

Optimal Hospital Admission Control of Two Multi-Server Loss Queues in Series

by

Fehmi Mert Gürel

A Dissertation Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in

Industrial Engineering



July 25, 2019

**Optimal Hospital Admission Control of Two Multi-Server Loss Queues in
Series**

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Fehmi Mert Gürel

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

Prof. Lerzan Örmeci

Prof. Refik Güllü

Prof. Fikri Karaesmen

Date: _____



To my family & my Pocahontas

ABSTRACT

The prolongation of human life in parallel with the developments in medicine increases the number of patients who need intensive care. Costs of critical care medicine are 13.4% of hospital costs and 0.66% of the gross domestic product in United States. Despite these expenditures, 90% of intensive care units do not have enough capacity. Under these circumstances, most intensive care units face high occupancy rates. These high occupancy rates might cause people in need of intensive care to be rejected and increase the risk of mortality. An efficient admission control policy may enhance the service processes significantly.

In this research, two Markov decision process models which represent intensive care units and related wards in tandem are provided. Both models consist of two types of patients who are intensive care unit and ward patients. In model-1, recovery process of patients who are blocked in the ICU due to lack of available beds in the ward continues, while in model-2 this process is assumed to stop till there is an available bed in the ward. Using these models, the effects of the interaction between wards and intensive care units on optimal admission policies can be understood. Preferred class is a group of patients who are always admitted when there is an available bed. Certain special conditions must be met for the existence of preferred classes. I present sufficient conditions for each patient type to be preferred in both models. At the end, numerical analyses are presented to demonstrate the characteristics of the models. A variation of model-1 is also included in the numerical analysis to observe the effect of an additional type of patients, called post-operational patients, on the system.

ÖZETÇE

Tıpta yaşanan gelişmelere paralel olarak insan yaşamının uzaması, yoğun bakıma ihtiyacı olan hastaların sayısını arttırmaktadır. Kritik bakım maliyetleri hastane masraflarının %13,4'ünü kapsamaktadır ve ABD'nin gayri safi yurtiçi hasılatının %0,66'sını oluşturmaktadır. Bu harcamalara rağmen, yoğun bakım ünitelerinin %90'ı yeterli kapasiteye sahip değildir. Bu şartlar altında, yoğun bakım ünitelerinin çoğu yüksek doluluk oranları ile karşılaşmaktadır. Bu yüksek doluluk oranları, yoğun bakıma ihtiyacı olan kişilerin reddedilmesine ve ölüm risklerinin artmasına sebep olmaktadır. Verimli bir kabul kontrol politikası, servis süreci kalitesini önemli ölçüde arttırabilir.

Bu araştırmada, yoğun bakım ünitelerini ve bunlara eşlik eden servisleri seri şekilde temsil eden iki Markov karar süreci modeli incelenmiştir. İki model de iki tip hastadan oluşmaktadır. Bu hastalar yoğun bakım hastaları ve koğuş hastalarıdır. Model-1'de bloklanan hastalar servis görmeye devam ederken, Model-2'de bu hastalar yatak boşalana kadar beklemektedirler. Bu iki modeli kullanarak, servisler ve yoğun bakım üniteleri arasındaki etkileşimin optimal kabul politikalarına etkisini anlayabiliriz. Tercih edilen sınıf, uygun bir yatak bulunduğu kabul edilen bir hasta grubudur. Tercih edilen sınıfların varlığı için belirli özel şartlar sağlanmalıdır. Bu çalışmada, her iki model için de bir hasta tipinin tercih edilen tip olması için gerekli şartlar sunulmuştur. En son olarak, modellerin özelliklerini göstermek için sayısal bir analiz sunulmuştur. Model-1'in operasyon sonrası hasta olarak adlandırılan yeni bir hasta tipini içeren bir varyasyonu, bu yeni hasta tipinin sistem üzerindeki etkilerini gözlemlemek amacıyla sayısal analiz bölümünde sunulmuştur.

ACKNOWLEDGMENTS

Thanks cordially to my supervisor, Prof. Lerzan Örmeci for encouraging and supporting me not only during the preparation of this thesis but also during my whole university life, special thanks to my dear friend Berk Sırmaçek for his valuable support and to Nil Burumođlu & Rızacan Çilođlu for their great assistance.

TABLE OF CONTENTS

List of Tables	ix
List of Figures	x
Nomenclature	xi
Chapter 1: Introduction	1
Chapter 2: Literature Review	3
2.1 Healthcare Operations	3
2.2 Queuing Models	4
Chapter 3: Problem Description and Mathematical Model	9
3.1 Model-1	9
3.2 Model-2	12
3.3 Effect of an Additional Job	14
Chapter 4: Structure of the Optimal Policies	15
4.1 Preliminary Results for Model-1	15
4.2 Preliminary Results for Model-2	17
4.3 Existence of a Preferred Class for Model-1 and Model-2	18
4.3.1 A sufficient condition for type-2 to be preferred	19
4.3.2 A sufficient condition for type-1 to be preferred	21
4.4 Variation of Model-1 Including k Types of ICU Patients	24

Chapter 5: Numerical Analysis	26
5.1 Parameter Estimation	26
5.2 Base Case	28
5.3 Variation of Model-1 Including Post-Op Patients	33
5.4 Sensitivity Analysis	38
5.5 Properties of the Value Function	40
Chapter 6: Conclusion	42

LIST OF TABLES

5.1	Parameters taken from de Bruin et al. (2010), with presented data from 2006	27
5.2	Base Case Parameters	28
5.3	The difference in the value function between model-1 and model-2	32
5.4	Optimal policy for the base case of model-1 and model-2	33
5.5	The difference in the value functions of Model-1 and the variation including post-op patients when $x_3 = 0$	37
5.6	Sensitivity of the actions by the change in cost ratio between ICU and wards	38

LIST OF FIGURES

3.1	Arrival and departure processes	10
5.1	Changes in the value function for certain x_1	29
5.2	Changes in the value function for certain x_2	30
5.3	An example of theorem 2 for the base case of model-1	31

NOMENCLATURE

C_i : Number of beds per station

R_i : Accepting rewards of Type- i patient

λ_i : Arrival rate of Type- i patient

μ_i : Service time of Type- i patient

α : Discount rate

P : Probability that Type-1 patients do not die

Chapter 1

INTRODUCTION

As the population gets older, demand for intensive care units (ICUs) increases. ICUs are one of the most congested departments in hospitals. Frequent overloading of inpatient units contributes to denial of ICU admission Kim et al. (2016) and higher risk of mortality Kuntz et al. (2014) among other consequences. Moreover, a great portion of expenditures stems from ICUs. Despite the increase of bed capacity, most of the ICUs do not have enough bed. Increasing the bed capacity continuously is not feasible because beds and rooms are limited resources. It would be a better approach to find a policy that helps using these limited resources more efficiently.

There has been prior research done to understand characteristics and efficiency of ICU processes. The prolongation of human life in parallel with the developments in medicine increases the number of patients who need intensive care. Costs of critical care medicine are 13.4% of hospital costs and 0.66% of the gross domestic product in United States. Despite these expenditures, 90% of intensive care units do not have enough capacity. Under these circumstances, most intensive care units face high occupancy rates. These high occupancy rates might cause people in need of intensive care to be rejected and increase the risk of mortality. An efficient admission control policy may enhance the service processes significantly. In this study, two variations of a system consisting of an ICU and a related ward in series are modeled as Markov decision processes (MDP) to find the optimal policy for patient admissions. I consider the system explained above as a two-stage queuing system with no waiting room and two classes of patients. The ICU and the ward are the first and the second stages of the systems respectively. There are multiple identical servers in both stages. Type-1

patients may die during their stay in the ICU and leave the system, i.e., these patients only use the first stage. Other type-1 patients move to the wards to completely recover so that they complete their treatment after the second stage. Upon service completion of type-1 patients in the first stage if there is no free server in the second stage, type-1 patients who finish the first stage recovery are blocked. In model-1 blocked type-2 patients receive the ward services so that their recovery process continues as if they have moved to the ward. Meanwhile, in model-2 they have to wait for a free server for the second stage. All type-2 patients depart after completing their services.

The next chapter presents a literature survey of healthcare operations and related queuing models. In chapter 3, I present model-1 and model-2. In chapter 4, I provide some preliminary results and the sufficient conditions for a patient type to be preferred. Numerical analyses of model-1, model-2 and the variation of model-1 are given in chapter 5. Finally, the study is concluded in chapter 6.

Chapter 2

LITERATURE REVIEW

There has been work on optimal admission policies in production and service systems. In healthcare operations, patient flow can be represented by queuing models. Beds in ICUs act like servers. The focus of this study is on dynamic control of queuing systems using MDP models. I will review related literature about healthcare operations and Markovian queuing models.

2.1 Healthcare Operations

ICUs are among the busiest departments of hospitals. Admission and discharge processes are of great importance for the quality of treatment and efficient use of resources. Patients who are admitted to ICUs encounter transfer delays especially during busy periods. They unnecessarily occupy intensive care beds and thereby delay inpatient admissions due to bed shortage. Long and Mathews (2018) conduct an empirical analysis about prolonged boarding time of the ICU patients. Boarding time is the amount of time that patients physically remain in the ICU bed but no longer receive high-intensity services. Service means recovery in dedicated unit bed. According to their findings, occupancy levels do not affect service time. On the other hand, boarding time is prolonged when hospital occupancy levels are high.

Helm et al. (2011) study the hospital admission control policy by grouping patients into three categories called emergent, scheduled and expedited. Expedited patients are a type of emergent patient who are not really in an emergency. All patients share the same hospital beds and have the same service rate. Patients can enter to the hospital in three ways which are emergency, scheduled and expedited ways. Emergent patients are accepted immediately. Scheduled patients can be admitted or

rejected based on availability of beds. There is no queue for emergent and scheduled patients, while expedited patients are placed in a queue and the patients in the queue can be admitted to the hospital any time. This study reduces variability in hospital occupancy and its negative effects via characterizing an optimal admission threshold policy and controlling expedited patient queue. In Helm et al. model all patients share a ward meanwhile there are two wards in series in the models presented in this thesis. Another difference is that in Helm et al. (2011) all patients have the same service rate while in the models presented in this thesis, type-1 patients receive service with hypo-exponential rate and type-2 patients receive service with exponential rate.

Bekker et al. (2017) provide different policies for allocation of beds which consist of both flexibility and specialization. They compare the effectiveness of threshold control, earmarking, separate wards and simple merging policies. According to separate ward policy, each patient type has dedicated beds. In contrast, all beds are shared by all patient types in simple merging. Earmarking also provides dedicated beds to each patient type. In contrast to separate wards policy in earmarking, all patient types share a joint ward of overflow with fully flexible beds. In threshold policy, all beds are fully flexible however there are priorities among patient types. Each patient type is only admitted when the amount of free beds is more than the specified threshold. They show that simple merging is more beneficial by the economies of scale but it's difficult to manage especially in large systems. Earmarking policy is a good alternative for large systems with little flexibility. Threshold control is effective to prioritize each patient type with different length of stay.

2.2 Queuing Models

High setup and operational costs of healthcare coupled with insufficient resources result in high congestion in healthcare facilities. Under these circumstances, patients have to wait and form queues. This is the reason why queuing models are important to increase the efficiency of healthcare operations. Queuing models have different characteristics. There are single and multiple server cases. These models either have

a waiting room or do not. Multiple servers can be designed in parallel or tandem.

Xu et al. (1992) consider a system that has two stations each consisting of parallel servers. There are two types of customers who arrive according to independent and identically distributed Poisson processes and receive exponential service with a common rate. The objective is to minimize the expected discounted holding cost by dynamically assigning customers to idle servers. Class-2 customers can use both stations, but class-1 customers can use only station-1. Moreover, class-1 customers have a higher holding cost. When there is a free server, the optimal policy assigns class- j customers to station- j and a class-2 customer should be assigned to an idle server in station-1 only when no class-1 customers are waiting and the number of class-2 customers in line is more than a critical value. This value is increasing in the number of busy servers in the first station. Xu et al. (1992) investigates stations in parallel in their study whereas in this thesis stations in series are investigated. Moreover in this thesis, I also investigate the effect of blocking when station-2 is full.

Örmeci et al. (2001), Altman et al. (2001) and Savin et al. (2005) consider a loss system which consist of two classes of customers with different service rates. They prove that a threshold type policy is optimal. Örmeci et al. (2001) and Savin et al. (2005) show the of a preferred class under certain conditions. Örmeci et al. (2001) shows that the threshold is monotone under very restrictive conditions. Altman et al. (2001) demonstrate that the customer classes can be ordered under some circumstances. Altman et al. (2001) and Savin et al. (2005) present fluid models which work well for the large capacity cases. Savin et al. (2005) investigate the effect of capacity allocation on optimal fleet size. Örmeci et al. (2002) consider a similar model to Örmeci et al. (2001) but this time the rewards are non-negative random variables. Sufficient conditions to have preferred jobs imply that the system prefers to serve the faster jobs if both classes offer the same rewards. However, the criterion of average rewards favor slow jobs because they provide steady returns for longer periods. Örmeci and van der Wal (2006) generalize the Örmeci et al. (2001) when interarrival times have a general distribution. They prove the existence of an optimal threshold

policy. They were not able to prove the monotonicity of thresholds but they did not find a counterexample either. They stated that an optimal policy has to utilize all the servers in the system. They also prove that at least one of the classes is accepted in each state in a system consisting of two servers. They also provide examples where both classes are rejected in systems with more than two servers. In contrast to the model presented in this thesis, models mentioned in this paragraph consider a single ward. They only have a single stage service and they do not include blocking.

Ulukus et al. (2011) investigate a loss system consisting of two types of jobs with Poisson arrivals. Ulukus et al. (2011) consider accepting an arriving job by terminating a job of the other type in the system. They show the existence of optimal threshold policy for both admission and termination. They also present sufficient conditions which ensure the existence of preferred or strongly preferred classes. Kurt (2014) extends the model presented in Ulukus et al. (2011) and proposes a MDP model to find the optimal policy about controlling inpatient arrivals by admission, early discharge and overflow decisions. The system consists of two types of patients. Severe patients require two stage recovery. After the first stage recovery, they are classified as mild patients. There are also external mild patient arrivals and they discharge after the second stage recovery. The model assumes that there are two hospital wards which are the main ward and the overflow ward. Both patient types are served in the main ward, whereas the overflow ward only accommodates mild patients. She shows that early discharge action is not optimal unless all beds are occupied in the main ward. Since there is no early discharge action within models presented in this thesis, acceptance or rejection actions are taken instead of early discharge. Therefore, no strongly preferred class exists in these models while Kurt (2014) provides sufficient conditions to show the existence of preferred and strongly preferred classes. In contrast to the model presented in this thesis, the models mentioned in this paragraph include early discharge action and existence of strongly preferred classes. These models also do not consider blocking. In the model of Kurt (2014), availability of the overflow ward is decided by the optimal control policy meanwhile in the models presented in thesis

availability of the wards is permanent.

Ghoneim and Stidham Jr (1985) consider two queues in series. Each queue has one server. Type-1 customers enter the system through the first server. After the first stage service, they move to the second stage to complete their service. They need to receive service in the second server before departure. Type-2 customers are only served at the second server. Customers incur a higher holding cost in the first queue than the second. The objective is maximizing the average expected net benefit by accepting or rejecting arriving customers. They show that an optimal policy is monotone in the following senses: Adding a customer to either queue makes it less likely that a new customer will be accepted in both queues; moreover, moving a customer from the first queue to the second makes it more likely that they will accept a new customer into the first queue. When the queues from which the customer moves switches, the likelihood of acceptance decreases instead of increasing. In the model of Ghoneim and Stidham Jr (1985) there is a single server in each station in contrast to the models presented in this thesis with multi-server in each station. They also do not consider blocking.

Millhiser and Burnetas (2013) consider dynamic admission control of multiple job classes in a production system consisting of N stations in series with finite buffers. All jobs enter the system through the first station and must complete their service in all N stages before departure. Jobs are blocked in all stages when they complete their service in one station but there is no room in next one except for the last. They show that the net benefit of admitting a job declines monotonically with the system congestion. Conditions for the existence of a preferred class are derived. The interaction between blocking and admission control is explored by analyzing the effect of blocking on the optimal admission policy and profits. In the model of Millhiser and Burnetas (2013) all arrivals are received in the first stage meanwhile in the models presented in this thesis there are separate arrivals for each stage. In their model, a job needs to complete service in every stage before discharging from the system in contrast to models presented in this thesis. Furthermore, in their model blocked jobs

do not receive service while in model-1 service for blocked patients continue.

Ku and Jordan (1997) consider two queues in tandem with two types of customers. Each station has multiple servers. Type-1 customers are admitted to the first station while type-2 customers are admitted to the second. Type-2 customers receive service only at the second station. On the other hand, type-1 customers can be served at both servers. There is a positive probability that type-1 customers are lost while they are flowing to the second service. They show that the value function is submodular and concave on its second variable and type 1 customers are preferred in both stage 1 and 2. The optimal admission policy is given by a switching curve. Ku and Jordan (2002) present a system consisting of parallel multi-server loss queues in two stages. All stations in the system have external customer arrivals. In the first stage, there are N independent queues. After the first stage customers either leave the system or enter the second stage. In the second stage, all customers admitted from the first stage receive service by one station while rejected customers are lost. In the models presented in this thesis, blocked patients are not discharged from the system in contrast to the models mentioned in this paragraph. Moreover, in the models mentioned in the paragraph, customers bring reward while transferring from stage-1 to stage-2 while in the models presented in this thesis they do not.

Chapter 3

PROBLEM DESCRIPTION AND MATHEMATICAL MODEL

In this section, I provide two MDP models which represent a part of hospital consisting of an ICU and a ward in series. Thanks to these models, I aim to prevent prolonged boarding time of ICU patients and blocking to improve the efficiency of ICU usage. The system consists of two types of patients who are ICU patients and ward patients. Patients who need to complete ICU service before discharge are classified as ICU patients. ICU patients who already completed their service or patients who directly apply to the ward are classified as ward patients. They are represented as type-1, type-2 patients respectively. The models find an optimal admission control policy for the arrivals of type-1 and type-2 patients.

3.1 Model-1

In this section, I develop a MDP model consisting of two decisions which are admission and rejection. Model-1 is a two class loss system. There are two stations in series consisting of C_1 and C_2 servers. Patients arrive according to Poisson Processes with rates λ_1 and λ_2 . Type-1 patients enter the system through the first station. Recovery of type-1 patients is exponentially distributed with rate μ_1 . There is a positive probability $1 - P$ that type-1 patients die during service so, type-1 patients enter the second station with probability P . Service times of type-1 patients coming from the first station and type-2 patients admitted are exponentially distributed with rate μ_2 . Patients who are blocked after the first service are also type-2 patients. Blocked patients in the first station occupy ICU beds until they are admitted to the second station and receive service similar to the ward.

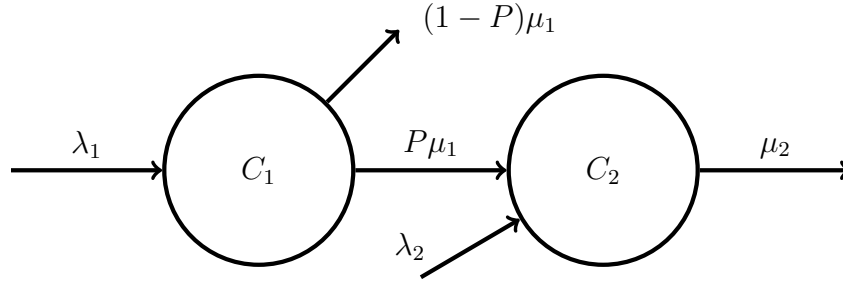


Figure 3.1: Arrival and departure processes

I denote the states as $\mathbf{x} = (x_1, x_2)$, where x_i is the number of type- i patients in the system. The state space is $S = \{x : 0 \leq x_1 \leq C_1, 0 \leq x_2, x_1 + x_2 \leq C_1 + C_2\}$.

I let α be the discount rate. Discounting is interpreted as exponential failures. That means the system closes down exponentially with rate α . Arrivals and departures occur in continuous time. Since the events of my model are independent and time intervals between them are exponentially distributed, the expected time until the next event is an exponential random variable with rate $\lambda_1 + \lambda_2 + \mu_1 x_1 + \mu_2 x_2 + \alpha$ in state (x_1, x_2) . I assume that service at the first station requires more time. Therefore, I can normalize the time unit so that $\lambda_1 + \lambda_2 + \mu_2(C_1 + C_2) + \alpha = 1$ and I can observe the system in exponential time intervals with rate 1 Lippman (1975). Decision epochs are the arrival times. $(C_2 + C_1 - x_2)\mu_2 - x_1\mu_1$ is the probability of fictitious events which are used to keep the rate constant in any state and do not cause any state change.

A period is the amount of time until the next event occurs. $V_n(x_1, x_2)$ is the maximum discounted cost for the system starting at state (x_1, x_2) and n periods remaining. I set $V_0(x_1, x_2) = 0$ for mathematical convenience, which does not affect the solution of the infinite horizon problem.

When type-1 or type-2 patients arrive to the system, they can either be accepted with a corresponding reward or be rejected. Let e_i denote the standard basis elements of \mathbb{R}^2 . When a type- i patient is admitted, the state moves from \mathbf{x} to $\mathbf{x} + e_i$ for $i = 1, 2$. If a type- i patient rejected, the state remains the same for $i = 1, 2$. Possible decisions

depend on the number of patients in the system. Four possible cases are explained below. Note that since $x_i \geq 0$, I set $x_i - 1$ to 0 whenever $x_i - 1 = -1$.

Case 1: $0 \leq x_1 < C_1, 0 \leq x_2 < C_2$

$$\begin{aligned}
V_{n+1}(x_1, x_2) = & \lambda_1 \max\{V_n(x_1, x_2), V_n(x_1 + 1, x_2) + R_1\} + \\
& \lambda_2 \max\{V_n(x_1, x_2), V_n(x_1, x_2 + 1) + R_2\} + \\
& Px_1\mu_1 V_n(x_1 - 1, x_2 + 1) + \\
& (1 - P)x_1\mu_1 V_n(x_1 - 1, x_2) + \\
& x_2\mu_2 V_n(x_1, x_2 - 1) + \\
& ((C_2 + C_1 - x_2)\mu_2 - x_1\mu_1)V_n(x_1, x_2)
\end{aligned} \tag{3.1}$$

Case 2: $0 \leq x_1 < C_1, x_2 \geq C_2, x_1 + x_2 < C_1 + C_2$

$$\begin{aligned}
V_{n+1}(x_1, x_2) = & \lambda_1 \max\{V_n(x_1, x_2), V_n(x_1 + 1, x_2) + R_1\} + \\
& \lambda_2 V_n(x_1, x_2) + \\
& Px_1\mu_1 V_n(x_1 - 1, x_2 + 1) + \\
& (1 - P)x_1\mu_1 V_n(x_1 - 1, x_2) + \\
& x_2\mu_2 V_n(x_1, x_2 - 1) + \\
& ((C_2 + C_1 - x_2)\mu_2 - x_1\mu_1)V_n(x_1, x_2)
\end{aligned} \tag{3.2}$$

Case 3: $0 \leq x_1, 0 < x_2, x_1 + x_2 = C_1 + C_2$

$$\begin{aligned}
V_{n+1}(x_1, x_2) = & \lambda_1 V_n(x_1, x_2) + \\
& \lambda_2 V_n(x_1, x_2) + \\
& Px_1\mu_1 V_n(x_1 - 1, x_2 + 1) + \\
& (1 - P)x_1\mu_1 V_n(x_1 - 1, x_2) + \\
& x_2\mu_2 V_n(x_1, x_2 - 1) + \\
& ((C_2 + C_1 - x_2)\mu_2 - x_1\mu_1)V_n(x_1, x_2)
\end{aligned} \tag{3.3}$$

Case 4: $x_1 = C_1, 0 < x_2 < C_2$

$$\begin{aligned}
V_{n+1}(x_1, x_2) = & \lambda_1 V_n(x_1, x_2) + \\
& \lambda_2 \max\{V_n(x_1, x_2), V_n(x_1, x_2 + 1) + R_2\} + \\
& P x_1 \mu_1 V_n(x_1 - 1, x_2 + 1) + \\
& (1 - P) x_1 \mu_1 V_n(x_1 - 1, x_2) + \\
& x_2 \mu_2 V_n(x_1, x_2 - 1) + \\
& ((C_2 + C_1 - x_2) \mu_2 - x_1 \mu_1) V_n(x_1, x_2)
\end{aligned} \tag{3.4}$$

In the first case, there are available beds in both the ICU and the ward. Thus, admitting and rejecting both arrivals is possible. In the second case, all ward beds are occupied so, admit action is not a possibility for type-2 patients anymore. In contrast, type-1 patients can be admitted to an empty bed in the ICU. In the third case, there is no empty bed in the system. For that reason, all of the arrivals have to be rejected. In the last case, type-1 patients use all of the ICU beds, so type-1 arrivals cannot be admitted. On the other hand, type-2 arrivals can be admitted to an empty bed in the ward.

3.2 Model-2

Items in production systems usually do not maintain their service when they are blocked. Although this is not the case for patient service, I also investigate the model when blocked patients do not maintain their service. In this section, a variation of model-1 in which the blocked patients are not served is presented. The state space of the model-2 is the same with model-1.

Case 1: $0 \leq x_1 < C_1, 0 \leq x_2 < C_2$

$$\begin{aligned}
V_{n+1}(x_1, x_2) = & \lambda_1 \max\{V_n(x_1, x_2), V_n(x_1 + 1, x_2) + R_1\} + \\
& \lambda_2 \max\{V_n(x_1, x_2), V_n(x_1, x_2 + 1) + R_2\} + \\
& Px_1\mu_1 V_n(x_1 - 1, x_2 + 1) + \\
& (1 - P)x_1\mu_1 V_n(x_1 - 1, x_2) + \\
& \min(x_2, C_2)\mu_2 V_n(x_1, x_2 - 1) + \\
& ((C_2 + C_1 - \min(x_2, C_2))\mu_2 - x_1\mu_1)V_n(x_1, x_2)
\end{aligned} \tag{3.5}$$

Case 2: $0 \leq x_1 < C_1, x_2 \geq C_2, x_1 + x_2 < C_1 + C_2$

$$\begin{aligned}
V_{n+1}(x_1, x_2) = & \lambda_1 \max\{V_n(x_1, x_2), V_n(x_1 + 1, x_2) + R_1\} + \\
& \lambda_2 V_n(x_1, x_2) + \\
& Px_1\mu_1 V_n(x_1 - 1, x_2 + 1) + \\
& (1 - P)x_1\mu_1 V_n(x_1 - 1, x_2) + \\
& \min(x_2, C_2)\mu_2 V_n(x_1, x_2 - 1) + \\
& ((C_2 + C_1 - \min(x_2, C_2))\mu_2 - x_1\mu_1)V_n(x_1, x_2)
\end{aligned} \tag{3.6}$$

Case 3: $0 \leq x_1, 0 < x_2, x_1 + x_2 = C_1 + C_2$

$$\begin{aligned}
V_{n+1}(x_1, x_2) = & \lambda_1 V_n(x_1, x_2) + \\
& \lambda_2 V_n(x_1, x_2) + \\
& Px_1\mu_1 V_n(x_1 - 1, x_2 + 1) + \\
& (1 - P)x_1\mu_1 V_n(x_1 - 1, x_2) + \\
& \min(x_2, C_2)\mu_2 V_n(x_1, x_2 - 1) + \\
& ((C_2 + C_1 - \min(x_2, C_2))\mu_2 - x_1\mu_1)V_n(x_1, x_2)
\end{aligned} \tag{3.7}$$

Case 4: $x_1 = C_1, 0 < x_2 < C_2$

$$\begin{aligned}
V_{n+1}(x_1, x_2) = & \lambda_1 V_n(x_1, x_2) + \\
& \lambda_2 \max\{V_n(x_1, x_2), V_n(x_1, x_2 + 1) + R_2\} + \\
& P x_1 \mu_1 V_n(x_1 - 1, x_2 + 1) + \\
& (1 - P) x_1 \mu_1 V_n(x_1 - 1, x_2) + \\
& \min(x_2, C_2) \mu_2 V_n(x_1, x_2 - 1) + \\
& ((C_2 + C_1 - \min(x_2, C_2)) \mu_2 - x_1 \mu_1) V_n(x_1, x_2)
\end{aligned} \tag{3.8}$$

There is only one difference between model-1 and model-2. All type-2 patients are serviced in model-1 whereas only C_2 number of type-2 patients are serviced at most and the others wait until there is an available bed in the ward in model-2.

3.3 Effect of an Additional Job

I define $D_n(ij)(\mathbf{x})$ as the difference in the total n -period discounted rewards between system A and system B if system A starts in state $\mathbf{x} + e_i$ and system B starts in $\mathbf{x} + e_j$ as presented in Örmeci et al. (2001). $D_n(10)$, $D_n(20)$, $D_n(12)$ and $D_n(21)$ are the four $D_n(ij)$ functions needed. These functions have the same definition in both models. $D_n(i0)$ is the expected net benefit of an additional type- i patient. Type- i patients are preferred if $D_n(i0) \geq -R_i$ for all $x \in S$ and all n .

Chapter 4

STRUCTURE OF THE OPTIMAL POLICIES

4.1 Preliminary Results for Model-1

Having one more type- i patient in the system reduces the expected benefit. Thus, the difference between the two systems caused by an additional type- i patient is always negative.

Theorem 1: For all \mathbf{x} with $\mathbf{x} + e_i \in S$, the value function is monotonically decreasing in x_i , *i.e.*, $V_n(\mathbf{x} + e_i) - V_n(\mathbf{x}) \leq 0$ for $i = 1, 2$.

Proof: I use a sample path argument to prove this statement. Let system A be in state $(\mathbf{x} + e_i)$ and system B be in state (\mathbf{x}) in period n . Then I couple them, except for the additional patient in System A. System A follows the optimal policy and System B follows the actions of System A, which is always possible. This policy is represented by π . Then, the difference in the expected returns of the two systems is only due to the additional patient in System A:

$$V_n(\mathbf{x} + e_i) - V_n(\mathbf{x}) \leq V_n(\mathbf{x} + e_i) - V_n^\pi(\mathbf{x}) = 0$$

For $i = 2$, when the additional type-2 patient in system A recovers, then the two systems couple. That means either the rewards of two systems are equal or system A has less. For $i = 1$, if the additional type-1 patient dies after the first stage, system A and B couple immediately. For $i = 1$, while system B is in state (x_1, x_2) , after the first stage recovery the additional type-1 patient in system A becomes a type-2 patient and system A moves into state $(x_1, x_2 + 1)$. This corresponds to the case with $i = 2$. Therefore, after all arrivals and service completions, $V_n(\mathbf{x} + e_i) - V_n(\mathbf{x}) \leq 0$ is true. ■

The following theorem shows the upper bound on the benefit of having one more type-1 patient instead of type-2 patient to the system.

Theorem 2: For all \mathbf{x} with $\mathbf{x} + e_i \in S$, $D_n(12)(\mathbf{x}) = V_n(\mathbf{x} + e_1) - V_n(\mathbf{x} + e_2) \leq R_2$.

Proof: This theorem is also proven via sample path argument. Let system A start in state $(\mathbf{x} + e_1)$ and system B start in state $(\mathbf{x} + e_2)$. Assume that system A follows the optimal policy and system B imitates the actions of system A which is policy π . For $x_2 + 1 < C_2$ and $x_2 \geq C_2$ System B can always imitate the actions of system A so, the difference in revenues due to arrivals remain the same as:

$$V_{n-1}(\mathbf{x} + e_1) - V_{n-1}(\mathbf{x} + e_2) \leq V_{n-1}(\mathbf{x} + e_1) - V_{n-1}^\pi(\mathbf{x} + e_2) = 0$$

For $x_2 + 1 = C_2$ system A may accept an incoming type-2 patient while system B has to reject. Then the difference between the two systems becomes $R_2 + D_n(10)(\mathbf{x})$. $D_n(10)(\mathbf{x})$ has the upper bound 0. Therefore:

$$V_{n-1}(\mathbf{x} + e_1) - V_{n-1}(\mathbf{x} + e_2) \leq V_{n-1}(\mathbf{x} + e_1) - V_{n-1}^\pi(\mathbf{x} + e_2) \leq R_2$$

With probability $(1 - P)\mu_1$, the additional type-1 patient dies and the additional type-2 patient in system B leaves the hospital. Then the difference between the two systems becomes zero. When the additional type-1 patient in system A completes the first stage recovery and the additional type-2 patient completes the service with probability $P\mu_1$, the difference becomes $V_{n-1}(\mathbf{x} + e_2) - V_{n-1}(\mathbf{x}) \leq 0$ by Theorem 1. With probability $\mu_2 - \mu_1$, the additional type-2 patient leaves the system while the additional type-1 patient continues his service. In this case, the difference becomes $V_{n-1}(\mathbf{x} + e_1) - V_{n-1}(\mathbf{x}) \leq 0$ by Theorem 1. Therefore, in all service completions, Theorem 2 holds. ■

In this model there is no monotonicity property for $D_n(12)(\mathbf{x})$ and $D_n(21)(\mathbf{x})$ as provided by Kurt (2014). Since in Kurt's model, type-1 and type-2 patients share the same ward, the additional type-1 patient in system A compensates the additional type-2 patient in system B, so system B can always imitate the actions of system

A. Kurt's model does not have a restriction that is similar to the above case when $x_2 + 1 = C_2$.

4.2 Preliminary Results for Model-2

Model-2 generally yields to similar results to those for model-1. However, there are certain differences as well. For example, Theorem 4 can be proven under the assumption $P = 1$.

Having one more type- i patient in the system reduces the expected benefit. Thus, the difference between the two systems caused by an additional type- i patient is always negative.

Theorem 3: For all \mathbf{x} with $\mathbf{x} + e_i \in S$, the value function is monotonically decreasing in x_i , i.e., $V_n(\mathbf{x} + e_i) - V_n(\mathbf{x}) \leq 0$ for $i = 1, 2$.

Proof: For $x_2 < C_2$, the proof is identical to the proof of Theorem 1 of **Section 4.1**. Since system B can always follow the actions of system A, the difference due to arrival remains the same. For $x_2 \geq C_2$ and $i = 2$, the additional type-2 patient in system A does not recover and the difference remains the same. The rest of the proof is again identical to the proof of Theorem 1. ■

The following theorem shows the upper bound on the benefit of having one more type-1 patient instead of type-2 patient to the system.

Theorem 4: For all $n \geq 0$, if $P = 1$ for all \mathbf{x} with $\mathbf{x} + e_i \in S$, $D_n(12)(\mathbf{x}) = V_n(\mathbf{x} + e_1) - V_n(\mathbf{x} + e_2) \leq R_2$.

Proof: The only difference from the proof of Theorem 2 in **Section 4.1** is service completion of type-2 patients in the case $x_2 \geq C_2$. Therefore I only consider service completions of that case here. If the additional type-1 patient in system A completes the first stage recovery and the additional type-2 patient remains with probability $P\mu_1 = \mu_1$ the difference becomes zero. With probability $\mu_2 - \mu_1$ both patients continue their service. Then the difference remains the same in all arrivals and service

completions. Therefore, Theorem 4 holds. ■

With probability $(1 - P)\mu_1$, the additional type-1 patient dies and the additional type-2 patient in system B remains. Since I assume $P = 1$ this never occurs and I can skip this case. If P had been smaller than 1, then the difference between two systems becomes $V_{n-1}(\mathbf{x}) - V_{n-1}(\mathbf{x} + e_2)$. This difference makes the original proof in this case incorrect. Millhisser and Burnetas (2013) state that mid-production abandonment violates the condition needed for the monotonicity properties. That is why I add the additional assumption of $P = 1$ to prove the theorem for the ideal case with no deaths.

4.3 Existence of a Preferred Class for Model-1 and Model-2

In this section, I show that under certain conditions, type- i patients are always admitted to the system whenever there is an available bed. This type is named as preferred class.

For a class to be preferred, its reward must be greater than its burden on future rewards in all states. While investigating sufficient conditions for being preferred, cases with $x_1 < C_1$ and $x_2 < C_2$ are considered for type-1 and type-2 patients respectively. When $x_2 < C_2$, there is no blocking effect so that Model-1 and Model-2 behave similarly. Therefore, the sufficient conditions for each class to be preferred, as well as their derivation, are the same in both models.

Let d_i be a lower bound on $D_n(i0)(\mathbf{x})$ for all $\mathbf{x} \in S$, $n \geq 0$ and $i = 1, 2$ which can be represented as $d_i \leq \min_{\mathbf{x} \in S} D_n(i0)(\mathbf{x})$. d_{ij} is defined as the lower bound on $D_n(ij)(\mathbf{x})$ for all $\mathbf{x} \in S$, $n \geq 0$ and $(ij) \in \{(12), (21)\}$. According to Theorem 2 and Theorem 4, the lower bound on d_{21} is set to $-R_2$.

4.3.1 A sufficient condition for type-2 to be preferred

In this section, I present a theorem which demonstrates the sufficient conditions for type-2 patients to be preferred by establishing a lower bound on $D_n(20)$. Let system A be in state $(\mathbf{x} + e_2)$, system B in state (\mathbf{x}) and they are coupled. System B follows the optimal policy and system A follows the policy π which consists of same actions as long as $x_2 < C_2 - 1$ or $x_2 \geq C_2, x_1 + x_2 + 1 < C_1 + C_2$. In this scenario, the difference between two systems is bounded by d_2 . When $x_2 + 1 = C_2$ and a type-2 patient arrives to the system A, he has to be rejected. On the other hand, system B may accept him and move to the state $(\mathbf{x} + e_2)$. Then, two systems couple and the difference between them becomes $-R_2$. For $x_2 \geq C_2, x_1 + x_2 + 1 = C_1 + C_2$ if a type-1 patient arrives system B accepts it and moves to state $(x_1 + 1, x_2)$ and system A has to reject that arrival and stays in state $(x_1, x_2 + 1)$. Then, the difference between two systems is bounded by $-R_1 + d_{21}$. Finally, I give an expression for the lower bound on $V_n(\mathbf{x} + e_2) - V_n(\mathbf{x})$:

$$\begin{aligned}
D_n(20)(\mathbf{x}) &= V_n(\mathbf{x} + e_2) - V_n(\mathbf{x}) \geq V_n^\pi(\mathbf{x} + e_2) - V_n(\mathbf{x}) \\
&\geq \lambda_1 \min\{D_{n-1}(21) - R_1, D_{n-1}(20)\} + \lambda_2 \min\{-R_2, D_{n-1}(20)\} \\
&\quad + P\mu_1 x_1 (V_{n-1}(x_1 - 1, x_2 + 2) - V_{n-1}(x_1 - 1, x_2 + 1)) \\
&\quad + (1 - P)\mu_1 x_1 (V_{n-1}(x_1 - 1, x_2 + 1) - V_{n-1}(x_1 - 1, x_2)) \\
&\quad + \mu_2 x_2 (V_{n-1}(x_1, x_2) - V_{n-1}(x_1, x_2 - 1)) \\
&\quad + \mu_2 (V_{n-1}(x_1, x_2) - V_{n-1}(x_1, x_2)) \\
&\quad + ((C_2 + C_1 - x_2 - 1)\mu_2 - x_1\mu_1)(V_{n-1}(x_1, x_2 + 1) - V_{n-1}(x_1, x_2)) \\
&\geq \lambda_1 \min\{d_{21} - R_1, d_2\} + \lambda_2 \min\{d_2, -R_2\} + (C_1 + C_2 - 1)\mu_2 d_2
\end{aligned} \tag{4.1}$$

Note that $d_2 < 0$ due to Theorem 1 and Theorem 3. I find the first inequality via the policy which system B follows, the second inequality is written as a result of coupling and the last one is driven by the definition of d_2 and d_{21} . Therefore, I construct the following inequality which guarantees that d_2 is a lower bound on

$D_n(20)(\mathbf{x})$ for all $\mathbf{x} \in S$:

$$\lambda_1 \min\{d_{21} - R_1, d_2\} + \lambda_2 \min\{d_2, -R_2\} + (C_1 + C_2 - 1)\mu_2 d_2 \geq d_2 \quad (4.2)$$

I simplify the given inequality via uniformization i.e., $\lambda_1 + \lambda_2 + (C_1 + C_2)\mu_2 + \alpha = 1$, then the inequality becomes:

$$\lambda_1 \min\{d_{21} - R_1, d_2\} + \lambda_2 \min\{d_2, -R_2\} \geq d_2(\lambda_1 + \lambda_2 + \mu_2 + \alpha) \quad (4.3)$$

I remove d_{21} from the given inequality above by replacing its lower bound which is $-R_2$. The inequality becomes:

$$\lambda_1 \min\{-R_1 - R_2, d_2\} + \lambda_2 \min\{d_2, -R_2\} \geq d_2(\lambda_1 + \lambda_2 + \mu_2 + \alpha) \quad (4.4)$$

In the following part of this section I derive d_2 for all cases.

Case 1: $d_2 \geq -R_2$

From the inequality (4.4):

$$\frac{-\lambda_1(R_1 + R_2) - \lambda_2 R_2}{\lambda_1 + \lambda_2 + \mu_2 + \alpha} \geq d_2, \quad (4.5)$$

From which I set:

$$d_2 = \frac{-\lambda_1(R_1 + R_2) - \lambda_2 R_2}{\lambda_1 + \lambda_2 + \mu_2 + \alpha} \quad (4.6)$$

d_2 is a valid lower bound as long as

$$\frac{\lambda_1}{\mu_2} \leq \frac{R_2}{R_1}$$

Case 2: $-R_2 \geq d_2 \geq -(R_1 + R_2)$ By using inequality (4.4) and case conditions, I set:

$$d_2 = \frac{-\lambda_1(R_1 + R_2)}{\lambda_1 + \mu_2 + \alpha}, \quad (4.7)$$

$$\text{If } \frac{R_2}{R_1} \leq \frac{\lambda_1}{\mu_2} \text{ holds,}$$

d_2 is a valid lower bound.

When the case $d_2 > -R_1 - R_2$ is checked, it is observed that, this case always holds. In the case $R_1 - R_2 > d_2$, since R_1 and R_2 are strictly positive, $d_2 = 0$ leads to a contradiction.

Lemma 1: For all $\mathbf{x} \in S$ and $n \geq 0$

If $\frac{R_2}{R_1} \leq \frac{\lambda_1}{\mu_2}$ holds, $D_n(20)(\mathbf{x}) \geq \frac{-\lambda_1(R_1+R_2)}{\lambda_1+\mu_2}$ is valid.

If $\frac{R_2}{R_1} \geq \frac{\lambda_1}{\mu_2}$ holds, $D_n(20)(\mathbf{x}) \geq \frac{-\lambda_1(R_1+R_2)-\lambda_2 R_2}{\lambda_1+\lambda_2+\mu_2}$ is valid.

Proof: I prove Lemma 1 by induction. For $n = 0$, $D_0(20)(\mathbf{x}) = 0$. The rest of the proof can be found via following the steps above. ■

Theorem 5: For all $n \geq 0$, if $R_2\mu_2 \geq R_1\lambda_1$, then, type-2 patients are preferred.

Proof: For a class to be preferred, its reward must be greater than its burden on future rewards. For type-2 to be preferred, R_2 has to be greater than or equal to $-d_2$. Therefore, only case 1, where d_2 is given by equation (4.6), can be used to determine that type-2 is a preferred class. ■

4.3.2 A sufficient condition for type-1 to be preferred

In this section I derive the lower bound on $D_n(10)(\mathbf{x})$. Then I conclude with a theorem, which represents sufficient conditions for type-1 patients to be preferred.

Assume that system A is in the state $(\mathbf{x} + e_1)$ and system B is in state (\mathbf{x}) . I couple system A and system B. Assume that system B follows the optimal policy and system A imitates the decisions of system B, which is always possible, except for the case where system A does not have an empty bed in the ICU, i.e. $x_1 + 1 = C_1$ or $x_1 + x_2 + 1 = C_1 + C_2$. Then, the difference between two systems is bounded by d_1 .

In the case $x_1 + 1 = C_1$ or $x_1 + x_2 + 1 = C_1 + C_2$ when a type-1 patient arrives, if system B accepts the incoming patient, system A rejects the patient. Then, system A and system B couple hence system B has an extra reward of R_1 . According to

information given, I give an expression for the lower bound on $V_n(\mathbf{x} + e_1) - V_n(\mathbf{x})$:

$$\begin{aligned}
D_n(10)(\mathbf{x}) &= V_n(\mathbf{x} + e_1) - V_n(\mathbf{x}) \geq V_n^\pi(\mathbf{x} + e_1) - V_n(\mathbf{x}) \\
&\geq \lambda_1 \min\{-R_1, D_{n-1}(10)\} + \lambda_2 D_{n-1}(10) \\
&\quad + P\mu_1 x_1 (V_{n-1}(x_1, x_2 + 1) - V_{n-1}(x_1 - 1, x_2 + 1)) \\
&\quad + (1 - P)\mu_1 x_1 (V_{n-1}(x_1, x_2) - V_{n-1}(x_1 - 1, x_2)) \\
&\quad + \mu_2 x_2 (V_{n-1}(x_1 + 1, x_2 - 1) - V_{n-1}(x_1, x_2 - 1)) \\
&\quad + P\mu_1 (V_{n-1}(x_1, x_2 + 1) - V_{n-1}(x_1, x_2)) \\
&\quad + (1 - P)\mu_1 (V_{n-1}(x_1, x_2) - V_{n-1}(x_1, x_2)) \\
&\quad + ((C_2 + C_1 - x_2)\mu_2 - (x_1 + 1)\mu_1) (V_{n-1}(x_1 + 1, x_2) - V_{n-1}(x_1, x_2)) \\
&\geq \lambda_1 \min\{-R_1, d_1\} + \lambda_2 d_1 + P\mu_1 d_2 + ((C_1 + C_2)\mu_2 - \mu_1) d_1
\end{aligned} \tag{4.8}$$

The first inequality comes from the policy system A follows. The second inequality follows via coupling and the last one is derived by the definition of d_1 and d_2 . The following inequality guarantees that d_1 is a lower bound on $D_n(10)(\mathbf{x})$ for all $\mathbf{x} \in S$:

$$d_1 \leq \lambda_1 \min\{-R_1, d_1\} + \lambda_2 d_1 + P\mu_1 d_2 + ((C_1 + C_2)\mu_2 - \mu_1) d_1 \tag{4.9}$$

The inequality above is simplified via uniformization i.e., $(C_1 + C_2)\mu_2 + \lambda_1 + \lambda_2 + \alpha = 1$, then the inequality becomes:

$$(\lambda_1 + \mu_1 + \alpha) d_1 \leq P\mu_1 d_2 + \lambda_1 \min\{-R_1, d_1\} \tag{4.10}$$

I remove d_2 from the inequality above by replacing it by its value.

Case 1.a: $d_2 \geq -R_2$, $-R_1 \leq d_1$

$$(\lambda_1 + \mu_1 + \alpha) d_1 \leq \frac{P\mu_1(-\lambda_1(R_1 + R_2) - \lambda_2 R_2)}{\lambda_1 + \lambda_2 + \mu_2} - \lambda_1 R_1 \tag{4.11}$$

From which I set:

$$d_1 = \frac{-P\mu_1(\lambda_1(R_1 + R_2) + \lambda_2 R_2)}{(\lambda_1 + \mu_1 + \alpha)(\lambda_1 + \lambda_2 + \mu_2)} - \frac{\lambda_1 R_1}{(\lambda_1 + \mu_1 + \alpha)} \tag{4.12}$$

d_1 is a valid lower bound as long as

$$\frac{\lambda_1(1 - P) + \lambda_2 + \mu_2}{(\lambda_1 + \lambda_2)P} \geq \frac{R_2}{R_1} \geq \frac{\lambda_1}{\mu_2}$$

Case 1.b: $-R_2 \geq d_2 \geq -(R_1 + R_2)$, $-R_1 \leq d_1$

$$(\lambda_1 + \mu_1 + \alpha)d_1 \leq \frac{P\mu_1(-\lambda_1(R_1 + R_2))}{\lambda_1 + \mu_2} - \lambda_1 R_1 \quad (4.13)$$

From which I set:

$$d_1 = \frac{P\mu_1(-\lambda_1(R_1 + R_2))}{(\lambda_1 + \mu_1 + \alpha)(\lambda_1 + \mu_2)} - \frac{\lambda_1 R_1}{(\lambda_1 + \mu_1 + \alpha)} \quad (4.14)$$

d_1 is a valid lower bound as long as

$$\min\left(\frac{\lambda_1(1-P) + \mu_2}{\lambda_1 P}, \frac{\lambda_1}{\mu_2}\right) \geq \frac{R_2}{R_1}$$

Case 2.a: $d_2 \geq -R_2$, $-R_1 > d_1$

$$(\lambda_1 + \lambda_2 + \mu_1 + \alpha)d_1 \leq \mu_1 d_2 + \lambda_1 d_1 + \lambda_2 d_1 \quad (4.15)$$

$$d_1 \leq d_2 \quad (4.16)$$

From which I set

$$d_1 = \frac{-P(\lambda_1(R_1 + R_2) + \lambda_2 R_2)}{\lambda_1 + \lambda_2 + \mu_2 + \alpha} \quad (4.17)$$

d_1 is a valid lower bound as long as

$$\max\left(\frac{\lambda_1(1-P) + \lambda_2 + \mu_2}{(\lambda_1 + \lambda_2)P}, \frac{\lambda_1}{\mu_2}\right) \leq \frac{R_2}{R_1}$$

Case 2.b: $-R_1 \geq d_2 \geq -R_2$, $-R_1 > d_1$

$$(\lambda_1 + \lambda_2 + \mu_1 + \alpha)d_1 \leq \mu_1 d_2 + \lambda_1 d_1 + \lambda_2 d_1 \quad (4.18)$$

$$d_1 \leq d_2 \quad (4.19)$$

From which I set

$$d_1 = \frac{-\lambda_1 P(R_1 + R_2)}{\lambda_1 + \mu_2 + \alpha} \quad (4.20)$$

d_1 is a valid lower bound as long as

$$\frac{\lambda_1(1-P) + \mu_2}{\lambda_1 P} \leq \frac{R_2}{R_1} \leq \frac{\lambda_1}{\mu_2}$$

Theorem 6: For all $n \geq 0$, if $\mu_2 \leq \lambda_1 P$, then type-1 is preferred whenever $R_1 \geq \frac{R_2 \lambda_1 P}{\lambda_1(1-P)+\mu_2}$; if $\mu_2 > \lambda_1 P$, then type-1 is preferred whenever $R_1 \geq \frac{R_2(\lambda_1+\lambda_2)P}{\lambda_1(1-P)+\lambda_2+\mu_2}$.

Proof: For a class to be preferred, its reward must be greater than its burden on future rewards. For type-1 to be preferred class, R_1 has to be greater than or equal to $-d_1$. Therefore only case 1 (4.12-4.14) is used to determine that type-1 is a preferred class. ■

4.4 Variation of Model-1 Including k Types of ICU Patients

In this section, a variation of model-1 including k types of ICU patients is provided, and type $k + 1$ represents type-2 in the original model. $R_1 > R_2 > \dots > R_{k+1}$ is assumed.

Case 1: $0 \leq x_1 < C_1, 0 \leq x_2 < C_2$

$$\begin{aligned}
V_{n+1}(x_1, x_2) = & \sum_{i=1}^k \lambda_i \max\{V_n(x_1, x_2), V_n(x_1 + 1, x_2) + R_i\} + \\
& \lambda_{k+1} \max\{V_n(x_1, x_2), V_n(x_1, x_2 + 1) + R_{k+1}\} + \\
& P x_1 \mu_1 V_n(x_1 - 1, x_2 + 1) + \\
& (1 - P) x_1 \mu_1 V_n(x_1 - 1, x_2) + \\
& x_2 \mu_2 V_n(x_1, x_2 - 1) + \\
& ((C_2 + C_1 - x_2) \mu_2 - x_1 \mu_1) V_n(x_1, x_2)
\end{aligned} \tag{4.21}$$

Case 2: $0 \leq x_1 < C_1, x_2 \geq C_2, x_1 + x_2 < C_1 + C_2$

$$\begin{aligned}
V_{n+1}(x_1, x_2) = & \sum_{i=1}^k \lambda_i \max\{V_n(x_1, x_2), V_n(x_1 + 1, x_2) + R_i\} + \\
& \lambda_{k+1} V_n(x_1, x_2) + \\
& P x_1 \mu_1 V_n(x_1 - 1, x_2 + 1) + \\
& (1 - P) x_1 \mu_1 V_n(x_1 - 1, x_2) + \\
& x_2 \mu_2 V_n(x_1, x_2 - 1) + \\
& ((C_2 + C_1 - x_2) \mu_2 - x_1 \mu_1) V_n(x_1, x_2)
\end{aligned} \tag{4.22}$$

Case 3: $0 \leq x_1, 0 < x_2, x_1 + x_2 = C_1 + C_2$

$$\begin{aligned}
V_{n+1}(x_1, x_2) = & \sum_{i=1}^k \lambda_i V_n(x_1, x_2) + \\
& \lambda_{k+1} V_n(x_1, x_2) + \\
& P x_1 \mu_1 V_n(x_1 - 1, x_2 + 1) + \\
& (1 - P) x_1 \mu_1 V_n(x_1 - 1, x_2) + \\
& x_2 \mu_2 V_n(x_1, x_2 - 1) + \\
& ((C_2 + C_1 - x_2) \mu_2 - x_1 \mu_1) V_n(x_1, x_2)
\end{aligned} \tag{4.23}$$

Case 4: $x_1 = C_1, 0 < x_2 < C_2$

$$\begin{aligned}
V_{n+1}(x_1, x_2) = & \sum_{i=1}^k \lambda_i V_n(x_1, x_2) + \\
& \lambda_{k+1} \max\{V_n(x_1, x_2), V_n(x_1, x_2 + 1) + R_{k+1}\} + \\
& P x_1 \mu_1 V_n(x_1 - 1, x_2 + 1) + \\
& (1 - P) x_1 \mu_1 V_n(x_1 - 1, x_2) + \\
& x_2 \mu_2 V_n(x_1, x_2 - 1) + \\
& ((C_2 + C_1 - x_2) \mu_2 - x_1 \mu_1) V_n(x_1, x_2)
\end{aligned} \tag{4.24}$$

The cases of the model and the proofs of theorems are generally similar to those for model-1. Hence, I provide only the statements and omit the proofs. I also exclude the result on the conditions for ICU patients to be preferred, since these conditions are too complicated and can be derived only for the class with the highest revenue, i.e., class-1 ICU patients.

Theorem 7: For all \mathbf{x} with $\mathbf{x} + e_i \in S$, the value function is monotonically decreasing in x_i , i.e., $V_n(\mathbf{x} + e_i) - V_n(\mathbf{x}) \leq 0$ for $i \in \{1, \dots, k + 1\}$.

Theorem 8: For all \mathbf{x} with $\mathbf{x} + e_i \in S$, $V_n(\mathbf{x} + e_i) - V_n(\mathbf{x} + e_{k+1}) \leq R_{k+1}$ for $i \in \{1, \dots, k\}$.

Theorem 9: For all $n \geq 0$, if $R_{k+1} \mu_2 \geq \sum_{i=1}^k R_i \lambda_i$, then, type-2 patients are preferred.

Chapter 5

NUMERICAL ANALYSIS

5.1 *Parameter Estimation*

In this chapter I implement model-1 and model-2 presented in Chapter 3. The parameters used to construct the base case are taken from de Bruin et al. (2010). The data used has been collected from 24 units of a university medical center, including both intensive care and normal care between 2004-2006. The data has been used to calculate number of daily admissions, average length of stay and the number of beds for each unit. Data includes scheduled and unscheduled arrivals. If the data is separated by weekday arrivals and weekend arrivals, arrivals of half of the wards fit the exponential distribution.

Post-op patients are patients who have undergone surgery. Post-op patients usually spend a single day in ICU before being moved to a normal ward. Therefore, their length of stay does not fit the exponential distribution and they require less attention and resource. Due to these, it is reasonable to partition ICU patients into post-op patients and regular patients. Intensive care unit medical has chosen as the ICU in the numerical analysis. Hematology, internal medicine unit-1 and, internal medicine unit-2 departments have been combined to create the ward so that the amount of post-op patients included is minimized.

According to Table 5.1 the ward consists of 61 beds. Number of daily arrivals of the ward is the sum of daily arrivals of hematology and internal medicine units which is 14.64. This number is the number of external arrivals. The patients who are transferred from the ICU are not included in it. I find the average length of the stay of the ward by weighted averaging the average length of stays of hematology and internal medicine units with the number patients used as weights. It is calculated to

Departments	Number of Operational Beds	Number of Daily Arrivals	Average Length of Stay (Days)	σ
ICU Medical	14	2.14	5.147	0.3734
Hematology	21	7.57	2.763	0.7146
Internal Medicine Unit-1	20	2.9	6.354	0.1961
Internal Medicine Unit-2	20	4.17	4.852	0.3027

Table 5.1: Parameters taken from de Bruin et al. (2010), with presented data from 2006

be 4.0694.

According to Halpern and Pastores (2015) ICU costs per day vary with ICU types, patient types and the length of stay of patients. ICU to non-ICU daily cost ratio changes between 2 to 5. For the last four decades, the ratio is usually assumed to be 3. Russell equation is used to find the national critical care medicine costs in US by multiplying the ICU cost per day with number of ICU days per year. Pastores et al. (2012) also assumed the ratio of ICU to non-ICU costs to be 3 while calculating the critical care medicine costs for their study. According to data shared by Canadian Institute for Health Information, in 2013-2014 staying at ICU costs about 2 times more than staying at non-ICU units in Canada. Due to these, the cost ratio between ICU and non-ICU is assumed to be 3 in this research. I use the cost ratio as a proxy to determine the rewards. Different values of this cost ratio will be implemented in the sensitivity analysis. The rewards are determined by multiplying the daily cost of stay with the average length of stay. I assume the unit cost for the wards.

$$R_2 = \frac{1}{\mu_2}$$

ICU patients receive service from both of the departments, each with their respective average length of stay. Rewards are calculated separately by the same method with R_2 for each stage and then summed to find R_1 . For the first stage the reward is $3 \times \frac{1}{\mu_1}$

and for the second stage it is $1 \times \frac{1}{\mu_2}$. Therefore

$$R_1 = \frac{3}{\mu_1} + \frac{1}{\mu_2}$$

Siddiqui (2015) presents data collected between 2010-2015 in a large public hospital in Singapore. It includes mortality results of medical, surgical and coronary intensive care units. According to this study, the average mortality rate of medical intensive care units is 7%. Based on this, I assume the survival rate in the models to be 93%.

Notation	Parameters	Estimated Value
λ_1	Arrival rate of x_1	2.14
λ_2	Arrival rate of x_2	14.64
$1/\mu_1$	Average length of stay of x_1	5.147
$1/\mu_2$	Average length of stay of x_2	4.0694
C_1	Total number of beds in station-1	14
C_2	Total number of beds in station-2	61
R_1	Accepting reward of x_1	17.1364
R_2	Accepting reward of x_2	4.0694
α	Discount rate	0.9
P	Probability of survival	0.93

Table 5.2: Base Case Parameters

Table 5.2 demonstrates the parameters used to construct the base case.

5.2 Base Case

The models are solved by the value iteration algorithm. Figures 5.1 and 5.2 demonstrate how the output of the value function, total expected discounted reward, changes with x_1 and x_2 .

As proved in Theorem 1, the output of the value function is monotonically non-increasing when x_1 and x_2 increase.

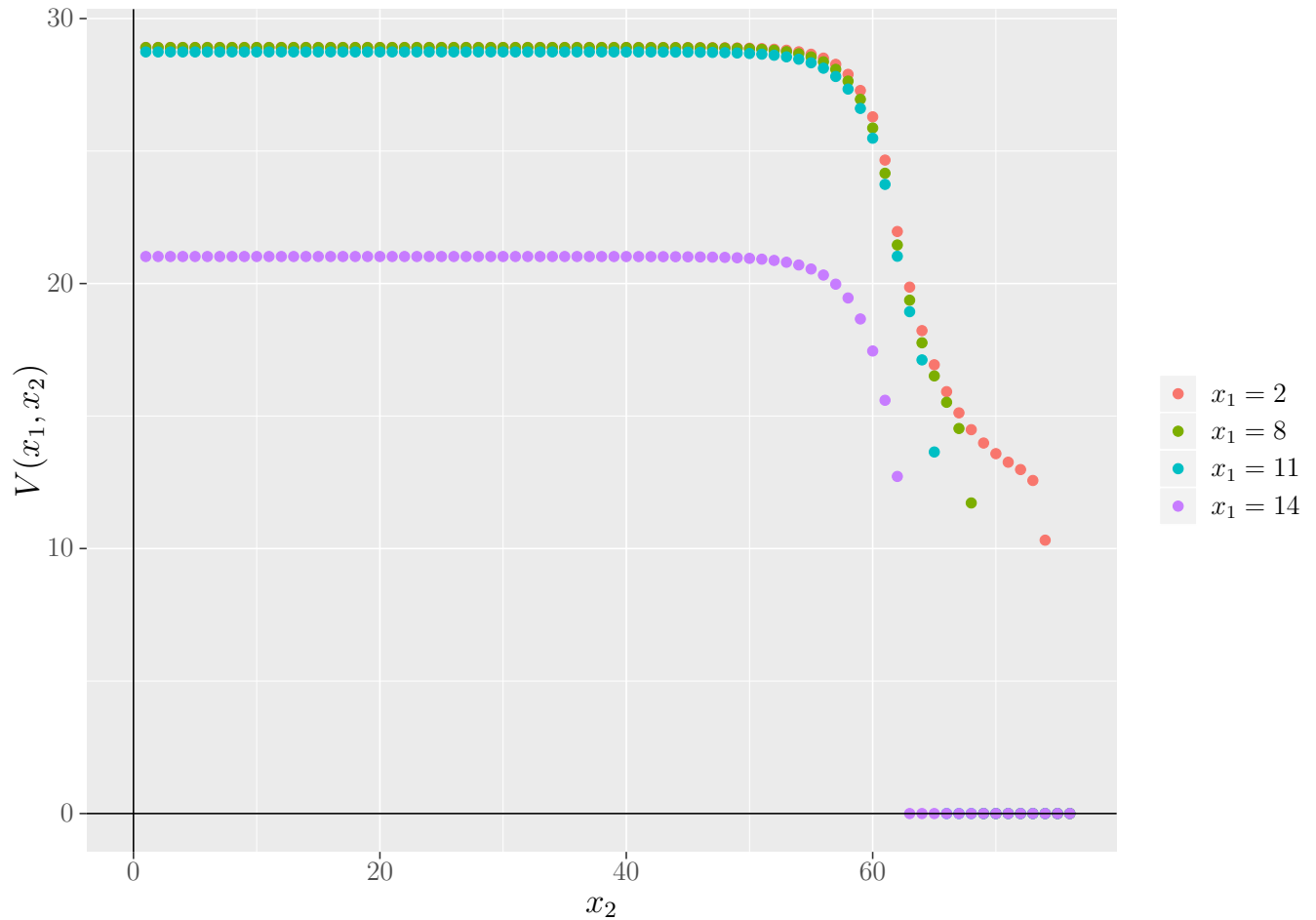
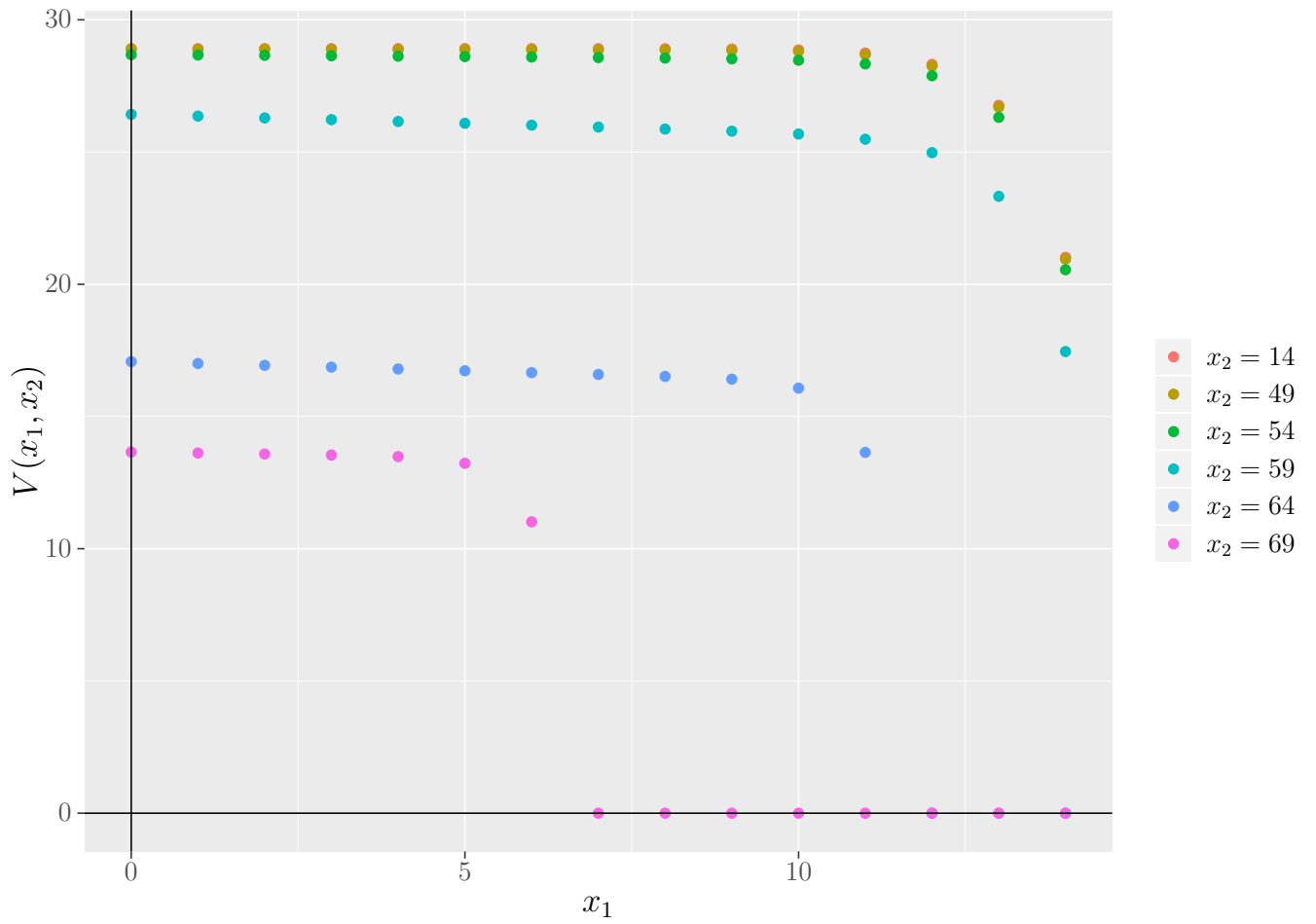


Figure 5.1: Changes in the value function for certain x_1

Figure 5.3 shows the difference between having an additional type-1 and having an additional type-2 patient can be at most R_2 as proven in Theorem 2.

Table 5.3 demonstrates the difference in the value function, total expected discounted reward, between model-1 and model-2 for certain states. Difference in the omitted states are approximately 0. There is a significant difference between the models only in states close to having blocked patients. This can be understood intuitively by considering that when $x_2 \leq C_2$, model-2 behaves similarly to model-1 due to the low risk of having blocked patients. The effect of having blocked patients when $x_2 > C_2$ can be observed in model-2.

Figure 5.2: Changes in the value function for certain x_2

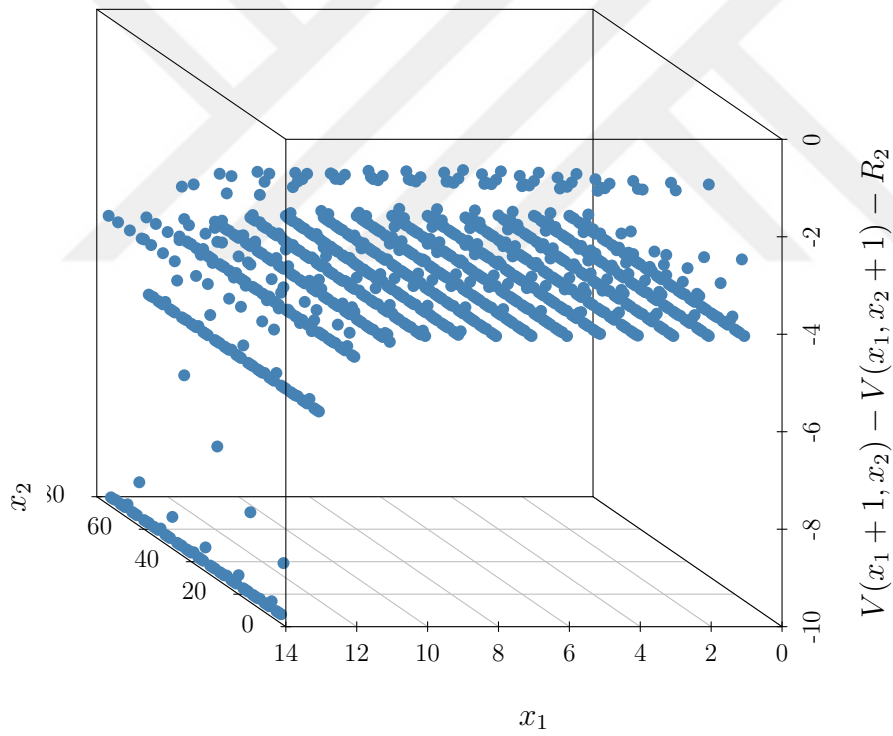


Figure 5.3: An example of theorem 2 for the base case of model-1

x_2, x_1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
59															0.01
60														0.01	0.01
61										0.01	0.01	0.01	0.01	0.01	0.01
62	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.04	0.04	0.04	0.07	
63	0.06	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.09	0.14		
64	0.10	0.10	0.10	0.10	0.10	0.11	0.11	0.11	0.11	0.11	0.13	0.21			
65	0.13	0.13	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.17	0.27				
66	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.19	0.33					
67	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.21	0.37						
68	0.19	0.19	0.19	0.18	0.18	0.19	0.23	0.41							
69	0.19	0.19	0.19	0.19	0.19	0.24	0.45								
70	0.19	0.19	0.19	0.19	0.24	0.48									
71	0.19	0.18	0.19	0.24	0.50										
72	0.18	0.19	0.24	0.52											
73	0.18	0.24	0.54												
74	0.24	0.56													
75	0.58														

Table 5.3: The difference in the value function between model-1 and model-2

Table 5.4 shows the optimal admission policy for type-1 and type-2 patients. The actions for which $x_2 < 60$ are the same with the action for $x_2 = 60$ therefore, Table 5.4 presents the optimal policy only for $x_2 \geq 57$. According to Table 5.4 the system accepts all patients whenever there is an available bed in the base case. Therefore, both types of patients are preferred.

x_2, x_1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
57	1	1	1	1	1	1	1	1	1	1	1	1	1	1	4
58	1	1	1	1	1	1	1	1	1	1	1	1	1	1	4
59	1	1	1	1	1	1	1	1	1	1	1	1	1	1	4
60	1	1	1	1	1	1	1	1	1	1	1	1	1	1	4
61	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2
62	3	3	3	3	3	3	3	3	3	3	3	3	3	2	
63	3	3	3	3	3	3	3	3	3	3	3	3	2		
64	3	3	3	3	3	3	3	3	3	3	3	2			
65	3	3	3	3	3	3	3	3	3	3	2				
66	3	3	3	4	3	3	3	3	3	2					
67	3	3	3	3	3	3	3	3	2						
68	3	3	3	3	3	3	3	2							
69	3	3	3	3	3	3	2								
70	3	3	3	3	3	2									
71	3	3	3	3	2										
72	3	3	3	2											
73	3	3	2												
74	3	2													
75	2														

Decisions:

- (1) Accept type-1 and type-2,
- (2) Reject type-1 and type-2,
- (3) Accept type-1 and reject type-2,
- (4) Reject type-1 and accept type-2

Table 5.4: Optimal policy for the base case of model-1 and model-2

5.3 Variation of Model-1 Including Post-Op Patients

The models presented previously do not include post-op patients. I present a variation of model-1 including post-op patients for comparison. In this variation, a new variable x_3 and its associated parameters, R_3 and $1/\mu_3$, representing the post-op patients, acceptance reward and their average length of stay respectively are added. These patients are referred as type-3 patients. The state space for this variation is $S = \{x : 0 \leq x_1, 0 \leq x_2, 0 \leq x_3, x_1 + x_3 \leq C_1, x_1 + x_2 + x_3 \leq C_1 + C_2\}$. Usually, post-op patients are transferred into ICU due to precautionary measures instead of their need for intensive care. The suitable place for post-op patients is a step down

unit. Since this type of patient usually does not require intensive care, they cost less in ICU. According to Eachempati et al. (2004) patient to nurse ratio is 2 in ICUs meanwhile it is 4 in step down unit. I use this ratio to approximate the cost ratio of type-3 patients to type-2 patients. Thus, the cost ratio of type-3 patients to type-2 is calculated to be 1.5. In the same article, the average length of stay in the step down unit is given as 2.3 days. Type-3 patients reward, R_3 , has been calculated by multiplying their average length of stay with their cost ratio. $R_3 = 1.5/\mu_3$. Type-3 patients are usually kept in the ICU for a short period of time before transferring to a designated ward, thus being removed from the system. Based on this, I assume $\mu_2 < \mu_3$. Then, I normalize the time unit so that $\lambda_1 + \lambda_2 + C_1\mu_3 + C_2\mu_2 + \alpha = 1$.

In this model, type-3 patients are transferred into a surgical ward after they complete their first stage recovery. Since this ward is not included in the combined wards, they are modelled as being removed from the system even though they do not leave the hospital.

Case 1: $0 \leq x_1, 0 \leq x_3, x_1 + x_3 < C_1, 0 \leq x_2 < C_2$

$$\begin{aligned}
V_{n+1}(x_1, x_2, x_3) = & \lambda_1 \max\{V_n(x_1, x_2, x_3), V_n(x_1 + 1, x_2, x_3) + R_1\} + \\
& \lambda_2 \max\{V_n(x_1, x_2, x_3), V_n(x_1, x_2 + 1, x_3) + R_2\} + \\
& \lambda_3 (V(x_1, x_2, x_3 + 1) + R_3) + \\
& Px_1\mu_1 V_n(x_1 - 1, x_2 + 1, x_3) + \\
& (1 - P)x_1\mu_1 V_n(x_1 - 1, x_2, x_3) + \\
& x_2\mu_2 V_n(x_1, x_2 - 1, x_3) + \\
& x_3\mu_3 V_n(x_1, x_2, x_3 - 1) + \\
& ((C_1 - x_3)\mu_3 + (C_2 - x_2)\mu_2 - x_1\mu_1)V_n(x_1, x_2, x_3)
\end{aligned} \tag{5.1}$$

Case 2: $0 \leq x_1, 0 \leq x_3, x_1 + x_3 < C_1, x_2 \geq C_2, x_1 + x_2 + x_3 < C_1 + C_2$

$$\begin{aligned}
V_{n+1}(x_1, x_2, x_3) = & \lambda_1 \max\{V_n(x_1, x_2, x_3), V_n(x_1 + 1, x_2, x_3) + R_1\} + \\
& \lambda_2 V_n(x_1, x_2, x_3) + \\
& \lambda_3 (V(x_1, x_2, x_3 + 1) + R_3) + \\
& P x_1 \mu_1 V_n(x_1 - 1, x_2 + 1, x_3) + \\
& (1 - P) x_1 \mu_1 V_n(x_1 - 1, x_2, x_3) + \\
& x_2 \mu_2 V_n(x_1, x_2 - 1, x_3) + \\
& x_3 \mu_3 V_n(x_1, x_2, x_3 - 1) + \\
& ((C_1 - x_3) \mu_3 + (C_2 - x_2) \mu_2 - x_1 \mu_1) V_n(x_1, x_2, x_3)
\end{aligned} \tag{5.2}$$

Case 3: $0 \leq x_1, 0 < x_2, 0 \leq x_3, x_1 + x_2 + x_3 = C_1 + C_2$

$$\begin{aligned}
V_{n+1}(x_1, x_2, x_3) = & \lambda_1 V_n(x_1, x_2, x_3) + \\
& \lambda_2 V_n(x_1, x_2, x_3) + \\
& \lambda_3 (V(x_1, x_2, x_3)) + \\
& P x_1 \mu_1 V_n(x_1 - 1, x_2 + 1, x_3) + \\
& (1 - P) x_1 \mu_1 V_n(x_1 - 1, x_2, x_3) + \\
& x_2 \mu_2 V_n(x_1, x_2 - 1, x_3) + \\
& x_3 \mu_3 V_n(x_1, x_2, x_3 - 1) + \\
& ((C_1 - x_3) \mu_3 + (C_2 - x_2) \mu_2 - x_1 \mu_1) V_n(x_1, x_2, x_3)
\end{aligned} \tag{5.3}$$

Case 4: $0 \leq x_1, 0 \leq x_3, x_1 + x_3 = C_1, 0 < x_2 < C_2$

$$\begin{aligned}
V_{n+1}(x_1, x_2, x_3) = & \lambda_1 V_n(x_1, x_2, x_3) + \\
& \lambda_2 \max\{V_n(x_1, x_2, x_3), V_n(x_1, x_2 + 1, x_3) + R_2\} + \\
& \lambda_3 (V(x_1, x_2, x_3)) + \\
& P x_1 \mu_1 V_n(x_1 - 1, x_2 + 1, x_3) + \\
& (1 - P) x_1 \mu_1 V_n(x_1 - 1, x_2, x_3) + \\
& x_2 \mu_2 V_n(x_1, x_2 - 1, x_3) + \\
& x_3 \mu_3 V_n(x_1, x_2, x_3 - 1) + \\
& ((C_1 - x_3) \mu_3 + (C_2 - x_2) \mu_2 - x_1 \mu_1) V_n(x_1, x_2, x_3)
\end{aligned} \tag{5.4}$$

Arrivals and departures of this variation are identical with model-1 except for type-3 patient flow. Since type-3 arrivals come from another unit within the hospital, they have to be accepted as long as there is an available bed. I can show the monotonicity of the value functions in this case as well.

Theorem 10: For all \mathbf{x} with $\mathbf{x} + e_i \in S$, the value function is monotonically decreasing in x_i , *i.e.*, $V_n(\mathbf{x} + e_i) - V_n(\mathbf{x}) \leq 0$ for $i = 1, 2, 3$.

Proof: For $i = 3$, when the additional type-3 patient recovers, the two systems couple. The rest of the proof is identical to the proof of Theorem 1. ■

Table 5.5 displays the difference between the value functions of model-1 and the variation calculated with discounted rewards and the base case parameters. The difference between the value functions increases as the number of type-2 patients in the systems declines. Therefore, I omit the states with $x_2 < 61$ for the sake of presentability. Table 5.5 suggests that even when there is no post-op patient in the system, the prospect of future post-op arrivals decreases the value function compared to model-1.

The reason of the decline in the value function is due to the lower acceptance reward of post-op patients. There is a curious case in the state with $x_2 = 75, x_1 = 0$

x_2, x_1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
61	7.25	7.24	7.24	7.23	7.22	7.21	7.21	7.20	7.19	7.17	7.14	7.05	6.78	5.88	2.77
62	7.09	7.08	7.07	7.06	7.05	7.04	7.03	7.02	7.01	6.99	6.95	6.86	6.51	4.83	
63	6.80	6.79	6.78	6.77	6.76	6.75	6.74	6.72	6.71	6.69	6.64	6.45	5.16		
64	6.44	6.43	6.42	6.41	6.40	6.39	6.38	6.36	6.35	6.32	6.17	5.00			
65	6.05	6.04	6.03	6.02	6.00	5.99	5.98	5.97	5.94	5.81	4.70				
66	5.63	5.62	5.61	5.60	5.59	5.57	5.56	5.54	5.42	4.34					
67	5.19	5.18	5.17	5.16	5.15	5.14	5.11	5.00	3.95						
68	4.73	4.72	4.71	4.70	4.69	4.67	4.56	3.54							
69	4.25	4.24	4.23	4.22	4.20	4.10	3.10								
70	3.74	3.73	3.72	3.70	3.60	2.63									
71	3.20	3.18	3.16	3.07	2.12										
72	2.60	2.58	2.49	1.56											
73	1.95	1.86	0.95												
74	1.16	0.28													
75	-0.48														

Table 5.5: The difference in the value functions of Model-1 and the variation including post-op patients when $x_3 = 0$

where the difference becomes negative. This is due to the fact that all of the beds in both the ICU and the ward are occupied by type-2 patients. The system cannot accept any patient until service completion of a type-2 patient. Incoming ICU patients would need both the first and the second stage service. Since there is no available bed in the ward, they keep occupying ICU beds for their second stage service. On the other hand, type-3 patients discharged from the system after completing the first stage recovery so they do not put a burden on the ward. Therefore, a variation of model-1 become preferable to model-1 just for the case $x_2 = 75, x_1 = 0$.

5.4 Sensitivity Analysis

In this section, I present how sensitive the optimal policy of model-1 and model-2 is to changes in $R_1, P, \lambda_1, \lambda_2$.

I only investigated the sensitivity in the cases where load is increased since in the base case the model already suggests accepting every arrival and decreasing load would not lead to a different outcome.

R_1	States Rejecting Type-2	Preferred Class
17.1364 — 107.01	None	Type-1, Type-2
261.42 — 518.77	(14, 60)	Type-1
2577.57 — 5151.07	(13, 60) (14, 60)	Type-1
25739.07 — 51474.07	(12, 60) (13, 60) (14, 60) (14, 59) (13, 59)	Type-1
$P_1 R_1$	States Rejecting Type-2	Preferred Class
1 17.1364	None	Type-1, Type-2
1 41.13	None	Type-1
1 518.77	(14, 60)	Type-1
1 37062.47	(12, 60) (13, 60) (14, 60) (14, 59)	Type-1

Table 5.6: Sensitivity of the actions by the change in cost ratio between ICU and wards

The first column of the first part of the table stores the different R_1 values by changing the cost ratio which is presented in **Section 5.1**. The second column indicates the states rejecting type-2 patients depending on the R_1 values. Type-1 patients are rejected in none of the states. The third column shows the preferred classes for the values of R_1 . The second part of the table is nearly identical to the first part with the only difference being a change in survival rate to 1.

Trying the same cases in Table 5.6 after changing λ_1 to 1.5 and 3 times itself and λ_2 to 2 times itself, results in no change to optimal policy.

This analysis shows that both models are insensitive to changes in rewards and arrival rates. The model-1 and model-2 do not reject type-2 patients in order to preserve empty beds for type-1 patients unless the difference between the rewards is very large.

In the cases with a large number of beds such as the base case, type-2 patients are not rejected to preserve beds for type-1 patients unless the difference between rewards is large enough. Inspecting the case with a small number of beds, the rejection of type-2 patients starts at lower values of the ratio of cost of ICU to cost of ward. For instance, in the base case the rejections start roughly around when the cost ratio of ICU to ward is 50. Meanwhile when the parameters $C_1 = 2$, $C_2 = 19$ are set, the rejection of type-2 patients starts with a cost ratio around 20.

In practice, departments for type-1 and type-2 patients are separated. When all beds in the type-2 department are occupied, type-1 patients who are no longer in need of intensive care do not have a bed to transfer into so they keep occupying beds in the ICU. This might prevent the acceptance of ICU arrivals. The only reason to reject type-2 patients would be a critically low number of available beds in the ICU, being preserved for type-1 patients. The probability of the system being in states with blocked patients in the long run has been calculated to be %17.13. Being in a state with blocked patients does not necessarily mean rejecting type-1 patients. The probability of being in a state with both blocked patients and unavailable beds in the ICU has been calculated to be %2.26. Even these states does not necessarily mean rejecting type-1 patients. If a type-2 service completion occurs before a type-1 arrival, one of the blocked patients may be transferred to the ward allowing the acceptance of the type-1 arrival. Note that type-1 service completions in those states do not give the same result as they turn into blocked patients and keep occupying ICU beds. Therefore in these states, the importance of service completion in the ward is higher than the importance of service completion in the ICU. Thus the probability of having to reject type-1 arrivals in the long run is fairly low. This implies that there is usually no reason to reject type-2 patients unless there is a big discrepancy in the rewards.

The results of this analysis seem to agree with this reasoning.

As shown in Table 5.2, optimal policies of model-1 and model-2 are the same. In model-2, effect of being blocked is more prevalent, since they cause the service to be ceased. Then it is reasonable to expect that the model will suggest to reject type-2 patients at lower cost ratios however, no practical difference is found between model-1 and model-2.

5.5 Properties of the Value Function

In this section, I present the results about submodularity and concavity properties of $V(x_1, x_2)$. In the case of parallel multi-servers in tandem, the event in which type-1 patients receive service fails to be submodular since the value function is not subconcave in x_2 for fixed x_1 .

$V(\mathbf{x})$ is subconcave in x_2 if for all $n \in \mathbb{N}_{\geq 0}$ and for all $\mathbf{x} + e_1 + e_2 \in S$, $\mathbf{x} + 2e_1 \in S$

$$V_n(\mathbf{x} + e_1) - V_n(\mathbf{x} + e_1 + e_2) - V_n(\mathbf{x} + e_2) + V_n(\mathbf{x} + 2e_2) \leq 0 \quad (5.5)$$

$V(\mathbf{x})$ is submodular if for all $n \in \mathbb{N}_{\geq 0}$ and for all $\mathbf{x} + e_1 + e_2 \in S$

$$V_n(\mathbf{x}) - V_n(\mathbf{x} + e_1) - V_n(\mathbf{x} + e_2) + V_n(\mathbf{x} + e_1 + e_2) \leq 0 \quad (5.6)$$

The benefit due to an additional x_1 decreases with increasing x_2 and vice versa in submodular functions. Submodularity of the value function guarantees that the optimal policy is of threshold type. Ghoneim and Stidham Jr (1985) in their lemma 1, show if a function is submodular and subconcave in x_i then it is also concave in x_i .

$V(\mathbf{x})$ is concave in x_i where $i \in \{1, 2\}$ if for all $n \in \mathbb{N}_{\geq 0}$ and for all $\mathbf{x} + 2e_i \in S$

$$V_n(\mathbf{x}) - 2V_n(\mathbf{x} + e_i) + V_n(\mathbf{x} + 2e_i) \leq 0 \quad (5.7)$$

The benefit due to an additional x_i decreases with increasing x_i , in concave functions. Concavity guarantees that the optimal threshold policy is monotone.

With base case parameters, submodularity and concavity fail to hold. I investigated cases where these properties hold by changing, $C_1, C_2, \lambda_1, \lambda_2$, and the cost ratio.

All combinations of values with $1 \leq C_1 \leq 14$, $3 \leq C_2 \leq 61$ and $1 \leq \text{Cost Ratio} \leq 50$ by increments 10 while adjusting λ_1, λ_2 to have a load close to the base case have been tried. Existence of these properties is more sensitive to C_1 . For $C_1 > 3$, submodularity and concavity do not hold. For $1 \leq C_1 \leq 3$ and value of C_2 varying between 2 to 61 with other parameters fixed to base case parameters; submodularity, concavity and subconcavity of $V(x_1, x_2)$ in x_1 for fixed x_2 are observed.

In trying to prove these properties, the event of transferring from the ICU to the ward is found to prevent them from holding. Cases these properties hold has been numerically searched. When C_1 is set to 1,2 or 3; the effect of this event is decreased and submodularity, concavity and subconcavity of $V(x_1, x_2)$ in x_2 for fixed x_1 can be observed with parameters taken from Table 5.2.

Chapter 6

CONCLUSION

In this study, two MDP models which represent an ICU and a ward in series are investigated. By comparing model-1 and model-2, effects of the service of blocked patients and the interaction between wards and intensive care units on admission policy are demonstrated.

I prove the monotonicity of the value function and demonstrate the difference between having an extra ICU patient or a ward patient. Then, I construct sufficient conditions for the existence of preferred classes.

In the numerical analysis, properties of the value function, the existence of preferred classes and rejections for model-1 and model-2 are observed with some parameter sets. A variation of model-1 including post-op patients has been investigated and effects of having post-op patients in the system are presented. A sensitivity analysis is shown, investigating the models under different parameters.

It has been observed that, ICU patients are always accepted whenever there is an available bed as they are assumed to be financially more beneficial due to the urgency of their situation. The only condition under which ward patients are rejected is when the number of available beds is critically low in both stations and the difference between rewards is large enough. This is due to the fact that ICU patients have greater reward so that it is beneficial to preserve beds for recovering or incoming ICU patients.

BIBLIOGRAPHY

- Altman, E., Jiménez, T., and Koole, G. (2001). On optimal call admission control in resource-sharing system. *IEEE Transactions on Communications*, 49(9):1659–1668.
- Bekker, R., Koole, G., and Roubos, D. (2017). Flexible bed allocations for hospital wards. *Health care management science*, 20(4):453–466.
- de Bruin, A. M., Bekker, R., Van Zanten, L., and Koole, G. (2010). Dimensioning hospital wards using the erlang loss model. *Annals of Operations Research*, 178(1):23–43.
- Eachempati, S. R., Hydo, L. J., and Barie, P. S. (2004). The effect of an intermediate care unit on the demographics and outcomes of a surgical intensive care unit population. *Archives of surgery*, 139(3):315–319.
- Ghoneim, H. A. and Stidham Jr, S. (1985). Control of arrivals to two queues in series. *European Journal of Operational Research*, 21(3):399–409.
- Halpern, N. A. and Pastores, S. M. (2015). Understanding the russell equation and projection estimates to describe critical care costs in the usa. *Intensive care medicine*, 41(10):1828–1830.
- Helm, J. E., AhmadBeygi, S., and Van Oyen, M. P. (2011). Design and analysis of hospital admission control for operational effectiveness. *Production and Operations Management*, 20(3):359–374.
- Kim, S.-H., Chan, C. W., Olivares, M., and Escobar, G. J. (2016). Association among icu congestion, icu admission decision, and patient outcomes. *Critical care medicine*, 44(10):1814–1821.

- Ku, C.-Y. and Jordan, S. (1997). Access control to two multiserver loss queues in series. *IEEE Transactions on Automatic Control*, 42(7):1017–1023.
- Ku, C.-Y. and Jordan, S. (2002). Access control of parallel multiserver loss queues. *Performance Evaluation*, 50(4):219–231.
- Kuntz, L., Mennicken, R., and Scholtes, S. (2014). Stress on the ward: Evidence of safety tipping points in hospitals. *Management Science*, 61(4):754–771.
- Kurt, N. E. (2014). A stylized bed allocation model for a turkish private hospital with a two class loss system.
- Lippman, S. A. (1975). Applying a new device in the optimization of exponential queuing systems. *Operations Research*, 23(4):687–710.
- Long, E. F. and Mathews, K. S. (2018). The boarding patient: Effects of icu and hospital occupancy surges on patient flow. *Production and Operations Management*, 27(12):2122–2143.
- Millhiser, W. P. and Burnetas, A. N. (2013). Optimal admission control in series production systems with blocking. *IIE Transactions*, 45(10):1035–1047.
- Örmeci, E. L., Burnetas, A., and van der Wal, J. (2001). Admission policies for a two class loss system.
- Örmeci, E. L., Burnetas, A. N., and Emmons, H. (2002). Dynamic policies of admission to a two-class system based on customer offers. *Iie Transactions*, 34(9):813–822.
- Örmeci, E. L. and van der Wal, J. (2006). Admission policies for a two class loss system with general interarrival times. *Stochastic models*, 22(1):37–53.
- Pastores, S. M., Dakwar, J., and Halpern, N. A. (2012). Costs of critical care medicine. *Critical care clinics*, 28(1):1–10.
- Savin, S. V., Cohen, M. A., Gans, N., and Katalan, Z. (2005). Capacity management in rental businesses with two customer bases. *Operations Research*, 53(4):617–631.

- Siddiqui, S. (2015). Mortality profile across our intensive care units: A 5-year database report from a singapore restructured hospital. *Indian journal of critical care medicine: peer-reviewed, official publication of Indian Society of Critical Care Medicine*, 19(12):726.
- Ulukus, M. Y., Güllü, R., and Örmeci, L. (2011). Admission and termination control of a two class loss system. *Stochastic Models*, 27(1):2–25.
- Xu, S. H., Richter, R., and Shanthikumar, J. G. (1992). Optimal dynamic assignment of customers to heterogeneous servers in parallel. *Operations Research*, 40(6):1126–1138.