

Customer Differentiated End-of-Life Inventory Problem

Morteza Pourakbar, Rommert Dekker

Erasmus School of Economics, Erasmus University Rotterdam

†Burgemeester Oudlaan 50, 3000DR Rotterdam, The Netherlands

pourakbar@ese.eur.nl, laan@rsm.nl, rdekker@ese.eur.nl

* Phone: (+31) 10 - 40 81274, Fax: (+31) 10 - 4089162

ECONOMETRIC INSTITUTE REPORT EI 2011-21

Abstract

This paper deals with the service parts end-of-life inventory problem in a circumstance that demands for service parts are differentiated. Customer differentiation might be due to criticality of the demand or based on various service contracts. In both cases, we model the problem as a finite horizon stochastic dynamic program and characterize the structure of the optimal policy. We show that when customers are differentiated based on the demand criticality then the optimal structure consists of time and state dependent threshold levels for inventory rationing. In case of differentiation based on service contracts, we show that in addition to rationing thresholds we also need contract extension thresholds by which the system decides whether to offer an extension to an expiring contract or not. By numerical experiments in both cases, we identify the value of incorporating such decisions in service parts end-of-life inventory management with customer differentiation. Moreover, we show that these decisions not only result in cost efficiency but also decrease the risk of part obsolescence drastically.

Keywords: End-of-life inventory, spare parts, customer differentiation

1 Introduction

With the rapid technology development, life cycles of products have become shorter. As a consequence, managing the inventories of service parts in order to fulfill service obligations and avoid obsolescence risk becomes a major challenge for companies. This becomes even more crucial as the production of a service part is discontinued when the part enters its final phase of service life cycle. One of the main tactics adopted by various industries to cope with the end-of-life inventory problem is placing a “final order quantity” at the beginning of the final phase. The final order should be placed in a quantity such that it balances the risk of obsolescence at the end of horizon versus the risk of failure to meet service agreements.

The service agreements oblige the company to provide its customers with a certain service level. It has been common in practice that companies offer different service levels versus different prices. As a result, customers are segmented according to the service level they choose. This introduces the idea of rationing available inventories for some customer classes. The practice of rationing inventory (or capacity) among different customer classes is an increasingly important tool for balancing supply with demand in environments where requirements for service vary widely. Basically a rationing policy issues stock to some customers while refuses or delays demand fulfillment for others. It is analogous to the highly successful yield management policies adopted by airlines and hotels. In capital-intensive goods industry there are several examples of implementing a customer differentiation scheme. For example defense systems (Deshpande et. al. 2003a and 2003b), semiconductor manufacturing equipment (Kranenberg and Van Houtum 2008), and mobile phone operating systems (Möllering and Thonemann 2008). Essentially, there are various situations that illustrate the importance of customer differentiation for inventory control, namely demand criticality based differentiation or service contract based differentiation. The criticality of the demand for the spare parts can be due to various reasons. First, a spare part can be critical in one place and non-critical in another. For example, breaking down of a mainframe computer at a stock exchange has more severe financial impact than when a mainframe computer in a library goes down. Another example is given by Dekker et. al. (1998). They consider the case of a

petrochemical plant in which one could distinguish between vital, essential and auxiliary equipments. Demands for service parts originating from equipments in each of these classes are given a different level of criticality. Another situation wherein demand criticality arises is identified where demands happen for both regular replenishment and emergency orders. Obviously, a demand for an emergency order is given a higher priority than a regular replenishment order.

In environments in which manufacturers of complex goods sell directly to customers, service parts are distributed primarily through service contracts. For OEMs and service providers that handle performance-based service contracts for customers, service payments are based on up-time performance or equipment availability, rather than actual support cost. This includes companies in the aerospace and defense, industrial chemical, semi-conductor equipment, and networking and telecom industries (Aberdeen Group, 2005). Service providers offer an array of service contracts for a single product. These service contracts impose different service requirements for service providers against different prices. Murthy and Blischke (2005) discuss the factors that form the distinction among various service contracts. In general one could argue that the distinctions are induced by varying: 1) Service responses and/or repair time commitments; and 2) Price and/or cost structures. The response times vary significantly from situation to situation. For example in mission critical situations, the service provider is always present on-site whereas in the same day delivery, the response time varies from 2 to 24 hours. Cohen et al. (2006) highlight that consumer products such as TVs and PCs have lower response time requirements. On the other hand, businesses computing machines, construction equipment and aircraft have typically very high response time requirements. The failure to meet service obligations may results in penalties that are negotiated during the contract formalization. In such a setting, service contract based differentiation has been an increasing trend for after-sale service providers.

With the technology developments and shortening of products life cycles, service parts enter the end-of-life phase sooner. Consequently, companies face a huge service responsibility while the parts are in the final phase and parts provisioning is not guaranteed any longer. Therefore, companies are mandated to take into account various types of service obligations they have while making

the final order quantity decisions. Considering this aspect, in the sequel of this paper, we study the inventory control problem of a capital-intensive product service part when the part production is discontinued. We proceed by characterizing the structure of an optimal inventory control policy in the final phase where customers are differentiated. We consider both demand criticality based differentiation and contract based differentiation schemes. Using Markov decision processes, we show that in these settings inventories should be controlled according to time dependent threshold levels that consider the level of available inventory. Moreover, we study the advantages of incorporating the critical level policies in this problem.

The rest of this paper is structured as follows. Section 2 proceeds by a brief overview of the literature on end-of-life inventory decisions and inventory problems with customer differentiation. Next, in section 3 we describe the problem. Section 4 formulates the problem when customers are differentiated based on demand criticality. Section 5 considers a service contract-based differentiation and section 5 concludes the paper.

2 Literature Review

In this paper, we consider the inventory control of service parts of a capital good manufacturer in the final phase of their service life cycle. The primary trade-off in this phase is balancing the risk of obsolescence and that of unmet service obligations. To do so, one of the main tactics used in practice is placing a final order. This problem is called final buy problem (FBP), or the end of production problem (EOP). There are three streams of research on the final phase inventory problem differentiated by the approach taken, namely cost driven, service driven and forecasting based approaches.

In a service-driven approach a service level should be optimized regardless of the cost incurred by the system. Fortuin (1980, 1981) describes a service level approach and addresses non-repairable items or consumable spare parts. He derives a number of curves by which the optimal final order quantity for a given service level can be obtained. Another service driven approach is

developed by van Kooten and Tan (2009) for a system in which parts are subject to the risk of condemnation. They build a transient Markovian model by which the corresponding optimal final order quantity can be obtained for a given service level.

Basically, a cost-driven approach decides on the quantity purchased by weighing the cost of ordering too many against the cost of buying too few or in other words a news-vendor problem approach. Among this category of works there are Teunter and Fortuin (1999), Teunter and Klein Haneveld (2002), Cattani and Souza (2003), Bradley and Guerrero (2009), Krikke and van der Laan (2010) and Pourakbar et. al. (2010, 2011). The latter also provides a general review on this subject.

Forecasting based approaches focus on forecasting demand for a discontinued product instead of dealing with the production or inventory problem. Moore (1971) is the first one to propose this approach followed by Ritchie and Wilcox (1977). Hong et. al. (2008) develops a stochastic forecasting model using the install-based information to forecast final order quantity.

The focus of the previously mentioned works are on spare parts planning. However, there are similar problems in the context of new product introduction that deal with placing a final order for products rather than parts. In other words, this category of works deals with the production and inventory planning when a product is replaced by its next generation counterpart. The main issues are related to the inventory planning of old and new generations of the product together with the timing of the release of the new product. We refer the interested readers to Li et. al. (2010) and Xu et. al. (2010) and references there in for an overview of this stream of literature.

Another stream of research related to this work is the issue of customer differentiation. In traditional inventory systems all customer demands are treated equally and fulfilled based on a first come first serve policy. However, in practical settings, managers often differentiate customers by assigning a priority based on the level of service customers are expecting. A very common approach employed in this setting is critical level policy. According to a critical level policy the system reserves some inventory for the higher priority customer. In other words, given the current level of inventory, the system may deliberately decide to deny some lower priority customers access

to the inventory in anticipation of higher priority demand arriving in the future. It is shown that this type of policy has considerable advantage over the traditional inventory policies in which it is assumed that all customers receive the same level of service. These types of problem have been studied extensively under various assumptions, see for example Veinott (1965), Topkis (1968), Ha (1997), Cattani and Souza (2002), Dekker et. al. (2002), De Vericourt et. al. (2002), Deshpande et. al. (2003), Frank et. al. (2003), Kranenburg and Van Houtum (2007, 2008), Teunter, Klein Haneveld (2008) and Benjaafar et. al. (2010).

However, to the best of our knowledge the issue of a customer differentiated inventory system in the final phase has not been dealt with in the literature. There are certain aspects that make this problem paramount. First of all, the end-of-life phase is considered the longest in the service parts life cycle. Therefore, we expect that service responsibilities are mostly stretched to this phase as well. However, another major complication in this phase is that the opportunity of ordering parts is not guaranteed and therefore companies should consider various service obligations at the moment the final order is placed. What distinguishes our work from the rest of the literature is that first, we consider customer differentiation in the final phase of the service life cycle. Next, we characterize the structure of the optimal inventory policy in this phase for demand criticality based differentiation and service contract based differentiation. Using a finite horizon Markov decision process, we show that in case of demand criticality based differentiation inventories should be controlled according to some time and state-dependent criticality levels. Moreover, for service contract based differentiation we show that we also need some time dependent contract extension thresholds. Characterizing the optimal policy structure, we study the value of the optimal policy by investigating similar systems without critical level or contract extension thresholds.

3 Problem Description

We consider the End-of-Life inventory decisions associated with a spare part used by different classes of customers in a finite horizon $[0, H]$. 0 denotes the time that the last time procurement

decision for service parts should be made . Parts can be either produced or purchased with a cost c_p per item. After this time, parts acquisition is not guaranteed any more. H signifies the end of the horizon and it is the time that the last service or warranty contract expires. All parts left at the end of the horizon are considered obsolete and should be disposed of. While deciding on the final order quantity, one should note that overage stock of service parts imposes high carrying and obsolescence costs while underage stock results in service failure. Therefore, one primary challenge is to balance the risk of obsolescence versus failure to meet service commitments.

We assume that customers are differentiated into S different categories. Demands of each category arrive according to a non-stationary Poisson distribution with mean value function $\Lambda_i(t)$. The non-stationarity in demand arrival is due to the fact that in this phase, the size of the installed-base available in the market is shrinking and so does the demand rates for spare parts.

We assume unmet demands are lost with a cost c_s where $s \in \{1, 2, \dots, S\}$. The lost sale assumption stems from the fact that as long as the system is in the final phase ordering parts is not guaranteed. Therefore if shortages are backlogged there is no assurance that system could satisfy them later in the course of final phase. Thus, lost-sale assumption is more reasonable. Furthermore, the higher the priority of a customer the larger is the lost sale cost, $c_s > c_{s+1}$ for $s \in \{1, 2, \dots, S - 1\}$. For each unit of serviceable inventory the system incurs a holding cost of h per unit per time . Moreover, all cost terms are discounted back to the beginning of the horizon with a rate α . All the remaining serviceable inventory at the end of the final phase are considered obsolete and should be disposed of with a cost d_s per unit.

One popular tactic in such a case to mitigate the risk of service failure is to ration the serviceable inventory. According to a rationing scheme the system might decide to deliberately avoid serving a lower priority customer in order to save the inventory in anticipation of future demand for higher priority customers. Moreover, if customers are segmented according to service contracts then denying to offer contract renewal to less valuable customers is another tactic, introduced in this paper, to ensure higher priority customers demand satisfaction.

4 Demand Criticality- Based Differentiation

In this setting customers are differentiated based on the criticality of their demands. One example is the case where a specific part is critical in one equipment and non-critical in another one. Another example is identified where there are demands for both emergency and regular replenishment shipment. Emergency shipment might be due to machine failure and therefore are signified more critical than regular shipment orders. In our model, we assume there are S different categories of demand and servicing a type i demand result in a revenue of amount r_i . Moreover, higher priority customers generate a larger revenue than lower priority ones, $r_i > r_{i+1}$ for $i \in \{1, 2, \dots, S - 1\}$.

Upon arrival of a customer of type i , the system should decide whether to satisfy or decline service in order to preserve the serviceable items for future demand of higher priority customers. In the rest of this section we formulate this problem as a finite horizon Markov decision process and using this formulation we characterize the structure of the optimal policy. First, without loss of generality we assume that the time interval $[0, H]$ is divided into periods of length one unit such that the probability of having more than one demand of any type i is negligible. This time unit can represent a month, a week, a day or an hour. Then the optimality equation at time t for a specific serviceable inventory level x is denoted by $\nu(t, x)$. In a case where $x \geq 1$ and $t < H$, it is formulated as

$$\begin{aligned} \nu(t, x) = & (1 - \sum_{i=1}^S \lambda_i(t)) \{hx + \alpha\nu(t + 1, x)\} \\ & + \sum_{i=1}^S \lambda_i(t) \{hx + \min\{\alpha\nu(t + 1, x) + c_i, \alpha\nu(t + 1, x - 1) - h - r_i\}\} \end{aligned} \quad (1)$$

The first term represents a situation in which no demand occurs during period t , and the second term explains a situation wherein a demand of type i arrives during t for which the system manager should decide whether to accept or deny it. In case $x = 0$ and $t < H$, all incoming demands are lost therefore we have

$$\nu(t, 0) = (1 - \sum_{i=1}^S \lambda_i(t))\alpha\nu(t + 1, 0) + \sum_{i=1}^S \lambda_i(t) \{\alpha\nu(t + 1, 0) + c_i + r_i\} \quad (2)$$

At time H all available serviceable inventory should be disposed of therefore the terminal value function is given by

$$\nu(H, x) = d_s x \quad (3)$$

We assume that it is always optimal to use a part to satisfy a demand than to dispose of it, i.e. $h + r_S + c_S > \alpha d_s$ where S corresponds to the least valuable customer. Note when there is on-hand serviceable stock, i.e. $x > 0$, at any time t it is optimal to satisfy an incoming class i demand if the increase in the system costs due to having one less unit of inventory is less than the associated unit lost sales cost and the generated revenue. The system rejects the class i incoming demand otherwise. This condition is formulated as

$$\alpha [\nu(t + 1, x - 1) - \nu(t + 1, x)] \leq c_i + r_i + h \quad (4)$$

This condition identifies the rationing threshold and has a simple interpretation. The system keeps satisfying demands from customers of type i as long as the discounted future loss from having one less unit of the serviceable inventory is less than the costs incurred by rationing. The following lemma is useful in characterizing the structure of the optimal policy.

Lemma 1 $\nu(t, x) - \nu(t, x - 1)$ is non-decreasing in x .

Proof. We prove this by induction. First, for the ease of exposition we define $\Delta_x \nu(t, x) = \nu(t, x) - \nu(t, x - 1)$ we note that $\Delta_x \nu(t, x)$ non-decreasing in x is equivalent to $\nu(t, x)$ being discretely convex. We also define $\Delta_x \nu(t, 0) = \alpha / (c_i + r_i + h)$. From the terminal value definition, $\nu(H, x) = d_s x$, we have $\Delta_x \nu(H, x) = d_s$. From the assumption $(h + r_S + c_S > \alpha d_s)$ we immediately have $\Delta_x \nu(H, x) > \Delta_x \nu(H, 1) > \Delta_x \nu(H, 0)$, $x \geq 1$. Therefore, $\Delta_x \nu(H, x)$ is non-decreasing in x .

Following the induction, we assume that $\Delta_x \nu(t, x)$ is non-decreasing in x , then we need to show that $\Delta_x \nu(t - 1, x)$ is non-decreasing in x . Relation (4) implies that $\exists r_i(t)$ such that it is optimal to deny an arriving demand of type i if $x < r_i(t)$ and fulfill it otherwise. Then using

equation (1) we have

$$\begin{aligned} \Delta_x \nu(t-1, x) = & (1 - \sum_{i=1}^S \lambda_i(t)) [h + \alpha \Delta_x \nu(t, x)] \\ & + \sum_{i=1}^S \lambda_i(t) \begin{cases} h + \alpha \Delta_x \nu(t, x) & x < r_i(t) \\ h + \alpha \Delta_x \nu(t, x-1) & x \geq r_i(t) \end{cases} \end{aligned} \quad (5)$$

Given that $\Delta_x \nu(t, x)$ is non-decreasing in x , it is clear from the above relation that $\Delta_x \nu(t-1, x)$ is also non-decreasing in x at any time t . \square

This lemma establishes that the value function $\nu(t, x)$ is convex in x for all time t . This lemma together with condition (4) leads to the optimality of a time-varying threshold policy as follows:

Theorem 2 *The optimal end-of-life inventory control policy when customers are differentiated based on demand criticality is a time-varying threshold policy. It can be characterized as follows:*

- i. *There exist threshold levels $r_1(t), r_2(t), \dots, r_S(t)$ such that if $x \geq r_s(t)$ an incoming demand of class s is satisfied and rejected otherwise.*
- ii. *At each time t we have $0 = r_1(t) \leq r_2(t) \leq \dots \leq r_S(t)$*
- iii. *$r_s(t)$ is non-increasing in t , $s \in \{1, 2, \dots, S\}$*

Proof.

i. We note that $\nu(t, x)$ is convex. Considering (4) for an arbitrary class i if $x \geq r_i(t)$ the minimizer of the objective function is to satisfy demand of class i . If $x < r_i(t)$ the minimizer is to reject any arriving demand of class i .

ii. This can be proved using the definition of $r_i(t)$ in relation (4). If $j \leq i$, $j, i \in \{1, 2, \dots, S\}$ then $\alpha [\nu(t+1, x-1) - \nu(t+1, x)] \leq c_i + r_i + h \leq c_j + r_j + h$. Thus $r_j(t) \geq r_i(t)$.

iii. We consider an arbitrary class i and show that $r_i(t)$ is non-increasing in t . The induction assumption in this case is $r_i(t) \geq r_i(t+1)$ and $\Delta_x \nu(t, x) \geq \Delta_x \nu(t+1, x)$, $\forall x \in [1, r_i(t)]$. Then we

need to show that $r_i(t-1) \geq r_i(t)$ and $\Delta_x \nu(t-1, x) \geq \Delta_x \nu(t, x), \forall x \in [1, r_i(t-1)]$. Considering the definition of $r_i(t)$ (4), we have $\alpha \Delta_x(t, x) \leq c_i + r_i + h$ if $x < r_i(t)$.

Using a contradiction approach we assume that $r_i(t-1) < r_i(t)$. Then we have $\alpha \Delta_x(t, r_i(t-1)) \leq c_i + r_i + h$ if $x < r_i(t)$. Then using equation (5), we have

$$\begin{aligned} \alpha \Delta_x(t-1, r_i(t)) - \alpha \Delta_x(t, r_i(t)) &= (1 - \sum_{i=1}^S \lambda_i(t)) [\alpha \Delta_x(t, r_i(t)) - \alpha \Delta_x(t+1, r_i(t))] \\ &+ \sum_{i=1}^S \lambda_i(t) \begin{cases} \alpha \Delta_x \nu(t, r_i(t)) - \alpha \Delta_x \nu(t+1, r_i(t)) & r_i(t) < r_i(t+1) \\ \alpha \Delta_x \nu(t, r_i(t)-1) - \alpha \Delta_x \nu(t+1, r_i(t)-1) & r_i(t) \geq r_i(t+1) \end{cases} \end{aligned} \quad (6)$$

By induction assumption, we know that the terms $\alpha \Delta_x \nu(t, r_i(t)) - \alpha \Delta_x \nu(t+1, r_i(t))$ and $\alpha \Delta_x \nu(t, r_i(t)-1) - \alpha \Delta_x \nu(t+1, r_i(t)-1)$ are non-negative. Thus $\alpha \Delta_x(t-1, r_i(t)) - \alpha \Delta_x(t, r_i(t))$ is also non-negative. Consequently $\alpha \Delta_x(t-1, r_i(t)) > \alpha \Delta_x(t, r_i(t)) > c_i + r_i + h$ which implies that $r_i(t-1) > r_i(t)$ which is a contradiction. Thus the desired result is shown.

Next, we show that $\Delta_x \nu(t-1, x) \geq \Delta_x \nu(t, x), \forall x \in [1, r_i(t-1)]$. Given that $r_i(t-1) \geq r_i(t)$, we can distinguish the following cases

Case I. $1 \leq x \leq r_i(t+1)$

In this case, we have

$$\begin{aligned} \alpha \Delta_x(t-1, r_i(t)) - \alpha \Delta_x(t, r_i(t)) &= (1 - \sum_{i=1}^S \lambda_i(t)) [\alpha \Delta_x(t, r_i(t)) - \alpha \Delta_x(t+1, r_i(t))] \\ &+ \sum_{i=1}^S \lambda_i(t) [\alpha \Delta_x \nu(t, r_i(t)) - \alpha \Delta_x \nu(t+1, r_i(t))] \end{aligned} \quad (7)$$

The right-hand side is non-negative therefore, $\alpha \Delta_x(t-1, x) - \alpha \Delta_x(t, x) \geq 0$ which shows the desired result.

Case II. $r_i(t+1) < x \leq r_i(t)$

The results can be shown similar to the previous case.

Case III. $r_i(t) < x \leq r_i(t-1)$

In this case we have according to the definition of the rationing levels, equation (4),

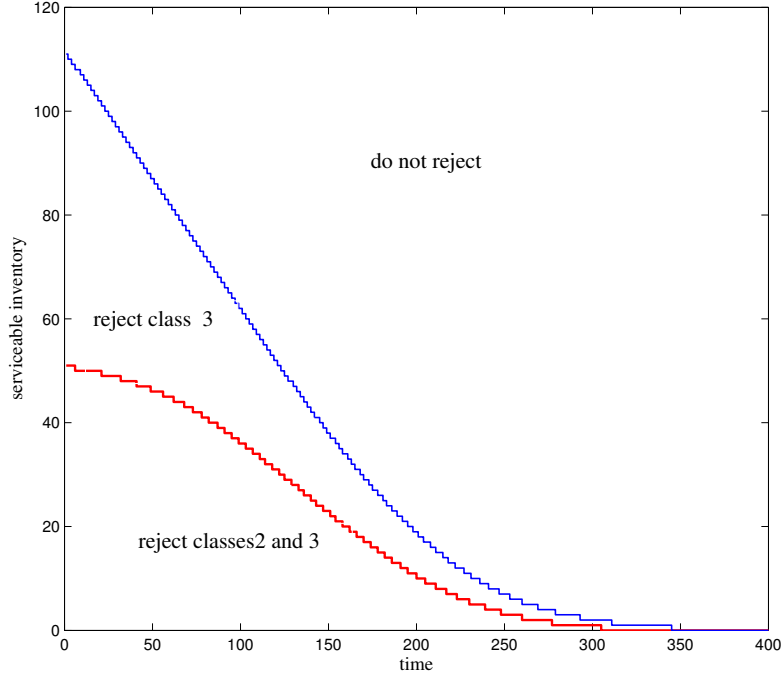


Figure 1: rationing threshold for a system with three classes of customers

$\alpha \Delta_x \nu(t, x) \leq c_i + r_i + h$ and $\alpha \Delta_x \nu(t-1, x) \geq c_i + r_i + h$ therefore $\Delta_x(t-1, x) - \Delta_x(t, x) \geq 0$.

We also need to show that the results hold for $t = H$. It is straightforward since intuitively $0 = r_i(H) \leq r_i(H-1)$. \square

Figure 1. demonstrates the optimal rationing policy for a system with three different customer classes. As it is observed the state space is divided into three different regions. The contour of each region is the rationing threshold. We observe that, when it is closer to the end of the horizon, the system tends to set a lower rationing level. The system just denies the demand from a lower priority class if the level of inventory is very low. In other words, the rationing levels are non-increasing in time. Moreover, if class i is more critical than class j , then the system sets a lower rationing threshold for class i than class j .

4.1 Final Order Quantity

So far the procurement and production costs are not accounted in the formulation since these can be considered sunk cost in making the rationing decisions. However this cost needs to be considered when we make decision over the final order quantity. The unit provisioning cost which is either the cost of purchasing or production is denoted by c_p . Then the net total cost at time 0 is defined by

$$TV(x) = \nu(0, x) + c_p x \quad (8)$$

As an immediate result from convexity of $\nu(t, x)$, we have $TV(x)$ is also convex and the optimal final order quantity can be obtained using a simple search algorithm.

4.2 Numerical Analysis

By conducting this numerical analysis we aim at quantifying the value of having a rationing decision in the course of the final phase. To do so, we consider a system without rationing. In other words, in case of available serviceable inventory the system always fulfill demands from all types of customers. For the base case scenario, we assume there are three different equipment criticality levels for which the associated total shortage and lost sale costs are $c_1 = 250$, $c_2 = 150$ and $c_3 = 50$. Holding cost is set as 0.05 per unit per time. Moreover, demand is a non-homogenous Poisson process with rates $\lambda_1(t) = 0.45/(1+\exp(0.025(t-200)))$, $\lambda_2(t) = 0.25/(1+\exp(0.025(t-200)))$ and $\lambda_3(t) = 0.20/(1 + \exp(0.025(t - 200)))$ respectively for class 1, 2 and 3 customers and $H = 400$. In order to study the value of rationing we compare the results of the system with rationing with the results of of a similar system without rationing. To do so the cost of a system with rationing is denoted by C_r and the one without rationing is denoted by C_{opt} . Then, the value of a rationing decision is calculated according to

$$\Delta\% = \frac{C_r - C_{opt}}{C_{opt}} \times 100 \quad (9)$$

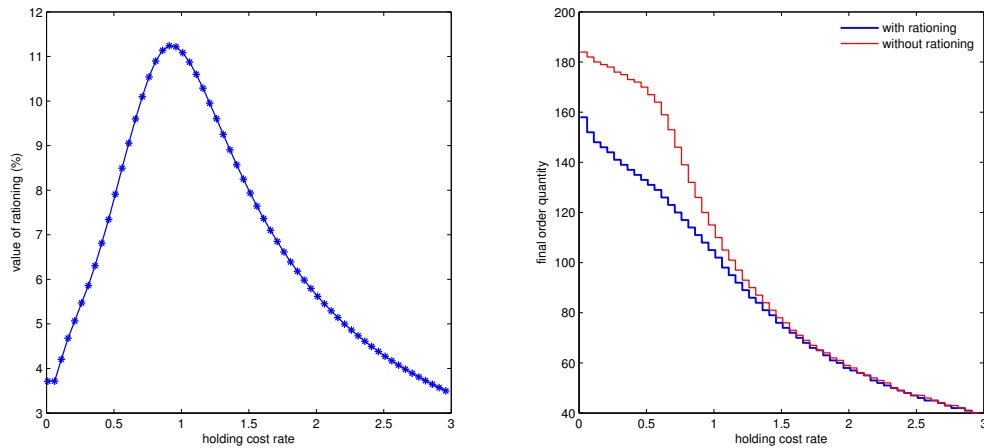


Figure 2: the value of a rationing option and its effect on final order quantity

Figure 2. depicts the effect of holding cost rate on the value of rationing and also the final order quantity. As we observe in this case, the final order quantity decreases as it becomes more expensive to hold inventory. Moreover, a system with rationing tends to place a smaller final order quantity than a system without rationing. As holding cost rate increases at first the value of rationing increases. It is due to the fact that the more expensive the items are to hold, the less inventory level is preferred and as a consequence rationing plays a more vital role. However, as the holding cost rate becomes higher it might become more advantageous to stock less and therefore deny demands for service rather than satisfying them thus, rationing loses its value.

Furthermore, as intuition dictates when the lost-sale ratio c_1/c_2 increases, the importance of meeting class 1 demands becomes more vital. As a consequence, having a rationing decision gains more value. We again observe that a system without rationing places a much larger order than a system with rationing. Therefore, an inherent consequence of a rationing option in the final phase would be an alleviated risk of obsolescence. By placing a smaller final order quantity while we have rationing, the system not only avoids extra carrying cost but also hedges upon the risk of obsolescence at the end of the final phase.

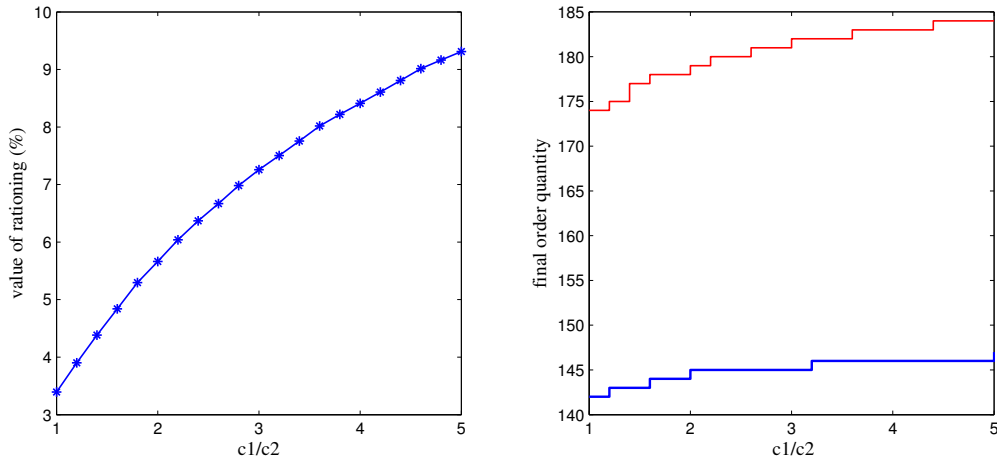


Figure 3: the value of rationing and its effect on final order quantity

5 Service Contract-based Differentiation

As mentioned earlier, over the past few decades, after-sale service contracts have become a common practice in service industries. It is due to the fact that in a highly competitive environment with increased and differentiated customer demands, service organizations pay more attention to service contracts as a source of predictable and renewable revenue streams. Once customers have signed or renewed their contracts, the service organization must meet its commitments consistently and cost effectively. In this setting, customers are differentiated based on the service contracts they choose. The service provider offers S different types of contracts such that type i contract generates a revenue of magnitude R_i and $R_1 \geq R_2 \geq \dots \geq R_S$ for a specific duration (for example a year). Customers with a more valuable service contract are given a higher priority. In such a setting the revenue is generated not per service given to the customer but per contract that runs for a certain period.

At the end of the service period the system should decide whether to offer a contract extension to a customer of type i or terminate this type of contract. Essentially the system should make this decision according to the level of serviceable inventory and the time remaining to the end of the horizon. If the inventory level is seen as critically low to satisfy the forthcoming demands for all active service contracts then it might be beneficial to consider not renewing lower priority

service contracts in order to preserve the inventory for higher priority ones. If it is decided not to extend a specific type of contract then the corresponding customers are not offered a renewal option once their contracts are expired. The times at which these decisions are made is denoted by $T = \{t_1, t_2, \dots, t_n\}$. Moreover, we should note that if the system decides not to extend a certain type of contract then from that time on, the demand rate for service decreases accordingly. In addition to the contract extension decision, the system also needs to decide whether to deny an arriving demand even if the associated service contract is running or satisfy it. In other words, the system needs to make rationing decision as well.

During the service period the service provider is mandated to satisfy customers demands. Demands which are unmet or denied, while the corresponding contract is active, are considered lost and a penalty cost is incurred. Moreover, failure to service higher priority customers leads to a higher lost sale cost, i.e. $c_1 \geq c_2 \geq \dots \geq c_S$.

The demand originated from each category of service contracts follows a non-homogenous Poisson process with intensity function $\lambda_s(t), s \in \{1, 2, \dots, S\}$. It is worth mentioning that, similar to the previous model we assume that satisfying a demand is always more beneficial than disposing of an item, i.e. $c_s > \alpha d_s$.

The optimality equation at time t , $\nu_s(t, x, \mathbf{s})$, is a function of inventory level, x and the status of various contract types, \mathbf{s} , where \mathbf{s} is a binary vector and $s_i = 1$ if the type i contract is active and 0 otherwise. The optimality equation at $t \in T$ where $x > 0$ is formulated as

$$\begin{aligned} \nu(t, x, \mathbf{s}) = & \min_{1 \leq j \leq \phi(\mathbf{s})} \left\{ (1 - \sum_{i=1}^j \lambda_i(t)) \left\{ hx - \sum_{k=1}^j R_k + \alpha \nu(t+1, x, \mathbf{s} - \sum_{h=j+1}^{\phi(\mathbf{s})} \mathbf{e}_h) \right\} \right. \\ & + \sum_{i=1}^j \left\{ \lambda_i(t) \left\{ h(x-1) - \sum_{k=1}^j R_k + \min \left\{ \alpha \nu(t+1, x, \mathbf{s} - \sum_{h=j+1}^{\phi(\mathbf{s})} \mathbf{e}_h) + c_i + h, \right. \right. \right. \\ & \left. \left. \left. \alpha \nu(t+1, x-1, \mathbf{s} - \sum_{h=j+1}^{\phi(\mathbf{s})} \mathbf{e}_h) \right\} \right\} \right\} \end{aligned} \quad (10)$$

where $\phi(\mathbf{s}) = \max\{i \mid s_i = 1, 1 \leq i \leq S\}$ finds the active type of contract with the lowest priority. The first term deals with the decision of which contract types to extend in period t where no demand arrives. The second term deals with the same decision together with a rationing decision.

Meaning that whether an arriving demand of type i should be denied or admitted even though the corresponding contract is still running. Furthermore, for $x = 0$ we have

$$\begin{aligned} \nu(t, 0, \mathbf{s}) = & \min_{1 \leq j \leq \phi(\mathbf{s})} \left\{ (1 - \sum_{i=1}^j \lambda_i(t)) \left\{ - \sum_{k=1}^j R_k + \alpha \nu(t+1, 0, \mathbf{s} - \sum_{h=j+1}^{\phi(\mathbf{s})} \mathbf{e}_h) \right\} \right. \\ & \left. + \sum_{i=1}^j \left\{ \lambda_i(t) \left\{ - \sum_{k=1}^j R_k + \alpha \nu(t+1, 0, \mathbf{s} - \sum_{h=j+1}^{\phi(\mathbf{s})} \mathbf{e}_h) + c_i \right\} \right\} \right\} \end{aligned} \quad (11)$$

Moreover, $\nu(t, x, \mathbf{s})$ where $t \in [0, H]$ and $t \notin T$ is defined similar to (1) and (2) as follows where $x \geq 1$

$$\begin{aligned} \nu(t, x, \mathbf{s}) = & (1 - \sum_{i=1}^{\phi(\mathbf{s})} \lambda_i(t)) \{hx + \alpha \nu(t+1, x, \mathbf{s})\} \\ & + \sum_{i=1}^{\phi(\mathbf{s})} \lambda_i(t) \{hx + \min\{\alpha \nu(t+1, x, \mathbf{s}) + c_i, \alpha \nu(t+1, x-1, \mathbf{s}) - h\}\} \end{aligned} \quad (12)$$

and when $x = 0$ we have

$$\nu(t, 0, \mathbf{s}) = (1 - \sum_{i=1}^{\phi(\mathbf{s})} \lambda_i(t)) \alpha \nu(t+1, 0, \mathbf{s}) + \sum_{i=1}^{\phi(\mathbf{s})} \lambda_i(t) \{ \alpha \nu(t+1, 0, \mathbf{s}) + c_i \} \quad (13)$$

In this setting we need a condition similar to (4) to assure that the system keeps satisfying incoming demand of type i as long as the discounted future loss from having one less unit of serviceable inventory is less than the lost sale cost, i.e.

$$\alpha [\nu(t+1, x-1, \mathbf{s}) - \nu(t+1, x, \mathbf{s})] \leq c_i + h, \quad i \in \{1, 2, \dots, S\}, \quad t \in [0, H] \quad (14)$$

Moreover, at contract extension decision times, $t \in T$, a specific contract of type i is extended only if the revenue generated is more than the future discounted cost of renewing the contract. This condition implies the contract extension thresholds and is formulated as

$$\alpha [\nu(t+1, x, \mathbf{s}) - \nu(t+1, x, \mathbf{s} - \mathbf{e}_i)] \leq R_i \quad i \in \{1, 2, \dots, S\}, \quad t \in T \quad (15)$$

Lemma 3 $\nu(t, x, \mathbf{s})$ is convex in $x \forall t \in [0, H]$.

Proof. In order to show this lemma we follow an induction approach similar to lemma 1. Thus, at time $t = H$ the result holds. We assume that the $\nu(t, x, \mathbf{s})$ is convex at time t and inductively establish the result for time $t - 1$. if time $t_j < t - 1 < t_{j+1}$ then the proof is similar to lemma 1. Therefore we assume that $t - 1 \in T$. $\Delta_x \nu(t, x, \mathbf{s}) = \nu(t, x, \mathbf{s}) - \nu(t, x - 1, \mathbf{s})$ similar to the previous definition. Using (10) we assume that an arbitrary index j is the minimizer, therefore we have

$$\begin{aligned} \Delta_x \nu(t, x, \mathbf{s}) = & (1 - \sum_{i=1}^j \lambda_i(t)) \left\{ h + \alpha \Delta_x \nu(t + 1, x, \mathbf{s} - \sum_{h=j+1}^{\phi(\mathbf{s})} \mathbf{e}_h) \right\} \\ & + \sum_{i=1}^j \lambda_i(t) \left\{ \begin{array}{ll} h + \alpha \Delta_x \nu(t + 1, x, \mathbf{s} - \sum_{h=j+1}^{\phi(\mathbf{s})} \mathbf{e}_h) & x < r_i(t) \\ h + \alpha \Delta_x \nu(t + 1, x - 1, \mathbf{s} - \sum_{h=j+1}^{\phi(\mathbf{s})} \mathbf{e}_h) & x \geq r_i(t) \end{array} \right. \end{aligned}$$

which establishes the convexity property at time $t - 1$. □

Having established that the value function is convex in x , together with conditions (14) and (15) we can show that a set of time dependent rationing and contract extension thresholds characterize the optimal policy as stated in the following theorem.

Theorem 4 *The optimal end-of-life inventory control policy when customers are differentiated based on service contracts is characterized by time-varying thresholds as follows:*

- i. There exists threshold levels $r_1(t), r_2(t), \dots, r_S(t)$ such that if $x \geq r_s(t)$ an incoming demand of class s is satisfied and rejected otherwise.*
- ii. There exists threshold levels $e_1(t), e_2(t), \dots, e_S(t)$ such that if $x \geq e_s(t)$ a contract of type s is extended and expired otherwise at time $t \in T$.*
- iii. At each time t we have $0 = r_1(t) \leq r_2(t) \leq \dots \leq r_S(t)$*
- iv. At each time $t \in T$ we have $0 = e_1(t) \leq e_2(t) \leq \dots \leq e_S(t)$*

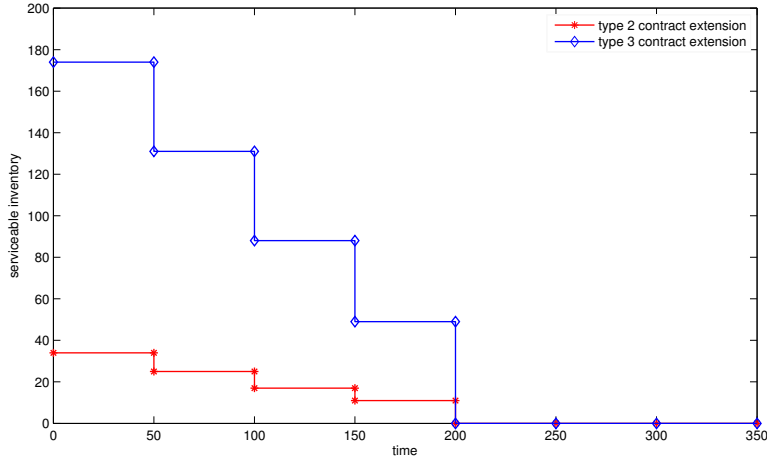


Figure 4: contract extension thresholds for a system with three classes of customers

v. $r_s(t)$ is non-increasing in t , $s \in \{1, 2, \dots, S\}$

vi. $e_s(t)$ is non-increasing in t , $s \in \{1, 2, \dots, S\}$

Proof. The proof is similar to the proof of theorem 2. □

In this setting, the optimal policy can be characterized by a set of rationing and a set of contract extension thresholds. The contract extension thresholds are shown in figure 4 for an instance of the problem. It is assumed that at every 50 periods the system has to revisit the decision of contract extensions. If at time t the serviceable inventory is above the blue contour and $s = (1, 1, 1)$ then the contracts of type 2 and 3 are extended and if the level of serviceable inventory is below the blue threshold then the contracts of type 3 should not be extended any more and s becomes $(1, 1, 0)$. The red contour represents the contract extension threshold for contracts of type 2. If at any time t serviceable inventory on-hand falls below this contour then the system does not offer contract renewal to both types 2 and 3. We observe that when it becomes closer to the end of the horizon, the system decides to stop extending contracts if the level of serviceable inventory is very low. In other words, the contract extension thresholds are non-increasing in time.

Moreover, the rationing thresholds are used to decide whether to service or decline a customer of type s when the corresponding service contract is still active. These rationing thresholds are also

dependent on the remaining time to the end of the horizon and the level of serviceable inventory and show similar behavior as in the previous model.

5.1 Numerical Analysis

The cost parameters are considered similar to the customer differentiation based on equipment criticality. Every 50 periods the system should decide which contracts to renew. The revenues associated with the extension of different service contracts are $R_1 = 150$, $R_2 = 100$ and $R_3 = 50$. In order to study the value of having a contract extension decision included in the model we consider a similar system but without contract extension decision in which service contracts are always extended. If we denote the expected total cost of the optimal policy by C_{opt} and that of the policy without contract extension decision with C_{nce} then the value of having a contract extension decision in the model can be calculated according to

$$\Delta\% = \frac{C_{nce} - C_{opt}}{C_{opt}} \times 100$$

As we observe in figure 6, with the increase of the holding cost rate the value of contract extension decision increases significantly. It is because, when the items are more expensive to hold the system decides to keep less items available in stock and resorts to not extending less profitable contracts when inventory level diminishes. Therefore, having a contract extension decision is more valuable for higher values of holding cost rate. Moreover, having a contract extension decision leads to setting a lower final order quantity inherently and together with rationing can be considered as tools to tame the risk of obsolescence. In figure 6 we also observe a jump down in the final order quantity diagram. This point represents the value of holding cost rates in which the system decides not to extend some contracts at the beginning of the final phase. Therefore, there is a step down in the size of the final order quantity.

In order to study the effect of R_i and c_i on the value of service contract extension decision, we define a coefficient k . Then all R_i and c_i are multiplied by k and the results are shown in figures

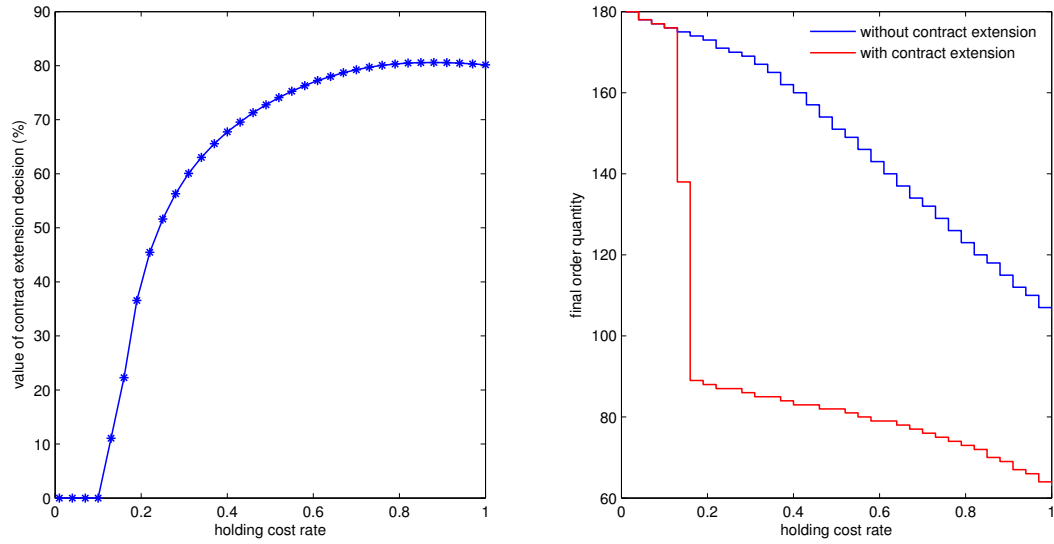


Figure 5: value of contract extension decision and its effect on final order quantity

6 and 7. Figure 6 shows that with the increase of R_i the contract extension decision loses its value due to the fact that the system tends to extend all contracts. As a result, the system places a larger final order quantity in order to assure sufficient serviceable items available during the course of the final phase. As intuition dictates, the final order quantity does not show any sensitivity to the service contract revenues where the system always extends the service contracts.

Figure 7 shows the value of having a contract extension decision with respect to different values of lost sale cost. With the increase of the lost sale costs, the value of having a contract extension decision first intuitively increases and then starts to decrease slowly. The reason for the contracts extension decision losing its value is that the shortage cost becomes so expensive that the system decides not to extend low priority contracts very early in the final phase. As a result, less contract extension decisions are made during the course of the final phase and therefore, the value of contract extension decision diminishes slowly with the increase of shortage costs. Moreover, as we expect with the increase of the lost sale cost the system places a higher final order quantity to avoid the shortage risk.

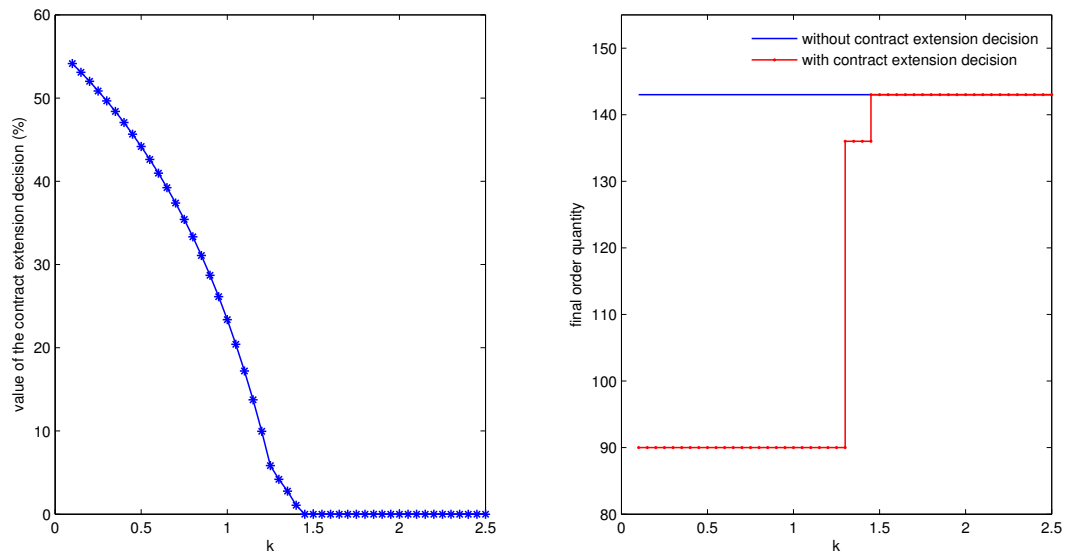


Figure 6: value of contract extension decision w.r.t. to R_i

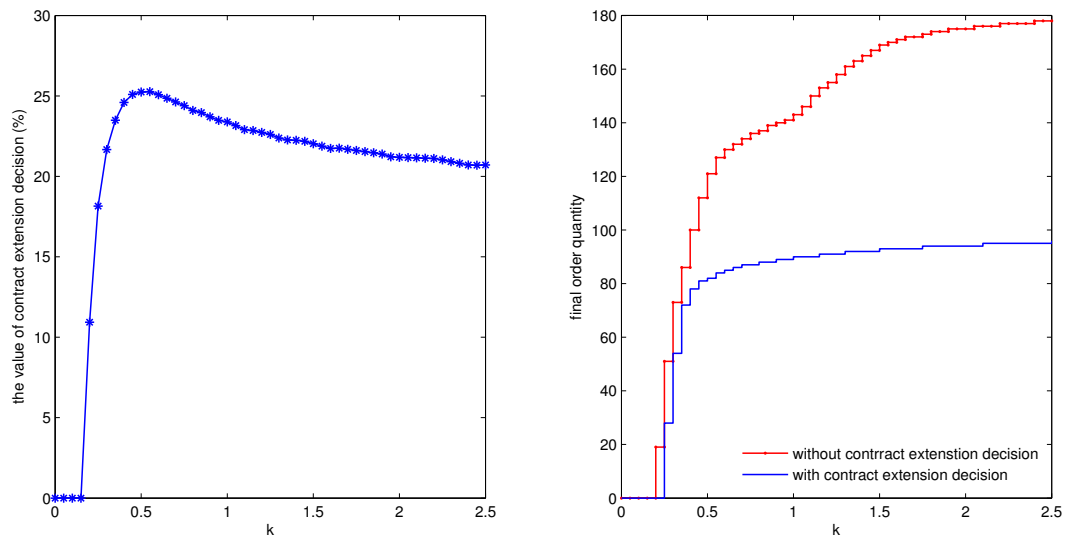


Figure 7: value of contract extension decision w.r.t. to c_i

5.2 Service Contracts for Multiple Parts

In practice, service contracts might apply to a certain equipment. For example, OEM agrees a certain uptime for a machine or is obliged to meet a specific availability for a specific machine service parts. In such a situation, the system manager needs to deal with a multiple part problem. Accordingly, while deciding upon extension of a specific contract we need to take into account the level of serviceable inventory on-hand for all parts. In other words, the system might resort to terminate a low priority contract because of low inventory level of some parts even though the on-hand stock of some other service parts abounds. We assume that the equipment considered has M different parts each one cost h_m , $m \in \{1, 2, \dots, M\}$ to hold per unit per time. λ_i^m denotes the demand intensity function of service contract of type i for part m . The state of the system is denoted by a m -dimensional vector, \mathbf{x} where the m -th element represent the inventory level of m -th part. The rest of the notations is similar to the previous model. In this setting the optimality equation is denoted by $\nu(t, \mathbf{x}, \mathbf{s})$ and is expressed according to (16), which can be obtained following the same logic as the previous case.

$$\begin{aligned}
\nu(t, \mathbf{x}, \mathbf{s}) &= \min_{1 \leq j \leq \phi(\mathbf{s})} \left\{ (1 - \sum_{m=1}^M \sum_{i=1}^j \lambda_i^m(t)) \left\{ \sum_{m=1}^M h_m x_m - \sum_{k=1}^j R_k + \alpha \nu(t+1, \mathbf{x}, \mathbf{s} - \sum_{h=j+1}^{\phi(\mathbf{s})} \mathbf{e}_h) \right\} \right. \\
&\quad + \sum_{m=1}^M \sum_{i=1}^j \left\{ \lambda_i^m(t) \left\{ h_m (x_m - 1) - \sum_{k=1}^j R_k \right. \right. \\
&\quad \left. \left. + \min \left\{ \alpha \nu(t+1, \mathbf{x}, \mathbf{s} - \sum_{h_m=j+1}^{\phi(\mathbf{s})} \mathbf{e}_h) + c_i + h_m, \alpha \nu(t+1, \mathbf{x} - \mathbf{e}_m, \mathbf{s} - \sum_{h=j+1}^{\phi(\mathbf{s})} \mathbf{e}_h) \right\} \right\} \mathbf{1}_{\{x_m > 0\}} \right. \\
&\quad \left. + \left\{ - \sum_{k=1}^j R_k + \alpha \nu(t+1, \mathbf{x}, \mathbf{s} - \sum_{h_m=j+1}^{\phi(\mathbf{s})} \mathbf{e}_h) + c_i + h_m \right\} \mathbf{1}_{\{x_m = 0\}} \right\} \\
&\hspace{15em} (16)
\end{aligned}$$

In this case it can be shown that at each time t the inventory should be controlled according to threshold levels $r_i^m(t)$ and $e_i(t)$, as rationing and contract extension thresholds, respectively. $r_i^m(t)$ and $e_i(t)$ are defined similar to relations (13) and (14).

Figure 8 shows an instance of contract extension thresholds at times 0 and 100. As it is observed at each time the state space is divided into three regions. When the inventory levels of both parts are low the optimal decision is just to extend the highest priority contract type. In rather higher

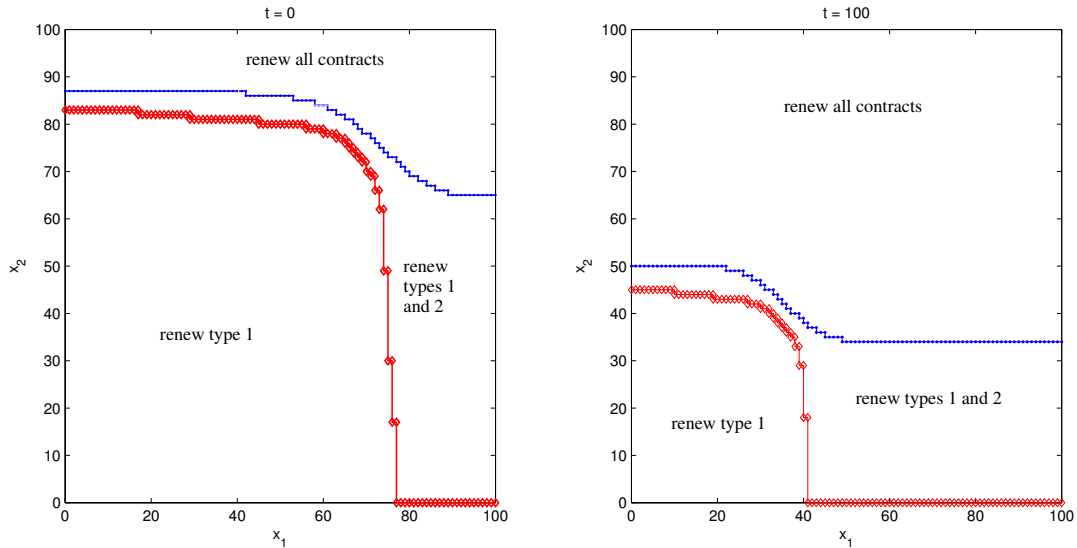


Figure 8: instances of contract extension thresholds

inventory levels the second highest priority contract is also renewed. In higher inventory levels all types of contracts are offered an extension. Furthermore, when it becomes closer to the end of the horizon, these contract extension thresholds are set lower. In other words, it is observed that contract extension thresholds are non-increasing in x_1 and x_2 .

6 Conclusion

In this paper we study the end-of-life inventory problem while customers are differentiated. Nowadays, technology development happens with a fast pace that leads to shorter product life cycle. As a result, parts and products enter their final phase earlier and system managers need to make a final order decision to mitigate the risk of obsolescence at the end of this phase. Customer differentiation has become a common practice in order to fulfill customer demands more efficiently. With the shortened product life cycle we expect a prominent share of service obligations be met while the production of parts is discontinued. Therefore, considering customer differentiation in the final phase can be of vital importance. We consider customer differentiation for both demand-criticality and service contracts-based differentiation. We show that inventory should be optimally

controlled according to threshold levels that consider both remaining time to the end of the horizon and the level of serviceable inventory. We show that considering these decisions in the final phase inventory management leads to remarkable cost improvements as well as obsolescence risk mitigation.

References

- [1] Aberdeen Group. 2005. Service Chain Management. Aberdeen Group. Available from: <http://www.aberdeen.com>.
- [2] Benjaafar, S., M. ElHafsi, C.Y. Lee, W. Zhou. 2010. Optimal control of assembly systems with multiple stages and multiple demand classes. *Operations Research*
- [3] Blischke W.R., D. Murthy. 1994. Warranty Cost Analysis. Marcel Dekker, Inc., New York.
- [4] Bradley, J.R. , H.H. Guerrero. 2009. Life-Time Buy Decisions with Multiple Parts. *Production and Operations Management*, 18(1).
- [5] Bulkeley, W.M. . IBM. 1999. Had 98 PC pretax loss of nearly \$ 1 billion, The Wall Street Journal, March 25.
- [6] Cattani, K.D., G.C. Souza. 2002. Inventory rationing and shipment flexibility alternatives for direct market firms. *Production and Operations Management* 11(4) 441-457.
- [7] Cattani K.D. and Souza G.C. 2003. Good buy? Delaying end-of-life purchases. *European Journal of Operational Research*, 146 216-228.
- [8] Cohen M.A., N. Agrawal, V. Agrawal. 2006. Winning in the Aftermarket. Harvard Business Review. 84(5) 129-138.
- [9] Dekker, R., R.M. Hill, M.J. Kleijn, R.H. Teunter. 2002. On the (s-1; s) lost sales inventory model with priority demand classes. *Naval Research Logistics*. 49 593-610

- [10] Dekker, R., M. J. Kleijn, P. J. de Rooij. 1998. A spare parts stocking policy based on equipment criticality. *International Journal of Production Economics*. 56 69-77.
- [11] Deshpande, V., M.A. Cohen, K. Donohue. 2003a. An empirical study of service differentiation for weapon system service parts. *Operations Research* 51(4) 518-530.
- [12] Deshpande, V., M.A. Cohen, K. Donohue. 2003b. A threshold inventory rationing policy for service-differentiated demand classes. *Management Science* 49(6) 683-703.
- [13] Fortuin, L. 1981. Reduction of All-time requirements for Spare Parts. *International Journal of Operations and Production Management*, 2(1), 29-37.
- [14] Fortuin, L. 1980. The All-Time Requirement of Spare Parts for Service After Sales-Theoretical Analysis and Practical Results, *International Journal of Operations and Production Management*, 1(1), 59-70.
- [15] Frank, K., R.Q. Zhang, I. Duenyas. 2003. Optimal policies for inventory systems with priority demand classes, *Operations Research*. 51(6) 993-1002.
- [16] Hong, J.S. , H.Y. Koo, C.S. Lee, J. Ahn. 2008. Forecasting service parts demand for a discontinued product, *IIE Transactions* 40 640-649.
- [17] Klein Haneveld, W.K. , R.H. Teunter. 1998. The Final Order Problem, *European Journal of Operational Research*, 107 35-44.
- [18] Kooten, J.P.J. van, T. Tan. 2008. The Final Order Problem for Repairable Spare Parts under Condemnation, *Journal of Operational Research Society*.
- [19] Kranenburg, A.A., G.J. Van Houtum. 2007. Cost optimization in the (s-1,s) lost sales inventory model with multiple demand classes, *OR Letters* 35 493-502.
- [20] Kranenburg, A.A., G.J. Van Houtum. 2008. Service differentiation in spare parts inventory management, *Journal of the Operational Research Society* 59 946-955.

- [21] Krikke H.R. and Laan E. van der. 2011. Last Time Buy and Control Policies With Phase-Out Returns: A Case Study in Plant Control Systems, *International Journal of Production Research*.
- [22] Li H, S. C. Graves, D. B. Rosenfield. 2010. Optimal Planning Quantities for Product Transition, *Production and Operations Management*, 19(2) 142-155.
- [23] Möllering, K. T., U.W. Thonemann. 2008. An optimal critical level policy for inventory systems with two demand classes, *Naval Research Logistics* 55 632-642.
- [24] Moore, J.R. 1971. Forecasting and Scheduling for Past-Model Replacement Parts, *Management Science* 18 B200-B213.
- [25] Murthy D., W.R. Blischke. 1995. Warranty management and Product Manufacture. Springer, New York.
- [26] Teunter, R.H. , L. Fortuin. 1998. End-of-life service: A case study, *European Journal of Operational Research*, 107 19-34.
- [27] Teunter, R.H. , W.K. Klein Haneveld. 2002. Inventory control of service parts in the final phase, *European Journal of Operational Research* 137 497-511.
- [28] Pourakbar, M. Frenk J.B.G., Dekker, R. 2010. End-of-life inventory decisions for consumer electronics service parts, econometric institute report series, Erasmus university Rotterdam.
- [29] Pourakbar, M. Laan, E. van der, Dekker, R. 2011. End-of-life inventory Problem with Phase-out Returns, econometric institute report series, Erasmus university Rotterdam.
- [30] Ritchie, E., P.Wilcox. 1977. Renewal Theory Forecasting for Stock Control, *Journal of operational Research Society* 1 90-93.
- [31] Teunter, R.H., W.K. Klein Haneveld. 2008. Dynamic inventory rationing strategies for inventory systems with two demand classes, poisson demand and backordering, *European Journal of Operational Research* 190 156 178.

- [32] Topkis, D.M. 1968. Optimal ordering and rationing policies in a non-stationary dynamic inventory model with n demand classes, *Management Science* 15(8) 160-176.
- [33] Veinott, A.F. 1965. Optimal policy in a dynamic, single product, non-stationary inventory model with several demand classes, *Operation Research* 13 761-778.
- [34] De Vericourt, F., F. Karaesmen, Y. Dallery. 2002. Optimal stock allocation for a capacitated supply system, *Management Science*. 48(11) 1486-1501.