YAŞAR UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

MASTER THESIS

CONTROL OF A MULTI-SERVER MAKE-TO-STOCK PRODUCTION SYSTEM WITH SETUP COSTS

Sinem ÖZKAN

Thesis Advisor: Asst. Prof. Dr. Önder BULUT

Department of Industrial Engineering

Presentation Date: 25.01.2016

Bornova-İZMİR
2016

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of master of science.

Asst. Prof. Dr. Önder BULUT (Supervisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of master of science.

Prof. Dr. M. Cemali DİNÇER

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of master of science.

Asst. Prof. Dr. Zeynep SARGUT

Prof. Dr. Cüneyt GÜZELİŞ
Director of the Graduate School

ABSTRACT

CONTROL OF A MULTI-SERVER MAKE-TO-STOCK

PRODUCTION SYSTEM WITH SETUP COSTS

ÖZKAN, Sinem
MSc in Industrial Engineering
Supervisor: Asst. Prof. Dr. Önder BULUT
February 2016, 61 pages

This study considers production and inventory control problems for a make-to-stock queue with production setup costs, several customer classes and lost sales. At any system state, production decision is to specify whether to activate new production channels or to continue with the currently active ones. If the decision is to activate new channels, a fixed/setup cost is incurred per channel. At the decision epochs where the system experiences demand from any customer class, the controller should also decide whether to satisfy the arriving demand or to reject it. The literature of the control of make-to-stock queues is extended by considering fixed system costs and multiple servers at the same time. Firstly, the structure of the optimal production and rationing policies are characterized and then new/alternative policies that have well-defined structures and are easier to apply are proposed. Numerical and theoretical studies are carried out to assess the performances of the proposed policies. The expected average cost of the optimal production policy for the single-server make-to-stock queue is obtained conducting a renewal analysis.

Keywords: Fixed/Setup cost, inventory and production control, stock rationing, make-to-stock, multiple servers, optimal control.

ÖZET

PARALEL ÜRETİM KANALLI VE HAZIRLIK MALİYETLİ STOĞA-ÜRETİM SİSTEMLERİNİN KONTROLÜ

Sinem ÖZKAN
Yüksek Lisans Tezi, Endüstri Mühendisliği Bölümü
Tez Danışmanı: Yrd. Doç. Dr. Önder BULUT
Şubat 2016, 61 sayfa

Bu çalışmada paralel üretim kanalları olan, birden çok müşteri sınıfına sahip, hazırlık ve kayıp satış maliyetli Stoğa-üretim sistemleri için üretim ve stok tayınlama kontrol problemleri ele alınmaktadır. Üssel dağılıma sahip üretim zamanı içeren bu sistem $M/M/$s stoğa-üretim kuyruk modeli olarak incelenmektedir. Herhangi bir sistem-durumunda, üretim kararı ya aktif olan üretim kanal sayısının arttırılmasını ya da aktif olan kanal sayısı ile devam edilmesini belirtir. Karar anlarında, eğer herhangi bir müşteri sınıfından talep gelirse, tayınlama kararı ya gelen talebin karşılanmasını ya da reddedilmesini sağlar. Bu çalışmayla, hazırlık maliyetli Stoğa-üretim sistemlerini şimdiye kadar tek bir üretim kanalıyla modelleyen çalışmaları içeren teknik yazına önemli katkıda bulunulmaktadır. Öncelikle, en iyi üretim ve tayınlama politikalarının yapıları belirlenmiş, daha sonra ise en iyi politikalara benzer ve uygulaması daha kolay olan alternatif politikalar sunulmuştur. Sunulan politikaların performanslarının ölçülebilmesi için nümerik ve teorik çalışmalar yapılmıştır. Tek kanallı sistem için en iyi üretim politikasının beklenen ortalama maliyeti hesabı için Yenileme Ödül Teoremi kullanılmıştır.

Anahtar sözcükler: Sabit/Hazırlık maliyeti, envanter ve üretim kontrol, stok tayınlama, Stoğa-üretim, paralel üretim kanalları, en iyi kontrol.

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my supervisor Asst. Prof. Dr. Önder Bulut for the continuous support of my graduate study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis.

Besides my supervisor, I would like to thank Prof. Dr. M. Murat Fadıloğlu for his insightful comments and encouragement.

I thankfully acknowledge the support of TUBITAK - Turkish Technological and Scientific Research Institute for during my master period. This research is a part of a project supported by TUBITAK project number 213M355.

The last but not the least, I would also like to thank my parents, sister and spouse. They were always there supporting me and encouraging me with their best wishes.

Sinem ÖZKAN
İzmir, 2016

TEXT OF OATH

I declare and honestly confirm that my study, titled "Control of a Multi-Server Make-to-Stock Production System with Setup Costs" and presented as a Master's Thesis, has been written without applying to any assistance inconsistent with scientific ethics and traditions, that all sources from which I have benefited are listed in the bibliography, and that I have benefited from these sources by means of making references.

TABLE OF CONTENTS

INDEX OF FIGURES

INDEX OF TABLES

INDEX OF SYMBOLS AND ABBREVIATIONS

Symbols	Explanations
K	Fixed cost
s	Number of available servers
h	Holding cost per unit
c_i	Lost sales cost of demand class i
λ_i	Demand rate of class i
μ	Production rate
α	Discount rate
$X(t)$	Inventory level at time t
$Y(t)$	Number of active channels at time t
$u_p(t)$	The production decision at time t
$u_r(t)$	The rationing decision at time t

Abbreviations

FCFS	First Come First Served

BSP	Base-Stock Policy
MDP	Markov Decision Process
RA	Renewal Analysis
PPP1	Proposed Production Policy 1
PPP2	Proposed Production Policy 2
PPP3	Proposed Production Policy 3

1 INTRODUCTION

The joint production and inventory control arises when there is limited number of production channels and the underlying system is make-to-stock. In general, production controllers desire to produce the ideal amount of products to meet random demand so as to balance holding and shortage costs. Today, in addition to these features, most of the production systems experience demands from the several different customer classes. Differentiation among customer classes is based on either the shortage costs or the service level requirements. If the customers have different shortage costs/service level requirements, reservation of inventory gains importance at this stage. It is a good strategy to reserve the inventory by not satisfying demand from lower priority classes in anticipation of future demand from the higher priority classes.

In this thesis, we consider the problem of production control and stock rationing of a single-item, make-to-stock production facility with parallel production channels and several demand classes. Contrary to the vast majority of the literature, this problem, which is considered in the lost sales environment, is considered under existence of fixed production cost. There are two essential decisions should be jointly addressed for managing such a system: *The first one* is the production control decision that dictates when to start production and how much to produce i.e., how many production channels should be activated. *The second control* is concerned with inventory rationing. *Stock-rationing* is mainly to reserve inventory for the future demand from more valuable demand classes by rejecting demands from less valuable demand classes. Thus, the joint management of production and inventory control has a huge potential to decrease the total production and inventory related costs.

In the most general setting, this joint problem is hard to solve and can be even intractable. The system-state should dynamically keep track of the number of outstanding production orders, the indices of the active channels (processing time distributions might be different for different channels), and the age information (or the remaining completion times) of all these random number of orders. Hence, the structure of the unknown optimal policy would be highly dynamic and would not possess a systematic behaviour for all practical purposes. Thus, in addition to characterize the optimal production and rationing policies as much as possible, introducing and analyzing new/alternative policies which are well-performing and have well-defined structures would be very beneficial for both literature and practice.

Although the general setting is too complex to analyze and we make some assumptions, in our model we still allow parallel channels which is not typical in the literature. That is, we have the flexibility of changing the number of channels/servers. This provides us to consider our problem under different settings that can be grouped in two main categories: *Capacitated* and *Uncapacitated systems*. For *Capacitated production-inventory systems*, the number of production channels is limited and most of the time all of them are utilized. The extreme of such a system has only a single production channel. In the production control literature, the optimal production policy for such a single-server system is to produce until a threshold inventory level and stop the production (Ha (1997a, 1997b)) which is the well-known *Base-Stock Policy*. Moderate values of number of production channels are analyzed as parallel production channels. For this case, the optimal production and rationing policies are more dynamic and the optimal rationing policy is state-dependent threshold type (Bulut and Fadıloğlu (2011)). As the number of servers approaches to infinity, capacitated production systems converge to *Uncapacitated systems*: At any time needed, a new order can be triggered. Hence, the infinite number of channels helps us to analyze typical inventory systems having ample supply.

Another problem addressed in this study is allocation of inventory among several demand classes from a common stock pool. The common stock pool provides not to have excessive inventory and to use system resources effectively. In inventory systems, rationing is a well-known strategy and is observed frequently in real life. Also, customer differentiation is often observed in different service industries such as hotels, banks or airline management etc. While we differentiate customers to manage the inventory, service industry uses customer differentiation strategy in queuing and revenue management.

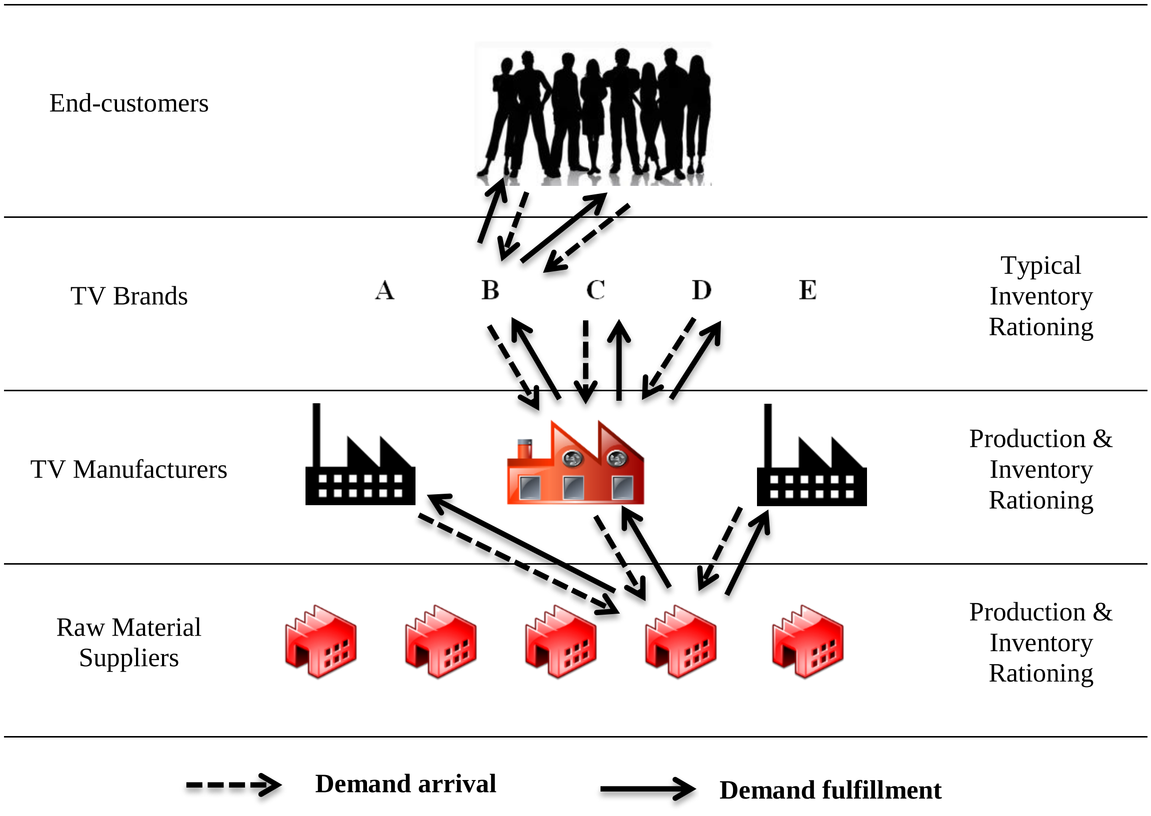

Figure 1.1 A Supply Chain Example

In order to better understand the problem, let us consider the following supply chain example depicted in Figure 1.1 where we can face with our problem at different echelons: Suppose there is a television manufacturer that produces televisions and gives services to many different brands or companies. It has parallel production lines that produce with a fixed cost to a common buffer stock. All the items in the buffer stock are identical and semi-finished. After a demand occurs, work in process is pulled and differentiated based on different requests of the customers. This strategy is known as *Delayed Differentiation* in the literature (Lee and Tang (1997)). TV manufacturer has many customers and each of them order televisions that have different priorities. For instance, one of the customers of the manufacturer might be the world's number one and hence this specific brand would have the highest priority for the TV manufacturer because lost sales cost of this brand would be the highest.

We can also consider the lower echelon of the supply chain from our problem perspective. In order to trigger the production, TV manufacturer needs raw materials. Hence, the TV manufacturer has raw material suppliers to supply them. If we consider one of these suppliers, our TV manufacturer is one of the customers of this raw material supplier. The raw material supplier should also have its own production and rationing strategies to meet demands from several TV manufacturers. On the other hand, at the higher echelon of the supply chain, end-customer demands come to any brand which is one of the customers of TV manufacturer. This company has to decide to satisfy or not to satisfy the demand based on customer differentiation. This decision is related with typical inventory rationing. Occurrence of end-customer demand leads to order the televisions from TV manufacturer. Then, our TV manufacturer starts to the differentiation of semi-products to satisfy the demand of the company. As seen from this example the problem we consider in this thesis can be experienced at different levels of the supply chains of many industries.

In this thesis, we first characterize the structural properties of the optimal production and rationing policies for a single product, multi-server, make-to-stock production system with fixed production costs, multiple customer classes and lost sales. We then propose well-performing alternative policies and conduct the performance analyses. In order to analyze such a system we assume the following: Demands from different customer classes are generated according to independent stationary Poisson processes. Processing times at each identical server are independent Exponential random variable with rate μ. Hence, we model production-inventory system as an $M/M/s$ make-to-stock queue. In this model, at any system state the controller decides the number of new channels/servers that should be activated in conjunction with the rationing decision. The cancellation of previously placed production orders is not permitted.

The remainder of this thesis is organized as follows: We review the related literature in Chapter 2. In Chapter 3 we present problem formulation and numerical characterization of the optimal production and rationing policies. In Chapter 4 we propose new policies and compare their structure, applicability and performances with the optimal ones. In this chapter, we also provide the renewal analysis for $M/M/1$ make-to-stock queue. We finally provide an overall summary of the thesis and discuss possible directions for future research in Chapter 5.

2 LITERATURE REVIEW

The related literature can be classified into two categories from both chronological and methodological perspectives. The first category is defined as the literature published before Ha (1997a) and the second category is defined as the literature published after Ha (1997a). In earlier studies such as Sobel (1968), Gavish and Graves (1980,1981), Graves and Keilson (1981), single-server, single-demand-class settings with setup costs are considered and analyzed using queueing theory techniques. In the second category, starting with Ha (1997a), in almost all the studies, problem is attacked using Markov Decision Process (MDP) analysis techniques. Interestingly, all such studies assume that setup/startup cost is zero. On the other hand, in this second research stream, studies consider more than single customer class and accordingly rationing policies. In parallel with the first category, the studies in this second category also assume a single processing channel with only one exception which is Bulut and Fadıloğlu (2011). Mainly, our study extends the literature before Ha (1997a) by increasing the number of production channels and the literature after Ha (1997a) by considering the fixed cost per each active production channel.

The earlier studies including fixed costs start with Sobel (1968) and analyze queueing systems with a single server. Sobel (1968) analyzes a service system under general service time distributions and general arrival processes with fixed start-up and shut-down costs. He presents a queue control strategy based on a given cost structure. He proves that two-critical-number policy, which resembles the (s, S) policy for the typical inventory systems, is the optimal service policy. This policy includes two control variables denoted by X^* and X^{**} such that $X^* < X^{**}$. If queue length is less than or equal to X^*, service is not provided until queue length reaches to X^{**} , whereupon service starts and continues until queue length drops to X^*, again. Based on the work of Sobel (1968), Gavish and Graves (1980) reconsiders a similar problem in a production-inventory setting. Actually, the queue length in Sobel's study correspond to the number of backorders in this study. They assume unit Poisson demand arrivals, constant demand size and fixed setup cost for production. If demand is rejected, it is backlogged. They model their system as $M/D/1$ queue. They derive cost expressions and propose an efficient search procedure to find the optimal policy parameters. Gavish and Graves (1981) extends the work of Gavish and Graves (1980) by allowing general service times. They generate a procedure to find the steady-state probabilities of each inventory level and used this procedure to search optimal X^* and X^{**} control levels. Graves and Keilson (1981) extends the previous works by including an exponential random variable which is the size of the demand requests. They generate a compensation method for the backordering case that finds a closed-form expression for the expected system cost. They used this procedure to find optimal X^* and X^{**} control levels. If inventory is less than X^*, production channel is triggered to start with a fixed setup cost and continues until inventory level reaches to X^{**}. While the production is continuing, the fixed setup cost is not incurred. If inventory level hits to X^{**}, the production is turned off and stays off until inventory level drops to below X^*.

Tijms (1980) studied the similar system with Poisson demands and random service times. The optimal control values are found by using Semi-Markov decision process techniques. Besides, the model also includes the time taken to activate the production channels. Altiok (1986) considers a system with compound Poisson demands and phase-type service times. Steady-state probabilities are calculated for each inventory level. This study can be adapted to both lost sales and backordering environments. Another important study is Lee and Srinivasan (1989) considers a system includes single item and a production channel with fixed start-up costs. The processing time has a general distribution and each demand arrives according to Poisson process. If there is not enough inventory, demands are backordered. The importance of this article is enabling to calculate the cost function by not calculating the steady-state probabilities. They succeed to achieve this using a renewal anaysis. Later on, Lee and Srinivasan (1991) extended the previous work by adding compound Poisson structure.

Ha (1997a) is the first to model the problem by using Markov Decision Process techniques. The setting includes an exponential server with no fixed cost, multiple demand classes with Poisson arrivals and lost sales. He shows that the optimal production policy is Base-stock policy for such a system. Also, the stock rationing problem in production environment was analyzed first by Ha (1997a) and he shows that a static threshold level policy is optimal for rationing. A stationary analysis is performed of a system with two demand classes. The comparisons between the performances of optimal rationing policy and FCFS policy are also conducted.

Ha (1997b) considered the same system with two demand classes, exponential production times and backordering. He defines the system-state with two variables. The first variable denotes to the inventory level. For zero and negative values of the first variable, it corresponds to the number of class 1 backorders. For positive values of the first variable, it corresponds to the units of inventory and the second variable denotes the number of class 2 backorders. He shows the optimal policies are characterized by a single monotone switching curve. Optimal production policy is Base-stock and optimal rationing policy is applied by a threshold level. Vericourt et al. (2002) included multiple demand classes to the work of Ha (1997b). The characterization of the optimal rationing policy is provided and an efficient algorithm is generated to compute optimal control parameters which are the optimal rationing levels for each demand class.

Bulut and Fadıloğlu (2011) included multiple production channels to the work of Ha (1997a). Until our study, this research was the unique study that considers multi-servers. They characterized the properties of the optimal cost function, the optimal production and rationing policies. The optimal production policy is a state-dependent base-stock policy, and the optimal rationing policy is threshold type.

Erlangian service times in the lost sales and backordering environments were considered by Ha (2000) and Gayon, Vericourt and Karaesmen (2009). Ha (2000) analyzes the optimal production and rationing policies in lost sales environment and defines a work storage level as a state of the system. Work storage level is number of completed production stages and has information about inventory level and the status of the production. There are optimal threshold levels for both production and rationing. If work storage level is less than a pre-determined level, the optimal is to produce until the work storage level hits to target value. Also, the work storage level is higher than target level, the optimal rationing decision is to satisfy the arrival demand class. Gayon et al. (2009) also considers the previous model in backordering environment. They show that work-storage type policy is optimal for rationing by providing a partial characterization.

Lee and Hong (2003) extends the work of Ha (1997a, 2000) by integrating setup cost to a production system with multi demand classes and lost sales. Demands arrive according to a Poisson process and processing time for each single item follows a 2-phase Coxian distribution. This system is modeled as continuous time Markov chain and the steady state probabilities are calculated by an efficient algorithm. For given optimal (s, S) obtained, they propose a heuristic algorithm to determine the threshold levels for stock rationing. Recently, Pang et al. (2014) study a system with several customer classes in a lost sales make-to-stock production environment with single channel and no fixed costs. In this study, they allow batch demand and consider general, phase-type and Erlang processing time distributions. It is shown that the optimal rationing policies are time-dependent threshold type.

The studies in stock rationing area were started in the 1960s. Most of the studies are for classical inventory systems. Critical level rationing is the first introduced by Veinott (1965). The model includes zero lead times and backordering costs. His model provides different service levels to several demand classes in periodic inventory system. These service levels provide to allocate on-hand inventory among distinct demand classes. Under the static rationing policy, Nahmias and Demmy (1981) consider two demand classes with Poisson arrivals and apply static rationing with zero lead time. The derivations to find expected number of backorders are performed. This study is the first to analyze stock rationing in continuous time. Deshpande et al. (2003) extend the work of , Nahmias and Demmy (1981) by providing the flexibility to number of outstanding orders. Arslan et al. (2007) consider a system with multiple demand classes, Poisson arrivals and constant lead time. They also provide a heuristic to obtain optimal rationing levels. Fadiloglu and Bulut (2010) proposes a dynamic rationing policy for continuous-review inventory systems. Here, dynamic rationing implies that the rationing levels are functions of the number of outstanding orders and their ages. In general, the optimal rationing policy would be of dynamic type. However, for sake of analysis most of the studies in inventory rationing literature assume static rationing levels.

Liu, Feng and Wong (2014) analyze different inventory rationing policies for an inventory system includes two demand classes. The steady-state probabilities are calculated and the performances of policies are explored. A heuristic algorithm is generated to find the optimal values of control variables. Wang and Tang (2014) address an inventory system with multiple demand classes of backorder and lost sales type. The penalty costs of backorders changes as time progresses. Hence, the importance of demand classes change with time. They analyze a dynamic rationing policy and model a Markov decision process (MDP) to observe optimal dynamic threshold levels for inventory rationing. The optimal rationing policy is shown to be a myopic base stock policy and dynamic rationing policy. A heuristic dynamic inventory policy is introduced to facilitate the solution of the complex problem. Liu and Zhang (2015) analyze an inventory system with two demand classes, Poisson demand with holding and penalty costs in backordering environment. They introduce an efficient method to obtain closed-form expressions for the dynamic threshold levels to overcome the computational complexity.

We would like to conclude the literature of the control of production-inventory systems with Table 2.1 that summarizes the related milestone works. In this table, the studies are classified on number of servers and fixed cost value. Our problem is closely related to the works of Bulut and Fadıloğlu (2011) and Graves and Keilson (1981). We extend the work of Bulut and Fadıloğlu (2011) by adding the fixed production cost per channel and the work of Graves and Keilson (1981) to multi-server case using a two-dimensional state space.

Table 2.1 Summary of the Related Literature on Production-Inventory Control

	Service Time	Single-server with Fixed Cost	Single-server without Fixed Cost	Multi-server with Fixed Cost	Multi-server without Fixed Cost
Backordering	**General**	Sobel (1965) Gavish et al.(1981) Lee et al. (1989) Lee et al. (1991)			
	Erlang		Ha (2000) Vericourt et al. (2002) Gayon et al. (2009)		
	Phase-type				
	Exponential	Graves et al.(1981)	Ha (1997b)		
	Deterministic	Gavish et al.(1980)			
Lost Sales	**General**		Pang et al. (2014)		
	Erlang		Pang et al. (2014)		
	Phase-type	Lee and Hong (2003)	Pang et al. (2014)		
	Exponential	**Our Study**	Ha (1997a)	**Our Study**	Bulut et al. (2011)
	Deterministic				

3 MODEL

In this chapter, we first characterize the structural properties of the optimal production and rationing policies for a single product, identical-multi-server, make-to-stock production with fixed production costs, multiple customer classes and lost sales. In order to analyze such a system we assume following: Demands from customer class i are generated according to a stationary Poisson process with rate λ_i, $i \in \{1,2,\dots,N\}$ and production times are independent Exponential random variables with mean $1/\mu$. Each demand class is prioritized by its lost sales cost c_i that incurred when class i demand is not satisfied. Without loss of generality, it is assumed that $c_1 \geq c_2 \dots \geq c_N$. The fixed cost to activate a new server is K, the holding cost per item in the inventory is h, the discount rate is α and order-cancellation is not allowed. If the cancellation is allowed, we have a flexibility to cancel all previously placed production orders. If almost all available production channels are active until the inventory hits a specific level and one of the servers is completed at that level; the rest of the orders can be cancelled. Based on these assumptions, we model production-inventory system under consideration of $M/M/s$ make-to-stock queue where s is the number of identical production channels/servers.

As mentioned in Chapter 2, there is no work that considers production control and stock rationing in a multi-server make-to-stock production system with fixed costs. Our problem is mostly related with the works of Bulut and Fadıloğlu (2011) and Graves and Keilson (1981). We extend Bulut and Fadıloğlu (2011) by adding the fixed production cost per channel and extend Graves and Keilson (1981) to multi-channel environment.

In the next subsection, our modeling approach, which is based on Markov Decision Analysis, and the corresponding dynamic programming formulation are provided. Afterwards, in Section 3.2, optimal production and rationing policies are characterized with numerical studies. The chapter concludes with Section 3.3 that provides a numerical study that illustrates how optimal policies respond to the changes in system parameters which define arrival and processing rates and cost structure.

3.1 Dynamic Programming Formulation

In order to formulate the problem, the system state should be defined with two variables: $X(t)$ and $Y(t)$. $X(t)$ is the inventory level and $Y(t)$ is the number of active channels at time t. Contrary to the systems with only a single-server, in addition to the inventory level information, we also ought to keep track of the number of active channels in order to identify the inventory replenishment rate at time t. As the number of active servers varies, the production completion times also change. The state space of our state vector $\left(X(t), Y(t)\right)$ is the following:

$$SS=\left\{\left(X(t), Y(t)\right) \mid X(t), Y(t) \in Z^{+} \cup \{0\}, Y(t) \leq s\right\}$$

At any decision epoch t, the system decides either to continue with the same number of active channels or to activate more. The production decision at time t is expressed as $u_p(t)$ such that $u_p(t) \in \{Y(t), Y(t)+1, \dots, s\}$. The second decision is related to the inventory rationing, if a class i demand occurs at time t, the system decides either or not to satisfy the arriving demand. The rationing decision for class i demand at time t is expressed as $u_{r_i}(t)$ such that $u_{r_i}(t) \in \{0,1\},\ i \in \{1,2,\dots,N\}$. If $u_{r_i}(t) = 1$, class i demand is satisfied; else, it is rejected. Given a control policy π, the process $\{(X^{\pi}(t),\ Y^{\pi}(t)) | t \geq 0\}$ is a continuous time Markov process where the transition rate at state (x, y) is $v_{(x,y)} = \sum_{i=1}^{N} \lambda_i u_{r_i} + u_p \mu$. Since the process is Markovian, it is possible to only focus on event occurrences (demand arrivals and production completions). Furthermore, using the uniformization technique proposed by Lippman (1975), we can obtain an equivalent discrete-time problem whose statistical characteristic is the same with statistical characteristic of the original continuous time problem. The uniform transition rate is defined as $v = \sum_{i=1}^{n} \lambda_i + s\mu$. Let α be the discount rate and the optimal cost-to-go function can be written as the following:

$$J(x,y)=\frac{1}{\alpha+v}\min_{y\le u\le s}\{hx+(u-y)K+(s-u)\mu J(x,u)$$
$$+u\mu\min\{J(x+1,u-1),J(x+1,u)\}$$
$$+T_{\mathrm{R}}(x,u)\} \quad (1)$$

where $T_R(x,y)=\sum_{i=1}^{N}T_{R_i}(x,y)$, $i\in\{1,2,\dots,N\}$ and

$$T_{R_i}(x,y)=\begin{cases}\lambda_i\min\{J(x-1,y),c_i+J(x,y)\} & ,x>0\\ \lambda_i(c_i+J(0,y)) & ,x=0\end{cases} \quad (2)$$

Equation (1) minimizes the expected discounted cost through deciding the number of active channels when there are x units on hand and y channels are active. Each time a production channel is switched on, a fixed cost K is incurred. The term $(s-u)\mu J(x,u)$ corresponds to the fictitious self-transitions due to uniformization. This term has a difference because of using production decision u instead of number of active channels y in the cost-to-go function. Because, production decision affects the number of active channels at any time t and changes y with u instantaneously. The term $u\mu\min\{J(x+1,u-1),J(x+1,u)\}$ corresponds to production completion. The minimization operator provides the continuation decision after production completion. That is, if there are active production channels, system can continue to produce with the same number of active channels without paying fixed cost. $T_{R_i}(x,y)$ corresponds to the rationing decision for class i. When a demand of class i arrives, system checks the inventory and decides whether to satisfy or reject the demand. If there is no on-hand inventory, all the arriving demands are lost.

3.2 Characterization of the Optimal Policies

In this section, we aim to characterize the structure of the optimal production and rationing policies. We achieve this aim via Value Iteration Algorithm coded in MATLAB. We run the value iteration algorithm for many different setting to obtain the optimal decisions and the corresponding system costs. Although we provide the DP formulation for the discounted cost criterion, we conduct the numerical analyses under the average system cost criterion using this formulation. The rationale behind this decision is twofold: *i*. Under the average cost criterion all the states converge to the same long-run average cost. However, the long-run discounted cost is state dependent and for such cases it is difficult to compare and interpret the performances of different policies. *ii*. We would like to exclude the discussion/decision about the determination of the discount rate α and directly focus on the impacts of the real system parameters. However, we utilize the modified version of the discounted DP formulation given in Equation (1) to calculate the average system cost using the value iteration algorithm. To find the average cost, we set the discount rate to zero and at each stage divide the value of the cost-to-go function to the completed number of stages (otherwise the cost-to-go function would not converge to the finite average value) in the algorithm. Afterwards, we propose three different termination criteria for the value iteration:

1. The maximum of the absolute values of the difference between the average costs of two consecutive steps for all states is smaller than epsilon (a small number representing the tolerance for termination),
2. For each state, the maximum of the absolute value of the difference between the averages of the all state costs at the current step and all state costs at the previous step is smaller than epsilon,
3. For any step, the absolute value of the difference between the average costs of all states is smaller than epsilon,

When we use one of these termination conditions, the system cost can become finite. Either one of the above criteria can be utilized but we have decided to use the final one. The third one is the most reliable algorithm that checks the differences of the average costs for all the states. Using the first and second algorithms is risky, because the algorithm can be terminated before the all states converge to the same average cost. Pseudo-code of three value iteration algorithms is given in the following where k is for the current stage/step, i represents whether we use the discounted cost criterion ($i = 0$) or the average ($i = 1$), and j is the input for the above three termination criteria when $i = 1$:

$Value\ iteration\ (i, j)$:
$k = 0$
$Assign\ an\ estimated\ value\ for\ J_0\ (for\ example: zero)$
$While\ (difference > epsilon)$
$k = k+1$
$Loop: For\ all\ states$
$Loop: u = \{set\ of\ all\ possible\ production\ decisions\}$
$TK = Rationing\ decision(u)$
$Calculate\ J_k^{cand}(J_{k-1}, state(u), TK)$
$End\ loop$
$J_k(state) = \min_u(J_k^{cand}(J_{k-1}, state(u), TK))$
$End\ loop$
$If\ i = 0$
$difference = max\ |J_k\ (state) - J_{k-1}(state)|$
$End\ loop$
$If\ i = 1$
$If\ j = 1$
$$difference = max\ \left|\frac{J_k\ (state)}{k} - \frac{J_{k-1}\ (state)}{k-1}\right|$$
$End\ loop$
$If\ j = 2$
$$difference = max\ \left|\left(\frac{1}{|\{set\ of\ the\ state\}|}\sum_{state}\frac{J_{k-1}(state)}{k-1}\right) - \frac{J_k\ (state)}{k}\right|$$
$End\ loop$
$If\ j = 3$
$$difference = \max_{state \in SS}\ \max_{state' \in SS/\{state\}}\left|\frac{J_k\ (state)}{k} - \frac{J_k(state')}{k}\right|$$
$End\ loop$
$End\ loop$

Figure 3.1 The Pseudo-code of Value Iteration Algorithms

Our value iteration algorithm is verified with Ha (1997a) and Bulut and Fadıloğlu (2011) in the literature. The optimal production and rationing decisions mentioned in their studies are checked with our results. Besides these studies, it is known that the optimal policy is two-critical number policy (X^{**}, X^{*}) for the exponential systems include the setup cost. For such a system, we decide to start the numerical analysis with the setting $(K, s, \mu, h,\ \lambda_1,\ \lambda_2,\ c_1,\ c_2) =$ (2, 1, 3, 1, 3, 1, 4, 1).

Table 3.1 Optimal Production Decisions u_p when s=1

	Fixed cost $K=0$		Fixed cost $K=2$		Fixed cost $K=4$	
Inventory Level	When $y=0$	When $y=1$	When $y=0$	When $y=1$	When $y=0$	When $y=1$
0	1	1	1	1	1	1
1	1	1	1	1	1	1
2	1	1	1	1	1	1
3	1	1	0	1	0	1
4	0	0	0	1	0	1
5	0	0	0	1	0	1
6	0	0	0	0	0	1
7	0	0	0	0	0	0
	X^* =	4	X^* =	3	X^* =	3
	X^{**} =	4	X^{**} =	6	X^{**} =	7

Table 3.1 shows the optimal production decisions when production is on $(s=y=1)$ and off $(y=0)$. If there is no fixed cost to activate a new channel, the optimal production policy behaves like Base-Stock ($X^*=X^{**}=S=4$). While the setup cost is a positive value, the system tries to extend the production process not to give extra cost to activate a channel again. There is a widening gap between the production control parameters while K increases. According to the results of the table, we have similar optimal production policy properties mentioned in the literature.

In the remaining part of this section and in the following sections, the parametric observations are analyzed and how the system parameters affect to the optimal decisions are shown using the base setting: $(K,s,h,\mu,\lambda_1,\lambda_2,c_1,c_2)=(2,4,1,1,3,1,4,1)$. The holding cost $(h=1)$, the production rate $(\mu=1)$, demand rates of customer classes $(\lambda_1=3,\lambda_2=1)$ and the lost sales cost of these classes $(c_1=4,c_2=1)$ are determined. The optimal decisions for this system are shown in Table 3.2. Each row shows the inventory level and each row shows the number of active channels. For example, the first cell of the first row shows that the inventory levels equals to zero $(x = 0)$ and the production is off $(y = 0)$. The value in the each cell of the decision matrix illustrates the number of servers is needed to be active.

Table 3.2 Optimal Production Decisions u_p

	$(K, s, h, \mu, \lambda_1, \lambda_2, c_1, c_2) = (2, 4, 1, 1, 3, 1, 4, 1)$				
	Number of Active Servers y				
Inventory Level x	0	1	2	3	4
0	4	4	4	4	4
1	3	3	3	3	4
2	2	2	2	3	4
3	2	2	2	3	4
4	0	1	2	3	4
5	0	1	2	3	4

If there is no active server and no inventory on-hand, the production decision is to use all of the limited capacity $u_p(x = 0, y = 0) = 4$. While the number of active servers is fixed and the inventory level is increasing one by one, the optimal production decisions decrease by one or more units. That means, there is no constant order-up-to level. Instead, the target inventory of any state is state dependent. There is an order-up-to level for each inventory level $S_x = x + u_p^*(x, 0)$. Because, the system does not want to give extra holding and setup cost by activating more channels.

For positive values of active servers, the cancellation cost of the production equals to infinity, so the production decisions decrease to at least the number of current active servers. For instance, when $y = 1$ and while the inventory level is increasing, the optimal decision cannot be lower than 1 because no-order cancellation. Thus, the system controls the production by deciding the trade-off between holding and lost sales costs.

Table 3.3 Optimal Inventory Rationing Policy for Class 1

	$(K, s, h, \mu, \lambda_1, \lambda_2, c_1, c_2) = (2, 4, 1, 1, 3, 1, 4, 1)$				
	Number of Active Servers y				
Inventory Level x	0	1	2	3	4
0	0	0	0	0	0
1	1	1	1	1	1
2	1	1	1	1	1
3	1	1	1	1	1

Table 3.4 Optimal Inventory Rationing Policy for Class 2

	$(K, s, h, \mu, \lambda_1, \lambda_2, c_1, c_2) = (2, 4, 1, 1, 3, 1, 4, 1)$				
	Number of Active Servers y				
Inventory Level x	0	1	2	3	4
0	0	0	0	0	0
1	0	0	0	0	0
2	0	0	0	0	1
3	1	1	1	1	1

Table 3.3 and 3.4 show the optimal rationing decisions for the customer class 1 and 2, respectively. In Table 3.3, if there is inventory on-hand, the system always satisfies an arriving demand of class 1. Because, the customer class 1 has the highest lost sales cost and is the most valuable class in the system. Table 3.4 shows that as inventory level and number of active servers increase, the willingness to satisfy a class 2 demand also increases. Hence, threshold inventory level for class 2 is a function of x and y. Therefore, the inventory rationing policy is provided for just class 2.

Stock rationing gains importance when there are more than one demand class and each of them has different lost sales cost. If the lost sales costs of the classes are the same, the customer of the system becomes just a one class. When the lost sales costs are different, the production and rationing decisions are made jointly.

Table 3.5 and 3.6 show the optimal production and rationing decisions under the discounted cost criterion. For this example, the discount rate (α) equals to 0.4. Even if the decisions in Table 3.2 and 3.5 are not the same, the characteristics of the production policies are not so different. This situation is also valid for stock rationing decisions.

Table 3.5.Optimal Production Decisions Under Discounted Cost Criterion

	Number of Active Servers y				
Inventory Level x	0	1	2	3	4
0	3	3	3	3	4
1	2	2	2	3	4
2	1	1	2	3	4
3	0	1	2	3	4
4	0	1	2	3	4
5	0	1	2	3	4

Table 3.6 Optimal Rationing Decisions Under Discounted Cost Criterion

	Number of Active Servers y				
Inventory Level x	0	1	2	3	4
0	0	0	0	0	0
1	0	0	0	0	0
2	0	0	0	0	1
3	1	1	1	1	1

3.3 The Impact of the System Parameters on the Optimal Policies

This section includes the analysis of how the system parameters affect the optimal policies of $M/M/s$ make-to-stock queues with fixed cost. The most important parameters for such a system are the fixed cost (K) and number of available servers (s). First of all, the impacts of the fixed cost and number of servers on optimal production and rationing decisions are analyzed, and then the impacts of the other parameters are shown in the following sections.

Table 3.7 The Impact of the Fixed Cost on Optimal Decisions

Optimal Production Decisions																									
Inventory Level x	Fixed Cost $K = 1$					Fixed Cost $K = 2$					Fixed Cost $K = 3$					Fixed Cost $K = 4$					Fixed Cost $K = 5$				
	Number of Active Servers y					Number of Active Servers y					Number of Active Servers y					Number of Active Servers y					Number of Active Servers y				
	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	4	3	3	3	3	4
1	3	3	3	3	4	3	3	3	3	4	3	3	3	3	4	3	3	3	3	4	3	3	3	3	4
2	2	2	2	3	4	2	2	2	3	4	2	2	2	3	4	2	2	2	3	4	2	2	2	3	4
3	1	1	2	3	4	2	2	2	3	4	2	2	2	3	4	2	2	2	3	4	2	2	2	3	4
4	0	1	2	3	4	0	1	2	3	4	1	1	2	3	4	1	1	2	3	4	1	1	2	3	4
5	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
Average Cost	5.021					5.280					5.481					5.602					5.636				
Optimal Rationing Decisions																									
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	1	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

In Table 3.7, when there is a small fixed cost to activate a new channel, the system operates with more channels than when there is a larger fixed cost to activate a new one. Because, the system chooses to pay the lost sales cost instead of paying the fixed cost. As K increases, the production continues not to give setup cost more than one, even if inventory level is higher. For example, when $K = 4$ and $(x, y) = (4,0)$ the optimal production decision is to continue to production with one active server. However, when $K = 1$ and $(x, y) = (4,0)$ the optimal decision equals to zero. Also, the average cost increases as K increases.

The impact of the fixed cost on optimal rationing policy is not very significant for this example. Because, the demand rate of class 2 is very low. If the demand rate increases, the effect of the fixed cost can be observed. For this example, the rationing decision for class 2 demand is expressed as u_r such that $u_r \in \{0,1\}$. If $u_r = 1$, class 2 demand is satisfied; else, it is rejected. When $K = 1$ and $(x, y) = (2,3)$ the optimal rationing decision is to satisfy the arriving class 2 demand. However, when $K > 1$ and $(x, y) = (2,3)$ the decision is not to reject the arriving class 2 demand. When there is a high fixed cost to activate a new channel, the system satisfies the demand in higher inventory position.

Table 3.8 The Impact of Number of Servers on Optimal Decisions

Optimal Production Decisions																														
Inventory Level x	Number of Servers $s = 3$				Number of Servers $s = 4$					Number of Servers $s = 5$						Number of Servers $s = 6$							Number of Servers $s = 7$							
	Number of Active Servers y				Number of Active Servers y					Number of Active Servers y						Number of Active Servers y							Number of Active Servers y							
	0	1	2	3	0	1	2	3	4	0	1	2	3	4	5	0	1	2	3	4	5	6	0	1	2	3	4	5	6	7
0	3	3	3	3	4	4	4	4	4	4	4	4	4	4	5	4	4	4	4	4	5	6	4	4	4	4	4	5	6	7
1	3	3	3	3	3	3	3	3	4	3	3	3	3	4	5	3	3	3	3	4	5	6	3	3	3	3	4	5	6	7
2	2	2	2	3	2	2	2	3	4	2	2	2	3	4	5	2	2	2	3	4	5	6	2	2	2	3	4	5	6	7
3	2	2	2	3	2	2	2	3	4	2	2	2	3	4	5	2	2	2	3	4	5	6	2	2	2	3	4	5	6	7
4	1	1	2	3	0	1	2	3	4	0	1	2	3	4	5	0	1	2	3	4	5	6	0	1	2	3	4	5	6	7
5	0	1	2	3	0	1	2	3	4	0	1	2	3	4	5	0	1	2	3	4	5	6	0	1	2	3	4	5	6	7
Average Cost	5.503				5.280					5.280						5.280							5.280							
Optimal Rationing Decisions																														
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	1	1	1	0	0	0	0	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 3.8 shows the effect of number of servers on optimal decisions. If available processing channels are scarce $(s = 3)$, the optimal production policy tries to use all of the limited capacity. When $s > 4$, the system is equivalent to a system with uncapacitated replenishment channel, because it always activates less than the available s channels and therefore additional channels do not provide further gain. Also, the average costs do not change because there is no difference in production decisions after $s = 4$.

If available processing channels are scarce ($s = 3$), the system satisfies an arriving class 2 demand in high inventory levels. As number of available servers increases, the system also reserves the inventory for class 2 demand. Because, if more inventory is needed, it can be replenished in a short time by using all limited capacity. Like production decisions after $s = 4$ there is no difference in rationing decisions.

Table 3.9 The Impact of Holding Cost on Optimal Decisions

Optimal Production Decisions															
Inventory Level x	Holding Cost $h = 1$					Holding Cost $h = 2$					Holding Cost $h = 3$				
	Number of Active Servers y					Number of Active Servers y					Number of Active Servers y				
	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
0	4	4	4	4	4	3	3	3	3	4	3	3	3	3	4
1	3	3	3	3	4	2	2	2	3	4	1	1	2	3	4
2	2	2	2	3	4	1	1	2	3	4	0	1	2	3	4
3	2	2	2	3	4	0	1	2	3	4	0	1	2	3	4
4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
5	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
Average Cost	5.278					7.500					9.076				
Optimal Rationing Decisions															
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
2	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

In Table 3.9, as the holding cost increases, the system operates with fewer channels. Because, the system does not prefer to stock not to give extra holding cost. This decision should be made by evaluating the balance of the holding and lost sales cost. While the holding cost increases, the average cost increases unsurprisingly.

When $h = 1$, an arriving demand class 2 is satisfied if the inventory level equals 3 or $(x, y) = (2,4)$. As the holding cost increases, the threshold level decreases and the system starts to satisfy an arriving class 2 demand in lower inventory levels.

Table 3.10 The Impact of Production Rate on Optimal Decisions

Optimal Production Decisions															
Inventory Level x	Production Rate $\mu = 1$					Production Rate $\mu = 2$					Production Rate $\mu = 3$				
	Number of Active Servers y					Number of Active Servers y					Number of Active Servers y				
	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
0	4	4	4	4	4	2	2	2	3	4	2	2	2	3	4
1	3	3	3	3	4	2	2	2	3	4	1	1	2	3	4
2	2	2	2	3	4	1	1	2	3	4	1	1	2	3	4
3	2	2	2	3	4	1	1	2	3	4	0	1	2	3	4
4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
5	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
Average Cost	5.280					5.044					4.837				
Optimal Rationing Decisions															
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1
2	0	0	0	0	1	0	0	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 3.10 shows the effect of increasing the production rate on optimal decisions and the average cost. As the production rate increases, the system does not activate all of the limited capacity. If the production rate is high, the processing time takes a short time. Hence, the system activates fewer channels. For example, when the production rate is small ($\mu = 1$), the system is willing to operate with all available channels. However when $\mu = 3$, the system does not need to activate all channels. Thus, the system activates more channels when the production rate is low otherwise; the system activates fewer channels when the production rate is high. To activate more channels when there is high production rate causes to increasing the average cost because of the fixed and holding costs.

When the production rate is low, an arriving class 2 demand is satisfied in higher inventory levels. The system wants to reserve the inventory in anticipation of future demand from class 1. However, the production rate is high ($\mu = 3$) the system starts to satisfy class 2 demand is fewer inventory levels. Because of low production time the production is completed in a short time and arriving class 1 demand can be also satisfied in a short time.

Table 3.11 The Impact of Demand Rates on Optimal Decisions

Optimal Production Decisions																									
Inventory Level x	Demand rate $\lambda_1 = 3, \lambda_2 = 1$					Demand rate $\lambda_1 = 2.5, \lambda_2 = 1.5$					Demand rate $\lambda_1 = 2, \lambda_2 = 2$					Demand rate $\lambda_1 = 1.5, \lambda_2 = 2.5$					Demand rate $\lambda_1 = 1, \lambda_2 = 3$				
	Number of Active Servers y					Number of Active Servers y					Number of Active Servers y					Number of Active Servers y					Number of Active Servers y				
	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
0	4	4	4	4	4	4	4	4	4	4	3	3	3	3	4	3	3	3	3	4	3	3	3	3	4
1	3	3	3	3	4	3	3	3	3	4	3	3	3	3	4	2	2	2	3	4	2	2	2	3	4
2	2	2	2	3	4	2	2	2	3	4	2	2	2	3	4	2	2	2	3	4	1	1	2	3	4
3	2	2	2	3	4	1	1	2	3	4	1	1	2	3	4	1	1	2	3	4	0	1	2	3	4
4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
5	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
Average Cost	5.280					5.007					4.680					4.359					4.112				
Optimal Rationing Decisions																									
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
2	0	0	0	0	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

In Table 3.11, when class 1 demand rate is higher than class 2 demand rate, the system tries to produce with more servers. Because, class 1 is the most valuable demand class and its demand occur in a short time. Hence, the system desires to satisfy class 1 demand by activating more production channels ($\lambda_1 = 3, \lambda_2 = 1$). If the system activates fewer channels, it may pay the lost sales cost. As class1 demand rate decreases and class 2 demand rate increases, the balance changes. Because, the amount of class 1 demand is less than the amount of class 2 demand in a unit time ($\lambda_1 = 1, \lambda_2 = 3$). Hence, the system does not activate all of the limited capacity because of a low demand of the most valuable class.

When the class 1 demand rate is high, class 2 demand is satisfied while increasing of inventory level and number of active channels. As the class 2 demand rate exceeds the demand rate of class 1, the rationing level decreases. The system starts to satisfy less valuable demand class in low inventory levels.

Table 3.12 The Impact of Lost Sales Costs on Optimal Decisions

Optimal Production Decisions															
Inventory Level x	Lost sales cost $c_1 = 3, c_2 = 1$ Number of Active Servers y					Lost sales cost $c_1 = 4, c_2 = 1$ Number of Active Servers y					Lost sales cost $c_1 = 5, c_2 = 1$ Number of Active Servers y				
	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
0	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4
1	3	3	3	3	4	3	3	3	3	4	3	3	3	3	4
2	3	3	3	3	4	2	2	2	3	4	3	3	3	3	4
3	1	1	2	3	4	2	2	2	3	4	2	2	2	3	4
4	0	1	2	3	4	0	1	2	3	4	1	1	2	3	4
5	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
Average Cost	4.862					5.280					5.664				
Optimal Rationing Decisions															
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	1	1	1	1	0	0	0	0	1	0	0	0	0	0
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 3.12 exhibits the effect of lost sales cost of class 1 on optimal production and rationing decisions. If the lost sales cost of class 1 is small, the optimal production decisions are less than the others. Even if the system cannot satisfy the demand of class 1, the lost sales cost will be paid is small ($c_1 = 3, c_2 = 1$). However, as the lost sales cost increases, the system starts to use all of the available servers in order to increase average inventory level and not to pay high lost sales cost per item ($c_1 = 5, c_2 = 1$). As the lost sales cost increases, the average cost increases as well.

While the lost sales cost of class 1 is close to the lost sales cost of class 2, the system satisfies an arriving class 2 demand in low inventory levels but as the difference of the lost sales cost increases, the system starts to reserve inventory for more valuable demand class in order not to pay the lost sales cost of class 1.

Based on above numerical observations the detailed characterization of optimal production and rationing polices can be provided with the following conjectures. For a given state (x, y), if the number of active channels is less than or equal to the optimal number of active channels at state $(x, 0)$, the production decision equals to

$u_p^*(x,0)$; else the production decision is not to change the number of active channels (i).

i. $u_p^*(x,y) = \begin{cases} u_p^*(x,0) & , y \le u_p^*(x,0) \\ y & , y > u_p^*(x,0) \end{cases}$

The optimal number of active channels at state $(x,0)$ decreases by one or more units for each unit increase in the on-hand inventory level and then it remains constant at zero. It is non-increasing in inventory on hand (*ii*). There are different order-up-to levels for each inventory level (*iii*). The optimal number of production channels is non-increasing in the inventory level (*iv*).

ii. $u_p^*(x,0) - u_p^*(x+1{,}0) \ge 1$.

iii. $S_x = x + u_p^*(x,0), \forall x.$

iv. $u_p^*(x,y) \ge u_p^*(x+k,y), \forall x, y, k \ge 1.$

The effects of the system parameters which are fixed cost (K), holding cost (h), production rate (μ) and traffic rates (λ_1, λ_2) on optimal production and rationing policies are shown as the following. $u_p^*(x,y)$ and $u_{r_j}^*(x,y)$ are expressed as optimal production and rationing decisions, respectively (*v*, *vi*, *vii*, *viii*).

v. $u_{p,K_1}^*(x,y) \ge u_{p,K_2}^*(x,y)$ and $u_{r_j,K_1}^*(x) \ge u_{r_j,K_2}^*(x)$

$, \forall x$ and $y,$ $K_1 \le K_2,$ $j = 1{,}2.$

vi. $u_{p,h_1}^*(x,y) \ge u_{p,h_2}^*(x,y)$ and $u_{r_j,h_1}^*(x) \le u_{r_j,h_2}^*(x)$

$, \forall x$ and $y,$ $h_1 \le h_2,$ $j = 1{,}2.$

vii. $u_{p,\mu_1}^*(x,y) \ge u_{p,\mu_2}^*(x,y)$ and $u_{r_j,\mu_1}^*(x) \le u_{r_j,\mu_2}^*(x)$

$, \forall x$ and $y,$ $\mu_1 \le \mu_2,$ $j = 1{,}2.$

viii. $u_{p,\lambda_1}^*(x,y) \le u_{p,\lambda_2}^*(x,y)$ $, \forall x$ and $y, \lambda_1 \le \lambda_2.$

ix. There exists a threshold inventory level $R_x^i(y)$ for demand class $i \ge 2$, which is a function of active production channels y. If the inventory level is more than $R_x^i(y)$, it is optimal to satisfy; else it should be rejected. $R_x^N(y) \geq R_x^{N-1}(y) \geq \cdots \geq R_x^2(y) \geq 0$ and $R_x^i(y+1) \leq R_x^i(y)$ for $i \in \{2, \ldots, N\}$.

x. There exists a threshold number of active channels $R_y^i(x)$ for demand class $i \geq 2$, which is a function inventory level x. If the number of active production channels is more than $R_y^i(x)$, it is optimal to satisfy; else it should be rejected. $R_y^N(x) \geq R_y^{N-1}(x) \geq \cdots \geq R_y^2(x) \geq 0$ and $R_y^i(x+1) \leq R_y^i(x)$ for $i \in \{2, \ldots, N\}$.

4 THE PROPOSED PRODUCTION AND RATIONING POLICIES

According to the numerical observations of the optimal policies, optimal policies are highly dynamic and do not have systematic behaviour for practical purposes. It is not so easy to be applied in the firms. Hence, proposing new/alternative policies whose performances are close to performances of the optimal policy and have a general behaviour to be applied can be beneficial. In this section, the characteristics and behaviours of the alternative production and rationing policies are deeply analyzed.

The rest of this section is organized as follows. In Section 4.1, we introduce three different production policies and discuss the properties of these policies. The Sub-section 4.1.2.2 includes Continuous Time Markov Chain analysis to get the steady-state probabilities of the production policy 2. In Sub-section 4.1.3.3, a renewal analysis utilizing renewal reward theorem is conducted to calculate the average cost for production policy 3 when s = 1, which is the optimal policy for single-server case. Also, we compare the performances of the proposed policies with optimal and Base-stock policies in the literature.

4.1 The Proposed Production Policies

The performances of the proposed policies can be compared with the Base-stock and optimal policies. Base-stock policy has only one critical level which is called as S. When inventory level drops to below S, the production is triggered and starts. In other words, the system turns to $(S, S-1)$. When inventory level hits to $S-1$, the production starts and continues until inventory level reaches to S. Base-stock policy is the optimal production policy for $M/M/1$ make-to-stock queues with no fixed cost.

The optimal production policy for $M/M/1$ make-to-stock queues with fixed cost is (X^*, X^{**}). Base-stock policy is a special case of (X^*, X^{**}) policy. If $X^{**} = X^* = S$, the production policy turns to Base-stock policy. However, when there is a fixed cost to activate a new channel, two-critical number policy is better than Base-stock. Because, the system wants to activate a new channel when the inventory is fewer and it wants to extend the production time with this activated channel. The decision function is shown in the following:

$$u(X,0) = \begin{cases} 1 & for\ X < X^* \\ 0 & for\ X \geq X^* \end{cases}$$

$$u(X,1) = \begin{cases} 1 & for\ X < X^{**} \\ 0 & for\ X \geq X^{**} \end{cases}$$

In this policy, if inventory level is less than the triggered point X^*, the production starts and continues to produce until inventory level hits to X^{**}. Since our model includes the fixed cost to activate a new channel, our policies are similar to the two-critical number policy but, our problem is analyzed on a system with s parallel production channels.

4.1.1 The Proposed Production Policy 1

The first proposed production policy has two inventory levels. These are X^* and X^{**} are called as trigger point and order-up-to level, respectively. Our policy has some differences with two-critical number policy mentioned in the literature. Our model includes s parallel production channels and checks the inventory position i.e. the information of inventory level and number of active servers. The system decides to or not to activate a server or servers by looking how many items is replenished in the inventory soon. The optimal decision at state (x, y) can be expressed as:

$$u(x,y) = \begin{cases} min(\boldsymbol{X}^{**} - x, s) & (x+y) < X^* \\ y & X^* \leq (x+y) < X^{**} \\ y & (x+y) \geq X^{**} \end{cases}$$

If the inventory position is less than X^*, the system tries to reach inventory level X^{**}. The function $min(X^{**} - x, s)$ serves to this purpose. If the inventory position is more than X^*, the production is not triggered and the number of active channels is not changed. At the epoch production completion occurs,

$$Next\ state = \begin{cases} (x+1, \min(\boldsymbol{X}^{**} - x - 1, s)) & (x+y) < X^* \\ (x+1, y) & X^* \leq (x+y) < X^{**} \\ (x+1, y-1) & (x+y) \geq X^{**} \end{cases}$$

The inventory level is increased one unit, the number of active servers can be changed according to the new inventory position. If the new inventory position hits to X^{**}, a server is turned off after production is completed. If a demand occurs, next states are shown in the following:

$$Next\ state = \begin{cases} (x-1, \min(\boldsymbol{X}^{**} - (x-1), s)) & (x+y) < X^* \\ (x-1, y) & X^* \leq (x+y) < X^{**} \\ (x-1, y-1) & (x+y) \geq X^{**} \end{cases}$$

4.1.1.1 The Behaviour of Proposed Production Policy 1

This section includes the analysis of how the most important system parameters affect the proposed policy (PPP1). As mentioned in the previous sections, the most important parameters are fixed cost (K) and number of available servers(s). For other system parameters, behaviour of the proposed production policy 1 behaves like the optimal policy.

Table 4.1 shows the effect of number of servers on PPP1. If available processing channels are scarce $(s = 3)$, PPP1 and the optimal policy activate all of the limited capacity. When $s > 4$, the system turns to an uncapacitated system. It should activate less than the available s channels because of no further gain. Also, the average costs of optimal policy do not change because there is no difference in production decisions after $s = 4$. As the optimal policy does not activate additional channels, PPP1 uses almost all available servers after $s = 4$. It also causes to increase the average cost. As activating more servers, the system has to pay the fixed cost per each server.

Table 4.1 The Impact of Fixed Cost on Optimal Production and PP1 Policies

Optimal Production Decisions																														
Inventory Level x	Number of Servers $s = 3$				Number of Servers $s = 4$					Number of Servers $s = 5$						Number of Servers $s = 6$							Number of Servers $s = 7$							
	Number of Active Servers y				Number of Active Servers y					Number of Active Servers y						Number of Active Servers y							Number of Active Servers y							
	0	1	2	3	0	1	2	3	4	0	1	2	3	4	5	0	1	2	3	4	5	6	0	1	2	3	4	5	6	7
0	3	3	3	3	4	4	4	4	4	4	4	4	4	4	5	4	4	4	4	4	5	6	4	4	4	4	4	5	6	7
1	2	2	2	3	2	2	2	3	4	2	2	2	3	4	5	2	2	2	3	4	5	6	2	2	2	3	4	5	6	7
2	1	1	2	3	1	1	2	3	4	1	1	2	3	4	5	1	1	2	3	4	5	6	1	1	2	3	4	5	6	7
3	0	1	2	3	0	1	2	3	4	0	1	2	3	4	5	0	1	2	3	4	5	6	0	1	2	3	4	5	6	7
Average Cost	12.743				**12.699**					**12.699**						**12.699**							**12.699**							
The Proposed Production Policy 1																														
0	3	3	3	3	4	4	4	3	4	5	5	5	3	4	5	5	5	5	3	4	5	-	5	5	5	3	4	5	-	-
1	3	3	2	3	4	4	2	3	4	5	5	2	3	4	5	4	4	2	3	4	-	-	4	4	2	3	4	-	-	-
2	3	1	2	3	4	1	2	3	4	4	1	2	3	4	-	3	1	2	3	-	-	-	3	1	2	3	-	-	-	-
3	0	1	2	3	0	1	2	3	-	0	1	2	3	-	-	0	1	2	-	-	-	-	0	1	2	-	-	-	-	-
X^*	3				3					3						3							3							
X^{**}	6				6					6						5							5							
Average Cost	12.864				**13.023**					**13.305**						**13.318**							**13.318**							

PPP1 decisions differ from the optimal policy decisions. As fixed cost and number of available servers increase, the cost difference between optimal and PPP1 increases as well. We can conclude that there are some drawbacks of proposed production policy. These are:

- For small values of s and K, there is no cost difference. As s and K increases, the percentage is getting higher.
- If X^{**} is much higher than s, the policy decides to activate almost at all available servers. (If $X^{**} - x > s$, the decision becomes number of available servers (s)).

4.1.2 The Proposed Production Policy 2

In order to avoid the undesirable behaviours, a new production policy is needed to be proposed (PPP2). Proposed Production Policy 2 (PPP2) has two control parameters as PPP1. We can define them as X^* and X^{**}. However, the production decision is made according to inventory level not inventory position as PPP1. If the inventory level is less than X^*, the system tries to the inventory level hits to X^* by using $min(X^* - x, s)$ function. This decision structure is similar to PPP1; however the primary objective is to reach X^* not X^{**} as in PPP1. If the inventory level is between X^* and X^{**}, the production is not triggered to activate additional channels. If the production starts below X^* and continues with one server until the inventory level hits to X^{**} otherwise if the inventory level hits to X^{**} and decreases with arrival demands the number of active servers should be zero. Lastly, when the inventory level reaches to X^{**}, the last server is turned off and non-production period starts.

$$u(x,y) = \begin{cases} min(X^* - x, s) & x < X^* \\ max(y-1,1) & X^* \leq x < X^{**} \\ y & x \geq X^{**} \end{cases}$$

At the epoch production completion occurs, one unit is added to the inventory and the number of active servers can be changed according to the new inventory level. If the new inventory level hits to X^{**}, the server is turned off after production completion

$$Next\ state = \begin{cases} (x+1, min(X^* - x - 1, s)) & x < X^* \\ (x+1, max(y-1,1)) & X^* \leq x < X^{**} \\ (x+1, y-1) & x \geq X^{**} \end{cases}$$

If a demand occurs, next states are shown in the following:

$$Next\ state = \begin{cases} (x-1, min(X^* - x, s)) & x < X^* \\ (x-1, max(y-1,1)) & X^* \leq x < X^{**} \\ (x-1, y-1) & x \geq X^{**} \end{cases}$$

4.1.2.1 The Behaviour of Proposed Production Policy 2

Table 4.2 depicts the decision matrixes of optimal, PP1 and PP2 policies as the fixed cost increases. When fixed cost is not incurred to activate a channel, there is no difference in decisions of policies. As the fixed cost increases, the average cost of PPP2 increases and passes the average cost of PPP1 for any given positive fixed cost. For larger values of K, the performance of PPP2 gets worse. The reason of this situation is to produce with one active server when the inventory level is between X^* and X^{**}. If many demand arrivals occur in this interval, the system tries to activate more servers. If the production completes, the server is turned off and turned on with paying fixed cost. In other words, a system which has a larger fixed cost should not apply PPP2 because of turning off the servers in small inventory levels.

Table 4.2 The Impact of Fixed Cost on Optimal, PP1 and PP2 Policies

Optimal Production Decisions																				
Inventory Level x	Fixed Cost $K = 0$					Fixed Cost $K = 1$					Fixed Cost $K = 2$					Fixed Cost $K = 3$				
	Number of Active Servers y					Number of Active Servers y					Number of Active Servers y					Number of Active Servers y				
	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	4
1	3	3	3	3	4	2	2	2	3	4	2	2	2	3	4	2	2	2	3	4
2	0	1	2	3	4	1	1	2	3	4	1	1	2	3	4	1	1	2	3	4
3	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
Average Cost	10.493					11.833					12.699					13.082				
PPP1 Decisions																				
0	4	4	4	4	4	4	4	4	3	4	4	4	4	3	4	4	4	4	3	4
1	3	3	3	3	-	4	4	2	3	4	4	4	2	3	4	4	4	2	3	4
2	2	2	2	-	-	3	1	2	3	-	4	1	2	3	4	4	1	2	3	4
3	1	1	-	-	-	0	1	2	-	-	0	1	2	3	-	0	1	2	3	-
4	0	-	-	-	-	0	1	-	-	-	0	1	2	-	-	0	1	2	-	-
5	-	-	-	-	-	0	-	-	-	-	0	1	-	-	-	0	1	-	-	-
X^*	4					3					3					3				
X^{**}	4					5					6					6				

Average Cost	10.613					12.180					13.023					13.684				
PPP2 Decisions																				
0	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	4
1	3	3	3	3	-	3	3	3	3	4	3	3	3	3	4	2	2	2	3	4
2	2	2	2	-	-	2	2	2	3	-	2	2	2	3	4	1	1	2	3	4
3	1	1	-	-	-	1	1	2	-	-	1	1	2	3	-	0	1	2	3	4
4	0	-	-	-	-	0	1	-	-	-	0	1	2	-	-	0	1	2	3	-
5	-	-	-	-	-	0	-	-	-	-	0	1	-	-	-	0	1	2	-	-
X^*	4					4					4					3				
X^{**}	4					5					6					7				
Average Cost	10.614					12.924					15.110					17.081				

The effects of number of available servers on optimal, PP1 and PP2 policies are shown in Table 4.3. We mention that PPP1 has some disadvantages in previous sections. These situations are also valid for PPP2. As number of server increases, the average costs of the policies are getting larger while the optimal policy does not change the decisions. According to these observations, we understand that we need to fine tune these policies.

Table 4.3 The Impact of Number of Servers on Optimal, PP1 and PP2 Policies

Optimal Production Decisions																		
Inventory Level x	Number of Servers $s = 2$			Number of Servers $s = 3$				Number of Servers $s = 4$					Number of Servers $s = 5$					
	Number of Active Servers y			Number of Active Servers y				Number of Active Servers y					Number of Active Servers y					
	0	1	2	0	1	2	3	0	1	2	3	4	0	1	2	3	4	5
0	2	2	2	3	3	3	3	4	4	4	4	4	4	4	4	4	4	5
1	2	2	2	2	2	2	3	2	2	2	3	4	2	2	2	3	4	5
2	1	1	2	1	1	2	3	1	1	2	3	4	1	1	2	3	4	5
3	0	1	2	0	1	2	3	0	1	2	3	4	0	1	2	3	4	5
4	0	1	2	0	1	2	3	0	1	2	3	4	0	1	2	3	4	5
Average Cost	13.900			12.743				12.699					12.699					
PPP1 Decisions																		

0	2	2	2	3	3	3	3	4	4	4	3	4	5	5	5	3	4	5
1	2	2	2	3	3	2	3	4	4	2	3	4	5	5	2	3	4	5
2	2	1	2	3	1	2	3	4	1	2	3	4	4	1	2	3	4	-
3	0	1	2	0	1	2	3	0	1	2	3	-	0	1	2	3	-	-
4	0	1	2	0	1	2	-	0	1	2	-	-	0	1	2	-	-	-
5	0	1	2	0	1	-	-	0	1	-	-	-	0	1	-	-	-	-
X^*	3			3				3					3					
X^{**}	6			6				6					6					
Average Cost	13.906			12.864				13.023					13.305					
PPP2 Decisions																		
0	2	2	2	3	3	3	3	4	4	4	4	4	4	4	4	4	4	5
1	2	2	2	3	3	3	3	3	3	3	3	4	3	3	3	3	4	5
2	2	2	2	3	3	3	3	2	2	2	3	4	2	2	2	3	4	-
3	2	2	2	2	2	2	3	1	1	2	3	-	1	1	2	3	-	-
4	1	1	2	1	1	2	-	0	1	2	-	-	0	1	2	-	-	-
5	0	1	-	0	1	-	-	0	1	-	-	-	0	1	-	-	-	-
X^*	5			5				4					4					
X^{**}	6			6				6					6					
Average Cost	14.181			14.103				15.110					15.110					

4.1.2.2 Continuous Time Markov Chain Analysis for PP2

This section exhibits a Continuous Time Markov Chain analysis to find steady state probabilities and the cost structure of the system. Each state is defined and named, balance equations are established and the probabilities are computed, respectively.

As mentioned previously, the state of the system is defined with two variables: $X(t)$ and $Y(t)$. $X(t)$ denotes the inventory level at time t and $Y(t)$ denotes the number of active channels at time t. However, we use only one variable instead of these two variables. The new system-state is defined as: $0,1,\dots,2X^{**}-X^{*}$. This variable denotes the inventory level until X^{**} state. The state transition diagram of proposed policy 1 is depicted in Figure 4.1. The below of the figure shows the production period and above of the figure shows the non-production period. The values between $X^{**}+1$ and $2X^{**}-X^{*}$ shows the non-production and the states that inventory level decreases one by one. For example, $X^{**}+1$ depicts the inventory level is $X^{**}-1$ and the number of active servers is zero. This situation equates $(X^{**}-1,0)$ in two-parameter system state. If the system-state is $2X^{**}-X^{*}$, the inventory level decreases as n units while the inventory level is X^{**}. Then, n is computed as $X^{**}+n=X^{**}+X^{**}-X^{*}=2X^{**}-X^{*}$. The transitions in the diagram are provided by λ and μ. λ_n denotes the demand rate when the inventory level is n. The system satisfies the demands from two different classes, so λ_n is either sum of λ_1 and λ_2 or just λ_1. The reason of this situation is value of the rationing level R. Hence, λ_n is equal to $\sum_{R_i \le n} \lambda_i$. R_i is the rationing level for demand class i and λ_i denotes the demand rate of class i.

P_n is determined as a steady-state probability for each state. The production system behaviour is modelled by Continuous Time Markov Chain and state balance equations are derived as given below.

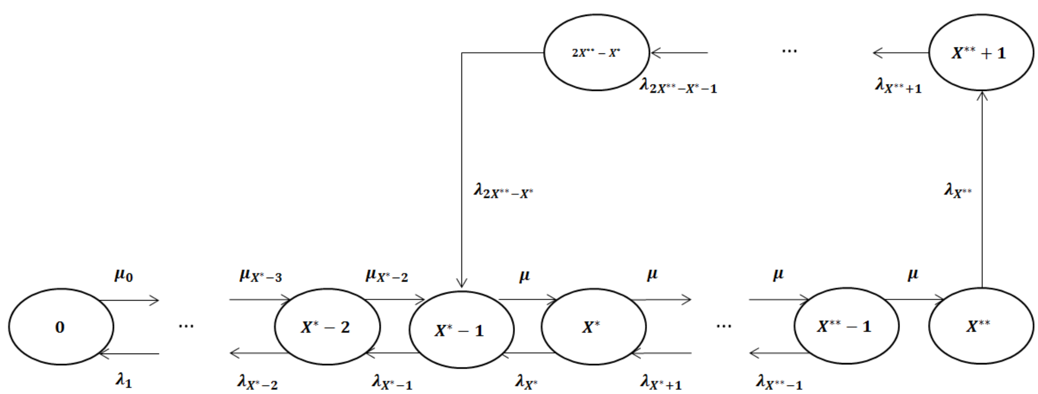

Figure 4.1 State Transition Diagram

(1) $P_0\mu_0 = P_1\lambda_1,$

(2) $P_n(\mu_n+\lambda_n) = P_{n-1}\mu_{n-1} + P_{n+1}\lambda_{n+1}, \quad n = 1,2,\dots,X^{*}-2,$

(3) $P_{X^{*}-1}(\mu_{X^{*}-1}+\lambda_{X^{*}-1}) = P_{X^{*}-2}\mu_{X^{*}-2} + P_{X^{*}}\lambda_{X^{*}} + P_{2X^{**}-X^{*}}\lambda_{2X^{**}-X^{*}},$

(4) $P_n(\mu_n+\lambda_n) = P_{n-1}\mu_{n-1} + P_{n+1}\lambda_{n+1}, \quad n = X^{*}, X^{*}+1,\dots,X^{**}-2,$

(5) $P_{X^{**}-1}(\mu_{X^{**}-1}+\lambda_{X^{**}-1}) = P_{X^{**}-2}\mu_{X^{**}-2},$

(6) $P_{X^{**}}\lambda_{X^{**}} = P_{X^{**}-1}\mu_{X^{**}-1},$

(7) $P_n\lambda_n = P_{n-1}\lambda_{n-1,}\ \ n = X^{**}+1, X^{**}+1, \dots, 2X^{**}-X^*,$

The Steady state probabilities, $P_n(n = 0, 1, \dots,\ 2X^{**}-X^*)$ are computed with using the balance equations above and the following normalization equation:

$$\sum_{n=0}^{2X^{**}-X^*} P_n = 1\,.$$

The state $(2X^{**}-X^*) = (X^*, 0)$ is set as a boundary state to solve the system. All the steady-state probabilities are computed as a linear function of the steady-state probability of $P_{2X^{**}-X^*}$ by deriving recursive equations. The steady-state probabilities are obtained by using MATLAB.

1. $P_n = \frac{(P_{2X^{**}-X^*})*(\lambda_{2X^{**}-X^*})}{\lambda_n}, n = 2X^{**}-X^*, \dots, X^{**}$.

2. $P_{X^{**}-1} = \frac{(P_{2X^{**}-X^*})*(\lambda_{2X^{**}-X^*})}{\mu_{X^{**}-1}} = (P_{2X^{**}-X^*}) * (\lambda_{2X^{**}-X^*})$.

3. $P_n = (P_{2X^{**}-X^*}) * (\lambda_{2X^{**}-X^*})[1 + \lambda_{n+1} + \lambda_{n+1}\lambda_{n+2} + \cdots + \lambda_{n+1}\lambda_{n+2} \dots \lambda_{X^{**}-1}]$,
   $n = X^{**}-2, X^{**}-3, \dots, X^*-1$.

4. $P_n = \frac{(P_{2X^{**}-X^*})*(\lambda_{2X^{**}-X^*})[\lambda_{n+1}\lambda_{n+2}+\lambda_{n+1}\lambda_{n+2}\lambda_{n+3}+\cdots+\lambda_{n+1}\lambda_{n+2}\dots\lambda_{X^{**}-1}]}{\mu_n\mu_{n+1}\cdots\mu_{X^*-2}}$,
   $n = X^*-2, X^*-3, \dots, 1, 0$.

4.1.3 The Proposed Production Policy 3

We propose two production policies in the previous sections. We mention the disadvantages of each policy. While the average costs of PPP1 is better than the average costs of PPP2, for larger values of the most important parameters (s and K) the performance of PPP1 deteriorates. Also continuing to production with one active server is not enough to satisfy the demand. Hence, we can propose a new policy which has the best aspects of the proposed policies. We use again two control variables which are X^* and X^{**}. When the inventory position is less than X^*, the system tries to reach X^*. The concept of inventory position comes from PPP1 and trying to reach X^* comes from PPP2. The optimal decision at any state (x, y),

$$u(x,y) = \begin{cases} min(\boldsymbol{X}^* - x, s) & (x+y) < X^* \\ y & X^* \leq (x+y) < X^{**} \\ y & (x+y) \geq X^{**} \end{cases}$$

If the inventory position is less than X^*(trigger point), the system tries the inventory position equate X^* value. When the inventory position is equal to trigger point, the production continues with the same number of active servers. If the inventory position hits to X^{**} (maximum stock level), each active server completes the production then is turned off. If a demand occurs and reduces the inventory position below X^*, servers are activated based on demand quantity. Between X^* and X^{**}, the production can continue with more than one server i.e. PPP2 is a special case of PPP3. At the epochs of production completion and a demand occurrence, next states can be shown respectively as:

$$Next\ state = \begin{cases} (x+1, min(\boldsymbol{X}^* - x - 1, s)) & (x+y) < X^* \\ (x+1, y) & X^* \leq (x+y) < X^{**} \\ (x+1, y-1) & (x+y) \geq X^{**} \end{cases}$$

$$Next\ state = \begin{cases} (x-1, min(\boldsymbol{X}^* - (x-1), s)) & (x+y) < X^* \\ (x-1, y) & X^* \leq (x+y) < X^{**} \\ (x-1, y-1) & (x+y) \geq X^{**} \end{cases}$$

4.1.3.1 The Behaviour of Proposed Production Policy 3

Table 4.4 shows the production decisions matrixes of optimal, PPP1, PPP2 and PPP3, respectively. The most similar proposed policy to the optimal is easily noticed that it is PPP3.

Table 4.4 Optimal and Proposed Production Policies Decisions

	Optimal Production Policy					PPP1					PPP2					PPP3				
	Number of Active Servers y					Number of Active Servers y					Number of Active Servers y					Number of Active Servers y				
Inventory Level x	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
0	**4**	4	4	4	4	**4**	4	4	3	4	**4**	4	4	4	4	**3**	3	3	3	4
1	**2**	2	2	3	4	**4**	4	2	3	4	**3**	3	3	3	4	**2**	2	2	3	4
2	**1**	1	2	3	4	**4**	1	2	3	4	**2**	2	2	3	4	**1**	1	2	3	-
3	0	1	2	3	4	0	1	2	3	-	**1**	1	2	3	-	0	1	2	-	-
X^*	-					3					4					3				
X^{**}	-					6					6					5				

Table 4.5 shows the effect of number of the active server for production decision and compares optimal and proposed policies decisions. As can be easily observed from the table, production decisions are similar and there is a low cost difference between optimal and proposed policy 3.

Table 4.5 Optimal and Proposed Policy 3 Production Decisions for larger s

$(K, s, h, \mu, \lambda_1, \lambda_2, c_1, c_2) = (2, 10, 1, 1, 3, 1, 4, 1)$												
Inventory Level x	Number of Active Servers y											Average Cost
	0	1	2	3	4	5	6	7	8	9	10	
0	4	4	4	4	4	5	6	7	8	9	10	5.2835
1	3	3	3	3	4	5	6	7	8	9	10	
2	2	2	2	3	4	5	6	7	8	9	10	
3	2	2	2	3	4	5	6	7	8	9	10	
4	0	1	2	3	4	5	6	7	8	9	10	
$(K, s, h, \mu, \lambda_1, \lambda_2, c_1, c_2, X^*, X^{**}) = (2, 10, 1, 1, 3, 1, 4, 1,4,9)$												
Inventory Level x	Number of Active Servers y											Average Cost
	0	1	2	3	4	5	6	7	8	9	10	
0	4	4	4	4	4	5	6	7	8	9	10	5.3073
1	3	3	3	3	4	5	6	7	8	9	10	
2	2	2	2	3	4	5	6	7	8	9	10	
3	1	1	2	3	4	5	6	7	8	9	10	
4	0	1	2	3	4	5	6	7	8	9	10	

Figure 4.2 shows the percentage of cost differences of the proposed policies based on optimal policy. The cost difference of Base-stock policy is increasing while K increases. Base-stock policy does not provide the continuation in the production. It is not appropriate policy for the systems include fixed cost for production. The best percentage of cost difference is shown in PPP3.

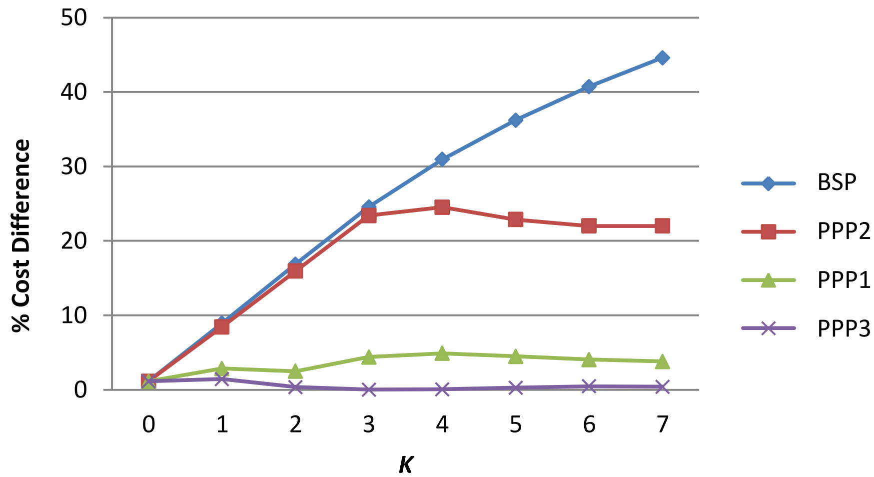

Figure 4.2 The Cost Differences of Optimal and Proposed Policies as *K* increases

Figure 4.3 shows the percentage of cost differences of the proposed policies based on optimal policy as number of available servers increases. It is obviously noticed that PPP3 has the best results among the other proposed policies. The performance of PPP3 is close to performance of the optimal production policy and has a general behaviour.

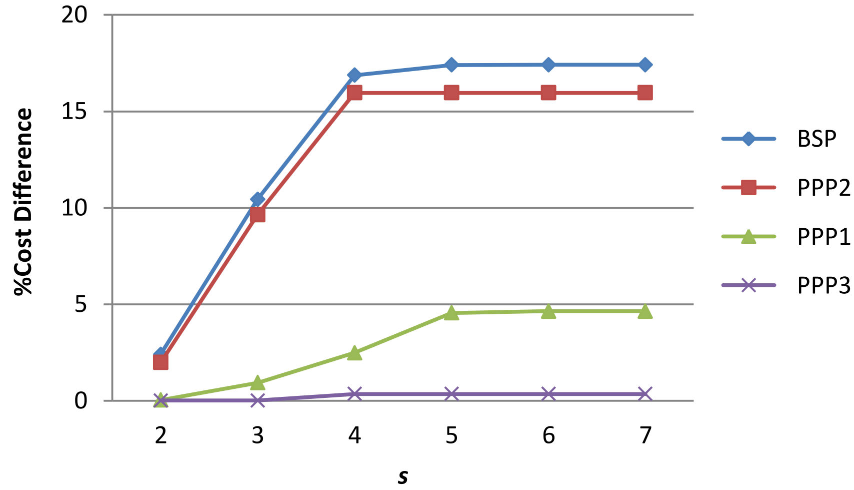

Figure 4.3 The Cost Differences of Optimal and Proposed Policies as *s* increases

Table 4.6 summarizes the proposed policies and shows the differences between them in terms of their structures. When inventory position drops to below X^*, the production is triggered and continued until inventory position hits to X^{**}. The Continuation column denotes the number of active servers as production is continued.

Table 4.6 Comparison of Production Policies in terms of their structures

Policies	Trigger Level	Order-up-to Level	Continuation	Maximum Stock Level
PP1	X^*	X^{**}	with more than or equal to one server	X^{**}
PP2	X^*	X^{**}	with one server	X^{**}
PP3	X^*	X^*	with more than or equal to one server	X^{**}

4.1.3.2 Continuous Time Markov Chain Analysis

It is obvious that the performance of PPP3 is closer to the performance of optimal production policy than the performances of the other proposed policies. Hence, CTMC analysis is conducted to obtain steady-state probabilities. The number of states and transition rates in PPP3 differ from the number of states and transition rates in PPP2. We mentioned that the production continues with one server when the inventory level is between X^* and X^{**} in PPP2. However, the number of active servers can differentiate when the inventory level is between this two critical variables. We foresee the diagrams and the process of analysis becoming complicated.

The primarily analyses are carried out for the largest and smallest X^* values. Figure 4.4 illustrates the state transition diagram when X^* equals to X^{**} (i.e. the largest value X^* can take). When $X^* = X^{**}$, the process converts to Birth-death process and the production policy behaves like Base-stock policy. For such a system, the steady-state probability of each state can be obtained via the recursive equations. Then, the average cost function is obtained by using the steady-state probabilities.

As mentioned in CTMC of PPP2, $\lambda(x,y)$ values are related to the rationing levels of demand classes. $\lambda(x,y)$ can be equal to either $\lambda_1 + \lambda_2$ or λ_1. However, this situation does not change the structure of Birth-death process but can change the decisions.

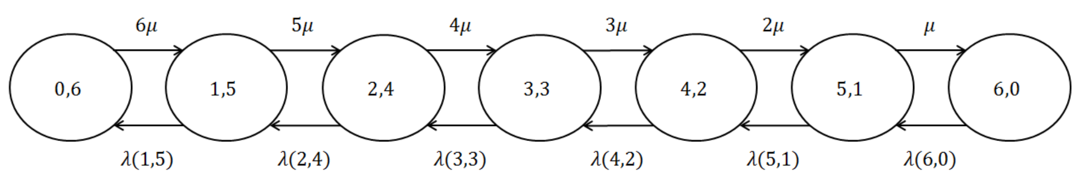

Figure 4.4 X^*=X^{**}=6 State Transition Diagram

The state transition diagram when X^* equals to 1 (i.e. the smallest value X^* can take) is depicted in Figure 4.5. When $X^* = 1$, the process has an analyzable structure and the diagram is not so complicated. Thus, the steady-state probabilities and the average cost can be obtained by calculating the recursive equations.

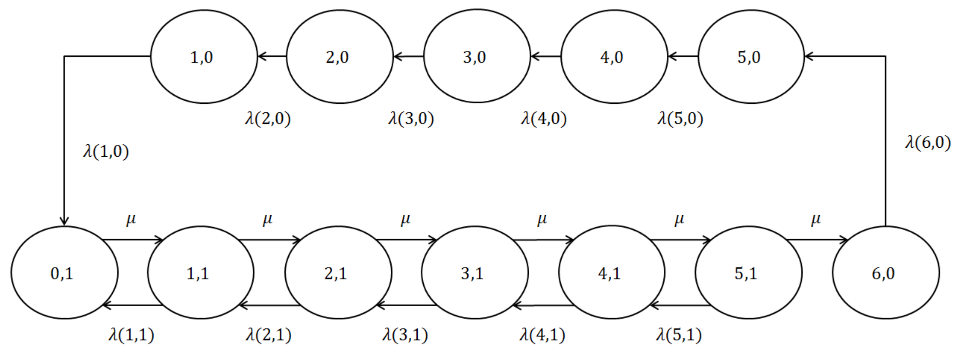

Figure 4.5 X^*=1, X^{**}=6 State Transition Diagram

If X^* does not equal to the minimum and maximum value, the diagram becomes complicated and can be depicted in Figure 4.6. This example is enough to sense the characteristics of the policy. The diagram is getting complex when X^* takes the moderate values. It is obvious that calculating the steady-state probabilities is very difficult and time-consuming for such a system. Actually, our objective is to find the average cost and we can obtain the expression of the average cost by not calculating the steady-state probabilities.

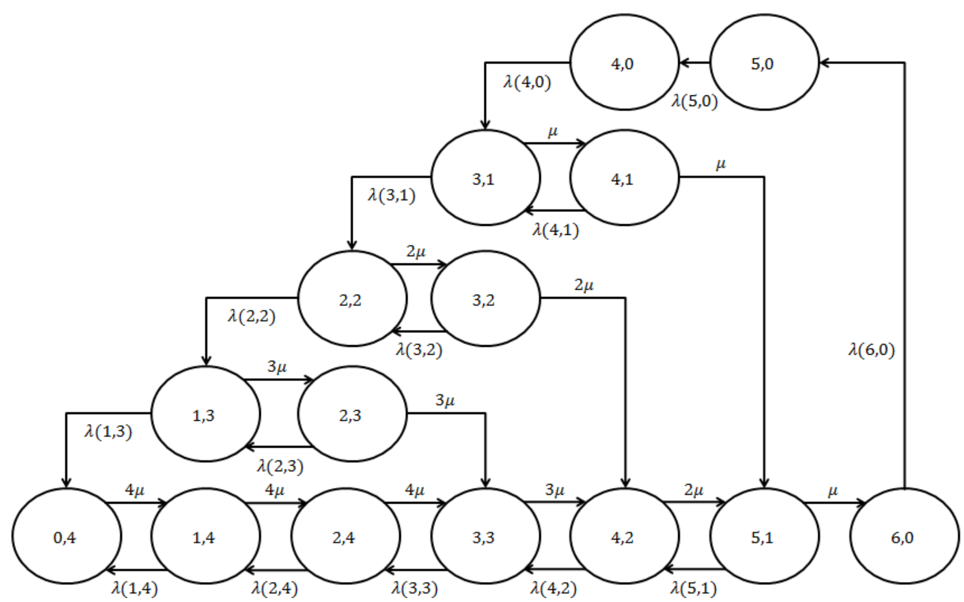

Figure 4.6 X^*=4, X^{**}=6 State Transition Diagram

As mentioned in the literature, Lee and Srinivasan (1989, 1991) calculate the average cost function by not using the steady-state probabilities. This method is called *Renewal Reward Theorem*. In order to use the theorem, we need a regeneration point which decomposes the complex system into two sub-systems.

Figure 4.7 illustrates the theorem in the (s, S) policy typical inventory system. The regeneration point is defined as every time inventory level reaches to S. As soon as the inventory level hits to specified S value, the production is turned off and non-production period starts. During a non-production period, the inventory level decreases because of demand arrivals. When the inventory level drops to s, the non-production period ends and the production period starts instantaneously. During a production period, the inventory level increases and sometimes decreases because of demand occurrences. When the inventory level reaches to S again, the production is completed and non-production period begins. Thus, it initiates a new cycle for such an inventory system.

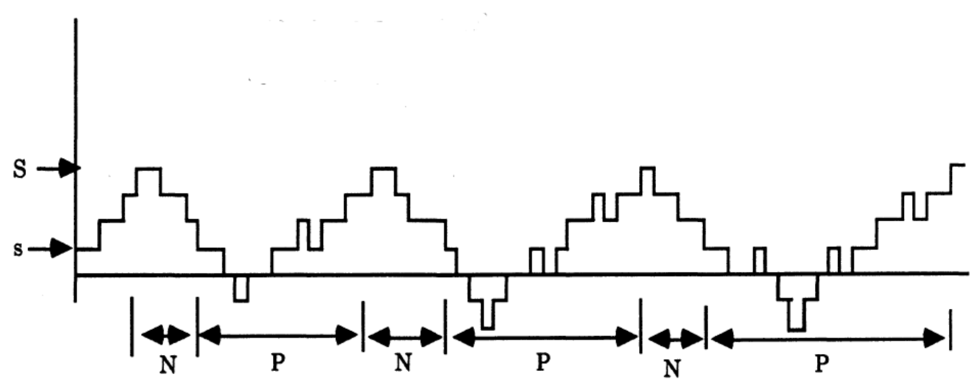

Figure 4.7 (*S*,*s*) Policy

In the following section, we find an expression for the average cost per unit time for $M/M/1$ make-to-stock queues with lost sales when the critical values are X^* and X^{**}.

4.1.3.3 Renewal Analysis for $M/M/1$

The control values are set to X^* for lower control value and X^{**} for upper control value. Notations are defined as:

K	=	Fixed cost,
h	=	holding cost,
c	=	lost sales cost,
U	=	the processing time to produce a product,
D	=	the number of demands which arrive during a processing time U,
d_j	=	the probability of j demands occur during a processing time i.e. $P(D = j) = d_j$, $j = 0,1,2, \ldots \infty$,
$C_N(X^*, X^{**})$	=	expected cost during a non-production period,
$C_P(X^*, X^{**})$	=	expected cost during a production period,
$L_N(X^*, X^{**})$	=	expected length of a non-production period,
$L_P(X^*, X^{**})$	=	expected length of a production period,
$AC(X^*, X^{**})$	=	average cost per unit time.

The purpose of this theorem is to find the expected cost per unit time when the control values are X^* and X^{**}. The regeneration point is defined as every time inventory level reaches to X^{**}. The average cost per unit time is obtained by

$$AC(X^*, X^{**}) = \frac{C_N(X^*, X^{**}) + C_P(X^*, X^{**}) + K}{L_N(X^*, X^{**}) + L_P(X^*, X^{**})} \tag{3}$$

By the Renewal Reward Theorem, the long-run average cost equals to the expected cost over a cycle divided by the expected cycle length (Equation 3). In order to find the average cost, we need to find $C_N(X^*, X^{**})$, $C_P(X^*, X^{**})$, $L_N(X^*, X^{**})$ and $L_P(X^*, X^{**})$ terms. First of all, we determine the expected cost during a non-production period. Let $g_{x,x-1}$ denotes the expected cost during a non-production period when the inventory level drops to $x-1$ from x (Equation (4)). Then, the total expected cost during a non-production period is calculated by the Equation (5).

$$g_{x,x-1} = \frac{hx}{\lambda} \tag{4}$$

$$C_N(X^*, X^{**}) = \sum_{x=X^*+1}^{X^{**}} g_{x,x-1} \tag{5}$$

In order to calculate the total expected cost during a production period, we need to determine the expected first passage cost during a production period when the inventory level increases to $x+1$ from x ($f_{x,x+1}$) and also $f_{x,x}=0$ for any x. The total expected cost during a production period is calculated by the Equation (6).

$$C_P(X^*, X^{**}) = \sum_{x=X^*}^{X^{**}-1} f_{x,x+1} \tag{6}$$

In Equation (7), E_x is defined as the expected cost during a processing time when the processing starts with x units in inventory. During a processing time, j demands occur with probability d_j. Number of demands can take any integer value between 0 and ∞. Hence, j demands take the inventory level to $\max(x-j,0)$ before production completion. The reason of using maximum function is the inventory level cannot take negative values because of lost sales environment. When the production completes, the outstanding order is added to the maximum function $\max(x-j,0)+1$.

$$f_{x,x+1} = E_x + \sum_{j=0}^{\infty} d_j f_{\max(x-j,0)+1,x+1} \tag{7}$$

Equation (7) has two unknown values. One of them is E_x and the other is d_j. The computation of d_j which is the probability of j demands occur during a processing time is expressed in the following equations:

$$d_j = \int_{u=0}^{\infty} P(N(u) = j \mid U = u)\, f_U(u) du \tag{8}$$

$$d_j = \int_{u=0}^{\infty} \frac{e^{-\lambda u}(\lambda u)^j}{j!} \mu e^{-\mu u} du \tag{9}$$

To obtain E_x, we use the expected arrival time of x^{th} demand during a processing time $\tau_x = E[T_x]$. If the number of products demanded during a processing time is less than x, x^{th} item is held during the whole processing time, otherwise is removed from inventory. The second term of Equation (10) corresponds to lost sales interval. After the arrival of x^{th} demand, all products are demanded and there is no product in inventory. Lost sales cost should be paid as the expected number of demand during $U - T_x$, i.e. $E[N(U - T_x)]$**.**

$$E_x = \sum_{i=1}^{x} h\tau_i + c\,\lambda(E(U) - \tau_x) \tag{10}$$

Let U_j be the length of the processing time given that j demands occurred within the processing time. Given j demand arrivals during (U_j), the joint distribution of these arrivals have the same distribution as the order statistics of j independent random variables uniformly distributed on $[0, U_j]$. If the number of arrival demands during a processing time is less than i, is denoted as j, i^{th} product is held in the inventory during whole processing time and the expected holding time of i^{th} product equals to u_j, otherwise i^{th} product is removed from inventory and the expected holding time of i^{th} product expressed as $\frac{i}{j+1} u_j$.

$$\tau_i = \sum_{j=0}^{i-1} d_j u_j + \sum_{j=i}^{\infty} d_j (\frac{i}{j+1} u_j) \tag{11}$$

$$u_j = E[U_j] = E[U \mid D = j]$$

$$= \int_u f_{U|D=j}(u|D=j)udu$$

$$= \int_u \frac{f_{U,D=j}(u, D=j)}{P(D=j)}udu$$

$$= \frac{1}{d_j}\int_u P(D=j \mid U=u)f_U(u)udu$$

$$u_j = \frac{1}{d_j}\int_u \frac{e^{-\lambda u}(\lambda u)^j}{j!}\mu e^{-\mu u}udu \tag{12}$$

We compute recursive expressions for variables to expedite our algorithm in MATLAB.

$$E_{x+1} = E_x + h\tau_{x+1} - c\left(\sum_{i=x+1}^{\infty} d_j\right) \tag{13}$$

$$\tau_{x+1} = \tau_x + \frac{1}{\lambda}\left(1 - \sum_{j=0}^{x} d_j\right) = \tau_x + \frac{1}{\lambda}\sum_{j=x+1}^{\infty} d_j \tag{14}$$

$$f_{x,x+1} = \frac{E_x + \sum_{i=0}^{x-1} f_{i,i+1}\left(\sum_{j=x}^{\infty} d_j\right) + \sum_{j=2}^{x-1} d_j \sum_{k=x+1-j}^{x} f_{k,k+1}}{d_0}\ ,\ f_{0,1} = E_0 \tag{15}$$

Expected length of a non-production period is easily computed. To complete a non-production period, the difference between X^{**} and X^{*} demands should occur. Hence, expected length of a non-production period is expressed as:

$$L_N(X^*, X^{**}) = \frac{X^{**} - X^*}{\lambda} \tag{16}$$

In order to calculate length of a production period, we carry out the first passage time analysis for our process. For a small example, suppose there 4 states from 0 to 3. The transition rate is defined as λ_{ij} from state i to j. The termination state is determines as state 3 and we try to find the expected first passage time from state 1 to 3 (m_{13}). Figure 4.8 shows the transition diagram for this example system. Equations are derived as follows:

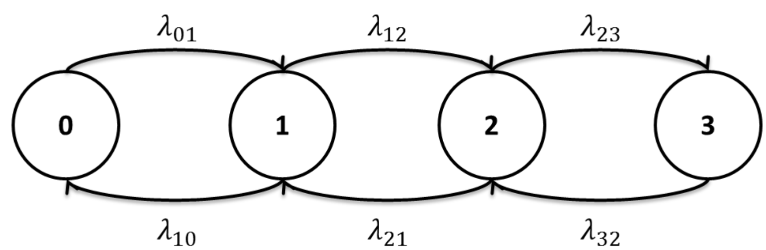

Figure 4.8 A sample state diagram

$$m_{13} = \frac{1}{\lambda_{10} + \lambda_{12}} + \frac{\lambda_{10}}{\lambda_{10} + \lambda_{12}} m_{03} + \frac{\lambda_{12}}{\lambda_{10} + \lambda_{12}} m_{23}$$

$$(\lambda_{10} + \lambda_{12}) m_{13} = 1 + \lambda_{10} m_{03} + \lambda_{12} m_{23}$$

$$0 = 1 + \lambda_{10} m_{03} + \lambda_{12} m_{23} - (\lambda_{10} + \lambda_{12}) m_{13}$$

$$0 = 1 + \lambda_{10} m_{03} + \lambda_{12} m_{23} + \lambda_{11} m_{13}$$

$$0 = 1 + \sum_{k \neq 3} \lambda_{1k} m_{k3} \tag{17}$$

The Equation (18) for any continuous time Markov process, the mean first passage times m_{ij} ($for\ i \neq j$) satisfy,

$$0 = 1 + \sum_{k \neq j} \lambda_{ik} m_{kj} \tag{18}$$

We can get the first passage time by using a matrix form. Firstly, the infinitesimal matrix Λ which denotes the rates departing from each state i and arriving in each state j is prepared. If our objective is to reach state j first time, j^{th} row and j^{th} column in Λ has to be changed with zeros and put 1 to (j, j) cell. Finally, we have a modified transition matrix Λ_j^+. These replacements are executed in order to invert the matrix to get m_j array. Then the matrix form of the equations is expressed in Equations (19, 20). The result m_j array is a column vector whose elements are the first passage times from each i to given j. If we are interested in find first passage time from any state i to j, we need to read the value of i^{th} element in the array.

$$0 = 1 + \Lambda_j^+ m_j \tag{19}$$

$$m_j = -(\Lambda_j^+)^{-1} \mathrm{I} \tag{20}$$

Finally, we calculate each part of the expected cost per unit time. In order to find the expected cost for any different system parameters and control values, the algorithms are coded in MATLAB for MDP and RA. We have considerable computation times by using this programming tool.

Table 4.7 CPU Times of MDP & RA

X^*	$X^{**} - X^*$	CPU Time of MDP	CPU Time of RA
1	9	18.33	0.22
3	7	20.20	0.23
5	5	21.23	0.23
7	3	22.15	0.23
9	1	23.15	0.24
10	0	23.90	0.24

Table 4.7 shows that as X^* increases, the computation times for $M/M/1$ is still quite short for RA. The difference between CPU times is getting larger as X^* increases for $M/M/1$. We can foresee that the computation times of MDP for different service times are distributed as Erlang or Coxian should be longer than the computation times of MDP for $M/M/1$. However, we have a robust tool which calculates the average cost in milliseconds. This study can be adapted for the systems with different service times such as Erlangian and Phase-type.

4.2 The Proposed Rationing Policy

Stock-rationing is actually an inventory policy which decides how much stock should be reserved in anticipation of future demand of more valuable demand class. If there is only one demand class, there is no meaning of stock rationing; otherwise the meaning of inventory rationing gains importance.

Customers can be classified according to their shortage costs or their service levels. For example, if one of the classes has the largest unit shortage cost, this demand class can be named by the highest priority demand class (i.e. demand class 1) or if one of them has the smallest unit shortage cost, the demand class is named by the lowest priority demand class (i.e. demand class N). In our problem, customer classes are prioritized based on different shortage costs of customers. Hence, rejecting the demand of less valuable classes provides to keep inventory for the future demand of a more valuable class.

Any rationing policy can be implemented by determining critical levels to satisfy the different demands. These critical levels can be either static or dynamic. Static rationing policies check only the inventory level and decide to satisfy the demand or not. If the inventory level is more than or equal to the critical level of arriving demand class, the demand is satisfied; otherwise the unsatisfied demand is lost. Dynamic rationing policies have different critical levels that change on a time. There are only a couple of studies that consider dynamic rationing levels. The one of them is mentioned before Fadiloglu and Bulut (2010) that analyze dynamic rationing policy is based on inventory level and number of outstanding orders. In our model, it can be easily seen that the optimal rationing decisions change according to inventory level and number of active servers at the same time. Hence, the critical level should be a function of inventory level (x) and number of active servers (y).

This section is addressed on an alternative inventory policy. The proposed policy uses the number of outstanding orders and on-hand inventory to decide whether a less valuable demand class should be satisfied or lost. Thanks to information technology, any firm can control the number of active server status and decide by incorporating this information. Therefore, the critical level R is defined and changes with the different values of x and y.

The first function at the following is rationing decision for class 1 customer. For the highest priority class, R equals to 1. Even if there is only one inventory on-hand, it is optimal to satisfy the demand of the highest customer class. For the other customer classes, the decision to meet or not to meet the demand can change according to the defined critical levels.

$$Rationing\ decision\ for\ class\ 1\ customer = \begin{cases} 1 & (x+ay) \geq R_y^x \\ 0 & (x+ay) < R_y^x \end{cases} \quad \left(R_y^x = 1\right)$$

$$Rationing\ decision\ for\ class\ i\ customer = \begin{cases} 1 & (x+ay) \geq R_y^x \\ 0 & (x+ay) < R_y^x \end{cases} \quad \left(R_y^x > 1\right)$$

$$for\ i\ \in \{1,2,\dots,N\}$$

Besides the inventory level and number of active servers, the critical level R includes another variable a that is a ratio, denotes the value of one outstanding order with respect to one unit of inventory. For instance, if $a = 0.5, x = 3$, and $y = 2$, two active servers indicates as one unit of inventory ($2\ active\ servers * 0.5\ inventory/ active\ server = 1\ inventory$). A modified inventory level is provided by summing both inventories ($2 + 1 = 3\ modified\ inventory$). Thus, this equation has the information of how many products will be replenished in the future.

4.2.1.1 The Behaviour of Proposed Rationing Policy

In this section, numerical analysis is conducted to show similarities and differences between optimal rationing policy and the other policies. First of all, Table 4.8 compares the rationing decisions of the static, dynamic, and FCFS policies with the optimal policy decisions. The dynamic rationing policy has similar behaviour with the optimal one. When inventory level is less than 2, the rationing decision is not to satisfy. However, when the inventory level equals to 2, the decisions are not the same for each number of active servers. While the number of active servers increases, the decision turns to satisfy the demand. Because the production rate is getting higher and the processing time is getting smaller.

Table 4.8 Optimal, Static, FCFS, and The Proposed Rationing Policy Decisions

$(K, s, h, \mu, \lambda_1, \lambda_2, c_1, c_2) = (2, 4, 1, 1, 3, 1, 4, 1)$										
Optimal Rationing Policy						When $a = a*$				
	Number of Active Servers y					Number of Active Servers y				
Inventory Level x	0	1	2	3	4	0	1	2	3	4
0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	0	0	0	0	1
3	1	1	1	1	1	1	1	1	1	1
When $a = 0$						FCFS Policy				
	Number of Active Servers y					Number of Active Servers y				
Inventory Level x	0	1	2	3	4	0	1	2	3	4
0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	1	1	1	1	1
2	0	0	0	0	0	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1

Table 4.9 shows the effect of number of the active server for rationing decision and compares optimal and proposed policy 3 decisions. As can be easily observed from the table, the threshold inventory rationing levels for demand class 2 are non-increasing as number of active servers increases and are decreasing as on-hand inventory increases. It means that if there are many active servers in production, arriving demands of class2 are satisfied in lower values of on-hand inventory because of high production rate.

Table 4.9 Optimal and Proposed Policy Rationing Decisions for larger s

$(K, s, h, \mu, \lambda_1, \lambda_2, c_1, c_2) = (2, 10, 1, 1, 3, 1, 4, 1)$												
Inventory Level x	Number of Active Servers y											Average Cost
	0	1	2	3	4	5	6	7	8	9	10	
0	0	0	0	0	0	0	0	0	0	0	0	5.2835
1	0	0	0	0	0	0	0	0	0	1	1	
2	0	0	0	0	1	1	1	1	1	1	1	
3	1	1	1	1	1	1	1	1	1	1	1	
$(K, s, h, \mu, \lambda_1, \lambda_2, c_1, c_2, a, R) = (2, 10, 1, 1, 3, 1, 4, 1, 0.4, 3)$												
Inventory Level x	Number of Active Servers y											Average Cost
	0	1	2	3	4	5	6	7	8	9	10	
0	0	0	0	0	0	0	0	0	0	0	0	5.2893
1	0	0	0	0	0	0	1	1	1	1	1	
2	0	0	0	1	1	1	1	1	1	1	1	
3	0	1	1	1	1	1	1	1	1	1	1	

Figure 4.10 shows the cost differences of FCFS, static and the proposed policies according to the optimal policy while the total traffic intensity is the same but the ratio of $\lambda_2/(\lambda_1 + \lambda_2) = p$ changes. In the extreme points of the graph, there is no cost difference. Because, there is only one customer class in the system and the stock rationing is not so significant for only one customer type. In the middle points of the graph, the figure starts to change since there are two different demand classes. One of them is more valuable than the other. Hence, the stock rationing becomes significant and dynamic rationing policy can be ideal for such a system.

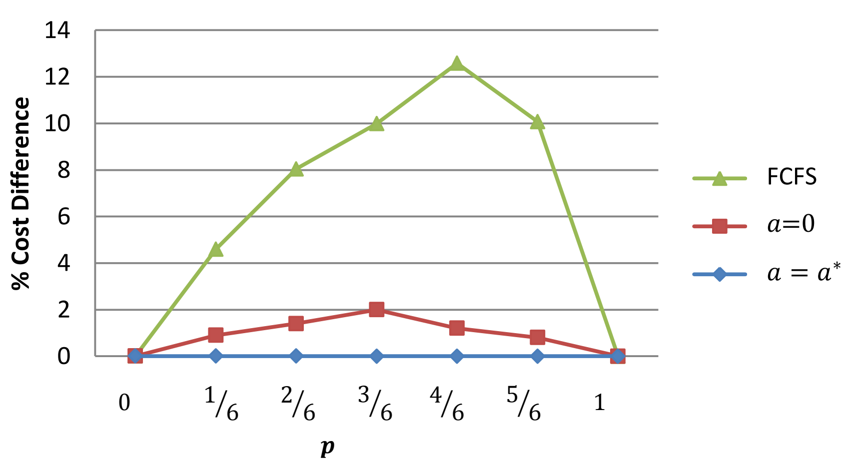

Figure 4.9 The Impact of the ratio of *p*

Figure 4.11 shows the cost differences of FCFS, static and the proposed policies according to the optimal policy while number of servers increases. For small values of s, there is no difference between the static and dynamic policies. While the number of servers increases, the performance of static policy is getting worse. The reason of this situation is the static policy does not use number of server information. Therefore, the proposed dynamic policy is the same as optimal policy for each number of servers.

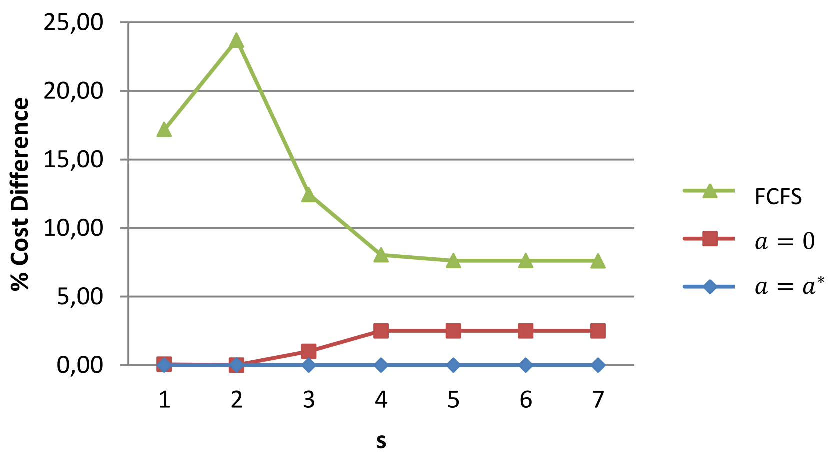

Figure 4.10 The Impact of Number of Servers

5 CONCLUSION

In this thesis, we consider a joint production and inventory problem in a make-to-stock production system with fixed start-up cost, limited parallel servers and several different customer classes at the same time. This work is a direct extension of the literature which considers a single server with fixed cost and multi-server with no fixed cost.

First of all, we model the system as an $M/M/s$ make-to-stock queue and characterize the structures of the optimal production and rationing policies via numerical studies. Because of dynamic structures of the optimal policies, we propose new policies which are well-defined, well-performing and easier to be applied in real life. We demonstrate three alternative production policies and one dynamic rationing policy. To assess the performances of the proposed policies, we analyze the impact of each system parameter on proposed policies. We also calculate the steady-state probabilities for proposed production policy 2 and conduct the renewal reward analysis for $M/M/1$ make-to-stock queue with fixed cost in lost sales environment. For such a system, the proposed production policy 3 is optimal. Renewal analysis enables us to calculate the expected average cost without calculating the steady-state probabilities. Hence, we can provide considerable computation times between RA and MDP.

A direct extension of our renewal analysis can be for multi-server or general processing times. It can provide us to calculate the expected average cost quickly. Another future extension of the study can be on make-to-stock production systems with multi-item. If multiple products are included in the problem, the sequence of the production of different products gains importance and change over costs can arise. Also, arriving demands can occur in batches and each server can complete production with batches.

REFERENCES

Altiok, T., 1989, (R, r) Production/Inventory Systems, Operations Research, **37** 266-276.

Arslan, H., Graves, S.C., and Roemer, T., 2007, A single product inventory model for multiple demand classes, Management Science, **53** 1486-1500.

Deshpande, V., Cohen, M.A., and Donohue, K., 2003, A threshold inventory rationing policy for service-differentiated demand classes, Management Science, **49** 683-703.

Fadiloglu, M.M., Bulut, O., 2010, A dynamic rationing policy for continuous-review inventory systems, European Journal of Operational Research, **202** 675-685.

Gavish, B., Graves, S.C., 1980, A one-product production/inventory problem under continuous review policy, Operations Research, **28** 1228-1236.

Gavish, B., Graves, S.C., 1981, Production/Inventory Systems with a Stochastic Production Rate under a Continuous Review Policy, Computers & Operations Research, **8** 169-183.

Gayon, J.P., Vericourt, F.D., and Karaesmen, F., 2009, Stock rationing in an $M/E_r/1$ multi-class make-to-stock queue with backorders, IIE Transactions, **41** 1096-1109.

Graves, S.C., Keilson, J., 1981, The compensation method applied to a one-product production/inventory problem, Mathematics of Operations Research, **6** 246-262.

Ha, A.Y., 1997a, Inventory rationing in a make-to-stock production system with several demand classes and lost sales, Management Science, **43** 1093-1103.

Ha, A.Y., 1997b, Stock rationing policy for a make-to-stock production system with two priority classes and backordering, Naval Research Logistics, **43** 458-472.

REFERENCES (continue)

Ha, A.Y., 2000, Stock Rationing in an $M/E_k/1$ make-to-stock queue, Management Science, **46** 77-87.

Lee, H.L., Tang, C.S., 1997, Modeling the costs and benefits of delayed product differentiation, Management Science, **43** 40-53.

Lee, H.S., Srinivasan, M.M., 1989, The Continuous Review (s, S) Policy for Production/Inventory Systems with Poisson Demands and Arbitrary Processing Times, Technical Report,87-33.

Lee, H.S., Srinivasan, M.M., 1991, Random Review Production/Inventory Systems with Compound Poisson Demands and Arbitrary Processing Times, Management Science, **37** 813-833.

Lee, J.E., Hong, M.M., 2003, A stock rationing policy in a (s,S)-controlled stochastic production system with 2-phase coxian processing times and lost sales, International Journal of Production Economics, **83** 299-307.

Liu, M., Feng, M., and Wong, C.Y., 2014, Flexible service policies for a Markov inventory system with two demand classes, International Journal of Production Research, **151** 180-185.

Liu, S., Zhang, C., 2015, An efficient inventory rationing strategy for systems with two demand classes and backordering, International Journal of Production Research, **53** 6136-6142.

Nahmias, S., Demmy, W., 1981, Operating characteristics of an inventory system with rationing, Management Science,, **27** 1236-1245.

Pang, Z., Shen, H., and Cheng, T.C.E., 2014, Inventory Rationing in a Make-to-Stock System with Batch Production and Lost Sales, Production and Operations Management, **151** 180-185.

REFERENCES (continue)

Sobel, M.J., 1968, Optimal average cost policy for a queue with start-up and shut-down costs, Operations Research, **17** 145-162.

Tijms, H.C., 1980, An Algorithm for Denumerable State Semi-Markov Decision Problems with Application to Controlled Production and Queueing Systems, Academic Press, 145-162.

Veinott, A.F., 1965, Optimal policy in a dynamic, single product, non-stationary inventory model with several demand classes, Operations Research, **13** 761-778.

Wang, D., Tang, D., 2014, Dynamic inventory rationing with mixed backorders and lost sales, International Journal of Production Economics, **149** 56-67.

CURRICULUM VITEA

Sinem Özkan was born in İzmir. She received her B.Sc. degree from Yaşar University in 2012 and started her graduate studies in 2013. Before she joined the M.Sc. program, she worked one year as a production engineer in a textile company. She is currently a research assistant in the Department of Industrial Engineering at Yaşar University where she has been working since 2013. During her graduate studies, she has taken many courses including System Simulation, Optimization Models and Algorithms, Probabilistic Analysis, Applied Stochastic Processes, Dynamic Programming, Supply Chain Processes and Management and Heuristic Optimization. She attended APS 2015 (Applied Probability Society Conference) in İstanbul, EURO 2015 (European Conference on Operational Research) in Glasgow and ORIE 2015 (Operational Research and Industrial Engineering Conference) in Ankara. Her research interests include stochastic modeling, production and inventory systems.