COPE WITH THE COVID-19 PANDEMIC:DYNAMIC BED ALLOCATION AND PATIENT SUBSIDIZATION IN A PUBLIC HEALTHCARE SYSTEM

XUE ZHAO

MPhil

The Hong Kong Polytechnic University

2021

The Hong Kong Polytechnic University

Department of Logistics and Maritime Studies

Cope with the COVID-19 Pandemic: Dynamic Bed Allocation and Patient Subsidization in a Public Healthcare System

Xue ZHAO

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Philosophy

May 2021

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____(Signed)

____Xue ZHAO____(Name of student)

Abstract

In many countries and territories, public hospitals play a major role in coping with the COVID-19 pandemic. For public hospital managers, on the one hand, they must best utilize their hospital beds to serve the COVID-19 patients immediately. On the other hand, they need to consider the need of bed resources from non-COVID-19 patients, including emergency and elective patients. In this work, we consider two control mechanisms for public hospital managers to maximize the overall utility of patients. One is the dynamic allocation of bed resources according to the evolution process of the COVID-19 pandemic. The other is the usage of a subsidy scheme to move elective patients from the public to private hospitals. We develop a dynamic programming model to study the effect of bed allocation and patient subsidization in serving three types of patients, COVID-19, emergency, and elective-care. We first demonstrate the multimodularity of the total expected cost function on the number of isolation beds and the length of waiting list, which assures the monotonicity of the optimal allocation decision (i.e., how many beds should be transferred between isolation beds and ordinary beds) and the optimal subsidization decision (i.e., how many elective patients should be moved to private hospitals) in the state variables in each period. We then show that the dynamic allocation between isolation and ordinary beds can provide a better utilization of bed resources, by cutting down at least 33.5% of the total cost compared with the static policy (i.e., keeping a fixed number of isolation beds) when facing a medium pandemic alert. Furthermore, we present that subsidizing elective patients and moving them to private hospitals is an efficient way to ease the overcrowded situation in public hospitals, as we numerically show that it could reduce the length of waiting list and the total expected cost at the same time.

Acknowledgements

First, I would like to express my deepest gratitude to my two "chief" supervisors, my prior chief supervisor, Prof. Pengfei GUO, and my current chief supervisor, Prof. Guang XIAO. They both give me continuous guidance and support in my Mphil study. We discussed frequently on the research problems. Their invaluable advice helped me overcome many difficulties in my research. Their rigorous attitude and great passion towards research influence me a lot. It's definitely a great honor to have them as my supervisors.

I gratefully acknowledge the professors who give me many helps. I would like to thank Prof. Yulan WANG, Prof. Miao SONG, Prof. Li JIANG for their wonderful lectures. I have learned a lot from those lectures. I would also like to thank Prof. Miao Song and Prof. Hengqing Ye for their valuable comments in my Mphil confirmation.

My special thanks go to Dr Xin Ma for his big help and support. Moreover, I would like to thank Dr. Yanli Tang. Each time I lose confidence, she always encourages me to march forward. Also, many thanks to all my classmates and friends in LMS, for their helps and companion in such a special time.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The coronavirus disease 2019 (COVID-19) has struck the world in a fast speed, affecting 222 countries, areas or territories in months; until 31st December 2020, there have been 80,773,033 confirmed cases (WHO, 2020). In coping with the COVID-19 pandemic, public hospitals play a major role in many counties and territories. For example, in Hong Kong (HK), COVID-19 patients are only treated by public hospitals appointed by the government. The public hospital managers are facing a significant challenge. On the one hand, with a surging number of COVID-19 patients, they are facing massive shortages of isolation beds. On the other hand, they still need to consider the need from the non-COVID-19 patients with other diseases such as cancer and leukemia. During the pandemic period, many non-COVID-19 patients, particular those emergency ones, still require hospitalization in public hospitals. How to optimally allocate the limited amount of hospital beds between COVID-19 and non-COVID-19 patients is a very critical management problem.

The inpatient beds for two types of patients are different. Since COVID-19 is very contagious, one person with COVID-19 can infect two or two-and-a-half healthy people (ABC News, 2020). Thus, to avoid cross-infection in a healthcare facility, the COVID-19 patients must be provided with the isolation beds under a negative pressure environment. In contrast, non-COVID 19 patients only need ordinary beds. Two types of beds (i.e., isolation and ordinary beds) need to be separately prepared to satisfy different demands from COVID-19 and non-COVID-19 patients. The two types of beds can be transformed into each other, after taking certain steps of modification work. In this work, we not only consider the bed reservation issue on

two types of beds, but also consider the transformation possibilities between them.

Different types of patients have different urgency levels on bed request. To prevent further spread of coronavirus, the adequate capacity of isolation beds should be prepared so that confirmed COVID-19 patients can be admitted immediately. The non-COVID-19 patients can be further divided into two classes, emergency non-COVID-19 patients (hereafter referred to emergency patients) and elective non-COVID-19 patients (hereafter referred to elective patients). The emergency patients must be admitted immediately upon their arrival, while the elective patients can wait. Therefore, ordinary beds shall be first allocated for emergency patients and the extra ones can be used to admit elective patients.

Upon the outbreak of COVID-19, the surging number of COVID-19 patients can cause big pressure to the public hospitals. Elective patients are often been put on waiting list, since reducing elective surgery is one of essential manners to utilize the limited medical capacity to cope with COVID-19 (7NEWS, 2020). For example, millions of operations in the UK had been postponed by National Health Service (NHS) to tackle the COVID-19 pandemic (The Guardian, 2020). Waiting in the queue can cause a deterioration in elective patients' health condition and a waiting cost is incurred for them. Therefore, besides serving the COVID-19 patients and emergency patients, reducing waiting time for elective patients is also a key issue for the public hospital managers. One effective way to reduce the waiting line is to subsidize elective patients and refer them to private hospitals for treatment. Even without pandemic, the subsidy mechanism has been widely adopted to relieve the pressure of the public hospitals by many countries and areas such as England, Australia and Hong Kong; see Qian et al. (2017). There are multiple merits by conducting this mechanism during the COVID-19 pandemic. One, elective patients can obtain faster treatment in private hospitals, without worrying about the expensive payment and the infection from COVID-19. Two, the overcrowded situation in public hospitals can be eased and the bed resource can be utilized to cope with the COVID-19 patients. Three, private hospitals may face a thinned patient flow in the pandemic period because some of their regular patients are afraid of going to hos-

pitals during pandemic. It is good for them to serve these subsidized patients, even with some price discounts. Indeed, between late February and April 15, 2020, the Hong Kong government subsidized part of non-COVID-19 patients, e.g., 66 babies, 5 cancer patients, and 15 pregnant women, and referred them to private hospitals for treatment (Cheung, 2020). However, subsidy scheme is not perfect because the government budget is limited and governments have to adopt some requirements on patients so that the subsidy is appropriately applied; see Qian et al. (2017) for the discussions of pros and cons of different types of subsidy schemes. We do not consider the detailed format of subsidy scheme here but instead we consider a general cost function associated with the number of subsidized patients. The policy makers concern about the following question: how many elective patients in public hospitals should be subsidized and be referred to the private hospitals during different phases of the COVID-19 pandemic?

We develop a dynamic programming model to find optimal decisions on bed allocation and patient subsidization in different time periods, by considering the evolution process of the COVID-19 pandemic. Compared with the literature results, our dynamic programming model has the following unique features. First, we take a more holistic approach by jointly considering the hospital bed allocation and waiting list management while past studies consider these two problems separately. Second, we consider three types of patients (COVID-19, emergency and elective patients) in our study while past studies normally consider two types (emergency and elective patients). Furthermore, the arrival processes of different type of patients are time-dependent, while they are often assumed to be stationary in the past studies. This time-dependence feature is driven by the evolution process of pandemic and its big impact on patients' behavior. In the outbreak phase of the pandemic, an increasing number of COVID-19 patients are presenting at public hospitals and the number of arrived elective patients are decreasing because they are worried being infected by the virus. In the post-peak phase, the number of COVID-19 patients is decreasing and more elective patients will seek medical care in public hospitals. Third, our bed allocation decision considers two types of bed reservation issues: iso-

lation beds and ordinary beds. Furthermore, we also consider the bed conversion issue: converting ordinary beds to isolation beds during the outbreak phase and converting the isolation beds back into ordinary beds during the post-peak phase, which provide a better utilization of bed resource.

Our results are helpful for hospital managers. First, we demonstrate the multimodularity of the total expected cost function on the number of reserved isolation beds and the length of waiting list. The multimodularity assures the monotonicity of the optimal allocation decision (i.e., how many beds should be transferred between isolation beds and ordinary beds) and the optimal subsidization decision (i.e., how many elective patients should be moved to private hospitals) in the state variables (i.e., the number of reserved isolation beds and the length of waiting list) in each period. Second, we show that our dynamic allocation rule can provide a better utilization of isolation and ordinary beds in coping with an infectious disease. Compared with the static policy (i.e., keeping a fixed number of isolation beds), our approach can greatly cut down the total expected cost. Furthermore, we find that our dynamic allocation rule works best under conditions when a region facing a medium pandemic situation, in which our approach can reduce the total expected cost by at least 33.5%. Third, our results show that subsidizing elective patients and referring them to private hospitals is an efficient way to ease the overcrowded situation in public hospitals. Specifically, by comparing the experiment outcomes with and without a subsidy scheme, we show that a subsidy scheme can shorten the length of waiting list by 58% and reduce the total expected cost by 38% . We also find that the cost saving effect of a subsidy scheme is convex in the hospital bed capacity. In short, these useful results are critical to hospital managers for determining the best strategies for serving multiple types of patients through the course of a pandemic.

The remainder of this paper is organized as follows. The related literature is reviewed in chapter 2. The basic setting of the hospital bed allocation and patient subsidization model is introduced in chapter 3. The research problem is formulated as a finite-horizon Markov decision process. And in chapter 4, we analyze

the structure properties of dynamic programming model, which establishes fundamental insights on the optimal scheduling rule. In chapter 5, numerical studies are conducted to validate our analytical results. Finally, conclusions and implications are discussed in chapter 6. All proofs can be found in Appendix A.

Chapter 2

Literature Review

This paper focuses on the dynamic hospital bed allocation and patient subsidization problem, which is related to the following streams of literature: infectious disease management, capacity planning in a healthcare system and dynamic multiproduct inventoty control problem.

2.1 Infectious disease management

Our work is related to the literature on infectious disease management. The allocation of scarce resources is a key issue to cope with infectious diseases, e.g., SARS, MERS, H1N1, and COVID-19. For example, Mamani et al. (2013) examine a contractual mechanism to increase vaccine supplies in advance of seasonal influenza rather than in response to a pandemic. Ekici et al. (2014) develop a facility location and resource allocation model to analyze food distribution when facing a severe influenza pandemic. Deo and Sohoni (2015) build an optimization model to study the allocation of point-of-care HIV diagnostic devices to mitigate the potential of long diagnostic delays. Long et al. (2018) formulate an optimization model to forecast the occurrence of new infections and to analyze spatial resource allocation. Blackmon et al. (2021) develop a decision support system to assist food distribution.

More recently, in response to COVID-19 epidemic, many researches are conducted to address various issues from rapid diagnosis to clinical management (see Phua et al. (2020) for review). In the operations management field, Kaplan (2020) develops scratch models to help support local level decisions. That work documents

problem faced, models developed, and advice offered during real-time response to the COVID-19 crisis at the local level. These problems include restricting the size of university events, stress testing COVID-19 intensive care capacity, university timing decisions, and tracking the outbreak and intervention scenarios. Therein, the most relevant part to our study is "stress test" of the hospital's Intensive Care Unit (ICU) capacity. They focus on determining the maximum arrival rates of COVID-19 patients the hospital could handle based on standard Erlang loss models. In their approach, they study different arrival scenarios for COVID-19 patients and non-COVID-19 patients and give sample findings that if the existing ICU capacity is enough to cope with such arrivals. Our work also studies the hospital beds capacity decision problem in face of COVID-19 and non-COVID-19 patients. We mainly focus on the dynamic isolation beds reservation decision with the progression of COVID-19.

2.2 Capacity planning in a healthcare system

Our work is also related to capacity planning in a healthcare system. Driven by the insufficient capacity in many hospitals, this stream of studies aim at fully utilizing the limited capacities to achieve the highest utility for hospitals. Considerable work has been done on various aspects, e.g., surgical scheduling (Cardoen et al., 2010; Gupta, 2007; Liu et al., 2019; Naderi et al., 2021), admission control (Helm and Van Oyen, 2014; Kim et al., 2015; Samiedaluie et al., 2017), beds configuration decision (Izady and Mohamed, 2019; Pinker and Tezcan, 2013); beds allocation strategy (Best et al., 2015; Thompson et al., 2009), and discharge decision (Chan et al., 2012). In these studies, patients are usually divided into several classes according to their characteristics, then limited capacities will be allocated to corresponding patients. For example, Liu et al. (2019) study a two-stage scheduling problem in which two classes of patients (i.e., emergency patients and elective patients) compete for the OR (Operating rooms) time and the beds capacity. They study the problem by dynamically deciding the number of elective patients to admit in each period. Samiedaluie et al. (2017) consider multiple types of patients with different medical

characteristics who demand for the insufficient neurology wards. They focus on the problem of patient admissions from the emergency department (ED) , developing rules for the allocation of inpatient beds among multiple types of patients. They formulate an infinite-horizon average cost dynamic program (DP) with the objective of minimizing the average opportunity cost of waiting and transferring. Izady and Mohamed (2019) propose a new beds configuration strategy to alleviate the shortage of inpatient beds, which is named clustered overflow configuration. Patients are partitioned into multiple types according to their medical characteristics and are allocated certain number of dedicated wards. In addition to dedicated wards, those patients who cannot be accommodated can also be admitted into the overflow wards, which are shared by all types of patients. They propose two different formulations for partitioning and bed allocation: one minimizing the sum of average daily costs, and another minimizing the number of patients turned away.

Among those studies, capacity reservation for a particular class of patients is the issue mostly related to our study. Gerchak et al. (1996) consider emergency cases and elective cases, and reserve capacity for emergency patients. They propose a dynamic programing model to decide how many of the additional request for elective surgery to assign for each day under the advance scheduling setting. Huh et al. (2013) study a multi-source allocation problem with two classes of patients (emergency patients and elective patients), the optimal amount of capacity reserved for emergency patients was decided through a dynamic programing model. Our work differs from this stream in two aspects. First, previous studies examine the optimal decisions of a single type of medical resources, e.g., hospital diagnostic facilities, intensive care unit capacity, and staff. Our work instead considers two types of hospital beds. Specifically, we investigate isolation beds reservation given the limited capacity and the transformation possibilities between two types of beds during different phases of the COVID-19 pandemic. Second, previous studies assume patients arrive according to a stochastic process, e.g., a Poison process, that is exogenous and independent of the current state of the system, see, Samiedaluie et al. (2017), and Liu et al. (2019). In contrast, we consider two arrival settings

for the COVID-19 and elective patients, which are assumed to be time-dependent according to the epidemiology features of the COVID-19 pandemic. In our model, due to the existing of COVID-19 patients, the capacity planning has to consider infection among patients, and hospital beds have to be separated into two types, which is more challenging in analysis.

2.3 Dynamic multiproduct inventory control problem

Our work is also related to the dynamic multiproduct inventory control problem which has been well studied in operations management literatures (Aviv and Federgruen, 2001; DeCroix and Arreola-Risa, 1998; Veinott, 1965). This stream of literature addresses the problem of finding the optimal production and inventory policy for periodic review, in a multiproduct, infinite or finite horizon production inventory systems, where the production of products share a limited capacity in each period due to the constrained resource. At the beginning of each decision period, the remaining inventory level of each product is observed and the production quantity for each product is decided with limited shared resources. Within each period, demand for each product is satisfied based on the replenished inventory level of the product. Holding costs and shortage costs are incurred in each period, and the excess demand is backordered (DeCroix and Arreola-Risa, 1998). Similar to this stream of literature, our hospital beds allocation problem considers isolation beds and ordinary beds with shared capacity, which is analogous to the shared production capacity of two products in the multiproduct inventory control problem. Moreover, COVID-19 patients and non-COVID-19 patients are distinct groups demanding for the two type of beds. If there is no spare capacity, patients waiting in a queue to be admitted into public hospitals can be regarded as backlogged demand. However, this dynamic scheme may not perform perfectly under the scenario of allocating beds to cope with COVID-19. Hence, different from the dynamic multiproduct inventory control problem, we propose a new dynamic scheme to further addresses the special features of healthcare delivery in face of a pandemic. For example, in our work, the

bed allocation presents a time-dependent feature. In the outbreak phase of the pandemic, we only transform ordinary beds into isolation beds. In the post-peak phase of the pandemic, we transform isolation beds back into ordinary beds. While in the dynamic multiproduct inventory problem, the production of multiple products is not related to time, but is only limited by the total capacity.

Chapter 3

The Model

In this chapter, we describe the model setting of a discrete-time dynamic programming model and introduce notation used throughout the paper. We consider the allocation of the limited bed capacity, i.e., isolation and ordinary beds, to admit both COVID-19 and non-COVID-19 patients in a finite time horizon t, where $t = 1, \ldots, T$. The summary of notation used in our paper is attached in Appendix B.

Demands of COVID-19, emergency, and elective patients over a period t are nonnegative integer-valued random variables, denoted by ϵ_t, λ_t, and δ_t, respectively, where $\epsilon_t, \lambda_t, \delta_t \in \mathbb{Z}$ ($\mathbb{Z}$ denotes the set of nonnegative integers). We assume the demand of emergency (non-COVID-19) patients, λ_t, is an independent and identically distributed (i.i.d.) random variable for $t = 1, \ldots, T$. In reality, all inpatients stay in the public system for a random number of days before being discharged from the public system. We assume random fractions ξ_t^c and ξ_t^n of COVID-19 and non-COVID-19 inpatients exit the public system at the end of period t, respectively, where $\xi_t^j \in (0, 1)$ and ξ_t^j is i.i.d. for $j \in \{c, n\}$. This assumption is in line with the literature, for example, Liu et al. (2019).

We denote by x_t the number of available ordinary beds at the beginning of period t, after non-COVID-19 inpatients discharged from the public system; y_t the number of available isolation beds (created in a negative pressure environment) for serving COVID-19 patients. Let w_t be the number of elective patients on the waiting list to receive public healthcare. The triplet (x_t, y_t, w_t) is the vector of state variables observable before any decision made in period t.

3.1 The dynamic of flow

In our model, the arrival processes of COVID-19 and elective patients are assumed to be time-dependent according to the epidemiology features of COVID-19 pandemic. Both COVID-19 patients and emergency patients are admitted immediately, whereas elective patients should join a waiting list if ordinary beds are unavailable. Meanwhile, to relieve the congested public health care system, elective patients can be removed from the waiting list by being offered with a financial subsidy. The patient flow, bed transformation and allocation are illustrated in Figure 3.1.

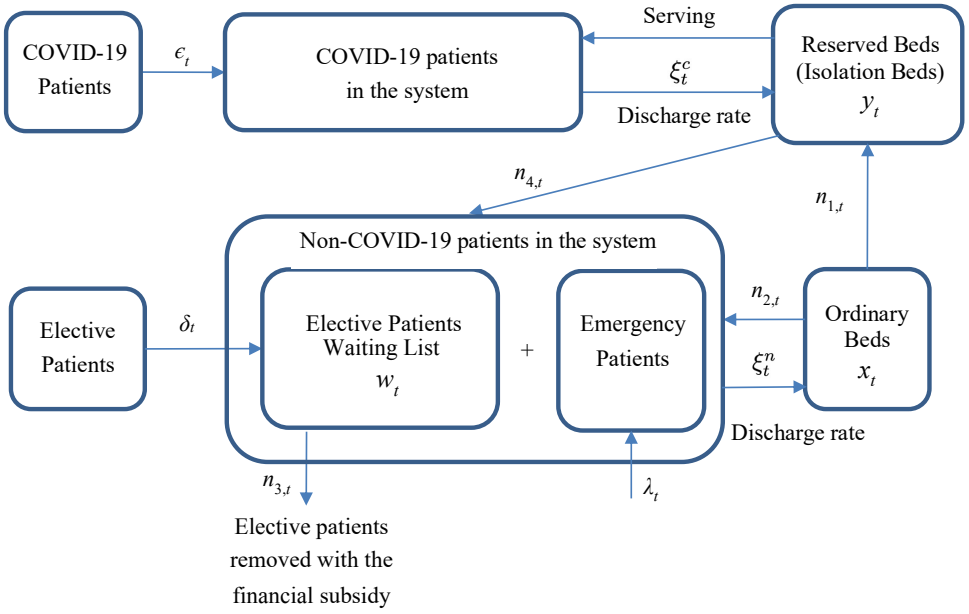

Figure 3.1: The illustration of patient flow and bed allocation

In each period, the decisions of the hospital manager are summarized as follows.

1. The hospital manager needs to *transform a number of ordinary beds to isolation beds*, $n_{1,t}$, to cope with COVID-19 patients, and *allocate the remaining ordinary beds*, $n_{2,t} = x_t - n_{1,t}$, to admit non-COVID-19 patients.

2. The hospital manager decides to *remove a number of elective patients on the waiting list*, $n_{3,t}$, to private hospitals, with a financial subsidy.

3. The hospital manager decides to *transform some of excessively reserved isolation beds back into ordinary beds*, $n_{4,t}$, to admit elective patients, in the post-peak phase.

In summary, there are three decision variables for the manager, $(n_{1,t}, n_{3,t}, n_{4,t})$. As it is not cost-efficient to reserve the isolation beds and reallocate the reserved isolation beds to admit non-COVID-19 patients simultaneously, $n_{1,t} \cdot n_{4,t} = 0$ holds in any period t.

The sequence of events in any period is summarized as follows.

1. At the beginning of period t, there are w_t elective patients on the waiting list and a waiting cost $M_t(w_t)$ is incurred. We assume $M_t(w_t)$ is an increasing convex function of the queue length, which is in line with the literature, for example, Mandelbaum and Stolyar (2004). The hospital manager obtains the number of available isolation and ordinary beds, y_t and x_t, respectively.

2. The hospital manager retrofits a number of ordinarily beds $n_{1,t}$ into isolation beds to admit COVID-19 patients. The unit retrofitting cost is denoted as c_b. The remaining capacity, $n_{2,t} = x_t - n_{1,t}$, can be allocated to admit non-COVID-19 patients, with emergency patients to be served first. The unit treatment cost for admitting a non-COVID-19 patient is denoted by c_n. If isolation beds are excessively reserved, the manager needs to determine how many of them, $n_{4,t}$, can be transferred back to ordinary beds to serve non-COVID-19 patients. The unit transferring cost is denoted as c_d.

3. The number of COVID-19 patients, emergency patients, and elective patients, represented by $(\epsilon_t,\ \lambda_t,\ \delta_t)$, is realized. The COVID-19 patients must be admitted immediately using the reserved isolation beds, y_t. And non-COVID-19 patients are allocated with the available ordinary beds, $n_{2,t}$. New incoming elective patients join the waiting list if there is no available beds for them. If too many non-COVID-19 patients are admitted, the limited isolation beds are unable to meet the demand of COVID-19 patients, which may cause big hazard for the society. We denote c_p as the unit penalty cost for the public hospital if COVID-19 patients cannot be served immediately. If too few non-COVID-19 patients are admitted, the waiting time of elective patients increases. Besides, if the excessively reserved isolation beds cannot be fully utilized, this is a waste of medical resource. There is an idling cost, c_o, inured for each reserved isolation bed. At the end of a period, we denote by

$F_t(y_t, w_t)$ the penalty/idling and waiting costs, where $F_t(y_t, w_t) = L_t(y_t) + M_t(w_t)$ and $L_t(y_t) = c_p y_t^- + c_o y_t^+$, where $x^- = min(0, x),\ x^+ = max(0, x)$. Note that we ignore the penalty cost for unmet emergency non-COVID-19 patients because we consider the penalty cost for an unmet COVID-19 patient is significantly higher than the penalty cost for an unmet emergency patient. There are normally multiple hospitals and if one public hospital is full, the emergency patient can be sent to another hospital via ambulance diversion (Allon et al. 2013).

4. If the waiting list is too long, the manager can offer an amount of financial subsidy, $c_s(n_{3,t})$, to some elective patients and refer them to private hospitals for treatment. This subsidy cost function is assumed to be increasing and convex in the number of subsidized elective patients $n_{3,t}$. The subsidy discipline is first-come-first-serve, which prioritizes patients based on their arrival times.

5. At the end of period t, the COVID-19 and non-COVID-19 patients are discharged from the public hospital with random proportions ξ_t^c and ξ_t^n, respectively, where $\xi_t^j \in (0, 1),\ j \in \{c, n\}$. Emptied beds are cleaned following cleaning and disinfecting procedures before admitting new patients.

The objective of the hospital manager is to identify a bed allocation rule and financial subsidy rule, to minimize the overall cost of the whole system.

3.2 Dynamic programming formulation

We now formulate the manager's decision problem as a Markov decision process (MDP). Based on the above analysis, we consider a finite planning horizon of T periods. The hospital manager reviews the system states and makes decisions in each period. Since the number of hospital beds and patients are integer, all the states in our model are assumed to be integer. The dynamic of states observed at the beginning of period t evolves into period $t + 1$ are formulated as follows.

The arrival of COVID-19 patients evolves following the equation:

$$\epsilon_{t+1} = \alpha_c \epsilon_t + u_t, \tag{3.1}$$

where the coefficient α_c indicates the increase of COVID-19 patients with time and $u_1, \cdots, u_T$ are independent random variables representing the noisy terms with mean

zero. As discussed in the introduction chapter, we denote by $\alpha_c > 1$ to describe the widespread human infection in the outbreak phase, and denote by $\alpha_c < 1$ to represent the post-peak phase with few human infections.

The number of COVID-19 patients in the system is updated by:

$$l_{t+1}^c = (1 - \xi_t^c) l_t^c + \epsilon_t. \tag{3.2}$$

The state of available isolation beds evolves following the equation:

$$y_{t+1} = y_t + \xi_t^c l_t^c - \epsilon_t + n_{1,t} - n_{4,t}. \tag{3.3}$$

The arrival of elective patients evolves following the equation:

$$\delta_{t+1} = \alpha_n \delta_t + k_t, \tag{3.4}$$

where the coefficient α_n indicates the time-dependent feature of the arrival process for elective patients, and $k_1, \cdots, k_T$ are independent random variables representing the noisy terms with mean zero. Fearing cross-infections in the public hospital, the arrival of elective patients presents a negative correlation to the arrival of COVID-19 patients. Thus, we denote by $\alpha_n < 1$ to show outbreak phase, and denote by $\alpha_n > 1$ to represent the post-peak phase.

The number of non-COVID-19 patients in the system is updated by:

$$l_{t+1}^n = (1 - \xi_t^n) l_t^n + \lambda_t + \min\{(n_{2,t} + n_{4,t} - \lambda_t)^+, w_t\}. \tag{3.5}$$

The number of available ordinary beds in the system is updated by:

$$x_{t+1} = x_t + \xi_t^n l_t^n - n_{1,t} - n_{2,t}. \tag{3.6}$$

The state of elective patients on the waiting list evolves following the equation:

$$w_{t+1} = (w_t + \delta_t - (n_{2,t} + n_{4,t} - \lambda_t)^+ - n_{3,t})^+. \tag{3.7}$$

Let $V_t(x_t, y_t, w_t, \mathbf{d}_t)$ be the total discounted cost incurred from periods t to T given the state variables $(x_t, y_t, w_t, \mathbf{d}_t)$, where $\mathbf{d}_t$ denotes the nonnegative vector

consisting of ϵ_t, λ_t, and δ_t. Thus, the Bellman equation can be established as follows:

$$V_t(x_t, y_t, w_t, \mathbf{d}_t) = \min_{\substack{n_{1,t}, n_{2,t}, \\ n_{3,t}, n_{4,t} \geqslant 0}} \{c_b n_{1,t} + c_n n_{2,t} + c_s(n_{3,t}) + c_d n_{4,t} + \mathbb{E}F_t(y_t - \epsilon_t + n_{1,t}^c - n_{4,t} + \xi_t^c l_t^c, w_t + \delta_t - (n_{2,t} + n_{4,t} - \lambda_t)^+ - n_{3,t}) + \gamma \mathbb{E}V_{t+1}(x_{t+1}, y_{t+1}, w_{t+1}, \mathbf{d}_{t+1})\} \tag{3.8}$$

where $n_{2,t} = x_t - n_{1,t}$, the expectation is taken over random variables, and the termination function $V_{T+1} = 0$. Note that on the right hand side of Equation (3.8), the first term in the brackets, $c_b n_{1,t}$, is the cost of retrofitting ordinary beds as isolation beds to cope with COVID-19 patients in period t; the second term, $c_n n_{2,t}$, is the treatment cost by admitting non-COVID-19 patients using ordinary beds in period t; the third term, $c_s(n_{3,t})$, is the financial subsidy offered to remove elective patients; the fourth term, $c_d n_{4,t}$, is the transferring cost by using the reserved isolation beds to serve non-COVID-19 patients; the fifth term, $\mathbb{E}F_t(\cdot, \cdot)$, is the expectation of penalty/idling and waiting costs; the last term, $\gamma \mathbb{E}V_{t+1}(\cdot, \cdot, \cdot, \cdot)$, is the minimum expected total discounted cost from period $t+1$ to T.

In the following chapters, we first analyze the structure properties of this problem and then find the optimal decisions on bed allocation and waiting list management.

Chapter 4

Structure Properties Analysis

In this chapter, we examine the characteristics of the total expected cost function in Equation (3.8). We adopt modularity analysis to obtain some structure properties of our decision problem. In our study, we jointly study the bed allocation issue and the waitlist management issue by considering three types of patients (i.e., COVID-19 patients, elective patients, emergency patients) and two types of beds (i.e., isolation beds, ordinary beds), which makes the modularity more challenging to analyze.

We first provide definitions and preliminaries analysis of modularity. Interested readers can refer to Topkis (1998), Murota (2005), and Li and Yu (2014) for comprehensive introduction of modularity analysis and its properties.

A partially ordered set $X = \{x', x''\}$ that contains the join $(x' \vee x'')$ and the meet $(x' \wedge x'')$ of each pair of its elements is a *lattice*. If X' is a subset of a lattice X and X' contains the join and the meet of each pair of elements of X', then, X' is a sublattice of X. Suppose that $f(x)$ is a real-valued function on a lattice X. A function $f(x)$ is *submodular* if $f(x') + f(x'') \geqslant f(x' \vee x'') + f(x' \wedge x'')$. Similarly, if $-f(x)$ is submodular, then $f(x)$ is supermodular. If $f(x)$ and $g(x)$ are supermodular on X, then $f(x) + g(x)$ is supermodular on X (Topkis 1998).

Let $X \in \mathbb{R}^n$ and $Y \in \mathbb{R}$ be polyhedral satisfying that the nonzero components of $\mathbf{a}_i$ are either consecutive ones or consecutive negative ones, where $X = \{\boldsymbol{x} \in \mathbb{R}^n | \mathbf{a}_i \cdot \boldsymbol{x} \geqslant b_i,\ i = 1, \ldots, m\}$, $b_i \in \mathbb{R}$. A function f: $X \to \mathbb{R}$ is multimodular if $g(\boldsymbol{x}, y) = h(x_1 - y, x_2 - x_1, \ldots, x_n - x_{n-1})$ is submodular on a lattice $S = \{(\boldsymbol{x}, y) \in \mathbb{R}^n \times \mathbb{R} |\ y \in Y,\ (x_1 - y, x_2 - x_1, \ldots, x_n - x_{n-1}) \in X\}$. (i) Suppose that $g(\boldsymbol{v})$ is multimodular, $g(\boldsymbol{v})$ has increasing differences; (ii) if $g(\boldsymbol{v}, d)$ is multimodular in $\boldsymbol{v}$,

for any given d and D is a random variable, then $\mathbb{E}g(\boldsymbol{v}, D)$ is multimodular in $\boldsymbol{v}$ (Li and Yu, 2014).

Proposition 1. *For $t = 1, \cdots, T$, $V_t(x_t, y_t, w_t)$ is increasing in w_t and jointly convex in (y_t, w_t).*

Proposition 1 shows that the optimal expected operations cost is greater if elective patients cannot be admitted to the public hospital but only waiting for treatment. Also, the convexity of $V_t(x_t, y_t, w_t)$ in y_t implies that increasing marginal effect of reserving one more isolation bed. Besides, the convexity of $V_t(x_t, y_t, w_t)$ in w_t implies increasing marginal effect associated with adding one more elective patient to the waiting list. The joint convexity implies that there exists a state status of (y_t, w_t) corresponding to a minimal cost-to-go function V_t.

Recall that we have three decision variables for each period. To obtain some structure properties of the optimal decisions, we take a sequential optimization approach on finding the optimal solutions of the three decision variables. First, the hospital manager determines $n_{1,t}$, namely the number of isolation beds reserved for admitting COVID-19 patients. Second, the hospital manager determines $n_{3,t}$, the number of elective patients subsidized from the waiting list and $n_{4,t}$, the number of excessively reserved isolation beds transferred back into ordinary bed for admitting non-COVID-19 patients. Following the backward induction, We first find the optimal subsidization and reallocation decisions given the reserved isolation beds in section 4.1; and then we find the optimal number of isolation beds reserved for COVID-19 patients in section 4.2.

4.1 Managing non-COVID-19 patients

Here, the hospital manager faces the decisions of managing elective patients given the reserved isolation beds. He can manage the 'demand' by subsidizing $n_{3,t}$ patients and move them away from the waiting list. He can also manage the 'supply' by increasing the supply of the ordinary beds through transferring $n_{4,t}$ isolation beds into ordinary beds. The optimal solutions can be recursively solved through the

following Bellman equation, for $t = 1, \cdots, T$,

$$H_t^*(x_t, y_t, w_t, \mathbf{d}_t) = \min_{\substack{0 \leqslant n_{3,t} \leqslant w_t \\ 0 \leqslant n_{4,t} \leqslant y_t}} \{c_s(n_{3,t}) + c_d n_{4,t} + \mathbb{E}F_t(y_t - \epsilon_t - n_{4,t} + \xi_t^c l_t^c, w_t + \delta_t - n_{3,t} - n_{4,t}) + \gamma \mathbb{E}V_{t+1}(x_{t+1}, y_{t+1}, w_{t+1}, \mathbf{d}_t)\}, \tag{4.1}$$

Proposition 2. *If $F_t(y_t, w_t)$ is multimodular in (y_t, w_t) and V_{t+1} is multimodular in (y_{t+1}, w_{t+1}), we have $H_t^*(x_t, y_t, w_t)$ is multimodular in (y_t, w_t).*

Proposition 2 shows that the property of multimodularity can be preserved under the minimization operation of two decision variables, $n_{3,t}$ and $n_{4,t}$, which also assures the increasing monotonicity of the optimal solutions $(n_{3,t}^*, n_{4,t}^*)$ in state variables y_t and w_t (Topkis 1998, Gong et al. 2014). This is quite reasonable: if the number of reserved isolation beds, y_t, is larger, then the manager shall subsidize more patients to move them to the private system (i.e., increasing $n_{3,t}^*$) and also it is more likely to transfer some isolation beds back into ordinary beds (i.e., increasing $n_{4,t}^*$). Also, if the waiting list, w_t, is longer, it is better for the manager to move away some patients from the waiting list and transfer some isolation beds back into ordinary beds.

The multimodularity property implies the joint convexity of H_t^* in (y_t, w_t). It also implies the diagonal dominance, i.e., $\frac{\partial^2 H_t^*}{\partial y_t^2} \geq \frac{\partial^2 H_t^*}{\partial y_t \partial w_t} \geq 0$ and $\frac{\partial^2 H_t^*}{\partial w_t^2} \geq \frac{\partial^2 H_t^*}{\partial y_t \partial w_t} \geq 0$. This shows that marginal impact on H_t^* due to a unit increase on y_t is more sensitive to the change in itself than the change in w_t and similar explanation holds for the change in w_t. We shall investigate the influence of y_t and w_t on the dynamic allocation policy, which is discussed in section 4.4.

4.2 Reserving isolation beds for COVID-19 patients

After obtaining the optimal demand- and supply-side decisions on managing elective patients, we now consider the optimal bed reservation decision for COVID-19 patients.

At the beginning of period t, non-COVID-19 patients are discharged from the public hospital system and the total number of ordinary beds, x_t, can then be

obtained. The hospital manager needs to determine $n_{1,t}$, namely the number of ordinary beds to be retrofitted into isolation beds, to cope with COVID-19 patients. The remaining ordinary beds, $n_{2,t} = (x_t - n_{1,t})^+$, will be first allocated for emergency non-COVID-19 patients and only the extra ones can be used to admit elective patients.

The optimality equation for $t = 1, \cdots, T$ can be written as follows,

$$V_t(x_t, y_t, w_t, \mathbf{d}_t) = \min_{0 \leqslant n_{1,t} \leqslant x_t - \lambda_t} \{c_b n_{1,t} + c_n(x_t - n_{1,t}) + H_t^*(x_t, y_t + n_{1,t}, w_t - x_t + n_{1,t}, \mathbf{d}_t)\}. \tag{4.2}$$

We have the following property of the total expected cost function V_t.

Proposition 3. *If $H_t^*(x_t, y_t, w_t)$ is multimodular in (y_t, w_t), we have V_t is multimodular in (y_t, w_t) for any $x_t \geqslant 0$.*

Propositions 2 and 3 yield the following corollary.

Corollary 1. *For $t = 1, \cdots, T$, V_t is multimodular in (y_t, w_t) for any $x_t \geqslant 0$.*

This result establishes the existence of global minimizers of (y_t, w_t) for V_t. The multimodularity of V_t shows that the marginal change on the total cost of reserving one more isolation bed is increasing in the waiting list w_t. This implies that a longer waiting list of elective patients makes it harder for the manager to reserve more isolation beds. In addition, when facing too many reserved isolation beds, the hospital manager can manage the 'supply' side by transforming isolation beds back into ordinary beds to admit more elective patients.

4.3 Monotonicity of optimal decisions

With the above properties of the Bellman equation, we analyze the detailed characterizations of the optimal decisions. Since it is not cost-efficient to reserve isolation beds and reallocate the reserved isolation beds simultaneously, $n_{1,t} \cdot n_{4,t} = 0$. To achieve efficient bed allocation and patient subsidization with the minimum cost, we then analyze the impacts of number of elective patients on waiting list, w_t, and the number of available isolation beds, y_t, on the manager's optimal allocation and subsidization decisions. The results are summarized in the following proposition.

Proposition 4. *For $t = 1, \cdots, T$, suppose $(x_t, y_t, w_t, \mathbf{d}_t)$ is the initial state in period t,*

(i) the optimal number of ordinary beds transformed into isolation beds, $n^*_{1,t}$, is decreasing in y_t and w_t;

(ii) the optimal number of subsidized elective patients, $n^*_{3,t}$, is increasing in y_t and w_t;

(iii) the optimal number of excessively reserved isolation beds transformed into ordinary beds, $n^*_{4,t}$, is increasing in y_t and w_t.

Proposition 4 shows that the optimal decisions in beds allocation and waiting list management behave the monotonicity of state variables. The decreasing monotonicity of $n^*_{1,t}$ in y_t implies that less isolation beds should be reserved when there are more available isolation beds; the decreasing monotonicity of $n^*_{1,t}$ in w_t indicates that, if a waiting list is longer, it's harder to reserve more isolation beds. The part (ii) and (iii) show that if the number of reserved isolation beds, y_t, is larger and the waiting list, w_t, is longer, it is better to transfer more isolation beds back into ordinary beds and also to use subsidy to remove more elective patients from the waiting list.

In the following section, we will illustrate the bed allocation and patient subsidization decisions more specifically by analyzing the marginal cost of two decisions.

4.4 Decision analysis and implications

By characterizing the properties of the objective function, we now present a more straightforward way to show the optimal decisions. Facing the limited bed capacity and the COVID-19 pandemic , the hospital manager should manage the situation from both demand- and supply-side. In the outbreak phase, on the supply side, the manager can retrofit some ordinary beds into isolation beds facing the surging demand of COVID-19 patients. On the demand side, the manager can remove some elective patients from the waiting list by subsidy. In the post-peak phase, on the supply side, the hospital manager can transfer excessively reserved isolation beds into ordinary beds. While on the demand side, the manager can also use subsidy to

remove some elective patients from the waiting list if the waiting list is too long. For both outbreak phase and post-peak phases, we conduct comparison analysis of the marginal cost incurred by two decisions (allocating beds or moving patients away using subsidy) to investigate the optimal operational rule.

By Proposition 3, we learn that these decisions can be found recursively by solving the following equation, for $t = 1, \cdots, T$,

$$V_t(x_t, y_t, w_t, \mathbf{d}_t) = \min_{0 \leqslant n_{1,t} \leqslant x_t - \lambda_t} \{c_b n_{1,t} + c_n(x_t - n_{1,t}) + H_t^*(x_t, y_t + n_{1,t}, w_t - x_t + n_{1,t}, \mathbf{d}_t)\}, \tag{4.3}$$

where

$$H_t^*(x_t, y_t, w_t, \mathbf{d}_t) = \min_{n_{3,t}, n_{4,t} \geqslant 0} \{c_s(n_{3,t}) + c_d n_{4,t} + h_t(y_t, w_t, x_t)\}$$

and

$$h_t(y_t, w_t, x_t) = \mathbb{E}F_t(y_t - \epsilon_t - n_{4,t} + \xi_t l_t, w_t + \delta_t - n_{3,t} - n_{4,t}) + \gamma \mathbb{E}V_{t+1}(x_{t+1}, y_{t+1}, w_{t+1}, \mathbf{d}_{t+1}).$$

We discuss the marginal costs incurred by of bed allocation and subsidization decisions. For the *outbreak phase*, our decisions variables are $n_{1,t}$ (or $n_{2,t}$) and $n_{3,t}$. Given state variables (x_t, y_t, w_t) in period t, the marginal cost of retrofitting an ordinary bed into an isolation bed to cope with COVID-19 patients is $c_b + \partial h_t(x_t, y_t, w_t)/\partial y_t$. If $c_b + \partial h_t(x_t, y_t, w_t)/\partial y_t < 0$, it indicates that transferring one more ordinary bed into isolation bed is cost saving. According to Lemma 1 (Li & Yu, 2014), $c_b + \partial h_t(x_t, y_t, w_t)/\partial y_t$ is increasing in y_t and w_t. The marginal cost of subsidizing an elective patients from the waiting list is $c_s - \partial h_t(x_t, y_t, w_t)/\partial w_t$. If $c_s - \partial h_t(x_t, y_t, w_t)/\partial w_t < 0$, it indicates that subsiding one more elective patient from the waiting list is cost saving (i.e., reducing the length of waiting list by removing one more elective patient). To analyze the difference between the marginal cost of reserving isolation beds and subsidizing elective patients, we examine $c_b - c_s + \partial h_t(x_t, y_t, w_t)/\partial y_t + \partial h_t(x_t, y_t, w_t)/\partial w_t$. If $c_b - c_s + \partial h_t(x_t, y_t, w_t)/\partial y_t + \partial h_t(x_t, y_t, w_t)/\partial w_t < 0$, it indicates that retrofitting an ordinary bed to an isolation bed from the supply side can achieve a better cost saving effect than the strategy of subsiding an elective patient away. This cost difference is increasing in y_t and w_t. Next, we define three curves in (y_t, w_t) plane to illustrate the optimization process. We use superscript 'o' to denote the marginal costs incurred in the outbreak phase.

$$Y_t^o(w_t, x_t) = \sup\{y_t : c_b + \partial h_t(x_t, y_t, w_t)/\partial y_t < 0\},$$

$$W_t(y_t, x_t) = \inf\{w_t : c_s - \partial h_t(x_t, y_t, w_t)/\partial w_t < 0\},$$

$$Z_t^o(w_t, x_t) = \sup\{y_t : c_b - c_s + \partial h_t(x_t, y_t, w_t)/\partial y_t + \partial h_t(x_t, y_t, w_t)/\partial w_t < 0\}.$$

Here, $Y_t^o(w_t, x_t)$ is the largest y_t which satisfies the requirement that increasing isolation beds is cost effective in the outbreak phase; $W_t(y_t, x_t)$ is the smallest w_t which satisfies the condition that reducing the length of waiting list by subsidy is cost effective; $Z_t^o(w_t, x_t)$ is the largest y_t which satisfies the condition that retrofitting an ordinary bed to an isolation bed can achieve a better cost saving effect than the strategy of subsiding an elective patient away.

We adopt the same logic to analyze decisions made in the *post-peak phase*. For this phase, our decisions variables are $n_{3,t}$ and $n_{4,t}$. Given state variables (x_t, y_t, w_t) in period t, the marginal cost of transferring an isolation bed to an ordinary bed to admit a non-COVID-19 patient is $c_n + c_b - \partial h_t(x_t, y_t, w_t)/\partial y_t - \partial h_t(x_t, y_t, w_t)/\partial w_t$. If $c_n + c_b - \partial h_t(x_t, y_t, w_t)/\partial y_t - \partial h_t(x_t, y_t, w_t)/\partial w_t < 0$, it indicates that reducing y_t (i.e., transferring an isolation bed into an ordinary bed) is cost saving. The marginal cost of subsidizing an elective patients from the waiting list is $c_s - \partial h_t(x_t, y_t, w_t)/\partial w_t$. If $c_s - \partial h_t(x_t, y_t, w_t)/\partial w_t < 0$, it indicates that subsiding one more elective patient from the waiting list is cost saving (i.e., reducing the length of waiting list by removing one more elective patient away). The cost difference of these two decisions is $c_n + c_b - c_s - \partial h_t(x_t, y_t, w_t)/\partial y_t$, which is decreasing in y_t. Similarly, we define three curves in (y_t, w_t) plane to illustrate the optimization process. We use superscript 'p' to denote the marginal costs incurred in the post-peak phase.

$$Y_t^p(w_t, x_t) = \inf\{y_t : c_n + c_b - \partial h_t(x_t, y_t, w_t)/\partial y_t - \partial h_t(x_t, y_t, w_t)/\partial w_t < 0\},$$

$$W_t(y_t, x_t) = \inf\{w_t : c_s - \partial h_t(x_t, y_t, w_t)/\partial w_t < 0\},$$

$$Z_t^p(w_t, x_t) = \inf\{y_t : c_n + c_b - c_s - \partial h_t(x_t, y_t, w_t)/\partial y_t < 0\}.$$

Here, similarly, $Y_t^p(w_t, x_t)$ is the smallest y_t which satisfies the requirement that transferring isolation beds to ordinary beds is cost effective in the outbreak phase; $W_t(y_t, x_t)$ is the smallest w_t which satisfies the condition that reducing the length of waiting list by subsidy is cost effective; $Z_t^o(w_t, x_t)$ is the smallest y_t which satisfies

the condition that transferring an isolation bed into an ordinary bed can achieve a better cost saving effect than the strategy of subsiding an elective patient away.

Let $(\hat{y}_t, \hat{w}_t)$ ($(\tilde{y}_t, \tilde{w}_t)$, resp.) denote the minimizer of V_t in the outbreak phase (the post-peak phase, resp.). Since the analyze of decisions in the post-peak phase is similar to the ones in the outbreak phase, we only show the main results in the outbreak phase. As the pandemic progresses, the hospital manager can make dynamic adjustments to minimize the total expected cost.

Proposition 5. *In period t, we have*

(i) As $h_t(x_t, y_t, w_t)$ is multimodular in (y_t, w_t) for $x_t \geqslant 0$, then $Y_t^o(w_t, x_t)$ and $Z_t^o(w_t, x_t)$ are decreasing in the number of elective patients on the waiting list;

(ii) If $y_t \leqslant \hat{y}_t$, $Z_t^o(w_t, x_t)$ is located to the left of $Y_t^o(w_t, x_t)$ and above $W_t(y_t, x_t)$; if $y_t > \hat{y}_t$, the curve $Z_t^o(w_t, x_t)$ is located under $W_t(y_t, x_t)$ and to the right of $Y_t^o(y_t, x_t)$.

Proposition 5(i) shows that, in outbreak phase, the maximum number of isolation beds subjects to the cost saving condition is sensitive to the number of elective patients on the waiting list. That is, when is waiting list is long, to be cost effective, we cannot retrofit a large number of ordinary beds into isolation beds. While if there is less elective patients on the waiting list, we can retrofit more ordinary beds into isolation beds, which is still cost effective.

Furthermore, Proportion 5 reveals the monotone property of three curves in the (y_t, w_t) plane which is presented in Figure 4.1. Because both y_t and w_t are non-negative, these curves locate on the first quadrant of the (y_t, w_t) plane. We divide the first quadrant of the (y_t, w_t) plane into six regions. For a state located in different regions, we have the following decision guidelines, which can help the hospital managers make decisions.

In region I , *hospital manager should retrofit more ordinary beds into isolation beds to cope with COVID-19 patients.* For any point in this area, e.g., $p_1(y_1, w_1)$, we draw a horizontal line that intersects with $Y_t^o(w_t, x_t)$. The intersection point is denoted by (y_1^o, w_1). The hospital manager should retrofit $y_1^o - y_1$ ordinary beds into

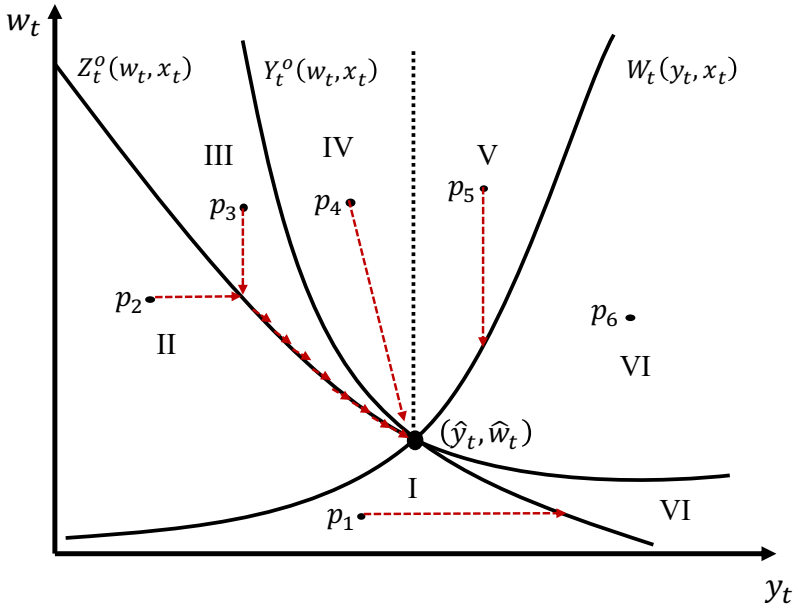

Figure 4.1: The changes of marginal costs on w_t and y_t

isolation beds to achieve a cost saving effect. Because $p_1(y_1, w_1)$ is located under the line $W_t(y_t, x_t)$, it's not cost saving to further subsidy patients away.

In region II, *hospital manager should primarily retrofit more ordinary beds into isolation beds and then follow the line of* $Z_t^o(w_t, x_t)$ *until it reaches the state* $(\hat{y}_t, \hat{w}_t)$. For example, for a state $p_2(y_2, w_2)$, since it is located at the left side of $Y_t^o(w_t, x_t)$ and above the line of $W_t(y_t, x_t)$, increasing y_t and reducing w_t are both cost saving strategies. Furthermore, $p_2(y_2, w_2)$ is located at the left side of $Z_t^o(w_t, x_t)$, thus, increasing isolation beds can achieve a better cost saving effect than reducing the length of waiting list. Until it reaches the line $Z_t^o(w_t, x_t)$, it can further follow the line $Z_t^o(w_t, x_t)$ downwards until it reaches $(\hat{y}_t, \hat{w}_t)$. Note that we only provide a possible way to reach the optimal state $(\hat{y}_t, \hat{w}_t)$. Other ways also work only if it can reach the state $(\hat{y}_t, \hat{w}_t)$. It holds for other regions as well.

In region III, *hospital manager should primarily subsidize patients away and then follow the line of* $Z_t^o(w_t, x_t)$ *until it reaches the state* $(\hat{y}_t, \hat{w}_t)$. For example, for a state $p_3(y_3, w_3)$, since it is located at the left side of $Y_t^o(w_t, x_t)$ and above the line of $W_t(y_t, x_t)$, increasing y_t and reducing w_t are both cost saving strategies. Furthermore, $p_3(y_3, w_3)$ is located at the right side of $Z_t^o(w_t, x_t)$, thus, reducing the length of waiting list can achieve a better cost saving effect than increasing isolation

beds. Until it reaches the line $Z_t^o(w_t, x_t)$, it can further follow the line $Z_t^o(w_t, x_t)$ downwards until it reaches $(\hat{y}_t, \hat{w}_t)$.

In region IV, *hospital manager should retrofit more ordinary beds into isolation beds and subsidize patients away at the same time until it reaches the state* $(\hat{y}_t, \hat{w}_t)$. For any point in this region, it can not reach the line of $Z_t^o(w_t, x_t)$, thus, the hospital manage can adopt any optimization process to reach the state $Z_t^o(w_t, x_t)$.

In region V, *hospital manager should subsidize patients away to reduce the length of waiting list.* For any point in this area, e.g., $p_5(y_5, w_5)$, we draw a vertical line that intersects with $W_t(y_t, x_t)$. The intersection point is denoted by (y_5, w_5^o). The hospital manager should subsidize t $w_5 - w_5^o$ patients and move them from the waiting list to achieve a cost saving effect. Because $p_5(y_5, w_5)$ is located at the right side of the line $Y_t^o(w_t, x_t)$, it's not cost saving to further creating more isolation beds. Furthermore, in the outbreak phase, hospital managers also cannot transfer excessive isolation beds back into ordinary beds, thus, they should just keep those isolation beds.

In region VI, *hospital manager should keep the current state and do nothing.* For any point in this area, e.g., $p_6(y_6, w_6)$, it is located at the right side of the line $Y_t^o(w_t, x_t)$, it's not cost saving to further creating more isolation beds. In the outbreak phase, hospital managers also cannot transfer excessive isolation beds back into ordinary beds, thus, they should just keep those isolation beds. Also, $p_6(y_6, w_6)$ is located below the line $W_t(y_t, x_t)$, it is not cost saving to further reduce the length of waiting list by subsidy.

In summary, applying Proportion 5 we conclude that the hospital manager should dynamically adjust the number of isolation beds by retrofitting and manage the length of waiting list by subsidy to reduce the total expected cost.

Chapter 5

Numerical Studies

In this chapter, we conduct numerical experiments to illustrate theoretical outcomes and to investigate the performance of the dynamic allocation rule. In reality, the COVID-19 situation varies in different countries and regions. (Han et al. 2020). In some countries and regions, the situation is very serious, where the transmission of virus is so fast that large amount of people get infected in a short time (e.g. Wuhan in China). Under such scenario, high-level pandemic alert is implemented in response to the serious outbreak of COVID-19. While in some countries and regions, the number of confirmed cases is increasing at a low speed, under which scenario, a low-level pandemic alert is enough to cope with the pandemic.

In numerical experiments, we classify the pandemic alerts of different countries and regions into three categories, namely weak, medium, and serious. Under each scenario, we show the proposed dynamic allocation rule numerically. Then, to validate the effectiveness of our dynamic allocation rule, we compare the expected operation cost between the static policy (i.e., keeping a fixed number of isolation beds) and our proposed dynamic allocation rule. We are particularly interested in under which scenario the dynamic allocation rule works the best when compared with static policy. In addition, to examine the effect of subsidy schemes, we compare the experiment outcomes with and without subsidy schemes. In the end, we conduct the robust analysis, investigating the impacts of hospital beds capacity and the unit waiting cost on the total operation cost.

Table 5.1: Numerical Settings

Parameters	Value
Total number of decision periods T	16
Total number of inpatient beds t_b	120
The initial number of isolation beds y_1	20
The initial number of elective patients on the waiting list w_1	50
The arrival rate of COVID-19 patients ϵ_t	$\alpha_c \in \{0.8,\ 1.4,\ 1.7,\ 2\}, u_t \sim N(0,2)$
The arrival rate of elective patients δ_t	$\alpha_n = 0.85, k_t \sim N(0,2)$
The arrival rate of emergency patients λ_t	Following a normal distribution $N(10,2)$
The number of COVID-19 patients being discharged from the system ξ_t^c	Uniformly distributed over $[0.4, 0.5]$
The number of non-COVID-19 patients being discharged from the system ξ_t^n	Uniformly distributed over $[0.4, 0.5]$
The unit idling cost of an isolation bed c_o	3
The unit cost of retrofitting an ordinary bed into an isolation bed c_b	10
The unit cost for admitting a non-COVID-19 patient c_n	1
The unit penalty cost if a COVID-19 patient is not timely admitted c_p	500
The unit penalty cost if a COVID-19 patient is not timely admitted c_d	2
The unit waiting cost c_w	1
The financial subsidy $c_s(n_{3,t}) = a_1 n_{3,t}^2 + a_2 n_{3,t}$	$(a_1, a_2) \in \{(1,19),(2,18),(3,17),(4,16)\}$
The discount factor γ	0.9

Note: The normal distribution is referred to as $N(\mu, \sigma)$ with mean μ and variance σ.

The parameter settings are summarized in Table 5.1. In numerical experiments, we assume the number of time periods is 16, and each period lasts for one week. We make such an assumption since the first wave and second wave of COVID-19 pandemic in Hong Kong both lasted about 3 months. In practice, the duration for the outbreak of COVID-19 in certain region can be estimated by experts.

5.1 Numerical outcomes of three examples

In this section, we use three examples to present the optimal outcomes of our dynamic beds allocation and patients subsidization rule under three pandemic alerts. Recall that we consider two phases (i.e., outbreak phase and post-peak phase) in the progression of COVID-19 pandemic. To describe different level of pandemic alerts, in the outbreak phase, we assume $\alpha_c = 1.4$ for the weak pandemic alert; $\alpha_c = 1.7$ for the medium pandemic alert; $\alpha_c = 2$ for the serious pandemic alert. Regarding the post-peak phase, we assume $\alpha_c = 0.8$ for all the three scenarios. The results of the simulation under three scenarios are summarized as follows:

5.1.1 Scenario under a weak pandemic alert

Figure 5.1 shows the beds allocation results under a weak pandemic alert ($\alpha_c = 1.4$).

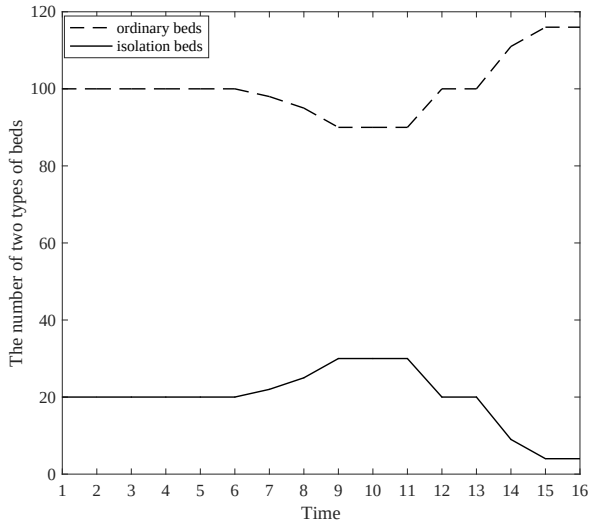

Figure 5.1: Bed allocation under a weak pandemic alert ($\alpha_c = 1.4$)

As we can see from Figure 5.1, when the outbreak of pandemic is mild, holding a certain number of isolation beds is able to cope with emerging COVID-19 patients in the first few periods. With the increase of COVID-19 patients, a small amount of ordinary beds need to be transformed into isolation beds. Specifically, total number of 10 ordinary beds are retrofitted into the isolation beds given the total number of beds is 120. After the peak phase, given the increasing demand of elective patients and more vacated isolation beds, a total number of 29 isolation beds are transformed back into ordinary beds to admit the elective patients.

Figure 5.2 shows how the length of waitlist changes with the progression of pandemic. At the beginning, the length of waitlist is quite long (which is 50 given the total waitlist capacity is 70) in a public hospital, which incurs a high waiting cost. Thanks to the subsidization scheme, some elective patients are given an amount of subsidy and go to private hospital for a quicker healthcare service, which leads to the waitlist gets shorter and shorter. In the later periods, since there are more ordinary beds to admit elective patients, the number of elective patients on the waitlist is almost zero.

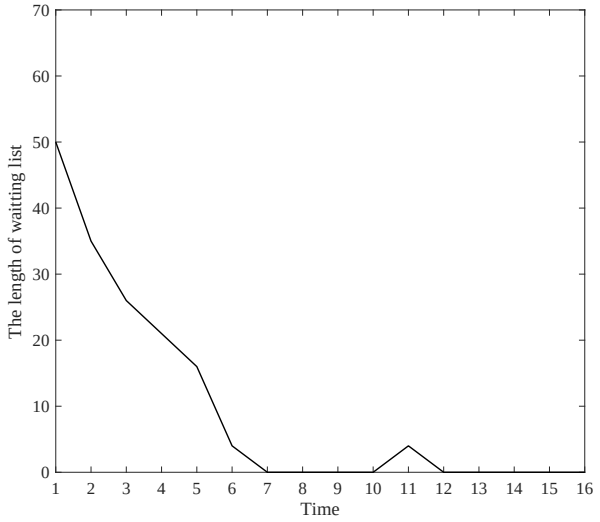

Figure 5.2: The length of waitlist at each period ($\alpha_c = 1.4$)

5.1.2 Scenario under a medium pandemic alert

Figure 5.3 shows bed allocation results when a region is facing a medium pandemic alert ($\alpha_c = 1.7$). In the first 4 periods, the initial reserved isolation beds can cope with those newly arrived COVID-19 patients. Thus, there is no significant change in terms of the number of two types of beds. However, with more confirmed COVID-19 cases arising in the following periods, more ordinary beds are transformed into isolation beds to admit those COVID-19 patients. Since period 11, the excessively reserved isolation beds are gradually transformed back into ordinary beds to admit non-COVID patients. Also, we note that the number of reserved isolation beds does not decline immediately after the peak phase. The reason is that, to avoid the potential risk of the second wave of the COVID-19 pandemic, it is necessary to ensure the adequate number of isolation beds within several time periods. In this case, 80 isolation beds are used to cope with the potential second wave of the COVID-19 pandemic.

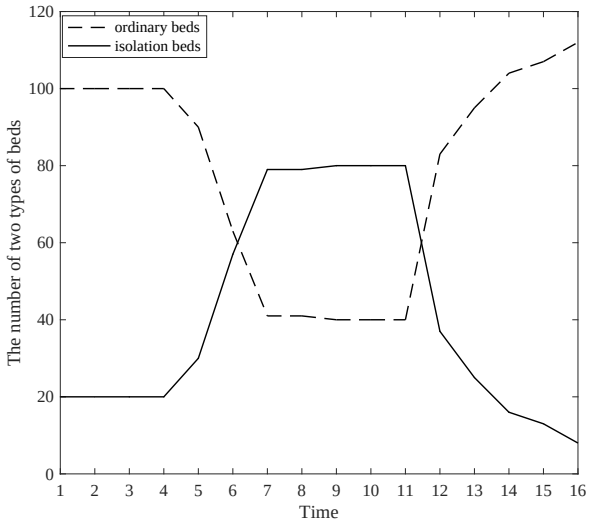

Figure 5.3: Bed allocation under a weak pandemic alert ($\alpha_c = 1.7$)

Figure 5.4 shows the change of the length of waitlist when a region is facing the medium pandemic situation. The length of waiting list first decreases, due to the subsidy scheme and the decreasing number of arrivals from elective patients. In the following periods, the length increases again. The reason may be that the number of

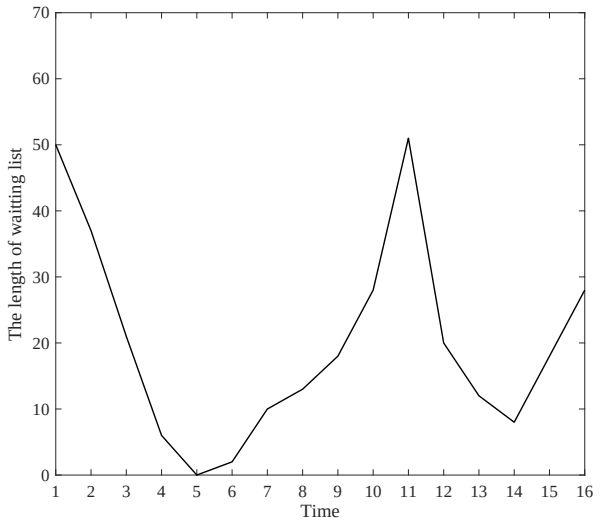

Figure 5.4: The length of waitlist at each period ($\alpha_c = 1.7$)

ordinary beds gets smaller, and fewer elective patients are admitted, which lengthens the waiting list. After some periods, the waiting list gets shorter. This can be caused by the increasing number of ordinary beds as shown in Figure 5.3. In periods 14-16, the waiting list gets slightly longer, due to the increasing arrivals of elective patients in the post-peak phase.

5.1.3 Scenario under a serious pandemic alert

Figure 5.5 presents optimal bed allocation results when facing a serious pandemic alert. Compared to the medium case, this scenario shows a different trend, especially during the outbreak phase. As we can see from Figure 5.5, there exists a sharp increase since the 4th period. This is because that the COVID-19 appears highly contagious in a short time. Therefore, the hospital manager must take immediate action, i.e., creating more isolation beds by retrofitting the ordinary beds, to deal with a surging number of COVID-19 patients. This experiment result shows that 96 ordinary beds are retrofitted as isolation beds. Similarly, a total of 116 isolation beds are reserved within three periods to avoid the potential risk of second wave. When facing a stabilized situation in the post-peak phase, the excessively reserved reservation beds are transformed back to admit the elective patients.

Figure 5.6 shows the change of the length of waitlist when a region is facing a

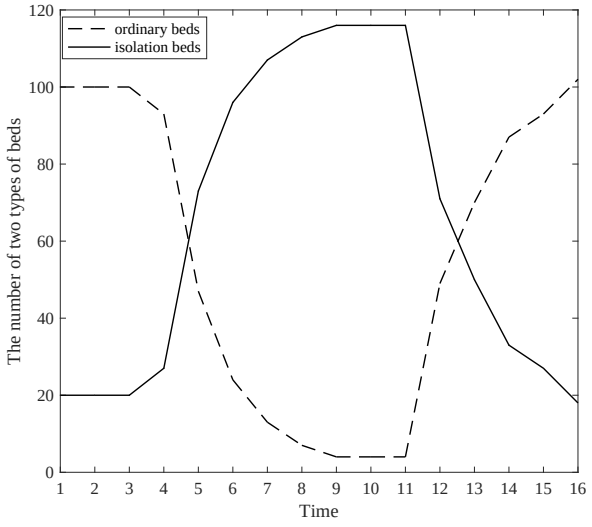

Figure 5.5: Bed allocation under a serious pandemic alert ($\alpha_c = 2.0$)

serious pandemic situation. The trend of the change is similar to the one under medium alert. What's the difference is that the overall length under a serious alert is longer than that under a medium alert. For instance, at period 11, the length of waitlist under a serious scenario is over 60, which is longer than the peak point which is 50 under the scenario with a medium alert.

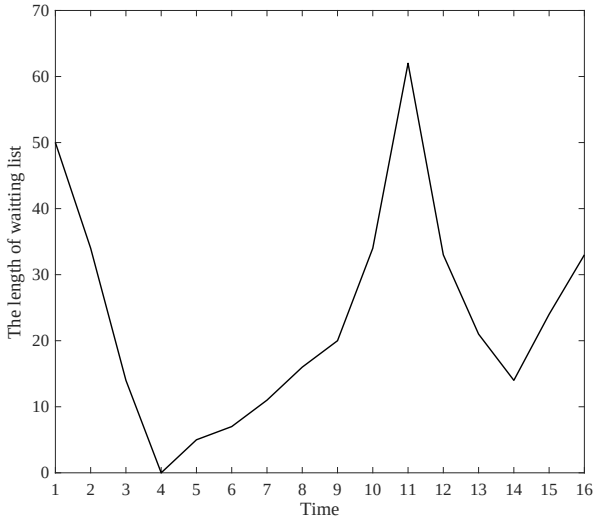

Figure 5.6: The length of waitlist at each period ($\alpha_c = 2.0$)

5.2 Performance comparison with a static bed allocation policy

In this experiment, we consider a benchmark case with a fixed number of isolation beds over the whole planning periods and with subsidy scheme provided. We define a ratio V_c^d/V_c^s, where V_c^s (V_c^d, resp.) is the expected cost incurred using the static policy (the dynamic scheduling rule, resp.). The smaller the cost ratio, the better our dynamic allocation rule performs, when compared with the static rule. Under each scenario (i.e., $\alpha_c = 1.4, 1.7, 2.0$), we compute the cost ratio with the number of fixed isolation beds changing from 10 to 120. The numerical outcomes are summarized in Table 5.2. From the Table 5.2 we can see that all the cost ratios are less then 1, which demonstrates that our dynamic approach always predominates over the static policy.

Table 5.2: The cost ratio between the static policy and the dynamic scheduling rule

Fixed Isolation Beds	10	20	30	40	50	60	70	80	90	100	110	120
$\alpha_c = 1.4$	0.284	<u>0.814</u>	0.725	0.575	0.439	0.343	0.272	0.218	0.179	0.160	0.157	0.156
$\alpha_c = 1.7$	0.269	0.336	0.418	0.497	0.583	0.653	<u>0.665</u>	0.628	0.525	0.471	0.460	0.458
$\alpha_c = 2.0$	0.297	0.342	0.387	0.436	0.485	0.530	0.579	0.621	0.664	0.701	0.776	<u>0.872</u>

What's more, if we observe each row in Table 5.2, we can see that for scenarios $\alpha_c = 1.4$ and $\alpha_c = 1.7$, the cost ratio is firstly increasing and then decreasing in the number of isolation beds under static policy, and for the scenario $\alpha_c = 2$, the cost ratio is increasing in the number of isolation beds under static policy. We find that the optimal fixed number of isolation beds are 20, 70, 120, respectively, which achieve the biggest cost ratio under each scenario. In addition, the peak number of COVID-19 patients in the public hospital system under three scenarios are 23, 80 and 142. It shows that if a public hospital has to adopt a static policy, it's optimal to reserve a fixed number of isolation beds that can almost deal with the peak number of expected COVID-19 patients.

Even comparing our dynamic policy with best static policy under each scenario, we still obtain good cost ratios, which are 0.814 for scenario $\alpha_c = 1.4$, 0.665 for sce-

nario $\alpha_c = 1.7$, and 0.872 for scenario $\alpha_c = 2$. It indicates that our dynamic approach can at least cut down the expected cost by 19.6%, 33.5%, and 12.8%. Furthermore, the three ratios also show that our dynamic policy has the best performance under the scenario $\alpha_c = 1.7$ when compared with static policy. The underlying reason may be as follows. For a region facing a weak pandemic alert (e.g., $\alpha_c = 1.4$), reserving a fixed and smaller number is enough to cope with the COVID-19 patients. If using dynamic approach, the number of beds transformed between two types of beds in each period is quite small. Therefore, the dynamic approach will not lead to a significant cost cutting effect under this scenario. For a region facing a serious pandemic alert (e.g., $\alpha_c = 2$), even though putting aside all the beds as isolation beds, the public hospital still cannot handle the surging number of COVID-19 patients. In such a serious situation, holding a large amount of isolation beds through the whole horizon is a reasonable manner to cope with COVID-19 pandemic. Similarly, if using dynamic policy, it still needs to quickly allocate large amount of beds as isolation beds and hold those isolation beds for periods, which performs slightly better than the static policy. However, for a region facing a medium pandemic alert (e.g., $\alpha_c = 1.7$), dynamically allocating isolation beds according to the evaluation process of COVID-19 can effectively cut down the total cost by reserving enough isolation beds for admitting COVID-19 patients in the outbreak phase and transforming excessive isolation beds to ordinary beds for admitting non-COVID-19 patients in the post-peak phase, which cannot be achieved by a static policy. Therefore, under this scenario, our dynamic policy performs best by significantly reducing 33.5% of the total expected cost.

5.3 Performance comparison of different subsidy schemes

In this section, we compare the experiment outcomes with and without subsidy schemes. The results are listed in Table 5.3. For the subsidy scheme, we assume the subsidy amount offered for each patient is a linear function of the patient's position in the waiting list. For example, if the subsidy rule is $s = 20 + 2(w_t - position)$,

it means that each patient can get at least a subsidy amount of 20 units, and the patient at the head of the queue can obtain 2 more units than the following one. We then design four subsidy rules listed in Table 5.3: all have a fixed part of 20, and their linear slopes are 2, 4, 6, and 8, respectively. In all these subsidy schemes, the total subsidy amount the system needs to pay is an increasing convex function of the number of subsidized patients, which is in line with our original assumption.

Table 5.3: The effects of the subsidy scheme ($\alpha_c = 2$)

Subsidy Rule		Total Cost		Average Queue Length	
		Value	Δ_c	Value	Δ_q
Without Subsidy	-	120,108	-	67.9	-
Subsidy Rule 1	$c_s = n_{3,t}^2 + 19n_{3,t}$	36,897	0.31	11.3	0.17
Subsidy Rule 2	$c_s = 2n_{3,t}^2 + 18n_{3,t}$	50,294	0.42	17.9	0.26
Subsidy Rule 3	$c_s = 3n_{3,t}^2 + 17n_{3,t}$	62,987	0.52	23.6	0.35
Subsidy Rule 4	$c_s = 4n_{3,t}^2 + 16n_{3,t}$	75,080	0.62	28.5	0.42

To measure the differences with and without subsidy, we use two ratios $\Delta_c = V_c^s/V_c^n$ and $\Delta_q = L_s/L_n$, where V_c^s (V_c^n, resp.) is the total expected cost with a subsidy scheme (without a subsidy scheme, resp.), and L_s (L_n, resp.) is the average queue length with a subsidy scheme (without a subsidy scheme, resp.). As we can see from Table 5.3, subsidy scheme yields a lower total cost and a shorter queue length. By comparison of four subsidy schemes, the value of Δ_c is decreasing when a more generous amount of subsidy is offered to an elective patient. That is, because subsidy cost gets higher when transferring the same number of patients in the waiting list; the effect of subsidy scheme gets less significant. However, even under the most generous subsidy rule (i.e., subsidy rule 4) , the existence of subsidy scheme can still significantly cut down the total expected cost by 38% and reduce the queue length by 58% compared with the experiment outcome without subsidy. Furthermore, patients waiting on the list are more likely to accept a relatively smaller amount of subsidy and go to a private hospital because they are fear of being infected by virus in a public hospital during the COVID-19 pandemic. Therefore, the subsidy scheme can be extremely effective when facing COVID-19 pandemic.

In addition, we also investigate the impact of hospital bed capacity on the ef-

fectiveness of a subsidy scheme. This sensitivity analysis result can help managers to consider whether or not to do capacity expansion. In the previous experiments, we set the bed capacity as 120. We are interested in under what bed capacity can a subsidy scheme work best. We change the bed capacity from 100 to 240 in increments of 20, under which capacity the cost ratio V_c^s/V_c^n is calculated based on subsidy rule 4 (i.e., $c_s = 4n_{3,t}^2 + 16n_{3,t}$) . The results are showed in Figure 5.7.

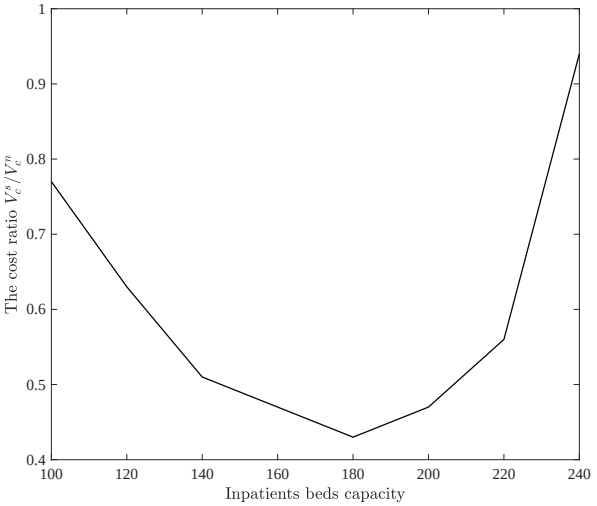

Figure 5.7: The cost ratio under different bed capacity

The numerical result shows the subsidy scheme works best with a moderate size of bed capacity. In this example, it is 180 and the total cost reduced by 57% due to subsidy scheme. In a scenario with ample bed capacity, the subsidy scheme does not work that well because elective patients can be directly admitted into the system. Also, when bed capacity is scarce, relying on subsidy scheme is not enough as the waiting list is too long. In that situation, the manager may have to consider to solve the problem in another approach such as building mobile cabin hospitals.

Chapter 6

Conclusion

In this paper, we study the issue of hospital bed allocation and waiting list management to serve three classes of patients: COVID-19, emergency and elective patients. Our work is motivated by problems in capacity planning given a number of pandemics in the 21st century, especially the most recent COVID-19 pandemic. We consider the two-way bed allocation decisions on the supply side and managing the waiting list of elective patients on the demand side. In the outbreak phase, facing the surging number of COVID-19 patients, the manager can consider to retrofit ordinary beds into isolation beds with negative pressure facilities; in the post-peak phase, the manager can transfer some excessively reserved isolation beds back into ordinary beds. The extensive numerical examples are conducted to gain insights on managing the system.

We demonstrate the multimodularity of the total expected cost function on the number of reservation beds and the length of waiting list. We show that a longer waiting list makes it harder to reserve more isolation beds. Hence it is better to use subsidy scheme to shorten the waiting list. By characterizing the marginal costs of key decisions in our model, we also graphically demonstrate the optimal decisions on bed allocation and patient subsidy, facing different combinations of the number of reserved isolation beds and the length of waiting list.

We conduct some numerical experiments to compare the effects of our dynamic allocation policy with that of a static policy (namely, a policy with a fixed number of isolation beds). We find that our dynamic allocation policy works best when facing a medium-alert pandemic situation and it can reduce the total cost by at least 33.5%.

In a serious-alert situation, beds are all reserved for COVID-19 patients and hence there is little room left for dynamic adjustment of bed capacities. In a weak-alert situation, demand for isolation bed is small and hence dynamic allocation also does not work well.

Through our numerical studies, we find that a subsidy scheme can greatly reduce the total cost and shorten the waiting list of elective patients, in comparing with the scenario without subsidy. Our numerical study shows that, with subsidy, the total cost can be reduced by 38% and the waiting list of elective patients can be reduced by 58% compared to the outcome without a subsidy scheme. We also find that the subsidy scheme works best when facing moderate bed capacity.

Our work presents a stepping stone to further study questions associated with dynamic medical resource allocation facing the COVID-19 pandemic. It would be interesting to use real data to identify the time-dependent arrival rates of COVID-19 and elective patients in different phases of pandemic. Another research question is to study the coordination issues between public and private hospitals in the pandemic.

Appendix A

Proofs for Chapter 4

Proof of Proposition 1

The results of this proposition is proved by induction on t. Given $V_{T+1} = 0$, the results obviously hold for $t = T + 1$. We assume the proposition holds for period $t + 1$ and we prove that it also holds for period t.

We first prove $V_t(x_t, y_t, w_t, \mathbf{d}_t)$ is jointly convex in (y_t, w_t). Based on the optimality equation 3.8, we have $L_t(y_t - \epsilon_t + n_{1,t} - n_{4,t} + \xi_t^c l_t^c) = c_p(y_t - \epsilon_t + n_{1,t} - n_{4,t} + \xi_t^c l_t^c)^- + c_o(y_t - \epsilon_t + n_{1,t} - n_{4,t} + \xi_t^c l_t^c)^+$. Since $L_t(\cdot)$ is a convex function, it is easy to verify that $L_t(y_t - \epsilon_t + n_{1,t} - n_{4,t} + \xi_t^c l_t^c)$ is jointly convex in $(y_t, n_{1,t})$ (Theorem 5.7 in Rockafellar, 1972). Due to the convexity can be hold for the linear transformation and $M_t(\cdot)$ is a convex function, F_t is jointly convex in (y_t, w_t), where $F_t = L_t + W_t$. As we assume V_{t+1} is jointly convex in its arguments, we know that $\mathbb{E}V_{t+1}(x_{t+1}, y_{t+1}, w_{t+1}, \mathbf{d}_{t+1})$ is also convex (Theorem 5.7 of Rockafellar, 1972). In addition, the first two and the fourth terms in equation 3.8 are linear functions; thus, these three terms are convex. Due to $V_t(x_t, y_t, w_t, \mathbf{d}_t)$ is minimized by a linear transformation of four convex functions, the convexity of V_t also holds; it follows from Theorem A.4 in Porteus (2002). By induction, the results hold for all periods.

In what follows, we prove $V_t(x_t, y_t, w_t, \mathbf{d}_t)$ increases in the length of waiting list w_t. Define $w_t > w_t^{'}$. Since $V_{t+1}(x_{t+1}, y_{t+1}, w_{t+1}) \geqslant V_{t+1}(x_{t+1}, y_{t+1}, w_{t+1}^{'})$ and $M_t(w_t)$ is increasing convex function, we have $F_t(y_t, w_t) \geqslant F_t(y_t, w_t^{'})$. Thus, it follows from Equation (3.8) that $V_t(x_t, y_t, w_t) \geqslant V_t(x_t, y_t, w_t^{'})$. Q.E.D.

Proof of Proposition 2

The local optimality condition for global optimality is rectified by Murota (2005) for multimodular functions. Based on our assumptions, we know that $F_t(y_t, w_t) + \gamma V_{t+1}(x_{t+1}, y_{t+1}, w_{t+1}, \mathbf{d}_{t+1})$ is multimodular in (y_t, w_t) for any $x_t, \mathbf{d}_t \geqslant 0$; since multimodular is preserved by taking expectation (Lemma 2 (iv), Li and Yu, 2014), thus, $\mathbb{E}F_t(y_t - \epsilon_t + \xi_t l_t, w_t) + \gamma \mathbb{E}V_{t+1}(x_{t+1}, y_{t+1} - \epsilon_{t+1} + \xi^c_{t+1} l^c_{t+1}, w_{t+1}, \mathbf{d}_{t+1})$ is multimodular in (y_t, w_t). In what follows, it is sufficient to prove that $H^*_t(x_t, y_t - \mu, w_t - y_t)$ is submodular in (y_t, w_t, μ) subject to $w_t \geqslant y_t$ for any $x_t \geqslant 0$. For convenience, we define

$$h_t(y_t, w_t, x_t) = \mathbb{E}F_t(y_t - \epsilon_t + \xi_t l_t, w_t) + \gamma \mathbb{E}V_{t+1}(x_{t+1}, y_{t+1} - \epsilon_{t+1} + \xi^c_{t+1} l^c_{t+1}, w_{t+1}, \mathbf{d}_{t+1}). \tag{A.1}$$

Then, the optimality equation (4.1) can be rewritten as, for $t = 1, \cdots, T$,

$$\begin{aligned} H^*_t(x_t, y_t - \mu, w_t - y_t) &= \min\{c_s(n_{3,t}) + c_d n_{4,t} + h_t(y_t - \mu - n_{4,t}, w_t - n_{3,t} - n_{4,t} - y_t, x_t)\} \\ &= \min\{c_s(n_{3,t}) + c_d n_{4,t} + h_t(\bar{y}_t - \bar{\mu}, w_t - \bar{y}_t)\}, \end{aligned} \tag{A.2}$$

where $\bar{y}_t = y_t + n_{3,t} + n_{4,t}$ and $\bar{\mu} = \mu + n_{3,t} + 2n_{4,t}$. It is note that $c_s(n_{3,t}) + c_d n_{4,t} + h_t(\bar{y}_t - \bar{\mu}, w_t - \bar{y}_t)$ is submodular in $(\bar{y}_t, \bar{\mu}, w_t, n_{3,t}, n_{4,t})$ for any $x_t, \delta_t, \epsilon_t \geqslant 0$ in a sublattice formed by $\bar{y}_t \geqslant w_t$, $n_{3,t} \geqslant 0$, and $n_{4,t} \geqslant 0$; since $y_t = \bar{y}_t - n_{3,t} - n_{4,t}$ and $\mu = \bar{\mu} - n_{3,t} - 2n_{4,t}$, by applying Corollary 1 (Chen et al., 2013), $H^*_t(x_t, y_t - \mu, w_t - y_t)$ is submodular in (y_t, μ). In addition, $H^*_t(x_t, y_t - \mu, w_t - y_t)$ can be re-formulated as follows:

$$H^*_t(x_t, y_t - \mu, w_t - y_t) = \min\{c_s(n_{3,t}) + c_d n_{4,t} + h_t(y_t - \bar{\mu}, \bar{w}_t - y_t, x_t)\}, \tag{A.3}$$

where $\bar{w}_t = w_t - n_{3,t} - n_{4,t}$ and $\bar{\mu} = \mu + n_{4,t}$. Here, $c_s(n_{3,t}) + c_d n_{4,t} + h_t(y_t - \bar{\mu}, \bar{w}_t - y_t, x_t)$ is submodular in $(\bar{w}_t, \bar{\mu}, y_t, n_{3,t}, n_{4,t})$ for any $x_t, \delta_t, \epsilon_t \geqslant 0$ in a sublattice formed by $y_t \geqslant \bar{w}_t$, $n_{3,t} \geqslant 0$, and $n_{4,t} \geqslant 0$; since $w_t = \bar{w}_t + n_{3,t} + n_{4,t}$ and $\mu = \bar{\mu} - n_{4,t}$, thus $H^*_t(x_t, y_t - \mu, w_t - y_t)$ is submodular in (w_t, μ). Now, we approve $H^*_t(x_t, y_t - \mu, w_t - y_t)$ is submodular in (y_t, w_t). With fixed μ, for $\theta_i,\ i = 1, 2$, let $(n'_{3,t}, n'_{4,t})$ and $(n''_{3,t}, n''_{4,t})$ respectively represent the optimizers at $((y_t + \theta_1) - \mu, (w_t + \theta_2) - (y_t + \theta_1))$ and

$((y_t+\theta_1)-\mu, w_t-(y_t+\theta_1))$. Then, we have

$$
\begin{aligned}
&H_t^*(x_t,(y_t+\theta_1)-\mu,(w_t+\theta_2)-(y_t+\theta_1))-H_t^*(x_t,(y_t+\theta_1)-\mu,w_t-(y_t+\theta_1))\\
&=c_s(n'_{3,t})+c_d n'_{4,t}+h_t(y_t+\theta_1-\mu-n'_{4,t},(w_t+\theta_2)-(y_t+\theta_1)-n'_{3,t}-n'_{4,t},x_t)\\
&-[c_s(n''_{3,t})+c_d n''_{4,t}+h_t(y_t+\theta_1-\mu-n''_{4,t},w_t-(y_t+\theta_1)-n''_{3,t}-n''_{4,t},x_t)]\\
&=c_s(n'_{3,t})+c_d(n'_{4,t}+\theta_1)+h_t(y_t-\mu-(n'_{4,t}+\theta_1),(w_t+\theta_2)-y_t-n'_{3,t}-(n'_{4,t}+\theta_1),x_t)\\
&-[c_s(n''_{3,t})+c_d(n''_{4,t}+\theta_1)+h_t(y_t-\mu-(n''_{4,t}+\theta_1),w_t-y_t-n''_{3,t}-(n''_{4,t}+\theta_1),x_t)]\\
&=H_t^*(x_t,y_t-\mu,w_t+\theta_2-y_t)-H_t^*(x_t,y_t-\mu,w_t-y_t)
\end{aligned}
\tag{A.4}
$$

where the last equality holds due to $(n'_{3,t},n'_{4,t})$ and $(n''_{3,t},n''_{4,t})$ are optimizers. The result indicates $H_t^*(x_t,y_t-\mu,w_t-y_t)$ is decreasing difference of (y_t,w_t). By Topkis (1998), $H_t^*(x_t,y_t-\mu,w_t-y_t)$ is submodular in (y_t,w_t). We obtain that $H_t^*(x_t,y_t-\mu,w_t-y_t)$ is submodular in (y_t,μ), (w_t,μ) and (y_t,w_t), respectively; thus, $H_t^*(x_t,y_t-\mu,w_t-y_t)$ is submodular in (y_t,w_t,μ), which implies $H_t^*(x_t,y_t,w_t)$ is multimodular in (y_t,w_t). Q.E.D.

Proof of Proposition 3

By replacing y_t and w_t with $y_t=y_t-\mu$ and $w_t=w_t-y_t$, the optimality equation (4.2) can be rewritten as, for $t=1,\cdots,T$,

$$
\begin{aligned}
&V_t(x_t,y_t-\mu,w_t-y_t,\mathbf{d}_t)\\
&=\min_{0\leqslant n_{1,t}\leqslant x_t-\lambda_t}\{c_b n_{1,t}+c_n(x_t-n_{1,t})+H_t^*(y_t-\mu+n_{1,t},w_t-y_t+\lambda_t-x_t+n_{1,t},x_t)\}\\
&=\min_{0\leqslant n_{1,t}\leqslant x_t-\lambda_t}\{c_n x_t+(c_b-c_n)n_{1,t}+H_t^*(\bar{y}_t-\mu,w_t-\bar{y}_t+\lambda_t-x_t,x_t)\},
\end{aligned}
\tag{A.5}
$$

where $w_t\geqslant y_t$ and $\bar{y}_t=y_t+n_{1,t}$. Given $c_n x_t+(c_b-c_n)n_{1,t}+H_t^*(\bar{y}_t-\mu,w_t-\bar{y}_t+\lambda_t-x_t,x_t)$ is submodular in $(y_t,w_t,\bar{y}_t,\mu)$ for $x_t\geqslant 0$ in a sublattice formed by $y_t\geqslant w_t$ and $0\leqslant n_{1,t}\leqslant x_t-\lambda_t$, $V_t(x_t,y_t-\mu,w_t-y_t,\mathbf{d}_t)$ is multimodular in (y_t,w_t) for any $x_t\geqslant 0$. Therefore, V_t is multimodular in (y_t,w_t) for any $x_t\geqslant 0$. Q.E.D.

Proof of Proposition 4

By Corollary 1, we have $V_t(x_t,y_t,w_t)$ is multimodular in (y_t,w_t). According to Lemma 1 (Li & Tu, 2014), $V_t(x_t,y_t,w_t)$ is submodular in (y_t,w_t). Thus, by applying

Theorem 2.8.2 in Topkis (1998), $n^*_{1,t}$ is decreasing in y_t and w_t; $n^*_{3,t}$ and $n^*_{4,t}$ are both increasing in y_t and w_t. Q.E.D.

Proof of Proposition 5

For a differential function $f(x_1, \cdots, x_i, \cdots, x_n)$ on $\mathbb{R}^n$, the partial derivative of a function f with respect to variable x_i is denoted by $f^i(x_1, \cdots, x_i, \cdots, x_n) = \frac{\partial f(x_1, \cdots, x_i, \cdots, x_n)}{\partial x_i}$. Thus, we define $h^2_t(x_t, y_t, w_t) = \frac{\partial h_t(x_t, y_t, w_t)}{\partial y_t}$ and $h^3_t(x_t, y_t, w_t) = \frac{\partial h_t(x_t, y_t, w_t)}{\partial w_t}$.

The outbreak phase:

(i) We first prove the monotonicity of $Y^o_t(w_t, x_t)$ and $Z^o_t(w_t, x_t)$ with respect to w_t. Suppose $w_t^{'} > w_t \geqslant 0$, for any $x_t \geqslant 0$, we have

$$c_b + h^2_t(x_t, Y^o_t(w_t^{'}, x_t), w_t) \leqslant c_b + h^2_t(x_y, Y^o_t(w_t^{'}, x_t), w_t^{'}), \tag{A.6}$$

where the inequality holds as $h_t(x_t, y_t, w_t)$ is multimodular in (y_t, w_t).

Since $c_b + h^2_t(x_t, y_t, w_t)$ is increasing in y_t, we have $Y^o_t(w_t^{'}, x_t) \leqslant Y^o_t(w_t, x_t)$, and $Y^o_t(w_t, x_t)$ is decreasing in w_t. Similarly, we have

$$c_b - c_s + h^2_t(x_t, Z^o_t(w_t^{'}, x_t), w_t) + h^3_t(x_t, Z^o_t(w_t^{'}, x_t), w_t) \leqslant c_b - c_s + h^2_t(x_t, Z_t(w_t^{'}, x_t), w_t^{'}) + h^3_t(x_t, Z_t(w_t^{'}, x_t), w_t^{'}),$$

the inequality holds since $h_t(x_t, y_t, w_t)$ is convex and multimodular in (y_t, w_t).

In addition, due to $c_b - c_s + h^2_t(x_t, y_t, w_t) + h^3_t(x_t, y_t, w_t)$ is increasing in y_t, we have $Z^o_t(w_t^{'}, x_t) \leqslant Z^o_t(w_t, x_t)$, and thus $Z^o_t(w_t, x_t)$ is decreasing in w_t for any $x_t \geqslant 0$.

(ii) We now prove the monotone properties of three curves. Considering the outbreak of COVID-19 pandemic, more ordinary beds are reserved as isolation beds which results in inadequate capacity to admit elective patients. That is, more elective patients have to join the waiting list. In addition, some elective patients could be subsidized from the waiting list. Thus, $W_t(y_t, x_t)$ may not display the monotone trend. Nevertheless, $W_t(y_t, x_t)$ must cross over the intersection of $Y^o_t(w_t, x_t)$ and $Z^o_t(w_t, x_t)$. It is critical to analyze the property of these curves, i.e., how the curves are entering into and leaving from the intersection.

More specifically, we aim to analyze how $Y^o_t(w_t, x_t)$ and $Z^o_t(w_t, x_t)$ behave to the left (or right) of $W_t(y_t, x_t)$ when $W_t(y_t, x_t)$ cross over the intersection point of

$Y_t^o(w_t, x_t)$ and $Z_t^o(w_t, x_t)$. The positional information discloses the optimal rules of reservation and subsidization policy in the initial and outbreak phases (i.e., before the peak phase).

Let $(\hat{y}_t, \hat{w}_t)$ be the intersection point, i.e., $\hat{y}_t = Y_t^o(\hat{w}_t, x_t) = Z_t^o(\hat{w}_t, x_t)$. Since $Y_t^o(w_t, x_t)$ and $Z_t^o(w_t, x_t)$ are decreasing in w_t, when they approach $(\hat{y}_t, \hat{w}_t)$ from the left side of $\hat{y}_t$, the corresponding w_t are going down to approach $\hat{w}_t$. For any $w_t \geqslant \hat{w}_t$, it is note that both $Y_t^o(w_t, x_t)$ and $Z_t^o(w_t, x_t)$ locate on the left of vertical line $y_t = \hat{y}_t$ and $Y_t^o(w_t, x_t) \leqslant Y_t^o(\hat{w}_t, x_t) = \hat{w}_t$.

Thus, we have $h_t^2(x_t, Y_t^o(w_t, x_t), w_t) + h_t^3(x_t, Y_t^o(w_t, x_t), w_t) \geqslant h_t^2(x_t, \hat{y}_t, \hat{w}_t) + h_t^3(x_t, \hat{y}_t, \hat{w}_t)$, where the inequality holds since $h_t^2(x_t, y_t, w_t) + h_t^3(x_t, y_t, w_t)$ is increasing in w_t and decreasing in y_t. In addition, since $w_t \geqslant W_t(Y_t(w_t, x_t), x_t)$ and $h_t^2(x_t, y_t, w_t) + h_t^3(x_t, y_t, w_t)$ is decreasing in w_t, we have

$$c_b - c_s + h_t^2(x_t, Y_t^o(w_t, x_t), w_t) + h_t^3(x_t, Y_t^o(w_t, x_t), w_t) \geqslant$$
$$c_b - c_s + h_t^2(x_t, Y_t^o(w_t, x_t), W_t(Y_t^o(w_t, x_t), x_t)) + h_t^3(x_t, Y_t^o(w_t, x_t), W_t(Y_t^o(w_t, x_t), x_t)).$$

It shows that $Y_t^o(w_t, x_t) \geqslant Z_t^o(w_t, x_t)$; that is, if $w_t > \hat{w}_t$, curve $Y_t^o(w_t, x_t)$ is above curve $Z_t^o(w_t, x_t)$. Based on these analysis, we have

$$\begin{aligned} & c_s - h_t^3(x_t, Z_t^o(w_t, x_t), w_t) \\ &= c_s + h_t^2(x_t, Z_t^o(w_t, x_t), w_t) - [h_t^2(x_t, Z_t^o(w_t, x_t), w_t) + h_t^3(x_t, Z_t^o(w_t, x_t), w_t)] \\ &= c_b + h_t^2(x_t, Z_t^o(w_t, x_t), w_t) \leqslant c_b + h_t^2(x_t, Y_t^o(w_t, x_t), w_t). \end{aligned}$$

Thus, $w_t \geqslant W_t(Z_t^o(w_t, x_t), x_t)$. Based on the analysis above, curve $W_t(y_t, x_t)$ is below curve $Z_t^o(w_t, x_t)$ if $w_t \geqslant \hat{w}_t$. By the same logic, we can show $W_t(y_t, x_t)$ is above curve $Z_t^o(w_t, x_t)$ when $w_t < \hat{w}_t$. For $w_t < \hat{w}_t$, we know that both $Y_t^o(w_t, x_t)$ and $Z_t^o(w_t, x_t)$ are greater than or equal to $\hat{y}_t$. We have $h_t^2(x_t, Y_t^o(w_t, x_t), w_t) + h_t^3(x_t, Y_t^o(w_t, x_t), w_t) \leqslant h_t^2(x_t, \hat{y}_t, \hat{w}_t) + h_t^3(x_t, \hat{y}_t, \hat{w}_t)$, where the inequality holds since $h_t^2(x_t, y_t, w_t) + h_t^3(x_t, y_t, w_t)$ is increasing in w_t and decreasing in y_t. Therefore, we have $w_t \leqslant W_t(Y^o(w_t, x_t), x_t)$, then,

$$c_b - c_s + h_t^2(x_t, Y_t^o(w_t, x_t), w_t) + h_t^3(x_t, Y_t^o(w_t, x_t), w_t) \leqslant$$
$$c_b - c_s + h_t^2(x_t, Y_t^o(w_t, x_t), W_t(Y_t^o(w_t, x_t), x_t)) + h_t^3(x_t, Y_t^o(w_t, x_t), W_t(Y_t^o(w_t, x_t), x_t)).$$

By definition, $Y_t^o(w_t, x_t) \leqslant Z_t^o(w_t, x_t)$ and curve $Y_t^o(w_t, x_t)$ is below curve $Z_t^o(w_t, x_t)$

when $w_t < \hat{w}_t$. In addition,

$$\begin{aligned} & c_s - h_t^3(x_t, Z_t^o(w_t, x_t), w_t) \\ &= c_s + h_t^2(x_t, Z_t^o(w_t, x_t), w_t) - [h_t^2(x_t, Z_t^o(w_t, x_t), w_t) + h_t^3(x_t, Z_t^o(w_t, x_t), w_t)] \\ &= c_b + h_t^2(x_t, Z_t^o(w_t, x_t), w_t) \geqslant c_b + h_t^2(x_t, Y_t^o(w_t, x_t), w_t). \end{aligned} \tag{A.7}$$

Thus, $w_t < W_t(Z^o(w_t, x_t), x_t)$ and curve $W_t(y_t, x_t)$ is above curve $Z_t(w_t, x_t)$.

The post-peak phase:

We adopt the similar logic to prove the monotone property of three curves in the post-peak phase.

(i) We first prove the monotonicity of $Y_t^p(w_t, x_t)$ and $Z_t^p(w_t, x_t)$ with respect to w_t. Suppose $w_t^{'} > w_t \geqslant 0$, for any $x_t \geqslant 0$, we have

$$\begin{aligned} & c_n + c_d - h_t^2(x_t, Y_t^p(w_t^{'}, x_t), w_t) - h_t^3(x_t, Y_t^p(w_t^{'}, x_t), w_t) \\ & \leqslant c_n + c_d - h_t^2(x_t, Y_t^p(w_t^{'}, x_t), w_t') - h_t^3(x_t, Y_t^p(w_t^{'}, x_t), w_t') \end{aligned} \tag{A.8}$$

where the inequality holds as $h_t(x_t, y_t, w_t)$ is multimodular in (y_t, w_t). Since $c_n + c_d - h_t^2(x_t, y_t, w_t) - h_t^3(x_t, y_t, w_t)$ is increasing in y_t, we have $Y_t^p(w_t^{'}, x_t) \leqslant Y_t^p(w_t, x_t)$, and $Y_t^p(w_t, x_t)$ is decreasing in w_t. Similarly, we have

$$c_n + c_b - c_s - h_t^2(x_t, Z_t^p(w_t^{'}, x_t), w_t) \leqslant c_n + c_b - c_s - h_t^2(x_t, Z_t^p(w_t^{'}, x_t), w_t) \tag{A.9}$$

the inequality holds since $h_t(x_t, y_t, w_t)$ is convex and multimodular in (y_t, w_t). In addition, due to $c_n + c_b - c_s - h_t^2(x_t, y_t, w_t)$ is increasing in y_t, we have $Z_t^p(w_t^{'}, x_t) \leqslant Z_t^p(w_t, x_t)$, and thus $Z_t^p(w_t, x_t)$ is decreasing in w_t for any $x_t \geqslant 0$.

(ii) As addressed above, $W_t(y_t, x_t)$ may not display the monotone trend. Nevertheless, $W_t(y_t, x_t)$ must cross over the intersection of $Y_t^p(w_t, x_t)$ and $Z_t^p(w_t, x_t)$. It is critical to analyze the property of these curves, i.e., how the curves are entering into and leaving from the intersection.

We aim to analyze how $Y_t^p(w_t, x_t)$ and $Z_t^p(w_t, x_t)$ behave to the left (or right) of $W_t(y_t, x_t)$ when $W_t(y_t, x_t)$ cross over the intersection point of $Y_t^p(w_t, x_t)$ and $Z_t^p(w_t, x_t)$. The positional information discloses the optimal rules of reservation and subsidization policy in post-peak phase.

Let $(\tilde{y}_t, \tilde{w}_t)$ be the intersection point, i.e., $\tilde{y}_t = Y_t^p(\tilde{w}_t, x_t) = Z_t^p(\tilde{w}_t, x_t)$. Since $Y_t^p(w_t, x_t)$ and $Z_t^p(w_t, x_t)$ are decreasing in w_t, when they approach $(\tilde{y}_t, \tilde{w}_t)$ from the left side of $\tilde{y}_t$, the corresponding w_t are going down to approach $\tilde{w}_t$. For any $w_t \geqslant \tilde{w}_t$, it is note that both $Y_t^p(w_t, x_t)$ and $Z_t^p(w_t, x_t)$ locate on the left of vertical line $y_t = \tilde{y}_t$ and $Y_t^p(w_t, x_t) \leqslant Y_t^p(\tilde{w}_t, x_t) = \tilde{w}_t$.

Thus, we have $h_t^2(x_t, Y_t^p(w_t, x_t), w_t) + h_t^3(x_t, Y_t^p(w_t, x_t), w_t) \geqslant h_t^2(x_t, \tilde{y}_t, \tilde{w}_t) + h_t^3(x_t, \tilde{y}_t, \tilde{w}_t)$, where the inequality holds since $h_t^2(x_t, y_t, w_t) + h_t^3(x_t, y_t, w_t)$ is increasing in w_t and decreasing in y_t. In addition, since $w_t \geqslant W_t(Y_t^p(w_t, x_t), x_t)$ and $h_t^2(x_t, y_t, w_t) + h_t^3(x_t, y_t, w_t)$ is decreasing in w_t, we have

$$c_n + c_d - c_s - h_t^2(x_t, Y_t^p(w_t, x_t), w_t) \geqslant c_n + c_d - c_s - h_t^2(x_t, Y_t^p(w_t, x_t), W_t(Y_t^p(w_t, x_t), x_t)). \tag{A.10}$$

It shows that $Y_t^p(w_t, x_t) \geqslant Z_t^p(w_t, x_t)$; that is, if $w_t > \tilde{w}_t$, curve $Y_t^p(w_t, x_t)$ is above curve $Z_t^p(w_t, x_t)$. Based on these analysis, we have

$$\begin{aligned}
&c_s - h_t^3(x_t, Z_t^p(w_t, x_t), w_t) \\
&= c_s + h_t^2(x_t, Z_t^p(w_t, x_t), w_t) + h_t^3(x_t, Z_t^p(w_t, x_t), w_t) - h_t^2(x_t, Z_t^p(w_t, x_t), w_t) \\
&= c_n + c_d + h_t^3(x_t, Z_t^p(w_t, x_t), w_t) - h_t^2(x_t, Z_t^p(w_t, x_t), w_t) \\
&\leqslant c_n + c_d + h_t^3(x_t, Y_t^p(w_t, x_t), w_t) - h_t^2(x_t, Y_t^p(w_t, x_t), w_t).
\end{aligned} \tag{A.11}$$

Thus, $w_t \geqslant W_t(Z_t^p(w_t, x_t), x_t)$. Based on the analysis above, curve $W_t(y_t, x_t)$ is below curve $Z_t^p(w_t, x_t)$ if $w_t \geqslant \tilde{w}_t$. By the same logic, we can show $W_t(y_t, x_t)$ is above curve $Z_t^p(w_t, x_t)$ when $w_t < \tilde{w}_t$. For $w_t < \tilde{w}_t$, we know that both $Y_t^p(w_t, x_t)$ and $Z_t^p(w_t, x_t)$ are greater than or equal to $\tilde{y}_t$. We have $h_t^2(x_t, Y_t^p(w_t, x_t), w_t) + h_t^3(x_t, Y_t^p(w_t, x_t), w_t) \leqslant h_t^2(x_t, \tilde{y}_t, \tilde{w}_t) + h_t^3(x_t, \tilde{y}_t, \tilde{w}_t)$, where the inequality holds since $h_t^2(x_t, y_t, w_t) + h_t^3(x_t, y_t, w_t)$ is increasing in w_t and decreasing in y_t. Therefore, we have $w_t \leqslant W_t(Y^p(w_t, x_t), x_t)$, then,

$$c_n + c_d - c_s - h_t^2(x_t, Y_t^p(w_t, x_t), w_t) \leqslant c_n + c_d - c_s - h_t^2(x_t, Y_t^p(w_t, x_t), W_t(Y_t^p(w_t, x_t), x_t)) \tag{A.12}$$

By definition, $Y_t^p(w_t, x_t) \leqslant Z_t^p(w_t, x_t)$ and curve $Y_t^p(w_t, x_t)$ is below curve $Z_t^p(w_t, x_t)$

when $w_t < \tilde{w}_t$. In addition,

$$\begin{aligned}
&c_s - h_t^3(x_t, Z_t^p(w_t, x_t), w_t) \\
&= c_s + h_t^2(x_t, Z_t^p(w_t, x_t), w_t) + h_t^3(x_t, Z_t^p(w_t, x_t), w_t) - h_t^2(x_t, Z_t^p(w_t, x_t), w_t) \\
&= c_n + c_d + h_t^3(x_t, Z_t^p(w_t, x_t), w_t) - h_t^2(x_t, Z_t^p(w_t, x_t), w_t) \\
&\geqslant c_n + c_d + h_t^3(x_t, Y_t^p(w_t, x_t), w_t) - h_t^2(x_t, Y_t^p(w_t, x_t), w_t).
\end{aligned} \tag{A.13}$$

Thus, $w_t < W_t(Z^p(w_t, x_t), x_t)$ and curve $W_t(y_t, x_t)$ is above curve $Z_t(w_t, x_t)$. Q.E.D.

Appendix B

Summery of Notation

Table B.1: Summary of Notation

	Description
t	The index of planning horizon, $t \in \{1, 2, ..., T\}$
x_t	The number of ordinary beds in period t
y_t	The number of reserved isolation beds in period t
w_t	The number of elective patients on the waiting list in period t
ϵ_t	The demand of COVID-19 patients in period t
δ_t	The demand of elective patients in period t
λ_t	The demand of emergency patients in period t
$n_{1,t}$	The number of retrofitted ordinary beds into isolation beds in period t
$n_{2,t}$	The number of ordinary beds allocated to serve non-COVID-19 patients in period t
$n_{3,t}$	The number of subsidized elective patients from the waiting list in period t
$n_{4,t}$	The number of isolation beds transformed back into ordinary beds in period t
ξ_t^c	The number of COVID-19 patients being discharged from the system in period t
ξ_t^n	The number of non-COVID-19 patients being discharged from the system in period t
l_t^c	The number of COVID-19 patients in the public hospital in period t
l_t^n	The number of non-COVID-19 patients in the public hospital in period t
c_o	The unit idling cost of an isolation bed
c_b	The unit cost of retrofitting an ordinary bed into an isolation bed
c_n	The unit cost for admission a non-COVID-19 patient
c_p	The unit penalty cost if a COVID-19 patient is not timely admitted
c_d	The unit transferring cost of using an isolation bed to admit non-COVID-19 patients
$c_s(\cdot)$	The financial subsidy offered to the elective patients
$M(\cdot)$	The waiting cost incurred if elective patients are on the waiting list
γ	The discount factor
V_t	The total expected cost incurred from periods t to T

References

[1] 7NEWS. Coronavirus latest cases: Staggering rise in NSW ICU admissions. 7news.com.au/lifestyle/health-wellbeing/coronavirus-latest-cases-staggering-rise-in-nsw-icu-admissions-c-900435

[2] ABC News. How does coronavirus COVID-19 compare to flu? https://www.abc.net.au/news/health/2020-03-20/how-coronavirus-covid-19-compares-to-flu/12073696

[3] Allon G., Deo S., & Lin W. 2013. The impact of size and occupancy of hospital on the extent of ambulance diversion: Theory and evidence. *Operations Research*, 61(3), 544-562.

[4] Aviv, Y., A. Federgruen. 2001. Capacitated multi-item inventory systems with random and seasonally fluctuating demands: implications for postponement strategies.*Management Science*, 47(4), 512-531.

[5] Best T.J., Sandıkçı B., Eisenstein D.D., & D.O. Meltzer. 2015. Managing hospital inpatient bed capacity through partitioning care into focused wings. *Manufacturing & Service Operations Management*, 17(2), 157-176.

[6] Blackmon L., et al. 2021. Rapid Development of a Decision Support System to Alleviate Food Insecurity at the Los Angeles Regional Food Bank amid the COVID-19 Pandemic. *Production and Operations Management.* https://doi.org/10.1111/poms.13365

[7] Cardoen B., Demeulemeester E., & J. Beliën. 2010. Operating room planning and scheduling: A literature review. *European Journal of Operational Research*,

201(3), 921-932.

[8] Chan C. W., Farias V. F., Bambos N., & G. J. Escobar. 2012. Optimizing intensive care unit discharge decisions with patient readmissions. *Operations Research*, 60(6), 1323-1341.

[9] Cheung E. 2020. Coronavirus pandemic exposes Hong Kong's inadequate links between public, private hospitals as patients left waiting for treatment. https://www.scmp.com/news/hong-kong/health-environment/article/3082527/coronavirus-pandemic-exposes-hong-kongs

[10] DeCroix G.A., A. Arreola-Risa. 1998. Optimal production and inventory policy for multiple products under resource constraints. *Management Science*, 44(7), 950-961.

[11] Deo S., M. Sohoni. 2015. Optimal decentralization of early infant diagnosis of HIV in resource-limited settings. *Manufacturing & Service Operations Management*, 17(2):191-207.

[12] Ekici A., Keskinocak P., & J.L. Swann. 2014. Modeling influenza pandemic and planning food distribution. *Manufacturing & Service Operations Management*, 16(1), 11-27.

[13] Gerchak Y., Gupta D., M. Henig. 1996. Reservation planning for elective surgery under uncertain demand for emergency surgery. *Management Science*, 42(3), 321-334.

[14] Gong X., Chao X., S. Zheng. 2014. Dynamic pricing and inventory management with dual suppliers of different lead times and disruption risks. *Production and Operations Management*, 23(12): 2058-2074.

[15] Gupta D. 2007. Surgical suites' operations management. *Productions and Op-*

erations Management, 16(6), 689-700.

[16] Han E., Tan M.J., Turk E., Sridhar D., Leung G.M., Shibuya K., Asgar N., Oh J., García-Basteiro A.L., Hanefeld J., Cook A.R., Hsu L.Y., Teo Y.Y., Heymann D., Clark H., McKee M., H. Legido-Quigley. 2020. Lessons learnt from easing COVID-19 restrictions: an analysis of countries and regions in Asia Pacific and Europe. The Lancet (British edition), 396 (10261), 1525-1534.

[17] Helm J. E., & M. P. Van Oyen. 2014. Design and optimization methods for elective hospital admissions. *Operations Research*, 62(6), 1265-1282.

[18] Huh, W. T., Liu N., & V.A. Truong. 2013. Multi-resource allocation scheduling in dynamic environments. *Manufacturing & Service Operations Management*, 15(2), 280-291.

[19] Izady N., & I. Mohamed. 2019. A clustered overflow configuration of inpatient beds in hospitals. *Manufacturing & Service Operations Management*, forthcoming.

[20] Kaplan E.H. 2020. COVID-19 Scratch models to support local decisions. *Manufacturing & Service Operations Management*, 22(4), 645-655.

[21] Kim S. H., Chan C. W., Olivares M., & G. Escobar. 2015. ICU admission control: An empirical study of capacity allocation and its implication for patient outcomes. *Management Science*, 61(1), 19-38.

[22] Long E.F., Nohdurft E., S. Spinler. 2018. Spatial resource allocation for emerging epidemics: A comparison of greedy, myopic, and dynamic policies. *Manufacturing & Service Operations Management*, 20(2): 181-198.

[23] Li Q. & P. Yu. 2014. Multimodularity and its applications in three stochastic dynamic inventory problems. *Manufacturing & Service Operations Management*, 16(3), 455-463.

[24] Liu, N., Truong, V-A., & X., Wang. 2019. Integrated scheduling and capacity planning with considerations for patients' length-of-stays. *Production and Operations Management*, 28(7), 1735-1756.

[25] Mamani H., Chick S.E., & D. Simchi-Levi. 2013. A game-theoretic model of international influenza vaccination coordination. *Management Science*, 59(7), 1650-1670.

[26] Mandelbaum A., & A.L., Stolyar. 2004. Scheduling flexible servers with convex delay costs: Heavy-traffic optimality of the generalized $c\mu$-Rule. *Operations Research*, 52(6), 836-855.

[27] Murota, K. 2005. Note on Multimodularity and L-Convexity. *Mathematics of Operations Research*, 30(3): 658-661.

[28] Naderi B. et al. 2021. Increased surgical capacity without additional resources:Generalized operating room planning and scheduling. *Productions and Operations Management*. http://doi.org/doi:10.1111/poms.13397

[29] Phua, J., Weng, L., Ling, L., Egi, M., Lim, C.-M., Divatia, J. V., Shrestha, B. R., Arabi, Y. M., Ng, J., Gomersall, C. D., Nishimura, M., Koh, Y., & Du, B. 2020. Intensive care management of coronavirus disease 2019 (COVID-19): challenges and recommendations. *The Lancet Respiratory Medicine*, 8(5), 506-517.

[30] Pinker E., & T. Tezcan, 2013. Determining the optimal configuration of hospital inpatient rooms in the presence of isolation patients. *Operations Research*, 61(6), 1259-1276.

[31] Qian Q., Guo P., & R. Lindsey. 2017. Comparison of subsidy schemes for reducing waiting times in healthcare systems. *Productions and Operations Management*, 26(11), 2033-2049.

[32] Samiedaluie S., Kucukyazici B., Verter V., & D. Zhang. 2017. Managing patient admissions in a neurology ward. *Operations Research*, 65(3), 635-656.

[33] The HK Hospital Authority. 2020. 400 more isolation beds ready. https://www.news.gov.hk/eng/2020/03/20200329/20200329_173432_499.html?type=ticker

[34]Thompson S., Nunez M., Garfinkel R., & M. D. Dean. 2009. Efficient short-term allocation and reallocation of patients to floors of a hospital during demand surges. *Operations Research*, 57(2), 261-273.

[35]The Guardian. 2020. NHS to postpone millions of operations to tackle coronavirus. www.thegua
rdian.com/society/2020/mar/17/nhs-postpone-millions-operations-tackle-coronavirus

[36] Veinott A.F. 1965. Optimal policy for a multi-Product, dynamic, nonstationary inventory problem. *Management Science*, 12(3), 206-222.

[37] Topkis, D.M. 1998. Supermodularity and Complementatity. Princeton University Press.

[38] WHO. 2009. WHO phase of pandemic alert for Pandemic (H1N1). https://www.who.int/csr/disease/swineflu/phase/en/

[39] WHO. 2020. Numbers of coronavirus disease (COVID-19) pandemic. www.who.int/emergencies/diseases/novel-coronavirus-2019.