u_i)\right)}{2\alpha^2(h_i + u_i)}\right)} - u_i e^{\left(\frac{u_i\left(h_i \sigma_i^2 (h_i + u_i) - 2\alpha^2 \ln(u/h)\right)}{2\alpha^2(h_i + u_i)}\right)}.\end{aligned}$$

Hence, examining conditions under which previous expression is greater than zero,

$$\begin{aligned}&h_i e^{\left(\frac{h_i\left(2\alpha^2 \ln(u_i/h_i) + \sigma_i^2 u_i (h_i + u_i)\right)}{2\alpha^2(h_i + u_i)}\right)} \geq u_i e^{\left(\frac{u_i\left(h_i \sigma_i^2 (h_i + u_i) - 2\alpha^2 \ln(u/h)\right)}{2\alpha^2(h_i + u_i)}\right)}\\ &\implies h_i\left(2\alpha^2 \ln\left(u_i/h_i\right) + \sigma_i^2 u_i (h_i + u_i)\right)\end{aligned}$$

$$- u_i \left(h_i\sigma_i^2(h_i+u_i) - 2\alpha^2 \ln(u_i/h_i)\right) + 2\alpha^2(h_i+u_i)\ln(u_i/h_i) \geq 0$$

However, the expression on the left hand side of the inequality is equal to zero. Hence, since we have found an x_i which gives positive derivative, we have found an upper bound on the order quantity. The proofs for the case when $\lambda w_i > 1$ follow naturally by exchanging $\leq$ for $\geq$ and vice versa.

Now to the case when $\lambda w_i = 1$, it is easy to see that when $x_i^\alpha = \mu_i$, we have that

$$\begin{aligned}\Phi\left(\frac{x_i-\mu_i}{\sigma_i} + \frac{\sigma_i h_i}{\alpha}\right) &= \Phi\left(\frac{\sigma_i}{\alpha}\right)\\ &= \Phi\left(\frac{\mu_i - x_i}{\sigma_i} + \frac{\sigma_i u_i}{\alpha}\right),\end{aligned}$$

and furthermore,

$$\frac{h_i}{\alpha}e^{\frac{h_i(\sigma_i^2 h_i)}{2\alpha^2}} = \frac{u_i}{\alpha}e^{\frac{u_i(\sigma_i^2 u_i)}{2\alpha^2}},$$

which implies we have a zero to the derivative.

□

Lemma B.3 (**Converging Upper/Lower Bounds**). *For $i = 1, \dots, n$, if $h_i \leq u_i$, $\hat{x}_i^\alpha \leq \max\{\bar{x}_i + \frac{u_i\sigma_i^2}{\alpha}, s_i\}$. Otherwise if $h_i \geq u_i$, $\hat{x}_i^\alpha \geq \max\{\bar{x}_i - \frac{h_i\sigma_i^2}{\alpha}, s_i\}$.*

Proof. Suppressing the indices for the duration of the proof, in the case that $u \geq h$, using the substitutions $z = (x-\mu)/\sigma$ and $\theta = \sigma/\alpha$, and letting $\bar{z} = \Phi^{-1}(u/(u+h))$. Consider the value $u\theta + \bar{z}$. We want to show that this is an upper bound. Via equation (B.8), it suffices to show

$$h\exp\left(\frac{h^2\theta^2}{2} + h\theta z\right)\Phi(h\theta+z) \geq u\exp\left(\frac{u^2\theta^2}{2} - u\theta z\right)\Phi(u\theta - z).$$

Now, to compare the terms and show that $h\Phi(h\theta+z) \geq u\Phi(u\theta - z)$, we can see that

$$h\Phi(h\theta + u\theta + \bar{z}) \geq \frac{hu}{u+h}$$

$$
\begin{aligned}
&= u\Phi(-\bar{z})) \\
&= u\Phi(\theta u - (u\theta + \bar{z})).
\end{aligned}
$$

Furthermore, to the exponential terms and show that

$$\exp\left(\frac{h^2\theta^2}{2} + h\theta z\right) \geq \exp\left(\frac{u^2\theta^2}{2} - u\theta z\right),$$

examining the terms reveals

$$
\begin{aligned}
\frac{h^2\theta^2}{2} + h\theta z &= \frac{h^2\theta^2}{2} + h\theta(u\theta + \bar{z}) \\
&\geq -\frac{u^2\theta^2}{2} - u\theta\bar{z} \\
&= \frac{u^2\theta^2}{2} - \theta u z,
\end{aligned}
$$

which proves that $u\theta + \bar{z}$ is an upper bound.

The case when $u \geq h$ holds in a similar fashion: via equation (B.8), it suffices to show

$$h \exp\left(\frac{h^2\theta^2}{2} + h\theta z\right) \Phi(h\theta + z) \leq u \exp\left(\frac{u^2\theta^2}{2} - u\theta z\right) \Phi(u\theta - z).$$

Now, to compare the terms and show that $h\Phi(h\theta + z) \leq u\Phi(u\theta - z)$, we can see that

$$
\begin{aligned}
u\Phi(h\theta + u\theta - \bar{z}) &\geq \frac{hu}{u+h} \\
&= h\Phi(\bar{z})) \\
&= u\Phi(h\theta + (\bar{z} - h\theta)).
\end{aligned}
$$

Furthermore, to the exponential terms and show that

$$\exp\left(\frac{h^2\theta^2}{2} + h\theta z\right) \leq \exp\left(\frac{u^2\theta^2}{2} - u\theta z\right),$$

examining the terms reveals

$$\frac{h^2\theta^2}{2} + h\theta z = \frac{h^2\theta^2}{2} + h\theta(-h\theta + \bar{z})$$

$$
\begin{aligned}
&= -\frac{h^2\theta^2}{2} + h\theta\bar{z} \\
&\leq \frac{u^2\theta^2}{2} + h\theta^2 u - u\theta\bar{z} \\
&= \frac{u^2\theta^2}{2} - \theta u z,
\end{aligned}
$$

which proves that $\bar{z} - h\theta$ is an lower bound.

□

Lemma B.4 (**Critical Fractile Upper/Lower Bounds**). *For $i = 1, \ldots, n$, if $h_i \leq u_i$, $\hat{x}_i^\alpha \geq \max\{\bar{x}_i, s_i\}$. Otherwise if $h_i \geq u_i$, $\hat{x}_i^\alpha \leq \max\{\bar{x}_i, s_i\}$.*

Proof. Suppressing the indices for the duration of the proof, it is immediate that $\hat{x}^\alpha \geq s_i$. In the case that $u \geq h$, using the substitutions $z = (x - \mu)/\sigma$ and $\theta = \sigma/\alpha$, and letting $\bar{z} = \Phi^{-1}\left(u/(u+h)\right)$, in the case that $\bar{z}\sigma + \mu > s_i$, we want to show that $\bar{z}$ is a lower bound. Via equation (B.8), it suffices to show

$$
h \exp\left(\frac{h^2\theta^2}{2} + h\theta z\right)\Phi(h\theta + z) \leq u \exp\left(\frac{u^2\theta^2}{2} - u\theta z\right)\Phi(u\theta - z).
$$

Now, this implies that

$$
\frac{h\Phi(h\theta + z)}{u\Phi(u\theta - z)} \exp\left(\frac{h^2\theta^2}{2} + h\theta z - \frac{u^2\theta^2}{2} + u\theta z\right) \leq 1. \tag{B.9}
$$

Letting $q = u/(u+h)$, $\hat{h} = \theta h$, and $\hat{u} = \theta u$, it is equivalent to show that

$$
\frac{\hat{h}\Phi(\hat{h} + z)}{\hat{u}\Phi(\hat{u} - z)} \exp\left(\frac{\hat{h}^2}{2} + \hat{h}z - \frac{\hat{u}^2}{2} + \hat{u}z\right) \leq 1.
$$

Now, $\hat{h} = \frac{\hat{u}(1-q)}{q}$, and

$$
\exp\left(\frac{\hat{h}^2}{2} + \hat{h}z - \frac{\hat{u}^2}{2} + \hat{u}z\right) = \exp\left(\frac{(1-2q)\hat{u}^2}{2q^2} + \frac{\bar{z}}{q\hat{u}}\right)
$$

which is a decreasing function of $\hat{u}$. Furthermore,

$$
\frac{\Phi(\hat{h} + z)}{\Phi(\hat{u} - z)} = \frac{\Phi(\hat{u}/q - \hat{u} + \bar{z})}{\Phi(\hat{u} - \bar{z})}
$$

is a decreasing function of $\hat{u}$. Hence, as θ increases, the expression

$$\frac{h\Phi(h\theta+z)}{u\Phi(u\theta-z)}\exp\left(\frac{h^2\theta^2}{2}+h\theta z-\frac{u^2\theta^2}{2}+u\theta z\right)$$

decreases. Therefore, since we know that at $\theta = 0$, the inequality (B.9) is satisfied due to

$$\frac{h\Phi(\bar{z})}{u\Phi(-\bar{z})}=\frac{h\frac{u}{u+h}}{u\frac{h}{u+h}}=1,$$

we know that (B.9) is satisfied for all $\theta > 0$, which shows that $\bar{z}$ is a lower bound.

The case when $u \leq h$ is identical with ($\leq$) replaced with ($\geq$).

□

Proof of Corollary 3.1. Suppressing the indices for the duration of the proof, since the denominator of G_x is always non-negative, we can focus on its numerator, which we denote $\bar{G}_x$. Using the substitutions $z = \frac{x-\mu}{\sigma}$ and $\theta = \frac{\sigma}{\alpha}$, we can write the numerator of G_x as

$$\bar{G}_x(z,\theta) = h\exp\left(\frac{h^2\theta^2}{2}+h\theta z\right)\Phi(h\theta+z) - u\exp\left(\frac{\theta^2u^2}{2}-\theta uz\right)\Phi(\theta u-z).$$

When we let $z = (u-h)\,\theta$, the expression becomes

$$e^{-\frac{1}{2}h\theta^2(h-2u)}\left(h\Phi(\theta u)-u\Phi(h\theta)e^{\frac{1}{2}\theta^2\left(h^2-u^2\right)}\right).$$

Attending to the case of upper bound first, when we let $\bar{G}_x \geq 0$ we gain the expression

$$\frac{e^{\frac{h^2\theta^2}{2}}\Phi(h\theta)}{h}\leq\frac{e^{\frac{\theta^2u^2}{2}}\Phi(\theta u)}{u}. \tag{B.10}$$

Now, the sides of the inequality are identical except for u and h, hence investigating the derivative

$$\frac{d}{dt}\frac{e^{\frac{\theta^2t^2}{2}}\Phi(\theta t)}{t}=\frac{e^{\frac{\theta^2t^2}{2}}\left(\left(\theta^2t^2-1\right)\Phi(\theta t)+\theta t\Phi'(\theta t)\right)}{t^2},$$

so long as

$$\left(\theta^2 t^2 - 1\right) \Phi(\theta t) + \theta t \Phi'(\theta t) \geq 0,$$

when we substitute u and h for t in the above, the inequaltity (B.10) is true since the function is increasing in t at these values. Now, the derivative is positive for $\theta t \geq 0.8399$ and negative for $\theta t \leq 0.8399$, hence (B.10) is true so long as $u\sigma/\alpha > h\sigma/\alpha > 0.8399$ or $u\sigma/\alpha < h\sigma/\alpha < 0.8399$. To prove the lower bound case, it is easy to see that when ($\leq$) is replaced for ($\geq$) in (B.10), so long as

$$\left(\theta^2 t^2 - 1\right) \Phi(\theta t) + \theta t \Phi'(\theta t) \leq 0,$$

when we substitute u and h for t in the above, the lower bound is true.

□

Proposition B.3 (**Monotonic Ordering in μ_i**). *The quantity $x_i^{*\alpha}$ is non-decreasing in μ_i.*

Proof. Examining the partial derivative G_{x_i} defined in (B.8), increasing μ_i to $\mu_i + \delta$ for $\delta > 0$ is identical to decreasing x_i to $x_i - \delta$. Since (B.7) is convex in x_i, and the derivatives of convex functions are monotone, clearly (B.8) is decreasing in μ_i. Thus by Theorem 3.1, increases in μ_i decrease G_{x_i}, and hence must increase $x_i^{*\alpha}$.

□

Proof of Proposition 3.3. We first prove that the proof that $\frac{\delta G_{x_i}}{\delta \sigma_i} \leq 0$ when $u \geq h$ for all $x_i \geq \mu_i$. The partial derivative of G_{x_i} with respect to σ_i can be expressed

$$\frac{\partial G_{x_i}}{\partial \sigma_i} = (h+u)\exp\left(\frac{h^2\sigma^2 + 2\alpha h(x-\mu) + u\left(-2\alpha\mu + \sigma^2 u + 2\alpha x\right)}{2\alpha^2}\right)$$
$$\frac{\left(\Phi\left(\frac{h\sigma}{\alpha} + \frac{x-\mu}{\sigma}\right)\left(\sigma^3\left(h^2 - u^2\right)\Phi\left(\frac{u\sigma}{\alpha} + \frac{\mu-x}{\sigma}\right) - \alpha\left(-\alpha\mu + \sigma^2 u + \alpha x\right)\Phi'\left(\frac{u\sigma}{\alpha} + \frac{\mu-x}{\sigma}\right)\right) + \alpha\left(\alpha\mu + h\sigma^2 - \alpha x\right)\Phi\left(\frac{u\sigma}{\alpha} + \frac{\mu-x}{\sigma}\right)\Phi'\left(\frac{h\sigma}{\alpha} + \frac{x-\mu}{\sigma}\right)\right)}{\alpha^2\sigma^2\left(\Phi\left(\frac{h\sigma}{\alpha} + \frac{x-\mu}{\sigma}\right)\exp\left(\frac{h^2\sigma^2 + 2\alpha h(x-\mu) + 2\alpha u(x-\mu)}{2\alpha^2}\right) + e^{\frac{\sigma^2 u^2}{2\alpha^2}}\Phi\left(\frac{u\sigma}{\alpha} + \frac{\mu-x}{\sigma}\right)\right)^2}$$

which implies that when we let $\frac{\partial G_{x_i}(x_i,\alpha)}{\partial \sigma_i} \leq 0$, the expression simplifies to

$$h^2\sigma^3 + \alpha\left(\alpha\mu + h\sigma^2 - \alpha x\right)\frac{\Phi'\left(\frac{h\sigma}{\alpha} + \frac{x-\mu}{\sigma}\right)}{\Phi\left(\frac{h\sigma}{\alpha} + \frac{x-\mu}{\sigma}\right)} \leq \sigma^3 u^2 + \alpha\left(-\alpha\mu + \sigma^2 u + \alpha x\right)\frac{\Phi'\left(\frac{\sigma u}{\alpha} + \frac{\mu-x}{\sigma}\right)}{\Phi\left(\frac{\sigma u}{\alpha} + \frac{\mu-x}{\sigma}\right)}.$$

Letting $z = \frac{x-\mu}{\sigma}$, we can rewrite the expression to

$$h^2\sigma^3 + \left(-z\alpha\sigma + h\sigma^2\right)\frac{\Phi'\left(\frac{h\sigma}{\alpha} + z\right)}{\Phi\left(\frac{h\sigma}{\alpha} + z\right)} \leq u^2\sigma^3 + \left(z\alpha\sigma + \sigma^2 u\right)\frac{\Phi'\left(\frac{\sigma u}{\alpha} - z\right)}{\Phi\left(\frac{\sigma u}{\alpha} - z\right)}. \tag{B.11}$$

Now, when $z = 0$ we have

$$h^2\sigma^3 + \alpha h\sigma^2\frac{\Phi'\left(\frac{h\sigma}{\alpha}\right)}{\Phi\left(\frac{h\sigma}{\alpha}\right)} \leq \sigma^3 u^2 + \alpha\sigma^2 u\frac{\Phi'\left(\frac{\sigma u}{\alpha}\right)}{\Phi\left(\frac{\sigma u}{\alpha}\right)}.$$

When $u = h$, we have equality in the above. If the gap increases in u, then we know that at $z = 0$, we have shown that the inequality (B.11) is true. To show this, we show that

$$\sigma^3 u^2 + \alpha\sigma^2 u\frac{\Phi'\left(\frac{\sigma u}{\alpha}\right)}{\Phi\left(\frac{\sigma u}{\alpha}\right)} = \sigma^2 u\left(\alpha\frac{\Phi'\left(\frac{\sigma u}{\alpha}\right)}{\Phi\left(\frac{\sigma u}{\alpha}\right)} + \sigma u\right) \tag{B.12}$$

is increasing in u. Now, if

$$\alpha\frac{\Phi'\left(\frac{\sigma u}{\alpha}\right)}{\Phi\left(\frac{\sigma u}{\alpha}\right)} + \sigma u$$

is increasing in u, then, certainly we have the increasing property for (B.12). Now, via the substitution $t = \frac{\sigma u}{\alpha}$ we can rewrite this as $\alpha\frac{\Phi'(t)}{\Phi(t)} + \alpha t$, so if

$$\frac{\Phi'(t)}{\Phi(t)} + t$$

is increasing in t, we have our monotonicity result. Now,

$$\begin{aligned}\frac{d}{dt}\frac{\Phi'(t)}{\Phi(t)} + t = 1 + \frac{\Phi''(t)}{\Phi(t)} - \frac{\Phi'(t)^2}{\Phi(t)^2} &= 1 - \frac{e^{-t^2/2}t}{\int_{-\infty}^{t} e^{-s^2}ds} - \frac{e^{-t^2/2}}{\left(\int_{-\infty}^{t} e^{-s^2}ds\right)^2} \\ &\leq -\sqrt{\frac{2}{\pi}}te^{-\frac{t^2}{2}} - \frac{2}{\pi}e^{-\frac{t^2}{2}} + 1,\end{aligned}$$

since

$$\int_{-\infty}^{t} e^{-s^2}ds = \sqrt{\frac{\pi}{2}}\left(\text{erf}\left(\frac{t}{\sqrt{2}}\right) + 1\right) \geq \sqrt{\frac{\pi}{2}},$$

for all $t \geq 0$. Now,

$$\max_{t\in\mathbb{R}} \left\{ -\sqrt{\frac{2}{\pi}} t e^{-\frac{t^2}{2}} - \frac{2}{\pi} e^{-\frac{t^2}{2}} + 1 \right\} \approx 0.0642$$

when $t = \frac{\sqrt{1+2\pi}-1}{\sqrt{2\pi}}$. Hence, (B.12) is increasing in u. Now, we know that

$$\left(z\alpha\sigma + \sigma^2 u\right) \frac{\Phi'\left(\frac{\sigma u}{\alpha} - z\right)}{\Phi\left(\frac{\sigma u}{\alpha} - z\right)}$$

is increasing in z for all $0 \leq z \leq \frac{\sigma h}{\alpha}$, since $\Phi'\left(\frac{\sigma u}{\alpha} - z\right)$ and $\Phi\left(\frac{\sigma u}{\alpha} - z\right)$ are respectively increasing and decreasing from $0 \leq z \leq \frac{\sigma u}{\alpha}$, and certainly $(z\alpha\sigma + \sigma^2 u)$ is increasing for all $z \geq 0$.

Furthermore, we have that

$$\left(-z\alpha\sigma + h\sigma^2\right) \frac{\Phi'\left(\frac{h\sigma}{\alpha} + z\right)}{\Phi\left(\frac{h\sigma}{\alpha} + z\right)}$$

is decreasing in z for all $0 \leq z \leq \frac{\sigma h}{\alpha}$ since $\Phi'\left(\frac{h\sigma}{\alpha} + z\right)$ and $\Phi\left(\frac{h\sigma}{\alpha} + z\right)$ are respectively decreasing and increasing from $0 \leq z \leq \frac{\sigma u}{\alpha}$, and certainly $(-z\alpha\sigma + h\sigma^2)$ is decreasing for all $z \geq 0$. Thus, for all $0 \leq z \leq \frac{\sigma h}{\alpha}$, we have that

$$\left(-z\alpha\sigma + h\sigma^2\right) \frac{\Phi'\left(\frac{h\sigma}{\alpha} + z\right)}{\Phi\left(\frac{h\sigma}{\alpha} + z\right)} \leq \left(z\alpha\sigma + \sigma^2 u\right) \frac{\Phi'\left(\frac{\sigma u}{\alpha} - z\right)}{\Phi\left(\frac{\sigma u}{\alpha} - z\right)}.$$

Now, when $z \geq \frac{\sigma h}{\alpha}$,

$$\left(-z\alpha\sigma + h\sigma^2\right) \frac{\Phi'\left(\frac{h\sigma}{\alpha} + z\right)}{\Phi\left(\frac{h\sigma}{\alpha} + z\right)} \leq 0$$

but

$$\left(z\alpha\sigma + \sigma^2 u\right) \frac{\Phi'\left(\frac{\sigma u}{\alpha} - z\right)}{\Phi\left(\frac{\sigma u}{\alpha} - z\right)} \geq 0$$

hence, we have that

$$\left(-z\alpha\sigma + h\sigma^2\right) \frac{\Phi'\left(\frac{h\sigma}{\alpha} + z\right)}{\Phi\left(\frac{h\sigma}{\alpha} + z\right)} \leq \left(z\alpha\sigma + \sigma^2 u\right) \frac{\Phi'\left(\frac{\sigma u}{\alpha} - z\right)}{\Phi\left(\frac{\sigma u}{\alpha} - z\right)}.$$

for all $z \geq 0$.

When $h \geq u$, the proof that $\frac{\delta G_{x_i}}{\delta \sigma_i} \geq 0$ for all $x_i \leq \mu_i$ follows in a similar manner. Letting $\frac{\delta G_{x_i}}{\delta \sigma_i} \geq 0$, the expression that must be satisified can be written

$$h^2\sigma^3 + \left(-z\alpha\sigma + h\sigma^2\right)\frac{\Phi'\left(\frac{h\sigma}{\alpha}+z\right)}{\Phi\left(\frac{h\sigma}{\alpha}+z\right)} \geq u^2\sigma^3 + \left(z\alpha\sigma + \sigma^2 u\right)\frac{\Phi'\left(\frac{\sigma u}{\alpha}-z\right)}{\Phi\left(\frac{\sigma u}{\alpha}-z\right)}. \tag{B.13}$$

When $z = 0$, we have

$$h^2\sigma^3 + \alpha h\sigma^2\frac{\Phi'\left(\frac{h\sigma}{\alpha}\right)}{\Phi\left(\frac{h\sigma}{\alpha}\right)} \geq \sigma^3 u^2 + \alpha\sigma^2 u\frac{\Phi'\left(\frac{\sigma u}{\alpha}\right)}{\Phi\left(\frac{\sigma u}{\alpha}\right)}.$$

When $u = h$, we have equality in the above. If this gap increases in h, we know that at $z = 0$, we have shown that the inequality (B.13) is true. To show this, we show that

$$\sigma^2 h\left(\alpha\frac{\Phi'\left(\frac{\sigma h}{\alpha}\right)}{\Phi\left(\frac{\sigma h}{\alpha}\right)} + \sigma h\right)$$

is increasing in h. However, this is identical to our claim that (B.12) is increasing in u, hence when $z = 0$, (B.13) holds.

Now, we know that

$$\left(-z\alpha\sigma + h\sigma^2\right)\frac{\Phi'\left(\frac{h\sigma}{\alpha}+z\right)}{\Phi\left(\frac{h\sigma}{\alpha}+z\right)}$$

is decreasing in z for all $-\frac{\sigma u}{\alpha} \leq z \leq 0$, since $\Phi'\left(\frac{\sigma h}{\alpha}+z\right)$ and $\Phi\left(\frac{\sigma h}{\alpha}+z\right)$ are respectively decreasing and increasing from $-\frac{\sigma u}{\alpha} \leq z \leq 0$, and certainly $(-z\alpha\sigma + h\sigma^2)$ is decreasing for all $z \leq 0$.

Furthermore, we have that

$$\left(z\alpha\sigma + \sigma^2 u\right)\frac{\Phi'\left(\frac{\sigma u}{\alpha}-z\right)}{\Phi\left(\frac{\sigma u}{\alpha}-z\right)}$$

is increasing in z for all $-\frac{\sigma u}{\alpha} \leq z \leq 0$ since $\Phi'\left(\frac{\sigma u}{\alpha}-z\right)$ and $\Phi\left(\frac{\sigma u}{\alpha}-z\right)$ are respectively increasing and decreasing from $-\frac{\sigma u}{\alpha} \leq z \leq 0$, and certainly $(z\alpha\sigma + \sigma^2 u)$ is increasing for all $z \leq 0$. Thus, for all $-\frac{\sigma u}{\alpha} \leq z \leq 0$, we have that

$$\left(-z\alpha\sigma + h\sigma^2\right) \frac{\Phi'\left(\frac{h\sigma}{\alpha} + z\right)}{\Phi\left(\frac{h\sigma}{\alpha} + z\right)} \geq \left(z\alpha\sigma + \sigma^2 u\right) \frac{\Phi'\left(\frac{\sigma u}{\alpha} - z\right)}{\Phi\left(\frac{\sigma u}{\alpha} - z\right)}.$$

Now, when $z \leq -\frac{\sigma u}{\alpha}$,

$$\left(z\alpha\sigma + \sigma^2 u\right) \frac{\Phi'\left(\frac{\sigma u}{\alpha} - z\right)}{\Phi\left(\frac{\sigma u}{\alpha} - z\right)} \leq 0$$

but

$$\left(-z\alpha\sigma + h\sigma^2\right) \frac{\Phi'\left(\frac{h\sigma}{\alpha} + z\right)}{\Phi\left(\frac{h\sigma}{\alpha} + z\right)} \geq 0$$

hence, we have that

$$\left(-z\alpha\sigma + h\sigma^2\right) \frac{\Phi'\left(\frac{h\sigma}{\alpha} + z\right)}{\Phi\left(\frac{h\sigma}{\alpha} + z\right)} \geq \left(z\alpha\sigma + \sigma^2 u\right) \frac{\Phi'\left(\frac{\sigma u}{\alpha} - z\right)}{\Phi\left(\frac{\sigma u}{\alpha} - z\right)}.$$

for all $z \leq 0$.

□

Proof of Proposition 3.4. Suppressing the indices for the duration of the proof, when $u \geq h$, we want to show that $\hat{x}^\alpha$ is decreasing in α. Clearly, if $\hat{x}^\alpha = s$, if the zeros associated with (B.8) are decreasing in α, the result holds, hence we turn our attention the zeros of (B.8) and refer to $\bar{x}^\alpha$ to these zeros. Letting $z = \frac{\bar{x}^\alpha - \mu}{\sigma}$, via equation (B.8), it suffices to show

$$h \exp\left(\frac{h^2\theta^2}{2} + h\theta z\right) \Phi(h\theta + z) \leq u \exp\left(\frac{u^2\theta^2}{2} - u\theta z\right) \Phi(u\theta - z)$$

for all $\theta \geq \sigma/\alpha$. Now, this implies that

$$\frac{h\Phi(h\theta + z)}{u\Phi(u\theta - z)} \exp\left(\frac{h^2\theta^2}{2} + h\theta z - \frac{u^2\theta^2}{2} + u\theta z\right) \leq 1, \tag{B.14}$$

for all $\theta \geq \sigma/\alpha$. Letting $q = u/(u+h), \hat{h} = \theta h$, and $\hat{u} = \theta u$, it is equivalent to show that

$$\frac{\hat{h}\Phi(\hat{h} + z)}{\hat{u}\Phi(\hat{u} - z)} \exp\left(\frac{\hat{h}^2}{2} + \hat{h}z - \frac{\hat{u}^2}{2} + \hat{u}z\right) \leq 1.$$

Now, $\hat{h} = \frac{\hat{u}(1-q)}{q}$, and

$$\exp\left(\frac{\hat{h}^2}{2} + \hat{h}z - \frac{\hat{u}^2}{2} + \hat{u}z\right) = \exp\left(\frac{(1-2q)\hat{u}^2}{2q^2} + \frac{\bar{z}}{q\hat{u}}\right)$$

which is a decreasing function of $\hat{u}$. Furthermore,

$$\frac{\Phi(\hat{h}+z)}{\Phi(\hat{u}-z)} = \frac{\Phi(\hat{u}/q - \hat{u} + \bar{z})}{\Phi(\hat{u}-\bar{z})}$$

is a decreasing function of $\hat{u}$. Hence, as θ increases, the expression

$$\frac{h\Phi(h\theta+z)}{u\Phi(u\theta-z)} \exp\left(\frac{h^2\theta^2}{2} + h\theta z - \frac{u^2\theta^2}{2} + u\theta z\right)$$

decreases. Therefore, since we know that at $\theta = \sigma/\alpha$, the inequality (B.14) is satisfied by definition, we know that (B.14) is satisfied for all $\theta > \sigma/\alpha$.

The case when $u \leq h$ is identical with ($\leq$) replaced with ($\geq$).

□

Proof of Corollary 3.2. By Theorem 3.1, it suffices to show that $\pi_i^{\alpha}(x_i^{*\alpha'})$ remains negative for all $\alpha < \alpha'$ since under optimality, if the carrier constraint is tight, $\pi_i^{\alpha'}(x_i^{*\alpha'}) < 0$ for all nonzero orders within $i = 1, \ldots, n$, and if the refrigeration constraint is tight, $\pi_i^{\alpha'}(x_i^{*\alpha'}) < 0$ for all nonzero orders in $i \in \mathcal{O}_r$. That is, all marginal benefit functions are either negative or zero for non-zero orders. Hence, if we can show that all of the marginal benefit functions (of non-zero orders) remain negative for all $\alpha < \alpha'$, the capacity constraint must remain tight.

Now, since (B.7) is convex and by Proposition 3.2, the α that makes $\pi_i^{\alpha}(x_i^{*\alpha'})$ zero must lie above α', hence each π is negative for every α less than α'. Thus, since π_i^{α} is monotonically increasing in x_i, and since either all of π_i^{α} are zero or all are negative under optimality via Theorem 3.1, this concludes the proof.

□

Proof of Proposition 3.5. Part (i) is immediate from the bounds established in Proposition 3.2 since the upper bound when $u_i < h_i$ tends to s_i when α approaches zero.

To prove (ii) and (iii), for vaccines with $u_i > h_i$, we first prove that $\lim_{\alpha\to 0} \pi^\alpha(x_i) = -u_i/w_i$. To accomplish this, we show that there exists α such that $G_{x_i}(x_i, \alpha) \leq -u_i + \delta$ for any $\delta > 0$. This is equivalent to

$$(-\delta + h + u)e^{\frac{h^2\sigma^2}{2\alpha^2} + \frac{h(x-\mu)}{\alpha}} \Phi\left(\frac{h\sigma}{\alpha} + \frac{x-\mu}{\sigma}\right) \leq \delta e^{\frac{\sigma^2 u^2}{2\alpha^2} + \frac{u(\mu - x)}{\alpha}} \Phi\left(\frac{\sigma u}{\alpha} + \frac{\mu - x}{\sigma}\right)$$
$$\implies 2\alpha^2 \ln\left(\frac{(-\delta + h + u)\Phi\left(\frac{h\sigma}{\alpha} + \frac{x-\mu}{\sigma}\right)}{\delta\Phi\left(\frac{\sigma u}{\alpha} + \frac{\mu - x}{\sigma}\right)}\right) + (h+u)\left(-2\alpha\mu + \sigma^2(h-u) + 2\alpha x\right) \leq 0$$

and since all of these terms tend toward zero with the exception of $\sigma^2(h-u)$, hence $\lim_{\alpha\to 0}\pi^\alpha(x_i) = -u_i/w_i$ for any order quantity x_i.

To prove (ii), since $\lim_{\alpha\to 0}\pi^\alpha(x_i) = -u_i/w_i$, for α small enough, any feasible order $\mathbf{x} \in \mathcal{X}(\mathbf{s})$ will result in $\pi_i^\alpha(x_i) > \gamma_c$ for vaccines that do not have $-u_i/w_i = \gamma_c$. As seen in Theorem 3.1, this implies that such vaccines become zero-order since all non-zero orders have $\pi_i^\alpha(x_i^{*\alpha}) = \min_{j\in\mathcal{O}_r} \pi_j^\alpha(x_j^{*\alpha})$ for all $i \in \mathcal{O}_r$ and likewise $\pi_i^\alpha(x_i^{*\alpha}) = \min_{j\in\mathcal{O}_c} \pi_j^\alpha(x_j^{*\alpha})$ for all $i \in \mathcal{O}_c$.

Now, since $\gamma_c \leq \gamma_r$, this implies that there exists a vaccine $i \in \mathcal{O}_c$ such that $u_i/w_i = \gamma_c$. Furthermore, since $\pi_i^\alpha(x_i) < 0$, this implies that the vaccine carrier is filled, hence $\lim_{\alpha\to 0} b_c - \sum_{i=1}^n (x_i^{*\alpha} - s_i)w_i = 0$.

The proof for (iii) follows in a similar manner. Since $\lim_{\alpha\to 0}\pi^\alpha(x_i) = -u_i/w_i$, for α small enough, any feasible order $\mathbf{x} \in \mathcal{X}(\mathbf{s})$ will result in $\pi_i^\alpha(x_i) > \gamma_r$ for vaccines that do not have $-u_i/w_i = \gamma_r$. As seen in Theorem 3.1, this implies that all vaccines in $\mathcal{O}_r$ that do not have $-u_i/w_i = \gamma_r$ become zero-order since all non-zero orders have $\pi_i^\alpha(x_i^{*\alpha}) = \min_{j\in\mathcal{O}_r} \pi_j^\alpha(x_j^{*\alpha})$ for all non-zero orders $i \in \mathcal{O}_r$.

Furthermore, since $\pi_i^\alpha(x_i^{*\alpha}) < 0$ for some $i \in \mathcal{O}_r$, this implies that either the vaccine carrier or the refrigeration is filled (whichever has less capacity). In the case that $\gamma_c < 0$ and there is remaining carrier capacity, the problem is the same as case

(ii), and the carrier is filled with vaccines that have $-u_i/w_i = \gamma_c$ for $i \in \mathcal{O}_c$.

□

Proposition B.4 (**Monotone Cost in σ_i and μ_i.**). *The robust objective* (3.8) *is increasing in σ_i and convex in μ_i for all $i \in \mathcal{N}$. Furthermore, if $\mathbf{s} = \mathbf{0}$,* (3.8) *is increasing in μ_i for all $i \in \mathcal{N}$.*

Proof. To show that G is non-decreasing in μ_i, we first note that since we assume non-negative demand $x_i^{*,\alpha} \geq 0$. Using the substitutions $z_i = (x_i - \mu_i)/\sigma_i$ and $\theta_i = \sigma_i/\alpha$, it is easy to show that the components of $\mathrm{E}_{\hat{f}}\left[e^{H_i(x_i,V_i)/\alpha}\right]$ can be expressed

$$e^{h_i z_i \theta_i + \frac{\theta_i^2 h_i^2}{2}} \Phi\left(z_i + \theta_i h_i\right) + e^{-z_i u_i \theta_i + \frac{\theta_i^2 u_i^2}{2}} \Phi\left(-z_i + \theta_i u_i\right).$$

Therefore, the cost of a system when μ_i is increased by any $\delta > 0$ does not increase so long as x_i can also be increased by δ. The only difference between these two systems is in the restrictiveness of the constraints. Hence, since $\mu_i + \delta$ experiences strictly tighter constraints than the system under μ_i, the cost is non-increasing in μ_i. Now, to show that G is convex in μ_i, we can easily see from above that shifting μ_i to $\mu_i + \delta$ results in an identical cost to shifting x_i to $x_i - \delta$ since $z_i = \frac{x_i - (\mu_i + \delta)}{\sigma_i} = \frac{(x_i - \delta) - \mu_i}{\sigma_i}$. Therefore, since G is convex in x_i, G must also be convex in μ_i.

To show that G is non-decreasing in σ_i, the partial derivative of G with respect to σ_i can be expressed:

$$\begin{aligned}\frac{\partial G}{\partial \sigma_i} = \alpha\Bigg(& \frac{h^2 \sigma e^{\frac{h\left(-2\alpha\mu + h\sigma^2 + 2\alpha x\right)}{2\alpha^2}} \Phi\left(\frac{h\sigma}{\alpha} + \frac{x-\mu}{\sigma}\right)}{\alpha^2} + e^{\frac{h\left(-2\alpha\mu + h\sigma^2 + 2\alpha x\right)}{2\alpha^2}} \left(\frac{h}{\alpha} + \frac{\mu - x}{\sigma^2}\right) \Phi'\left(\frac{h\sigma}{\alpha} + \frac{x - \mu}{\sigma}\right) \\ & + \frac{\sigma u^2 e^{\frac{u\left(2\alpha\mu + \sigma^2 u - 2\alpha x\right)}{2\alpha^2}} \Phi\left(\frac{\sigma u}{\alpha} + \frac{\mu - x}{\sigma}\right)}{\alpha^2} + \frac{e^{\frac{u\left(2\alpha\mu + \sigma^2 u - 2\alpha x\right)}{2\alpha^2}} \left(-\alpha\mu + \sigma^2 u + \alpha x\right) \Phi'\left(\frac{\sigma u}{\alpha} + \frac{\mu - x}{\sigma}\right)}{\alpha\sigma^2}\Bigg) \\ & \Bigg/ \left(e^{\frac{h\left(-2\alpha\mu + h\sigma^2 + 2\alpha x\right)}{2\alpha^2}} \Phi\left(\frac{h\sigma}{\alpha} + \frac{x - \mu}{\sigma}\right) + e^{\frac{u\left(2\alpha\mu + \sigma^2 u - 2\alpha x\right)}{2\alpha^2}} \Phi\left(\frac{\sigma u}{\alpha} + \frac{\mu - x}{\sigma}\right)\right)\end{aligned}$$

Setting $\frac{\partial G}{\partial \sigma_i} > 0$, we can simplify the above expression to

$$\frac{h^2\sigma e^{\frac{h\left(-2\alpha\mu+h\sigma^2+2\alpha x\right)}{2\alpha^2}}\Phi\left(\frac{h\sigma}{\alpha}+\frac{x-\mu}{\sigma}\right)}{\alpha^2}+e^{\frac{h\left(-2\alpha\mu+h\sigma^2+2\alpha x\right)}{2\alpha^2}}\left(\frac{h}{\alpha}+\frac{\mu-x}{\sigma^2}\right)\Phi'\left(\frac{h\sigma}{\alpha}+\frac{x-\mu}{\sigma}\right)$$
$$+\frac{\sigma u^2 e^{\frac{u\left(2\alpha\mu+\sigma^2 u-2\alpha x\right)}{2\alpha^2}}\Phi\left(\frac{\sigma u}{\alpha}+\frac{\mu-x}{\sigma}\right)}{\alpha^2}+e^{\frac{u\left(2\alpha\mu+\sigma^2 u-2\alpha x\right)}{2\alpha^2}}\left(\frac{u}{\alpha}+\frac{x-\mu}{\sigma^2}\right)\Phi'\left(\frac{\sigma u}{\alpha}+\frac{\mu-x}{\sigma}\right)>0$$

which, by substituting $z=\frac{x-\mu}{\sigma}$ and $\theta=\frac{\sigma}{\alpha}$ can be expressed

$$e^{\frac{h\left(-2\alpha\mu+h\sigma^2+2\alpha x\right)}{2\alpha^2}}\left(h^2\sigma^2\Phi\left(\frac{h\sigma}{\alpha}+z\right)+\alpha(h\sigma-\alpha z)\Phi'\left(\frac{h\sigma}{\alpha}+z\right)\right)$$
$$+e^{\frac{u\left(2\alpha\mu+\sigma^2 u-2\alpha x\right)}{2\alpha^2}}\left(\sigma^2u^2\Phi\left(\frac{\sigma u}{\alpha}-z\right)+\alpha(\sigma u+\alpha z)\Phi'\left(\frac{\sigma u}{\alpha}-z\right)\right)>0$$

which is equivalent to

$$e^{\frac{1}{2}h\theta(h\theta+2z)}\left(h^2\theta^2\Phi(h\theta+z)+(h\theta-z)\Phi'(h\theta+z)\right)$$
$$+e^{\frac{1}{2}\theta u(\theta u-2z)}\left(\theta^2u^2\Phi(\theta u-z)+(\theta u+z)\Phi'(\theta u-z)\right)>0$$

which is equivalent to

$$\sqrt{\pi}\theta h^2 e^{\frac{1}{2}(h\theta+z)^2}\operatorname{erfc}\left(-\frac{h\theta+z}{\sqrt{2}}\right)+\sqrt{\pi}\theta u^2 e^{\frac{1}{2}(z-\theta u)^2}\operatorname{erfc}\left(\frac{z-\theta u}{\sqrt{2}}\right)+\sqrt{2}(h+u)>0$$

which is obviously true due to all positive components.

□

Proof of Proposition 3.6 and Corollary 3.3. When $\mathcal{N}=\mathcal{O}_c$, the proposition follows immediately since $\mathbf{s}^t=\mathbf{0}$ at each period, regardless a decision-maker's action. When vaccine return is enabled with $b_c=\infty$, even if $\mathbf{s}^t\neq\mathbf{0}$, the decision-maker can choose to return all vaccines and order up to any given value. Hence it is equivalent to assume that $\mathbf{s}^t=\mathbf{0}$ at each period.

Similarly, when vaccines are delivered to an IHC, this can be viewed as two separate single-period problems, one with refrigerated vaccines, and the other with OCC vaccines. If the OCC vaccines are treated with fictitious "refrigeration" constraint b_c and have an infinite fictitious transportation constraint, while the refrigerated vaccines treat b_r as their usual refrigerated constraint with infinite transportation constraint, the problem instance is recovered. Since order quantities of $\mathcal{O}_c$ and $\mathcal{O}_r$ do not limit each other, the problem can be solved up to the term α using two separate problem instances. □

To show basestock optimality in systems with backordering, we generate a dynamic formulation whose statespace includes the previous p forecasts and inventory state. We let $\hat{W} \in \mathbb{R}^{p \times n}$ be composed of the previous p forecasts, let $\aleph : \mathbb{R}^{p\times n} \to \mathbb{R}^{p \times n}$ which updates $\hat{W}$ to include the current forecast and discard the pth forecast. Then, letting $f_{\hat{W}}$ denote the forecast with respect to p previous forecasts $\hat{W}$, we define the following dynamic program with backordering

$$\mathrm{J}_t(\mathbf{s}, \hat{W}) = \min_{\mathbf{x} \in \mathcal{X}(\mathbf{s})} \max_{f \in \mathbb{D}(f_{\hat{W}}, \eta)} \mathrm{E}_f \left[\sum_{i=1}^{n} H_i(x_i, V_i) + \beta \int_{\mathbf{v} \geq \mathbf{0}} \mathrm{J}_{t-1}(\mathbf{x} - \mathbf{V}, \aleph(\hat{W})) f(\mathbf{v}) d\mathbf{v} \right] \tag{B.15}$$

with terminating condition $\mathrm{J}_0(\mathbf{s}, \hat{W}) = 0$. Also, we introduce the equation

$$\hat{\mathrm{J}}_t(\mathbf{x}, \hat{W}) = \max_{f \in \mathbb{D}(f_{\hat{W}}, \eta)} \mathrm{E}_f \left[\sum_{i=1}^{n} H_i(x_i, V_i) + \beta \int_{\mathbf{v} \geq \mathbf{0}} \mathrm{J}_{t-1}(\mathbf{x} - \mathbf{V}, \aleph(\hat{W})) f(\mathbf{v}) d\mathbf{v} \right] \tag{B.16}$$

Proof of Proposition 3.7. The proof follows in a nearly identical fashion to the Decroix and Arreola-Risa (1998). There exists an infinite-horizon policy that has finite expected discounted cost (see, e.g., Lemma B.5). This implies that $\mathrm{J}_t(\mathbf{s}, \hat{W})$ is bounded above by a function $\hat{\mathrm{M}}(\mathbf{s}, \hat{W})$.

$\mathrm{J}_t(\mathbf{s}, \hat{W})$ is clearly increasing in t, hence, there exists a function $\mathrm{M}(\mathbf{s}, \hat{W})$ that satisfies

$$\mathrm{M}(\mathbf{s}, \hat{W}) = \lim_{t \to \infty} \mathrm{J}_t(\mathbf{s}, \hat{W}).$$

$\hat{\mathrm{J}}_t(\hat{\mathbf{s}}, \hat{W})$ and $\mathrm{J}_t(\mathbf{s}, \hat{W})$ are convex for each t (see, e.g., Lemma B.6). Also, $\mathrm{J}_t(\mathbf{s}, \hat{W}) \to \infty$ as $||\mathbf{s}|| \to \infty$ and $\hat{\mathrm{J}}_t(\hat{\mathbf{s}}, \hat{W}) \to \infty$ as $||\hat{\mathbf{s}}|| \to \infty$. By taking limits, $\mathrm{K}(\mathbf{s}, \hat{W})$ is convex and $\mathrm{K}(\mathbf{s}, \hat{W}) \to \infty$ as $||\mathbf{s}|| \to \infty$.

We now show that $\mathrm{K}(\mathbf{s}, \hat{W})$ satisfies the inventory functional equation (B.15), thereby establishing that $\mathrm{J}(\mathbf{s}, \hat{W}) = \mathrm{K}(\mathbf{s}, \hat{W})$, i.e., $\mathrm{K}(\mathbf{s}, \hat{W})$ is the infinite-horizon minimum-cost function. We establish equality in (B.15) by showing that the inequality holds in both directions. First, note that

$$\begin{aligned}
\mathrm{K}(\mathbf{s}, \hat{W}) &= \lim_{t\to\infty} \mathrm{J}_t(\mathbf{s}, \hat{W}) \\
&= \lim_{t\to\infty} \left\{ \min_{\mathbf{x}\in\mathcal{X}(\mathbf{s})} \max_{f\in\mathbb{D}(f_{\hat{W}},\eta)} \mathrm{E}_f \left[\sum_{i=1}^{n} H_i(x_i, V_i) + \beta \int_{\mathbf{v}\geq\mathbf{0}} \mathrm{J}_{t-1}(\mathbf{x}-\mathbf{V}, \aleph(\hat{W})) f(\mathbf{v}) d\mathbf{v} \right] \right\} \\
&\leq \lim_{t\to\infty} \left\{ \min_{\mathbf{x}\in\mathcal{X}(\mathbf{s})} \max_{f\in\mathbb{D}(f_{\hat{W}},\eta)} \mathrm{E}_f \left[\sum_{i=1}^{n} H_i(x_i, V_i) + \beta \int_{\mathbf{v}\geq\mathbf{0}} \mathrm{K}(\mathbf{x}-\mathbf{V}, \aleph(\hat{W})) f(\mathbf{v}) d\mathbf{v} \right] \right\} \\
&= \min_{\mathbf{x}\in\mathcal{X}(\mathbf{s})} \max_{f\in\mathbb{D}(f_{\hat{W}},\eta)} \mathrm{E}_f \left[\sum_{i=1}^{n} H_i(x_i, V_i) + \beta \int_{\mathbf{v}\geq\mathbf{0}} \mathrm{K}(\mathbf{x}-\mathbf{V}, \aleph(\hat{W})) f(\mathbf{v}) d\mathbf{v} \right]
\end{aligned}$$

Now since $\mathrm{J}_t(\mathbf{s}, \hat{W})$ converges monotonically to $\mathrm{K}(\mathbf{s}, \hat{W})$, the Monotone Convergence Theorem implies that $\hat{\mathrm{J}}_t(\hat{\mathbf{s}}, \hat{W})$ converges monotonically to $\hat{\mathrm{J}}(\hat{\mathbf{s}}, \hat{W})$. Since $\hat{\mathrm{J}}_t(\hat{\mathbf{s}}, \hat{W})$ and $\hat{\mathrm{J}}(\hat{\mathbf{s}}, \hat{W})$ are continuous everywhere, $\hat{\mathrm{J}}_t(\hat{\mathbf{s}}, \hat{W})$ converges uniformly to $\hat{\mathrm{J}}(\hat{\mathbf{s}}, \hat{W})$ on the compact set

$$\mathcal{Y}(\mathbf{u}) = \{\mathbf{y} \in \mathbb{R}^n_+ : y_i \geq u_i, \mathbf{w}'(\mathbf{y}-\mathbf{u}) \leq b_c, \mathbf{r}'\mathbf{y} \leq b_r\}.$$

From the earlier discussion,

$$\mathrm{K}(\mathbf{s}, \hat{W}) \geq \mathrm{J}_t(\mathbf{s}, \hat{W}) = \min_{\hat{\mathbf{s}}\geq\mathbf{s}, \hat{\mathbf{s}}\in\mathcal{X}(\mathbf{s})} \hat{\mathrm{J}}_t(\hat{\mathbf{s}}, \hat{W}).$$

For any $\epsilon > 0$, there exists T such that for all $t \geq T$, $0 \leq \hat{\mathrm{J}}(\hat{\mathbf{s}}, \hat{W}) - \hat{\mathrm{J}}_t(\hat{\mathbf{s}}, \hat{W}) < \epsilon$ for all $\hat{\mathbf{s}} \in \mathcal{Y}(\mathbf{s})$. Let $\mathbf{z}^t$ be a minimizer of $\hat{\mathrm{J}}_t(\hat{\mathbf{s}}, \hat{W})$ and $\mathbf{z}$ be a minimizer of $\hat{\mathrm{J}}(\hat{\mathbf{s}}, \hat{W})$ on $\mathcal{Y}(\mathbf{s})$. Then

$$0 \leq \hat{\mathrm{J}}(\mathbf{z}^t, \hat{W}) - \hat{\mathrm{J}}_t(\mathbf{z}^t, \hat{W}) < \epsilon,$$

and clearly

$$\hat{\mathrm{J}}(\mathbf{z}, \hat{W}) - \hat{\mathrm{J}}(\mathbf{z}^t, \hat{W}) \leq 0 \text{ and } \hat{\mathrm{J}}(\mathbf{z}, \hat{W}) - \hat{\mathrm{J}}_t(\mathbf{z}^t, \hat{W}) \geq 0,$$

so we have that

$$\begin{aligned}
0 &\leq \hat{\mathrm{J}}(\mathbf{z}, \hat{W}) - \hat{\mathrm{J}}_t(\mathbf{z}^t, \hat{W}) \\
&= \left[\hat{\mathrm{J}}(\mathbf{z}, \hat{W}) - \hat{\mathrm{J}}(\mathbf{z}^t, \hat{W})\right] + \left[\hat{\mathrm{J}}(\mathbf{z}^t, \hat{W}) - \hat{\mathrm{J}}_t(\mathbf{z}^t, \hat{W})\right] \\
&< 0 + \epsilon = \epsilon,
\end{aligned}$$

i.e.,

$$\min_{\mathbf{z} \in \mathcal{Y}(\mathbf{s})} \hat{\mathrm{J}}(\mathbf{z}, \hat{W}) < \min_{\mathbf{z} \in \mathcal{Y}(\mathbf{s})} \hat{\mathrm{J}}_t(\mathbf{z}, \hat{W}) + \epsilon,$$

for all $t \geq T$. Taking the limit of the right-hand side as $t \to \infty$ and then as $\epsilon \to 0$ yields

$$\min_{\mathbf{z} \in \mathcal{Y}(\mathbf{s})} \hat{\mathrm{J}}(\mathbf{z}, \hat{W}) \leq \lim_{t \to \infty} \min_{\mathbf{z} \in \mathcal{Y}(\mathbf{s})} \hat{\mathrm{J}}_t(\mathbf{z}, \hat{W})$$

Combining this with $\mathrm{K}(\mathbf{s}, \hat{W}) \geq \mathrm{J}_t(\mathbf{s}, \hat{W})$ yields

$$\mathrm{K}(\mathbf{s}, \hat{W}) \geq \min_{\mathbf{x} \in \mathcal{X}(\mathbf{s})} \max_{f \in \mathbb{D}(f_{\hat{W}}, \eta)} \mathrm{E}_f \left[\sum_{i=1}^{n} H_i(x_i, V_i) + \beta \int_{\mathbf{v} \geq \mathbf{0}} \mathrm{K}(\mathbf{x} - \mathbf{V}, \aleph(\hat{W})) f(\mathbf{v}) d\mathbf{v}\right]$$

Therefore

$$\mathrm{J}(\mathbf{s}, \hat{W}) = K(\mathbf{s}, \hat{W}) = \lim_{t \to \infty} \mathrm{J}_t(\mathbf{s}, \hat{W}),$$

so $\mathrm{J}(\mathbf{s}, \hat{W})$ is convex and $\mathrm{J}(\mathbf{s}, \hat{W}) \to \infty$ as $||\mathbf{s}|| \to \infty$. As a result, $\hat{\mathrm{J}}(\hat{\mathbf{s}}, \hat{W})$ is convex and $\hat{\mathrm{J}}(\hat{\mathbf{s}}, \hat{W}) \to \infty$ as $||\hat{\mathbf{s}}|| \to \infty$. Therefore, $\hat{\mathrm{J}}(\hat{\mathbf{s}}, \hat{W})$ achieves its minimum at some finite $\hat{\mathbf{s}}$, which implies our base-stock optimality conditions. □

Lemma B.5 (**Finite Expected Cost**). *There exists an infinite-horizon policy in the backorder case that has finite expected cost.*

Proof. Consider the policy that always orders zero vaccines. For any state $\mathbf{s} \in \mathbb{R}^n$, we have already shown that the cost of

$$\max_{f \in \mathbb{D}(f_{\hat{W}}, \eta)} \mathrm{E}_f \left[\sum_{i=1}^{n} H_i(x_i, V_i)\right]$$

for any order quantity $\mathbf{x} \in \mathcal{X}(\mathbf{s})$ is finite. Hence, the costs due to underage demands occurring in period t can be bounded by some c^u_t for each period, and likewise the overage costs can be bounded by c^o for each period. Denote $c^u = \sup_{t\geq 1} c^u_t$. Then the costs can be bounded by

$$\sum_{t=1}^{\infty} \beta^t \sum_{\tau=1}^{t} (c^u_t + c^o) \leq \sum_{t=1}^{\infty} \beta^t t (c^u + c^o) = \frac{\beta}{(1-\beta)^2}(c^u + c^o)$$

which completes the proof.

□

Lemma B.6 (J **and** $\hat{\mathrm{J}}_t$ **Convexity**). *$\hat{\mathrm{J}}_t(\mathbf{x}, \hat{W})$ is convex in $\mathbf{x}$ and $\mathrm{J}_t(\mathbf{s}, \hat{W})$ is convex in $\mathbf{s}$.*

Proof. We prove via induction. Now, in the base case, when $t = 1$,

$$\hat{\mathrm{J}}_1(\mathbf{x}, \hat{W}) = \max_{f \in \mathbb{D}(f_{\hat{W}}, \eta)} \mathrm{E}_f \left[\sum_{i=1}^{n} H_i(x_i, V_i) \right] \tag{B.17}$$

Now, obviously, each $H_i(x_i, v_i)$ is convex in x_i, hence $\sum_{i=1}^n H_i(x_i, v_i)$ is convex in $\mathbf{x}$. Moreover, for any fixed f, $\mathrm{E}_f\left[\sum_{i=1}^n H_i(x_i, V_i)\right]$ is convex in $\mathbf{x}$ since

$$\begin{aligned}
\mathrm{E}_f \left[\sum_{i=1}^{n} H_i(\lambda x_i^1 + (1-\lambda) x_i^2, V_i) \right] &\leq \mathrm{E}_f \left[\sum_{i=1}^{n} \lambda H_i(x_i^1, V_i) + (1-\lambda) H_i(x_i^2, V_i) \right] \\
&= \lambda \mathrm{E}_f \left[\sum_{i=1}^{n} H_i(x_i^1, V_i) \right] + (1-\lambda) \mathrm{E}_f \left[\sum_{i=1}^{n} H_i(x_i^2, V_i) \right]
\end{aligned}$$

for any $\mathbf{x}^1 = (x_1^1 \ldots, x_n^1)'$, $\mathbf{x}^2 = (x_1^2 \ldots, x_n^2) \in \mathbb{R}^n$, and $\lambda \in [0,1]$. Since the maximum of convex functions is convex, the base case is complete. By similar reasoning, we can show that $\mathrm{J}_1(\mathbf{s}, \hat{W})$ is convex in $\mathbf{s}$:

$$\begin{aligned}
&\min_{\mathbf{x} \in \mathcal{X}(\mathbf{s}^1)} \max_{f \in \mathbb{D}(f_{\hat{W}}, \eta)} \lambda \mathrm{E}_f \left[\sum_{i=1}^{n} H_i(x_i, V_i) \right] + \min_{\mathbf{x} \in \mathcal{X}(\mathbf{s}^2)} \max_{f \in \mathbb{D}(f_{\hat{W}}, \eta)} (1-\lambda) \mathrm{E}_f \left[\sum_{i=1}^{n} H_i(x_i, V_i) \right] \\
&\geq \min_{\substack{\mathbf{x}^1 \in \mathcal{X}(\mathbf{s}^1) \\ \mathbf{x}^2 \in \mathcal{X}(\mathbf{s}^2)}} \max_{f \in \mathbb{D}(f_{\hat{W}}, \eta)} \mathrm{E}_f \left[\lambda \sum_{i=1}^{n} H_i(x_i^1, V_i) + (1-\lambda) \sum_{i=1}^{n} H_i(x_i^2, V_i) \right]
\end{aligned}$$

$$\geq \min_{\substack{\mathbf{x}^1 \in \mathcal{X}(\mathbf{s}^1) \\ \mathbf{x}^2 \in \mathcal{X}(\mathbf{s}^2)}} \max_{f \in \mathbb{D}(f_{\hat{W}}, \eta)} \mathrm{E}_f \left[\sum_{i=1}^{n} H_i(\lambda x_i^1 + (1-\lambda) x_i^2, V_i) \right]$$

$$= \min_{\mathbf{x} \in \mathcal{X}(\mathbf{s})} \max_{f \in \mathbb{D}(f_{\hat{W}}, \eta)} \mathrm{E}_f \left[\sum_{i=1}^{n} H_i(x_i, V_i) \right]$$

since $\lambda \mathbf{x}^1 + (1-\lambda)\mathbf{x}^2 \in \mathcal{X}(\lambda \mathbf{s}^1 + (1-\lambda)\mathbf{s}^2)$.

For the inductive step, we assume both J_t and $\hat{\mathrm{J}}_t$ are true up to $t-1$. For the inductive step to

$$\min_{\mathbf{x} \in \mathcal{X}(\mathbf{s}^1)} \max_{f \in \mathbb{D}(f_{\hat{W}}, \eta)} \lambda \mathrm{E}_f \left[\sum_{i=1}^{n} H_i(x_i, V_i) \right] + \lambda \beta \mathrm{E}_f \left[\mathrm{J}_{t-1}(\mathbf{x} - \mathbf{V}, \aleph(\hat{W})) \right]$$

$$+ \min_{\mathbf{x} \in \mathcal{X}(\mathbf{s}^2)} \max_{f \in \mathbb{D}(f_{\hat{W}}, \eta)} (1-\lambda) \mathrm{E}_f \left[\sum_{i=1}^{n} H_i(x_i, V_i) \right] + (1-\lambda) \beta \mathrm{E}_f \left[\mathrm{J}_{t-1}(\mathbf{x} - \mathbf{V}, \aleph(\hat{W})) \right]$$

$$\geq \min_{\substack{\mathbf{x}^1 \in \mathcal{X}(\mathbf{s}^1) \\ \mathbf{x}^2 \in \mathcal{X}(\mathbf{s}^2)}} \max_{f \in \mathbb{D}(f_{\hat{W}}, \eta)} \mathrm{E}_f \left[\lambda \sum_{i=1}^{n} H_i(x_i^1, V_i) + (1-\lambda) \sum_{i=1}^{n} H_i(x_i^2, V_i) \right]$$

$$+ \mathrm{E}_f \left[\lambda \mathrm{J}_{t-1}(\mathbf{x}^1 - \mathbf{V}, \aleph(\hat{W})) + (1-\lambda) \mathrm{J}_{t-1}(\mathbf{x}^2 - \mathbf{V}, \aleph(\hat{W})) \right]$$

$$\geq \min_{\substack{\mathbf{x}^1 \in \mathcal{X}(\mathbf{s}^1) \\ \mathbf{x}^2 \in \mathcal{X}(\mathbf{s}^2)}} \max_{f \in \mathbb{D}(f_{\hat{W}}, \eta)} \mathrm{E}_f \left[\sum_{i=1}^{n} H_i(\lambda x_i^1 + (1-\lambda) x_i^2, V_i) \right]$$

$$+ \mathrm{E}_f \left[\lambda \mathrm{J}_{t-1}(\lambda \mathbf{x}^1 + (1-\lambda)\mathbf{x}^2 - \mathbf{V}, \aleph(\hat{W})) \right]$$

$$= \min_{\mathbf{x} \in \mathcal{X}(\mathbf{s})} \max_{f \in \mathbb{D}(f_{\hat{W}}, \eta)} \mathrm{E}_f \left[\sum_{i=1}^{n} H_i(x_i, V_i) \right] + \mathrm{E}_f \left[\mathrm{J}(\mathbf{x} - \mathbf{V}, \aleph(\hat{W})) \right]$$

since $\lambda \mathbf{x}^1 + (1-\lambda)\mathbf{x}^2 \in \mathcal{X}(\lambda \mathbf{s}^1 + (1-\lambda)\mathbf{s}^2)$. The case for $\hat{\mathrm{J}}(\mathbf{x}, \hat{W})$ follows in a nearly identical fashion.

□

References

Decroix, G., Arreola-Risa, A. 1998. Optimal Production and Inventory Policy for Multiple Products Under Resource Constraints. *Management Science* **44**(7) 950-961.

APPENDIX C

DISTRICT-MANGED VACCINE SUPPLY NETWORKS

C.1 Heuristic Details

To further detail our heuristic approach (in particular, our routing strategies), we let

$$\mathcal{T}(i) = \{t \in \mathbb{N} | t \bmod i = 0 \bigcap t \leq \bar{\tau}\}, \tag{C.1}$$

and let $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_{\bar{\tau}})$ represent the number of each class in a given grouping. Then, if s is the targeted class of reduction and $\nu_s \bmod m + \sum_{i=s+1}^{\bar{\tau}+1} \nu_i \geq m$, (where $\nu_{\bar{\tau}+1} = 0$ for notational convenience), we can employ the following MIP to make routing decisions. We let $T = s_2/s_1$, then we define the following two MIPs:

$$\begin{aligned}
\eta_1(\boldsymbol{\nu}, \mathcal{M}) = \min \quad & \sum_{i \in \mathcal{M} \bigcup \{0\}} \sum_{\substack{j \in \mathcal{M} \bigcup \{0\} \\ j \neq i}} d_{i,j} u_{i,j,t} & \\
\text{s.t.} \quad & \sum_{j \in \mathcal{M}} (u_{0,j} + u_{j,0}) = 2 & \\
& \sum_{\substack{i \in \mathcal{M} \bigcup \{0\} \\ i \neq j}} u_{i,j} = y_j & j \in \mathcal{M} \\
& \sum_{\substack{j \in \mathcal{M} \bigcup \{0\} \\ i \neq j}} u_{i,j} = y_i, & i \in \mathcal{M} \\
& b_i - b_j + n u_{i,j} \leq n - 1 + n\big(2 - (y_i + y_j)\big) & i, j \in \mathcal{M}, i \neq j \\
& \sum_{i \in \mathcal{M}(j)} y_i = \nu_j & j = 1, \ldots, \bar{\tau} \\
& y_i \in \mathbb{B}, u_{i,j} \in \mathbb{B}, b_i \in \mathcal{N} &
\end{aligned} \tag{C.2}$$

$$\begin{aligned}
\eta_2(s_1, s_2, \nu_{s_2}, \mathcal{W}_1, \mathcal{W}_2) = \min \quad & \sum_{t=0}^{T-1} \sum_{i \in \mathcal{W} \bigcup \{0\}} \sum_{\substack{j \in \mathcal{W} \bigcup \{0\} \\ j \neq i}} d_{i,j} u_{i,j,t} & \\
\text{s.t.} \quad & \sum_{j \in \mathcal{W}} (u_{0,j,t} + u_{j,0,t}) = 2 & t = 0, \ldots, T-1 \\
& \sum_{\substack{i \in \mathcal{W} \bigcup \{0\} \\ i \neq j}} u_{i,j,t} = y_{j,t} & j \in \mathcal{W}, t = 0, \ldots, T-1
\end{aligned} \tag{C.3}$$

$$\sum_{\substack{j \in \mathcal{W} \bigcup \{0\} \\ i \neq j}} u_{i,j,t} = y_{i,t}, \qquad i \in \mathcal{W}, t = 0, \ldots, T-1$$

$$b_{i,t} - b_{j,t} + m u_{i,j,t} \leq m - 1 + n\big(2 - (y_{i,t} + y_{j,t})\big)$$

$$i, j \in \mathcal{W}, i \neq j, t = 0, \ldots, T-1$$

$$\sum_{i \in \mathcal{W}_2} y_{i,t} = \nu_{s_2}/T \qquad t = 0, \ldots, T-1$$

$$\sum_{t=0}^{T-1} y_{i,t} \leq 1 \qquad i \in \mathcal{W}_2$$

$$y_{i,t} = 1 \qquad i \in \mathcal{W}_1, t = 0, \ldots, T-1$$

$$y_{i,t} \in \mathbb{B}, u_{i,j} \in \mathbb{B}, b_i \in \mathcal{N}$$

Otherwise, if $\nu_s \bmod m + \sum_{i=s+1}^{\bar{\tau}+1} \nu_i < m$, routes can be determined by ψ. Here, $y_{i,t} = 0$ signifies that IHC i is visited and the routes are determined via the $u_{i,j,t}$. Though this handles the fully-observed case, the Bayesian case is handled nearly identically by denoting $\hat{\eta}_1$ and $\hat{\eta}_2$ as η_1 and η_2 with θ replaced with $\hat{\theta}$.

Combining these tools allows us to create a heuristic which works to create high-density routes while visiting each IHC near their class. This heuristic, expressed in Algorithm 3, iteratively groups and provides routing policies for IHCs with a priority on routing smaller classes and is based on fixed-τ policies like (4.19) and (4.20). Here, W is an $n \times \bar{\tau}$ matrix with components $w_{i,t}$ that act as storage for the routing strategies and $\bar{\mathbf{y}}$ is the n-vector that stores the fixed-τ policy for each IHC. In the fully-observed case, $w_{i,t} = j$ if the edge from i to j is used whenever the current period $\hat{t} = t \bmod \bar{\tau} + 1$ and $\bar{y}_i = \tau$ if IHC $i \in \mathcal{N}$ is served every τ periods. In the Bayesian case, as discussed in Section 4.6, we consider routing decisions for periods $\hat{t} < T_1$, and similarly let $w_{i,t} = j$ if edge i to j is used in period $\hat{t} = t \bmod \bar{\tau} + 1$, and $\bar{y}_i = \tau$ if IHC $i \in \mathcal{N}$ is served every τ periods.

Algorithm 3 Heuristic Route Assignment

```
function HEURISTIC_ASSIGNMENT
    $\mathcal{M} \leftarrow \mathcal{N}$
    for $s = 1, s++$, while $s \leq \bar{\tau}$ and $\mathcal{M} \neq \emptyset$ do
        $\bar{\mathbf{y}} \leftarrow \mathbf{0}$
        $\boldsymbol{a} \leftarrow$ SOFT_REDUCE$((|\mathcal{M}(1)|, \ldots, |\mathcal{M}(\bar{\tau})|), s)$
        $(\boldsymbol{a}, \boldsymbol{u}) \leftarrow$ HARD_REDUCE$(\boldsymbol{a}, s)$
        $\boldsymbol{b} \leftarrow (|\mathcal{M}(1)|, \ldots, |\mathcal{M}(\bar{\tau})|) - \boldsymbol{a}$
        if $b_s \bmod m = 0$ then
            $z_i \leftarrow (s, s)$ for all $i \in \mathcal{M}(s)$
            $\mathcal{M} \leftarrow \mathcal{M} \setminus \mathcal{M}(s)$
        else if $s < \bar{\tau}$ and $b_s \bmod m + \sum_{\tau=s+1}^{\bar{\tau}} b_\tau \leq m$ then
            Solve $\eta_1((0, \ldots, 0, b_s \bmod m, b_{s+1}, \ldots, b_{\bar{\tau}}), \mathcal{M})$
            $\mathcal{W} = \{i \in \mathcal{M} | y_i > 0\}$ and $\mathcal{M} \leftarrow \mathcal{M} \setminus (\mathcal{W} \bigcup \mathcal{M}(s))$
            $z_i \leftarrow (s, s)$ for all $i \in \mathcal{W} \bigcup \mathcal{M}(s)$
        else if $s < \bar{\tau}$ and $\sum_{i=s}^{\bar{\tau}} b_i \geq m$ then
            Solve $\eta_1((0, \ldots, 0, b_s \bmod m, b_{s+1}, \ldots, b_{\bar{\tau}}), \mathcal{M})$
            $\mathcal{W} = \{i \in \mathcal{M} | y_i > 0\}$ and $\mathcal{M} \leftarrow \mathcal{M} \setminus \mathcal{W}$
            Solve $\eta_1((0, \ldots, 0, b_s \bmod m, u_{s+1}, \ldots, u_{\bar{\tau}}), \mathcal{W})$
            $\mathcal{W}_1 = \{i \in \mathcal{W} | y_i > 0\}$ and $\mathcal{W}_2 = \mathcal{W} \setminus \mathcal{W}_1$
            $z_i \leftarrow (s, s)$ for all $i \in \mathcal{W}_1 \bigcup \mathcal{M}(s)$
            for $s_2 = s + 1, s_2++$, while $s_2 \leq \bar{\tau}$ do
                if $s_2 \bmod s = 0$ and $|\mathcal{W}_2(s_2)| > 0$ then
                    Solve $\eta_2(s, s_2, |\mathcal{W}_2(s_2)|, \mathcal{W}_1, \mathcal{W}_2(s_2))$
                    $\bar{y}_i \leftarrow \max_{t=0,\ldots,s_2/s-1} s(t+1) y_{i,t}$ for all $i \in \mathcal{W}_2(s_2)$
                    $z_i \leftarrow (\bar{y}_i, \theta(i))$ for all $i \in \mathcal{W}_2(s_2)$
            $\mathcal{M} \leftarrow \mathcal{M} \setminus \mathcal{M}(s)$
        else
            $z_i \leftarrow (s, s)$ for all $i \in \mathcal{M}$
            break
    return $z_1, z_2, \ldots, z_n$
```

C.2 Vehicle Capacity Restrictions

If the manager wishes to modify the assumption of a m IHC visits per excursion for the purpose of reflecting the vehicle's capacity, this can easily be accomplished via our

MIP formulations if we restrict the policy space to only fixed-τ policies. Therefore, letting $m \in \mathbb{N}$ now refer to the number of demand units a vehicle holds and letting $a_{j,\tau} \in \mathbb{N}$ for $j \in \mathcal{N}$ and $\tau \in \{1, \ldots, \bar{\tau}\}$ denote the level of capacity that is taken up by IHC j when it is visited every τ time periods. For example, if a manager insists that a vehicle must carry at least 1.5 times the mean demand (within the associated order fixed-τ ordering interval), and $\lambda_1 = 150$ or $\alpha_1/\beta_1 = 150$, if vehicle capacity $m = 10$ (in hundreds of demand units), then $a_{1,1} = 2$, $a_{1,2} = 5$, and $a_{1,3} = 7$.

Then, changing the constraint

$$b_{i,t} - b_{j,t} + mu_{i,j,t} \leq m - 1, \qquad i, j \in \mathcal{N}, i \neq j, t = 1, ..., T$$

in equations (4.19) and (4.20) (where T represents $T_{\bar{\tau}}$ in the fully-observed case and $T_1 - 1$ in the Bayesian case) to

$$b_{i,t} - b_{j,t} + mu_{i,j,t} \leq m - \sum_{\tau=1}^{\bar{\tau}} y_{j,\tau,t \bmod \tau} a_{j,\tau}, \qquad i, j \in \mathcal{N}, i \neq j, t = 1, ..., T$$

results will generate vehicle routes that never exceed the capacity constraints implied by $a_{i,\tau}$, and can be accomplished with the same number of constraints as those in equations (4.19) and (4.20). However, since m now represents demand, the variables $b_{i,t}$ may be required to take on larger values, which can moderately increase the complexity of the programs.

C.3 Proofs of Propositions, Lemmas, and Theorems

LEMMA C.1 (**Dynamic Program Equivalence**). $\mathrm{V}_T(\lambda)$ *and* $\hat{\mathrm{V}}_T(\alpha, \beta)$ *correspond to the fully-observed and learning objectives* (4.3) *and* (4.5)*.*

Proof. The base case in both (4.3) and (4.5) (where $t = T$) is immediate so long as

$$c_i(\tau\lambda) = \mathrm{E}_\lambda\left[\sum_{i=1}^{\tau}(Y_i - s)^+\right] \text{ and } \hat{c}_i(\alpha, \beta, \tau) = \mathrm{E}_g\left[\sum_{i=1}^{\tau}(Y_i - s)^+\right], \tag{C.4}$$

which we will show at the end of the proof. Therefore, we first show that $\mathrm{V}_t(\lambda)$ is equivalent to (4.3). When $t < T$ in (4.3), this corresponds to the case where $t > 0$ in $\mathrm{V}_t(\lambda)$. Now,

$$\begin{aligned}
\mathrm{E}\left[\sum_{\bar{t}=t}^{T} Z_{\bar{t}}^{\pi}\right] - \mathrm{V}_t(\boldsymbol{\tau}) &= \sum_{\bar{t}=t}^{T} \nu \sum_{i=1}^{n} a_{it}^{\pi}\Big(\sum_{\hat{t}=t-\tau_{it}^{\pi}}^{t} X_{i\hat{t}} - q_i\Big)^+ \\
&\quad - \sum_{i=1}^{n} a_i c_i(\lambda_i \tau_i) - \mathrm{V}_{t-1}(((1-a_1)\tau_1 + 1, \ldots, (1-a_n)\tau_n + 1)), \\
&= \nu \sum_{i=1}^{n} a_{it}^{\pi}\Big(\sum_{\hat{t}=t-\tau_{it}^{\pi}}^{t} X_{i\hat{t}} - q_i\Big)^+ - \sum_{i=1}^{n} a_i c_i(\lambda_i \tau_i),
\end{aligned}$$

by the inductive hypothesis, hence so long as (C.4) holds, the dynamic program follows. $\hat{\mathrm{V}}_T(\alpha, \beta)$ hold in an identical fashion.

For showing the equivalence of $c_i(\tau\lambda)$, it is well known that $\sum_{i=1}^{\tau} Y_i$ is Poisson with parameter $\tau\lambda$ since it is the sum of τ independent Poisson random variables. Letting $\mu = \tau\lambda$ and Y denote a Poisson random variable with parameter μ, as shown in Geyer (2017),

$$\mathrm{E}_\mu\left[Y \mathbb{1}(Y > k)\right] = \mu(1 - F(k-1, \mu)),$$

hence

$$\begin{aligned}
\mathrm{E}_\mu\left[(Y - s)^+\right] &= \mathrm{E}_\mu\left[(Y - s)\mathbb{1}(Y > s)\right] \\
&= \mu(1 - F(s-1, \mu)) - s(1 - F(s, \mu)),
\end{aligned}$$

which proves our assertion for the λ fully-known case.

Similarly, for showing the equivalence of $c_i(\alpha,\beta,\tau)$, when $Y=\sum_{i=1}^{\tau} Y_i$ is Poisson with Gamma prior, it is well known that the posterior distribution is negative binomial. As shown in Geyer (2017),

$$\mathrm{E}_g\left[Y\mathbb{1}(Y>k)\right]=\frac{(k+\tau\alpha)(1-p)h(k)+\tau\alpha(1-p)(1-H(k))}{p},$$

for negative binomial random variables where $p=\beta/(1+\beta)$, hence,

$$\begin{aligned}\mathrm{E}_\mu\left[(Y-s)^+\right]&=\mathrm{E}_\mu\left[(Y-s)\mathbb{1}(Y>s)\right]\\&=\frac{(H(s,\tau\alpha,\beta)-1)(p(\tau\alpha+s)-\tau\alpha)-(p-1)h(s,\tau\alpha,\beta)(\tau\alpha+s)}{p}\\&=\frac{(s+\tau\alpha)h(s,\tau\alpha,\beta)+(\beta s-\tau\alpha)(-1+H(s,\tau\alpha,\beta))}{\beta}\end{aligned}$$

which proves our assertion. □

Lemma C.2 (**Full-Observed Convexity**). *Cost function $c(\tau\lambda)$ is convex increasing in λ.*

Proof. It suffices to show the lemma is true for $c_i(\lambda)$ with $\nu=1$. Investigating the second derivative of c_i,

$$\frac{d^2}{d\lambda^2}c_i(\lambda)=\frac{e^{-\lambda}\lambda^{-1+q}\left(q(1-q+\lambda)\Gamma(q-1)+(q-\lambda)\Gamma(q)\right)}{\Gamma(q-1)\Gamma(q)}$$

and observing that

$$\begin{aligned}q(1-q+\lambda)\Gamma(q-1)+(q-\lambda)\Gamma(q)&=\Gamma(q-1)(q(1-q+\lambda)+(q-\lambda)(q-1))\\&=\Gamma(q-1)\lambda>0,\end{aligned}$$

hence $c_i(\tau\lambda)$ is convex in λ. □

Proof of Proposition 4.1. To point (i), in the infinite horizon case, it suffices to show that per-period cost over any horizon is dominated by our proposed policy. Let

$\hat{\tau} = \arg\min_{\tau \geq 1} \frac{1}{\tau}(k + 2d_{1,0} + c_1(\tau\lambda))$. Each $\hat{\tau}$ order results in $\frac{1}{\hat{\tau}}(k + 2d_{1,0} + c_1(\tau\lambda))$ per period over the course of $\hat{\tau}$ periods. Comparing the policy that always orders τ to the policy that orders $\hat{\tau}$ at some time t, the $\hat{\tau}$ policy obviously experiences smaller per period costs by definition.

To point (ii), when $T + 1 \bmod \hat{\tau} = 0$, it can be observed that a manager can order according to $\hat{\tau}$ at every period, which allows for the average per period cost of $\frac{1}{\hat{\tau}}(k + 2d_{1,0} + c_1(\hat{\tau}\lambda))$, which is a lower bound on per period costs (in accordance with the proof of (i)), which also proves (iii).

□

Proof of Proposition 4.2 and 4.5. Via Proposition 4.1, a lower bound on the per-period cost for a single IHC i can $\frac{1}{\hat{\tau}}(k + 2d_{i,0} + c_1(\hat{\tau}\lambda))$ if $m = 1$. When $m > 1$, if m IHCs are located at exactly the same position as IHC i and served only when IHC i is served, minimizing costs with respect to IHC in the same manner as Proposition 4.1 will yield

$$\min_{\tau \geq 1} \frac{1}{\tau}\left(\frac{k + 2d_{0,i}}{m} + c_i(\tau\lambda_i)\right)$$

of costs per period since IHC i only contributes $(k + 2d_{0,i})/m$ to the total travel costs in this (potentially fictitious) scenario. Hence, this serves as a lower bound on per-period costs possible for IHC i, and considering this for each IHC proves the proposition. The proof for the Bayesian case follows in exactly the same manner using Proposition 4.3. □

Proof of Proposition 4.3. For Part (i), we first show that $c_i(\tau\alpha/\beta) \leq \hat{c}_i(\alpha, \beta, \tau)$. This is true since $c_i(\tau\alpha/\beta)$ is convex in α/β, and $\frac{\tau\alpha}{\beta} = \mathrm{E}_g[\tau\lambda]$, hence by Jensen's inequality we have

$$c_i(\tau\alpha/\beta) \leq \mathrm{E}_g[c_i(\tau\lambda)] = \hat{c}_i(\alpha, \beta, \tau).$$

Now, moving forward by induction, the base case, where $t = 1$ is automatically satisfied by $c_i(\tau\alpha/\beta) \leq \hat{c}_i(\alpha, \beta, \tau)$. For the inductive step, suppose that τ, associated π, is the best fixed-τ policy for $\mathrm{V}_t^\pi(\alpha, \beta)$. Then if $\tau < t$,

$$\begin{aligned}
\mathrm{V}_t^\pi(\alpha/\beta) &= k + 2d_{1,0} + c_1(\hat{\tau}\alpha/\beta) + \mathrm{V}_{t-\hat{\tau}}^\pi(\alpha/\beta) \\
&\leq k + 2d_{1,0} + c_1(\hat{\tau}\alpha/\beta) + \sum_{y=0}^{\infty} \mathrm{V}_{t-\hat{\tau}}^\pi\left(\frac{\alpha+y}{\beta+\hat{\tau}}\right) h(y, \alpha, \beta) \\
&\leq k + 2d_{1,0} + \hat{c}_1(\alpha, \beta, \hat{\tau}) + \sum_{y=0}^{\infty} \hat{\mathrm{V}}_{t-\hat{\tau}}^\pi(\alpha + y, \beta + \hat{\tau}) h(y, \alpha, \beta)
\end{aligned}$$

where the first line to second occurs via Jensen's inequality since $\mathrm{V}_{t-\hat{\tau}}^\pi(\alpha/\beta)$ is a convex function in α/β. The second to third inequality takes place due to the inductive hypothesis. Otherwise, if $\tau \geq t$, the proof is the same as the base case.

To show Part (ii), note that as $\lim_{\beta\to\infty} h(\tau, \beta\lambda, \beta) = f(\tau\lambda)$ since negative binomial is known to converge to the Poisson distribution in this limit. As such, $\lim_{\beta\to\infty} \hat{c}_i(\beta\lambda, \beta, \tau) = \hat{c}(\tau\lambda)$, hence, proceeding by induction, the base case when $t = 1$ is obvious. To the inductive step, suppose that the optimal action for $\lim_{\beta\to\infty} \hat{\mathrm{V}}_t(\lambda_1\beta, \beta)$ is associated with $\tau < t$. Then,

$$\begin{aligned}
\lim_{\beta\to\infty} \hat{\mathrm{V}}_t(\beta\lambda, \beta) &= \lim_{\beta\to\infty} k + 2d_{1,0} + \hat{c}_1(\beta\lambda, \beta, \tau) + \sum_{y=0}^{\infty} \hat{\mathrm{V}}_{t-\tau}(\lambda\beta + y, \beta) h(y, \tau\beta\lambda, \beta) \\
&= k + 2d_{1,0} + c(\tau\lambda) + \sum_{y=0}^{\infty} \mathrm{V}_{t-\tau}(\lambda) f(y, \lambda) \\
&\geq \mathrm{V}_t(\lambda),
\end{aligned}$$

which achieves equality if τ minimizes $\mathrm{V}_t(\lambda)$. Hence in this case, τ must be identical to the τ that minimizes $\mathrm{V}_t(\lambda)$. Otherwise, if $\tau \geq t$, the proof is the same as the base case.

Finally, to show part (iii), we note that the left hand side of the inequality is immediate since this is the case of fully observed λ. The right hand side of the inequality is the evaluation of the potentially suboptimal fixed-τ policy, hence it is an upper bound to the optimal policy for $\hat{\mathrm{V}}_t(\alpha, \beta)$. □

Proof of Corollary 4.1. By Proposition 4.3, $\frac{t+1}{\hat{\tau}}(k+2d_{1,0}+\hat{c}_1(\tau,\alpha,\beta))$ and $\mathrm{E}_g\left[\mathrm{V}_{t-\tau}(\lambda)\right]$ act as upper and lower bounds to $\mathrm{V}_t(\alpha,\beta)$ and hence, so long as $\mathrm{E}_{\hat{g}}\left[\mathrm{V}_{t-\tau}(\lambda)\right]$ acts as a lower bound to $\sum_{y=0}^{\infty}\hat{\mathrm{V}}_{t-\tau}(\alpha+y,\beta+\tau)h(y,\alpha,\beta)$, if there exists a τ that satisfies the conditions of the Corollary, an upper bound to $\hat{\mathrm{V}}_t(\alpha,\beta)$ has less cost than a lower bound after first action, which proves the assertion. To show that this is a lower bound,

$$\begin{aligned}
\sum_{y=0}^{\infty}\hat{\mathrm{V}}_{t-\tau}(\alpha+y,\beta+\tau)h(y,\tau\alpha,\beta) &\geq \sum_{y=0}^{\infty}\mathrm{E}_{\alpha+y,\beta+\tau}\left[\mathrm{V}_{t-\tau}(\lambda)\right]h(y,\tau\alpha,\beta)\\
&= \sum_{y=0}^{\infty}\left(\int_{\lambda=0}^{\infty}\mathrm{V}_{t-\tau}(\lambda)g(\lambda,\alpha+y,\beta+\tau)d\lambda\right)h(y,\tau\alpha,\beta)\\
&= \int_{\lambda=0}^{\infty}\mathrm{V}_{t-\tau}(\lambda)\left(\sum_{y=0}^{\infty}g(\lambda,\alpha+y,\beta+\tau)h(y,\tau\alpha,\beta)\right)d\lambda\\
&= \mathrm{E}_{\hat{g}}\left[\mathrm{V}_{t-\tau}(\lambda)\right]
\end{aligned}$$

where, as discussed in Xekalaki (1981), summing gamma over negative binomial distribution yields distributions of the form $\hat{g}$ given by

$$\frac{\lambda^{\alpha-1}\left(\frac{\beta}{\beta+1}\right)^{\alpha\tau}(\beta+\tau)^{\alpha}e^{-\lambda(\beta+\tau)}\,{}_1F_1\left(\alpha\tau;\alpha;\frac{\lambda(\beta+\tau)}{\beta+1}\right)}{\Gamma(\alpha)}$$

where ${}_1F_1$ is the hypergeometric function which concludes the proof so long as $\tau < t$. If $\tau \geq t$, the problem reduces to the myopic case, and hence is immediate. □

Proof of Proposition 4.4. Consider two policies: π_1 chooses action τ at time t and τ again at time $t-\tau$, then follows the optimal policy for all remaining decision epochs. The other policy π_2 chooses action 2τ at time t, then follows the optimal policy for all remaining decision epochs. Obviously, both π_1 and π_2 yield identical costs to go from time $t - 2\tau$, yet π_1 experiences $2k + 4d_{1,0} + 2\hat{c}_1(\tau, \alpha, \beta)$ whereas π_2 experiences $k + 2d_{1,0} + c(2\tau, \alpha, \beta)$ from periods t to $t - 2\tau$. Thus, if

$$\begin{aligned} 2k + 4d_{1,0} + 2\hat{c}_1(\tau, \alpha, \beta) &\leq k + 2d_{1,0} + \hat{c}(2\tau, \alpha, \beta) \\ \implies k + 2d_{1,0} &\leq \hat{c}_1(2\tau, \alpha, \beta) - 2\hat{c}_1(\tau, \alpha, \beta) \end{aligned}$$

this implies that two fills can occur for less cost than a single fill over the course of 2τ periods, hence for any policy calling for an action larger than 2τ, there exists a policy with less cost that orders twice during that interval. □