2014

Joint Optimization of Asset and Inventory Management in a Product–Service System

Xiang Wu
Iowa State University

Sarah M. Ryan
Iowa State University, smryan@iastate.edu

Joint Optimization of Asset and Inventory Management in a Product–Service System

Abstract

We propose an integrated model of the asset management decisions for a fleet of identical product units and the inventory management decisions for a closed-loop supply chain in the context of a product-service system, in which the two types of decisions are closely coupled. A joint optimization technique is developed to obtain the parameters of the operational policy for the integrated model that minimize the long run average cost rate. A numerical example is provided to illustrate the computational procedures. In addition, the effect of a simplifying assumption that the replaced products have no quality difference is evaluated and the results suggest that as long as the quality difference between the preventively replaced products and failure replaced products is not too big, the simplification to treat them as one category is reasonable.

Keywords

decision support systems, closed-loop supply chain, computational procedures, integrated modeling, inventory management, long-run average cost, management decisions, product-services systems, inventory control

Disciplines

Industrial Engineering | Systems Engineering

Comments

The Version of Record of this manuscript has been published and is available in The Engineering Economist 2014, http://www.tandfonline/ 10.1080/0013791X.2013.873844. Posted with permission.

JOINT OPTIMIZATION OF ASSET AND INVENTORY MANAGEMENT IN A PRODUCT-SERVICE SYSTEM

ABSTRACT

We propose an integrated model of the asset management decisions for a fleet of products and the inventory management decisions for a closed-loop supply chain in the context of a product-service system, in which the two types of decisions are closely coupled. A joint optimization technique is developed to obtain the parameters of the operational policy for the integrated model that minimize the long run average cost per unit time. A numerical example is provided to illustrate the computational procedures. In addition, the effect of a simplifying assumption that the replaced products have no quality difference is evaluated and the results suggest that as long as the quality difference between the preventively replaced products and failure replaced products is not too big, the simplification to treat them as one category is reasonable.

Keywords: Joint optimization, product-service system, preventive maintenance, inventory management, closed-loop supply chain

1 INTRODUCTION

A product-service system (PSS), or servicizing[1], is a strategy in which producers provide the use as well as the maintenance of products while retaining ownership. Prospective customers who become the clients pay fees for receiving the services or functions of products rather than purchasing them, and so are free of the risk, responsibility and cost burdens that are commonly associated with ownership. Since the introduction of this attractive concept in 1999 (Goedkoop et al., 1999, White et al., 1999), a diverse range of PSS examples in the literature have demonstrated its economic success, but most have tended

[1] We use the terms "servicizing" and "PSS" interchangeably.

to emphasize its significant environmental benefits and social gains (Luiten et al., 2001, Manzini et al., 2001). Although a variety of tools and methodologies have been developed for designing a servicizing system, such as those in Manzini et al., 2001), Maxwell and van der Vorst, 2003), and Van Halen et al., 2004), how to effectively structure an organization to be competent at designing, making and delivering PSS is still difficult (Baines and Lightfoot, 2007). Most literature in this area provides qualitative description and analysis of servicizing. There is a lack of in-depth and rigorous research to develop models, methods and theories; to assess the implications of competitiveness; and to help manufacturers configure their products, technologies, operations, and supply chain (Baines and Lightfoot, 2007).

The motivation for this research is to improve the economic viability of PSS. The economic impacts may include:

- Lower overall investment in equipment because incentive realignment reduces the manufacturer’s motive to sell more units and because existing units may be redeployed rather than sitting idle;
- A shift from manufacturing to remanufacturing with increased modularity of design allowing easy upgrade;
- Higher operational and maintenance costs due to more intensive equipment usage; and
- Increased investment in sensing and monitoring equipment to mitigate the risk to the manufacturer of equipment abuse by the customer.

We analyze and model the operation of the PSS and identify optimal parameters of an operational policy to minimize the long-run average replacement cost and inventory management cost incurred in the PSS when condition monitoring can drive replacement decisions.

We focus on “customer-user servicizing,” in which the customer operates the equipment while the manufacturer retains responsibility for design, manufacturing, ownership, maintenance, repair, redeployment, reclamation and disposal (Toffel, 2008). The service contracts usually include replacement of the initial machines with newer or better ones, and the machines coming off lease are remanufactured extensively (Thierry et al., 1995). Service providers must balance the cost of built-in durability and reusability against the lifecycle cost savings, choose when to take old products out of service, and decide whether to remanufacture them or to replace them with newly manufactured products. Servicizing motivates the use of condition monitoring; i.e., using sensors, information and communication technology to increase visibility of the product's condition and performance in the field, so as to improve asset utilization and make better maintenance decisions (Baines and Lightfoot, 2007). Under servicizing, the remanufacturing facilities frequently operate together with a manufacturing plant to satisfy the demand. Such systems are known as hybrid manufacturing and remanufacturing systems, and involve both forward and reverse flows of products.

For the service paradigm to be viable from the provider's perspective, the fee for service must allow for a profit margin over the cost of providing the service. The cost of service provision depends largely on the ability to manage and maintain products effectively in a closed-loop system. In particular, manufacturers who servicize must engage in reverse as well as forward logistics; and in addition, they must make maintenance decisions for their products. Unlike the common closed-loop supply chain for sold products, a distinctive feature of the closed-loop supply chain in PSS is that the demands are driven by maintenance actions on the products and/or a capacity expansion requirement, and the returns are generated by out-of-service products, replaced either preventively or due to failure. In other words, the demands and returns are controllable by the servicizing manufacturers via the maintenance decision. On one hand, replacement costs are affected by the inventory management policy; on the other hand, the inventory management cost is affected by maintenance decisions. Therefore, the maintenance decisions are closely coupled with the inventory management decisions of this closed-loop supply chain. This coupling makes the decision making under servicizing significantly more complicated than that under traditional product sales.

Maintenance policies for deteriorating systems have been studied extensively for decades (Aven and Bergman, 1986, Lam and Yeh, 1994, Liu et al., 2010, Giorgio et al., 2011). The recent research effort has been focused on the problem of optimal replacement when certain condition information about the system is available (which is often the case in PSS), such as temperature, humidity, vibration levels, or the amount of metal particles in a lubricant (Banjevic et al., 2001, Ghasemi et al., 2007, Kharoufeh et al., 2010, Wu and Ryan, 2010). Some authors have also studied the effect of technological advances on the optimal lifetime and the optimal replacement of the assets (Yatsenko and Hritonenko, 2009, Mardinab and Araib, 2012). As to the inventory management policies, a rapidly growing body of research in the operations management of closed-loop supply chains recognizes and tries to mitigate the complexities of managing the supply chain involving remanufacturable products under traditional product sales (Fleischmann et al., 1997, Guide, 2000, Aras et al., 2004, Guide and Van Wassenhove, 2009).

However, little research has been done to consider the joint optimization of the maintenance policy and the closed-loop supply chain inventory management, which is needed to develop optimal decisions in the context of PSS. Some relevant work appears in the context of production inventory systems. For example, Das and Sarkar, 1999) considered the optimal maintenance policies for a production inventory system where inventory is controlled according to an (S, s) policy. Rezg et al., 2004) studied the joint optimization problem of preventive maintenance and inventory control in a production line using simulation, and proposed a methodology combining simulation with genetic algorithms to obtain the optimal policy. In those cases, the maintenance is applied to the machines in the production line, rather than the service products in the fleet under PSS. Thus, their problems differ in nature from the one in PSS.

More relevant work in the existing literature investigates the joint optimization of maintenance and inventory policies for deteriorating systems with spare-part inventory. In particular, Armstrong and Atkins, 1996) examined the age replacement and ordering decisions for a system subject to random failure and with room for only one spare in stock, and several extensions have been made to generalize the cost terms and the order lead time in their subsequent paper (Armstrong and Atkins, 1998). Brezavscek and Hudoklin, 2003) considered the problem of joint optimization of block replacement and periodic review spare-provisioning policy for deteriorating systems to minimize the expected total cost per unit time. A joint condition-based maintenance and spare part inventory control strategy is suggested by Rausch and Liao, 2010) to minimize the expected total operating cost with constraints on stockout probability, production lot size and due date. Berthaut et al., 2011) proposed a modified block replacement policy combining with the hedging point inventory control policy for a failure-prone single machine and confirmed the superiority of their proposed model. Still, those studies differ fundamentally from the one we conduct in this paper, because they do not involve the production process of the spare parts.

In this work, we present an integrated model that takes into account both the maintenance decisions and the inventory management decisions of a closed-loop supply chain in the context of a product-service system to minimize the total cost per unit time. For maintenance, we consider a condition-based replacement policy that uses the proportional hazards model (PHM) with a semi-Markovian covariate process to model the degradation of the products (Wu and Ryan, 2011). For inventory management of the closed-loop supply chain, a continuous review base stock policy is adopted due to its easy implementation and proven effectiveness in practice. Identifying and formulating the couplings between asset and inventory management in this context, we develop an optimization technique to obtain the optimal parameters for the two policies simultaneously in the integrated model.

This paper is organized as follows. Section 3 presents the development and mathematical formulation of the integrated model. In section 4, an optimization technique is developed and a two-step algorithm is presented to obtain the optimal policy parameters and cost. This is followed by a numerical example in Section 5 to illustrate the computational procedures of the optimization algorithm. Section 6 revisits the single return category assumption and evaluates its impact on the optimal cost by comparing with the analysis of a system with two categories of returns. Section 7 concludes with a discussion of future research directions.

2 System Description

We assume the service provider has a fleet of N identical products in service. The objective is to develop an operational policy for the PSS to keep every product in the fleet in working condition at all times with minimum cost. The products deteriorate with age and operation, and are subject to failure. When a product is preventively replaced or fails, it is collected for remanufacturing and replaced by a new product. The output of the remanufacturing facility may not be able to fulfill all the demand for new products. We assume a manufacturing facility exists with sufficient capacity to cover any unsatisfied demand.

The PSS consists of two subsystems: a service subsystem (SS) and a remanufacturing subsystem (RS) which is supplemented as needed by the manufacturing facility. The service subsystem employs products to provide service to clients and sends replaced products to the remanufacturing subsystem. The (hybrid) remanufacturing subsystem satisfies the demands of the service subsystem for replacement products through remanufacturing or manufacturing.

The flow of products through the whole system is depicted in Figure 1. In the SS, the operational conditions of the products are continuously monitored and a condition-based replacement policy is applied to each product independently. The demand of the remanufacturing system is driven by the replacement of products in the fleet. When a product is taken out of service due to preventive replacement or failure, it is replaced immediately with either a remanufactured or a newly manufactured product.

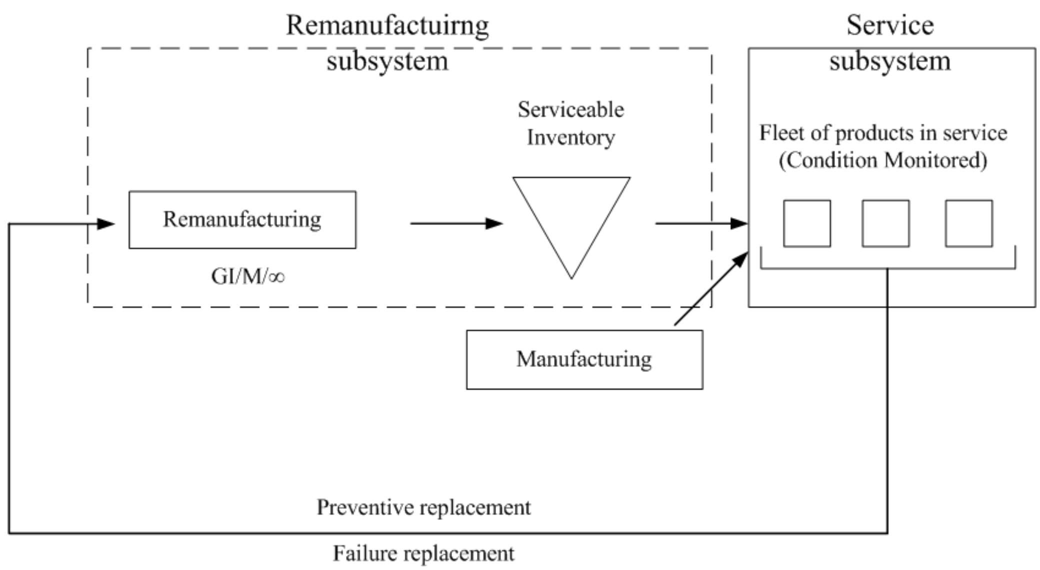

Figure 1 Product flow through the whole system

The RS is a remanufacturing facility that replenishes serviceable inventory. The replaced products directly go to the remanufacturing process if needed to maintain a base stock level; otherwise, they are discarded to save storage costs. We focus on the remanufacturing facility and do not represent the manufacturing plant in detail. Priority is given to remanufactured products when satisfying demand. Based on the conventional wisdom that remanufacturing is cheaper than manufacturing, newly manufactured products are viable only when the serviceable inventory is unable to fulfill the demand (i.e., is empty). . We assume manufactured products are available whenever necessary.

The goal of the study is to investigate the replacement problem of the SS and the inventory management of the RS jointly in the context of PSS. Considering the coupling between two subsystems, an integrated model is built to address the uncertainties residing not only in the replacement problem but also in the inventory management, and a joint operational policy is developed to minimize the long-run average cost incurred in the whole system per unit time. In what follows, we shall first introduce the replacement policy for the SS and the inventory management policy for the RS respectively, and then present the integrated model. First, we summarize the notations and assumptions used in this paper here.

2.1 Notation

Input parameters

N : Number of products in the fleet.

C_1 : The cost of preventive replacement with a remanufactured product in the SS, which is also the unit remanufacture (production) cost in the RS; $C_1 > 0$.

C_2 : The cost of preventive replacement with a manufactured product in the SS, which is also the aggregate acquisition cost for a newly manufactured product in the RS; $C_2 > C_1$.

K : The additional cost for a failure replacement; $K > 0$.

$Z = \{Z_t, t \geq 0\}$: A continuous semi-Markov process with a finite state space $S = \{0,1,...,n-1\}$ and $Z_0 \equiv 0$, which depicts the evolution of the working condition of a product.

$h_0(t)$: The baseline hazard rate, which depends only on the age of the product.

$\Psi(\cdot)$: A link function; $\Psi : S \mapsto \Re$.

T : The time to failure of the product.

μ : Processing rate for remanufacturing.

h_S : Unit serviceable inventory holding cost.

h_W : Unit remanufacturing work in process (WIP) holding cost.

Internal variables

$\delta = \{t_0, t_1, \cdots, t_{n-1}\}$: A replacement policy which replaces at failure or at age t_i when in state $i \in S$, whichever occurs first.

$M(\delta)$: The expected length of a replacement cycle under policy δ .

$Q(\delta)$: The probability of failure under policy δ .

$I(t)$: The number of products in the serviceable inventory at time t .

$W(t)$: The number of products in WIP at time t .

c : Base stock level of the serviceable inventory position.

p_L : The proportion of time that the serviceable inventory is empty.

Output variables

δ^* : Optimal replacement policy

c^* : Optimal base stock level

2.2 Assumptions

1 The service provider has a large fleet of identical products in service and maintains an inventory of serviceable products to satisfy the demand for replacements.

2 Manufactured and remanufactured products are perfectly substitutable; that is, remanufactured products are considered as good as new.

3 Setup cost for remanufacturing is negligible and there is no holding cost associated with the remanufacturable inventory.

4 Remanufacturing capacity is unlimited and the time required to remanufacture a replaced product is exponentially distributed with rate μ.

5 The newly manufactured products are always available and there is no lead time for acquiring one.

6 We consider the replaced products as one category, whether they are replaced preventively or due to failure. In section 6, we will re-evaluate this assumption.

7 The baseline hazard rate, $h_0(t)$, is strictly increasing with the product age and unbounded as the age approaches infinity; that is, the product deteriorates with time. In addition, $h_0(0)=0$.

8 The covariate process Z changes state according to a pure birth process; i.e., whenever a transition occurs, the state of the process always increases by one, and state $n-1$ is absorbing.

9 The link function, $\Psi(\cdot)$, is non-decreasing with $\Psi(0)\equiv 1$.

10 The fleet of products must be kept in working order at all times. Replacement is instantaneous.

3 MODEL DEVELOPMENT AND FORMULATION

3.1 Replacement policy for the service subsystem

In a PSS, the service provider retains ownership and maintains direct access to its products. This allows it to continuously collect data on the condition of products in service using condition monitoring technologies. Such data can help the service provider to improve the performance of products, lower failure probability, improve asset utilization and so reduce the total cost. For systems under continuous monitoring, a condition-based maintenance policy is natural.

Assume that replacement is the only maintenance option in our PSS setting. The condition-based replacement policy developed in Wu and Ryan, 2011) well suits the service subsystem, where the PHM (Cox and Oakes, 1984) is employed to account for the impact of dynamic working conditions on the failure process of the system. Herein we adopt the policy described in Wu and Ryan, 2011) as the replacement policy for the SS. Because the replacement policy is applied to each product independently, we first consider the replacement policy for a single product.

We assume that $Z = \{Z_t, t \geq 0\}$ is a continuous-time semi-Markov covariate process that depicts the evolution of the working condition of the product, and is under continuous monitoring. Under the proportional hazards model, the hazard rate of the product at time t is expressed as

$$h(t) \equiv h_0(t)\Psi(Z_t), \quad t \geq 0 \tag{1}$$

Denote the replacement policy as $\delta = \{t_0, t_1, \cdots, t_{n-1}\}$, $t_0 \geq t_1 \geq \cdots \geq t_{n-1} \geq 0$, where t_i is the threshold age for replacement if the covariate process of the product is in state i. According to renewal theory (Ross, 2003), the long run average replacement cost per unit time for a single product can be expressed as the ratio of the expected cost per replacement cycle to the expected length of a replacement cycle, which is given by

$$\phi_R = \frac{(C_1 + KQ(\delta))(1 - p_L) + (C_2 + KQ(\delta))\,p_L}{M(\delta)} = \frac{C_1 + (C_2 - C_1)\,p_L + KQ(\delta)}{M(\delta)} \tag{2}$$

Here p_L is the proportion of products in the fleet replaced with manufactured products, which will be discussed further in the context of the remanufacturing subsystem. The explicit expressions for the expected length of a replacement cycle, $M(\delta)$, and the failure probability, $Q(\delta)$, in terms of $t_0, t_1, \cdots, t_{n-1}$ given $Z(t)$, $h_0(t)$ and $\Psi(\cdot)$ can be found using the method described in Wu and Ryan, 2011). We detail those expressions for a three-state Z process and their partial derivatives with respect to t_i in Appendix A for convenience. Obviously, $M(\delta)$ is an increasing function of the threshold age for replacement t_i, $\forall i$.

3.2 Inventory policy for the remanufacturing subsystem

For systems involving remanufacturing, two inventory control strategies are generally applied: "push" and "pull". Under the push strategy, the returned products are batched and pushed into the remanufacturing process as soon as the remanufacturable inventory has sufficient products. Under the pull strategy, the timing of the remanufacturing process depends on the demands as well as inventory positions. Van der Laan et al., 1999) shows that pull control is preferable if the holding cost in remanufacturable inventory is lower than the holding cost in the serviceable inventory, which is true in most practical situations.

Based on above findings, the inventory in the remanufacturing subsystem is assumed to be managed by a continuous review base stock policy. This policy aims at keeping the serviceable inventory position at a base stock level c at all times, which is achieved by pulling returned products into the remanufacturing process each time a demand is served from the serviceable inventory. The serviceable inventory position at time t includes the on-hand serviceable inventory $I(t)$ and the work in process (WIP) of remanufacturing $W(t)$. Thus we have

$$I(t)+W(t)=c \qquad \forall t \geq 0. \tag{3}$$

The policy is easy to implement and is efficient when the setup cost for remanufacturing is negligible, which we assume is true in our PSS setting.

Under the continuous review base stock policy, the remanufacturing subsystem is a pull system. An execution flowchart is shown in Figure 2. Priority is given to remanufactured products when satisfying demand. Following the flowchart, we can see that p_L defined in Section 3.1 equals the probability that the serviceable inventory is empty; i.e., $P(I(t)=0)$.

The cost C_1 defined in Section 3.1 is equivalent to the unit remanufacturing cost, and C_2 is equivalent to the unit acquisition cost for a manufactured product, which is incurred every time we resort to manufacturing.

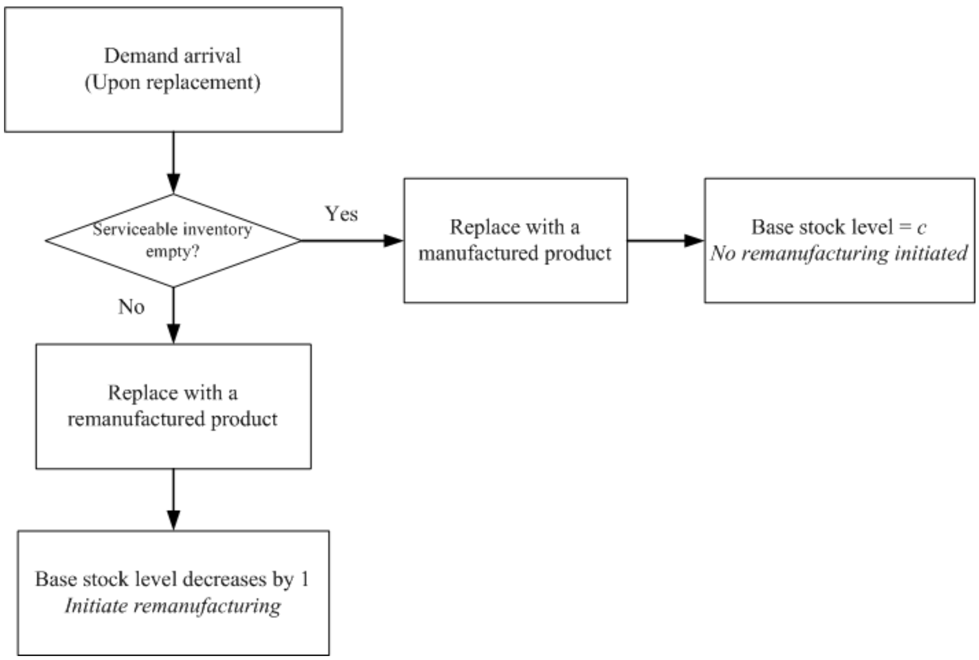

Figure 2 Flowchart of the remanufacturing subsystem

For the serviceable inventory and remanufacturing WIP, we adopt a similar holding cost structure to that in Aras et al., 2004). The unit holding cost rate for serviceable inventory is $h_S = h + \alpha C_1$, and the unit holding cost rate for remanufacturing WIP is $h_W = h + \beta\alpha C_1$ ($\beta < 1$). Here, h denotes the basic holding cost and α denotes the opportunity cost of capital. WIP is considered to have approximately 100β% value added and the serviceable inventory has all the value added.

In addition, we assume there is no capacity limitation on the remanufacturing process, so it could be modeled as an infinite-server station. The time for remanufacturing is highly variable due to various conditions of the replaced products. Thus, the service time of each server is assumed to be exponentially distributed with rate μ. In fact, since the WIP in the remanufacturing subsystem is bounded by the base stock level c, only c servers are needed to avoid blocking in the remanufacturing process, and the subsystem can achieve steady state. Then the loss probability, p_L is the limiting value

$$p_L = \lim_{t\to\infty} P\left(I\left(t\right) = 0\right) \tag{4}$$

and the long run average cost incurred in the remanufacturing subsystem is given by

$$\phi_I = \lim_{t\to\infty} h_S E\left(I\left(t\right)\right) + h_W E\left(W\left(t\right)\right) \tag{5}$$

In the following, we assume steady state and suppress the t in the notation for $I(t)$ and $W(t)$. Only inventory costs are considered for the RS because we already account for the costs C_1 and C_2 in the SS.

We consider the returns as a single category, regardless of whether they are preventively replaced or replaced due to failure. We do not differentiate the returned products in terms of inventory cost, processing time and cost. Therefore the remanufacturing node together with the serviceable inventory node can be modeled as a single-stage produce-to-stock system with a single product type.

Examining our system carefully, we find that the inventory is virtually controlled by a target-level production authorization mechanism with lost sales as discussed in Buzacott and Shanthikumar, 1993). In our case, the target-level is c. Production authorization is transmitted to the remanufacturing facility when the inventory position falls by one. A "lost sale" occurs when the serviceable inventory is empty, in which situation we resort to manufactured products and no new remanufacturing is authorized. According to Buzacott and Shanthikumar, 1993), the performance of this produce-to-stock system may be obtained from the analysis of a fictitious $G/M/c/c$ queue, also known as $G/M/c$ loss system. The correspondence between our system and the fictitious system is as follows:

- The demand process to the RS is the arrival process to the fictitious queue.
- The probability that a demand is satisfied by manufacturing, p_L, is the loss probability of the fictitious queue.
- The WIP of the RS is the number of products in the fictitious queue.

The demand process of the RS is generated from the replacements of products in the SS. Since the replacement policy is applied to each product independently, the replacement flow of each individual product is a renewal process. The demand process, which is a pool of N such renewal processes, is called a superposed renewal process (SRP) in the literature. In general, if the number of products in a service fleet, N, is sufficiently large, then the superposed renewal process can be approximated by a Poisson process with rate λ (Cinlar and Lewis, 1972). Then the fictitious queue may be approximated by a $M/M/c$ loss system, also known as the Erlang loss system.

For each product, the renewal rate $r = \dfrac{1}{M(\delta)}$. Then the overall arrival rate is

$$\lambda = Nr = \frac{N}{M(\delta)} \tag{6}$$

Let L be the steady-state average number of products in the $M/M/c$ loss system and q_n be the steady-state probability that there are n products in the queue. From the performance of $M/M/c$ loss system in steady state, we have

$$q_0 = 1/\sum_{k=0}^{c}\frac{(\lambda/\mu)^k}{k!}$$
$$q_n = \frac{(\lambda/\mu)^n}{n!}q_0, \quad n = 1, 2, ., c$$
$$L = \frac{\lambda}{\mu}(1 - q_c)$$

For our system,

$$E(W) = L \tag{7}$$

$$E(I) = c - L \tag{8}$$

From equations (5)-(8),

$$\phi_I = h_S c + (h_W - h_S)\frac{\lambda}{\mu}(1 - p_L) \tag{9}$$

where the probability that a demand must be satisfied by manufacturing is

$$p_L \equiv p_L(c,\delta) = \lim_{t\to\infty} P(I(t)=0) = q_c = \frac{(\lambda/\mu)^c}{c!\sum_{k=0}^{c}\frac{(\lambda/\mu)^k}{k!}} \tag{10}$$

with $\lambda = N / M(\delta)$.

We state two important properties of p_L as below.

Lemma 1 $\frac{\partial p_L}{\partial M} < 0$.

Proof.

The loss probability p_L is a increasing function of the arrival rate λ, and thus a decreasing function of $M(\delta)$. Therefore $\frac{\partial p_L}{\partial M} < 0$. ■

Lemma 2 $\frac{\partial p_L}{\partial M}$ is increasing in its parameter $t_i, \forall i$.

Proof.

According to Proposition 3 of Harel, 1990), for fixed number of servers and fixed arrival rate, the Erlang loss formula is strictly convex in service rate. The symmetric positions of μ and $M(\delta)$ in the loss formula (10) imply that p_L is strictly convex in $M(\delta)$ for fixed c and μ. Thus $\frac{\partial^2 p_L}{\partial M^2} > 0$, which means $\frac{\partial p_L}{\partial M}$ is increasing in $M(\delta)$. And since $M(\delta)$ is an increasing function of t_i, $\frac{\partial p_L}{\partial M}$ is increasing in t_i, $\forall i$. ■

3.3 *Integrated model*

The service subsystem (SS) and the remanufacturing subsystem (RS) discussed above are closely coupled in terms of the demands and returns. The RS satisfies the demand of the SS, while the SS generates returns to the RS. A distinctive feature of this closed-loop system is that the demands and the returns are generated simultaneously.

In particular, the decision making process couples the two subsystem. The decision variable in the SS, the replacement policy δ, affects the demand and return flows of the supply chain in the RS and, thus, affects the average cost incurred in the RS. On the other hand, the base stock level c has a direct impact on p_L, the proportion of products replaced with manufactured products and thus, influences the average cost incurred in the SS.

To account for the coupling, we must optimize the decision variables of the two subsystems simultaneously. Treating them separately would lead to an inferior solution.

Integrating the costs of subsystems, the total cost per unit time for the whole system can be expressed as

$$\phi(c,\delta) = \phi_I + N\phi_R = h_S c + N\frac{C_1 + (h_W - h_S)\mu^{-1} + KQ(\delta) + [C_2 - C_1 - (h_W - h_S)\mu^{-1}]p_L}{M(\delta)} \qquad (11)$$

which incorporates the production cost for remanufactured products (C_1 per unit), the acquisition cost for manufactured products (C_2 per unit), additional cost for failure replacement (K per unit) and the inventory holding costs (h_S per unit per unit time for serviceable inventory and h_W per unit per unit time for WIP).

4 OPTIMIZATION TECHNIQUE

Our objective function is the total cost per unit time given by (11). The decision variables are the parameters of the replacement policy $\delta = \{t_0, t_1, \cdots, t_{n-1}\}$ and the base stock level c, where

$t_0 \geq t_1 \geq \cdots \geq t_{n-1} \geq 0$ and c is a positive integer. The key variables are p_L, expressed in (10); and $M(\delta)$ and $Q(\delta)$, whose expressions can be obtained from Appendix A.

The objective function is a complicated function that appears to lack "nice" structure (such as convexity) and closed-form solutions are hard to achieve. To obtain the global minimum, we resort to a special optimization method -- the lambda minimization technique (Aven and Bergman, 1986) which is summarized in Appendix B. For simplicity, we consider a three-state covariate process Z to illustrate the optimization technique, which can be generalized to any number of states.

In what follows, we find the optimal parameters c^* and $\delta^* = \{t_0^*, t_1^*, t_2^*\}$, and the global minimum through a two step process:

1) For a fixed c, we find the optimal $\delta^c = \{t_0^c, t_1^c, t_2^c\}$ to minimize the objective using the lambda minimization technique.
2) From the objective value obtained in last step, we find an upper bound for c. By enumerating from the minimum base stock level ($c = 1$) to the upper bound, we can find the optimal parameters and the global minimum of the objective function.

With c fixed, minimizing (11) is equivalent to minimizing

$$v(\delta) = \frac{\phi(c,\delta) - h_S c}{N} = \frac{b_1 + KQ(\delta) + b_2 p_L}{M(\delta)}. \quad (12)$$

where $b_1 = C_1 + \frac{(h_W - h_S)}{\mu}, b_2 = C_2 - C_1 - \frac{(h_W - h_S)}{\mu} > 0$.

To apply the lambda minimization technique, define the γ-function (analogous to the λ-function in Appendix B) as

$$u(\gamma, \delta) = b_1 + b_2 p_L(c, M(\delta)) + KQ(\delta) - \gamma M(\delta). \quad (13)$$

For a fixed γ, with n = 3 states we have the following optimization problem, where $\delta = \{t_0, t_1, t_2\}$:

$$\min u(\gamma,\delta)$$
$$s.t.\ \ t_0 \ge t_1 \ge t_2 \ge 0$$

Taking partial derivatives of (13) with respect to t_0, t_1, t_2 and setting them to 0, we have

$$\frac{\partial u}{\partial t_i} = b_2 \frac{\partial p_L}{\partial M}\frac{\partial M}{\partial t_i} + K\frac{\partial Q}{\partial t_i} - \gamma\frac{\partial M}{\partial t_i} = 0, \quad i = 0,1,2\,,$$

which is the system of equations that determines the critical point of u. With the partial derivatives of $M(\delta)$ and $Q(\delta)$ developed in Appendix A and equation (24), the above system of equations can be reduced to

$$b_2 \frac{\partial p_L}{\partial M} + K h_0(t_0)\Psi(0) - \gamma = 0 \tag{14}$$

$$b_2 \frac{\partial p_L}{\partial M} + K h_0(t_1)\Psi(1) - \gamma = 0 \tag{15}$$

$$b_2 \frac{\partial p_L}{\partial M} + K h_0(t_2)\Psi(2) - \gamma = 0 \tag{16}$$

From (14)-(16), we have

$$\frac{h_0(t_0)}{h_0(t_1)} = \frac{\Psi(1)}{\Psi(0)} \quad \Rightarrow \quad t_1(t_0) = h_0^{-1}\left[\frac{\Psi(0)}{\Psi(1)} h_0(t_0)\right] \quad \text{and}$$

$$\frac{h_0(t_0)}{h_0(t_2)} = \frac{\Psi(2)}{\Psi(0)} \quad \Rightarrow \quad t_2(t_0) = h_0^{-1}\left[\frac{\Psi(0)}{\Psi(2)} h_0(t_0)\right]$$

Since $h_0(\cdot)$ is monotonically increasing, $t_1(t_0), t_2(t_0)$ are also monotonically increasing functions of t_0. Upon substituting them back into (14), we get an univariate equation in t_0, which is

$$b_2 \frac{\partial p_L}{\partial M}(t_0, t_1(t_0), t_2(t_0)) + K h_0(t_0)\Psi(0) - \gamma = 0 \tag{17}$$

Note, $\frac{\partial p_L}{\partial M}$ is a function of the tuple $\delta = \{t_0, t_1, t_2\}$, which has been suppressed in its notation.

Lemma 3 For a given γ, the multivariate function $u(\gamma,\delta)$ has a unique critical point.

Proof.

From Lemma 1 and Lemma 2, we know that $\frac{\partial p_L}{\partial M}(t_0, t_1(t_0), t_2(t_0))$ is always negative and is a increasing function of t_0. In addition, $h_0(\cdot)$ is an increasing function with $h_0(0)=0$ and is unbounded as its parameter approaches infinity. Thus, the function

$$b_2 \frac{\partial p_L}{\partial M}(t_0, t_1(t_0), t_2(t_0)) + Kh_0(t_0)\Psi(0)$$

equals the positive constant γ at a unique point.

Therefore equation (17) has a unique solution, which means that for a given γ, $u(\gamma,\delta)$ has a unique critical point, denoted as $\delta^{\gamma} = \{t_0^{\gamma}, t_1^{\gamma}, t_2^{\gamma}\}$. ■

The following theorem shows how to find the global minimum of $u(\gamma,\delta)$.

Theorem 1 For a given γ, function $u(\gamma,\delta)$ achieves global minimum at its critical point $\delta^{\gamma} = \{t_0^{\gamma}, t_1^{\gamma}, t_2^{\gamma}\}$.

The proof of Theorem 1 is in Appendix C.

In light of Theorem 1 and the lambda minimization technique, we state the following algorithm to find the optimal δ that minimizes $v(\delta)$ for a given c.

Algorithm I

1. Initialize the iteration counter $m=0$ and $\gamma=\gamma^0$.
2. For γ^m, use Theorem 1 to find $\delta^{c,m} = \{t_0^{c,m}, t_1^{c,m}, t_2^{c,m}\} = \arg\min_{\delta} u(\gamma^m, \delta)$.
3. Use the replacement policy $\delta^{c,m} = \{t_0^{c,m}, t_1^{c,m}, t_2^{c,m}\}$ obtained in step 2 and equation (12) to update $\gamma^{m+1} = v(\delta^{c,m})$.
4. If $\gamma^{m+1} = \gamma^m$, stop with $v^c = \gamma^{m+1}$ and $\delta^c = \delta^{c,m}$; otherwise, set $m \leftarrow m+1$ and go to step 2.

In addition, an upper bound on the optimal stock level c can be obtained. Denote the optimal parameters as $\delta^* = \{t_0^*, t_1^*, t_2^*\}$ and c^*. Let $\phi_0 = \phi(c_0, \delta^{c_0})$, which is the optimal cost when c is fixed at c_0. Then we have

$$\phi_0 \geq \phi(c^*, \delta^*) \geq h_S c^* + N \frac{C_1 + (h_W - h_S)\mu^{-1} + KQ(\delta^*)}{M(\delta^*)} \geq h_S c^* + N \frac{C_1 + (h_W - h_S)\mu^{-1} + KQ(\delta')}{M(\delta')}$$

where the second inequality follows by omitting the p_L term in $\phi(c^*, \delta^*)$, and the third inequality holds if δ' is the replacement policy that minimizes the term

$$\frac{C_1 + (h_W - h_S)\mu^{-1} + KQ(\delta)}{M(\delta)}.$$

Using the methods developed in (Wu and Ryan, 2011), we can obtain δ' as an optimal policy for the condition-based replacement model described there with preventive replacement cost $C_1 + (h_W - h_S)\mu^{-1}$ and additional failure cost K. Thus an upper bound for c^* is given by

$$\overline{c} = \left\lfloor \left(\phi_0 - N \frac{C_1 + (h_W - h_S)\mu^{-1} + KQ(\delta')}{M(\delta')} \right) h_S^{-1} \right\rfloor \quad (18)$$

In light of the above discussion, the following algorithm is presented to find the optimal parameters and the global minimum of (11).

Algorithm II

1 Initialize $c = c_0$.

2 For fixed c_0, using Algorithm I to find the optimal parameters $t_i^{c_0}, i = 0,1,2$. Set $\phi_0 = \phi(c^0, t_0^{c_0}, t_1^{c_0}, t_2^{c_0})$.

3 Obtain an upper bound $\overline{c}$ for c^* using equation (18).

4 For $c = 1, \cdots, \overline{c}$, find the optimal $t_i^c, i = 0,1,2$ and the corresponding optimal cost $\phi_c^* = \phi(c, t_0^c, t_1^c, t_2^c)$. Then the optimal stock level $c^* = \arg\min_c \phi_c^*$, the optimal replacement policy $\delta^* = \{t_0^*, t_1^*, t_2^*\}$ and the global minimum is $\min \phi_c^*$.

5 NUMERICAL EXAMPLE

Suppose the covariate process $Z(t)$ is a pure birth process with three states $\{0,1,2\}$ and transition rates $v_0 = v_1 = -\ln(0.4), v_2 = 0$. Let $h_0(t) = 2t$ and $\Psi(Z_t) = \exp(2Z_t)$. Assume

$$C_1 = 5, C_2 = 15, K = 25, N = 10, \alpha = 0.2, \beta = 0.5, h = 0.5, \mu = 5.$$

The total cost per unit time is

$$\phi(c,\delta) = 1.5c + 10\frac{4.9 + 10.1p_L(c,\delta) + 25Q(\delta)}{M(\delta)}.$$

Let $c_0 = 10$. In Algorithm I step 1, initialize $\gamma_0 = 12$. As shown in Table 1, the lambda-minimization converges after four iterations with $v^{c_0} = 24.7330$ and optimal parameter $\delta^{c_0} = \{0.5440,\ 0.0736,\ 0.0100\}$. The corresponding total cost $\phi_0 = Nv^{c_0} + h_S c_0 = 262.33$.

Table 1 Illustration of Algorithm I and the lambda-minimization process

m	γ	Critical point $\delta^{c,m}$	Value of u at critical point	$v(\delta^{c,m})$
0	12	(0.3853, 0.0521, 0.0070)	4.4924	26.4098
1	26.4908	(0.5705, 0.0772, 0.0104)	-0.6750	24.7573
2	24.7573	(0.5444, 0.0737, 0.0100)	-0.0096	24.7330
3	24.7330	(0.5440, 0.0736, 0.0100)	0	24.7330

Based on methods developed in (Wu and Ryan, 2011), the replacement policy that minimizes term $\frac{4.9 + 25Q(\delta)}{M(\delta)}$ is $\delta' = \{0.4826, 0.0653, 0.0088\}$ and the minimal value of the term is 24.1302. Thus from equation (18), an upper bound for c is

$$\bar{c} = \left\lfloor \left(262.33 - 10\frac{4.9 + 25Q(\delta')}{M(\delta')}\right)\frac{1}{1.5} \right\rfloor = 14.$$

In step 4 of Algorithm II, when c ranges from 1 to 14, the resulting total costs are shown in Figure 3. The optimal parameters are $c^* = 12$ and $\delta^* = \{0.5048,\ 0.0683,\ 0.0092\}$. The global minimal cost is $\phi^* = 260.827$.

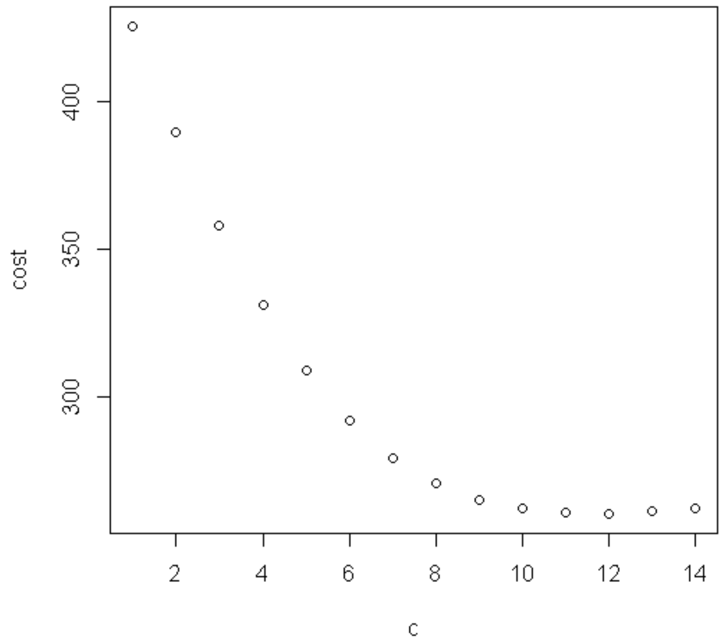

Figure 3 The minimized total cost when c varies from 1 to its upper bound

6 EVALUATION OF THE SINGLE CATEGORY RETURN ASSUMPTION

In the previous analysis, we consider the returns as a single category. However, in reality, there is usually a quality difference between the preventively replaced products and failure replaced products in terms of remanufacturing time and remanufacturing cost.

To understand the effect of the single category assumption, in this section, we will examine the case when we categorize the returns into two types: Type 1 (T1), preventively replaced products and Type 2 (T2), failure replaced products, estimate its cost and compare to the cost of the no categorization case.

In general, T1 products have better quality than T2 products, so it typically requires less remanufacturing effort to bring them to the "as good as new" condition. To quantify the quality difference, assume the statement “T1 is x% better than T2 in quality” implies that

1 The unit remanufacturing cost of T1, C_{11}, is x% lower than that of T2, C_{12}; i.e., $C_{11} = C_{12}(1 - x\%)$.

2 The remanufacturing time of T1, $1/\mu_1$, is x% shorter than that of T2, $1/\mu_2$; i.e., $1/\mu_1 = (1 - x\%)/\mu_2$.

6.1 Model analysis

Let $W_1(t), W_2(t)$ be the number of T1, T2 products, respectively, in WIP at time t .Then

$$W_1(t) + W_2(t) + I(t) = c.$$

And let λ_1, λ_2 be the arrival rate of T1, T2 products, respectively. Assume the product mix that enters the remanufacturing process is the same as the product mix that enters the remanufacturable inventory at all times. Then under replacement policy δ,

$$\lambda_1 = \lambda(1 - Q(\delta)),\ \lambda_2 = \lambda Q(\delta)$$

where $\lambda = N / M(\delta)$.

It is not hard to see that $(W_1(t), W_2(t) : t \geq 0)$ consists of a continuous-time Markov chain with a finite state space

$$S = \{(i, j) : i + j \leq c, i = 0,1,\ldots,c, j = 0,1,\ldots,c\}$$

A typical portion of the transition diagram among those states is shown in Figure 4. Although on the boundaries, one or more of the states depicted in the figure may not exist, the transition rates for the rest are valid. It can be verified that this continuous-time Markov chain is irreducible and ergodic, so it has a limiting distribution.

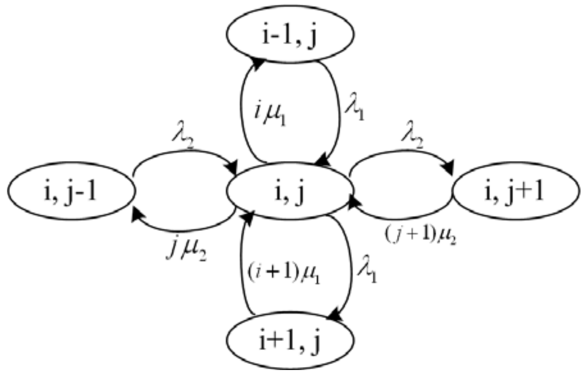

Figure 4: Part of transition diagram

Define the indicator function

$$I_S(i,j) = \begin{cases} 1 & \text{if } (i,j) \in S \\ 0 & \text{if } (i,j) \notin S \end{cases}$$

Then the balance equations of the limiting probabilities are

$$\begin{aligned} &\lambda_2 P(i,j-1) I_S(i,j-1) + \lambda_1 P(i-1,j) I_S(i-1,j) \\ &\quad + (j+1)\mu_2 P(i,j+1) I_S(i,j+1) + (i+1)\mu_1 P(i+1,j) I_S(i+1,j) \\ &= [\, j\mu_2 I_S(i,j-1) + i\mu_1 I_S(i-1,j) \\ &\quad + \lambda_2 I_S(i,j+1) + \lambda_1 I_S(i+1,j)] P(i,j) \quad \text{for all } (i,j) \in S \end{aligned}$$

Let $a = \lambda_1 / \mu_1, b = \lambda_2 / \mu_2$. The solution to the balance equations is

$$P(0,0) = \frac{1}{\sum_{i=0}^{c} \sum_{j=0}^{c-i} \frac{a^i b^j}{i!\, j!}} \tag{19}$$

$$P(i,j) = \frac{a^i b^j}{i!\, j!} P(0,0) \quad \text{for all } (i,j) \in S \tag{20}$$

Therefore

$$P(W_1(t) + W_2(t) = m) = \sum_{i=0}^{m} P(i, m-i) = \frac{(a+b)^m}{m!} P(0,0)$$

$$p_{L,cat} = P(W_1(t) + W_2(t) = c) = \frac{(a+b)^c}{c!} P(0,0) \tag{21}$$

$$E(W)_{cat} = E(W_1(t) + W_2(t)) = P(0,0) \sum_{k=0}^{c-1} \frac{(a+b)^{k+1}}{k!} \tag{22}$$

With categorization, the replacement cost per unit time for a single product is

$$\phi_{R,cat} = \frac{(1-p_{L,cat})[(1-Q(\delta))(C_{11}+KQ(\delta))+Q(\delta)(C_{22}+KQ(\delta))]+p_{L,cat}(C_2+KQ(\delta))}{M(\delta)}$$
$$= \frac{C_{11}+(K+C_{12}-C_{11})Q(\delta)+p_{L,cat}[C_2-C_{11}-Q(\delta)(C_{12}-C_{11})]}{M(\delta)}$$

and the total cost per unit time is

$$\phi(c,\delta)_{cat} = N\phi_{R,cat} + h_S c + (h_W - h_S)E(W)_{cat} . \tag{23}$$

We can follow the two step process as described in section 4 to optimize this objective function. However, since this objective function involves more complicated expressions of the loss probability and mean WIP than that of (11), minimizing the γ-function in the lambda minimization technique is challenging. Thus for step 1, we resort to some numerical optimization methods to minimize the objective function for a given c. Because first derivatives are available, we adopt the BFGS method (Fletcher, 1987), which is generally considered as the best quasi-Newton method.

The BFGS method cannot guarantee the global optimality of the obtained policy. However, for the single category case, we have verified that the policy and cost obtained using the BFGS method is the same as the optimal policy and cost obtained using lambda minimization technique. Since objective function for the two category case shares the same basic structure with the objection function for single category case, we use the result of BFGS method to approximate the optimal policy and cost in the two category case.

6.2 Cost impact of the single category assumption

Here we illustrate the cost impact of the single category assumption through a numerical example.

Assume T1 products are 20% better than T2 products in quality. Let $C_{11} = 5, \mu_1 = 5$. Then $C_{12} = 6.125, \mu_2 = 4$. Assume all the other parameters stay the same. Then the optimal cost $\phi^*_{cat} = 265.789$ and the optimal parameters are $c^*_{cat} = 12$, $\delta^*_{cat} = \{0.4911, 0.0620, 0.0091\}$.

If the decision maker uses the single category assumption, then first he must estimate the equivalent unit remanufacturing cost

$$C_1 = C_{11}(1-Q(\delta)) + C_{12}Q(\delta)$$

and equivalent unit processing rate

$$\mu = \frac{\lambda}{E(W)}(1-p_L)$$

where $\lambda = N / M(\delta)$. The estimation requires a realization of the operation policy $\{c, \delta\}$. Assume the policy maker can observe the results of $\{c^*_{cat}, \delta^*_{cat}\}$. Then under policy $\{c^*_{cat}, \delta^*_{cat}\}$

$$Q(\delta) = 0.1619\,,\ M(\delta) = 0.3707\,,\ p_L = 0.0075\,,\ E(W) = 5.5710$$

and the estimation would be

$$C_1 = 5.182, \mu = 5.000\,.$$

With those parameters, the optimal policy under the single category assumption is

$$c_{no\_cat} = 12\,,\ \delta_{no\_cat} = \{0.5120, 0.0693, 0.0094\}\,.$$

Under this policy, the actual total cost is $\phi_{no\_cat} = 265.916$, which is 0.05% bigger than δ^*_{cat}. This negligible cost difference indicates that the single category return assumption is acceptable in this instance.

Intuitively, the cost difference between the categorized and non-categorized cases depends on the quality difference between T1 and T2. To further evaluate the impact of single category assumption, we vary the quality difference, while keeping $C_{11} = 5, \mu_1 = 5$ unchanged, and then obtain the corresponding cost differences following a similar procedure as above. The results are summarized in Table 2.

As expected, for bigger quality difference, the additional cost introduced by the single category assumption is more substantial. And we can see that for our example, as long as the quality difference is below 50%, the cost error caused by the single category assumption is under 1%. Another observation is that as the quality difference increases, we tend to perform the preventive replacement more frequently and keep the stock level higher in the two category case, which are reasonable because, with a wider quality difference, it is more costly and time consuming to remanufacture a failure-replaced product.

Table 2 The impact of single category assumption under various quality difference between the two types of products

Quality difference x%	c^*_{cat}	δ^*_{cat}	ϕ^*_{cat}	c_{no_cat}	δ_{no_cat}	ϕ_{no_cat}	Cost difference $(\phi_{no\_cat} - \phi^*_{cat})/\phi^*_{cat}$
20%	12	{0.4911, 0.0620, 0.0091}	265.789	12	{0.5120, 0.0693, 0.0094}	265.916	0.05%
40%	13	{0.4697, 0.0638, 0.0088}	276.206	12	{0.5387, 0.0729, 0.0105}	277.673	0.53%
50%	13	{0.4603, 0.0622, 0.0087}	283.546	12	{0.5323, 0.0720, 0.0102}	285.766	0.78%
60%	14	{0.4364, 0.0591, 0.0081}	294.168	12	{0.5710, 0.0773, 0.0016}	301.289	2.42%
80%	17	{0.3634, 0.0492, 0.0069}	341.987	13	{0.6146, 0.0832, 0.0116}	374.794	9.60%

7 CONCLUSION

This paper investigates a joint operation problem in the context of a product-service system, which to the best of our knowledge has not been addressed in the literature. The system consists of a service subsystem and a remanufacturing subsystem where the replacement decision and the inventory management decision must be made at the same time. Identifying and formulating the couplings between the two subsystems, an integrated model aiming to minimize the total cost per unit time of the system is developed and an algorithm is presented to jointly optimize the replacement policy and the inventory management policy. Then we evaluate the cost impact of treating the preventively replaced products and products replaced due to failure as one category. A numerical example demonstrates that as long as the quality difference between the two types of replaced products is not too large, where how large depends on other parameters in the model, the single category assumption is reasonable.

In this paper, for illustration the covariate process is assumed to have three states, which could be characterized as “like new,” “deteriorated,” and “critical.” It is straightforward to generalize our model to accommodate a finer-grained approximation of a continuous state space by adding more discrete states. The additional effort required for formulation mainly lies in obtaining the explicit expressions of the mean replacement time and the failure probability for a given replacement policy, which is discussed in detail in Wu and Ryan, 2011). Correspondingly, the additional computational effort lies in the evaluation of the mean replacement time and the failure probability. In particular, for an *n*-state covariate process, the expressions of the mean replacement time and the failure probability consist of several *n*-fold integrals. Monte Carlo integration methods are essential to evaluate them efficiently (Press et al., 2007).

Other possible extensions to this paper are as follows. First, in our analysis, the demand process of the fleet for new products, which is a superposed renewal process, is approximated by a Poisson process assuming that the number of products in the fleet is sufficiently large. Evaluating the impact of this approximation in the situation of moderate or small fleet sizes is a possible extension of this research. Second, considering the capacity expansion problem of service subsystem in addition to maintenance would be an interesting and challenging problem, which is a natural generalization of the model presented in this study. Third, in a hybrid business model, where the producers operate traditional product sales as well as a PSS, the external returns in addition to the internal replaced products will become part of the input to the remanufacturing system. In this case, the inventory model needs to be reconsidered. Last but not least, the accelerated failure time (AFT) model (Meeker and Escobar, 1998) is considered as a strong competitor to the proportional hazards model when incorporating the covariate information into system failure time estimation. In case it is hard to decide which model to use, the general proportional hazards model (Bagdonavicius and Nikulin, 2001), which includes PHM and the AFT model as special cases, might be appropriate.

ACKNOWLEDGEMENT

This material is based upon work supported by the National Science Foundation under grant CNS-0540293.

REFERENCES

Aras, N., Boyaci, T., and Verter, V. (2004). The effect of categorizing returned products in remanufacturing. *IIE Transactions*, 36(4):319–331.

Armstrong, M. J. and Atkins, D. R. (1996). Joint optimization of maintenance and inventory policies for a simple system. *IIE Transactions*, 28:415–424.

Armstrong, M. J. and Atkins, D. R. (1998). A note on joint optimization of maintenance and inventory. *IIE Transactions*, 30:143–149.

Aven, T. and Bergman, B. (1986). Optimal replacement times – a general set-up. *Journal of Applied Probability*, 23:432–442.

Bagdonavicius, V. and Nikulin, M. (2001). *Accelerated life models: modeling and statistical analysis*. Chapman and Hall/CRC.

Baines, T. S. and Lightfoot, H. W. (2007). State-of-the-art in product-service systems. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 221:1543–1552.

Banjevic, D., Jardine, A. K. S., Makis, V., and Ennis, M. (2001). A control-limit policy and software for condition-based maintenance optimization. *INFOR*, 39:32–50.

Berthaut, F., Gharbi, A., and Dhouib, K. (2011). Joint modified block replacement and production/inventory control policy for a failure-prone manufacturing cell. *Omega*, 39:642–654.

Brezavscek, A. and Hudoklin, A. (2003). Joint optimization of block-replacement and periodic-review spare-provisioning policy. *IEEE Transactions on Reliability*, 52:112 – 117.

Buzacott, J. A. and Shanthikumar, J. G. (1993). *Stochastic Models of Manufacturing Systems*. Prentice-hall, Inc., Englewood Cliffs, NJ.

Cinlar, E. and Lewis, P. A. W. (1972). Superposition of point processes. In *Stochastic Point Processes*, pages 546–606. Wiley-Interscience, New York.

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*. Chapman & Hall, London.

Das, T. K. and Sarkar, S. (1999). Optimal preventive maintenance in a production inventory system. *IIE Transactions*, 31:537–551.

Fleischmann, M., Bloemhof-Ruwaard, J. M., Dekker, R., Van der Laan, E., and Van Numen, J. A. E. E. (1997). Quantitative models for reverse logistics: A review. *European Journal of Operational Research*, 103:1–17.

Fletcher, R. (1987). *Practical Methods of Optimization*. John Wiley & Sons Ltd, 2nd edition.

Ghasemi, S., Yacout, S., and Ouali, M. S. (2007). Optimal condition based maintenance with imperfect information and the proportional hazards model. *International Journal of Production Research*, 45(4):989–1012.

Giorgio, M., Guida, M., and Pulcini, G. (2011). An age- and state-dependent Markov model for degradation processes. *IIE Transactions*, 43:621–632.

Goedkoop, M., van Halen, C., and te Riele, H. (1999). Product service-systems, ecological and economic basics. Report for Dutch ministries of environment (VROM) and economic affairs (EZ). http://www.pre.nl/download/ProductService.zip.

Guide, V. D. R. (2000). Production planning and control for remanufacturing: Industry practice and research needs. *Journal of Operations Management*, 18:467–483.

Guide, V. D. R. and Van Wassenhove, L. N. (2009). The evolution of closed-loop supply chain research. *Operations Research*, 57(1):10–18.

Harel, A. (1990). Convexity properties of the erlang loss formula. *Operations Research*, 38:499–505.

Kharoufeh, J. P., Solo, C. J., and Ulukus, M. Y. (2010). Semi-Markov models for degradation-based reliability. *IIE Transactions*, 42:599–612.

Lam, C. T. and Yeh, R. H. (1994). Optimal maintenance policies for deteriorating systems under various maintenance strategies. *IEEE Transactions on Reliability*, 43(3):423–430.

Liu, Y., Li, Y., Huang, H.-Z., Zuo, M. J., and Sun, Z. (2010). Optimal preventive maintenance policy under fuzzy bayesian reliability assessment environments. *IIE Transactions*, 42:734–745.

Luiten, H., Knot, M., and van der Horst, T. (2001). Sustainable product-service-systems: the kathalys method. In *Proceedings of the Second International Symposium on Environmentally conscious design and inverse manufacturing*, pages 190–197.

Manzini, E., Vezzoli, C., and Clark, G. (2001). Product service-systems: using an existing concept as a new approach to sustainability. *Journal of Design Research*, 1(2).

Mardinab, F. and Araib, T. (2012). Capital equipment replacement under technological change. *The Engineering Economist*, 57(2):119–129.

Maxwell, I. and van der Vorst, R. (2003). Developing sustainable products and services. *Journal of Clearner Production*, pages 883–895.

Meeker, W. Q. and Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. Wiley, New York.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and P., F. B. (2007). *Numerical recipes: the art of scientific computing*. Cambridge University Press, third edition.

Rausch, M. and Liao, H. (2010). Joint production and spare part inventory control strategy driven by condition based maintenance. *IEEE Transactions on Reliability*, 59.

Rezg, N., Xie, X., and Mati, Y. (2004). Joint optimization of preventive maintenance and inventory control in a production line using simulation. *International Journal of Production Research*, 42:2029–2046.

Ross, S. M. (2003). *Introduction to Probability Models*. Academic Press, San Diego, CA, 8th edition.

Stewart, J. (1999). *Calculus*. Brooks/Cole Pub Co, 4th edition.

Thierry, M. C., Salomon, M., and Van Nunen, J. A. E. E. (1995). Strategic production and operations management issues in product recovery management. *California Management Review*, 37:114–135.

Toffel, M. W. (2008). Contracting for servicizing. Harvard Business School Technology & Operations Mgt. Unit Research Paper No. 08-063. Available at SSRN: http://ssrn.com/abstract=1090237 or http://dx.doi.org/10.2139/ssrn.1090237.

Van der Laan, E., Salomon, M., Dekker, R., and Wassenhove, L. V. (1999). Inventory control in hybrid systems with remanufacturing. *Management Science*, 45(5):733–747.

Van Halen, C., Vezzoli, C., and Wimmer, R. (2004). *MEPSS Handbook*. Royal Van Gorcum, Assen, Netherlands.

White, A., Stoughton, M., and Feng, L. (1999). *Servicizing: The Quiet Transition to Extended Producer Responsibility*. Tellus Institute, Boston.

Wu, X. and Ryan, S. M. (2010). Value of condition monitoring for optimal replacement in the proportional hazards model with continuous time degradation. *IIE Transactions*, 42:553–563.

Wu, X. and Ryan, S. M. (2011). Optimal replacement in the proportional hazards model with semi-Markovian covariate process and continuous monitoring. *IEEE Transactions on Reliability*, 60:580–589.

Yatsenko, Y. and Hritonenko, N. (2009). Technological breakthroughs and asset replacement. *The Engineering Economist*, 54(2):81–100.

APPENDIX A THE EXPLICIT EXPRESSIONS OF $M(\delta)$ AND $Q(\delta)$ FOR PH MODEL WITH THREE-STATE COVARIATE PROCESS AND THEIR PARITICAL DERIVATIVES

Assume the covariant process $Z(t)$ has three states {0, 1, 2}. Let S_i be the product age at which the state changes from i to $i+1$, $i=0,1$. And let $g_0(s_0)$ be the pdf of S_0, $g_1(s_0,s_1)$ be the joint pdf of S_0,S_1. Denote replacement policy $\delta=\{t_0,t_1,t_2\}$. Then the expected length of a replacement cycle $M(\delta)$ and the failure probability $Q(\delta)$ in terms of t_0,t_1,t_2 are

$$\begin{aligned}M(\delta) &= \int_{t_0}^{\infty} M_0(t_0)g_0(s_0)ds_0 + \int_{t_1}^{t_0} M_1(s_0)g_0(s_0)ds_0 + \int_{t_1}^{\infty}\int_0^{t_1} M_2(s_0,t_1)g_1(s_0,s_1)ds_0ds_1\\ &\quad + \int_{t_2}^{t_1}\int_0^{s_1} M_3(s_0,s_1)g_1(s_0,s_1)ds_0ds_1 + \int_0^{t_2}\int_0^{s_1} M_4(s_0,s_1,t_2)g_1(s_0,s_1)ds_0ds_1,\end{aligned}$$

$$\begin{aligned}Q(\delta) &= \int_{t_0}^{\infty} Q_0(t_0)g_0(s_0)ds_0 + \int_{t_1}^{t_0} Q_1(s_0)g_0(s_0)ds_0 + \int_{t_1}^{\infty}\int_0^{t_1} Q_2(s_0,t_1)g_1(s_0,s_1)ds_0ds_1\\ &\quad + \int_{t_2}^{t_1}\int_0^{s_1} Q_3(s_0,s_1)g_1(s_0,s_1)ds_0ds_1 + \int_0^{t_2}\int_0^{s_1} Q_4(s_0,s_1,t_2)g_1(s_0,s_1)ds_0ds_1\end{aligned}$$

where

$$M_0(t_0) = \int_0^{t_0} tdF_0(t) + t_0(1-F_0(t_0))$$

$$M_1(s_0) = \int_0^{s_0} tdF_0(t) + s_0(1-F_0(s_0))$$

$$M_2(s_0,t_1) = \int_0^{s_0} tdF_0(t) + \int_{s_0}^{t_1} tdF_1(s_0,t) + t_1(1-F_1(s_0,t_1))$$

$$M_3(s_0,s_1) = \int_0^{s_0} tdF_0(t) + \int_{s_0}^{s_1} tdF_1(s_0,t) + s_1(1-F_1(s_0,s_1))$$

$$M_4(s_0,s_1,t_2) = \int_0^{s_0} tdF_0(t) + \int_{s_0}^{t_1} tdF_1(s_0,t) + \int_{s_1}^{t_2} tdF_2(s_0,s_1,t) + t_2(1-F_2(s_0,s_1,t_2))$$

$$Q_0(t_0) = F_0(t_0) \qquad Q_1(s_0) = F_0(s_0)$$

$$Q_2(s_0,t_1) = F_1(s_0,t_1) \qquad Q_3(s_0,s_1) = F_1(s_0,s_1)$$

$$Q_4(s_0,s_1,t_2) = F_2(s_0,s_1,t_2)$$

and

$$F_0(t) = 1 - exp\left(-\Psi(0)\int_0^t h_0(u)du\right), t \le s_0$$

$$F_1(s_0,t) = 1 - exp\left(-\Psi(0)\int_0^{s_0} h_0(u)du - \Psi(1)\int_{s_0}^t h_0(u)du\right), s_0 \le t \le s_1$$

$$F_2(s_0,s_1,t) = 1 - exp\left(-\Psi(0)\int_0^{s_0} h_0(u)du - \Psi(1)\int_{s_0}^{s_1} h_0(u)du - \Psi(2)\int_{s_1}^t h_0(u)du\right), s_1 \le t.$$

The partial derivatives of $M(\delta)$ and $Q(\delta)$ with respect to t_i are

$$\frac{\partial M(\delta)}{\partial t_0} = \left(1 - \int_0^{t_0} g_0(s_0)ds_0\right)(1-F_0(t_0))$$

$$\frac{\partial M(\delta)}{\partial t_1} = \int_{t_1}^{\infty}\int_0^{t_1}(1-F_1(s_0,t_1))g_1(s_0,s_1)ds_0 ds_1$$

$$\frac{\partial M(\delta)}{\partial t_2} = \int_0^{t_2}\int_0^{s_1}(1-F_2(s_0,s_1,t_2))g_1(s_0,s_1)ds_0 ds_1$$

$$\frac{\partial Q(\delta)}{\partial t_0} = h_0(t_0)\Psi(0)\left(1 - \int_0^{t_0} g_0(s_0)ds_0\right)(1-F_0(t_0))$$

$$\frac{\partial Q(\delta)}{\partial t_1} = h_0(t_1)\Psi(1)\int_{t_1}^{\infty}\int_0^{t_1}(1-F_1(s_0,t_1))g_1(s_0,s_1)ds_0 ds_1$$

$$\frac{\partial Q(\delta)}{\partial t_2} = h_0(t_2)\Psi(2)\int_0^{t_2}\int_0^{s_1}(1-F_2(s_0,s_1,t_2))g_1(s_0,s_1)ds_0 ds_1$$

Note that

$$\frac{\partial Q(\delta)}{\partial t_i} = h_0(t_i)\Psi(i)\frac{\partial M(\delta)}{\partial t_i}, \forall i\,. \tag{24}$$

Appendix B Lambda minimization technique

In this section, we will give a brief introduction to the lambda minimization technique developed in Aven and Bergman, 1986), aiming to minimizing a function with the following form

$$B(X) = \frac{M(X)}{S(X)}, \tag{25}$$

where $X \in \Re^n$ and $S(X) > 0, \forall X$.

Define the λ-function

$$C(X,\lambda) = M(X) - \lambda S(X), \qquad \lambda \in (-\infty, +\infty).$$

For each λ, denote the value of X that minimizes $C(X,\lambda)$ as X_λ.

Now we will show how the problem of minimizing $B(X)$ can be solved by minimizing the λ-function $C(X,\lambda)$. Aven and Bergman proved the following proposition which associates the optimality of $B(X)$ with the optimality of $C(X,\lambda)$.

Proposition 1: *If X_λ minimizes $C(X,\lambda)$ and $C(X_\lambda,\lambda) = 0$, then X_λ is optimal for* (25) *and the optimal value of $B(X)$ is $B(X_\lambda) = \lambda \equiv \lambda^*$.*

Aven and Bergman then proved another important proposition, stated below, which leads to an iterative algorithm that always produces a sequence converging to λ^*.

Proposition 2: *Choose any λ_1 and set iteratively $\lambda_{n+1} = B(X_{\lambda_n})$, $n = 1, 2, 3, \cdots$. Then*

$$\lim_{n\to\infty} \lambda_n = \lambda^*.$$

Propositions 1 and 2 imply that the minimization of $B(X)$ can be transformed into the problem of minimizing $C(X,\lambda)$ plus a succession of iterations. This is the essence of the lambda minimization technique. This technique is very suitable for situations where it is easy to find the optimal solutions to the λ-function $C(X,\lambda)$ while it is hard to minimize $B(X)$ directly; this is often the case in replacement/maintenance applications. The optimal solution and the optimal value of $B(X)$ can be attained simultaneously when the algorithm converges.

APPENDIX C PROOF OF THEOREM 1

For readability, first we list all the monotonicity properties of various functions that are related to the proof of Theorem 1 in the following.

1) $M(\delta)$ is increasing in t_i, $\forall i$.
2) $\frac{\partial p_L}{\partial M}$ is increasing in t_i, $\forall i$.
3) $h_0(\cdot)$ is increasing in its parameter.
4) t_1 is increasing in t_0 if $t_1 = h_0^{-1}\left[\frac{\Psi(0)}{\Psi(1)} h_0(t_0)\right]$ and t_2 is increasing in t_0 if $t_2 = h_0^{-1}\left[\frac{\Psi(0)}{\Psi(2)} h_0(t_0)\right]$.

Since γ is fixed, in the following discussion, it is suppressed in the notation of u.

The feasible region of $u(\delta)$ is $R = \{\delta : t_0 \geq t_1 \geq t_2 \geq 0\}$. Divide this region into two sets: a closed and bounded set $D = \{\delta : \Lambda \geq t_0 \geq t_1 \geq t_2 \geq 0\}$ where Λ is an arbitrary large positive number, and set $B = R \setminus D$; i.e., $\{B, D\}$ is a partition of R. Define $t_i = +\infty$ represent failure replacement in state *i.*

Lemma C.1 $u(\delta)$ achieves its minimum at the critical point $\delta^\gamma = \{t_0^\gamma, t_1^\gamma, t_2^\gamma\}$ in D.

Proof

Since u is a continuous function on the closed and bounded region D, according to the extreme value theorem for multivariate functions (Stewart, 1999), u has a global minimum which happens either at its critical point or a certain point on the boundary.

The boundaries of $u(\delta)$ are where $t_0 = t_1$, $t_1 = t_2$, $t_2 = 0$ or $t_0 = \Lambda$ within the feasible region $D = \{\delta : \Lambda \geq t_0 \geq t_1 \geq t_2 \geq 0\}$. We prove by contradiction that points on the boundaries are not optimal to $u(\delta)$.

The Version of Record of this manuscript has been published and is available in The Engineering Economist 2014, http://www.tandfonline/ 10.1080/0013791X.2013.873844. Posted with permission.

1) Points on the boundary where $t_0 = t_1$.

Suppose an optimal point exists on this boundary and denote it as $\delta = \{a_0, a_0, a_2\}$, $a_0 \geq a_2 \geq 0$.

Recall that the partial derivative of $u(\delta)$ with respect to t_0, t_1, t_2 are

$$\frac{\partial u}{\partial t_i} = \frac{\partial M}{\partial t_i}\left(b_2 \frac{\partial p_L}{\partial M}(t_0, t_1, t_2) + Kh_0(t_i)\Psi(i) - \gamma \right), \quad i = 0,1,2,$$

and $\frac{\partial M}{\partial t_i} > 0, \forall i$.

i. If $b_2 \frac{\partial p_L}{\partial M}(a_0, a_0, a_2) + Kh_0(a_0)\Psi(0) - \gamma \geq 0$

it follows that

$$b_2 \frac{\partial p_L}{\partial M}(a_0, a_0, a_2) + Kh_0(a_0)\Psi(1) - \gamma > 0.$$

According to the definition of a continuous function,

$$\exists a_1 \in (a_2, a_0), \text{ s.t. } b_2 \frac{\partial p_L}{\partial M}(a_0, a_1, a_2) + Kh_0(a_1)\Psi(1) - \gamma > 0.$$

From assumption #7 and Lemma 2, we know that $b_2 \frac{\partial p_L}{\partial M}(t_0, t_1, t_2) + Kh_0(t_1)\Psi(i)$ is increasing in t_1.

Thus

$$\frac{\partial u}{\partial t_1}(a_0, t_1, a_2) > 0 \;\; \forall t_1 \in [a_1, a_0],$$

which means u is an increasing function for $t_1 \in [a_1, a_0]$ with fixed $t_0 = a_0, t_2 = a_2$. Therefore $u(\{a_0, a_1, a_2\}) < u(\{a_0, a_0, a_2\})$. Contradiction.

ii. If $b_2 \frac{\partial p_L}{\partial M}(a_0, a_0, a_2) + Kh_0(a_0)\Psi(0) - \gamma < 0$

It follows that

$$\exists a_0' > a_0, \text{ s.t. } b_2 \frac{\partial p_L}{\partial M}(a_0', a_0, a_2) + Kh_0(a_0')\Psi(0) - \gamma < 0.$$

Thus $\frac{\partial u}{\partial t_0}(t_0, a_0, a_2) < 0$, $\forall t_0 \in [a_0, a_0']$, which means u is an decreasing function for $t_0 \in [a_0, a_0']$ with fixed $t_1 = a_0, t_2 = a_2$. Therefore $u(\{a_0', a_0, a_2\}) < u(\{a_0, a_0, a_2\})$. Contradiction.

2) Points on the boundary where $t_1 = t_2$

Using the same argument as in 1), we can prove that points on this boundary are not optimal.

3) Points on the boundary where $t_2 = 0$

Assume $(a_0, a_1, 0)$ is optimal.

Since $\frac{\partial p_L}{\partial M} < 0$, $h_0(0) = 0$ and $\gamma > 0$, $b_2 \frac{\partial p_L}{\partial M}(a_0, a_1, 0) + Kh_0(0)\Psi(2) - \gamma < 0$

It follows

$$\exists a_2 \in (0, a_1) \text{, s.t. } b_2 \frac{\partial p_L}{\partial M}(a_0, a_1, a_2) + Kh_0(a_2)\Psi(2) - \gamma < 0 .$$

Thus $\frac{\partial u}{\partial t_2}(a_0, a_1, t_2) < 0$, $\forall t_2 \in [0, a_2]$, which mean u is an decreasing function for $t_2 \in [0, a_2]$ with fixed $t_0 = a_0, t_1 = a_1$. Therefore $u(\{a_0, a_1, a_2\}) < u(\{a_0, a_1, 0\})$. Contradiction.

4) Points on the boundary where $t_0 = \Lambda$

Since K is arbitrary large and $h_0(\cdot)$ is increasing and unbounded, then we have

$$b_2 \frac{\partial p_L}{\partial M}(\Lambda, a_1, a_2) + Kh_0(\Lambda)\Psi(0) - \gamma > 0$$

It follows

$$\exists a_0 \in [a_1, \Lambda) \text{, s.t. } b_2 \frac{\partial p_L}{\partial M}(a_0, a_1, a_2) + Kh_0(a_0)\Psi(0) - \gamma > 0 .$$

Thus $\frac{\partial u}{\partial t_0}(t_0, a_1, a_2) > 0 \; \forall t_0 \in [a_0, \Lambda)$, which means u is an increasing function for $t_0 \in [a_0, \Lambda)$ with fixed $t_1 = a_1, t_2 = a_2$. Therefore $u(\{a_0, a_1, a_2\}) < u(\{\Lambda, a_1, a_2\})$. Contradiction.

In summary, in region D, $u(\delta)$ can not achieve its global minimum at its boundaries; which means $u(\delta)$ can only achieve its global minimum at its unique critical point $\delta^{\gamma} = \{t_0^{\gamma}, t_1^{\gamma}, t_2^{\gamma}\}$. ■

Lemma C.2 $\forall \delta \in B, \exists \delta' \in D$, such that $u(\delta') < u(\delta)$.

Proof

There are only three types of points in B: $\delta = \{\infty, a_1, a_2\}$, $\delta = \{\infty, \infty, a_2\}$ and $\delta = \{\infty, \infty, \infty\}$, $\infty > a_1 > a_2 \geq 0$.

1) For points in the form of $\delta = \{\infty, a_1, a_2\}$

Since $h_0(\cdot)$ is increasing and unbounded, then we have

$$\lim_{t_0 \to \infty} b_2 \frac{\partial p_L}{\partial M}(t_0, a_1, a_2) + K h_0(t_0)\Psi(0) - \gamma > 0.$$

Thus

$$\exists a_0 \in [a_1, \infty), \text{ s.t. } b_2 \frac{\partial p_L}{\partial M}(a_0, a_1, a_2) + K h_0(a_0)\Psi(0) - \gamma > 0.$$

Thus $\frac{\partial u}{\partial t_0}(t_0, a_1, a_2) > 0 \ \forall t_0 \geq a_0$, which means u is an increasing function for $t_0 \geq a_0$ with fixed $t_1 = a_1, t_2 = a_2$. Therefore if we let $\delta' = \{a_0, a_1, a_2\}$, then $\delta' \in D$ and $u(\delta') < u(\delta)$.

2) For points in the form of $\delta = \{\infty, \infty, a_2\}$

Using the similar argument as in 1), there exists $a_1 \in [a_2, \infty)$, such that $\frac{\partial u}{\partial t_1}(\infty, a_1, a_2) > 0 \ \forall t_1 \geq a_1$. Therefore $u(\{\infty, a_1, a_2\}) < u(\delta)$. And based on the result in 1), $\exists \delta' \in D$, such that $u(\delta') < u(\{\infty, a_1, a_2\}) < u(\delta)$.

3) For points in the form of $\delta = \{\infty, \infty, \infty\}$

Similarly, we can prove that

$$\exists \delta' \in D, a_2 \in [0, \infty), \text{ such that } u(\delta') < u(\{\infty, \infty, a_2\}) < u(\delta).$$

To sum up, $\forall \delta \in B, \exists \delta' \in D$, such that $u(\delta') < u(\delta)$. ■

From Lemma C.1 and Lemma C.2, we conclude that for a given γ, function $u(\gamma,\delta)$ achieves a global minimum at its unique critical point, $\delta^{\gamma}=\{t_0^{\gamma},t_1^{\gamma},t_2^{\gamma}\}$.