OPTIMAL CONTROL POLICIES FOR MAKE-TO-STOCK PRODUCTION SYSTEMS WITH SEVERAL PRODUCTION RATES AND DEMAND CLASSES

WEI LIN

NATIONAL UNIVERSITY OF SINGAPORE

2004

OPTIMAL CONTROL POLICIES FOR MAKE-TO-STOCK PRODUCTION SYSTEMS WITH SEVERAL PRODUCTION RATES AND DEMAND CLASSES

WEI LIN

(B. Eng. HUST)

A THESIS SUBMITTED

FOR THE DEGREE OF MASTER OF ENGINEERING

DEPARTMENT OF INDUSTRIAL AND SYSTEMS ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2004

Acknowledgement

I would like to express my profound gratitude to my supervisors, Dr. Chulung Lee and Dr. Wikrom Jaruphongsa, for their invaluable advice and guidance throughout the whole course.

My sincere thanks are conveyed to the National University of Singapore for offering me a Research Scholarship and the Department of Industrial and Systems Engineering for usage of its facilities, without any of which it would be impossible for me to complete the work reported in this dissertation.

I am highly indebted to many friends, Mr. Bao Jie, Mr. Gao Wei, Mr. Li Dong, Mr. Liang Zhe, Mr. Liu Bin, Ms. Liu Rujing, Mr. Xu Zhiyong, Ms. Yang Guiyu and Mr. Zhang Jun who have contributed in one way or another towards the fulfillment of this dissertation.

I am grateful to my parents and parents-in-law for their continuous concern and moral support.

Finally, I would like to express my special great gratitude to my wife for her understanding, patience, and encouragement throughout the course of my research.

Table of Contents

Acknowledgement **i**

Summary **iv**

Nomenclature **vi**

List of Figures **vii**

1 Introduction and Literature Review **1**

2 A Make-to-Stock Production System with Multiple Production Rates, One Demand Class and Backorders **10**

2.1 The Stochastic Model and Optimal Control 10

2.1.1 Dynamic Programming Formulation 11

2.1.2 The Optimal Control Policy 17

2.2 Stationary Analysis of the Production System 21

2.3 Numerical Study 27

2.4 Production System with Multiple Production Rates 33

2.5 Conclusions 36

3 A Make-to-Stock Production System with Two Production Rates,

N Demand Classes and Lost Sales **37**

3.1 The Stochastic Model and Optimal Control 37

3.1.1 Dynamic Programming Formulation 39

3.1.2 The Optimal Control Policy 42

3.2 Stationary Analysis of the Production System 45

3.3 Numerical Study 53

3.4 Conclusions 59

4 A Make-to-Stock Production System with Two Production Rates, Two Demand Classes and Backorders **61**

4.1 The Stochastic Model and Optimal Control 61

4.1.1 Dynamic Programming Formulation 62

4.1.2 The Optimal Control Policy 65

4.2 Conclusions 78

5 Conclusions and Future Study **79**

Bibliography **81**

Summary

In this dissertation, we develop the optimal control policies for make-to-stock production systems under different operating conditions. First, we consider a make-to-stock production system with a single demand class and two production rates. With the assumptions of Poisson demands and exponential production times, it is found that the optimal control policy, denoted later as (S_1, S_2) policy, is characterized by two critical inventory levels S_1 and S_2. Then, under the (S_1, S_2) policy, an $M/M/1/S$ queueing model with state-dependent arrival rates is developed to compute the expected total cost per unit time. To show the benefits of employing the emergency rate, numerical studies are carried out to compare the expected total costs per unit time between the production system with two rates and the one with a single rate. Moreover, the developed model is extended to consider N production rates and the optimal control policy with certain conditions satisfied is shown to be characterized by N critical inventory levels. Second, we consider a make-to-stock production system with N demand classes and two production rates for a lost-sale case. It is found that the optimal control policy is a combination of the (S_1, S_2) policy and the so-called stock reservation policy. Similarly, under this optimal control policy, an $M/M/1/S$ queueing model with state-dependent arrival rates and service rates is developed to compute the expected total cost per unit time. Then, the results of numerical studies are provided to show the benefits of employing the emergency production rate. Finally, we study a make-to-stock production system with two demand classes and two production rates for a backorder case. The optimal control policy is shown to be characterized by three monotone curves.

(Normal/Emergency Production Rates; Make-to-Stock Production System; Dynamic Programming; Inventory Control)

Nomenclature

$\boldsymbol{A}$	Transition Rate Matrix
b_i	Backorder Cost of Class i Demand
B_i	Expected Number of Class i Backorders
c	Cost Difference between Normal and Emergency Rate
c_i	Unit Production Cost of i^{th} Production Rate
C	Expected Total Cost Per Unit Time
CS	Cost Saving
f	The Minimal Expected Total Discounted Cost
h	Inventory Holding Cost
H	The Operater
I	Expected On-Hand Inventory Level
L_i	Probability of Lost Sales for Class i Demand
P_i	Probability of i^{th} Production Rate Employed
$P(i,j)$	Transition probability from state i to j
R_i	Critical Inventory Level
S_i	Critical Inventory Level
TRC	Relevant Expected Total Cost Per Unit Time

v	Function belonging to the Set V
V	The Set of Structured Functions
X_i	Continuous-time Markov Process
X_i'	Converted Continuous-time Markov Process
α	The Interest Rate
λ_i	Arrival Rate of Class i Demand
Λ	Transition Rate of Converted Markov Processes
μ_i	i^{th} Production Rate
p_i	Unit Lost-Sale Cost of Class i Demand
$\pi(n)$	Steady State Probability of State n
ρ_1	Ratio between λ and μ_1
ρ_2	Ratio between λ and μ_2
ρ_{11}	Ratio between λ_1 and μ_1
ρ_{12}	Ratio between λ_1 and μ_2
Z	The set of integers

List of Figures

2.1 Transition process for the Markov process X'_1 12

2.2 The illustration of the (S_1, S_2) policy 21

2.3 Rate diagram for the $M/M/1/S$ queueing system 22

2.4 The effect of ρ_1 over cost saving 29

2.5 The effect of μ_2/μ_1 over cost saving 31

2.6 The effect of c over cost saving 31

2.7 The effect of h over cost saving 32

2.8 The effect of b over cost saving 32

3.1 Transition process for the Markov process X'_3 39

3.2 Rate diagram for the $M/M/1/S$ queueing system if $S_2 \geq R_2$ 47

3.3 Rate diagram for the $M/M/1/S$ queueing system if $S_2 < R_2$ 50

3.4 Cost saving versus μ_2/μ_1 55

3.5 Cost saving versus ρ_1 56

3.6 Cost saving versus λ_2/λ_1 56

3.7 Cost saving versus h 57

3.8 Cost saving versus c_2/c_1 58

3.9 Cost saving versus p_1/p_2 59

4.1 Transition process for the Markov process X_4' 64

4.2 The optimal policy characterized by $R(y)$, $S(y)$ and $B(x)$ 77

Chapter 1

Introduction and Literature Review

Inventory systems with two replenishment modes are becoming increasingly common in practice nowadays [25]. For such inventory systems, a slower replenishment mode is normally used except when the stock supply needs to be expedited where the emergency production mode is employed. In this dissertation, we first consider a make-to-stock production system with two production rates: normal and emergency. The normal production rate is the main resource for the stock supply. However, when the inventory level becomes difficult to satisfy the anticipated demands, the emergency production rate is employed to prevent costly stock-outs. The normal production rate incurs lower production cost but with lower throughput while the emergency production rate increases throughput at the expense of higher production cost. This production system can be considered as an inventory system with two replenishment modes, which can be met in the real life. For example, for the remanufacturable-products, such as some parts of automobiles, the remanufactured-items are normally used to satisfy the incoming demands. However, when there are not enough remanufactured items, newly manufactured items may be used to avoid costly stock-outs. The most important operational decision, which significantly affects the total system cost, is to determine the optimal production rate given the inventory levels. Such decisions must be carefully made to minimize the system cost. This problem is referred to as the production control problem. Despite its importance, the production control problem for the production system with two production rates has yet received its due attention in the literature.

This dissertation is closely related to the literature of inventory systems with two replenishment modes, which were discussed as early as in 1960s. Since then, many articles in this area have been published. Inventory systems studied in these articles can be divided into two groups: those with continuous-review control policies and those with periodic-review control policies. Almost all the earlier papers studied inventory systems with periodic-review control policies. In a seminal paper, Barankin [1] developed a single-period inventory model with normal and emergency replenishments whose lead-times are one period and zero, respectively. Daniel [7] and Neuts [23] extended Barankin's for multiple periods and obtained an optimal control policy with similar forms. Fukuda [10] further generalized Daniel's model by considering fixed order costs and allowing normal and emergency replenishments to be placed simultaneously. However, still the assumptions that lead-time of normal replenishments is one period and that of emergency replenishments is zero are not relaxed. Whittmore and Saunders [28] obtained the optimal control policy for a multiple planning period model where lead-times for normal and emergency replenishments can take any multiple of the review period. However, the policy developed is too complex to be implemented in practice. The explicit results are able to be obtained only for the case where two replenishment lead-times differ by one period only.

Chiang and Gutierrez [3] developed a model where lead times of normal and emergency replenishments can be shorter than the review period. At any review epoch, either normal or emergency replenishments can be placed to raise the inventory level to an order-up-to level. Unit purchasing costs are same for normal and emergency replenishments, but emergency replenishments have fixed order costs which normal replenishments do not have. It is found that for any given non-negative order-up-to level, either only normal replenishments are used all the time, or there exists an indifference inventory level such that if the inventory level at the review epoch is below the indifference inventory level, emergency replenishments are placed and normal replenishments are placed otherwise. In a subsequent paper, Chiang and Gutierrez [4] allowed emergency replenishments to be placed at any time within a review period while normal replenishments may be placed only at review epochs. In addition, the order-up-to level of emergency replenishments depends on the time remaining until the next normal replenishment arrives. They analyzed the problem within the framework of a stochastic dynamic programming and derive an optimal control policy. However, this control policy is quite complex, especially if lead-times of normal replenishments and emergency replenishments differ by more than one time unit.

Tagaras and Vlachos [25] also studied an inventory system where lead times can be shorter than the review period. Normal replenishments may be placed only at review epochs based on an order-up-to level policy. Emergency replenishments are placed at most once per cycle and are expected to arrive just before the arrival of the normal replenishment placed in this cycle when the likelihood of stock-outs is highest.

For the case where lead-times of emergency replenishments are only one unit time, an approximate total cost is obtained.

Inventory systems with continuous-review control policies have been studied only in recent years. Moinzadeh and Nahmias [20] proposed a heuristic control policy for an inventory model with two replenishment modes. This control policy, which is a natural extension of the standard (Q, R) policy, can be specified by (Q_1, R_1, Q_2, R_2) where $Q_1 > Q_2$ and $R_1 > R_2$. A normal replenishment with lot size Q_1 is placed when the inventory level reaches R_1 and an emergency replenishment of lot size Q_2 is placed when the inventory level falls below R_2. An approximate expected total cost per unit time is derived with the assumptions that there is never more than a single outstanding replenishment of each type and that an emergency replenishment is placed only if it will arrive before the scheduled arrival of the outstanding normal replenishment. Fixed order costs for normal and emergency replenishments are considered. However, the backorder cost only consists of fixed shortage cost per unit backlogged. Essentially, this is equal to the lost sale problem because there is no incentive to satisfy the backorders once they occur. The parameters Q_1, R_1, Q_2 and R_2 are obtained numerically by applying simple search procedures. At last, simulation is employed to check the validity of the control policy. The results obtained shows that for certain parameters combinations, the cost saving might be 10–30%, in some cases even larger.

Johansen and Thorstenson [11] developed a similar model to Moinzadeh and Nahmias [20] where instead Q_2 and R_2 vary with the time remaining until the outstanding normal replenishment arrives, i.e., Q_2 and R_2 are state-dependent. The backorder cost now consists of both fixed shortage cost per unit backlogged and backordering cost per unit backlogged per unit time. A tailor-made policy-iteration algorithm is developed and implemented to minimize the approximate expected total cost per unit time. In addition, a simplified control policy is considered for comparative purposes where Q_2 and R_2 are constant instead of varied. The results of numerical studies show that there is only a small extra gain from using the state-dependent Q_2 and R_2.

Moinzadeh and Schmidt [19] considered an inventory system with Poisson demands and two replenishment modes. The control policy implemented is an extension of the standard $(S-1, S)$ policy. When a demand occurs, a replenishment is placed immediately no matter whether the demand is satisfied or backlogged. However, what kind of replenishment to be placed depends on the ages of all the outstanding replenishments and the inventory level at the time of the demand arrival. If the inventory level is above a critical level, normal replenishments are placed. If the inventory level is less than the critical level but enough outstanding replenishments will arrive within the lead time of normal replenishments to increase the inventory level beyond the critical level, normal replenishments are still employed; emergency replenishments are employed otherwise. Under this control policy, they obtain several optimality properties for the steady-state behavior and provide some computational results.

Kalpakam and Sapna [15] considered a lost sale inventory model with renewal demands and state-dependent lead times based on an extension of the (Q, R) policy. When the inventory level reaches R from above and no order is outstanding, an order of size Q is placed. Moreover, whenever the inventory level drops to zero, an order of size R (or size Q) is placed if an order of size Q (or size R) is outstanding. The lead times of the two replenishments modes depend on the order size and the number of outstanding orders. Simulation is employed to check the validity of their model.

This dissertation also has a close relationship with the literature of inventory systems with rationing. Veinott [27] considered a periodic-review, nonstationary, multiperiod inventory model in which there are N classes of demand for a single product. He is the first one who introduces the concept of a critical level policy, i.e., demand from a particular class is satisfied only if the inventory level is above the critical level associated with this demand class. In a model formulated similar to Veinott's, Topkis [26] broke down the review period into a finite number of intervals and assumes that all demands are observed before making any rationing decision. He proves the optimality of the critical level policy for an interval for both backordering case and lost sale case. Evans [9] and Kaplan [16] derived essentially the same results, but for two demand classes. Nahmias and Demmy [22] considered a single period inventory model with two demand classes. With the assumptions that demand occurs at the end of the review period and high priority demands are filled first, they develop an approximate expression of the expected backorder rate for each demand class under the critical level policy. They also generalized the results to an infinite horizon, multi-period inventory model, where stock is replenished under (s, S) policy and lead time is zero. Later, Moon and Kang [21] generalized Nahmias and Demmy's results for multiple demand classes. Cohen et al. [6] considered a periodic review (s, S) inventory model in which there are two priority demand classes. However, the critical level policy is not employed in the model. In each period, inventory is issued to meet high-priority demand first and the remaining is then available to satisfy low-priority demand.

Nahmias and Demmy [22] is the first to consider continuous-review inventory model with inventory rationing. They analyzed a (Q, R) inventory model with two demand classes and positive deterministic leadtime. Assuming that there is never more than a single replenishment outstanding, an approximate expected backordering rate for each demand class is obtained. Dekker et al. [8] considered a $(S-1, S)$ inventory model with two demand classes, Poisson demand and fixed lead time. The main result is the approximate expressions for the service levels of the two demand classes.

Ha [12] considered a make-to-stock production system for the lost sale case in which there are N demand classes for a single item. With the assumptions of Poisson demand and exponential production time, it is found that the optimal control policy is essentially a combination of the base-stock policy controlling the production process and the critical level policy controlling the inventory rationing. Based on $M/M/1/S$ queueing system, the expected total cost per unit time is computed for a case with two demand classes. The results of numerical studies show that remarkable benefits can be generated by the critical level policy relative to the first-come-first-served policy.

Ha [14] considered a make-to-stock production system for the backordering case with two demand classes, Poisson demand and exponential production time. He proves that the critical level policy is still optimal for inventory rationing. The critical level decreases as the number of backorders of low-priority demand increases.

In Chapter 2, we first consider a make-to-stock production system with two production rates, one demand class and backorders. The two production rates are characterized by different production times and unit production costs, i.e., the faster the production is, the larger the unit production cost is. With the assumptions of Poisson demand and exponential production time, it is found that the optimal control policy is characterized by two critical levels S_1 and S_2. We refer to this control policy later as the (S_1, S_2) policy. If the inventory level reaches S_1, production is stopped. If the inventory level is between S_1 and S_2, production is performed by employing the smaller production rate. If the inventory level is less than S_2, production is performed by employing the larger production rate. In addition, we extend the production system for considering N production rates. From the foregoing literature review, all the previous works considering inventory systems with alternative replenishment modes focus on the situation where lead times of normal and emergency replenishments are constant. Moreover, supply processes of those works are of an infinite capacity. But in this chapter, lead times of the normal and emergency production rate, which are exponentially distributed, are stochastic. Meanwhile, supply process of the production system is capacitated. Therefore, our model is different from the models in the literature.

In Chapter 3, we consider a make-to-stock production system with two production rates, N demand classes and lost sales. It is found that the optimal control policy is a combination of the (S_1, S_2) policy controlling the production process and the critical level policy controlling inventory allocation. There is a critical level associated with each demand class. An incoming demand of this particular class will be satisfied if the inventory level is above the critical level, and is rejected otherwise.

In Chapter 4, we consider a make-to-stock production system with two production rates, two demand classes and backorders. The optimal control policy is characterized by three monotone switch curves, which partition the state space of the system into four areas each of which corresponds to a different production decision.

As shown above, exponential production times are assumed throughout this thesis to make our problems tractable. While this assumption may not be realistic in most production systems, we believe that the insights of our results are still useful when it is relaxed. Without this assumption, the properties of Markov process, on which our analysis mainly depends on, are lost. This will make our problem much more complex.

Chapter 2

A Make-to-Stock Production System with Multiple Production Rates, One Demand Class and Backorders

2.1. The Stochastic Model and Optimal Control

In this chapter, we consider a single-item, make-to-stock production facility with two production rates: normal and emergency. Production times for the normal and emergency rates are independent and exponentially distributed with means $1/\mu_1$ and $1/\mu_2$, respectively. The unit production cost for the normal rate is c_1 and that for the emergency rate is c_2. For notational convenience, let $\mu_0 = 0$ and $c_0 = 0$ be the parameters for the case when there is no production. Naturally, we assumed that $\mu_0 < \mu_1 < \mu_2$ and $c_0 < c_1 < c_2$. Customer demands arise as a Poisson process with mean rate λ and unsatisfied demands are backlogged with penalty costs incurred.

At an arbitrary point of time, we have three possible production decisions to make given the current inventory level: i) not to produce, ii) to produce normally, and iii) to produce urgently. Due to the exponential production times and Poisson demands assumptions, the current inventory level possesses all the necessary information for decision-making (Memoryless Property). Thus, although we allow the production rate to be varied at any time, the optimal production rate is reviewed only when the inventory level changes, i.e., when demand arrives or production completes. A control policy specifies the production rate at any time given the current inventory level. We develop an optimal control policy for the objective of minimizing the expected total discounted cost over an infinite time horizon. This expected total discounted cost is computed by the following cost components: the inventory holding cost h per unit per unit time, the normal production cost c_1 per unit, the emergency production cost c_2 per unit, and the backorder cost b per unit backordered per unit time.

In the next subsection, the optimality equation is obtained which is satisfied by the minimal expected total discounted cost and the optimal control policy is identified by analyzing this optimality equation.

2.1.1. Dynamic Programming Formulation

Let $X_1(t)$ be the net inventory level at time t. For any given Markovian control policy u, $X_1 = \{X_{1u}(t) : t \geq 0\}$ is a continuous-time Markov process with the state space Z, where Z represents integers. For the Markov process X_1, transitions occur when demand arrives or production completes. Denote $P(i, j)$ as the transition probability from state i to j. Given the current state x and the production rate employed at this stage μ_k, $k = 0, 1, 2$, the transition probabilities of the Markov process X_1 are $P(x, x+1) = \mu_k/(\mu_k + \lambda)$ and $P(x, x-1) = \lambda/(\mu_k + \lambda)$. It can be seen that the transition probabilities take different values for different production rates employed upon jumping into state x. Especially, the transition probabilities are $P(x, x+1) = 0$ and $P(x, x-1) = 1$ when there is no production employed. For the Markov process X_1, the time between successive transitions is influenced by both the exponential production process and the Poisson demand process. It follows that the time between successive transitions follows an exponential distribution with mean $1/(\mu_k + \lambda)$ (see Çinlar [5]). The mean $1/(\mu_k + \lambda)$ is variable and dependent on control policies employed. This will significantly increase the complexity of computing the expected total discounted cost, from which the optimal control policy will be identified.

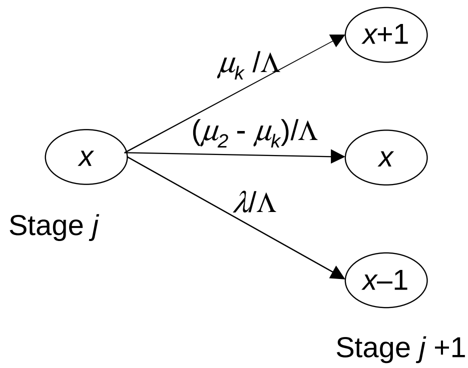

Figure 2.1: Transition process for the Markov process X_1'

To simplify the problem, we follow the procedure proposed by Lippman [18] to convert the Markov process X_1 to X_1' where the transition rate Λ is defined by $\lambda + \mu_2$. Accordingly, the transition probabilities of the converted Markov process X_1' becomes $P'(x, x) = (\mu_2 - \mu_k)/\Lambda$, $P'(x, x+1) = \mu_k/\Lambda$ and $P'(x, x-1) = \lambda/\Lambda$, i.e., a transition taking place at the end of the stage turns out to be no event with the probability $(\mu_2 - \mu_k)/\Lambda$, to be a production completion with the probability of μ_k/Λ, and to be a demand arrival with the probability of λ/Λ. Figure 2.1 shows the transition process for the Markov process X_1'. With the newly defined transition rate and transition probabilities, the underlying stochastic processes of the Markov processes X_1 and X_1' are essentially the same, which will be shown next.

For the Markov process X_1, transitions occur with mean rate $\mu_k + \lambda$. When a transition occurs, the system will definitely jump out from the current state . Thus, the transition rates matrix $\boldsymbol{A}$ of the Markov process X_1 are as follows:

$$A(x, x+1) = (\mu_k + \lambda)\, P(x, x+1) = \mu_k \tag{2.1}$$

$$A(x, x-1) = (\mu_k + \lambda)\, P(x, x-1) = \lambda \tag{2.2}$$

$$A(x, x) = -\left[A(x, x+1) + A(x, x-1)\right] = -\mu_k - \lambda \tag{2.3}$$

For the Markov process X_1', transitions occur with mean rate Λ. When a transition occurs, the system jumps out from the current state x with the probability of $1 - P'(x,x)$ and stays in state x with the probability of $P'(x,x)$. Thus, the mean rate of jumping out of state x is $\Lambda\left[1 - P'(x,x)\right]$ and that of staying in state x is $\Lambda P'(x,x)$. Moreover, if the system jumps out of state x, the probability of entering state $x+1$ is $P'(x,x+1)/\left[1 - P'(x,x)\right]$ and that of entering state $x-1$ is $P'(x,x-1)/\left[1 - P'(x,x)\right]$. Therefore, the Markov process X_1' has the transition rates matrix $\boldsymbol{A'}$ as follows:

$$A'(x, x+1) = \Lambda\left[1 - P'(x,x)\right] P'(x, x+1)/\left[1 - P'(x,x)\right] = \mu_k \tag{2.4}$$

$$A'(x, x-1) = \Lambda\left[1 - P'(x,x)\right] P'(x, x-1)/\left[1 - P'(x,x)\right] = \lambda \tag{2.5}$$

$$A'(x,x) = -\left[A'(x,x+1) + A'(x,x-1)\right] = -\mu_k - \lambda \tag{2.6}$$

It can be seen that the Markov processes X_1 and X_1' have the same transition rates matrices (see Çinlar [5]). Given a transition rates matrix, one continuous-time Markov process can be uniquely determined. Therefore, the underlying stochastic processes of the Markov processes X_1 and X_1' are the same and thus X_1' has the same optimal control policy and then the same optimal return function to that of X_1; see Lippman [18]. For the Markov process X_1', the mean time length between successive transitions Λ is constant and independent of states and control policies employed. Henceforth, we analyze X_1' to identify the optimal control policy.

Denote α as the interest rate. First, we compute as follows the expected discounted cost incurred during one-stage transition of the Markov process X_1' where the current state is x and the current production rate employed is μ_k, $k = 0, 1, 2$.

$$\begin{aligned}
& \int_0^\infty \int_0^T e^{-\alpha t}(h[x]^+ + b[x]^-)\Lambda e^{-\Lambda T} dt dT + \frac{\mu_k}{\Lambda}\int_0^\infty \left(e^{-\alpha T}c_k\right)\Lambda e^{-\Lambda T} dT \\
= \; & (h[x]^+ + b[x]^-)\int_0^\infty \Lambda e^{-\Lambda T} dT \int_0^T e^{-\alpha t} dt + \mu_k c_k \int_0^\infty e^{-(\alpha+\Lambda)T} dT \\
= \; & \frac{(h[x]^+ + b[x]^-)}{\alpha}\int_0^\infty \Lambda e^{-\Lambda T}(1 - e^{-\alpha T}) dT + \frac{\mu_k c_k}{\alpha + \Lambda} \\
= \; & \frac{(h[x]^+ + b[x]^-)}{\alpha}\left(\int_0^\infty \Lambda e^{-\Lambda T} dT - \int_0^\infty \Lambda e^{-(\alpha+\Lambda)T} dT\right) + \frac{\mu_k c_k}{\alpha + \Lambda} \\
= \; & \frac{(h[x]^+ + b[x]^-)}{\alpha}\left(1 - \frac{\Lambda}{\Lambda + \alpha}\right) + \frac{\mu_k c_k}{\alpha + \Lambda} \\
= \; & \frac{h[x]^+ + b[x]^- + \mu_k c_k}{\Lambda + \alpha}
\end{aligned} \tag{2.7}$$

where $[x]^+ = \max\{0, x\}$, $[x]^- = \max\{0, -x\}$.

Now, we consider the first n stages of the infinite horizon problem by truncation. Denote $f_j^n(x)$ as, evaluated at the beginning of stage j with $1 \leq j \leq n$, the minimal expected total discounted cost in stages j through n given that the starting state is x. Let $f_{n+1}^n(x)$ be the terminal value function applied at the end of stage n if the ending state is x. Given that the state at stage j is x and the production rate employed is μ_k, the expected total discounted cost in stages $j+1$ through n is given by

$$
\begin{aligned}
& \frac{\lambda}{\Lambda}\int_0^\infty f_{j+1}^n(x-1)e^{-\alpha T}\Lambda e^{-\Lambda T}dT + \frac{\mu_k}{\Lambda}\int_0^\infty f_{j+1}^n(x+1)e^{-\alpha T}\Lambda e^{-\Lambda T}dT \\
& + \frac{\mu_2-\mu_k}{\Lambda}\int_0^\infty f_{j+1}^n(x)e^{-\alpha T}\Lambda e^{-\Lambda T}dT \\
= \; & \lambda f_{j+1}^n(x-1)\int_0^\infty e^{-(\alpha+\Lambda)T}dT + \mu_k f_{j+1}^n(x+1)\int_0^\infty e^{-(\alpha+\Lambda)T}dT \\
& + (\mu_2-\mu_k)f_{j+1}^n(x)\int_0^\infty e^{-(\alpha+\Lambda)T}dT \\
= \; & \frac{\lambda}{\alpha+\Lambda}f_{j+1}^n(x-1) + \frac{\mu_k}{\alpha+\Lambda}f_{j+1}^n(x+1) + \frac{\mu_2-\mu_k}{\alpha+\Lambda}f_{j+1}^n(x)
\end{aligned}
$$

Because we can always re-scale the time unit, without loss of generality, it is assumed that $\Lambda + \alpha = 1$. To minimize the expected total discounted cost at stage j, $f_j^n(x)$ is computed recursively as follows.

$$
\begin{aligned}
f_j^n(x) &= \min_{k=0,1,2}\Big\{h[x]^+ + b[x]^- + \mu_k c_k + \lambda f_{j+1}^n(x-1) \\
&\qquad + \mu_k f_{j+1}^n(x+1) + (\mu_2-\mu_k)f_{j+1}^n(x)\Big\} \\
&= h[x]^+ + b[x]^- + \lambda f_{j+1}^n(x-1) \\
&\quad + \mu_2 f_{j+1}^n(x) + \min\left\{\begin{array}{l} \mu_1\left[f_{j+1}^n(x+1) - f_{j+1}^n(x) + c_1\right] \\ \mu_2\left[f_{j+1}^n(x+1) - f_{j+1}^n(x) + c_2\right] \\ 0 \end{array}\right\}
\end{aligned} \tag{2.8}
$$

Let $f(x)$ be the minimal expected total discounted cost over an infinite horizon with the starting state x. According to Theorem 11.3 of Porteus [24], $f(x) = \lim_{n\to\infty} f_j^n(x)$ and $f(x)$ satisfies the following optimality equation.

$$\begin{aligned} f(x) &= h[x]^+ + b[x]^- + \lambda f(x-1) \\ &\quad + \mu_2 f(x) + \min \left\{ \begin{array}{l} \mu_1 \left[f(x+1) - f(x) + c_1\right] \\ \mu_2 \left[f(x+1) - f(x) + c_2\right] \\ 0 \end{array} \right\} \end{aligned} \tag{2.9}$$

It is easy to see that the optimal control decision is to select the production rate that minimizes Equation 2.9. By analyzing the last term of this equation, the following lemma is obtained.

Lemma 2.1 *The optimal control decision is*

1. *not to produce if* $f(x) - f(x+1) \leq c_1$*,*

2. *to produce normally if* $c_1 \leq f(x) - f(x+1) \leq (\mu_2 c_2 - \mu_1 c_1)/(\mu_2 - \mu_1)$*, and*

3. *to produce urgently if* $f(x) - f(x+1 \geq (\mu_2 c_2 - \mu_1 c_1)/(\mu_2 - \mu_1)$*.*

Proof. Since $\mu_2 c_2 - \mu_1 c_1 > \mu_2 c_2 - \mu_1 c_2$, it follows that $(\mu_2 c_2 - \mu_1 c_1)/(\mu_2 - \mu_1) > c_2 > c_1$. By analyzing the last term of Equation 2.9, it is optimal not to produce if 0 is the minimum item, which is equivalent to $f(x) - f(x+1) \leq c_1$. It is optimal to produce normally if $\mu_1 \left[f(x+1) - f(x) + c_1\right]$ is the minimum one instead, which is equivalent to $c_1 \leq f(x) - f(x+1) \leq (\mu_2 c_2 - \mu_1 c_1)/(\mu_2 - \mu_1)$. Similarly, it is optimal to produce urgently if $\mu_2 [f(x+1) - f(x) + c_2]$ is the minimum one, which is equivalent to $f(x) - f(x+1) \geq (\mu_2 c_2 - \mu_1 c_1)/(\mu_2 - \mu_1)$. □

The emergency production rate μ_2 can be viewed as a combination of the normal rate μ_1 and an additional rate $\mu_2 - \mu_1$. Due to the lower unit production cost c_1, the normal rate is always employed to produce. However, if needed, an additional production rate $\mu_2 - \mu_1$ can be added in with a higher unit production cost $(\mu_2 c_2 - \mu_1 c_1)/(\mu_2 - \mu_1)$ to expedite stock replenishment. In Lemma 2.1, the difference $f(x) - f(x+1)$ is the cost saving when the net inventory level is increased by one. If the cost saving does not justify the unit normal production cost c_1, we should not produce at all; otherwise, the system cost would not be minimized. If the cost saving exceeds the unit normal production cost c_1, we should produce either normally or urgently. If the cost saving is smaller than $(\mu_2 c_2 - \mu_1 c_1)/(\mu_2 - \mu_1)$, i.e., the cost saving can not justify the higher production cost for an additional production rate, we should produce normally. If the cost saving is greater than $(\mu_2 c_2 - \mu_1 c_1)/(\mu_2 - \mu_1)$, the emergency production rate should be employed to expedite inventory replenishment.

2.1.2. The Optimal Control Policy

Let V be the collection of the real-valued convex functions defined on Z. Define H as the operator applied on $v \in V$ such that

$$Hv(x) \quad = \quad h[x]^+ + b[x]^- + \lambda v(x-1)$$

$$+\mu_2 v(x) + \min \left\{ \begin{array}{l} \mu_1 \left[v(x+1) - v(x) + c_1 \right] \\ \mu_2 \left[v(x+1) - v(x) + c_2 \right] \\ 0 \end{array} \right\} \tag{2.10}$$

Lemma 2.2 shows that the operator H preserves the convexity of the function v.

Lemma 2.2 *If v is convex, then Hv is also convex.*

Proof. First, $h[x]^+ + b[x]^- + \lambda v(x-1)$ is convex since v is assumed to be convex. Then, we only need to show that $\mu_2 v(x) + \min_{k=0,1,2} \{\mu_k [v(x+1) - v(x) + c_k]\}$, defined as $F(x)$, is convex. Let

$$\bar{k} = \arg \min_{k=0,1,2} \{\mu_k [v(x+3) - v(x+2) + c_k]\}$$

$$\underline{k} = \arg \min_{k=0,1,2} \{\mu_k [v(x+1) - v(x) + c_k]\}$$

Then,

$$\begin{aligned} & F(x+2) - F(x+1) \\ = \ & \mu_2 v(x+2) + \mu_{\bar{k}} [v(x+3) - v(x+2) + c_{\bar{k}}] \\ & - \mu_2 v(x+1) - \min_{k=0,1,2} \{\mu_k [v(x+2) - v(x+1) + c_k]\} \\ \geq \ & \mu_2 v(x+2) + \mu_{\bar{k}} [v(x+3) - v(x+2) + c_{\bar{k}}] \\ & - \mu_2 v(x+1) - \mu_{\bar{k}} [v(x+2) - v(x+1) + c_{\bar{k}}] \\ = \ & \mu_2 [v(x+2) - v(x+1)] \\ & + \mu_{\bar{k}} \{[v(x+3) - v(x+2)] - [v(x+2) - v(x+1)]\} \end{aligned}$$

and

$$
\begin{aligned}
& F(x+1) - F(x) \\
= \;& \mu_2 v(x+1) + \min_{k=0,1,2} \{\mu_k [v(x+2) - v(x+1) + c_k]\} \\
& - \mu_2 v(x) - \mu_{\underline{k}} \left[v(x+1) - v(x) + c_{\underline{k}}\right] \\
\leq \;& \mu_2 v(x+1) + \mu_{\underline{k}} \left[v(x+2) - v(x+1) + c_{\underline{k}}\right] \\
& - \mu_2 v(x) - \mu_{\underline{k}} \left[v(x+1) - v(x) + c_{\underline{k}}\right] \\
= \;& \mu_2 [v(x+1) - v(x)] + \mu_{\underline{k}} \{[v(x+2) - v(x+1)] - [v(x+1) - v(x)]\}
\end{aligned}
$$

Thus,

$$
\begin{aligned}
& [F(x+2) - F(x+1)] - [F(x+1) - F(x)] \\
\geq \;& \mu_2 [v(x+2) - v(x+1)] + \mu_{\bar{k}} \{[v(x+3) - v(x+2)] - [v(x+2) - v(x+1)]\} \\
& - \mu_2 [v(x+1) - v(x)] - \mu_{\underline{k}} \{[v(x+2) - v(x+1)] - [v(x+1) - v(x)]\} \\
= \;& \left(\mu_2 - \mu_{\underline{k}}\right) \{[v(x+2) - v(x+1)] - [v(x+1) - v(x)]\} \\
& + \mu_{\overline{k}} \{[v(x+3) - v(x+2)] - [v(x+2) - v(x+1)]\} \\
\geq \;& 0
\end{aligned}
$$

The last inequality comes from $\mu_2 - \mu_{\underline{k}} \geq 0$ and the convexity of v. Hence, $F(x)$ is convex, and it follows that Hv is also convex. □

Based on Lemmas 2.1 and 2.2, we have the following theorem:

Theorem 2.1 *1. The minimal expected total discounted cost function $f(x)$ is convex with respect to the net inventory level x.*

2. Define

$S_1 = \min\{\, x : f(x) - f(x+1) \le c_1 \,\}$

$S_2 = \min\{\, x : f(x) - f(x+1) \le (\mu_2 c_2 - \mu_1 c_1)/(\mu_2 - \mu_1) \,\}$

There exists a stationary optimal policy, denoted as (S_1, S_2) policy, such that it is optimal not to produce if the net inventory level is at or above S_1, to produce normally if the net inventory level is below S_1 and at or above S_2, and to produce urgently if the net inventory level is below S_2.

Proof. We prove this theorem based on Theorem 11.5 of Porteus [24]. Define the set of structured decision rules as all the decision rules with the form given by part 2 of the theorem while S_1 and S_2 can take any integers. Define the set of structured value functions as all the convex functions, which essentially is the set V. Because the limit of any convergent sequence of functions in V will be in V as well, the set V is complete. Moreover, from Lemma 2.2, the operator H preserves the structure of V. Therefore, the optimal return function f must be structured, i.e., it is convex as well. From the optimality equation 2.9, it can be seen that the structured decision rule with S_1 and S_2 defined in the theorem is optimal for the one-stage minimization problem. Thus, the control policy developed in the theorem is optimal. Because the production system is stationary, i.e., the system equation, the cost per stage, the demand process, and the production process do not change from one stage to the next, the optimal control policy is stationary. □

Figure 2.2 illustrates the (S_1, S_2) control policy. Due to the convexity of $f(x)$, $f(x) - f(x+1)$ is non-increasing with respect to x. The state space Z of the production system is partitioned into three areas by the pair (S_1, S_2), each of which corresponds

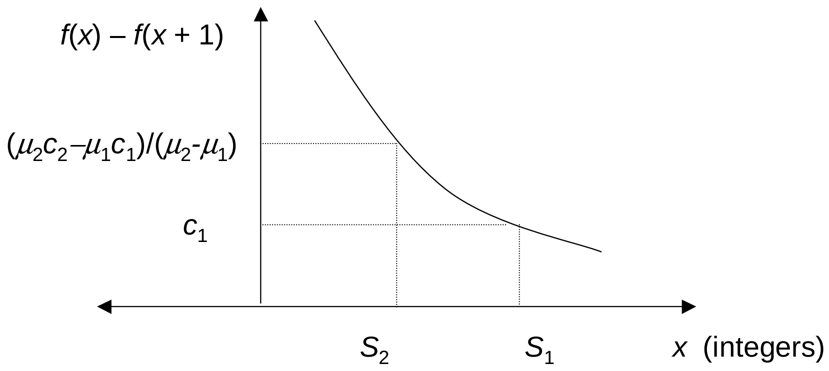

Figure 2.2: The illustration of the (S_1, S_2) policy

to a different production decision respectively: not to produce, to produce normally and to produce urgently.

2.2. Stationary Analysis of the Production System

In this section, the expected total cost per unit time is computed for the production system developed in the previous section. Under the (S_1, S_2) policy, this production system can be considered as an $M/M/1/S$ queueing system with state-dependent arrival rates. In this queueing system, the net inventory level is considered as the number of customers waiting for service except that it may take on negative integers. Production completion is represented as arrival to the queueing system and customer demand is modelled as service of the system. The service rate is equivalent to the customer demand rate. The arrival rate to the queueing system corresponds to the production rate, which varies with the net inventory level, i.e., the arrival rate is 0 if the net inventory level is greater than or equal to S_1, μ_1 if the net inventory level drops below S_1 but greater than or equal to S_2, and μ_2 if the net inventory level drops below S_2. Figure 2.3 shows the rate diagram of this $M/M/1/S$ queue.

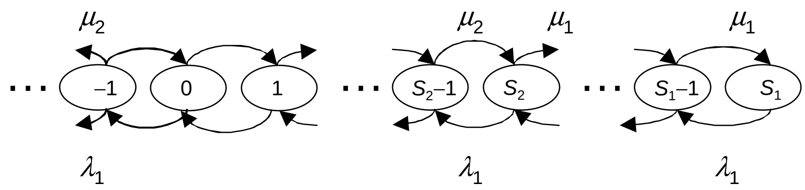

Figure 2.3: Rate diagram for the $M/M/1/S$ queueing system

Let $\rho_1 = \lambda/\mu_1$ and $\rho_2 = \lambda/\mu_2$ be the utilizations for the normal and emergency production rate, respectively. We assume that $\rho_2 < 1$ to guarantee the existence of steady states. Define $\pi(n)$ as the steady-state probability where the net inventory level is n. It can be obtained that $\pi(n)$ is given by

$$\pi(n) = \begin{cases} \rho_1^{S_1-n}\pi(S_1), & \text{for } S_2 \le n < S_1 \\ \rho_2^{S_2-n}\rho_1^{S_1-S_2}\pi(S_1), & \text{for } n < S_2 \end{cases} \tag{2.11}$$

where

$$\pi(S_1) = \frac{(1-\rho_1)(1-\rho_2)}{1-\rho_2+\rho_1^{S_1-S_2}(\rho_2-\rho_1)} \tag{2.12}$$

To compute the expected total cost per unit time, the performance measures of the queueing model are needed. Under the (S_1, S_2) policy, define

$I(S_1, S_2) = \sum_{n=1}^{S_1} n\pi(n)$ as the expected on-hand inventory level,

$B(S_1, S_2) = \sum_{n=1}^{\infty} n\pi(-n)$ as the expected number of backorders,

$P_1(S_1, S_2) = \sum_{n=S_2}^{S_1-1} \pi(n)$ as the probability of the normal rate employed, and

$P_2(S_1, S_2) = \sum_{n=-\infty}^{S_2-1} \pi(n)$ as the probability of the emergency rate employed.

Now we compute the expected on-hand inventory level $I(S_1, S_2)$. Because S_2 can be negative, there are two cases to be considered next.

1. If $S_2 < 0$, then

$$\begin{aligned} I(S_1,S_2) &= \pi(1)+2\pi(2)+\cdots+(S_1-1)\pi(S_1-1)+S_1\pi(S_1) \\ &= \rho_1^{S_1-1}\pi(S_1)+2\rho_1^{S-2}\pi(S_1)+\cdots+(S_1-1)\rho_1\pi(S_1)+S_1\pi(S_1) \end{aligned}$$

and

$$\rho_1 I(S_1,S_2)=\rho_1^{S_1}\pi(S_1)+2\rho_1^{S_1-1}\pi(S_1)+\cdots+(S_1-1)\rho_1^2\pi(S_1)+S_1\rho_1\pi(S_1)$$

Thus,

$$\begin{aligned} I(S_1,S_2) &= \pi(S_1)\left[S_1-\left(\rho_1^{S_1}+\rho_1^{S_1-1}+\cdots+\rho_1\right)\right]/(1-\rho_1) \\ &= \pi(S_1)\left[\frac{S_1}{1-\rho_1}-\frac{\rho_1\left(1-\rho_1^{S_1}\right)}{(1-\rho_1)^2}\right] \end{aligned}$$

2. If $S_2 \geq 0$, then $I(S_1,S_2)=\sum_{n=1}^{S_2-1} n\pi(n)+\sum_{n=S_2}^{S_1} n\pi(n)$. Let $G_1=\sum_{n=1}^{S_2-1} n\pi(n)$ and $G_2=\sum_{n=S_2}^{S_1} n\pi(n)$.

$$\begin{aligned} G_1 &= \pi(1)+2\pi(2)+\cdots+(S_2-1)\pi(S_2-1) \\ &= \rho_2^{S_2-1}\rho_1^{S_1-S_2}\pi(S_1)+2\rho_2^{S_2-2}\rho_1^{S_1-S_2}\pi(S_1)+\cdots+(S_2-1)\rho_2\rho_1^{S_1-S_2}\pi(S_1) \end{aligned}$$

Then,

$$\rho_2 G_1=\rho_2^{S_2}\rho_1^{S_1-S_2}\pi(S_1)+2\rho_2^{S_2-1}\rho_1^{S_1-S_2}\pi(S_1)+\cdots+(S_2-1)\rho_2^2\rho_1^{S_1-S_2}\pi(S_1)$$

Thus,

$$\begin{aligned} G_1 &= \rho_2\rho_1^{S_1-S_2}\pi(S_1)\left[S_2-\left(\rho_2^{S_2-1}+\cdots+\rho_2+1\right)\right]/(1-\rho_2) \\ &= \frac{S_2(1-\rho_2)-\left(1-\rho_2^{S_2}\right)}{\left(1-\rho_2\right)^2}\rho_2\rho_1^{S_1-S_2}\pi(S_1) \end{aligned}$$

Similarly,

$$\begin{aligned} G_2 &= S_2\pi(S_2)+(S_2+1)\pi(S_2+1)+\cdots+S_1\pi(S_1) \\ &= S_2\rho_1^{S_1-S_2}\pi(S_1)+(S_2+1)\rho_1^{S_1-S_2-1}\pi(S_1)+\cdots+S_1\pi(S_1) \end{aligned}$$

Then,

$$\rho_1 G_2 = S_2 \rho_1^{S_1-S_2+1}\pi(S_1) + (S_2+1)\rho_1^{S_1-S_2}\pi(S_1) + \cdots + S_1\rho_1\pi(S_1)$$

Thus,

$$\begin{aligned} G_2 &= \pi(S_1)\left[S_1 - S_2\rho_1^{S_1-S_2+1} - \left(\rho_1^{S_1-S_2} + \cdots + \rho_1\right)\right]/(1-\rho_1) \\ &= \frac{S_1(1-\rho_1) - S_2\rho_1^{S_1-S_2+1}(1-\rho_1) - \rho_1\left(1-\rho_1^{S_1-S_2}\right)}{(1-\rho_1)^2}\pi(S_1) \end{aligned}$$

So,

$$\begin{aligned} I(S_1,S_2) &= \frac{S_2(1-\rho_2) - (1-\rho_2^{S_2})}{(1-\rho_2)^2}\rho_2\rho_1^{S_1-S_2}\pi(S_1) \\ &\quad + \frac{\left(S_1 - S_2\rho_1^{S_1-S_2+1}\right)(1-\rho_1) - \rho_1\left(1-\rho_1^{S_1-S_2}\right)}{(1-\rho_1)^2}\pi(S_1) \end{aligned}$$

Therefore, combining the two cases together, it is obtained that

$$I(S_1,S_2) = \begin{cases} \left[\begin{array}{l} \frac{\left[S_2(1-\rho_2)-(1-\rho_2^{S_2})\right]\rho_2\rho_1^{S_1-S_2}}{(1-\rho_2)^2} + \\ \frac{\left(S_1 - S_2\rho_1^{S_1-S_2+1}\right)(1-\rho_1) - \rho_1\left(1-\rho_1^{S_1-S_2}\right)}{(1-\rho_1)^2} \end{array}\right]\pi(S_1), & \text{if } S_2 \geq 0 \\ \left[\frac{S_1}{1-\rho_1} - \frac{\rho_1\left(1-\rho_1^{S_1}\right)}{(1-\rho_1)^2}\right]\pi(S_1), & \text{if } S_2 < 0 \end{cases} \tag{2.13}$$

Next, we compute the expected number of backorders $B(S_1, S_2)$. Similarly, two cases are to be considered:

1. If $S_2 < 0$, then $B(S_1,S_2) = \sum_{n=S_2}^{-1}(-n)\pi(n) + \sum_{n=-\infty}^{S_2-1}(-n)\pi(n)$. Let $G_3 = \sum_{n=S_2}^{-1}(-n)\pi(n)$ and $G_4 = \sum_{n=-\infty}^{S_2-1}(-n)\pi(n)$.

$$\begin{aligned} G_3 &= (-S_2)\pi(S_2) + (-S_2-1)\pi(S_2+1) + \cdots + \pi(-1) \\ &= (-S_2)\rho_1^{S_1-S_2}\pi(S_1) + (-S_2-1)\rho_1^{S_1-S_2-1}\pi(S_1) + \cdots + \rho_1^{S_1+1}\pi(S_1) \end{aligned}$$

Then,

$$\rho_1 G_3 = (-S_2)\rho_1^{S_1-S_2+1}\pi(S_1) + (-S_2-1)\rho_1^{S_1-S_2}\pi(S_1) + \cdots + \rho_1^{S_1+2}\pi(S_1)$$

Thus,

$$\begin{aligned}
G_3 &= \rho_1^{S_1+1}\pi(S_1)\left[S_2\rho_1^{-S_2} + \rho_1^{-S_2-1} + \cdots + 1\right]/(1-\rho_1) \\
&= \pi(S_1)\rho_1^{S_1+1}\left[\frac{S_2\rho_1^{-S_2}}{1-\rho_1} + \frac{1-\rho_1^{-S_2}}{(1-\rho_1)^2}\right]
\end{aligned}$$

Similarly,

$$\begin{aligned}
G_4 &= (-S_2+1)\pi(S_2-1) + (-S_2+2)\pi(S_2-2) + \cdots \\
&= (-S_2+1)\rho_2\rho_1^{S_1-S_2}\pi(S_1) + (-S_2+2)\rho_2^2\rho_1^{S_1-S_2}\pi(S_1) + \cdots
\end{aligned}$$

Then,

$$\rho_2 G_4 = (-S_2+1)\rho_2^2\rho_1^{S_1-S_2}\pi(S_1) + (-S_2+2)\rho_2^3\rho_1^{S_1-S_2}\pi(S_1) + \cdots$$

Thus,

$$\begin{aligned}
G_4 &= \rho_2\rho_1^{S_1-S_2}\pi(S_1)\left[-S_2 + 1 + \rho_2 + \rho_2^2 + \cdots\right]/((1-\rho_2)) \\
&= \pi(S_1)\rho_2\rho_1^{S_1-S_2}\left[\frac{-S_2}{1-\rho_2} + \frac{1}{(1-\rho_2)^2}\right]
\end{aligned}$$

Therefore,

$$B(S_1,S_2) = \pi(S_1)\left[\frac{S_2\rho_1^{S_1-S_2+1}}{1-\rho_1} + \frac{\rho_1^{S_1+1} - \rho_1^{S_1-S_2+1}}{(1-\rho_1)^2} - \frac{S_2\rho_1^{S_1-S_2}\rho_2}{1-\rho_2} + \frac{\rho_1^{S_1-S_2}\rho_2}{(1-\rho_2)^2}\right]$$

2. If $S_2 \geq 0$, then

$$\begin{aligned}
B(S_1,S_2) &= \pi(-1) + 2\pi(-2) + 3\pi(-3) + \cdots \\
&= \rho_2^{S_2+1}\rho_1^{S_1-S_2}\pi(S_1) + 2\rho_2^{S_2+2}\rho_1^{S_1-S_2}\pi(S_1) + 3\rho_2^{S_2+3}\rho_1^{S_1-S_2}\pi(S_1) + \cdots
\end{aligned}$$

and

$$\rho_2 B(S_1,S_2) = \rho_2^{S_2+2}\rho_1^{S_1-S_2}\pi(S_1) + 2\rho_2^{S_2+3}\rho_1^{S_1-S_2}\pi(S_1) + 3\rho_2^{S_2+4}\rho_1^{S_1-S_2}\pi(S_1) + \cdots$$

Thus,

$$\begin{aligned} B(S_1,S_2) &= \rho_2^{S_2+1}\rho_1^{S_1-S_2}\pi(S_1)\left[1+\rho_2+\rho_2^2+\cdots\right]/(1-\rho_2) \\ &= \frac{\rho_2^{S_2+1}\rho_1^{S_1-S_2}}{(1-\rho_2)^2}\pi(S_1) \end{aligned}$$

Therefore, combining the two cases together, it can be obtained that

$$B(S_1,S_2)=\begin{cases} \frac{\rho_2^{S_2+1}\rho_1^{S_1-S_2}}{(1-\rho_2)^2}\pi(S_1), & \text{if } S_2\geq 0 \\ \left[\begin{array}{l} \frac{S_2\rho_1^{S_1-S_2+1}}{1-\rho_1}+\frac{\rho_1^{S_1+1}-\rho_1^{S_1-S_2+1}}{(1-\rho_1)^2} \\ -\frac{S_2\rho_1^{S_1-S_2}\rho_2}{1-\rho_2}+\frac{\rho_1^{S_1-S_2}\rho_2}{(1-\rho_2)^2} \end{array}\right]\pi(S_1), & \text{if } S_2<0 \end{cases} \tag{2.14}$$

Finally, the probabilities of the normal and emergency rate employed $P_1(S_1,S_2)$ and $P_2(S_1,S_2)$ are computed as follows:

$$P_1(S_1,S_2)=\pi(S_1)\left(\rho_1+\rho_1^2+\cdots+\rho_1^{S_1-S_2}\right)=\frac{\rho_1\left(1-\rho_1^{S_1-S_2}\right)}{1-\rho_1}\pi(S_1) \tag{2.15}$$

$$P_2(S_1,S_2)=\pi(S_1)\rho_1^{S_1-S_2}\rho_2\left(1+\rho_2+\rho_2^2+\cdots\right)=\frac{\rho_1^{S_1-S_2}\rho_2}{1-\rho_2}\pi(S_1) \tag{2.16}$$

It can be seen that $\mu_1P_1(S_1,S_2)$ and $\mu_2P_2(S_1,S_2)$ are the expected numbers per unit time of the normal production and the emergency production, respectively. Therefore, the expected total cost per unit time is computed as follows:

$$C(S_1,S_2)=h\,I(S_1,S_2)+bB(S_1,S_2)+c_1\mu_1P_1(S_1,S_2)+c_2\mu_2P_2(S_1,S_2) \tag{2.17}$$

2.3. Numerical Study

In this section, we investigate the benefit of the production system with two rates over the one with a single rate under different operating conditions. In this study, we set the normal rate of the production system with two rates equal to the rate of the production system with a single rate. That is, the benefit can be viewed as a cost saving of providing the single rate production system with an emergency production rate.

The cost formula for the production system with two rates is obtained in the previous section. For the production system with a single rate, it is well known that the base-stock policy is optimal; see Li [17]. Let S be the base-stock level for the production system with a single rate, the expected total cost per unit time can be computed in a straightforward manner through $M/M/1/S$. To guarantee the existence of steady states, it is assumed that $\rho_1 = \lambda/\mu_1 < 1$. Define $I(S)$ and $B(S)$ as the expected on-hand inventory level and the expected number of backorders with the base-stock level S, respectively. Then,

$$I(S) = S - \frac{\rho_1 - \rho_1^{S+1}}{1 - \rho_1} \tag{2.18}$$

$$B(S) = \frac{\rho_1^{S+1}}{1 - \rho_1} \tag{2.19}$$

Therefore, the expected total cost is

$$C(S) = h\,I(S) + bB(S) + c_1\lambda \tag{2.20}$$

It is easy to show that $C(S)$ is convex with respect to the base-stock level S.

Define $c = c_2 - c_1$ as the difference of unit production costs between the normal and emergency production rate. For the backordering case, all demands must be satisfied; thus, $\mu_1 P_1(S_1, S_2) + \mu_2 P_2(S_1, S_2) = \lambda$. Therefore, Equation 2.17 becomes

$$C(S_1, S_2) = h\, I(S_1, S_2) + bB(S_1, S_2) + c_1\lambda + c\mu_2 P_2(S_1, S_2) \tag{2.21}$$

By dropping the cost components $c_1\lambda$ from Equations 2.20 and 2.21, the corresponding total relevant costs *TRC* affected by control policies are as follows.

$$TRC(S_1, S_2) = h\, I(S_1, S_2) + bB(S_1, S_2) + c\mu_2 P_2(S_1, S_2) \tag{2.22}$$

$$TRC(S) = h\, I(S) + bB(S) \tag{2.23}$$

For a given set of parameters, let S_1^* and S_2^* be the optimal critical inventory levels for the production system with two rates, and let S^* be the optimal base-stock level for the production system with a single rate. Define the relative cost saving, *CS*, as the following percentage:

$$CS = \frac{TRC(S^*) - TRC(S_1^*, S_2^*)}{TRC(S^*)} \times 100\% \tag{2.24}$$

which is a function of the parameters μ_1, μ_2, λ, h, b and c. The larger *CS* is, the more beneficial it is to employ the emergency production rate. After some manipulations, *CS* can also be expressed in terms of μ_1, ρ_1, μ_2/μ_1, h, b and c. Without loss of

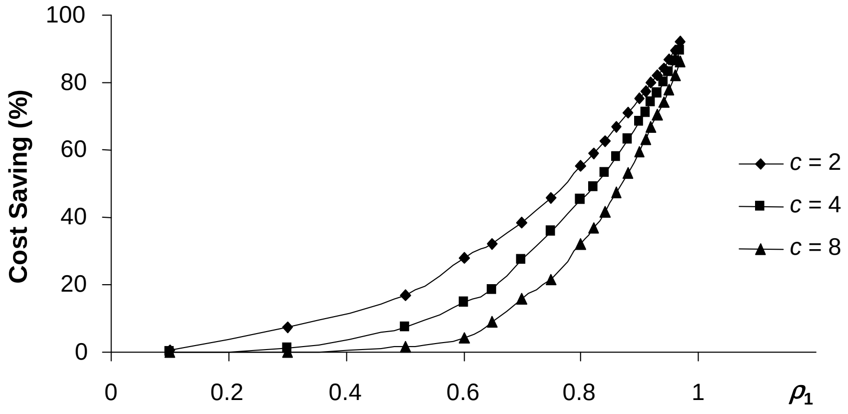

Figure 2.4: The effect of ρ_1 over cost saving

generality, it is assumed that $\mu_1 = 1$. We seek to find how the parameters ρ_1, μ_2/μ_1, h, b and c affect CS and try to identify those having significant influences on CS under different operating conditions.

For a given problem instance, the optimal solutions (S_1^*, S_2^*) and S^* can be found by exhaustive search over a large range of S_1, S_2, and S. However, $TRC(S_1, S_2)$ appears to be convex in the three-dimension graphs plotted although we can not prove its convexity analytically. To make the search simpler and more efficient, the solver function in Microsoft Excel is employed, which uses the Generalized Reduced Gradient method. It is found that results can be obtained very quickly on a personal computer. Initially, we set that $\rho_1 = 0.95$, $\mu_2/\mu_1 = 1.8$, $h = 1$, $b = 2$ and $c = 4$. Based on the initial setting, we compute CS over a range of 20 values of each of the five parameters for three different values of another parameter, while the other three parameters remain unchanged. The results are shown in Figures 2.4 to 2.8.

Figure 2.4 shows that the cost saving increases as the parameter ρ_1 increases. At a large ρ_1, the production system with only normal rate keeps high inventory, i.e. the base-stock level is high. Even with a high base-stock level, the expected number of backorders is large because demands backordered can not be satisfied quickly due to the limited production capacity. After the introduction of the emergency production rate, the production system reduces both the base-stock level and the expected number of backorders significantly. Consequently, the total cost is greatly reduced. For example, with $\rho_1 = 0.95$ and $c = 8$, the base-stock level and the expected number of backorders of the production system with a single rate are 21 and 6.47 compared to 4 and 0.56 of the production system with two rates. This results in a cost saving of 78.1%. At a small ρ_1, the production system with only normal rate maintains low inventory and the expected number of backorders is small. Although the emergency production rate is available, it is barely used. Thus, introducing an emergency production does not reduce the base-stock level and expected number of backorders significantly. Therefore, the cost saving is insignificant. For example, with $\rho_1 = 0.5$ and $c = 8$, the base-stock level and the expected number of backorders of the production system with a single rate are 1 and 0.5 compared to 1 and 0.44 which results in a cost saving of 1.5%.

It is intuitive that the cost saving increases as the production rates ratio μ_2/μ_1 increases (see Figure 2.5). The cost saving can be significant even for small values of μ_2/μ_1. For example, when $\mu_2/\mu_1 = 1.05$, the cost saving is 39% for the case where $\rho_1 = 0.95$. It can be seen from the figure that once the ratio μ_2/μ_1 is large enough beyond a certain threshold, there is only a small improvement in the cost saving as we continue to increase the emergency production rate. This threshold is useful for selecting an appropriate emergency production rate in the real life.

Figure 2.6 shows that the cost saving decreases as the unit production costs difference c increases. This is because the emergency production rate is used less frequently

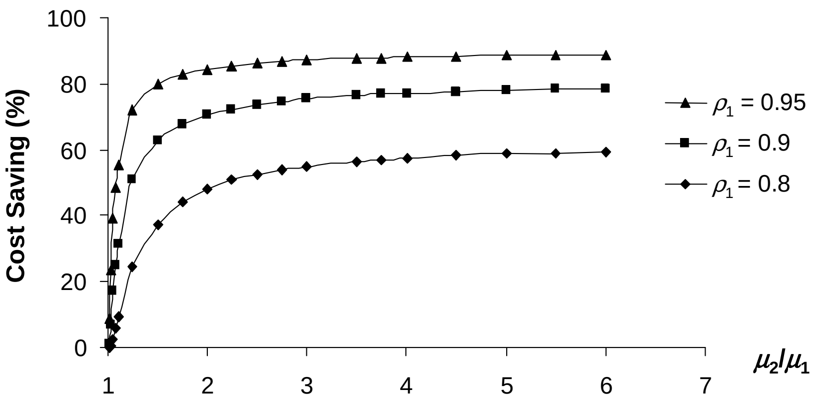

Figure 2.5: The effect of μ_2/μ_1 over cost saving

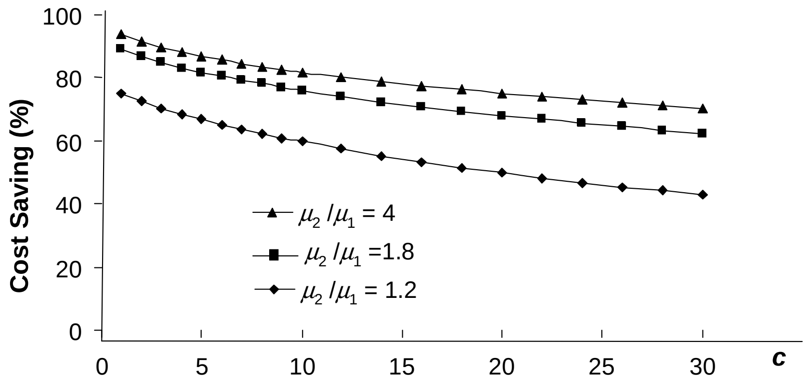

Figure 2.6: The effect of c over cost saving

because of the higher production cost. Interestingly, even at a high value of c, the cost saving can be significant. For example, with $c = 30$ and $\mu_2/\mu_1 = 1.2$, the cost saving is as high as 43%. This is because the emergency production rate still can reduce the base-stock level and expected number of backorders greatly from 21 and 6.47 to 12 and 1.82.

Finally, Figures 2.7 and 2.8 show that the cost saving increases as the inventory holding cost rate h or the backordering cost rate b increases. When h is large, the system with a single rate try to avoid the high holding cost with the expense of backordering cost. The system with two rate can reduce the number of backordered more efficiently. As a result, the base-stock level and the expected number of backorders

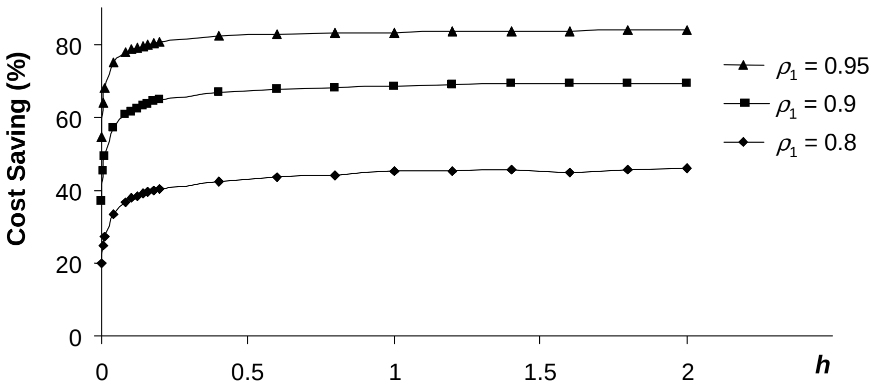

Figure 2.7: The effect of h over cost saving

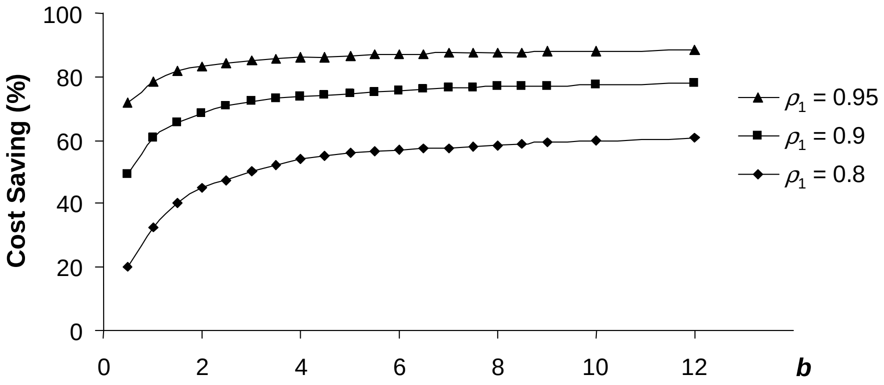

Figure 2.8: The effect of b over cost saving

are reduced. Thus, the cost saving achieved is substantial. For example, with $h = 2$ and $\rho_1 = 0.95$, the base-stock level and the expected number of backorders of the system with a single rate are 13 and 9.75 compared to 2 and 0.82 which results in a cost saving of 83.75%. According to the experiment, the cost saving can be significant even with a small value of h. On the other hand, if b is large, the system with a single rate try to avoid a costly backordering cost by holding a high amount of inventory. The system with two rate can significantly reduce the base-stock level because it can reduce the number of backorders more quickly. Thus, the cost saving achieved is substantial as well.

To summarize, we find that providing the production system with a higher production rate can always reduce the expected number of backorders because backorders can be satisfied more quickly. Consequently, it also brings down the base-stock level. The magnitude of the cost saving depends on how much the expected number of backorders and the base-stock level can be reduced. In this numerical study, the average cost saving over all problems tested is more than 60%.

2.4. Production System with Multiple Production Rates

In this section, the optimal control policy for the production system with N production rates is studied. Suppose that the production time of the k^{th} production rate is exponentially distributed with mean $1/\mu_k$ and the corresponding unit production cost is $c_k, k = 0, 1, 2, \ldots, N$. Without loss of generalization, it is assumed that $\mu_0 = 0 < \mu_1 < \mu_2 < \cdots < \mu_N$ and $c_0 = 0 < c_1 < c_2 < \cdots < c_N$.

For any given Markovian control policy u, $X_2 = \{X_{2u}(t) : t \geq 0\}$ is a continuous-time Markov process with the state space Z. Similarly, we convert the Markov process X_2 into X_2' where the transition rate Λ is defined by $\lambda+\mu_N$. Without loss of generality, it is assumed that $\Lambda + \alpha = 1$. Let $f(x)$ be the minimal expected total discounted cost over an infinite horizon with the starting net inventory level x. Similarly to Section 2.1, $f(x)$ exists and satisfies the following optimality equation.

$$
\begin{aligned}
f(x) \quad &= \quad h[x]^+ + b[x]^- + \mu_N f(x) + \lambda f(x-1) \\
&\quad + \min_{k=0,1,2,\ldots,N} \{\mu_k [f(x+1) - f(x) + c_k]\}
\end{aligned}
$$

We are not able to characterize the optimal production decision completely for the basic model. Henceforth, we will consider a special case in which $(\mu_{k+1}c_{k+1} - \mu_k c_k)/(\mu_{k+1} - \mu_k)$ increases in k, that is, $(\mu_{k+1}c_{k+1} - \mu_k c_k)/(\mu_{k+1} - \mu_k) > (\mu_k c_k - \mu_{k-1}c_{k-1})/(\mu_k - \mu_{k-1}), k = 1, 2, \ldots, N-1$. Similarly to Subsection 2.1.1, the production rate μ_{k+1} can be viewed as a combination of μ_k and $\mu_{k+1} - \mu_k$ and then $(\mu_{k+1}c_{k+1} - \mu_k c_k)/(\mu_{k+1} - \mu_k)$ is the unit production cost for the additional production rate $\mu_{k+1} - \mu_k$. It is obvious that the assumption introduced here is intuitive. The following lemma shows the optimal production decision for the production system with N production rates.

Lemma 2.3 *The optimal control decision is*

1. *not to produce if* $f(x) - f(x+1) \le c_1$,
2. *to produce with the* k^{th} *rate if* $(\mu_k c_k - \mu_{k-1}c_{k-1})/(\mu_k - \mu_{k-1}) \le f(x) - f(x+1) \le (\mu_{k+1}c_{k+1} - \mu_k c_k)/(\mu_{k+1} - \mu_k), k = 1, 2, \ldots, N-1$, *and*
3. *to produce with the* N *th rate if* $f(x) - f(x+1) \ge (\mu_N c_N - \mu_{N-1}c_{N-1})/(\mu_N - \mu_{N-1})$.

Proof. The optimal production rate is the one which minimizes the right side of Equation 2.25. Then, it is optimal to employ the k^{th} production rate if

$$\mu_k[f(x+1) - f(x) + c_k] \le \mu_m[f(x+1) - f(x) + c_m], \text{ for all } m \ne k$$

which is equivalent to

$$f(x) - f(x+1) \ge (\mu_k c_k - \mu_m c_m)/(\mu_k - \mu_m), m = 0, 1, \ldots, k-1$$

$f(x) - f(x+1) \leq (\mu_m c_m - \mu_k c_k)/(\mu_m - \mu_k), m = k+1, k+2, \ldots, N$

Because

$$\frac{\mu_{m+1}c_{m+1} - \mu_m c_m}{\mu_{m+1} - \mu_m} > \frac{\mu_m c_m - \mu_{m-1}c_{m-1}}{\mu_m - \mu_{m-1}}, m = 1, 2, 3, \ldots, N-1$$

It is easy to show that

$$\frac{\mu_k c_k - \mu_{k-1}c_{k-1}}{\mu_k - \mu_{k-1}} > \frac{\mu_k c_k - \mu_m c_m}{\mu_k - \mu_m}, m = 0, 1, \ldots, k-2$$

and

$$\frac{\mu_{k+1}c_{k+1} - \mu_k c_k}{\mu_{k+1} - \mu_k} < \frac{\mu_m c_m - \mu_k c_k}{\mu_m - \mu_k}, m = k+2, \ldots, N$$

Thus, it is optimal to employ the k^{th} production rate if

$$\frac{\mu_k c_k - \mu_{k-1}c_{k-1}}{\mu_k - \mu_{k-1}} \leq f(x) - f(x+1) \leq \frac{\mu_{k+1}c_{k+1} - \mu_k c_k}{\mu_{k+1} - \mu_k}, k = 1, 2, \ldots, N-1$$

It can be checked that it is optimal not to produce if $f(x) - f(x+1) \leq (\mu_1 c_1 - \mu_0 c_0)/(\mu_1 - \mu_0) = c_1$. Similarly, it is optimal to produce with N th production rate if $f(x) - f(x+1) \geq (\mu_N c_N - \mu_{N-1}c_{N-1})/(\mu_N - \mu_{N-1})$. □

Based on Lemma 2.3, the following theorem can be easily obtained in a similar fashion to that of Theorem 2.1.

Theorem 2.2 *1. The minimal expected total discounted cost function $f(x)$ is convex with respect to the net inventory level x.*

2. Define

$S_1 = \min \{ x : f(x) - f(x+1) \leq c_1 \}$

$S_k = \min \{x : f(x) - f(x+1) \leq (\mu_k c_k - \mu_{k-1}c_{k-1})/(\mu_k - \mu_{k-1})\}, k = 2, \ldots, N$

There exists a stationary optimal policy such that it is optimal not to produce if the net inventory level is greater than or equal to S_1, to produce with k^{th} production rate if the net inventory level is below S_k and greater than or equal to S_{k+1}, $k = 1, 2, \ldots, N-1$, and to production with the N th production rate if the net inventory level is below S_N.

Theorem 2.2 shows that the optimal control policy is characterized by N critical inventory levels $S_1, S_2, \ldots, S_N$, denoted as $(S_1, S_2, \ldots, S_N)$ policy. This control policy is also stationary, i.e., all the critical inventory levels $S_1, S_2, \ldots, S_N$ do not change with time. It is easy to see that this control policy is a direct extension of the (S_1, S_2) control policy.

2.5. Conclusions

In this chapter, we first consider a make-to-stock production system with two production rates. With the assumptions of Poisson demand and exponential production time, it is found that the (S_1, S_2) policy is optimal for production control. Under this policy, it is optimal to stop production if the inventory level reaches S_1, to produce normally if the inventory level falls between S_1 and S_2, and to produce urgently if the inventory level drops below S_2. Later on, we consider a production system with N production rates for which the optimal control policy is shown to be the $(S_1, S_2, \ldots, S_N)$ policy. The numerical study shows that a significant cost saving can be achieved by employing an emergency production rate.

Chapter 3

A Make-to-Stock Production System with Two Production Rates, N Demand Classes and Lost Sales

3.1. The Stochastic Model and Optimal Control

In this chapter, we consider a single-item, make-to-stock production facility with two production rates: normal and emergency. Production times for the normal and emergency rate are independent and exponentially distributed with means $1/\mu_1$ and $1/\mu_2$, respectively. The unit production cost for the normal rate is c_1 and that for the emergency rate is c_2. Naturally, it is assumed that $0 < \mu_1 < \mu_2$ and $0 < c_1 < c_2$. Demands that can not be satisfied immediately are lost forever and lost-sale costs are incurred. There are N demand classes which are characterized by different lost-sale costs p_i with $p_1 > p_2 > \ldots > p_N$. Demand from class i follows an independent Poisson process with mean λ_i. We assume that $c_1 < p_N$, i.e., the unit normal production cost is less than the lowest lost-sale cost. This assumption is intuitive and demand from class N will never be satisfied otherwise.

At any arbitrary point of time, we have two types of operational decisions to make for this production system: production decision and inventory allocation decision. Production decision is to choose the optimal production rate given the on-hand inventory level. There are three choices available for production decision: i) not to produce, ii) to produce normally and iii) to produce urgently. Inventory allocation decision is to decide how to allocate limited inventory among different demand classes. Specifically, if there is on-hand inventory, we may choose either to satisfy an incoming demand of class i or to reject it. This is intuitive since demand classes have different lost-sale costs and thus different fulfillment priorities. To minimize the total cost, when the inventory level is low, a certain amount of on-hand inventory may be reserved for demand of classes with higher priorities by rejecting those with lower priorities. A control policy is to specify the production and inventory allocation decisions at any time given the on-hand inventory level. We develop an optimal control policy for the objective of minimizing the expected total discounted cost over an infinite horizon. This expected total discounted cost is computed by the following cost components: the inventory holding cost h per unit per unit time, the production cost of the normal rate c_1 per unit, the production cost of the emergency rate c_2 per unit, and the lost-sale cost p_i per lost demand from class i.

In the next subsection, the optimality equation is obtained which is satisfied by the minimal expected total discounted cost and the optimal control policy is identified

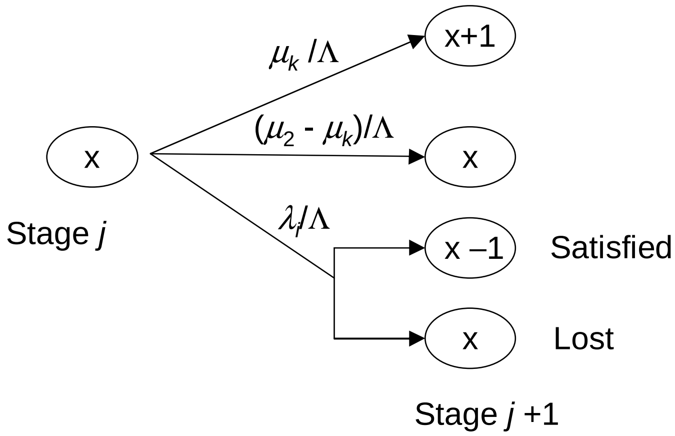

Figure 3.1: Transition process for the Markov process X'_3

by analyzing this optimality equation.

3.1.1. Dynamic Programming Formulation

Let $X_3(t)$ be the on-hand inventory level at time t. Given any Markovian control policy u, $X_3 = \{X_{3u}(t) : t \geq 0\}$ is a continuous-time Markov process with the state space Z^+, where Z^+ represents nonnegative integers. Because the sum of Poisson processes is a Poisson process as well, the aggregate demand from all demand classes follows a Poisson process with an aggregate mean $\lambda = \sum_{i=1}^{N} \lambda_i$. Similarly to Subsection 2.1.1, we convert the Markov process X_3 to X'_3 where the transition rate Λ is defined by $\lambda + \mu_2$. Figure 3.1 shows the transition process for the converted Markov process X'_3. Given that the current state is x and the production rate taken at the stage is $\mu_{k=0,1,2}$, the transition occurring at the next stage turns out to be no event at all with the probability of $(\mu_2 - \mu_k)/\Lambda$, to be a production completion with the probability of μ_k/Λ, and to be an arrival from class-i demand with the probability of λ_i/Λ, which is the product of λ/Λ and λ_i/λ.

Let α be the interest rate. Because we can always re-scale the time unit, without loss of generality, it is assumed that $\Lambda + \alpha = 1$. Now, we consider the first n stages of the infinite horizon problem by truncation. Denote $f_j^n(x)$ as, evaluated at the beginning of stage j with $1 \leq j \leq n$, the minimal expected total discounted cost in stages j through n given that the starting state is x. Let $f_{n+1}^n(x)$ be the terminal value function applied at the end of stage n if the ending state is x. Thus, $f_j^n(x)$ can be computed recursively as follows.

$$\begin{aligned} f_j^n(x) &= \min\left\{hx + \mu_k c_k + \mu_k f_{j+1}^n(x+1) + (\mu_2 - \mu_k) f_{j+1}^n(x)\right\} \\ &\quad + \sum_{i=1}^{N} \lambda_i \min\left\{f_{j+1}^n(x) + p_i, H_i f_{j+1}^n(x)\right\} \\ &= hx + \mu_2 f_{j+1}^n(x) + \sum_{i=1}^{N} \lambda_i \min\left\{f_{j+1}^n(x) + p_i, H_i f_{j+1}^n(x)\right\} \\ &\quad + \min_{k=0,1,2}\left\{\mu_k\left[f_{j+1}^n(x+1) - f_{j+1}^n(x) + c_k\right]\right\} \end{aligned} \tag{3.1}$$

where H_i is the operator defined by

$$H_i f_{j+1}^n(x) = \begin{cases} f_{j+1}^n(x) + p_i, & \text{if } x = 0 \\ f_{j+1}^n(x-1), & \text{Otherwise} \end{cases}$$

Let $f(x)$ be the minimal expected total discounted cost over an infinite horizon with the initial on-hand inventory level x. Based on Theorem 11.3 of Porteus [24], it follows that $f(x) = \lim_{n\to\infty} f_j^n(x)$ and $f(x)$ satisfies the following optimality equation.

$$\begin{aligned} f(x) &= hx + \mu_2 f(x) + \sum_{i=1}^{N} \lambda_i \min\{f(x) + p_i, H_i f(x)\} \\ &\quad + \min_{k=0,1,2}\{\mu_k[f(x+1) - f(x) + c_k]\} \end{aligned} \tag{3.2}$$

The decision process of this production system is as follows. Upon entering a new stage, production decision is made immediately based on the current on-hand inventory level. Since occurrence of transitions follows a Poisson process with mean Λ, only one event will take place at the end of this stage, i.e., production completion and demand arrival can not be happening simultaneously. Thus, we can consider separately production completion and demand arrival. If production completion occurs at the end of the stage, then the on-hand inventory level will definitely be increased by one. If demand arrival occurs instead, the inventory allocation decision is made and the on-hand inventory level changes accordingly. The optimal production and inventory allocation decision is to minimize the right side of Equation 3.3. The first minimization term corresponds to the inventory allocation decision. When there is no inventory held on hand, i.e., $x = 0$, any incoming demand has to be rejected and lost-sale cost is incurred. When there is on-hand inventory available, it is optimal to satisfy an incoming demand of class i if $f(x-1) \leq f(x)+p_i$, i.e., the incremental cost incurred $f(x-1) - f(x)$ is less than or equal to the corresponding lost-sale cost p_i, and reject it otherwise. The second minimization term corresponds to the production decision. It can be seen that Lemma 2.1 applies here, i.e., it is optimal not to produce if the cost saving $f(x) - f(x+1)$ is less than or equal to the unit production cost of the normal rate c_1, to produce normally if $f(x) - f(x+1)$ is greater than c_1 and less than or equal to $(\mu_2 c_2 - \mu_1 c_1)/(\mu_2 - \mu_1)$, and to produce urgently if $f(x) - f(x+1)$ is greater than $(\mu_2 c_2 - \mu_1 c_1)/(\mu_2 - \mu_1)$.

3.1.2. The Optimal Control Policy

Let V be the set of all the real-valued convex functions defined on Z^+ (the set of all non-negative integers) with the first difference bounded below by $-p_1$. Define H as the operator applied on $v \in V$ such that

$$\begin{aligned} Hv(x) &= hx + \mu_2 v(x) + \sum_{i=1}^{N} \lambda_i \min\left[v(x) + p_i, H_i v(x)\right] \\ &\quad + \min_{k=0,1,2} \left\{\mu_k \left[v(x+1) - v(x) + c_k\right]\right\} \end{aligned} \tag{3.3}$$

Lemma 3.1 shows that the operator H preserves the properties of the function v.

Lemma 3.1 *If $v \in V$, then $Hv \in V$.*

Proof. It is assumed that $v \in V$, then $v(x)$ is convex and its first difference is bounded below by $-p_1$. Define $F(x) = \mu_2 v(x) + \min_{k=0,1,2} \{\mu_k [v(x+1) - v(x) + c_k]\}$ and then $F(x+1) = \mu_2 v(x+1) + \min_{k=0,1,2} \{\mu_k [v(x+2) - v(x+1) + c_k]\}$. In Lemma 2.2, we have proved the convexity of $F(x)$. Now we need to develop the lower bound of the first difference of $F(x)$. Let $k^* = \arg\min_{k=0,1,2} \{\mu_k [v(x+2) - v(x+1) + c_k]\}$. Then,

$$\begin{aligned} & F(x+1) - F(x) \\ =\ & \mu_2 v(x+1) + \min_{k=0,1,2} \{\mu_k [v(x+2) - v(x+1) + c_k]\} \\ & - \mu_2 v(x) - \min_{k=0,1,2} \{\mu_k [v(x+1) - v(x) + c_k]\} \\ =\ & \mu_2 v(x+1) + \mu_{k^*} [v(x+2) - v(x+1) + c_{k^*}] \end{aligned}$$

$$
\begin{aligned}
& - \mu_2 v(x) - \min_{k=0,1,2} \{\mu_k [v(x+1) - v(x) + c_k]\} \\
\geq\ & \mu_2 v(x+1) + \mu_{k^*} [v(x+2) - v(x+1) + c_{k^*}] \\
& - \mu_2 v(x) - \mu_{k^*} [v(x+1) - v(x) + c_{k^*}] \\
=\ & \mu_2 [v(x+1) - v(x)] + \mu_{k^*} \{[v(x+2) - v(x+1)] - [v(x+1) - v(x)]\} \\
\geq\ & -\mu_2 p_1 + \mu_{k^*} \{[v(x+2) - v(x+1)] - [v(x+1) - v(x)]\} \\
\geq\ & -\mu_2 p_1
\end{aligned}
$$

The first inequality follows from the definition of k^*, the second from the bound of the first different of v and the last from the convexity of v.

Let $m_i(x) = \min [v(x) + p_i, H_i v(x)]$, $1 \leq i \leq N$. Ha [12] has proved that $m_i(x)$ is convex and its first difference is bounded below by $-p_1$. As $Hv(x)$ is just the sum of convex functions, it is also convex. Moreover,

$$
\begin{aligned}
& Hv(x+1) - Hv(x) \\
=\ & h + F(x+1) + \sum_{i=1}^{N} \lambda_i m_i(x+1) - F(x) - \sum_{i=1}^{N} \lambda_i m_i(x) \\
\geq\ & h + [F(x+1) - F(x)] + \sum_{i=1}^{N} \lambda_i [m_i(x+1) - m_i(x)] \\
\geq\ & h - \mu_2 p_1 - \sum_{i=1}^{N} \lambda_i p_1 \\
\geq\ & -p_1 \left(\mu_2 + \sum_{i=1}^{N} \lambda_i \right) \\
\geq\ & -p_1
\end{aligned}
$$

Hence, we get the results. □

Based on Lemma 3.1, we have the following theorem:

Theorem 3.1 *1. The minimal expected total discounted cost function $f(x)$ is convex with respect to the on-hand inventory level x and its first difference is* *bounded below by $-p_1$.*

2. *The (S_1, S_2) policy is optimal for production control where*

   $S_1 = \min\{\ x : f(x) - f(x+1) \leq c_1\ \}$

   $S_2 = \min\{\ x : f(x) - f(x+1) \leq (\mu_2 c_2 - \mu_1 c_1)/(\mu_2 - \mu_1)\ \}.$

   Specifically, it is optimal not to produce if the on-hand inventory level reaches S_1, to produce normally if the on-hand inventory level is below S_1 and at or above S_2, and to produce urgently if the on-hand inventory level is below S_2.

3. *The stock-reservation policy proposed by Ha [12] is optimal for inventory allocation where there exits rationing levels $R_1, R_2, \ldots, R_N$ defined by $R_i = \max\{\ x : f(x-1) - f(x) > p_i,\ i = 1, 2, \ldots, N\ \}$ such that it is optimal to satisfy an incoming demand of class i if the on-hand inventory level is above R_i, and reject it otherwise. Moreover, $S_1 \geq R_N \geq \cdots \geq R_1 = 0$.*

4. *There exits an optimal stationary policy.*

Proof. We prove this theorem based on Theorem 11.5 of Porteus [24]. Define the set of structured decision rules as all the decision rules with the form given by part 2 and part 3 while $S_1, S_2, R_1, \ldots, R_N$ can take any integers. Define the set of structured value functions as all the convex functions whose first difference is bounded below by $-p_1$. Essentially, the set of structured value functions is the set V. Because the limit of any convergent sequence of functions in V will be in V as well, the set V is complete. Moreover, from Lemma 3.1, the operator H preserves the structure of V. Therefore, the optimal return function f must be structured, i.e., it is convex and its first difference is bounded below by $-p_1$. From the optimality equation 3.3, it can be seen that the structured decision rule with $S_1, S_2, R_1, \ldots, R_N$ defined in the theorem is optimal for the one-stage minimization problem. Thus, the control policy developed in the theorem is optimal. Because the production system is stationary, i.e., the system equation, the cost per stage, the demand process, and the production process do not change from one stage to the next, the optimal control policy is stationary. □

The optimal policy shown in Theorem 3.1 is referred to as the $(S_1, S_2, R_1, \ldots, R_N)$ policy, which essentially is the combination of the (S_1, S_2) policy and the so-called stock-reservation policy also known as the $(R_1, R_2, \ldots, R_N)$ policy. The (S_1, S_2) policy controls the production process, where S_1 acts like a base-stock level and S_2 decides the switch between the normal rate and the emergency rate. The stock-reservation policy controls the inventory allocation, which is actually the critical level policy, the terminology normally used in the literature. When a demand of class i arrives, it is optimal to satisfy it if the on-hand inventory level is above R_i, i.e., the incremental cost of reducing the on-hand inventory by one is less than the lost-sale cost of demand class i, and reject it otherwise. Because the first difference of $f(x)$ is bounded below by $-p_1$, it is always optimal to satisfy the incoming demand of class 1 if the on-hand inventory is available.

3.2. Stationary Analysis of the Production System

In this section, the expected total cost per unit time is computed for the production system with two production rates and two demand classes under the optimal control policy of the form proposed in the previous section. For this production system, the optimal control policy can be specified completely by the critical inventory levels S_1, S_2 and R_2 since R_1 is always equal to zero. Similarly to Section 2.2, the production system can be considered as an $M/M/1/S$ queueing system with state-dependent arrival rates and service rates. In this queueing system, the on-hand inventory level is considered as the number of customers waiting for service. Production completion is represented as arrival to the queueing system and customer demand is modelled as service of the system. The service rate of the queueing system is equivalent to the customer demand arrival rate, which varies with the on-hand inventory level, i.e., the service rate is $\lambda_1 + \lambda_2$ if the on-hand inventory level is above R_2 and is λ_1 if the on-hand inventory level is at or below R_2. The arrival rate of the queueing system is equivalent to the production rate, which also varies with the on-hand inventory level, i.e., the arrival rate is 0 if the on-hand inventory level is at or above S_1, μ_1 if the on-hand inventory level is below S_1 but at or above S_2, and μ_2 if the on-hand inventory level drops below S_2.

To compute the expected total cost per unit time, the following performance measures are needed. Under the (S_1, S_2, R_2) policy, define

$I(S_1, S_2, R_2) = \sum_{n=0}^{S_1} n\pi(n)$ as the expected on-hand inventory level,

$P_1(S_1, S_2, R_2) = \sum_{n=S_2}^{S_1-1} \pi(n)$ as the probability of the normal rate employed,

$P_2(S_1, S_2, R_2) = \sum_{n=0}^{S_2-1} \pi(n)$ as the probability of the emergency rate employed,

$L_1(S_1, S_2, R_2) = \pi(0)$ as the probability of stock-outs of demand class 1, and

$L_2(S_1, S_2, R_2) = \sum_{n=0}^{R_2} \pi(n)$ as the probability of stock-outs of demand class 2.

Define $\pi(n)$ as the steady-state probability that the on-hand inventory level is n.

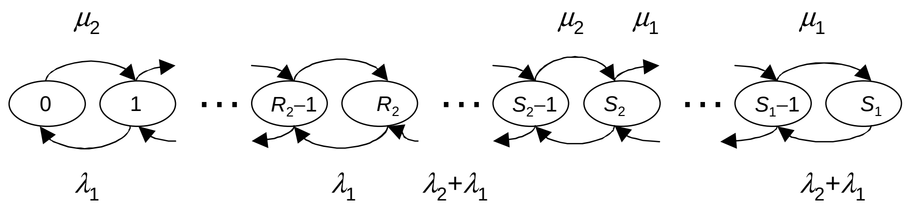

Figure 3.2: Rate diagram for the $M/M/1/S$ queueing system if $S_2 \geq R_2$

Let $\rho_1 = (\lambda_1 + \lambda_2)/\mu_1$, $\rho_2 = (\lambda_1 + \lambda_2)/\mu_2$, $\rho_{11} = \lambda_1/\mu_1$ and $\rho_{12} = \lambda_1/\mu_2$. To compute the above-defined performance measures, two cases are to be considered next.

1. For $S_2 \geq R_2$, Figure 3.2 shows the rate diagram of the $M/M/1/S$ queue. It can be obtained that

$$\pi_1(n) = \begin{cases} \rho_1^{S_1-n}\pi_1(S_1), & \text{for } S_2 \leq n < S_1 \\ \rho_2^{S_2-n}\rho_1^{S_1-S_2}\pi_1(S_1), & \text{for } R_2 \leq n < S_2 \\ \rho_{12}^{R_2-n}\rho_2^{S_2-R_2}\rho_1^{S_1-S_2}\pi_1(S_1), & \text{for } n < R_2 \end{cases} \tag{3.4}$$

where

$$\pi_1(S_1) = \left[\frac{1-\rho_1^{S_1-S_2}}{1-\rho_1} + \rho_1^{S_1-S_2}\frac{1-\rho_2^{S_2-R_2}}{1-\rho_2} + \rho_2^{S_2-R_2}\rho_1^{S_1-S_2}\frac{1-\rho_{12}^{R_2+1}}{1-\rho_{12}}\right]^{-1} \tag{3.5}$$

Now we compute the expected on hand inventory level $I(S_1, S_2, R_2)$. It is easy to see that $I(S_1, S_2, R_2) = \sum_{n=0}^{R_2} n\pi_1(n) + \sum_{n=R_2+1}^{S_2-1} n\pi_1(n) + \sum_{n=S_2}^{S_1} n\pi_1(n)$. Let $G_5 = \sum_{n=0}^{R_2} n\pi_1(n)$, $G_6 = \sum_{n=R_2+1}^{S_2-1} n\pi_1(n)$ and $G_7 = \sum_{n=S_2}^{S_1} n\pi_1(n)$. First, G_5 is computed, which is

$$\begin{aligned} G_5 &= \pi_1(1) + 2\pi_1(2) + \cdots + R_2\pi_1(R_2) \\ &= \rho_2^{S_2-R_2}\rho_1^{S_1-S_2}\pi_1(S_1)\left[\rho_{12}^{R_2-1} + 2\rho_{12}^{R_2-2} + \cdots + (R_2-1)\rho_{12} + R_2\right] \end{aligned}$$

Then,

$$\rho_{12}G_5 = \rho_2^{S_2-R_2}\rho_1^{S_1-S_2}\pi_1(S_1)\left[\rho_{12}^{R_2} + 2\rho_{12}^{R_2-1} + \cdots + (R_2-1)\rho_{12}^2 + R_2\rho_{12}\right]$$

Thus,

$$\begin{aligned} G_5 &= \rho_2^{S_2-R_2}\rho_1^{S_1-S_2}\pi_1(S_1)\left[R_2 - \left(\rho_{12}^{R_2} + \rho_{12}^{R_2-1} + \cdots + \rho_{12}^2 + \rho_{12}\right)\right]/(1-\rho_{12}) \\ &= \rho_2^{S_2-R_2}\rho_1^{S_1-S_2}\pi_1(S_1)\left[\frac{R_2}{1-\rho_{12}} - \frac{\rho_{12}(1-\rho_{12}^{R_2})}{(1-\rho_{12})^2}\right] \end{aligned}$$

Now we compute G_6, which is

$$\begin{aligned} G_6 &= (R_2+1)\pi_1(R_2+1) + R_2\pi_1(R_2+2) + \cdots + (S_2-1)\pi_1(S_2-1) \\ &= \rho_1^{S_1-S_2}\pi_1(S_1)\left[(R_2+1)\rho_2^{S_2-R_2-1} + (R_2+2)\rho_2^{S_2-R_2-2} + \cdots + (S_2-1)\rho_2\right] \end{aligned}$$

Then,

$$\rho_2 G_6 = \rho_1^{S_1-S_2}\pi_1(S_1)\left[(R_2+1)\rho_2^{S_2-R_2} + (R_2+2)\rho_2^{S_2-R_2-1} + \cdots + (S_2-1)\rho_2^2\right]$$

Thus,

$$\begin{aligned} G_6 &= \rho_1^{S_1-S_2}\pi_1(S_1)\left[S_2\rho_2 - R_2\rho_2^{S_2-R_2} - \left(\rho_2^{S_2-R_2} + \rho_2^{S_2-R_2-1} + \cdots + \rho_2\right)\right]/(1-\rho_2) \\ &= \rho_1^{S_1-S_2}\pi_1(S_1)\left[\frac{S_2\rho_2 - R_2\rho_2^{S_2-R_2}}{1-\rho_2} - \frac{\rho_2(1-\rho_2^{S_2-R_2})}{(1-\rho_2)^2}\right] \end{aligned}$$

Finally, we compute G_7, which is

$$\begin{aligned} G_7 &= S_2\pi_1(S_2) + (S_2+1)\pi_1(S_2+1) + \cdots + S_1\pi_1(S_1) \\ &= \pi_1(S_1)\left[S_2\rho_1^{S_1-S_2} + (S_2+1)\rho_1^{S_1-S_2-1} + \cdots + S_1\right] \end{aligned}$$

Then,

$$\rho_1 G_7 = \pi_1(S_1)\left[S_2\rho_1^{S_1-S_2+1} + (S_2+1)\rho_1^{S_1-S_2} + \cdots + S_1\rho_1\right]$$

Thus,

$$
\begin{aligned}
G_7 &= \pi_1(S_1)\left[S_1 - S_2\rho_1^{S_1-S_2+1} - \left(\rho_1^{S_1-S_2} + \rho_1^{S_1-S_2-1} + \cdots + \rho_1\right)\right]/(1-\rho_1) \\
&= \pi_1(S_1)\left[\frac{S_1 - S_2\rho_1^{S_1-S_2+1}}{1-\rho_1} - \frac{\rho_1(1-\rho_1^{S_1-S_2})}{(1-\rho_1)^2}\right]
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
I(S_1,S_2,R_2) &= \pi_1(S_1)\left[\frac{S_1 - S_2\rho_1^{S_1-S_2+1}}{1-\rho_1} - \frac{\rho_1\left(1-\rho_1^{S_1-S_2}\right)}{(1-\rho_1)^2}\right] \\
&\quad + \rho_1^{S_1-S_2}\pi_1(S_1)\left[\frac{S_2\rho_2 - R_2\rho_2^{S_2-R_2}}{1-\rho_2} - \frac{\rho_2\left(1-\rho_2^{S_2-R_2}\right)}{(1-\rho_2)^2}\right] \\
&\quad + \rho_2^{S_2-R_2}\rho_1^{S_1-S_2}\pi_1(S_1)\left[\frac{R_2}{1-\rho_{12}} - \frac{\rho_{12}\left(1-\rho_{12}^{R_2}\right)}{(1-\rho_{12})^2}\right] \qquad (3.6)
\end{aligned}
$$

Now we compute $P_1(S_1,S_2,R_2)$, $P_2(S_1,S_2,R_2)$, $L_1(S_1,S_2,R_2)$ and $L_2(S_1,S_2,R_2)$, respectively.

$$
\begin{aligned}
P_1(S_1,S_2,R_2) &= \pi_1(S_2) + \pi_1(S_2+1) + \cdots + \pi_1(S_1-1) \\
&= \rho_1^{S_1-S_2}\pi_1(S_1) + \rho_1^{S_1-S_2-1}\pi_1(S_1) + \cdots + \rho_1\pi_1(S_1) \\
&= \frac{\rho_1(1-\rho_1^{S_1-S_2})}{1-\rho_1}\pi_1(S_1) \qquad (3.7)
\end{aligned}
$$

$$
\begin{aligned}
P_2(S_1,S_2,R_2) &= 1 - P_1(S_1,S_2,R_2) - \pi_1(S_1) \\
&= 1 - \pi_1(S_1) - \frac{\rho_1(1-\rho_1^{S_1-S_2})}{1-\rho_1}\pi_1(S_1) \qquad (3.8)
\end{aligned}
$$

$$
L_1(S_1,S_2,R_2) = \pi(0) = \rho_{12}^{R_2}\rho_2^{S_2-R_2}\rho_1^{S_1-S_2}\pi_1(S_1) \qquad (3.9)
$$

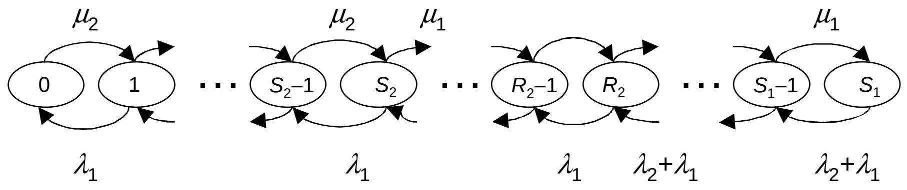

Figure 3.3: Rate diagram for the $M/M/1/S$ queueing system if $S_2 < R_2$

$$\begin{aligned} L_2(S_1,S_2,R_2) &= \pi_1(0)+\pi_1(1)+\cdots+\pi_1(R_2) \\ &= \rho_2^{S_2-R_2}\rho_1^{S_1-S_2}\pi_1(S_1)\left[\rho_{12}^{R_2}+\rho_{12}^{R_2-1}+\cdots+1\right] \\ &= \rho_2^{S_2-R_2}\rho_1^{S_1-S_2}\pi_1(S_1)\frac{1-\rho_{12}^{R_2+1}}{1-\rho_{12}} \end{aligned} \tag{3.10}$$

2. For $S_2 < R_2$, Figure 3.3 shows the transition process of the Markov process X_3'. It can be obtained that

$$\pi_2(n) = \begin{cases} \rho_1^{S_1-n}\pi_2(S_1), & \text{for } R_2 \le n < S_1 \\ \rho_{11}^{R_2-n}\rho_1^{S_1-R_2}\pi_2(S_1), & \text{for } S_2 \le n < R_2 \\ \rho_{12}^{S_2-n}\rho_{11}^{R_2-S_2}\rho_1^{S_1-R_2}\pi_2(S_1), & \text{for } n < S_2 \end{cases} \tag{3.11}$$

where

$$\pi_2(S_1) = \left[\frac{1-\rho_1^{S_1-R_2}}{1-\rho_1} + \rho_1^{S_1-R_2}\frac{1-\rho_{11}^{R_2-S_2}}{1-\rho_{11}} + \rho_{11}^{R_2-S_2}\rho_1^{S_1-R_2}\frac{1-\rho_{12}^{S_2+1}}{1-\rho_{12}}\right]^{-1} \tag{3.12}$$

Now we compute $I(S_1,S_2,R_2)$. It can be seen that $I(S_1,S_2,R_2) = \sum_{n=0}^{S_2-1} n\pi(n) + \sum_{n=S_2}^{R_2} n\pi(n) + \sum_{n=R_2+1}^{S_1} n\pi(n)$. Let $G_8 = \sum_{n=0}^{S_2-1} n\pi(n)$, $G_9 = \sum_{n=S_2}^{R_2} n\pi(n)$ and $G_{10} = \sum_{n=R_2+1}^{S_1} n\pi(n)$. First, we compute G_8, which is

$$\begin{aligned} G_8 &= \pi_2(1)+2\pi_2(2)+\cdots+(S_2-1)\pi_2(S_2-1) \\ &= \rho_{11}^{R_2-S_2}\rho_1^{S_1-R_2}\pi_2(S_1)\left[\rho_{12}^{S_2-1}+2\rho_{12}^{S_2-2}+\cdots+(S_2-1)\rho_{12}\right] \end{aligned}$$

Then,

$$\rho_{12}G_8 = \rho_{11}^{R_2-S_2}\rho_1^{S_1-R_2}\pi_2(S_1)\left[\rho_{12}^{S_2} + 2\rho_{12}^{S_2-1} + \cdots + (S_2-1)\rho_{12}^2\right]$$

Thus,

$$\begin{aligned} G_8 &= \rho_{11}^{R_2-S_2}\rho_1^{S_1-R_2}\pi_2(S_1)\left[S_2\rho_{12} - \rho_{12} - \cdots - \rho_{12}^{S_2-1} - \rho_{12}^{S_2}\right] / [1-\rho_{12}] \\ &= \rho_{11}^{R_2-S_2}\rho_1^{S_1-R_2}\pi_2(S_1)\left[\frac{S_2\rho_{12}}{1-\rho_{12}} - \frac{\rho_{12}\left(1-\rho_{12}^{S_2}\right)}{(1-\rho_{12})^2}\right] \end{aligned}$$

Now we compute G_9, which is

$$\begin{aligned} G_9 &= S_2\pi_2(S_2) + (S_2+1)\pi_2(S_2) + \cdots + R_2\pi_2(R_2) \\ &= \rho_1^{S_1-R_2}\pi_2(S_1)\left[S_2\rho_{11}^{R_2-S_2} + (S_2+1)\rho_{11}^{R_2-S_2-1} + \cdots + R_2\right] \end{aligned}$$

Then,

$$\rho_{11}G_9 = \rho_1^{S_1-R_2}\pi_2(S_1)\left[S_2\rho_{11}^{R_2-S_2+1} + (S_2+1)\rho_{11}^{R_2-S_2} + \cdots + R_2\rho_{11}\right]$$

Thus,

$$\begin{aligned} G_9 &= \rho_1^{S_1-R_2}\pi_2(S_1)\left[R_2 - S_2\rho_{11}^{R_2-S_2+1} - \rho_{11} - \cdots - \rho_{11}^{R_2-S_2}\right] / (1-\rho_{11}) \\ &= \rho_1^{S_1-R_2}\pi_2(S_1)\left[\frac{R_2 - S_2\rho_{11}^{R_2-S_2+1}}{1-\rho_{11}} - \frac{\rho_{11}(1-\rho_{11}^{R_2-S_2})}{(1-\rho_{11})^2}\right] \end{aligned}$$

Finally, we compute G_{10}, which is

$$\begin{aligned} G_{10} &= (R_2+1)\pi(R_2+1) + (R_2+2)\pi(R_2+2) + \cdots + S_1\pi(S_1) \\ &= \pi_2(S_1)\left[(R_2+1)\rho_1^{S_1-R_2-1} + (R_2+2)\rho_1^{S_1-R_2-2} + \cdots + S_1\right] \end{aligned}$$

Then,

$$\rho_1 G_{10} = \pi_2(S_1)\left[(R_2+1)\rho_1^{S_1-R_2} + (R_2+2)\rho_1^{S_1-R_2-1} + \cdots + S_1\rho_1\right]$$

Thus,

$$
\begin{aligned}
G_{10} &= \pi_2(S_1)\left[S_1 - R_2\rho_1^{S_1-R_2} - \rho_1 - \cdots - \rho_1^{S_1-R_2}\right]/(1-\rho_1) \\
&= \pi_2(S_1)\left[\frac{S_1 - R_2\rho_1^{S_1-R_2}}{1-\rho_1} - \frac{\rho_1(1-\rho_1^{S_1-R_2})}{(1-\rho_1)^2}\right]
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
I(S_1,S_2,R_2) &= \rho_{11}^{R_2-S_2}\rho_1^{S_1-R_2}\pi_2(S_1)\left[\frac{S_2\rho_{12}}{1-\rho_{12}} - \frac{\rho_{12}\left(1-\rho_{12}^{S_2}\right)}{\left(1-\rho_{12}\right)^2}\right] \\
&\quad + \rho_1^{S_1-R_2}\pi_2(S_1)\left[\frac{R_2 - S_2\rho_{11}^{R_2-S_2+1}}{1-\rho_{11}} - \frac{\rho_{11}\left(1-\rho_{11}^{R_2-S_2}\right)}{\left(1-\rho_{11}\right)^2}\right] \\
&\quad + \pi_2(S_1)\left[\frac{S_1 - R_2\rho_1^{S_1-R_2}}{1-\rho_1} - \frac{\rho_1\left(1-\rho_1^{S_1-R_2}\right)}{\left(1-\rho_1\right)^2}\right] \qquad (3.13)
\end{aligned}
$$

Now, we compute $P_1(S_1,S_2,R_2)$, $P_2(S_1,S_2,R_2)$, $L_1(S_1,S_2,R_2)$ and $L_2(S_1,S_2,R_2)$, respectively.

$$
\begin{aligned}
P_1(S_1,S_2,R_2) &= 1 - \pi_2(S_1) - P_2(S_1,S_2,R_2) \\
&= 1 - \pi_2(S_1) - \rho_{11}^{R_2-S_2}\rho_1^{S_1-R_2}\pi_2(S_1)\frac{\rho_{12}(1-\rho_{12}^{S_2})}{1-\rho_{12}} \qquad (3.14)
\end{aligned}
$$

$$
\begin{aligned}
P_2(S_1,S_2,R_2) &= \pi_2(0) + \pi(1) + \cdots + \pi(S_2-1) \\
&= \rho_{11}^{R_2-S_2}\rho_1^{S_1-R_2}\pi_2(S_1)\left[\rho_{12}^{S_2} + \cdots + \rho_{12}\right] \\
&= \rho_{11}^{R_2-S_2}\rho_1^{S_1-R_2}\pi_2(S_1)\frac{\rho_{12}(1-\rho_{12}^{S_2})}{1-\rho_{12}} \qquad (3.15)
\end{aligned}
$$

$$
L_1(S_1,S_2,R_2) = \rho_{12}^{S_2}\rho_{11}^{R_2-S_2}\rho_1^{S_1-R_2}\pi_2(S_1) \qquad (3.16)
$$

$$
\begin{aligned}
L_2(S_1,S_2,R_2) &= \pi_2(0)+\pi_2(1)+\cdots+\pi_2(S_2-1)+\pi_2(S_2)+\cdots+\pi_2(R_2)\\
&= \rho_{11}^{R_2-S_2}\rho_1^{S_1-R_2}\pi_2(S_1)\left[\rho_{12}^{S_2}+\rho_{12}^{S_2-1}+\cdots+\rho_{12}\right]\\
&\quad+\rho_1^{S_1-R_2}\pi_2(S_1)\left[\rho_{11}^{R_2-S_2}+\cdots+1\right]\\
&= \rho_1^{S_1-R_2}\pi_2(S_1)\left[\rho_{11}^{R_2-S_2}\frac{\rho_{12}(1-\rho_{12}^{S_2})}{1-\rho_{12}}+\frac{1-\rho_{11}^{R_2-S_2+1}}{1-\rho_{11}}\right] \qquad (3.17)
\end{aligned}
$$

It can be seen that $\mu_1P_1(S_1,S_2,R_2)$ and $\mu_2P_2(S_1,S_2,R_2)$ are the numbers per unit time of the normal production and the emergency production, respectively, and $\lambda_1L_1(S_1,S_2,R_2)$ and $\lambda_2L_2(S_1,S_2,R_2)$ are the numbers per unit time of stock-outs of demand class 1 and class 2, respectively. Therefore, the expected total cost per unit time is

$$
\begin{aligned}
C(S_1,S_2,R_2) &= hI(S_1,S_2,R_2)+c_1\mu_1P_1(S_1,S_2,R_2)+c_2\mu_2P_2(S_1,S_2,R_2)\\
&\quad+p_1\lambda_1L_1(S_1,S_2,R_2)+p_2\lambda_2L_2(S_1,S_2,R_2) \qquad (3.18)
\end{aligned}
$$

3.3. Numerical Study

In this section, we investigate the benefit of the production system with two production rates and two demand classes over the one with a single production rate and two demand classes. We set the normal rate of the former production system equal to the production rate of the latter one. Thus, the benefit can be viewed as a cost saving of providing with an emergency production rate the production system with a single production rate and two demand classes. The cost formula for the production system with two production rates and two demand classes has been obtained in the previous section. For the production system with a single rate and two demand classes, (S, R) policy proposed by Ha [12] is optimal. Under (S, R) policy, define

$I(S, R)$ as the expected on-hand inventory level,

$P(S, R)$ as the probability of production,

$L_1(S, R)$ as the probability of stock-outs of demand class 1, and

$L_2(S, R)$ as the probability of stock-outs of demand class 2

Thus, the expected total cost per unit time is given by

$$C(S, R) = hI(S, R) + c_1\mu_1 P(S, R) + p_1\lambda_1 L_1(S, R) + p_2\lambda_2 L_2(S, R) \tag{3.19}$$

where

$$\begin{aligned} I(S, R) \quad = \quad & S - \pi_3(S)\frac{\rho_1}{(1-\rho_1)^2}\left[1 - \rho_1^{S-R} - (1-\rho_1)(S-R)\rho_1^{S-R-1}\right] \\ & - \pi_3(S)\frac{\rho_{11}}{(1-\rho_{11})^2}\left(\frac{\rho_1}{\rho_{11}}\right)^{S-R}\left(\rho_{11}^{S-R} - \rho_{11}^{S+1}\right) \\ & - \pi_3(S)\frac{\rho_{11}^{S-R}}{1-\rho_{11}}(\frac{\rho_1}{\rho_{11}})^{S-R}\left[S - R - (S+1)\rho_{11}^{R+1}\right] \end{aligned} \tag{3.20}$$

$$P(S, R) = 1 - \pi_3(S) \tag{3.21}$$

$$L_1(S, R) = \rho_1^{S-R}\rho_{11}^R\pi_3(S) \tag{3.22}$$

$$L_2(S, R) = \frac{\left(1 - \rho_{11}^{R+1}\right)\rho_1^{S-R}}{1 - \rho_{11}}\pi_3(S) \tag{3.23}$$

where

$$\pi_3(S) = \frac{(1-\rho_1)(1-\rho_{11})}{(1-\rho_{11})\left(1-\rho_1^{S-R}\right) + (1-\rho_1)\rho_1^{S-R}\left(1-\rho_{11}^{R+1}\right)} \tag{3.24}$$

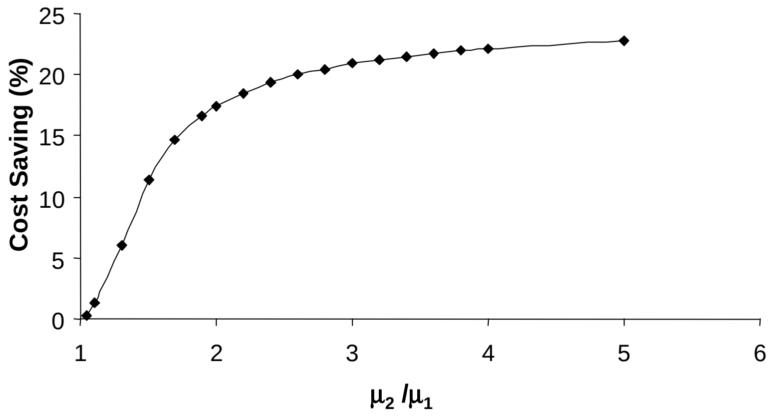

Figure 3.4: Cost saving versus μ_2/μ_1

For any given set of parameters, let S_1^*, S_2^* and R_2^* be the optimal critical inventory levels of the production system with two rates and two demand classes, and S^* and R^* be the optimal critical inventory levels of the production system with a single rate and two demand classes. Define the cost saving, CS, as the following percentage:

$$CS = \frac{C(S^*, R^*) - C(S_1^*, S_2^*, R_2^*)}{C(S^*, R^*)} \times 100\% \tag{3.25}$$

For each problem instance, we utilize the solver function in Microsoft Excel to search for the optimal solutions (S_1^*, S_2^*, R_2^*) and (S^*, R^*). Initially, we set that $c_1 = 1$, $\mu_1 = 1$, $p_2 = 2$, $\mu_2/\mu_1 = 1.5$, $\rho_1 = 1.4$, $\lambda_2/\lambda_1 = 1/1.8$, $h = 0.01$, $c_2/c_1 = 1.2$ and $p_1/p_2 = 5$. Based on the initial settings, we seek to find how the parameters μ_2/μ_1, ρ_1, λ_2/λ_1, h, c_2/c_1 and p_1/p_2 affect the cost saving achieved.

Figure 3.4 shows that the cost saving increases as the production rates ratio μ_2/μ_1 increases. This result is quite intuitive since the purpose of employing the emergency production rate is to reduce the probability of stock-outs and then the penalty cost rendered. The more capacity the emergency production rate can provide, the more significant the probability of stock-outs can be reduced and then the larger cost saving

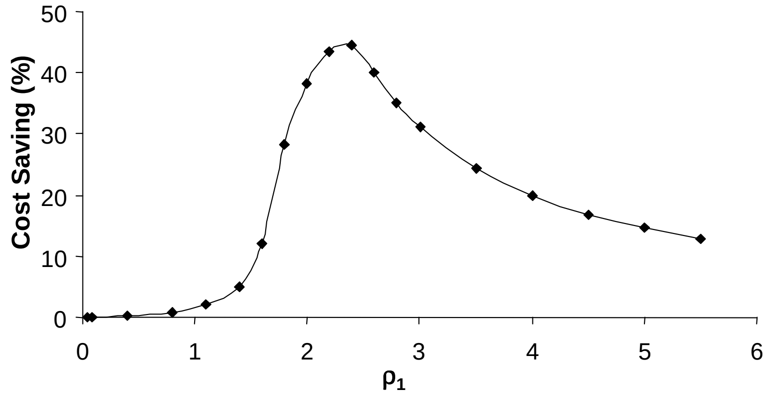

Figure 3.5: Cost saving versus ρ_1

can be achieved. In addition, it is found that once the ratio μ_2/μ_1 is larger enough beyond a certain threshold, there is only a small improvement in the cost saving as we continue to increase the emergency production rate. This threshold is useful for selecting an appropriate emergency production rate in the real life.

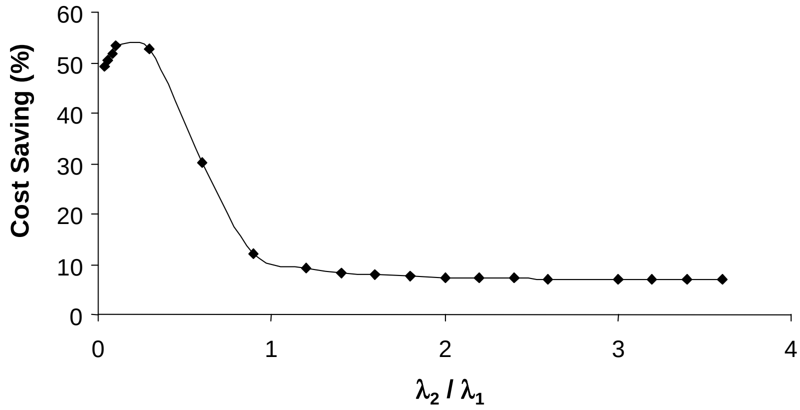

Figure 3.6: Cost saving versus λ_2/λ_1

Figure 3.5 shows that the cost saving due to the emergency production rate has a nonmonotone relationship with the traffic intensity ρ_1. When the normal rate has an excessive capacity to meet demand, i.e., ρ_1 is small, the emergency rate is seldom used and thus do not provide significant benefit. When capacity provided by the normal rate becomes small relative to demand, i.e., ρ_1 is large, stockouts becomes more frequently. The emergency production rate can reduce stockouts significantly. Thus, the cost saving achieved becomes significant. When the traffic intensity ρ_1 reaches a certain value, there is a maximum cost saving achieved. As the traffic intensity ρ_1 continues to drop, even capacity provided by the emergency rate is small relative to demand. Stockouts reduced by the emergency rate is less significant relative to the large stockouts produced. Thus, the benefit incurred becomes less significant.

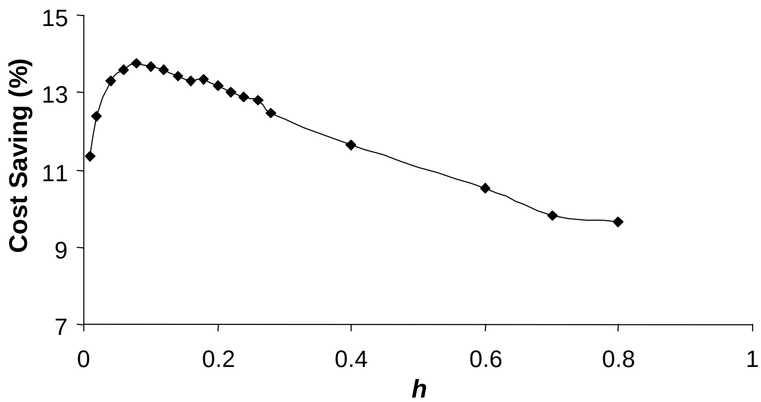

Figure 3.7: Cost saving versus h

Figure 3.6 shows that the demand rates ratio λ_2/λ_1 also has a nonmonotone relationship with the cost saving achieved. When the ratio λ_2/λ_1 is large, demand of class 1 is rare relative to that of class 2. Although stockouts of demand class 2 can be reduced significantly by employing the emergency rate, the cost saving is not so large due to the lower lost-sale cost of demand class 2. As the ratio λ_2/λ_1 decreases, demand of class 1 becomes more and more. Then, the emergency rate can reduce stockouts of both class 1 and class 2. Thus, the cost saving achieved becomes larger. However, as the ratio λ_2/λ_1 continues to decrease, demand of class 2 is dominated by that of class 1. Thus, the cost saving achieved decreases slightly.

Again, the cost saving achieved has a nonmonotone relationship with the inventory

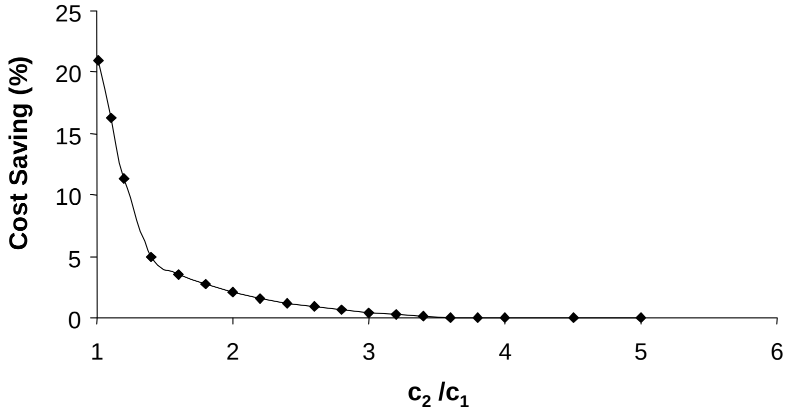

Figure 3.8: Cost saving versus c_2/c_1

hold cost rate h, as shown in Figure 3.7. When the rate h is large, the inventory holding cost plays an important role in the total cost. Although the emergency production rate provided can reduce stockouts, the inventory level is already low and can not be reduced substantially. Thus, the cost saving achieved is less significant. As the rate h decreases, the emergency production rate can reduce both stockouts and the inventory level substantially. Thus, the cost saving achieved becomes more significant. However, if the rate h continues to increase, both production systems can hold a larger number of inventory and stockouts become less. Then, the cost saving begins to decrease slightly.

Figure 3.8 shows that the cost saving decreases as the unit production costs ratio c_2/c_1 increases. When the ratio c_2/c_1 is small, the emergency production rate can be employed as frequently as possible to reduce stockouts without incurring a larger extra production cost. Thus, the cost saving achieved is significant. However, as the ratio c_2/c_1 increases, stockouts are reduced by the emergency rate at the expense of a higher production cost. Then, the cost saving achieved becomes less significant.

Figure 3.9 shows that the cost saving increases as the lost sale costs ratio p_1/p_2

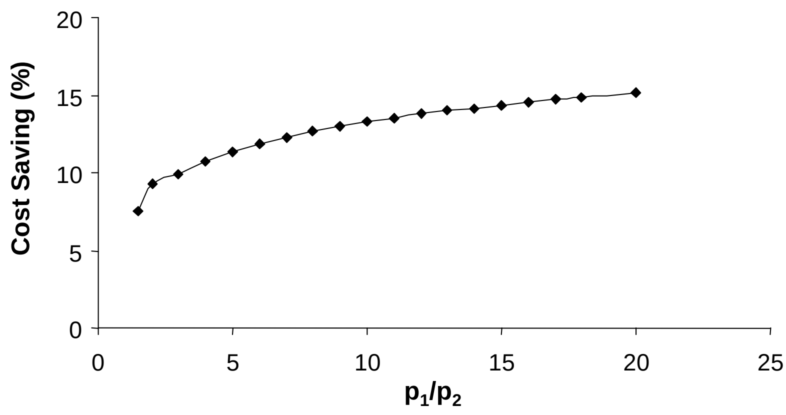

Figure 3.9: Cost saving versus p_1/p_2

increases. When the ratio p_1/p_2 is large, the emergency production rate is employed mainly to reduce stockouts of demand class 1 while that of demand class 2 remain non-increased. The cost saving achieved is significant due to the larger lost sale cost of class-1 demand. As the ratio p_1/p_2 decreases, i.e., the lost sale cost of demand class 1 decreases, it is intuitive to see that the cost saving achieved becomes less significant. However, even if the ratio p_1/p_2 is very small, we still have certain amount of benefits achieved.

To summarize, we find that the emergency production rate can produce remarkable benefits in most cases studied. The magnitude of the cost saving is affected significantly by the parameters μ_2/μ_1, ρ_1, λ_2/λ_1, h, c_2/c_1 and p_1/p_2. Specifically, larger values of μ_2/μ_1 and p_1/p_2 and small values of c_2/c_1 and h can produce a larger cost saving. In addition, ρ_1 and λ_2/λ_1 affect the achieved cost saving non-monotonically.

3.4. Conclusions

In this chapter, a make-to-stock production system is considered with two production rates, N demand classes and lost sales. It is found that the optimal control policy is the $(S_1, S_2, R_1, \ldots, R_N)$ policy, which is a combination of the (S_1, S_2) policy and the so-called stock reservation policy. The (S_1, S_2) policy is optimal for production control while the stock-reservation policy is used to control inventory allocation among N demand classes. Demand of class i is satisfied when the inventory level is above R_i and rejected otherwise. The numerical study shows that a significant cost saving can be achieved by employing an emergency production rate.

Chapter 4

A Make-to-Stock Production System with Two Production Rates, Two Demand Classes and Backorders

4.1. The Stochastic Model and Optimal Control

In this chapter, we consider a make-to-stock production system similar to the one studied in chapter 3. The difference is that the production system considered here has two demand classes only and demand that can not be satisfied immediately is backordered. The two demand classes, referred to as class 1 and class 2, incur different backordering cost b_1 and b_2, respectively and demands of the two classes arrive following independent Poisson processes with mean rates λ_1 and λ_2, respectively.

In this production system, we have three types of operational decisions to make.

First, there are three production modes to choose: i) not to produce at all, ii) to produce normally and iii) to produce urgently. Second, when a production completes, we first use this product to satisfy class-1 backorder if it exits. If there is no class-1 backorder, we have two choices for using the product: i) to increase the on-hand inventory level and ii) to satisfy class-2 backorder if available. Finally, we consider the incoming demand. If there is no on-hand inventory at all, demand from any class will be backordered. If on-hand inventory is available and class-1 demand arrives, we should satisfy it immediately. When class-2 demand arrives, we may either satisfy or backorder it even if there is on-hand inventory in order to minimize the total cost. By backordering the incoming class-2 demand when the on-hand inventory level is low, some inventory can be reserved in anticipation of the future class-1 demand. The optimal control policy must provide these operational decisions for the objective of minimizing the expected total discounted cost. In this chapter, we develop a stochastic model for this production system and the optimal control policy is identified. The expected total discounted cost includes the following cost components: the inventory holding cost h per unit per unit time, the additional production cost incurred by the emergency rate $c = c_2 - c_1$, the backordering cost b_1 per class-1 backorder per unit time, and the backordering cost b_2 per class-2 backorder per unit time. To simplify the problem, it is assumed that $c_1 = 0$.

4.1.1. Dynamic Programming Formulation

Define the following variables $X(t)$ and $Y(t)$:

$X(t) \geq 0$: the on-hand inventory level at time t,

$X(t) < 0$: the number of class-1 backorders at time t, and

$Y(t)$: the number of class-2 backorders at time t, $Y(t) \geq 0$

For any given Markovian control policy u, $X_4 = \{X_u(t), Y_u(t) : t \geq 0\}$ is a continuous-time Markov process with the state space $Z \times Z^+$. Similarly to Subsection 2.1.1, we convert the Markov process X_4 to X_4' where the transition rate Λ is defined as $\lambda_1 + \lambda_2 + \mu_2$. After the conversion, the mean time between successive transitions is constant and independent of system states and control policies employed. Because the underlying stochastic process remains unchanged, the Markov process X_4' has the same optimal control policy and then the same optimal return function to those of X_4. Henceforth, we analyze X_4' to characterize the optimal control policy. Figure 4.1 shows the transition process for the Markov process X_4'. Let x and y be the particular values of $X(t)$ and $Y(t)$, respectively. Given the current state (x, y) and the production rate employed at the stage μ_k, $k = 0, 1, 2$, a transition taking place at the end of the stage is a production completion with the probability of μ_k/Λ, an arrival of class-1 demand with the probability of λ_1/Λ, an arrival of class-2 demand with the probability of λ_2/Λ, and finally to be no event at all with the probability of $(\mu_2 - \mu_k)/\Lambda$. All the decisions regarding inventory allocation are made just before the end of the stage and the system state changes accordingly.

Define $f(x, y)$ as the minimal expected total discounted cost over an infinite horizon with the starting state (x, y). Without loss of generality, it is assumed that $\Lambda + \alpha = 1$. Similarly to Subsection 2.1.1, the optimal cost function $f(x, y)$ must satisfy the following optimality equation:

$$f(x, y) = c(x, y) + H_1 f(x, y) + \lambda_1 f(x - 1, y) + \lambda_2 H_2 f(x, y) \tag{4.1}$$

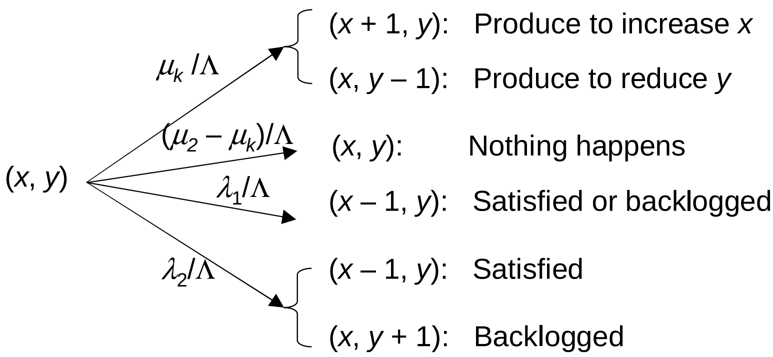

Figure 4.1: Transition process for the Markov process X_4'

where H_1 and H_2 are the operators applied on $f(x,y)$ and $c(x,y)$ is the expected discounted inventory holding and backordering cost for this stage.

$$c(x,y) = h[x]^+ + b_1[x]^- + b_2 y \tag{4.2}$$

$$H_1 f(x,y) = \mu_2 f(x,y) + \begin{cases} \min\left\{ \begin{array}{l} \mu_1 [f(x+1,y) - f(x,y)] \\ \mu_2 [f(x+1,y) - f(x,y) + c] \\ \mu_1 [f(x,y-1) - f(x,y)] \\ \mu_2 [f(x,y-1) - f(x,y) + c] \\ 0 \end{array} \right\}, & \text{if } y > 0 \\ \min\left\{ \begin{array}{l} \mu_1 [f(x+1,y) - f(x,y)] \\ \mu_2 [f(x+1,y) - f(x,y) + c] \\ 0 \end{array} \right\}, & \text{if } y = 0 \end{cases} \tag{4.3}$$

$$H_2 f(x,y) = \begin{cases} \min\{f(x-1,y), f(x,y+1)\}, & \text{if } x > 0 \\ f(x,y+1), & \text{if } x \le 0 \end{cases} \tag{4.4}$$

In $H_1 f(x,y)$, $f(x+1,y) - f(x,y)$ (resp. $f(x+1,y) - f(x,y) + c$) is the incremental cost incurred when the normal rate (resp. the emergency rate) is employed to increase

x by one. Similarly, $f(x, y-1) - f(x, y)$ (resp. $f(x, y-1) - f(x, y) + c$) is the incremental cost incurred when the normal rate (resp. the emergency rate) is employed to reduce y. The optimal production decision is to minimize the incremental cost at any time. In $H_2 f(x, y)$, $f(x-1, y)$ and $f(x, y+1)$ corresponds to the decisions of satisfying and backordering an incoming class-2 demand, respectively. The optimal allocation decision is to choose the minimum one between $f(x-1, y)$ and $f(x, y+1)$ at any time.

4.1.2. The Optimal Control Policy

To characterize the optimal control policy, we need to prove that the structural properties of the optimal return function $f(x, y)$ is preserved in the optimality equation 4.1. However, with the boundary $y = 0$ on the state space $Z \times Z^+$, it is difficult to make such a proof. For simplification, we follow an approach used by Ha [14] to extend $f(x, y)$ into an unconstrained function $f'(x, y)$ which is defined on the state space $Z \times Z$ and satisfies the following Equation 4.5.

$$f'(x, y) = c'(x, y) + H_1' f'(x, y) + \lambda_1 f'(x-1, y) + \lambda_2 H_2 f'(x, y) \tag{4.5}$$

where

$$c'(x, y) = \begin{cases} h[x]^+ + b_1[x]^- + b_2 y, & \text{if } y \geq 0 \\ +\infty, & \text{if } y < 0 \end{cases} \tag{4.6}$$

$$H_1'f'(x,y) = \mu_2 f'(x,y) + \min\left\{\begin{array}{l}\mu_1\left[f'(x+1,y)-f'(x,y)\right]\\ \mu_2\left[f'(x+1,y)-f'(x,y)+c\right]\\ \mu_1\left[f'(x,y-1)-f'(x,y)\right]\\ \mu_2\left[f'(x,y-1)-f'(x,y)+c\right]\\ 0\end{array}\right\} \quad (4.7)$$

In Equation 4.5, $c'(x,y)$ defined on the state space $Z \times Z$ is developed from $c(x,y)$ by imposing an infinite penalty cost on infeasible states $y < 0$. In addition, H_1 is modified into H_1' to make possible the transitions into the infeasible region $y < 0$. Essentially, $f'(x,y) = f(x,y)$ for $y \geq 0$ and $f'(x,y) = +\infty$ for $y < 0$. It is easy to see that an optimal control policy for the unconstrained problem will never allow the transitions into the infeasible region $y < 0$, and thus provides the same optimal control policy and the same optimal return function to the original, constraint problem.

Define V as the set of all the real-valued functions defined on $Z \times Z$ such that any $v \in V$ must satisfy the following four properties.

1. For $x < 0$,

$$v(x+1,y) \leq v(x,y-1) \text{ and } v(x+1,y) \leq v(x,y) \quad (4.8)$$

2. For $y > 0$,

$$v(x,y-1) \leq v(x,y) \quad (4.9)$$

3. Convexity

$$v(x+2,y) - v(x+1,y) \geq v(x+1,y) - v(x,y) \quad (4.10)$$

$$v(x, y+2) - v(x, y+1) \geq v(x, y+1) - v(x, y) \tag{4.11}$$

4. Submodularity/supermodularity

$$v(x+1, y+1) - v(x, y+1) \leq v(x+1, y) - v(x, y) \tag{4.12}$$

$$v(x+2, y) - v(x+1, y-1) \geq v(x+1, y) - v(x, y-1) \tag{4.13}$$

$$v(x+1, y+1) - v(x, y) \geq v(x+1, y) - v(x, y-1) \tag{4.14}$$

Let H be the operator applied on any function $v \in V$ such that

$$Hv(x, y) = c'(x, y) + H_1'v(x, y) + \lambda_1 v(x-1, y) + \lambda_2 H_2 v(x, y) \tag{4.15}$$

Lemmas 4.1– 4.3 show that the structure of the functions in V is preserved by H_1', H_2 and $c'(x, y)$.

Lemma 4.1 *If $v \in V$, then $H_1'v \in V$.*

Proof. First, we prove $H_1'v$ satisfying Equation 4.8. For any (x, y) with $x < 0$, we have $v(x+1, y) \leq v(x, y-1)$ and $v(x+1, y) \leq v(x, y)$. Therefore, we can get that

$$H_1'v(x, y) = \mu_2 v(x, y) + \min \left\{ \begin{array}{l} \mu_1 \left[v(x+1, y) - v(x, y) \right] \\ \mu_2 \left[v(x+1, y) - v(x, y) + c \right] \end{array} \right\}$$

$$H_1'v(x+1,y) = \mu_2 v(x+1,y) + \min \left\{ \begin{array}{l} \mu_1 \left[v(x+2,y) - v(x+1,y)\right] \\ \mu_2 \left[v(x+2,y) - v(x+1,y) + c\right] \\ \mu_1 \left[v(x+1,y-1) - v(x+1,y)\right] \\ \mu_2 \left[v(x+1,y-1) - v(x+1,y) + c\right] \\ 0 \end{array} \right\}$$

1. If $0 < v(x,y) - v(x+1,y) \leq \mu_2 c/(\mu_2 - \mu_1)$, then

$$\begin{aligned} & H_1'v(x+1,y) - H_1'v(x,y) \\ \leq\ & \mu_2 v(x+1,y) - \mu_2 v(x,y) - \mu_1 \left[v(x+1,y) - v(x,y)\right] \\ =\ & (\mu_2 - \mu_1)[v(x+1,y) - v(x,y)] \leq 0 \end{aligned}$$

2. If $v(x,y) - v(x+1,y) \geq \mu_2 c/(\mu_2 - \mu_1)$, then

$$\begin{aligned} & H_1'v(x+1,y) - H_1'v(x,y) \\ \leq\ & \mu_2 v(x+1,y) - \mu_2 v(x,y) - \mu_2 \left[v(x+1,y) - v(x,y) + c\right] \\ =\ & -\mu_2 c < 0 \end{aligned}$$

Thus, $H_1'v(x+1,y) \leq H_1'v(x,y)$.

For any (x,y) with $x < 0$, we have $v(x+1,y-1) \leq v(x,y-2)$ and $v(x+1,y-1) \leq v(x,y-1)$. Therefore, we can get that

$$H_1'v(x,y-1) = \mu_2 v(x,y-1) + \min \left\{ \begin{array}{l} \mu_1 \left[v(x+1,y-1) - v(x,y-1)\right] \\ \mu_2 \left[v(x+1,y-1) - v(x,y-1) + c\right] \end{array} \right\}$$

1. If $0 < v(x,y-1) - v(x+1,y-1) \leq \mu_2 c/(\mu_2 - \mu_1)$, then

$$H_1'v(x+1,y) - H_1'v(x,y-1)$$

$$
\begin{aligned}
&\leq \mu_2 v(x+1,y) + \mu_1 \left[v(x+1,y-1) - v(x+1,y)\right] \\
&\quad -\mu_2 v(x,y-1) - \mu_1 \left[v(x+1,y-1) - v(x,y-1)\right] \\
&= \mu_2[v(x+1,y) - v(x,y-1)] - \mu_1[v(x+1,y) - v(x,y-1)] \\
&= (\mu_2 - \mu_1)[v(x+1,y) - v(x,y-1)] \leq 0
\end{aligned}
$$

2. If $v(x,y-1) - v(x+1,y-1) > \mu_2 c/(\mu_2 - \mu_1)$, then

$$
\begin{aligned}
&H_1' v(x+1,y) - H_1' v(x,y-1) \\
&\leq \mu_2 v(x+1,y) + \mu_2 \left[v(x+1,y-1) - v(x+1,y) + c\right] \\
&\quad -\mu_2 v(x,y-1) - \mu_2 \left[v(x+1,y-1) - v(x,y-1) + c\right] \\
&= 0
\end{aligned}
$$

Thus, $H_1' v(x+1,y) \leq H_1' v(x,y-1)$.

Second, we prove that $H_1' v$ satisfies Equation 4.9. For any (x,y) with $y > 0$, we have $v(x,y-1) \leq v(x,y)$ and $v(x+1,y-1) \leq v(x+1,y)$. Then, we can get that

$$
H_1' v(x,y) = \mu_2 v(x,y) + \min \left\{ \begin{array}{l}
\mu_1 \left[v(x+1,y) - v(x,y)\right] \\
\mu_2 \left[v(x+1,y) - v(x,y) + c\right] \\
\mu_1 \left[v(x,y-1) - v(x,y)\right] \\
\mu_2 \left[v(x,y-1) - v(x,y) + c\right]
\end{array} \right\}
$$

and

$$
H_1' v(x,y-1) = \mu_2 v(x,y-1) + \min \left\{ \begin{array}{l}
\mu_1 \left[v(x+1,y-1) - v(x,y-1)\right] \\
\mu_2 \left[v(x+1,y-1) - v(x,y-1) + c\right] \\
\mu_1 \left[v(x,y-2) - v(x,y-1)\right] \\
\mu_2 \left[v(x,y-2) - v(x,y-1) + c\right] \\
0
\end{array} \right\}
$$

1. If $H_1'v(x,y) = \mu_2 v(x,y) + \mu_1 [v(x+1,y) - v(x,y)]$, then

$$\begin{aligned}
& H_1'v(x,y-1) - H_1'v(x,y) \\
\le\; & \mu_2 v(x,y-1) + \mu_1 [v(x+1,y-1) - v(x,y-1)] \\
& -\mu_2 v(x,y) - \mu_1 [v(x+1,y) - v(x,y)] \\
\le\; & \mu_2 v(x,y-1) + \mu_1 [v(x+1,y) - v(x,y-1)] \\
& -\mu_2 v(x,y) - \mu_1 [v(x+1,y) - v(x,y)] \\
=\; & (\mu_2 - \mu_1)[v(x,y-1) - v(x,y)] \le 0
\end{aligned}$$

2. If $H_1'v(x,y) = \mu_2 v(x,y) + \mu_2 [v(x+1,y) - v(x,y) + c]$, then

$$\begin{aligned}
& H_1'v(x,y-1) - H_1'v(x,y) \\
\le\; & \mu_2 v(x,y-1) + \mu_2 [v(x+1,y-1) - v(x,y-1) + c] \\
& -\mu_2 v(x,y) - \mu_2 [v(x+1,y) - v(x,y) + c] \\
=\; & \mu_2 [v(x+1,y-1) - v(x+1,y)] \le 0
\end{aligned}$$

3. If $H_1'v(x,y) = \mu_2 v(x,y) + \mu_1 [v(x,y-1) - v(x,y)]$, then

$$\begin{aligned}
& H_1'v(x,y-1) - H_1'v(x,y) \\
\le\; & \mu_2 v(x,y-1) - \mu_2 v(x,y) - \mu_1 [v(x,y-1) - v(x,y)] \\
=\; & (\mu_2 - \mu_1)[v(x,y-1) - v(x,y)] \le 0
\end{aligned}$$

4. If $H_1'v(x,y) = \mu_2 v(x,y) + \mu_2 [v(x,y-1) - v(x,y) + c]$, then

$$\begin{aligned}
& H_1'v(x,y-1) - H_1'v(x,y) \\
\le\; & \mu_2 v(x,y-1) - \mu_2 v(x,y) - \mu_2 [v(x,y-1) - v(x,y) + c] \\
=\; & -\mu_2 c < 0
\end{aligned}$$

Thus, $H_1'v$ satisfies Equation 4.9, i.e., $H_1'v(x,y-1) \le H_1'v(x,y)$ with $y > 0$.

It can be checked that convexity is implied by submodularity/supermodularity. Thus, it remains to prove that $H_1'v(x,y)$ satisfies Equation 4.12–4.14. First we prove that $H_1'v$ satisfies Equation 4.12. Define $w(u,x,y)$ as a function defined on $\{0,1,2,3,4\} \times Z \times Z$ such that

$$
\begin{aligned}
& w(u,x,y) \\
= \; & \frac{1}{4} \times \frac{1}{3} \times \frac{1}{2}(1-u)(2-u)(3-u)(4-u)\mu_2 v(x,y) \\
& + \frac{1}{3} \times \frac{1}{2} u(2-u)(3-u)(4-u)\left[\mu_1 v(x+1,y) + (\mu_2-\mu_1)v(x,y)\right] \\
& + \frac{1}{2} \times \frac{1}{2} u(u-1)(3-u)(4-u)\left[\mu_2 v(x+1,y) + c\mu_2\right] \\
& + \frac{1}{3} \times \frac{1}{2} u(u-1)(u-2)(4-u)\left[\mu_1 v(x,y-1) + (\mu_2-\mu_1)v(x,y)\right] \\
& + \frac{1}{4} \times \frac{1}{3} \times \frac{1}{2} u(u-1)(u-2)(u-3)\left[\mu_2 v(x,y-1) + c\mu_2\right] \\
= \; & \begin{cases}
\mu_2 v(x,y), & \text{if } u=0 \\
\mu_1 v(x+1,y) + (\mu_2-\mu_1)v(x,y), & \text{if } u=1 \\
\mu_2 v(x+1,y) + c\mu_2, & \text{if } u=2 \\
\mu_1 v(x,y-1) + (\mu_2-\mu_1)v(x,y), & \text{if } u=3 \\
\mu_2 v(x,y-1) + c\mu_2, & \text{if } u=4
\end{cases}
\end{aligned}
$$

Then, $H_1'v(x,y) = \min_{u \in \{0,1,2,3,4\}} w(u,x,y)$.

It can be seen that $w(u,x,y)$ is submodular in (x,y) for any given u. In addition, $w(u,x,y+1) - w(u,x,y)$ is decreasing as u increases and then w is submodular with respect to (u,y). Let u_1^* and u_2^* be the minimizers of $H_1'v$ at $(x,y+1)$ and $(x+1,y)$, respectively. If $u_1^* \le u_2^*$, then

$$
\begin{aligned}
& H_1'v(x,y+1) + H_1'v(x+1,y) = w(u_1^*,x,y+1) + w(u_2^*,x+1,y) \\
= \; & w(u_1^*,x,y+1) + w(u_1^*,x+1,y) + w(u_2^*,x+1,y) - w(u_1^*,x+1,y)
\end{aligned}
$$

$$\begin{aligned}
&\geq w(u_1^*, x+1, y+1) + w(u_1^*, x, y) + w(u_2^*, x+1, y) - w(u_1^*, x+1, y)\\
&\geq w(u_1^*, x+1, y) + w(u_1^*, x, y) + w(u_2^*, x+1, y+1) - w(u_1^*, x+1, y)\\
&= w(u_1^*, x, y) + w(u_2^*, x+1, y+1)\\
&\geq H_1'v(x+1, y+1) + H_1'v(x, y)
\end{aligned}$$

The first inequality comes from the submodularity of w in (x, y), the second comes from the submodularity of w in (u, y) and the last comes from the definition of H_1'.

If $u_1^* > u_2^*$, there are 10 possible cases for the ordered pair (u_1^*, u_2^*): (4, 3), (4, 2), (4, 1), (4, 0), (3, 2), (3, 1), (3, 0), (2, 1), (2, 0) and (1, 0). It can be checked that $H_1'v$ satisfies Equation 4.12 for each case. For example, if (u_1^*, u_2^*) takes (4, 3), then

$$\begin{aligned}
& H_1'v(x, y+1) + H_1'(x+1, y)\\
&= \mu_2 v(x, y) + c\mu_2 + \mu_1 v(x+1, y-1) + (\mu_2 - \mu_1) v(x+1, y)\\
&= \mu_2 \left[v(x, y) + v(x+1, y)\right] + \mu_1 \left[v(x+1, y-1) - v(x+1, y)\right] + c\mu_2
\end{aligned}$$

and

$$\begin{aligned}
& H_1'(x, y) + H_1'(x+1, y+1)\\
&\leq \mu_2 v(x, y) + \mu_1 \left[v(x, y-1) - v(x, y)\right] + \mu_2 v(x+1, y) + c\mu_2\\
&= \mu_2 \left[v(x+1, y) + v(x, y)\right] + \mu_1 \left[v(x, y-1) - v(x, y)\right] + c\mu_2\\
&\leq H_1'v(x, y+1) + H_1'(x+1, y)
\end{aligned}$$

By employing the similar method, we can prove that $H_1'v$ also satisfies Equations 4.13 and 4.14. Thus, Lemma 4.1 is obtained. □

Lemma 4.2 *If $v \in V$, then $H_2 v \in V$.*

Proof. See Ha [14]. □

Lemma 4.3 *$c' \in V$.*

Proof. See Ha [14]. □

From Lemmas 4.1–4.3, we can obtain the following Lemma 4.4 and then Theorem 4.1.

Lemma 4.4 *$f' \in V$ and $f \in V$.*

Proof. See Ha [14].

Theorem 4.1 *1. Define*

$$R(y) = \begin{cases} \min\{x : f(x+1,y) > f(x,y-1)\}, & \text{if } y > 0 \\ \min\{x : f(x+1,y) > f(x,y)\}, & \text{if } y = 0 \end{cases}$$

$$S(y) = \max\{x : f(x,y) - f(x+1,y) > \mu_2 c/(\mu_2 - \mu_1)\}$$

$$B(x) = \min\{y : f(x,y) - f(x,y-1) > \mu_2 c/(\mu_2 - \mu_1)\}$$

(a) $R(y) \geq 0$ and $R(y)$ is non-increasing as y increases.

(b) $S(y)$ is non-decreasing as y increases.

(c) $B(x)$ is non-decreasing as x increases.

2. Production Control Policy

(a) When there are class-1 backorders, it is always optimal to produce either normally if $x > S(y)$ or urgently if $x \leq S(y)$ to satisfy class-1 backorders.

(b) *When there are only class-2 backorders, it is optimal to produce to stock if the on-hand inventory level is below $R(y)$ and to satisfy class-2 backorders otherwise.*

i. *If producing to stock, it is optimal to produce normally if the on-hand inventory level is above $S(y)$, and to produce urgently otherwise.*

ii. *If satisfying class-2 backorders, it is optimal to produce normally if the number of class-2 backorders is below $B(x)$, and to produce urgently otherwise.*

(c) *When there is no any backorder, it is optimal to produce urgently to increase the on-hand inventory level if $x \leq S(0)$ and produce normally to increase the on-hand inventory level up to $R(0)$ if $x > S(0)$.*

3. *Inventory Allocation Policy*

It is optimal to satisfy an incoming class-2 demand from on-hand inventory if the on-hand inventory level is above $R(y+1)$ and to backorder this demand otherwise.

Proof. For part 1a, please refer to Ha [14]. For part 1b, we need to show $S(y+1) \geq S(y)$. Suppose the contrary that $S(y+1) < S(y)$. By the definition of $S(y+1)$, $f(S(y), y+1) - f(S(y)+1, y+1) \leq \mu_2 c/(\mu_2 - \mu_1)$. Similarly, by the definition of $S(y)$, $f(S(y), y) - f(S(y)+1, y) > \mu_2 c/(\mu_2 - \mu_1)$. Then, $f(S(y), y+1) - f(S(y)+1, y+1) < f(S(y), y) - f(S(y)+1, y)$. However, by Equation 4.12, it is shown that $f(S(y), y+1) - f(S(y)+1, y+1) \geq f(S(y), y) - f(S(y)+1, y)$, which is a contradiction. Therefore we must have $S(y+1) \geq S(y)$.

For part 1c, we need to show that $B(x) \geq B(x-1)$. Suppose the contrary that $B(x) < B(x-1)$. By the definition of $B(x-1)$, $f(x-1, B(x)) - f(x-1, B(x)-1) \leq \mu_2 c/(\mu_2 - \mu_1)$. And by the definition of $B(x)$, $f(x, B(x)) - f(x, B(x)-1) > \mu_2 c/(\mu_2 - \mu_1)$. Then, $f(x-1, B(x)) - f(x-1, B(x)-1) < f(x, B(x)) - f(x, B(x)-1)$. By Equation 4.12, $f(x-1, B(x)) - f(x-1, B(x)-1) \geq f(x, B(x)) - f(x, B(x)-1)$, which is a contradiction. Therefore, we must have $B(x) \geq B(x-1)$.

Consider part 2a. From Equation 4.8, it follows that $f(x+1, y) - f(x, y) \leq f(x, y-1) - f(x, y)$ and $f(x+1, y) - f(x, y) \leq 0$. Thus, $H_1 f(x, y)$ becomes

$$H_1 f(x,y) = \mu_2 f(x,y) + \min \left\{ \begin{array}{l} \mu_1 \left[f(x+1,y) - f(x,y) \right] \\ \mu_2 \left[f(x+1,y) - f(x,y) + c \right] \end{array} \right\}$$

It is obvious that we should always produce for increasing x if $x < 0$. In addition, it is optimal to produce normally if $f(x, y) - f(x+1, y) \leq \mu_2 c/(\mu_2 - \mu_1)$ and to produce urgently otherwise. Because of the definition of $S(y)$ and convexity of $f(x, y)$ with respect to x, $x \geq S(y)$ can guarantee that $f(x, y) - f(x+1, y) \leq \mu_2 c/(\mu_2 - \mu_1)$ and then it is optimal to produce normally and produce urgently otherwise.

Now we consider part 2b. From Equation 4.9, if $y > 0$, $f(x, y-1) \leq f(x, y)$, i.e., we never stop production if there are class-2 backorders. From the definition of $R(y)$ and Equation 4.13, $f(x+1, y) > f(x, y-1)$ for all $x \geq R(y)$. Then, $f(x+1, y) - f(x, y) > f(x, y-1) - f(x, y)$ and $H_1 f(x, y)$ becomes

$$H_1 f(x,y) = \mu_2 f(x,y) + \min \left\{ \begin{array}{l} \mu_1 \left[f(x,y-1) - f(x,y) \right] \\ \mu_2 \left[f(x,y-1) - f(x,y) + c \right] \end{array} \right\}$$

Thus, it is optimal to produce to reduce y if $x \geq R(y)$. By analyzing the last term of the above equation, it can be shown that it is optimal to produce normally if $f(x,y) - f(x,y-1) \leq \mu_2 c/(\mu_2 - \mu_1)$ and to produce urgently otherwise. Because of the definition of $B(x)$ and convexity of $f(x,y)$ with respect to y, $y < B(x)$ can guarantee that $f(x,y) - f(x,y-1) \leq \mu_2 c/(\mu_2 - \mu_1)$ and thus it is optimal to produce normally and produce urgently otherwise.

From the definition of $R(y)$ and Equation 4.13, $f(x+1,y) \leq f(x,y-1)$ if $x < R(y)$. Then, $f(x+1,y) - f(x,y) \leq f(x,y-1) - f(x,y)$ and $H_1 f(x,y)$ becomes

$$H_1 f(x,y) = \mu_2 f(x,y) + \min \left\{ \begin{array}{l} \mu_1 \left[f(x+1,y) - f(x,y) \right] \\ \\ \mu_2 \left[f(x+1,y) - f(x,y) + c \right] \end{array} \right\}$$

Thus, it is optimal to produce to increase x if $x < R(y)$. By analyzing the last term of the above equation, we can get that it is optimal to produce normally if $f(x,y) - f(x+1,y) \leq \mu_2 c/(\mu_2 - \mu_1)$ and to produce urgently otherwise.

Consider part 2c. If there is no any backorder, we can only produce to stock. Then, $H_1 f(x,y)$ becomes

$$H_1 f(x,y) = \min \left\{ \begin{array}{l} \mu_1 \left[f(x+1,y) - f(x,y) \right] \\ \\ \mu_2 \left[f(x+1,y) - f(x,y) + c \right] \\ \\ 0 \end{array} \right\}$$

Thus, it is optimal not to produce if $f(x,y) - f(x+1,y) \leq 0$, to produce normally if $0 < f(x,y) - f(x+1,y) \leq \mu_2 c/(\mu_2 - \mu_1)$ and to produce urgently if $f(x,y) - f(x+1,y) > \mu_2 c/(\mu_2 - \mu_1)$. From the definitions of $R(y)$ and $S(y)$ and convexity of $f(x,y)$

with respect to x, $x \geq R(y)$ guarantees that $f(x,y) - f(x+1,y) \leq 0$ and then it is optimal not to produce, $S(y) < x < R(y)$ guarantees that $0 < f(x,y) - f(x+1,y) \leq \mu_2 c/(\mu_2 - \mu_1)$ and then it is optimal to produce normally and $x < S(y)$ guarantees that $f(x,y) - f(x+1,y) > \mu_2 c/(\mu_2 - \mu_1)$ and then it is optimal to produce urgently.

For part 3, please see Ha [14]. □

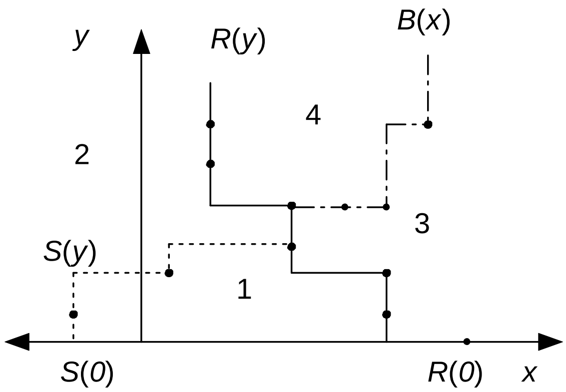

Figure 4.2: The optimal policy characterized by $R(y)$, $S(y)$ and $B(x)$

The form of the optimal control policy is illustrated in Figure 4.2. The state space $Z \times Z^+$ is partitioned into four areas, namely area 1, 2, 3 and 4, by the three critical inventory levels, $R(y)$, $S(y)$ and $B(x)$. If (x,y) falls in area 1, it is optimal to produce normally to increase x. If (x,y) falls in area 2, it is optimal to produce urgently to increase x. If (x,y) falls in area 3, it is optimal to produce normally to reduce y. If (x,y) falls in area 4, it is optimal to produce urgently to reduce y. For an incoming class-2 demand, it is optimal to satisfy it from on-hand inventory if the on-hand inventory level is above $R(y+1)$ and to backorder this demand otherwise.

4.2. Conclusions

In this chapter, we consider a make-to-stock production system with two production rates, two demand classes and backorders. The optimal control policy is shown to be characterized by three monotone switch curves $R(y)$, $S(y)$ and $B(x)$. The state space of the production system is partitioned by the three curves into four areas, each of which corresponds to a different production decision.

Chapter 5

Conclusions and Future Study

In this dissertation, optimal control policies are developed for make-to-stock production systems under different operating conditions. First, a make-to-stock production system with two production rates, one demand class, Poisson demand, exponential production time and backorders are considered. It is found the (S_1, S_2) control policy is optimal for the production system, where S_1 acts like a base-stock level and S_2 controls the switch between the normal and emergency production rate. Specifically, it is optimal not to produce if the net inventory level is at or above S_1, to produce normally if the net inventory level is below S_1 and at or above S_2 and to produce urgently if the net inventory level is below S_2. Later on, the developed model is generalized to consider N production rates, where the optimal control policy is the $(S_1, S_2, \ldots, S_N)$ policy. Specifically, it is optimal not to produce if the net inventory level is at or above S_1, to produce with k^{th} production rate if the net inventory level is below S_k and at or above S_{k+1}, $k = 1, 2, \ldots, N-1$, and to production with the N th production rate if the net inventory level is below S_N. An $M/M/1/S$ queueing model is developed as well to compute the expected total cost per unit time for the production system with two rates under the (S_1, S_2) policy. To show the benefits of employing the emergency rate, numerical studies are carried out to compare the expected total cost per unit time between the production system with two rates and the one with a single rate. The result obtained shows that the emergency production rate can generate a significant cost saving under most cases studied.

Second, a make-to-stock production system with two production rates, N demand classes, Poisson demand, exponential production time and lost sales are considered. It is found that the optimal control policy is the $(S_1, S_2, R_1, \ldots, R_N)$ policy, which is a combination of the (S_1, S_2) policy and the so-called stock reservation policy. The (S_1, S_2) policy is employed to control the production process while the stock-reservation policy is used to control inventory allocation among N demand classes. Demand of class i is satisfied when the inventory level is above R_i and rejected otherwise. An $M/M/1/S$ queueing model is also developed to compute the expected total cost per unit time for the production system with two production rates and two demand classes.

Finally, a make-to-stock production system with two production rates, two demand classes, Poisson demand, exponential production time and backorders are studied. The optimal control policy is shown to be characterized by three monotone switch curves $R(y)$, $S(y)$ and $B(x)$. The state space of the production system is partitioned by the three curves into four areas, each of which corresponds to a different production decision.

The main limitation of our models is the assumption of exponential production times, which might be difficult to be realized in reality. However, this assumption is important to make our problem tractable, without which the memoryless property is missing and the problems studied become much more complex. Nevertheless, one direction of the future research is to relax this assumption and allow production time to be of any kind of distribution. Another possible direction for the future research is to consider both backorders and lost sales simultaneously for a production system with multiple production rates and demand classes. For such a production system, when a demand arrives, we can satisfy, backorder or reject it. This must be making the problem much complicated.

Bibliography

[1] E. W. Barankin, A delivery-Lag inventory model with an emergency provision, Naval research Logistics Quarterly 8 (1961), 285–311.

[2] E. V. Bulinskaya, Some results concerning optimum inventory policies, Theory of Probability and Applications 9 (1964), 389–403.

[3] C. Chiang, and G. J. Gutierrez, A periodic review inventory system with two supply modes, European Journal of Operational Research 94 (1996), 527–547.

[4] C. Chiang, and G. J. Gutierrez, Optimal control policies for a periodic review inventory system with emergency orders, Naval Research Logistics 45 (1998), 187–204.

[5] E. Çinlar, Introduction to Stochastic Processes, Prentice-Hall, Englewood Cliffs, New Jersey, 1975.

[6] M. A. Cohen, P. R. Kleindorfer, and H. L. Lee, Service constrained (s, S) inventory systems with priority demand classes and lost sales, Management Science, 34 (1988), 482–499.

[7] K. H. Daniel, A delivery-lag inventory model with emergency order, Chapter 2, Multistage Inventory Models and Techniques, Scarf, Gilford, Shelly (Eds), Stanford University Press, Stanford, CA, 1962.

[8] R. Dekker, M. J. Kleijn, P. J. de Rooij, A spare parts stocking policy based on equipment criticality, International of Production Economics, 56–57 (1998), 69–77.

[9] R. V. Evans, Sales and restocking policies in a single item inventory system, Management Science, 14 (1968), 463–472.

[10] Y. Fukuda, Optimal policies for the inventory problem with negotiable lead-time, Management Science 10 (1964), 690–708.

[11] S. G. Johansen and A. Thorstenson, An inventory model with poisson demands and emergency orders, International Journal of Production Economics 56–57 (1998), 275–289.

[12] A. Ha, Inventory rationing in a make-to-stock production system with several demand classes and lost sales, Management Science 43 (1997), 1093–1103.

[13] A. Ha, Optimal dynamic scheduling policy for a make-to-stock production system, Operations Research 45 (1997), 42–53.

[14] A. Ha, Stock-rationing policy for a make-to-stock production system with two priority classes and backordering, Naval Research Logistics 44 (1997), 457–472.

[15] S. Kalpakam and K. P. Sapna, A two reorder level inventory system with renewal demands, European Journal of Operational Research 84(2) (1995), 402–415.

[16] A. Kaplan, Stock rationing, Management Science, 15 (1969), 260–267.

[17] L. Li, The role of inventory in delivery-time competition, Management Science 38 (1992), 182–197

[18] S. Lippman, Applying a new device in the optimization of exponential queueing system, Operation Research 23 (1975), 687–710.

[19] K. Moinzadeh and C. P. Schmidt, An $(S-1, S)$ inventory system with emergency orders, Operations Research 39 (1991), 308–321.

[20] K. Moinzadeh and S. Nahmias, A continuous review model for an inventory system with two supply modes, Management Science 34 (1988), 761–774.

[21] I. Moon and S. Kang, Rationing policies for some inventory systems, Journal of the Operational Research Society, 49 (1998), 509–518.

[22] S. Nahmias and S. Demmy, Operating characteristics of an inventory system with rationing, Managemant Science, 27 (1981), 1236–1245.

[23] M. F. Neuts, An inventory model with optional lag time, SIAM Journal of Applied Mathematics 12 (1964), 179–185.

[24] E. L. Porteus, Foundations of Stochastic Inventory Theory, Stanford University Press, Stanford, California, 2002.

[25] G. Tagaras and D. Vlachos, A periodic review inventory system with emergency replenishments, Management Science 47 (2001), 415–429.

[26] D. M. Topkis, Optimal ordering and rationing policies in a nonstationary dynamic inventory model with n demand classes, Management Science 15 (1968), 160–176.

[27] A. F. Jr. Veinott, Optimal policy in a dynamic, single product, nonstationary inventory model with several demand classes, Operations Research 13 (1965), 761–778.

[28] A. F. Whittemore and S. Saunders, Optimal inventory under stochastic demand with two supply options, SIAM Journal of Applied Mathematics 32 (1977), 293–305.