Optimal Reliability and Upgrading Decisions for Capital Goods

This thesis is number D134 of the thesis series of the Beta Research School for Operations Management and Logistics. The Beta Research School is a joint effort of the departments of Industrial Engineering & Innovation Sciences, and Mathematics and Computer Science at Eindhoven University of Technology and the Centre for Production, Logistics and Operations Management at the University of Twente.

A catalogue record is available from the Eindhoven University of Technology Library.

ISBN: 978-90-386-2288-0

Printed by University Printing Office, Eindhoven

This research has been funded by the Innovation-Oriented Research Programme 'Integrated Product Creation and Realization (IOP IPCR)' of the Netherlands Ministry of Economic Affairs.

Optimal Reliability and Upgrading Decisions for Capital Goods

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag van de
rector magnificus, prof.dr.ir. C.J. van Duijn, voor een
commissie aangewezen door het College voor
Promoties in het openbaar te verdedigen
op maandag 30 augustus 2010 om 16.00 uur

door

Kurtuluş Barış Öner

geboren te Tarsus, Turkije

Dit proefschrift is goedgekeurd door de promotoren:

prof.dr.ir. G.J.J.A.N. van Houtum
en
prof.dr. A.G. de Kok

Aileme...
özellikle,
yeğenime

To my family...
especially,
to my nephew

Acknowledgements

There are two major manners of telling an event or situation: Through a short story or a long story. My friends know... well, not only know but also complain about the fact that I tell long stories. If you ask something... anything... my response can start with something like the big bang theory. I will also tell a story regarding this thesis now. I will start the story with some parts from its end, just in case you are so busy (bored) and cannot read the full story.

I guess that all Ph.D. students get frustrated every now and then, as they often get lost in the chaotic clouds of thoughts and ideas. My first promoter, Geert-Jan van Houtum, has been a great guide, clarifying the paths that I could follow whenever I was lost. I have never had a question left without an answer. He helped me zoom in and out on problems. He also taught me a lot about how to write down my work properly. Thank you so so much Geert-Jan.

It has been a great fortune to me to have Gudrun Kiesmüller as my daily supervisor. Her office was just a few meters away from mine and she was so open to any discussion such that I often had the chance to knock on her door and start talking without an appointment. Her perspective, which was different from Geert-Jan's, improved my approach towards the problems considerably. I enjoyed the time that I spent with her in front of the board for formulations and proofs. Gudrun: Thank you for all your effort and friendly supervision.

I had a visit to Carnegie Mellon University (CMU) in Pittsburgh, USA, between March-June 2009. I worked under the supervision of Alan Scheller-Wolf. Again, a different perspective which was exhibited in a very positive manner. Alan had a very busy agenda... and I was always in that agenda. The third chapter of this thesis includes the research that I started at CMU. My stay there was an extraordinary experience, not only in terms of research. Things were different in the USA... different from what I was thinking of. Thank you Alan: First, for your supervision and cooperation; second, for providing me the opportunity to get rid of some of my prejudices and establish very good friendships there.

Now, back to the beginning...

I remember a conversation between me and my mom, some time after I started the first grade. She asked me how I was feeling about the school (apparently, after observing me for a while). I did not have any troubles with the classes. I was not crying after my mom or something. But I did not like the school. It was so boring. I told her that I still wanted to be a free kid, playing all the time, drawing pictures, etc, etc, etc. The school education started too early for me (I was 6)! The next day, I was a free kid. I thank to my parents, because not only they are my parents and take care of me (I am almost 33 but they are still on duty), but also they have always respected my opinions and decisions.

-Canım annem ve babam, sizlere çok çok teşekkür ediyorum. Sadece annem-babam olmanızdan dolayı değil, aynı zamanda küçücük çocukken dahi düşünce ve kararlarıma saygı göstermenizden dolayı.-

I restarted the primary school next year. This time, I was ready for it. I loved math from the beginning. I still love it. If you have a look at the further pages of this book, you will see some math. I am grateful to my primary school teacher, Ahmet Hüseyin Petek, for reinforcing my interests. I know quite many people who hated most of the things taught in classrooms. And I also know that a large portion of their hate stemmed from the manner that they were taught.

I also feel grateful to (almost) all my high school teachers, but I would like to mention two of them in particular. I did not get the best scores in English classes, but thanks to my English teacher, Serpil Mıstık, I can write a book fully in English now. I and many friends of mine have a great gratitude to our math teacher, Mustafa Özdemir, who made us study hard for the university entrance exam at an early stage compared to other students... and we all made it to very good universities.

I had a great time at Boğaziçi University. I was studying hard. I was also enjoying the life in the beautiful campus on the coast of Bosphorus. Thanks to all academic members at the Industrial Engineering Department of Boğaziçi University. Special thanks to my supervisor there, Kuban Altınel, who has a very large impact on my views on science. I will never forget his support which was not limited only to my academic life.

The education at the Industrial Engineering Department of Boğaziçi University provides a very strong theoretical background. However, I think that there is a lack of bridging theory and practice in the education (indeed, I think that this is a general problem in the education system in Turkey). This connection has been established throughout my Ph.D. research. Thanks to the Innovation-Oriented Research Programme 'Integrated Product Creation and Realization (IOP IPCR)' of the Netherlands Ministry of Economic Affairs, which funded my research and enabled me to cooperate with a number of companies. I would also like to thank to Ad Zephat and Guillaume Stollman from Philips Healthcare; and Wout Smans and Radj Bachoe from Vanderlande Industries. I can confidently state the motivation of the problems that I tackled in this thesis and the use of our research by their help.

I would like to thank my second promoter, Ton de Kok, for his high-level comments and ideas about my research. Thanks to Martin Newby for his valuable comments about my thesis. His remarks made me really think of the problems in a different way.

At the department, I sometimes walked around in the corridor annoyingly... distracting my colleagues (friends!). Thanks to all for their tolerance. But, unfortunately, no-one could save Michiel Jansen, my office-mate, from suffering my existence all the time. So, you understand that he deserves to be thanked separately. I would also like to thank Ingrid van Helvoort - Vliegen for being my 'supervisor' about the Dutch way of life and education system, especially in my first year here. She also helped me with the writing of this thesis.

There are three other friends that I have to mention here. Especially two of them would kill me if I do not do so. And, honestly, they have such a right... to some extent. First: Thank you Kostas Kevrekidis for our discussions and cooperation... and, of course, your great friendship. Then: Thank you Ola Jabali (Olla Gabali is a misspelling!) and Hayriye Çağnan, my angels who saved my life multiple times, one of which was related to the numerical precision problem that I had with the fourth chapter of this thesis. Now... you know who have... some right.

I told you, I tell long stories. But, if these people were not involved in my life and these events did not happen, you couldn't read this book, if you ever intended to do so (in the meanwhile, there were other steps in between which I also appreciate). I do not mean that there would not be any book written by me. But, it could not be this one. I have been a lucky man so far. And I hope that I will be lucky enough to tell some stories in the future.

At the very end, I would also like to thank Matthieu van der Heijden and Fred Langerak for participating in my Ph.D. committee.

Summary

Optimal Reliability and Upgrading Decisions for Capital Goods

Advanced technical systems, also called advanced capital goods (e.g. medical systems, material handling systems, defense systems, manufacturing systems, packaging lines, computer networks) are used in core processes by their users. By core processes, we mean the processes which are essential for operational continuity. For example, baggage handling at airports, transactions in a bank, data processing in a computer network, can be considered as core processes. Operational interruptions of these systems lead to significant losses for the users and keeping the systems up and running (availability of the systems) is crucial.

A high level of system availability can be provided by maintaining

- a low frequency of system failures, and/or
- a high speed of system repair activities (short downtime per system failure).

The frequency of failures of a system depends heavily on its design. The focus of this thesis is on two major design decisions in this context:

(i) reliability of components that compose the system, and

(ii) redundancy (i.e., having a number of identical components in parallel instead of a single component).

We refer to these decisions as *reliability decisions*.

The speed of system repair is commonly accelerated by using the *repair-by-replacement* concept during the exploitation phase. That is, if a part fails and leads to a system failure, the system is restored by replacing the failed part with a ready-for-use one. Spare parts are kept on stock at a short distance of the installed systems to prevent long downtimes. For a fixed system design, the spare parts inventory level is a key factor affecting the system availability. We take the spare parts inventory into account when investigating the optimal reliability decisions.

The primary goal of this thesis is to develop quantitative models and methods for optimal reliability decisions in the design phase. In Chapter 2 and 3, we study the optimal reliability level of a critical component and the redundancy optimization for serial systems, respectively. Typically, Original Equipment Manufacturers (OEMs) of capital goods are responsible for the availability of their systems in the field through service contracts. OEMs redesign components that fail too often and therefore have a strong negative effect on availability. It is then economical to improve the reliability of those components and upgrade the systems by replacing the old parts in the field with the redesigned ones. After the redesign, there are multiple policies that can be followed by an OEM for upgrading the systems. In Chapter 4, we study two common *upgrading policies* and investigate their optimality. In Chapter 1 and 5, an introduction and the conclusions are given. In Appendix A, we provide several results for the Erlang loss system which are motivated by the problem studied in Chapter 2.

In Chapter 2, we develop a model for the optimization of the reliability level of a critical component. In this model, portions of the Life Cycle Costs (LCC - total costs incurred throughout the lifetime of systems) of a general number of systems that are affected by component reliability and the spare parts inventory level are formulated. We develop an efficient solution procedure for the problem. By conducting a numerical experiment, we show that taking the spare parts inventory level into account for the optimization of component reliability in the design phase lead to significant cost reductions compared to solutions generated by sequential consideration the component reliability and the spare parts inventory level. The results of the experiment also reveal that the optimal component reliability is much higher for a cheap component than for an expensive component and increases as the number of the systems increases, the downtime penalty rate increases; and, the exploitation phase gets longer. We also show that the optimal LCC have negligible or limited sensitivity to the most of the major parameters in our model.

In Chapter 3, we introduce a redundancy optimization model for a capital good with a serial structure (from the reliability point of view). We refer to the units which are connected to each other in series in the capital good as *stages*. When there is no redundancy in a stage, the stage is composed of a single component. If a stage is designed with redundancy, then it includes two units of the same component which are connected to each other in parallel (from the reliability point of view). In the problem that we studied, three policies per stage are defined. Redundancy is included by one of the policies. Each of the three policies provides different levels of uptime (availability). We formulate the problem as the minimization of the Total Cost of Ownership (TCO - equivalent to LCC from the customer perspective) of a general number of systems under a defined constraint on the expected downtime of the systems throughout their life cycle. We decompose the problem into single-stage problems and show that a solution for the multi-stage problem can be generated by solutions of each of the single-stage problems. We develop an efficient procedure to find optimal solutions of the single-stage problems for varying levels of the downtime constraint. Solutions for the multi-stage problem for varying levels of the downtime constraint are generated efficiently by repeating this procedure for each stage. We derive the following major results through the analysis of the single-stage and multi-stage problem formulations:

- Single-stage: When level of the downtime constraint is decreased from a high value to zero; i.e., the constraint was initially loose and got tighter and tighter, the policy to include redundancy becomes optimal at a certain level and remains optimal for all smaller levels.
- Multi-stage:
  - One can generate an efficient frontier which reflects the trade-off between the uptime and the TCO .
  - An optimal ordering of the stages to include redundancy one-by-one can be generated.

In Chapter 4, we develop a model for studying the following two upgrading policies that an OEM may follow for multiple systems in the field after the redesign of a component (we denote the time just after the redesign by time 0):

- Policy 1 - Upgrade all systems preventively at time 0.
- Policy 2 - Upgrade systems one-by-one correctively.

Under Policy 2, new (improved) parts are kept on stock for upgrading while no inventory of new parts is kept under Policy 1. Under Policy 2, the initial supply quantity of new parts is a decision variable and new parts can be replenished in batches with a fixed size after the initial supply. The unit price of the new parts might increase after time 0.

We develop a problem formulation for the comparison of the two policies and perform exact analysis. We conduct a numerical study and find out that Policy 1 is favored by low values of the number of the systems, long lifetime of the systems, low values of the MTBF of the old parts (for fixed percentage improvement in MTBF), high values of the percentage improvement in MTBF, high values of the increase in the unit price of the new parts after time 0, large batch sizes for new parts under Policy 2, and high values of the downtime costs per failures. The reverse of each of these conditions favors Policy 2. Our numerical study showed that the optimal policy may change by varying any of the mentioned factors.

Contents

Chapter 1

Introduction

Advanced technical systems, also called advanced capital goods (e.g. medical systems, material handling systems, defense systems, manufacturing systems, packaging lines, computer networks) are used in core processes by their users. By core processes, we mean the processes which are essential for operational continuity. For example, baggage handling at airports, transactions in a bank, data processing in a computer network, can be considered as core processes. Operational interruptions of these systems lead to significant losses for the users and keeping the systems up in the field (availability of the systems) is crucial.

In many cases, Original Equipment Manufacturers (OEMs) are the service providers of their systems. Traditionally, after selling a system, OEMs are responsible for the availability of their systems at customer sites only during a warranty period (2-3 years) which is considerably shorter than the life cycle (lifetime) of these systems, which is 10-40 years. After the warranty period, they benefit from failures by charging customers for spare parts, labor, and other resources that are used during service activities. However, the market dynamics force OEMs to be responsible for the availability of their systems throughout their life cycle. Service contracts with payment terms based on performance of systems (availability) in the field rather than materials used for keeping systems up are becoming more and more common.

A high level of system availability can be provided by maintaining

- a low frequency of system failures, and/or
- a high speed of system repair activities (low downtime per system failure).

The frequency of failures of a system depends heavily on its design. The focus of this thesis is on two major design decisions in this context:

(i) reliability of components that compose the system, and

(ii) redundancy (i.e., having a number of identical components in parallel instead of a single component).

We refer to these decisions as *reliability decisions*.

The speed of system repair is commonly accelerated by using the *repair-by-replacement* concept during the exploitation phase. That is, if a part fails and leads to a system failure, the system is restored by replacing the failed part with a ready-for-use one. Spare parts are kept on stock at a short distance of the installed systems to realize the repair-by-replacement concept efficiently. For a fixed system design, the spare parts inventory level is a key factor affecting the system availability. We investigate the optimal reliability decisions taking into account the effect of spare parts inventory on availability.

In addition, periodically, OEMs redesign some components as they decide that it is more economical to improve the reliability of those components and upgrade the systems by replacing the old parts in the field with the redesigned ones. But, after the redesign, there are multiple policies that can be followed by an OEM for upgrading the systems. We also study a number of common *upgrading policies* and investigate their optimality.

The remainder of this chapter is organized as follows. We first motivate the need for investigating reliability related problems for capital goods in Section 1.1. In Section 1.2, we give the definitions of the problems that we study. Concepts such as maintenance, spare parts, availability, and critical components are fundamental for our research. We explain the relevance of these concepts in Section 1.3. In Section 1.4, we introduce terminology to make the descriptions in the remainder of the Thesis clearer. This thesis has connections to several research fields such as reliability optimization, warranty, and spare parts inventory control. We give an overview of the relevant literature in these fields in Section 1.5. We identify that the models in the literature are incapable of providing satisfactory solutions for the needs in the capital goods industry and list our contributions accordingly in Section 1.6. Finally, we give the outline of the rest of the thesis in Section 1.7.

1.1. Motivation

The research in this thesis is a part of an IOP-IPCR[1] project entitled "Life cycle oriented design of capital goods: System availability and integral costs". The goal of this project is to develop quantitative methods for integrated decision making to balance availability and Life Cycle Costs (LCC) of capital goods in the design phase. We carried out our research in cooperation with a number of companies, in particular with Philips Healthcare and Vanderlande Industries, and all the problems addressed in this thesis are motivated by practice.

1.1.1 Life Cycle Costs

The life cycle of a system is composed of four phases: design/development, production, exploitation and disposal; see Figure3 . LCC includes costs that are incurred in each of these phases. A number of definitions for LCC can be found in different publications which date back to the 1960's; see Gupta and Chow (1985); Asiedu and Gu (1998); Christensen et al. (2005). In this thesis, we use the following definition given by the US Department of Energy in 1995; see Barringer and Weber (1996): "LCC are the total costs estimated to be incurred in the design/development, production, operation, maintenance, support, and final disposition of a major system over its anticipated useful life span." Operation costs, maintenance costs, and support costs are incurred in the exploitation phase. We also incorporate downtime costs explicitly into the exploitation phase costs.

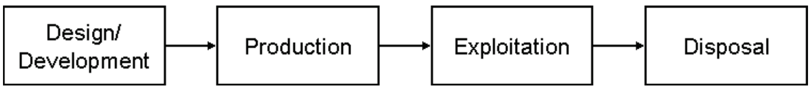

Figure 1.1 Life cycle of a system

LCC can be calculated with two different perspectives: the manufacturer's perspective or the customer's perspective. The costs that are included in LCC depend on the perspective. For example, a manufacturer distinguishes the design/development and production costs for LCC while these costs are incorporated in the acquisition costs for a customer. The LCC of a system from a customer perspective is also known as Total Cost of Ownership (TCO); see Ellram (1994).

[1] In Dutch, 'Innovatiegerichte Onderzoeksprogramma's - Integrale Productcreatie en Realisatie' or 'Innovation oriented research programs - Integrated Product Creation and Realization', which is a research programme of the Netherlands Ministry of Economic Affairs

1.1.2 Downtime Costs and Maintenance Costs are High for Advanced Capital Goods

In general, advanced capital goods are very expensive and their utilization is usually high. Downtime costs (stemming from losses in revenues, penalties, idle employees, etc. during the downtime) can be considerably high as the operations of users depend heavily on the availability of these goods. For example, downtime costs of computer systems of large e-commerce companies and brokerage companies can be $100,000-$1,000,000 per hour (see Patterson (2002); cnet news (2001); see also Downtime-Central (2009) for other examples). Intense maintenance activities are carried out to keep downtime as small as possible, which results in high maintenance costs.

We studied the TCO of an engineer-to-order type of system and found that the TCO is distributed as given in Table 1.1 (see Öner et al. (2007) for the details of the measurement). The downtime costs account for almost 50% of TCO and the downtime and maintenance costs together constitute 75% of TCO. These figures are in line with the other studies conducted within the scope of the IOP-IPCR project (see Basten (2006); Meutstege (2007)) and figures in the literature (see Gupta (1983); Saranga and Kumar (2006)).

Table 1.1 Distribution of TCO for an engineer-to-order system

Acquisition costs	23%
Maintenance costs	27%
Operations costs	2%
Downtime costs	48%

1.1.3 OEMs Focus on After-Sales Service Business

As stated earlier, OEMs are usually the primary service providers of their systems; the conditions in the capital goods market force OEMs to shift their focus from pure manufacturing to servicing systems. Below, we explain these conditions and their effects on OEMs.

After-sales service is a big business: A number of recent studies revealed the high volume of after-sales service business. The research firm Aberdeen Group reported that spare parts and after-sales services accounted for 8% of the annual gross domestic product in the United States in 2003 and the total annual global spending on after-sales services was over $1.5 trillion (see AberdeenGroup (2003)). Deloitte Consultancy states that the revenues from service business covers 25% of total business of many of the globally leading OEMs see (Deloitte (2006)). Similarly, this share lies between 20-30% according to a report by Aberdeen Group (see AberdeenGroup (2006)). Profit margins for after-sales services and parts range from 25% to 1000% higher than margins for initial products, which makes after-sales services account for about 40% of profits for most companies (see AberdeenGroup (2006) and Deloitte (2006)). A benchmark study by Deloitte Consulting (Deloitte (2006)) which included many of the world's largest manufacturing companies revealed that the average growth of the service businesses is about 10% higher than that of the business units overall.

Downtime costs and maintenance costs become concerns of OEMs: Downtime costs and maintenance costs have been concerns for customers only until recently. An OEM sells a system with a warranty and she incurs maintenance costs only during the warranty period. After the warranty period, the customer and OEM agree on a so-called *material contract* and the customer pays the OEM for spare parts, labor, and other resources that are used during service activities (see Kim et al. (2007b)). The OEM benefits from failures and downtime with such an agreement, let alone that she is bothered with them.

Nevertheless, OEMs increasingly feel the pressure to decrease downtime costs and maintenance costs of their systems. There are two main reasons for this change. First, TCO (LCC) of a system is increasingly becoming the primary criterion for a customer in her purchasing decision rather than price (acquisition costs). Previously, customers have tended to concentrate on acquisition costs when purchasing systems. However, they have gradually recognized the fact that seeking low prices in the short-run might lead to high exploitation phase costs in the long-run. Thus, they ask for TCO estimates during purchasing. As we described, a significantly large portion of TCO may be constituted by downtime costs and maintenance costs. But estimation of downtime costs is usually a nontrivial task, if ever possible. Thus, in many cases, availability estimates are demanded by customers together with TCO estimates disregarding downtime costs.

Second, *performance-based* and *power-by-the-hour* business models are becoming more common means of service provision (see Cohen et al. (2006)) as the service priority is very high in the capital goods industry. In a performance-based model, customers pay for services according to the performance of systems (e.g. through a contract which contains a service level agreement with respect to the uptime of the system(s)); while in a power by the hour model, customers pay for the services used. In both models, the OEM directly suffers from the losses due to downtime and incurs maintenance costs.

Customer demand for services increases: Nowadays, customers ask for near-100-% asset availability and better customer support. These requirements are translated into shorter lead times and/or higher service levels in the service supply chains (see AberdeenGroup (2005); Deloitte (2006)).

Oliva and Kallenberg (2003) study 11 OEMs and identify major reasons for the shift of the focus of OEMs to services. Together with the economic arguments and increasing customer demand for services (explained above), they also state a competitive argument as one of the reasons: Services are difficult to imitate and, thus, become a sustainable competitive advantage.

1.1.4 Long-Term Approach During the Design Phase

Although OEMs benefit from the large-scale service business, there are still unutilized opportunities, primarily due to different characteristics of service supply chains which make them more difficult to manage than manufacturing supply chains. Cohen et al. (2006) state these characteristics and propose a procedure for high-level management of service supply chains. Oliva and Kallenberg (2003) also propose a process model for the transition of orientation from manufacturing to service, which helps OEMs in changing their organization and processes.

The trends in the capital goods market brings further challenges. As stated earlier, OEMs are becoming responsible for the life cycle and availability management of their products. Despite its difficulty, this responsibility also brings the opportunity to grasp full handling of products from the beginning (i.e. their design phase). A long-term perspective which takes the effect of design decisions on major performance measures and costs in the exploitation phase is a necessity to benefit from this opportunity.

The reliability decisions of a system (reliability of its components and redundancy) are key factors that affect availability, downtime costs, and maintenance costs of the system and they are determined in its design phase. The spare parts inventory levels are the other main determinant of availability, downtime costs, and maintenance costs and they are managed during the exploitation phase. These decisions are typically made not only at different points in time, but also by different departments. Design departments aim at meeting a target reliability level by keeping design and production costs as small as possible rather than taking all costs affected by the reliability decisions into account. In general, there is a trade-off among these costs: components with higher reliability and redundancy have higher design and production costs and lower maintenance and downtime costs. Thus, ignoring downtime costs, maintenance costs and the effects of spare parts inventory might lead to suboptimal solutions. A long-term approach for reliability decisions, which incorporates the effects of spare parts inventory will help companies to adapt their design processes to the market trends.

1.2. Reliability Optimization and Upgrading Policy Problem

In this section, we give general definitions of the problems that we study in this thesis.

1.2.1 Reliability Optimization

The reliability of a system is defined as the probability that the system will perform its intended function for a predetermined mission period under a given set of environmental conditions (see Lewis (1996), and Blischke and Murthy (2000)). The main determinants of system reliability are the reliability of its parts and its structure (e.g. simple series, simple parallel, series-parallel, parallel-series, etc.). In general, during the design, for each component that will constitute the system, there exists a set of options with different reliability levels such that the unit costs (prices) of these options increase with their reliability levels. That is, an option is less costly than a more reliable option. Given a system structure, a high system reliability level can be achieved by

- selecting options with high reliability levels, and/or
- redundancy - using a subsystem composed of a number of identical parts in parallel instead of a single part,

In both cases, an increase in system reliability is provided by a higher investment in reliability during the design.

Two main reliability optimization problems are defined with respect to the decision studied:

- Reliability allocation problem: the optimal selection of a design option for each part in a system.
- Redundancy allocation problem: the optimal number of identical parts placed in parallel in each subsystem.

In this thesis, we study a component-level reliability optimization problem which is closely related to the reliability allocation problem. Notice that the reliability allocation problem is on system-level by definition. We also study the redundancy allocation problem. We use availability of a system rather than its reliability as a performance measure.

1.2.2 Upgrading Policy Problem

In some cases, systems in the field do not satisfy certain availability requirements and/or significant maintenance costs and downtime costs are incurred during the exploitation phase. In such a case, the OEM might choose to redesign one or multiple components to improve their reliability. The following problems then have to be solved by the OEM for the redesign:

1. Selection of component(s) for redesign.
2. Determination of the level up to which the reliability of the component(s) will be improved.
3. Determination of the policy for upgrading the systems in the field (e.g., replacing all old parts with the improved ones at once after the redesign or replacing an old part with an improved one only when the old part fails).

We refer to the third problem as the *upgrading policy problem*.

Studying these three problems together would result in very complex models. Thus, we study only the upgrading policy problem. Its solution may be a basis for studying the first and second problem.

1.3. Key Concepts

There are four key concepts that play a role in the problems that we study: Maintenance, spare parts supply, availability, and critical components. Below, we will give necessary definitions related to these concepts and explain their relations.

1.3.1 Maintenance, Repair-by-Replacement, and Repair-on-Site

Maintenance can be defined as a set of actions necessary to sustain and restore the performance, reliability and safety of a system (see Kumar et al. (2000)). The main objective of maintenance is to assure that a system is available for operation when required. Maintenance actions which are planned to avoid unexpected failures and downtime are known as *preventive maintenance* actions while those which are taken whenever a failure occurs are known as *corrective maintenance* actions (see Coetzee (2004)). We focus on corrective maintenance in this thesis as the companies that we cooperated with reported that preventive maintenance has very little impact on the failures of many parts and downtime costs stemming from failures are very high compared to those incurred during preventive maintenance as preventive maintenance is scheduled beforehand.

As we mentioned before, the repair-by-replacement concept is commonly used for system repair. That is, spare parts are kept on stock for a set of components of a capital good and if a part belonging to that set fails, it is replaced with a ready-for-use one from the inventory. However, some parts (e.g. X-ray chain in an X-ray machine) are repaired on customer site rather than being replaced with a ready-for-use one when they fail. The major reasons for applying the *repair-on-site* concept can be listed as follows:

- The replacement of a failed part with a ready-for-use one is more costly and/or technically more difficult than repairing it on site.
- The owner(s) of systems prefer(s) to keep their original parts rather than replacing them with spare ones.

Obviously, spare parts are not kept for parts which are repaired on site.

In this thesis, we study the reliability optimization problems for the situations in which only the repair-by-replacement is applied and the upgrading policy problem for situations in which only the repair-on-site is applied.

1.3.2 Spare Parts Supply, Ordinary Procedure, and Emergency Procedure

In practice, the activities that are executed upon a failure of a part for which repair-by-replacement is applied depend primarily on the status of the spare parts supply and the location(s) where spare parts are stored. If there is a ready-for-use part available from the inventory, the failed part is replaced with the ready-for-use one independently of the status of the system (i.e., whether the system is down or not). We refer to the procedure of such a replacement with a part from inventory as an *ordinary procedure*. In case of an out-of-stock situation, if the system fails or the probability of a system failure becomes significantly high due to the failure of the part, an *emergency procedure* is carried out to replace the failed part; that is, other means of supply for a ready-for-use part are exploited. For example, rather than waiting for a part to be finished at the repair facility, a part may be shipped from a more distant warehouse.

The emergency procedure becomes particularly crucial when a failure of a part leads to a system failure, as typically downtime will be significantly longer if it is not applied. However, the emergency procedure is more costly and takes a longer time than the ordinary procedure (i.e., when parts are in stock). Hence, for a fixed system design, the spare parts inventory level is a key factor affecting the system availability and exploitation phase costs as this influences the need to execute the emergency procedure. Of course, the replacement times and costs in the ordinary procedure and emergency procedure play important roles as well.

1.3.3 Availability

Availability can be defined as the proportion of the time a system is available for operation to the total time that it is required to be in operation (see Moss (1985); Thompson (1999); Birolini (2007)). It is used to measure the combined effect of reliability, maintenance and logistic support on the operational effectiveness of systems. Different types of availability, such as *inherent availability*, *achieved availability*, and *operational availability*, are defined to measure effects of different factors (see Kumar et al. (2000) and Sherbrooke (2004)). In this thesis, we formulate total expected downtime or downtime costs stemming from corrective maintenance actions (the ordinary procedure and emergency procedure) throughout the life cycle of systems rather than using any defined availability of systems with respect to the existing definitions.

1.3.4 Critical Components

A system is composed of a number of components. Some of these components are vital for the functioning of the system (i.e., the failure of a vital part leads to a system failure) while others are not. We refer to the vital components as *critical components*. The focus in reliability optimization and spare parts inventory models is on critical components.

All critical components in a system can be represented as a serial structure from the reliability point of view. The serial structure shows that a failure of any component results in a system failure. This representation does not necessarily mean that the corresponding parts are connected to each other physically in series. For example, a car cannot run if any of its tires is flat, so the tires are connected to each other in series when its reliability is considered while they are not physically in series.

In this thesis, we focus on critical components and consider a serial structure when we deal with multi-component problems.

Table 1.2 Terms

Abstract	Concrete
Capital Good	System
Stage	Subsystem
Component	Part

1.4. Terminology

Terms used during system design represent abstract concepts as a physical system does not exist yet. In general, the same terms are also used for the concrete counterparts of those concepts after the design. The term “system” is a good example for this situation. It may refer to an abstract representation of an object during the design, while it refers to a physical object afterwards. Our problem includes such abstract concepts and their concrete counterparts. In the remainder of this thesis, we use different terms for the abstract and concrete versions of several key concepts for the precision of our descriptions. These terms are given in Table 1.2.

Within the context of this thesis, each unit in a capital good with a serial structure is referred to as a stage. When there is no redundancy in a stage, the stage is composed of a single component. If a stage is designed with redundancy, then it includes a number of units of the same component which are connected to each other in parallel from the reliability point of view. In Figure 1.2, you can see the illustration of a capital good with four stages in a serial structure. Stage 2 is designed with redundancy and has two identical units in parallel, while the other stages are designed without redundancy.

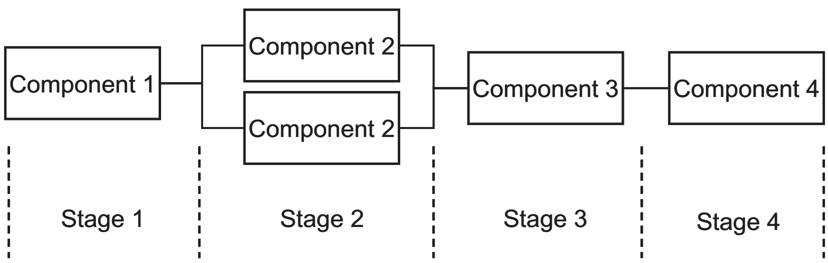

Figure 1.2 The representation of a capital good with four stages in a serial structure

1.5. Literature

There are three major streams of research relevant to the reliability decisions that we focus in this thesis: Reliability optimization, warranty, and spare parts inventory. In this section, we first give an overview of the literature in these streams. Then, we discuss the literature on the upgrading policy problem.

1.5.1 Reliability Optimization

There exists a large number of papers in the reliability optimization literature (see review papers by Kuo and Prasad (2000) and Kuo and Wan (2007) and references therein). The models in these papers deal with either the reliability allocation problem, or the redundancy allocation problem, or both. The reliability of a system (survival probability of a system throughout a predetermined mission period - as defined in Subsection 1.2.1) is used as the performance measure in models for nonmaintained systems, while availability is the performance measure in models for maintained systems. In some cases, one of the availability measures defined in Subsection 1.3.3 can be used (see Vintr and Holub (2001), and Elegbede and Adjallah (2003)), while case-specific availability measures have to be derived in others (see Sharma and Misra (1988)).

In a typical formulation of any model, the objective is the maximization of system reliability/availability against certain constraints, e.g. a budget constraint, a total weight constraint, a total volume constraint. Several formulations include the minimization of acquisition cost (or design cost and production cost) of a system under a reliability/availability constraint (reliability/availability must be greater than or equal to a target level) together with other mentioned constraints. Denoting the decision variables by vector $\vec{x}$, a general formulation for the existing models can be given as

$$
\begin{array}{llll}
(P_0) & \text{min/max} & \pi(\vec{x}) & \\
 & \text{s.t.} & g_i(\vec{x}) \leq b_i & \text{for } i \in \{1,2,...,z\} \\
 & & \vec{x} \in X. &
\end{array}
$$

where $\pi(\vec{x})$ represents the system reliability/availability or acquisition cost of a system, and $g_i(\vec{x}) \leq b_i$, $j \in \{1,2,...,z\}$, represents the relevant constraints. Multiobjective formulations in which maximization of reliability/availability and minimization of acquisition cost are the main objectives have also been introduced. A number of models also employ maximization of percentile life of a system (maximum mission time for which system reliability satisfies at least a certain level) as the objective to cope with uncertain mission times. See Kuo and Prasad (2000) and Kuo and Wan (2007) for extensive review of the models.

Reliability optimization problems are known to be NP-hard (see Chen (1992)). As a result, a large number of papers is devoted to finding efficient optimization algorithms rather than models themselves. The reviews by Kuo and Prasad (2000) and Kuo and Wan (2007) also provide an organized report of algorithms existing in the literature (e.g. heuristics, metaheuristics, exact methods).

An important aspect of this stream of research is that the cost factors are limited to acquisition costs or design and production costs. Quantitative models that incorporate maintenance costs - repair costs in particular - exist mainly in the warranty literature.

1.5.2 Warranty

Normally, a system (product) is sold together with a base warranty and a customer can obtain an additional warranty period against a premium payment. Warranties have different aspects in terms of management, marketing, engineering, logistics and accounting. As a consequence of these various aspects, warranties have been investigated by researchers from different fields (see Blischke and Murthy (1996)). Blischke and Murthy (1992) and Murthy and Blischke (1992a,b) provide an extensive review of the studies conducted on warranty until 1992. The review by Murthy and Djamaludin (2002) covers the later period until 2002.

Quantitative models constitutes an important part of the warranty literature (see Murthy and Blischke (1992b) and Blischke (1990)). These models may differ with respect to warranty policies (see Blischke and Murthy (1992) for a taxonomy for warranty policies), the viewpoint taken (OEM's or customer's), cost elements included, whether the items are repairable or not, etc. In the models, the primary focus is on the optimal length of the warranty period. A general lifetime distribution for items is given and failures throughout the warranty period are modeled as renewal processes. Costs are derived through cost parameters and formulations obtained from the renewal processes. These models were mostly developed for base warranties, however, they also have been used as a basis for long-term warranties (see Murthy and Djamaludin (2002), Rahman and Chattopadhyay (2006), and Chattopadhyay and Rahman (2008)).

As warranty costs depend on the reliability of systems, reliability optimization is also studied in warranty literature. Models developed by Nguyen and Murthy (1988), Hussain and Murthy (2003), Huang et al. (2007) can be considered as reliability allocation models while those introduced by Hussain and Murthy (1998), Monga and Zuo (1998) can be considered as redundancy allocation models. Nevertheless, in all these reliability optimization models, it is assumed that ready-for-use parts that are used for replacement of failed parts are always available and spare parts inventory is not incorporated.

1.5.3 Joint Optimization of Reliability and Spare Parts Inventory

The research on spare parts inventory is extensive. See Muckstadt (2005) and Sherbrooke (2004) for a broad overview of the models and the methods for spare parts inventory. The focus in this thesis is on reliability decisions rather than spare parts inventory. In Chapter 2 and 3, we incorporate the spare parts inventory level into our models as it is a crucial determinant of maintenance costs, downtime costs and availability. In this subsection, we discuss the literature in which reliability and spare parts inventory are considered jointly as is the case in our models.

Kim et al. (2007a,b) study the spare parts inventory and reliability of a single-component system in game-theoretic settings in order to compare certain service contract types. In Kim et al. (2007b), the reliability level is incorporated into the model explicitly and the trade-off between investing in reliability and investing in spare parts is evaluated. The reliability level is indirectly included in the model in Kim et al. (2007a). As the authors' objective is to derive high level managerial insights about the contract types, they develop stylized models in which an overall reliability level for a system is represented rather than the reliability of its components, incorporating redundancy.

Sharma and Misra (1988) consider redundancy and spare parts jointly for a single system with subsystems in a serial structure. Within a subsystem, multiple parts of the same type might be required to function simultaneously for system operation (subsystems with k-out-of-n structure) and the parts are repairable. The decision variables are redundancy level (the number of parts in parallel), the number of spare parts to be bought for each subsystem and the repair capacity. The objective is the maximization of availability of the system subject to several constraints. They develop an algorithm for the solution of the Mixed Integer Program (MIP) arising from their model; their algorithm can solve a formulation with linear constraints. Later, in Misra and Sharma (1991), they generalize their algorithm to a wider range of MIP models for reliability/availability optimization.

1.5.4 Upgrading Policy

The models that are related to the upgrading problem that we study were first introduced for replacement decisions of a part or a system due to technological obsolescence. In practice, new units (parts or systems) which have the same functionality as the old ones in use but with a higher performance often become available in the market. The higher performance could be in terms of reliability, efficiency, energy consumption, purchase cost, etc. In general, the replacement problems are formulated periodically. At each period, one has to decide whether to replace an old unit with one of the available improved ones. Sethi and Chand (1979) and Chand and Sethi (1982) introduce models for deterministic technological changes; that is, the timing and the nature of changes are known with certainty. Nair and Hopp (1992), Nair (1995), and Rajagopalan and M.R. Singh (1998) models cases in which stochasticity in the timing and/or the nature of change is involved.

Mercier and Labeau (2004) study replacement policies that largely overlap with the upgrading policies that we consider (see Section 1.6). We refer to the units which are in use just before the technological change, and improved units provided by the new technology, as *old units* and *new units*, respectively. Mercier and Labeau (2004) investigate situations in which failure rates for both old units and new units are constant. The new units have a lower failure rate and lower energy consumption rate (cost per unit time) compared to that of the old units. They introduce a so-called K *strategy* for a general number, N, of identical and independent units on some finite time interval $[0, T]$. Under this strategy, failed old units are replaced with new ones correctively until the K^{th} failure of the old units, $K \in \{0, 1, ..., N\}$. After the K^{th} failure, the failed part is replaced correctively and the remaining $N - K$ old units are replaced preventively. $K = 0$ and $K = N$ correspond to strategies under which all old units are replaced preventively at time 0 and each old unit is replaced correctively (no preventive replacement), respectively. A new unit is replaced with another new unit with zero lead time when it fails. They calculate the mean total cost over $[0, T]$, which includes replacement costs and energy consumption costs. They discount the total costs to time zero. They show that only three strategies can be optimal: the strategies with $K = 0$, $K = 1$, and $K = N$,respectively.

Mercier (2008) extends the model introduced by Mercier and Labeau (2004) for general failure rates (e.g., with degradation) and show that the optimal strategy can be different than $K = 0$, $K = 1$, and $K = N$; and it depends on the time horizon T.

In Mercier and Labeau (2004) and Mercier (2008), inventory decisions are not incorporated into the models. The new items are available at any instant. In recent papers by Clavareau and Labeau (2009b,a), inventory decisions are incorporated into a Petri net model and a simulation model, respectively, which are developed to investigate the K *strategy*. The inventory is managed by the so-called point command method in Clavareau and Labeau (2009b), while a modified version of the Economic Order Quantity is used in Clavareau and Labeau (2009a). These models also include other details (e.g., different types of maintenance actions, limited maintenance capacity, priority rules for different actions, effectiveness of a repair, etc.). The interaction between the inventory decisions and the optimal strategy is not established explicitly in these models.

As a final remark in this section, we use the term upgrading rather than replacement to avoid the confusion with the repair-by-replacement concept. Within the context of this thesis, repair-by-replacement means replacement of a part with another of the same type; that is, the parts which is used for replacement is not an improved one.

1.6. Contributions of the Thesis

The main goal of this thesis is to develop quantitative models and methods for the optimal reliability decisions for advanced capital goods. As we stated before, we study the reliability optimization problems for situations in which repair-by-replacement is used, which means that spare parts inventory is kept and it is a key factor affecting availability and exploitation phase costs of capital goods.

In practice, OEMs and their customers often only consider the initial costs (design and production costs or acquisition costs) for their reliability decisions. A similar approach is also followed in the reliability optimization literature. However, our exploratory studies and LCC calculations at companies involved in the IOP-IPCR project revealed that the exploitation phase costs (maintenance costs and downtime costs) can be considerably higher than the initial costs of the systems. Models and methods which take into account the initial costs and the exploitation phase costs for reliability decisions, are not only relevant but also necessary for both OEMs and their customers due to the market trends.

As the reliability optimization models in the warranty literature include costs from the warranty claims during the exploitation phase of systems, they have the potential to assist companies in their decisions. However, these models lack the focus on availability and they do not incorporate spare parts inventory.

The existing models that do consider reliability and spare parts jointly and/or incorporate exploitation phase costs (Kim et al. (2007a,b), Sharma and Misra (1988), and Misra and Sharma (1991)) have several limitations for the cases that we consider. First, the situations in these models do not involve any emergency procedures. That is, when a part in a system fails and there is no ready-for-use part available from a warehouse, the system is down until a part is available from a repair facility, meaning that downtime can be considerably long. This is unrealistic within the companies we studied. As stated before, Kim et al. (2007b) and Kim et al. (2007a) use stylized models for comparison of certain service contract types. Sharma and Misra (1988) and Misra and Sharma (1991) do not include maintenance costs. In addition, these models are developed for a single system and spare parts are dedicated to this single system. In practice, spare parts are usually stocked for multiple systems at a central location and there is a pooling effect on spare parts which is not captured in these models.

Consequently, there is a need to develop models for the reliability optimization problems which include the following attributes:

- maintenance costs
- downtime costs or availability (or downtime) constraints
- spare parts inventory
- an emergency procedure

Such a reliability optimization problem can be a single-stage problem or a multi-stage problem. While studying a multi-stage problem, once its relation to relevant single-stage problems can explicitly be established (e.g., decomposition of the multi-stage problem into single-stage problems), one can first analyze and solve the single-stage problems and use these findings to analyze and solve the multi-stage problem. We thus start with a single-stage problem in Chapter 2 and contribute to the literature with the followings:

- We develop a reliability optimization model for critical components and an efficient solution procedure for the resulting problem formulation.
- We derive insights about how certain factors, such as component type (cheap, medium, expensive), the size of the installed base, the downtime penalty rate, and the lifetime of the system, affect the optimal reliability decision.

The approach followed for single-stage problems in Chapter 2 serves as a basis to analyze and solve the related multi-stage problem.

Next, we study a multi-stage redundancy allocation problem for capital goods in Chapter 3. Our contribution in Chapter 3 can be listed as follows:

- We develop a redundancy allocation model for capital goods and establish its relation to the relevant single-stage problems.
- We develop an algorithm to solve the single-stage problems and the redundancy allocation problem.
- We provide insights on the optimality of having redundancy by deriving results for the single-stage and the multi-stage problems.

The two models introduced in Chapter 2 and Chapter 3 include the key aspects such as maintenance costs, spare parts inventory, and emergency procedure (listed above) and they are developed for multiple systems. The effect of downtime is incorporated differently in the two models. In the reliability optimization model for critical components (Chapter 2), downtime costs are included in the model, while there is a constraint on the total uptime (or downtime) throughout the lifetime of a number of systems in the redundancy allocation model (Chapter 3). These two approaches are consistent with each other.

Reliability decisions during the redesign of components for improvement also fit within the broad scope of our research. We focus on the upgrading policy problem of critical components for which repair-on-site is applied in Chapter 4. We study two major policies that OEMs follow for the upgrading of N systems, each of which includes a single unit of the old parts (we denote the time just after the redesign by time 0):

- Policy 1 - Upgrade all systems preventively at time 0: All the old parts are preventively replaced with the redesigned components immediately after the redesign.
- Policy 2 - Upgrade systems one-by-one correctively: An inventory of redesigned component is kept. As an old component in the field fails, it is correctively replaced with a redesigned one from the inventory.

Notice that Policy 1 and 2 are the special cases of the K strategy introduced in Mercier and Labeau (2004) with $K = 0$ and $K = N$, respectively (see Subsection 1.5.4). We investigate a situation in which the initial order quantity (initial supply quantity) for the inventory under Policy 2 is one of the main factors that affects the costs incurred for upgrading the systems. Remember that Mercier and Labeau (2004) and Mercier (2008) study the K strategy without inventory considerations. Clavareau and Labeau (2009b,a) do incorporate inventory decisions into their investigation of the K strategy, but as mentioned in Subsection 1.5.4, they use predefined methods and fix the order quantities with respect to certain parameter values rather than formulating the costs affected by these decisions and optimizing them. Furthermore, the effect of inventory decisions cannot be explicitly observed in their models as there are a number of other details incorporated into these models.

Our contribution in Chapter 4 is as follows:

- We develop a quantitative model for the upgrading problem with Policy 1 and Policy 2. We formulate the interaction between the initial supply quantity and the costs affected by the initial supply quantity under Policy 2 explicitly.
- We develop an efficient solution procedure for the optimal initial supply quantity in Policy 2.
- We derive insights on conditions under which each policy is optimal.

1.7. Outline of the Thesis

This thesis is composed of two parts, one devoted for the reliability optimization and the other for the upgrading policy problem. Part 1 consists of Chapter 2 and Chapter 3, in which we study the optimal reliability level of critical components and redundancy allocation for serial systems, respectively. The models that we introduce in these chapters include the following attributes that are essential for capital goods: maintenance costs, downtime costs or a downtime constraint, spare parts inventory and an emergency procedure. Part 2 is constituted by Chapter 4, in which we investigate the optimality of the two upgrading policies that are mentioned in Section 1.6. In Appendix A, we provide monotonicity and supermodularity results for the Erlang Loss System, which are motivated by the reliability optimization problem that we study in Chapter 2.

The research presented in Chapters 2, 3, 4, and Appendix A is based on Öner et al. (2010b,a,c, 2009), respectively.

Chapter 2

Optimization of Component Reliability

2.1. Introduction

As we mentioned in Chapter 1, we focus on critical components in this thesis. In this chapter, we start our investigation of reliability decisions for capital goods by studying the optimization of the reliability of critical components. We present a quantitative model to support the decision on the reliability level of a critical repairable component during the design phase of a capital good. We investigate a situation in which an OEM will sell a number of units of the same system together with a Performance-Based (PB) service contract which covers the life time of a system. The PB contract specifies multiple service aspects including a downtime penalty; that is, the OEM pays a certain amount of money to its customers per unit of downtime. The systems are installed in one region that is served by a single spare parts inventory stock point which is at a sufficiently close distance from all systems. Our objective is the minimization of the portion of the system's Life Cycle Costs (LCC) which is affected by the component's reliability, as measured by its Mean Time Between Failures (MTBF) and the spare parts inventory level.

In order to cover all costs, we need to formulate the maintenance costs and downtime costs as a function of the reliability level. However, the spare parts inventory level is also a crucial determinant of these costs. It is not a given parameter during the design phase,but a decision variable which is set (optimized) with respect to the reliability level later. We aim at providing a decision support model which makes the best use of available data/information in the simplest way during the design phase: Our joint optimization helps the OEM foresee the effect of the reliability level decision on spare parts inventory level, and, ultimately, the total maintenance and downtime costs.

In a recent paper, Murthy et al. (2004) highlight the current issues and challenges in product warranty logistics. They underline the need for linking the spare parts inventory levels to failures of parts, i.e., to component reliability. As stated in Subsection 1.5.3, the spare parts inventory level and component reliability have been jointly studied in recent work by Kim et al. (2007b,a). Recall though that the models in these papers are stylized ones which do not include any emergency procedures. This aspect is important for capital goods as it has a large impact on downtimes and thus the related costs. Incorporation of this aspect leads to a more complex model than the models in Kim et al. (2007a,b). Furthermore, to simplify the analysis, the normal approximation for the lead-time demand is used in these previous papers. We provide an exact analysis for the LCC function, which enables us to derive an exact optimization procedure.

The contributions of this chapter can be stated as follows:

- First, we propose a new decision support model to determine the reliability of a critical component in the design phase. In this model, we explicitly formulate the relationship between the reliability level of the component and its spare parts inventory level, incorporating design costs, production costs and service costs (including downtime costs).
- Second, we perform an exact analysis on the LCC and we derive several of its analytical properties.
- Third, we provide an efficient optimization algorithm.
- Fourth, we provide managerial insights through a numerical experiment which is based on real-life data. We compare costs obtained under our joint optimization method to costs obtained via a non-integrated method. In our experiment, we show that joint optimization leads to an average cost reduction of 44.3% and the optimal reliability level significantly depends on component type, the size of the installed base, the downtime penalty rate, and the lifetime of the system. We also perform sensitivity analysis and show that the average extra costs that would be incurred is negligibly small for most of the cases with parameter values even ±50% off the values that we have in our numerical experiment.

Our model in this chapter is closely related to reliability allocation models. Through the discussions in Chapter 1, we can deduce that the following attributes are fundamenta in reliability optimization models for advanced capital goods:

Table 2.1 Comparison of papers

Attribute	Hussain and Murthy (2003)	Huang et al. (2007)	Nguyen and Murthy (1988)	Kim et al. (2007a,b)	This chapter
Maintenance costs	X	X	X	X	X
Downtime costs					X
Multiple systems		X		X	X
Spare parts				X	X
Emergency Procedure					X

- Maintenance costs (due to their high magnitude),
- Downtime costs or availability targets (requirements-constraints) (due to high downtime costs),
- Multiple systems (due to the pooling effect on spare parts),
- Spare parts (due to its large effect on maintenance costs and availability of systems), and
- an emergency procedure (as it is a common practice which limits the downtime significantly).

Several of these attributes are covered in reliability allocation models in the warranty literature and the contracting literature; see subsections 1.5.2 and 1.5.3. In Table 2.1, you can see a comparison of this chapter and the most related papers from these literatures. We include the papers which have at least one of the given attributes.

The problem in this chapter is on a component-level. Remember that critical components in a capital good form a serial structure (see Subsection 1.3.4). Within the context of this chapter, a critical component corresponds to a stage in a capital good with series structure, as redundancy is not considered. Thus, we can equivalently state that the problem in this chapter is a single-stage problem. The problem can serve as a building block for a multi-stage problem which can be decomposed into single-stage problems, each of which corresponds to an instance of the problem studied in this chapter. Once the decomposition is established, one can analyze and solve the multi-stage problem by making use of the analysis and the solution procedure developed in this chapter.

In Chapter 3, we work on a redundancy allocation problem, which is a multi-stage problem by definition, and show its relation to the single-stage problems that are generated by its decomposition. In that chapter, we first derive results for the single-stage problems and then provide results for the redundancy allocation problem by making use of the results of the single-stage problems. The multi-stage problem analogous to the problem introduced in this chapter could be tackled in a similar way as followed in Chapter 3.

The outline of this chapter is as follows. In Section 2.2, we present our model assumptions and problem formulation. We derive the LCC function and provide a number of analytical properties and an optimization procedure in Section 2.3. We give the setting and the results of our numerical experiment in Section 2.4. We conclude the chapter by Section 2.5.

2.2. Model

An OEM is designing/developing a critical repairable component for a capital good (system). The OEM estimates that he will sell N units of the same system ($N \in \mathbb{N} = \{1, 2, 3, ...\}$) after the design of the capital good (system). The OEM expects that he will sell each system with a service contract. The lifetime (exploitation phase) of each system is estimated to be T, which is in the order of 10-30 years and considerably longer than the design and production phases (e.g. at least ten times longer). We assume that design costs and production costs are incurred at time 0. The OEM plans to offer the service contracts for the estimated lifetime of the systems (T).

We assume that the N systems are all sold at time $t = 0$. During the exploitation phase of the systems, which we denote by $[0, T]$, the OEM will keep a spare parts inventory of the critical component. The OEM will produce s parts for the spare parts inventory at time $t = 0$; s is a decision variable. When the part in a system fails, the system will be repaired by replacement of the failed part with a ready-for-use one when it is available from the inventory. The OEM cooperates with a secondary supplier for the emergency replenishment of a part in an out-of-stock situation.

2.2.1 Failure and Repair Processes

In general, failures of the repairable parts in the systems will depend on the parts' quality, how the systems will be used and the usage conditions (e.g. environmental conditions). We assume that no systematic failure will occur due to deficiencies in production of the systems, so the parts, will satisfy all quality specifications at the beginning of their lives. We denote the MTBF of the component by τ which is a decision variable to be fixed during the design of the component. We assume that there is a lower bound $\underline{\tau}$ for the MTBF. This lower bound can have two interpretations: It can be a target reliability level which is determined with respect to customer / market expectations; or it can be the reliability level of an already existing component in the market. (Note that, $\underline{\tau}$ is not the minimum MTBF that can be achieved (technically) in any case.) We also assume that there is a limit $\bar{\tau} > \underline{\tau}$ that the MTBF can be improved up to; i.e., there is a limitation on the level to which the reliability can be improved.

The systems will be supported by a single warehouse where all spare parts are stocked. There is a single repair facility where defective parts will be repaired.

Upon the failure of a system at a customer site, one of the following two procedures will be applied depending on the availability of a part from the spare parts inventory:

1. *Ordinary Procedure*: If a ready-for-use part is available from the inventory, it will be transported to the customer site. A service engineer will visit the customer site and repair the system by replacing the defective part with the ready-for-use one. The defective part will be transported to the repair facility for a repair. After repair, the part will be restored to an as-good-as-new condition and added to the spare parts inventory.

2. *Emergency Procedure*: If there is an out-of-stock situation, an as-good-as-new part will be replenished from the secondary supplier and it will be transported to the customer site as soon as possible. The failed part will be replaced with the replenished part and will be returned to the secondary supplier. We assume that the secondary supplier has ample supply of the parts.

The emergency procedure assures that downtimes are always short (i.e., even in out-of-stock situations). The spare parts inventory is affected only by the ordinary procedure. In this procedure, a part is removed form the inventory when a failure occurs, and each failed part is immediately sent to the repair and added to the inventory after its repair. In out-of-stock situations, demand is lost for the inventory. Hence the spare parts inventory position (the sum of pipeline stock and actual stock-on-hand) is kept at a constant and will be equal to the initial amount of spare parts that one stocks in the warehouse which is s. As a result, we may also say that the spare parts inventory is controlled by a continuous-review *basestock policy* with basetock level s; see books by Sherbrooke (2004) and Muckstadt (2005).

We assume that the total stream of system failures will follow a Poisson process with rate N/τ. The memoryless property of the Poisson failures implies that there is no aging (degradation) effect. This assumption is justified when the number of the systems (N) is sufficiently large or if lifetimes of parts are close to exponential. In reality, when a system is down due to a failure of the critical part in it, the failure rates of all of the parts decreases during the downtime of the system as there is one less system contributing for the total stream of failures. However, owing to the short downtimes and large number of systems, we neglect this effect and assume that the failure rate is constant, as is standard in the spare parts inventory literature. This simplifies the analysis considerably and has been demonstrated to be a benign assumption; see books by Sherbrooke (2004) and Muckstadt (2005).

We assume that when a part fails during the exploitation phase, it will be diagnosed with 100% accuracy in a negligibly short time. The system will become operational just after the replacement of the failed part. The downtime after a failure is equal to the replacement time of the part. Replacement times in the applications of the ordinary procedure are independently and identically distributed with mean μ_1. Similarly, replacement times in the applications of the emergency procedure are independently and identically distributed with mean μ_2. μ_1 and μ_2 are in the order of 1-48 hours and we assume that $\mu_1 \leq \mu_2$.

The repair facility has planned leadtimes for all repairs. This is also a standard assumption in the spare parts inventory literature (see Sherbrooke (2004) and Muckstadt (2005)). This fits best to situations in which the repair facility is outsourced and a limit on repair time to repair is specified in an agreement. We assume that repair times which include time to transport the part to and from the repair facility are independent and identically distributed with mean $U > 0$ which is in the order of 1-4 months. The orders of magnitude imply that μ_2 is significantly shorter than U, which reflects the users incentive to apply the emergency procedure.

2.2.2 Problem Formulation

We formulate our problem as

$$
\text{(P)} \qquad \begin{array}{ll} \min & \pi(\tau, s) \\ \text{s.t.} & \underline{\tau} \leq \tau \leq \bar{\tau} \\ & s \in \mathbb{N}_0 = \{0, 1, 2, ...\}. \end{array}
$$

where $\pi(\tau, s)$ is the expected Net Present Value (NPV) of the LCC of the N systems affected by the reliability of a part, τ, and the spare parts inventory position, s. In our model, LCC consists of design costs, production costs, spare parts inventory costs, repair costs, and downtime costs. We assume that design costs and production costs are incurred at time 0 and formulate the NPVs of the other costs, which occur throughout $[0, T]$, at time 0. We denote the discount rate by $\alpha > 0$; a cost of 1 at time t contributes $e^{-\alpha t}$ to the NPV (notice that $\alpha = 0$ would correspond to no discounting). We use the following notation to refer to these costs and cost parameters relevant to these costs.

$K(\tau)$:	The expected NPV of the design costs of the component.
$P(\tau)$:	The expected NPV of the production costs of the parts that will be installed in the N systems.
$S(\tau, s)$:	The expected NPV of the spare parts costs
$R(\tau, s)$:	The expected NPV of the system repair costs that stem from the failures of the repairable parts.
$D(\tau, s)$:	The expected NPV of the downtime costs that stem from the failures of the repairable parts.
h:	The storage cost rate per part ($h > 0$).
r_1:	Expected cost of an ordinary repair.
r_2:	Expected cost of an emergency repair.
p:	Downtime penalty rate ($p > 0$).
d_1:	Expected downtime penalty incurred because of a failure leading to an ordinary repair.
d_2:	Expected downtime penalty incurred because of a failure leading to an emergency repair.

Obviously,

$$
\pi(\tau, s) = K(\tau) + P(\tau) + S(\tau, s) + R(\tau, s) + D(\tau, s) \tag{2.1}
$$

The factors r_1 and r_2 include all costs originating from the corresponding procedures, such as administrative costs, costs of one or more visits of a service engineer, transportation costs, repair costs of a failed part, and storage costs during the repair lead time at the repair facility. We assume that $hU \leq r_1 \leq r_2$, as emergency repairs require more expensive activities than ordinary repairs do and r_1 and r_2 include storage costs at the repair facility. We also assume that r_1 and r_2 are immediately incurred when a failure occurs.

Upon a system failure, an average downtime cost of $d_1 = p\mu_1$ is incurred if an ordinary repair is performed and an average downtime cost of $d_2 = p\mu_2$ is incurred if an emergency repair is performed. We assume that downtime costs are immediately incurred when a failure occurs. As $\mu_1 \leq \mu_2$, $d_1 \leq d_2$.

Since $\underline{\tau}$ is the target reliability that the manufacturer has to provide, we define the function $K(\tau)$ for the extra design cost that would be incurred to improve reliability to τ, $\underline{\tau} \leqslant \tau \leqslant \bar{\tau}$. Thus, $K(\underline{\tau}) = 0$. Design costs of a component can be derived by analyzing data of previous versions of the component or data of a similar component. In general, design costs are assumed to be an increasing convex function of the reliability level (see Mettas *et al.* Mettas (2000) and Kim *et al.* Kim et al. (2007b)). We also assume that $K(\tau)$ is an increasing convex function of τ.

Production costs include all the costs incurred for the production of these N components that will be installed in the systems. The production cost per part is $c(\tau)$ which is an increasing convex function of τ. Then $c(\underline{\tau})N$ is the baseline production cost. This fixed amount will be invested by the manufacturer regardless of the choice τ. Thus, we include $P(\tau) = \Big[c(\tau) - c(\underline{\tau})\Big] N$ in our model. This is the extra production costs that is incurred when N parts are produced with an MTBF of τ instead of $\underline{\tau}$.

2.3. Analysis

In this section, we first give the formulations of the spare parts costs, the repair costs, and the downtime costs. Next, we derive a number of analytical properties of the LCC function. We finalize the section by providing an optimization algorithm based on those analytical properties.

2.3.1 Analysis of the Cost Functions

The average spare parts inventory, the number of ordinary and emergency repairs and the downtime throughout $[0, T]$ depend on the out-of-stock probability of the spare parts inventory. This out-of-stock probability is denoted by $G(\tau, s)$ and we start by determining this function.

The factors affecting the stock-on-hand process of the spare parts inventory and their effects are as follows. Demands arrive at the spare parts inventory according to a Poisson process with rate N/τ. Upon the arrival of a demand, if a ready-for-use part is available from the inventory, the demand is satisfied (a part is taken from the inventory) and a part is added to the inventory after a generally distributed lead time with mean U. If there is an out-of-stock situation, the emergency procedure is applied and the demand is lost. Remember that the spare parts inventory position (sum of stock on hand and pipeline) is kept at a constant level which is set to s at time 0. Consequently, the stock-on-hand process of the spare parts inventory is identical to the process for the number of free servers in an Erlang loss system (also denoted as the $M/G/s/s$ queueing system) with an arrival rate N/τ, mean service time U, and s servers. Hence, $G(\tau, s)$ is equal to the Erlang loss probability (see Cooper (1982)), and we obtain

$$G(\tau, s) = \frac{\frac{(NU/\tau)^s}{s!}}{\sum_{i=0}^{s} \frac{(NU/\tau)^i}{i!}}. \tag{2.2}$$

Later on we will exploit the following property of $G(\tau, s)$.

Property 2.1 $G(\tau, s)$ is strictly decreasing and strictly convex in τ.

Proof: Our formulation of the Erlang loss probability $g(\tau, s)$ is mathematically equivalent to the Erlang loss probability given by Harel (1990) with $\lambda = NU$ and $\mu = \tau$, where λ is the arrival rate and μ is the service rate in their notation. In Harel (1990), Harel shows that the Erlang loss probability is strictly decreasing and strictly convex in μ. This is equivalent to $G(\tau, s)$ being strictly decreasing and strictly convex in τ. $\square$

The spare parts costs $S(\tau, s)$ are the sum of spare parts investment costs, $S_1(\tau, s) = c(\tau)s$, and spare parts storage costs, $S_2(\tau, s)$. The formulations of $S_2(\tau, s)$, repair costs $R(\tau, s)$, and downtime costs $D(\tau, s)$ are given in Lemma 2.2 below. After that, in Lemma 2.3 and Lemma 2.4 we provide the monotonicity properties of $S(\tau, s)$, $R(\tau, s)$, and $D(\tau, s)$. In the proof of Lemma 2.2, we will exploit the property stated in Lemma 2.1, which therefore is presented first.

Lemma 2.1 The numbers of applications of the ordinary procedure and that of the emergency procedure performed throughout $[0, T]$ have Poisson distributions with means $(N/\tau)T[1 - G(\tau, s)]$ and $(N/\tau)TG(\tau, s)$, respectively.

Proof: Let Q be the random variable representing the length of time when there is a part available from the inventory throughout $[0, T]$. Then, $E[Q] = T\left[1 - g\left(\tau, s\right)\right]$. Let M_1 denote the random variable for the number of ordinary repairs throughout $[0, T]$. As the failures follow a Poisson process with rate $N\tau^{-1}$, for a given $Q = x$, M_1 has Poisson distribution with mean $N\tau^{-1}x$ and

$$E[M_1] = \int_0^T E[M_1|Q = x]dx = \int_0^T \frac{N}{\tau}xdx = \frac{N}{\tau}E[Q] = \frac{N}{\tau}TG\left[1 - (\tau, s)\right].$$

The proof for the number of emergency repairs can be obtained similarly. □

Lemma 2.2 It holds that:

(i)

$$S_2(\tau, s) = \frac{h}{\alpha}\left(1 - e^{-\alpha T}\right)\left[s - \frac{NU}{\tau} + \frac{NU}{\tau}G(\tau, s)\right]. \tag{2.3}$$

(ii)

$$R(\tau, s) = \left[1 - G\left(\tau, s\right)\right]\frac{N}{\tau}\frac{r_1}{\alpha}\left(1 - e^{-\alpha T}\right) + G\left(\tau, s\right)\frac{N}{\tau}\frac{r_2}{\alpha}\left(1 - e^{-\alpha T}\right). \tag{2.4}$$

(iii)

$$D(\tau, s) = \left[1 - G(\tau, s)\right]\frac{N}{\tau}\frac{d_1}{\alpha}(1 - e^{-\alpha T}) + G(\tau, s)\frac{N}{\tau}\frac{d_2}{\alpha}(1 - e^{-\alpha T}). \tag{2.5}$$

Proof: See the Appendix at the end of the chapter. □

The equations in Lemma 2.2 can be explained as follows: In equation (2.3), $\frac{h}{\alpha}\left(1 - e^{-\alpha T}\right)$ can be interpreted as the NPV of the cost of storing one unit of spare parts throughout $[0, T]$ while $\left[s - \frac{NU}{\tau} + \frac{NU}{\tau}G(\tau, s)\right]$ is the average inventory in the steady state. The factor $\frac{r_1}{\alpha T}\left(1 - e^{-\alpha T}\right)$ is the NPV of the cost of an arbitrary ordinary repair and its multiplication with the expected number of ordinary repairs throughout the lifetime, $[1 - G\left(\tau, s\right)](N/\tau)T$, results in the first term of equation (2.4), which is the expected NPV of the cost of ordinary ordinary repairs throughout $[0, T]$. The second term of equation (2.4) can be explained in the same way as the expected NPV of the costs of ordinary repairs. The downtime costs given by equation (2.5) have the same interpretation as the repair costs.

By Lemma 2.2, we can write equation (2.1) as

$$\begin{aligned}\pi(\tau, s) \quad = \quad & K(\tau)+\left[c(\tau)-c(\underline{\tau})\right] N+c(\tau) s+\frac{h}{\alpha}\left(1-e^{-\alpha T}\right)\left[s-\frac{N U}{\tau}+\frac{N U}{\tau} G(\tau, s)\right] \\ & +[1-G(\tau, s)] \frac{N}{\tau} \frac{r_1+d_1}{\alpha}\left(1-e^{-\alpha T}\right) \\ & +G(\tau, s) \frac{N}{\tau} \frac{r_2+d_2}{\alpha}\left(1-e^{-\alpha T}\right) . \end{aligned} \tag{2.6}$$

Lemma 2.3 For a fixed value of s, the following monotonicity properties hold:

(i) $S(\tau, s)$ is strictly increasing in τ for $s > 0$; $S(\tau, s) = 0$ for $s = 0$.

(ii) $R(\tau, s)$ and $D(\tau, s)$ are strictly decreasing in τ.

Proof: See the Appendix at the end of the chapter. □

Recall the assumption that $K(\tau)$ and $c(\tau)$ are increasing in τ. Together with this assumption, Lemma 2.3 reflects the conflicting behavior of the costs included in $\pi(\tau, s)$ for varying reliability levels: Some parts of $\pi(\tau, s)$ are increasing in τ while others are decreasing in τ.

Lemma 2.4 For a fixed value of τ, the following monotonicity properties hold:

(i) $S(\tau, s)$ is strictly increasing in s.

(ii) $R(\tau, s)$ and $D(\tau, s)$ are decreasing in s.

Proof: See the Appendix at the end of the chapter. □

Lemma 2.4 shows that there is also a trade-off among costs for varying spare parts level.

2.3.2 Solution Procedure for Problem (P)

We give the properties of $\pi(\tau, s)$ that we exploit for its optimization in Lemma 2.5 below.

Lemma 2.5 $\pi(\tau, s)$ has the following properties:

(i) For a fixed $\tau \in [\underline{\tau}, \bar{\tau}]$, $\pi(\tau, s)$ is strictly convex in s.

(ii) For all $\tau \in [\underline{\tau}, \bar{\tau}]$, $\lim_{s\to\infty} \pi(\tau, s) = \infty$.

(iii) For a fixed $s \in \mathbb{N}_0$, $\pi(\tau, s)$ is strictly convex in τ.

(iv) For a fixed $\tau \in [\underline{\tau}, \bar{\tau}]$, define $s^*(\tau) = \min\{\arg\min_{s\in\mathbb{N}_0} \pi(\tau, s)\}$, i.e., $s^*(\tau)$ is the smallest value of s under which $\pi(\tau, s)$ is minimized. Then, $s^*(\tau)$ is decreasing in τ.

Proof: See the Appendix at the end of the chapter. □

(i) and (ii) in Lemma 2.5 imply that $s^*(\tau)$ is finite for a fixed value of τ and can be found by standard procedures for optimization in one variable. Let (τ^*, s^*) be a minimizer of $\pi(\tau, s)$. By (iv), $s^*(\bar{\tau}) \leq s^* \leq s^*(\underline{\tau})$. (iii) implies that we can also find the optimal value of τ for a fixed value of s, which we denote by $\tau^*(s)$. Then, an optimal solution (τ^*, s^*) can be found by enumerating all solutions $(\tau^*(s), s)$ for $s^*(\bar{\tau}) \leq s \leq s^*(\underline{\tau})$.

Theorem 2.1 The following procedure determines an optimal solution of of problem (P):

1. Find $s^*(\bar{\tau})$ and $s^*(\underline{\tau})$.

2. For each $s = s^*(\bar{\tau}), s^*(\bar{\tau}) + 1, ..., s^*(\underline{\tau})$, solve the problem $\{\min\ \pi(\tau, s),$ s.t. $\underline{\tau} \leq \tau \leq \bar{\tau}\}$. Let $\tau^*(s)$ be an optimal τ for a given s.

3. $(\tau^*, s^*) = \arg\min_{(\tau^*(s),s)}\{\pi(\tau^*(s), s),\ s = s^*(\bar{\tau}), s^*(\bar{\tau}) + 1, ..., s^*(\underline{\tau})\}$ is an optimal solution and $\pi^* = \pi(\tau^*, s^*)$ is the corresponding minimum LCC.

2.4. Numerical Results

In this section, we present a numerical experiment which is based on healthcare systems data. We identify four main factors in our model, generate a testbed of 81 instances by having three choices for each factor, and compute the optimal decision parameters. Since our model is used during the design phase of a system and there is not much data or information available in this phase, the estimates of the parameters in our model are vulnerable to errors. Therefore, we also perform a sensitivity analysis.

2.4.1 Testbed

We use the following modified version of the design cost function introduced by Mettas (2000) (see Huang et al. (2007) as well) in our numerical experiments:

$$K(\tau) = B_1 \left[\exp\left(k \frac{\tau - \underline{\tau}}{\tau_\infty - \tau} \right) - 1 \right], \qquad \underline{\tau} \leq \tau \leq \bar{\tau},$$

where B_1 and k are strictly positive factors and τ_∞ is a given reliability level that exceeds $\bar{\tau}$ (i.e. $\bar{\tau} < \tau_\infty$). Notice that under this definition, reliability improvement becomes infeasible already before the costs become infinite. k is a parameter that represents the difficulty in increasing MTBF due to complexity, limited resources and technology, etc. Larger values of k correspond to more difficulties in increasing MTBF, so, higher design costs.

We formulate the unit production cost function as

$$c(\tau) = A + B_2(\tau^m - \underline{\tau}^m), \qquad \underline{\tau} \leq \tau \leq \bar{\tau},$$

where $A \geq 0$, $B_2 > 0$, and $m \geq 1$. This is a modified version of the unit production cost function used by Huang et al. (2007). In their paper, Huang *et al.* consider a situation in which production is carried out for a considerable duration and incorporate a learning effect on top of an initial unit cost in their unit production cost formulation. Since the production phase is negligibly short in our case, we omit the learning effect and our formulation is similar to their initial unit cost function.

We investigate the effect of four factors on optimal decisions: component type (explained below), number of systems N, downtime penalty rate p, and length of the exploitation phase T. We created 81 instances by all combinations of three choices of the four factors. The choices of the factors are given in Table 2.2.

Component type reflects the value of a component in monetary terms. At the smallest possible MTBF, $\underline{\tau}$, the unit production cost of an expensive component is larger than that of a cheaper one. Furthermore, a certain amount of improvement in MTBF leads to a higher increase in both the unit production cost and design cost of an expensive component compared to a cheaper component. We realize the choices of component type mainly through the parameter B_1 in $K(\tau)$ and the parameters A and B_2 in $c(\tau)$.

Table 2.2 Choices of the factors

Component type	N	T (months)	p ($ per hour)
cheap, medium, expensive	100, 500, 2500	60, 120, 240	100, 500, 2500

Table 2.3 Parameter values for component types

Comp. Type	Parameters					
	B_1 (\$)	A (\$/mt.)	B_2 (\$)	h (\$/mt. per pt.)	r_1 (\$ per rp.)	r_2 (\$ per rp.)
1 (ch)	200000	1000	10	20	600	1200
2 (med)	2000000	10000	100	200	1500	3000
3 (exp)	20000000	100000	1000	2000	10500	21000

h (the storage cost rate per part), r_1, and r_2 are also varied for different choices of component type since each include a variable part which is positively correlated with the design costs and unit production cost. We use $k = 1$ and $m = 1$ in $K(\tau)$ and $c(\tau)$, respectively, for all three types of components. The other parameter values for different component types are given in Table 2.3. In the table, comp., mt., pt., and rp. stands for component, month, part, and repair, respectively.

Table 2.4 shows the values of the other parameters, which are fixed.

2.4.2 Results and Managerial Insights

We call the method where decisions on reliability level and the spare parts inventory level are made separately the *non-integrated method*. In the non-integrated method, while deciding on the reliability level, the focus is only on the design costs and the production costs. Then, for any values of the parameters of the cost functions, the OEM sets MTBF at the target value $\underline{\tau}$ since the sum of design costs and production costs has its minimum value when $\tau = \underline{\tau}$. Next, the summation of the other cost terms in $\pi(\tau, s)$, which belong to the exploitation phase, are optimized by the optimal inventory level s for $\tau = \underline{\tau}$. Thus, the LCC found by the non-integrated method is $\pi_n = \pi(\underline{\tau}, s^*(\underline{\tau}))$. Denoting the optimal LCC in the joint optimization case as π^*, we define the *relative cost reduction* achieved by the joint optimization as

$$\Delta\pi = \frac{\pi_n - \pi^*}{\pi_n}(100\%).$$

We present the results of the experiment in Table 2.5. The numbers given in the

Table 2.4 Values of fixed parameters

$\underline{\tau}$ (mt.)	$\bar{\tau}$ (mt.)	τ_∞ (mt.)	μ_1 (hours)	μ_2 (hours)	L (mt.)	α (per year)
24	240	360	10	50	3	0.05

Table 2.5 Results of the experiment

		Avg. τ^*	min τ^*	max τ^*	Avg. $\Delta\pi$	min $\Delta\pi$	max $\Delta\pi$
Comp.	1 (ch)	162.63	68.91	240.00	72.6%	42.4%	88.4%
	2 (med)	82.21	31.99	183.38	43.2%	6.1%	76.5%
	3 (exp)	42.63	24.58	74.40	17.0%	0.1%	44.7%
N	100	79.96	24.58	202.92	39.0%	0.1%	84.3%
	500	99.18	28.17	240.00	45.8%	2.0%	87.3%
	2500	108.32	29.03	240.00	47.9%	2.7%	88.4%
p	100	62.18	24.58	148.68	29.7%	0.1%	70.6%
	500	91.82	27.36	225.89	43.2%	1.3%	82.7%
	2500	133.47	36.61	240	59.9%	11.5%	88.4%
T	60	79.82	24.58	240.00	35.9%	0.1%	85.4%
	120	96.21	30.61	240.00	44.7%	4.1%	87.4%
	240	111.44	36.78	240.00	52.1%	11.3%	88.4%
All		95.82	24.58	240.00	44.3%	0.1%	88.4%

columns named Avg. τ^* and Avg. $\Delta\pi_n$ are the average optimal MTBF values and the average relative cost reduction values, respectively. For example, the values 162.63 and 72.6% in the first row are the average optimal MTBF and the average relative cost reduction values found in the 27 instances with the cheap component. We also depict the minimum and maximum values of the optimal MTBF to show whether the lower bound $\underline{\tau}$ and the upper bound $\bar{\tau}$ are attained as optimal values. The minimum and maximum relative cost reductions are also reported to show to what extent the joint optimization can be advantageous.

We derive the following managerial insights by the results observed in our numerical experiment:

- Avg. τ^* is much higher for the cheap component than for the expensive component. For the cheap component, reliability improvement is favoured by the relatively low cost of reliability improvement and reductions in the repair costs and downtime costs achieved by reliability improvement.

- Avg. τ^* increases as the number of the systems increases. When the number of the systems increases, the frequency of system failures increases under the same τ, and, thus, repair costs and downtime costs increase. This constitutes an incentive to choose a higher reliability level.

- Avg. τ^* increases as the downtime penalty rate increases. The only effect of an increase in downtime penalty rate is an increase in downtime costs. A higher reliability level compensates this increase.

- Avg. τ^* increases as the exploitation phase gets longer. A longer exploitation phase implies that the OEM has to deal with a larger number of failures and suffer from higher repair costs and downtime costs. This provokes choosing a higher reliability level.

Note that, the increasing/decreasing effects of all factors on Avg. τ^* is significant. Avg. $\Delta\pi$ follows the same pattern as Avg. τ^*. Generally, the larger the distance between τ^* and $\underline{\tau}$, the larger the difference between π_n and π^*, and the larger $\Delta\pi$. Further, we should remark that the joint optimization leads to an average cost reduction of 44.3% in our experiment and it can even go up to 88.4%. These reductions correspond to large savings in absolute terms.

2.4.3 Sensitivity Analysis

We investigate the sensitivity of the optimal MTBF (τ^*) and the optimal LCC (π^*) to design costs ($K(\tau)$), unit production cost ($c(\tau)$), the number of system (N), downtime penalty rate (p), and the length of the exploitation phase (T). We see the 81 instances of the testbed of Subsection 2.4.1 as the true instances. Next, we generate modifications of each true instance by deviating $K(\tau)$, $c(\tau)$, N, p, and T by $\pm 20\%$ and $\pm 50\%$. We provide the deviations of the functions $K(\tau)$ and $c(\tau)$ by simply changing their coefficients. We identify each modification as a false case of the true instance where an estimation error is made for a function or a factor. Then we compare the optimal solutions of the true instances and that of their false cases.

Let $\pi_{true}(\tau, s)$ be the LCC function of a true instance (i.e., $\pi_{true}(\tau, s)$ has the correct parameter values for that instance). Let τ^*_{true}, s^*_{true}, and $\pi^*_{true} = \pi_{true}(\tau^*_{true}, s^*_{true})$ be the optimal MTBF, spare parts inventory level and optimal LCC of the instance, respectively. Let τ^*_{false} and s^*_{false} be the optimal MTBF and spare parts inventory level of a false modified instance. We use

$$\Delta\tau^* = \frac{\tau^*_{false} - \tau^*_{true}}{\tau^*_{true}}(100\%),$$

$$\Delta s^* = \frac{s^*_{false} - s^*_{true}}{s^*_{true}}(100\%),$$

and

$$\Delta\pi^* = \frac{\pi_{true}(\tau^*_{false}, s^*_{false}) - \pi^*_{true}}{\pi^*_{true}}(100\%)$$

Table 2.6 Results of the sensitivity analysis (1)

	$K(\tau)$			$c(\tau)$			N		
	$\bar{\Delta}\tau^*$	$\bar{\Delta}s^*$	$\bar{\Delta}\pi^*$	$\bar{\Delta}\tau^*$	$\bar{\Delta}s^*$	$\bar{\Delta}\pi^*$	$\bar{\Delta}\tau^*$	$\bar{\Delta}s^*$	$\bar{\Delta}\pi^*$
-50%	6.2%	-4.6%	0.5%	18.8%	-10.4%	2.5%	-6.8%	-39.1%	29.1%
-20%	2.3%	-1.7%	0.1%	6.3%	-4.0%	0.3%	-2.0%	-14.9%	3.7%
+20%	-1.8%	1.4%	0.0%	-5.6%	5.0%	0.3%	1.9%	14.2%	1.7%
+50%	-4.2%	3.6%	0.3%	-10.7%	8.2%	1.1%	3.7%	36.0%	6.3%

to measure the error in the optimal MTBF, the optimal spare parts inventory level, and the optimal LCC.

We calculate $\Delta\tau^*$, Δs^*, and $\Delta\pi^*$ for each of the false instances. We summarize the results of the analysis in Table 2.6 and Table 2.7, where $\bar{\Delta}\tau^*$, $\bar{\Delta}s^*$ and $\bar{\Delta}\pi^*$ denote the averages of the percent differences in the optimal MTBF and percent differences in LCC for the 81 instances, respectively. For example, the values -4.2%, 3.6%, and 0.3% in the last row of Table 2.6 are the averages of $\Delta\tau^*$, Δs^*, and $\Delta\pi^*$, respectively, when design costs are 50% higher. Our focus is on the $\Delta\tau^*$ and $\Delta\pi^*$ since we concern supporting the reliability decision in the design phase for optimal LCC. We use Δs^* to explain why $\Delta\tau^*$ and $\Delta\pi^*$ follow the same pattern with the deviations in the functions and factors.

In Table 2.6 and Table 2.7 large deviations in the LCC can only be observed with respect to the parameter N while for all other factors the model is robust against estimation errors. Especially when the expected number of systems to be sold is underestimated, the resulting costs will be high, because low values of N lead to less investment in reliability and spare parts inventory, which leads to low initial investment costs but higher costs in the exploitation phase. When such a situation is discovered in practice, replenishment of extra spare parts during the exploitation phase (possibly against a higher unit price than the initial price) can be an option to reduce the LCC.

The optimal MTBF is more sensitive to $c(\tau)$, p, and T compared to $K(\tau)$ and N. In

Table 2.7 Results of the sensitivity analysis (2)

	p			T		
	$\bar{\Delta}\tau^*$	$\bar{\Delta}s^*$	$\bar{\Delta}\pi^*$	$\bar{\Delta}\tau^*$	$\bar{\Delta}s^*$	$\bar{\Delta}\pi^*$
-50%	-12.8%	6.9%	2.2%	-18.5%	15.5%	3.5%
-20%	-4.6%	2.2%	0.2%	-6.1%	4.6%	0.3%
+20%	4.2%	-2.1%	0.2%	5.2%	-3.6%	0.2%
+50%	9.6%	-4.4%	0.8%	11.3%	-7.5%	0.9%

cases where $|\Delta\tau^*|$ is high, $\Delta\pi^*$ is significantly low. These cases also have high values of $|\Delta s^*|$; i.e., $|\Delta s^*|$ acts as a compensator when $|\Delta\tau^*|$ is high, too.

2.5. Conclusions

In this chapter, we introduced a reliability optimization model for critical components. We formulated the costs that are affected by the reliability level of the component and its spare parts inventory level throughout the life time of a number of systems (LCC). We showed certain analytical properties of the cost function and derived an optimization procedure based on these properties.

We conducted a numerical study based on real-life data to derive insights about how certain factors, such as component type (cheap, medium, expensive), the size of the installed base, the downtime penalty rate, and the lifetime of the system, affect the optimal reliability decision. We showed that our method leads to significant cost reductions compared to a non-integrated optimization method. The results of the experiment revealed that the optimal value of MTBF of a component depends on whether the component is cheap or expensive, the number of systems to be installed, downtime penalty rate and the length of exploitation phase. We also performed a sensitivity analysis for all the instances in our experiment and showed that the primary parameter to be concerned with is the size of the installed base and the optimal LCC have negligible or limited sensitivity to the other major parameters in our model.

As mentioned in Section 2.1, the findings in this chapter can be used as a foundation for a relevant multi-stage problem. In the next chapter, we will explain how to establish such a relation in detail for a redundancy allocation problem.

Appendix

Proof of Lemma 2.2: *(i)* Let $I(t)$ be the random variable denoting inventory on hand at time t, $f_{I(t)}(x)$ be its probability mass function and $h'(t)$ be the expected rate at which storage cost is incurred at time t. Then

$$h'(t) = h\sum_{x=0}^{\infty} x f_{I(t)}(x).$$

The expected NPV of storage cost throughout $[0,T]$, $S_2(\tau,s)$, is

$$S_2(\tau,s) = \int_0^T h'(t)e^{-\alpha t}dt = \int_0^T h\left(\sum_{x=0}^{\infty} x f_{I(t)}(x)\right)e^{-\alpha t}dt. \tag{2.7}$$

Because the stochastic process $\mathbf{I}= \{I(t) : t \geqslant 0\}$ is assumed to be in steady-state throughout $[0,T]$, (2.7) may be further rewritten as

$$\begin{aligned} S_2(\tau,s) &= \int_0^T h\left(\sum_{x=0}^{\infty} x f_I(x)\right)e^{-\alpha t}dt = \left(\int_0^T he^{-\alpha t}dt\right)\left(\sum_{x=0}^{\infty} x f_I(x)\right) \\ &= \frac{h}{\alpha}\left(1-e^{-\alpha T}\right)\left(\sum_{x=0}^{\infty} x f_I(x)\right), \end{aligned} \tag{2.8}$$

where by $f_I(x)$ denotes the steady-state distribution of $\mathbf{I}$. Let $\bar{I}(\tau,s)$ denote the expected steady-state inventory level for a given τ and s. Then $\bar{I}(\tau,s)$ is equal to the average number of idle servers in Erlang Loss System:

$$\bar{I}(\tau,s) = \sum_{x=0}^{\infty} x f_I(x) = s - \frac{NU}{\tau} + \frac{NU}{\tau}g(\tau,s) \tag{2.9}$$

where $g(\tau,s)$ is the probability of being out of stock (see (2.2) in Section 2.3). By substitution of this result into (2.8), we obtain

$$S_2(\tau,s) = \frac{h}{\alpha}\left(1-e^{-\alpha T}\right)\left[s - \frac{NU}{\tau} + \frac{NU}{\tau}g(\tau,s)\right]$$

(ii) Let M_1 and M_2 be the random variables representing the numbers of ordinary repairs and emergency repairs performed throughout $[0,T]$, respectively. By Lemma 2.1, M_1 and M_2 have Poisson distributions with means $[1-G(\tau,s)](N/\tau)T$ and $G(\tau,s)(N/\tau)T$, respectively.

Let W_1,..., W_{M_1} be the random variables representing the times of failures leading to instances of ordinary system repair throughout $[0,T]$ and V_1,..., V_{M_1} be the NPVs of the costs of the respective instances. W_i's are unordered times, that is, W_1 does not necessarily represent the time of the first failure, W_2 does not necessarily represent the time of the second failure, and so on. Since failures follow a Poisson process, W_i's are independent and uniformly distributed. That is, letting $f_{W_i}(t)$ denote the probability density function of W_i, $f_{W_i}(t) = 1/T$, $0 \leq t \leq T$. Then we can derive $E[V_i]$, the expectation of V_i, by conditioning on W_i.

$$E\left[V_i\right] = \int_0^T E\left[V_i|W_i = t\right] f_{W_i}(t)dt = \int_0^T r_1 e^{-\alpha t}\frac{1}{T}dt = \frac{r_1}{\alpha T}\left(1 - e^{-\alpha T}\right), \quad i = 1, ..., M_1.$$

Let P_1 be the NPV of the costs of instances of ordinary system repair that are performed throughout $[0, T]$.

$$E\left[P_1|M_1 = m\right] = E\left[\sum_{i=0}^{m} V_i|M_1 = m\right] = E\left[\sum_{i=0}^{m} V_i\right] = \sum_{i=0}^{m} E\left[V_i\right] = mE\left[V_1\right].$$

since $E[V_i]$'s are the same for all i. Let $f_{M_1}(m)$ denote the probability mass function of M_1. Then

$$\begin{aligned} E\left[P_1\right] &= \sum_{m=0}^{\infty} E\left[P_1|M_1 = m\right] f_{M_1}(m) = \sum_{m=0}^{\infty} mE\left[V_1\right] f_{M_1}(m) \\ &= E\left[M_1\right] E\left[V_1\right] = \left[1 - G\left(\tau, s\right)\right] \frac{N}{\tau}\frac{r_1}{\alpha}\left(1 - e^{-\alpha T}\right). \end{aligned}$$

We denote the NPV of the costs of instances of emergency system repair that are performed throughout $[0, T]$ by P_2. Using similar arguments, $E\left[P_2\right]$ can be derived as

$$E\left[P_2\right] = G\left(\tau, s\right) \frac{N}{\tau}\frac{r_2}{\alpha}\left(1 - e^{-\alpha T}\right).$$

Then

$$R(\tau, s) = \left[1 - G\left(\tau, s\right)\right] \frac{N}{\tau}\frac{r_1}{\alpha}\left(1 - e^{-\alpha T}\right) + G\left(\tau, s\right) \frac{N}{\tau}\frac{r_2}{\alpha}\left(1 - e^{-\alpha T}\right).$$

(iii) Follows from the same arguments used in the proof for $R(\tau, s)$ in part(ii).

Proof of Lemma 2.3: *(i)* $S(\tau, s) = S_1(\tau, s) + S_2(\tau, s)$ and it is trivial to show that $S(\tau, s) = 0$ when $s = 0$. For a fixed $s > 0$, $S_1(\tau, s) = c(\tau)s$ is increasing in τ as $c(\tau)$ is increasing in τ.

The Erlang loss probability $g(\tau, s)$ can be expressed as $B(s, a) = \frac{a^s}{s!} / \sum_{i=0}^{s} \frac{a^i}{i!}$, where $a = (N/\tau)L$; a represents the arriving workload in the corresponding Erlang loss system. Then the average inventory given in equation (2.9) can be rewritten as

$$\bar{I}\,(\tau, s) = \bar{I}\left(\frac{NU}{a}, s\right) = s - a[1 - B(s, a)].$$

$a[1 - B(s, a)]$ is known as the carried load in the Erlang loss system and is strictly increasing in a for a fixed value of $s > 0$ (see Appendix A). Thus, $\bar{I}\left(\frac{NU}{a}, s\right)$ is strictly decreasing in a, which implies that $\bar{I}(\tau, s)$ is strictly increasing in τ. As $\frac{h}{\alpha}\left(1 - e^{-\alpha T}\right) > 0$, this monotonicity property of $\bar{I}(\tau, s)$ implies that $S_2(\tau, s)$ is strictly increasing in τ for $s > 0$ (see equation (2.3)). Hence, $S(\tau, s)$ is strictly increasing in τ for $s > 0$.

(ii) $R(\tau, s)$ can be rewritten as

$$R(\tau, s) = \frac{r_1}{\alpha}\frac{N}{\tau}\left(1 - e^{-\alpha T}\right) + g(\tau, s)\frac{r_2 - r_1}{\alpha}\frac{N}{\tau}\left(1 - e^{-\alpha T}\right). \tag{2.10}$$

For its derivative with respect to τ, we find:

$$\frac{\partial R(\tau, s)}{\partial \tau} = -\frac{r_1}{\alpha}\frac{N}{\tau^2}\left(1 - e^{-\alpha T}\right) + N\frac{r_2 - r_1}{\alpha}\left(1 - e^{-\alpha T}\right)\left[\frac{\partial g(\tau, s)}{\partial \tau}\frac{1}{\tau} - \frac{1}{\tau^2}g(\tau, s)\right].$$

$\frac{\partial g(\tau,s)}{\partial \tau} < 0$ as $g(\tau, s)$ is strictly decreasing in τ (see Property 2.1). Thus, $\frac{\partial R(\tau,s)}{\partial \tau} < 0$. The proof for $D(\tau, s)$ follows from the same arguments as used for $R(\tau, s)$.

Proof of Lemma 2.4: *(i)* It is trivial that $S_1(\tau, s) = c(\tau)s$ is increasing in s (note that $S_1(\tau, s) = 0$ when $c(\tau) = 0$). Let $\Delta S_2(\tau, s) = S_2(\tau, s + 1) - S_2(\tau, s)$.

$$\Delta S_2(\tau, s) = \frac{h}{\alpha}\left(1 - e^{-\alpha T}\right)\left\{1 - [g(\tau, s) - g(\tau, s + 1)]\frac{NU}{\tau}\right\}. \tag{2.11}$$

The Erlang loss probability $G(\tau, s)$ is strictly decreasing and strictly convex in s (see Karush (1957); see also Remark 2 in Kranenburg and van Houtum (2007)). This implies that $\Delta g(\tau, s) = g(\tau, s) - g(\tau, s + 1)$ in equation (2.11) is strictly decreasing in s. The maximum value of $\Delta g(\tau, s)$ for a fixed τ is attained when $s = 0$, and

$$\Delta g(\tau, 0) = 1 - \frac{\frac{NU}{\tau}}{1 + \frac{NU}{\tau}} = \frac{\tau}{\tau + NU}. \tag{2.12}$$

Then

$$\Delta g(\tau, s)\frac{NU}{\tau} \leqslant \Delta g(\tau, 0)\frac{NU}{\tau} = \frac{NU}{\tau + NU} < 1, \tag{2.13}$$

which implies that

$$\Delta S_2(\tau, s) = \frac{h}{\alpha}\left(1 - e^{-\alpha T}\right)\left[1 - \Delta g(\tau, s)\frac{NU}{\tau}\right] > 0. \tag{2.14}$$

That is, $S_2(\tau, s)$ is strictly increasing in s. Hence, $S(\tau, s)$ is strictly increasing in s.

(ii) The first term of $R(\tau, s)$ in equation (2.10) is constant. As $g(\tau, s)$ is strictly decreasing in s, $(1 - e^{\alpha T}) > 0$, and $r_1 \leq r_2$, $R(\tau, s)$ is decreasing in s. The proof for $D(\tau, s)$ follows from the same arguments.

Proof of Lemma 2.5: *(i)* We rewrite equation (2.6) as

$$\begin{aligned} \pi(\tau, s) &= K(\tau) + \left[c(\tau) - c(\underline{\tau})\right] N + \frac{N}{\alpha}(r_1 + d_1)\left(1 - e^{-\alpha T}\right)\frac{1}{\tau} + c(\tau)s \\ &\quad + \frac{h}{\alpha}\left(1 - e^{-\alpha T}\right)\left[s - \frac{NU}{\tau}\right] \\ &\quad + \frac{N}{\alpha}\left(1 - e^{-\alpha T}\right)(hL + r_2 + d_2 - r_1 - d_1)\frac{1}{\tau} g(\tau, s). \end{aligned} \tag{2.15}$$

For a fixed value of τ:

- The first three terms of (2.15) are constant.
- The fourth and fifth term are linear in s.
- The last term in (2.15) is strictly convex in s because the Erlang loss probability $G(\tau, s)$ is strictly convex in s, $\left(1 - e^{-\alpha T}\right) > 0$, $r_1 \leq r_2$, and $d_1 \leq d_2$.

Hence $\pi(\tau, s)$ is strictly convex in s.

(ii) For a fixed value of τ:

- The first three terms of (2.15) are constant.
- $$\lim_{s \to \infty} \left\{ c(\tau)s + \frac{h}{\alpha}\left(1 - e^{-\alpha T}\right)\left[s - \frac{NU}{\tau}\right] \right\} = \infty.$$
- The last term depends on s via $g(\tau, s)$, but it is bounded from below by 0.

Hence $\lim_{s \to \infty} \pi(\tau, s) = \infty$.

(iii) We rewrite equation (2.6) for $\pi(\tau, s)$ as

$$\begin{aligned} \pi(\tau, s) &= -c(\underline{\tau})N + \frac{h}{\alpha}\left(1 - e^{-\alpha T}\right)s + K(\tau) + c(\tau)N + c(\tau)s \\ &\quad + \frac{N}{\alpha}\left(1 - e^{-\alpha T}\right)(r_1 + d_1 - hL)\frac{1}{\tau} \\ &\quad + \frac{N}{\alpha}\left(1 - e^{-\alpha T}\right)(hL + r_2 + d_2 - r_1 - d_1)\, g(\tau, s)\frac{1}{\tau} \end{aligned} \tag{2.16}$$

Recall that $hL \leq r_1 \leq r_2$, $h > 0$, $L > 0$, and $d_1 \leq d_2$. For a fixed value of s:

- The first two terms of (2.16) are constant.
- The following three terms are convex in τ.
- The sixth term is strictly convex in τ.
- The last term is also strictly convex in τ. Let

  $$f(\tau, s) = g(\tau, s)\frac{1}{\tau}.$$

  Then
  $$\frac{\partial^2 f(\tau,s)}{\partial \tau^2} = . \left[\frac{\partial^2 g(\tau,s)}{\partial \tau^2}\frac{1}{\tau} - 2\frac{1}{\tau^2}\frac{\partial g(\tau,s)}{\partial \tau} + \frac{2}{\tau^3}g(\tau,s)\right].$$

  As $g(\tau, s)$ is strictly decreasing and strictly convex in τ, $\frac{\partial g(\tau,s)}{\partial \tau} < 0$ and $\frac{\partial^2 g(\tau,s)}{\partial \tau^2} > 0$. So, $\frac{\partial^2 f(\tau,s)}{\partial \tau^2} > 0$, which implies the strict convexity of the last term in τ.

Hence, $\pi(\tau, s)$ is strictly convex in τ.

(iv) Let $\Delta\pi(\tau, s) = \pi(\tau, s + 1) - \pi(\tau, s)$. As we define $s^*(\tau) = \min\{\arg\min_{s\in\mathbb{N}_0} \pi(\tau, s)\}$, by parts (i) and (ii) imply that $s^*(\tau)$ is the smallest value of s satisfying $\Delta\pi(\tau, s) \geqslant 0$ and $s^*(\tau)$ is finite. The inequality $\Delta\pi(\tau, s) \geqslant 0$ may be shown to be equivalent to (use equation (2.15))

$$\frac{\Delta g(\tau,s)}{\tau} \quad \geqslant \quad \frac{h(1-e^{-\alpha T}) + \alpha c(\tau)}{N(1-e^{-\alpha T})(hL + r_2 + d_2 - r_1 - d_1)}, \tag{2.17}$$

where that $\Delta g(\tau, s) = g(\tau, s) - g(\tau, s + 1)$. Let $a = (N/\tau)L$ represents the arriving workload in the corresponding Erlang loss system, as in the proof of Lemma 2.3. Let $F_B(s + 1, a) = a\,[B(s, a) - B(s + 1, a)]$. $F_B(s + 1, a)$ is known as the load carried by the last server in the Erlang loss system (with $s + 1$ servers). Then, equation (2.17) may be rewritten as

$$\frac{hL(1-e^{-\alpha T}) + \alpha c(\frac{NU}{a})}{(1-e^{-\alpha T})(hL + r_2 + d_2 - r_1 - d_1)} - F_B(s+1, a) \quad \geq \quad 0. \tag{2.18}$$

$F_B(s + 1, a)$ is known to be increasing in a (see Appendix A). Hence, the lefthand side of inequality (2.18) is decreasing in a for each s. This implies that the first s for which inequality (2.18) is satisfied is increasing in a, and $s^*(\tau)$ (the first s for which inequality (2.17) is satisfied) is decreasing in τ.

Chapter 3

Redundancy Allocation for Serial Systems

3.1. Introduction

In Chapter 2, we worked on a single-stage reliability optimization problem. We focus on a redundancy allocation problem, which is a multi-stage problem by definition, in this chapter. As we stated in Chapter 2, the approach that we follow in this chapter will also help the reader to connect the single-stage problem in Chapter 2 to a corresponding multi-stage problem.

In this chapter, we investigate a situation in which a user buys a number of systems. Each system is composed of subsystems placed in a serial structure (i.e., a failure of any subsystem leads to a system failure). At most two identical, repairable parts can be used in a cold standby redundancy setting in each subsystem; that is, only one part is active in each subsystem and if there is a redundant part in a subsystem, it immediately becomes active when the active part in the subsystem fails.

We also extend the range of application of the emergency procedure in this chapter compared to the one defined in Chapter 2. The emergency procedure defined in Chapter 2 is applied only when there is an out-of-stock situation upon a failure while the one in this chapter can also be applied when stock-on-hand is one, as a preventive measure to avoid long downtime. The probability of stock-out events decreases significantly with the preventive application of the emergency procedure, meaning a failed part is almost always replaced with a ready-for-use one form the inventory. However, the activities performed in an application of the emergency procedure when stock-on-hand is one, slightly differ from the activities performed in an application of the emergency procedure in an out-of-stock situation. Thus, we only refer to the procedure applied when there is an out-of-stock situation as the emergency procedure and name the procedure applied when stock-on-hand is one the *provision procedure*, for the sake of clarity. We will give the details of these procedures in Subsection 3.2.1 and clarify their difference.

We consider the case in which the availability requirement is defined as a constraint on the total uptime throughout the lifetime of a number of systems. In the problem formulation, we translate this constraint to an equivalent downtime constraint.

The user can implement one of the following three policies per stage:

1. *Policy (0,0) - Do not choose redundancy and apply the emergency procedure when a failure occurs and there is an out-of-stock situation.*
2. *Policy (0,1) - Do not choose redundancy and apply the provision procedure when a failure occurs and stock-on-hand is 1.*
3. *Policy (1,0) - Choose redundancy and apply the emergency procedure when a failure occurs and there is an out-of-stock situation.*

In the name Policy (y, z), y represents whether redundancy is chosen or not and z represents whether the emergency procedure or the provision procedure is applied. These policies provide different total uptime against different TCO. If Policy (1,0) is chosen for a stage, 100% availability of the corresponding subsystems is attained. As a result, choosing redundancy and applying the provision procedure for a stage does not provide any further advantage in terms of availability and we exclude Policy (1,1) in our study.

Cold standby redundancy can be considered as a special strategy for keeping spare parts inventory: A number of spare parts dedicated to a certain subsystem of a system are kept in the system rather than at another location. Furthermore, when a single system is considered, a redundancy allocation model is equivalent to a spare parts inventory model under certain assumptions (e.g. negligible switching time for redundancy and negligible replacement time for spare parts, costs stemming from a standby part is equal to that stemming from a spare part); see Black and Proschan (1959). Due to this equivalence, Mizukami (1968), Bryant and Murphy (1983), and Wells and Bryant (1985) use the terms standby parts and spares interchangeably. Spare parts and redundancy are not equivalent in our case, as is common in a number of papers in the literature.

Sharma and Misra (1988) and Misra and Sharma (1991) present models similar to ours in which the equivalence of redundancy and spare parts does not hold. However, as explained in Section 1.6, these papers have the following limitations: They do not include maintenance costs and their models do not incorporate any emergency procedures, meaning considerably long downtime is allowed. In addition, these models are developed for a single system, so the pooling effect of multiple systems on spare parts does not exist.

The contributions of this chapter are:

- First, we develop a model for redundancy allocation of a capital good in the design phase. This model includes three different policies per stage as explained above. We develop a problem formulation in which the three policies per stage are represented. The formulation can be summarized as the minimization of the TCO of a general number of systems under a defined downtime constraint. TCO includes acquisition costs, spare parts costs, and repair costs. We explicitly relate the redundancy level and spare parts inventory level of each subsystem to these costs and the downtime of the systems.

- Second, we decompose the multi-stage problem into single-stage problems by using the Lagrangian relaxation method. The multi-stage problem has a combinatorial nature as any of the three policies can be applied per stage. Our decomposition enables the generation of optimal solutions for the multi-stage problem efficiently, without considering all possible combinations: The single-stage problems have the same formulations with different parameter values. We show that a solution for the multi-stage problem can be generated by finding solutions of each of the single-stage problems. We develop an efficient optimization procedure which can be used to solve any of the single-stage problems for varying resource levels of the downtime constraint. Thus, our procedure can be used to find solutions for the multi-stage problem for varying resource levels of the downtime constraint efficiently.

- Third, we compare the three policies at the stage level and provide conditions under which one policy outperforms the other. When the downtime constraint is loose, Policy (0,0) is optimal. As the constraint becomes tight (i.e., the resource level of the downtime constraint is decreased), one of the following two cases occurs depending on the values of the parameters such as the MTBF of a part, the number of the systems, the lifetime of the systems, the unit repair costs and the mean downtime stemming from the ordinary and the emergency procedures, the unit storage rate of a part, and the repair lead time of a part:

  - Policy (0,1) becomes optimal after a certain level, say D_1, of the constraint. As the constraint is further tightened, it remains optimal until another level, say $D_2 < D_1$, at which Policy (1,0) becomes optimal. Policy (1,0)

  remains optimal for all smaller resource levels afterwards; that is, it remains optimal for all recourse levels that lie within $[0, D_2]$.
  - Policy (1,0) becomes optimal after a certain level, say D_3, of the constraint and remains optimal for all smaller resource levels afterwards; that is, it remains optimal for all recourse levels that lie within $[0, D_3]$.

  We show that the values of the TCO and downtime (or uptime) when the optimal policy changes from one to the other can be easily computed. Notice that the conditions under which Policy (1,0) is optimal correspond to the situations in which having redundancy is optimal.

- Fourth, we provide the following multi-stage results:
  - We introduce a method to construct an efficient frontier which reflects the trade-off between the uptime and the TCO. The optimal value of the TCO and uptime, when the policy decision for a stage is changed from one to another, is found by this method.
  - We provide a method for ordering of the stages in a capital good reflecting the benefits of investing in redundancy. In many cases, one is interested in finding the optimal order to implement redundancy for stages, while making one-by-one decisions during the design due to some other considerations (e.g. there might be a budget limit and one might follow such an order until the budget limit is reached).

As we compared the model in Chapter 2 to a number of models existing in the literature, we also provide a comparison of the model in this chapter to the models existing in the literature in Table 3.1. The discriminative attributes that are incorporated in our model are identified as maintenance costs, multi-component (system-level), multiple systems, spare parts and the emergency procedure. We include the papers which have at least one of the given attributes.

This chapter is organized as follows. In Section 3.2, we present our model assumptions. Next, we formulate our problem, derive the cost functions and decompose the multi-stage problem into single-stage problems by using the Lagrangian relaxation method in Section 3.3. We finalize the chapter by comparing the three policies on stage level and providing the multi-stage results in Section 3.4.

3.2. Model

A capital good is being designed by an Original Equipment Manufacturer (OEM) for a user. The user will buy N ($N \in \mathbb{N} = \{1, 2, ...\}$) systems. We assume that the N

systems will start operating at the same time and it is estimated that they will be in use for a time length of T years, which is in the order of 10-30 years. We denote the exploitation phase of the systems by $[0, T]$. The user requires the uptime of at least $p \in (0, 1]$ of the total possible operational time (NT system-years). We will use p as an availability measure in Subsection 3.4.3.

The capital good includes m ($m \in \mathbb{N} = 1, 2, ...$) stages placed in a serial structure. Each component is included in only one stage, so there is a 1-1 relationship between stages and components. We index stages and components with the same indices; that is, stage i includes component i, $i \in M = \{1, 2, ..., m\}$. We refer to a subsystem of stage i as a *stage-i subsystem* and a part of component i as a *component-i part*. Each stage i, $i \in M$, includes at least one unit of component i; the user can choose to have a second unit of component i in a cold standby redundancy setting.

The user may keep spare parts inventory for each stage at a single stock point. She buys s_i component-i spare parts together with the systems at time 0. We refer to the purchase of the spare parts at time 0 as the *initial supply.* There will be a single repair facility for defective parts. The user also agrees with the OEM for replenishment of ready-for-use parts as soon as possible via a fast transportation mode (e.g. by plane) in case of need. We assume that the OEM has ample supply of the parts.

Table 3.1 Comparison of papers on redundancy allocation

Attribute	Sharma and Misra (1988)	Misra and Sharma (1991)	Hussain and Murthy (1998)	Monga and Zuo (1998)	This chapter
Maintenance costs			X	X	X
Multi-component	X	X		X	X
Multiple systems					X
Spare parts					X
Emergency Procedure					X

3.2.1 Failure and Repair Processes

We denote the MTBF of a component-i part by τ_i. The τ_i, $i \in M$, are typically in the order of 1-10 years and known by the user. We assume that the failures of a part in a system are independent of the failures of the other parts in the system and the total stream of failures of component-i parts follows a Poisson process with the constant rate $N\tau_i^{-1}$ throughout $[0, T]$. Remember that the memoryless property of the Poisson failures implies that there is no aging (degradation) effect. As stated in Subsection 2.2.1, this assumption is justified when the number of the systems (N) is sufficiently large or if lifetimes of parts are close to exponential. Also, remember that the short downtimes justify the assumption of constant failure rates.

Upon the failure of a component-i part at time $t \in [0, T]$, one of the ordinary procedure, the provision procedure or the emergency procedure defined below will be applied. These definitions are independent of choosing redundancy or not for a stage. That is, each procedure will be applied in the defined manner for the stage regardless of the redundancy decision of the stage. Which procedure is applied will depend on a predetermined threshold value $z_i \geq 0$ for the actual stock-on-hand, which we denote by $H_i(t)$, and $H_i(t)$ itself:

1. *Ordinary Procedure*: If $H_i(t) > z_i$, the failed part is replaced with a ready-for-use part from the inventory. The defective part is transported to the repair facility for a repair. After the repair, the part is restored to an as-good-as-new condition and added to the spare parts inventory.
2. *Provision Procedure*: If the $0 < H_i(t) \leq z_i$, the failed part is replaced with a ready-for-use part from the inventory. An as-good-as-new component-i part is replenished from the OEM and added to the inventory. The defective part is returned to the OEM.
3. *Emergency Procedure*: If the $H_i(t) = 0$ (out-of-stock situation), an as-good-as-new component-i part is replenished from the OEM and directly transported to the location of the failure. The failed part is replaced with the replenished part and returned to the OEM.

Notice that if $z_i = 0$, the provision procedure is never applied and we have the same ordinary procedure and emergency procedure defined in Chapter 2.

We assume that when a part fails during the exploitation phase, it will be diagnosed with 100% accuracy in a negligibly short time. If the failure occurs in a subsystem with a single part, the subsystem becomes operational just after the replacement of the failed part. The downtime after a failure of a part is equal to the replacement time of the part. We call a replacement in an application of the ordinary procedure or the provision procedure an *inventory-replacement* and a replacement in an application of the emergency procedure an *OEM-replacement* due to the replacement modes. Inventory-replacement times are independently and identically distributed with mean $\mu_{1,i} > 0$, for all $i \in M$. Similarly, OEM-replacement times are independently and identically distributed with mean $\mu_{2,i}$. $\mu_{1,i}$ and $\mu_{2,i}$ are in the order of 1-48 hours and $\mu_{1,i} \leq \mu_{2,i}$ as the stock point of the spare parts is at a close distance to the systems (the stock point and the systems may even be at the same site). The emergency procedure assures that downtimes are short even in out-of-stock situations as one does not have to wait until a ready-for-use part becomes available from the repair facility. The provision procedure serves as a preventive action to avoid longer downtimes that would rise in out-of-stock situations, in which OEM-replacements take place.

The repair facility has planned lead times for all repairs. Again, as stated in Subsection 2.2.1, this is a standard assumption in the spare parts inventory literature. We assume that repair times of component-i parts which include time to transport the part to and from the repair facility are independent and identically distributed with mean $U_i > 0$ which is typically in the order of 1-4 months. The orders of magnitude imply that $\mu_{2,i}$ is very small compared to U_i, which reflects the users incentive to apply the emergency procedure.

For $z_i > 0$, if a component-i part fails at time t and $0 < H_i(t) \leq z_i$, the part which is replenished by the provision procedure is added to the inventory after a random lead time with mean $\mu_{3,i}$. $\mu_{3,i}$ is also in the order of 1-48 hours (the distance between the OEM's site and the user's stock point is comparable to the distance between the OEM's site and the location(s) of the systems). Notice that this lead time has no relation to the replacement times.

Upon a failure, the spare parts inventory of component i is affected as follows. If the ordinary procedure or the provision procedure is applied, a part is removed form the inventory and a part is added to the inventory after some lead time. If the emergency procedure is applied, demand is lost for the inventory; i.e., no parts are removed or added to the inventory. As a result, the assumed procedures imply that the inventory position (the sum of pipeline stock and actual stock-on-hand) of component-i is kept at a constant level which is equal to the initial supply amount s_i. Similar to the situation in Subsection 2.2.1, for component i, we may also say that the spare parts inventory is controlled by a continuous-review *basestock policy* with basetock level s_i.

3.2.2 Policies

For each stage i, $i \in M$, we denote the redundancy decision by y_i: $y_i = 0$ means that no redundancy is chosen and $y_i = 1$ means that redundancy is chosen. The user can implement one of the three policies which are defined couples (y_i, z_i) as follows:

1. *Policy (0,0) - Do not choose redundancy ($y_i = 0$) and apply only the ordinary procedure and the emergency procedure.*
2. *Policy (0,1) - Do not choose redundancy ($y_i = 0$) and apply the ordinary procedure and the provision procedure with $z_i = 1$.*
3. *Policy (1,0) - Choose redundancy ($y_i = 1$) and apply only the ordinary procedure and the emergency procedure.*

Notice that Policy (0,1) requires $s_i \geq 1$. Also notice that the number of policies on the system level is m^3. Now, we compare downtimes under the policies.

Under Policy (0,0), downtimes arise during both inventory-replacements and OEM-replacements.

We assume that $\mu_{3,i}$ is very small compared to τ_i/N, which is the mean time between events in the process of the total stream of failures of component-i parts. Consequently, under Policy (0,1), if a component-i part fails at time t and $H_i(t) = 1$ (the provision procedure is applied), the probability that a component-i part in another system would fail during the replenishment lead time is negligibly small. Based on this, we assume that no failure of component-i parts occurs during the replenishment lead time, which means that there is at least one component-i part available from the inventory whenever a component-i part fails. As a result, the emergency procedure is never applied under Policy (0,1) and downtimes arise only during inventory-replacements.

Similarly, under Policy (1,0), at most one of the two component-i parts in each stage-i subsystem will be active at any instant during the exploitation phase. We assume that the switchover will be perfect and instantaneous in all stage-i subsystems; that is, when an active part fails, the standby part will take over the functionality in a negligibly small time without any failure (this is a standard assumption in the redundancy allocation literature). As $\mu_{1,i} \leq \mu_{2,i}$ and $\mu_{2,i}$ is very small compared to τ_i (see their orders of magnitude), the failed part will be replaced by the ordinary procedure or the emergency procedure in a negligibly short time compared to the MTBF of the standby part. Thus, the failure probability of the subsystem will be negligibly small. Hence, we assume that the redundancy is sufficient for 100%

availability of the subsystem (no downtime), and limit the redundancy setting to two parts.

Under Policy (0,0) and Policy (0,1), when a subsystem fails and leads to a system failure, the probability that another subsystem with a single part in the same system would fail before the failed part is replaced by the ordinary procedure or the emergency procedure is negligibly small as $\mu_{1,i} \leq \mu_{2,i}$ and $\mu_{2,i}$ is very small compared to τ_j for all $i, j \in M$. Thus, we assume that downtimes of different subsystems in a system do not overlap. That is, when a system is down, only one of its subsystems with a single part is down. This assumption leads to a simple formula for availability (or downtime).

3.2.3 Cost Factors

The objective is the minimization of the portions of TCO which are affected by the policy chosen for each stage and the initial supply amount of its spare parts inventory. These portions are the acquisition costs, the spare pasts costs, and the repair costs stemming from the applications of the three procedures. We assume that the acquisition costs of the N systems and spare parts are incurred at time 0. The other costs are incurred throughout $[0, T]$ and their Net Present Values (NPVs) at time 0 are taken into account. We denote the discount rate by $\alpha > 0$. We use the following notation to refer to the cost parameters in our model:

$c_{0,i}$: Unit acquisition cost of a component-i part during the initial supply.

$c_{1,i}$: Extra cost incurred for a stage-i subsystem with two component-i parts in a redundancy setting.

h_i: The storage cost rate per spare part of component i ($h_i > 0$ for all $i \in M$).

$r_{1,i}$: Expected costs incurred per application of the ordinary procedure for a component-i part ($r_{1,i} > 0$ for all $i \in M$).

$r_{2,i}$: Expected costs incurred per application of the emergency procedure for a component-i part ($r_{2,i} > 0$ for all $i \in M$).
Also expected costs incurred per application of the provision procedure for a component-i part; see the explanation below.

In the emergency procedure, a ready-for-use part is transported from the OEM's site to the location of the failure and the failed part is transported from the location of the failure to the OEM's site. In the provision procedure, a ready-for-use part is transported from the stock point to the location of the failure, another ready-for-use part is transported form the OEM's site to the stock point and the failed part is transported from the location of the failure to the OEM's site. Despite this difference, we assume that the costs incurred in the two cases are equal as the stock point is at a close distance to the systems.

The factor $r_{1,i}$ includes all costs originating from an instance of the ordinary procedure applied for a component-i part, which are administrative costs, costs of a visit of a service engineer, transportation costs, and repair costs of a failed component-i part. Similarly, the factor $r_{2,i}$ includes all costs originating from an instance of the emergency procedure or the provision procedure applied for a component-i part, which are administrative costs, costs of a visit of a service engineer, relevant transportation costs and replenishment costs of a component-i part from the OEM.

We assume that $r_{2,i} \geq r_{1,i}$; that is, an application of the emergency procedure costs at least as much as an application of the ordinary procedure, which is an incentive for the user to keep spare parts and apply the ordinary procedure. Generally, $r_{2,i}$ will be much larger than $r_{1,i}$.

The effect of redundancy in a subsystem on its acquisition cost appears as an extra cost compared to a subsystem without redundancy. The variable part of the acquisition costs of the subsystems originate from this extra cost. So, in our model, we only take into account this extra cost and treat the acquisition cost without redundancy as a fixed cost. Notice that, $c_{0,i}$ is defined as the unit acquisition cost of a component-i part but not as the unit acquisition cost of stage-i subsystem without redundancy. It serves in the formulation of the acquisition costs of spare parts.

The spare parts costs includes the spare parts investments costs (i.e. acquisition) and the spare parts storage costs. The resources for storing s_i units of component-i part are allocated permanently throughout $[0, T]$. That is, at any instant, the costs incurred for storing spare parts depend on the initial supply amount (s_i), not on actual stock-on-hand.

The repair costs stemming from the ordinary procedure and the emergency procedure depend on the stock-on-hand processes of the spare parts. We assume that the spare parts stock-on-hand process of each component is in the steady state from time 0.

3.3. Problem Formulation

In this section, we first give the multi-stage problem formulation and several preliminary results. Next, we derive the cost and downtime functions which constitute the problem formulation. We finalize the section by decomposing the multi-stage problem into single-stage problems by the Lagrangian relaxation.

Our problem formulation is as follows:

$$
\begin{array}{lll}
\text{(Q0)} & \min & \pi(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{s}) \\
& \text{s.t.} & D(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{s}) \leq D_0 \\
& & (y_i, z_i) \in \{(0,0), (0,1), (1,0)\} \text{ for all } i \in M \\
& & s_i \in \mathbb{N}_0 = \{z_i, z_i + 1, z_i + 2, ...\} \text{ for all } i \in M,
\end{array}
$$

where $\boldsymbol{y}$, $\boldsymbol{z}$, and $\boldsymbol{s}$ are the vectors of y_i, z_i, and s_i, respectively, $i \in M$. That is, $\boldsymbol{y} = (y_1, ..., y_m)$, $\boldsymbol{z} = (z_1, ..., z_m)$, and $\boldsymbol{s} = (s_1, ..., s_m)$. $\pi(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{s})$ is the expected NPV of TCO of the N systems, $D(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{s})$ is the expected downtime of all N systems together throughout $[0, T]$ (in system-years), and $D_0 = (1 - p)NT$ is the maximum downtime that can be tolerated as uptime of at least $p \in (0, 1]$ of the NT system-years is required. We give the notation for the cost functions and the downtime per stage as below:

$P_i(y_i)$:	The expected NPV of extra acquisition costs of the N subsystems of stage i due to redundancy (remember that we take into account only the extra acquisition costs incurred when redundancy is chosen)
$S_{1,i}(s_i)$:	The expected NPV of spare parts investment costs of component i
$S_{2,i}(s_i)$:	The expected NPV of spare parts storage costs of component i
$S_i(s_i)$:	The expected NPV of spare parts costs of component i incurred throughout $[0, T]$. $S_i(s_i) = S_{1,i}(s_i) + S_{2,i}(s_i)$
$R_i(z_i, s_i)$:	The expected NPV of repair costs incurred for the stage-i subsystems throughout $[0, T]$.
$D_i(y_i, z_i, s_i)$:	The expected downtime stemming from failures of the N stage-i subsystems throughout $[0, T]$.
$\pi_i(y_i, z_i, s_i)$:	The expected NPV of total costs of stage i. $\pi_i(y_i, s_i) = P_i(y_i) + S_i(s_i) + R_i(z_i, s_i)$.

Obviously, $\pi(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{s}) = \sum_{i=1}^{m} \pi_i(y_i, z_i, s_i)$. As downtimes of different subsystems in a system do not overlap, total downtime of the N systems is $D(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{s}) = \sum_{i=1}^{m} D_i(y_i, z_i, s_i)$ (if they could overlap, $D(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{s})$ would be an upper bound for the downtime).

Different combinations of the the binary variables y_i and z_i, $i \in M$, correspond to different policies per stage. We limit the choices of the combinations to $(y_i, z_i) = \{(0,0), (0,1), (1,0)\}$ as $(y_i, z_i) = (1,1)$ which corresponds to choosing redundancy and applying the emergency procedure with $z_i = 1$ does not exist as a policy in our model.

3.3.1 Derivation of the Cost and Downtime Functions

The acquisition costs of stage-i subsystems only depend on the redundancy decision of stage i. As we assume that total stream of failures of stage-i subsystems occur according to a Poisson process with rate $N\tau_i^{-1}$ for all $i \in M$, the failures of subsystems of different stages are independent of each other. Independent failures also lead to independence of the spare parts usage (spare parts inventory) of different stages. Hence, acquisition costs $P_i(y_i)$, spare parts costs $S_i(s_i)$, and repair costs $R_i(z_i, s_i)$ can be formulated independently for each $i \in M$. Then, the expected NPV of total costs of stage i and the expected NPV of total costs of the N systems, which is our objective function, can be written as $\pi_i(y_i, z_i, s_i) = P_i(y_i) + S_i(s_i) + R_i(z_i, s_i)$ and $\pi(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{s}) = \sum_{i=1}^{m} \pi_i(y_i, z_i, s_i)$, respectively.

The objective function and the constraint of problem (Q0) are linear combinations of objectives and constraints per stage; hence (Q0) is separable. That is, if we consider the downtime as a resource, it can be committed into stages independently and the overall TCO is simply the sum of the total costs incurred for each independent stage.We make use of this separability of the multi-stage problem to decompose it into single-stage problems by the Lagrangian relaxation method in the next subsection. We continue this section with the derivation of the single stage cost functions and downtime.

As we assume that the acquisition costs of the N systems and spare parts are incurred at time 0,

$$P_i(y_i) = Nc_{1,i}y_i,$$

and

$$S_{1,i}(s_i) = c_{0,i}s_i.$$

The expected NPV of spare parts storage costs throughout $[0, T]$ is

$$S_{2,i}(s_i) = \int_0^T h_i s_i e^{-\alpha t} dt = \frac{h_i}{\alpha}(1 - e^{-\alpha T})s_i. \tag{3.1}$$

Hence,

$$S_i(s_i) = \left[c_{0,i} + \frac{h_i}{\alpha}(1 - e^{-\alpha T})\right] s_i.$$

To derive the repair costs and downtime of stage-i subsystems, we need to identify in which cases the emergency procedure and the provision procedure are applied mathematically (the ordinary procedure is applied in other cases). These cases depend on the stock-on-hand process of the spare parts inventory of component i. Demands arrive at the spare parts inventory of component i according to a Poisson process with rate N/τ_i. Upon the arrival of a demand at time $t \in [0, T]$, if $H_i(t) > 0$, the demand is satisfied (a part is taken from the inventory) and a part is added to the inventory after a generally distributed lead time with mean U_i. Remember that the spare parts inventory position (sum of stock on hand and pipeline) of component i is kept at a constant level which is set to s_i at time 0.

- Under Policy (0,0) and Policy (1,0) (for $z_i = 0$), the emergency procedure is applied if $H_i(t) = 0$ and the demand is lost for the inventory. Hence, the stock-on-hand process of the spare parts inventory of component i, $\{H_i(u) : u \in [0, T]\}$, is identical to the process for the number of free servers in an Erlang loss system (also denoted as the $M/G/s_i/s_i$ queueing system) with an arrival rate N/τ_i, mean service time U_i, and s_i servers.
- Under Policy (0,1) (for $z_i = 1$), the provision procedure is applied if $H_i(t) = 1$. As we assume that $\mu_{3,i}$ is very small compared to τ_i/N and no failures of component-i parts occur during the replenishment lead time, $H_i(u) \geq 1$ for all $u \in [0, T]$. Hence, $H_i(u) = 1 + \bar{H}_i(u)$ where $\{\bar{H}_i(u) : u \in [0, T]\}$ is a stochastic process identical to the process for the number of free servers in an Erlang loss system with an arrival rate N/τ_i, mean service time U_i, and $s_i - 1$ servers.

Under Policy (0,0) and Policy (1,0), the emergency procedure is applied if a failure occurs when the stochastic process $\{H_i(u) : u \in [0, T]\}$ is in state zero. Under Policy (0,1), the providence procedure is applied if a failure occurs when the stochastic process $\{\bar{H}_i(u) : u \in [0, T]\}$ is in state zero. As we assume that $\{H_i(u) : u \in [0, T]\}$ (the stock-on-hand process) is in equilibrium from the beginning, so is $\{\bar{H}_i(u) : u \in [0, T]\}$. Hence, the probabilities of $H_i(u) = 0$ and $\bar{H}_i(u) = 0$ are equal to the Erlang loss probabilities $B_i(s_i)$ and $B_i(s_i - 1)$, respectively, where

$$B_i(x) = \frac{\frac{\left(\frac{NU_i}{\tau_i}\right)^x}{x!}}{\sum_{q=0}^{x} \frac{\left(\frac{NU_i}{\tau_i}\right)^q}{q!}}.$$

(We do not use the notation that we used in Chapter 2 for the Erlang loss probability as its arguments are different in this chapter.)

By using this result, first, we derive the distribution of the numbers of applications of the ordinary procedure and applications of the emergency procedure for stage-i subsystems throughout $[0, T]$. Next, we derive the expected NPV of repair costs incurred for stage-i subsystems ($R_i(s_i)$) and the expected downtime stemming from failures of the stage-i subsystems throughout $[0, T]$ ($D_i(s_i)$). We introduce these derivations in Property 3.1.

Property 3.1 For all $i \in M$, the followings hold.

(i) Under Policy (0,0) and Policy (1,0) (for $z_i = 0$), the numbers of applications of the ordinary procedure and applications of the emergency procedure due to failures of component-i parts throughout $[0, T]$ have Poisson distributions with means $(N/\tau_i)T[1 - B_i(s_i)]$ and $(N/\tau_i)TB_i(s_i)$, respectively.

(ii) Under Policy (0,1) (for $z_i = 1$), the numbers of applications of the ordinary procedure and applications of the provision procedure due to failures of component-i parts throughout $[0, T]$ have Poisson distributions with means $(N/\tau_i)T[1 - B_i(s_i - 1)]$ and $(N/\tau_i)TB_i(s_i - 1)$, respectively.

(iii) The expected NPV of repair costs incurred for the stage-i subsystems throughout $[0, T]$ is given by

$$R_i(z_i, s_i) = \frac{N}{\alpha\tau_i}(1 - e^{-\alpha T})\Big\{r_{1,i} + (r_{2,i} - r_{1,i})\Big[(1 - z_i)B_i(s_i) + z_iB_i(s_i - 1)\Big]\Big\}. \tag{3.2}$$

(iv) The expected downtime stemming from failures of the stage-i subsystems throughout $[0, T]$ is given by

$$D_i(y_i, z_i, s_i) = \frac{NT}{\tau_i}(1 - y_i)\Big[\mu_{1,i} + (\mu_{2,i} - \mu_{1,i})(1 - z_i)B_i(s_i)\Big]$$

Proof: See Appendix at the end of this chapter. □

We reconstruct the storage costs function $S_{2,i}(s_i)$ in equation (3.1) and repair costs function $R_i(z_i, s_i)$ in equation (3.2) to simplify their representation and interpretation. Let

$$\hat{h}_i = \frac{h_i}{\alpha T}(1 - e^{-\alpha T}),\ \hat{r}_{1,i} = \frac{r_{1,i}}{\alpha T}(1 - e^{-\alpha T}),\ \text{and}\ \hat{r}_{2,i} = \frac{r_{2,i}}{\alpha T}(1 - e^{-\alpha T}).$$

Then,

$$S_{2,i}(s_i) = \hat{h}_i T s_i \tag{3.3}$$

and

$$R_i(z_i, s_i) = \frac{NT}{\tau_i}\{\hat{r}_{1,i} + (\hat{r}_{2,i} - \hat{r}_{1,i})[(1 - z_i)B_i(s_i) + z_iB_i(s_i - 1)]\}. \tag{3.4}$$

So, $\hat{h}_i$, $\hat{r}_{1,i}$ and $\hat{r}_{2,i}$ can be interpreted as the parameters which already include the discounting effect on h, $r_{1,i}$, and $r_{2,i}$ throughout $[0, T]$, respectively, as equations (3.3) and (3.4) are formulations without discounting when storage cost rate is $\hat{h}_i$, the expected cost of an application of the ordinary procedure is $\hat{r}_{1,i}$, and the expected cost of an application of the emergency procedure is $\hat{r}_{2,i}$. We will use these interpretations in the remainder of this chapter.

3.3.2 Decomposition into single-stage Problems

The Lagrangian function for Problem (Q0) is defined as

$$L(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{s}, \lambda) = \sum_{i=1}^{m} \pi_i(y_i, z_i, s_i) + \lambda \left(\sum_{i=1}^{m} D_i(y_i, z_i, s_i) - D_0 \right)$$

where $\lambda \geq 0$ is a Lagrange multiplier. As (Q0) is separable, the Lagrangian is also separable; that is, we can rewrite the Lagrangian as

$$L(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{s}, \lambda) = \sum_{i=1}^{m} L_i(y_i, z_i, s_i, \lambda) - \lambda D_0,$$

where

$$L_i(y_i, z_i, s_i, \lambda) = \pi_i(y_i, z_i, s_i) + \lambda D_i(y_i, z_i, s_i)$$

is the decentralized Lagrangian for stage i. Observe that the decentralized Lagrangian functions are connected to each other through a single Lagrange multiplier (λ) as there is only one constraint in our problem. For a given value of λ, if a solution $(y_i^*(\lambda), z_i^*(\lambda), s_i^*(\lambda))$ that minimizes the decentralized Lagrangian $L_i(y_i, z_i, s_i, \lambda)$ (i.e., $(y_i^*(\lambda), z_i^*(\lambda), s_i^*(\lambda)) = \arg\min_{(y_i,z_i,s_i)} \{L_i(y_i, z_i, s_i, \lambda)\}$) can be found for each stage i, the vectors $(\boldsymbol{y}^*(\lambda), \boldsymbol{z}^*(\lambda), \boldsymbol{s}^*(\lambda))$, $\boldsymbol{y}^*(\lambda) = (y_1^*(\lambda), ..., y_m^*(\lambda))$, $\boldsymbol{z}^*(\lambda) = (z_1^*(\lambda), ..., z_m^*(\lambda))$, $\boldsymbol{s}^*(\lambda) = (s_1^*(\lambda), ..., s_m^*(\lambda))$, will also minimize the Lagrangian $L(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{s}, \lambda)$ for that value of λ.

By the so-called *Everett result* (see Theorem 1 in Everett (1963)), for all $i \in M$, if a solution $(y_i^*(\lambda), z_i^*(\lambda), s_i^*(\lambda))$ that minimizes the decentralized Lagrangian $L_i(y_i, z_i, s_i, \lambda)$ can be found for a given $\lambda \geq 0$, then $(y_i^*(\lambda), z_i^*(\lambda), s_i^*(\lambda))$ is also an optimal solution to the Problem ($\mathrm{Q}_i(\lambda)$) given as

$$
\begin{array}{lll}
(\mathrm{Q}_i(\lambda)) \qquad & \min & \pi_i(y_i, z_i, s_i) \\
 & \text{s.t.} & D_i(y_i, z_i, s_i) \leq D_i(y_i^*(\lambda), z_i^*(\lambda), s_i^*(\lambda)) \\
 & & s_i \in \mathbb{N}_0 = \{0, 1, 2, ...\} \\
 & & (y_i, z_i) \in \{(0,0), (0,1), (1,0)\}.
\end{array}
$$

$(y_i^*(\lambda), z_i^*(\lambda), s_i^*(\lambda))$ will satisfy the downtime constraint in Problem ($\mathrm{Q}_i(\lambda)$) with equality. Furthermore, the vectors $(\boldsymbol{y}^*(\lambda), \boldsymbol{z}^*(\lambda), \boldsymbol{s}^*(\lambda))$ will also be a solution for the problem ($\mathrm{Q}(\lambda)$) formulated as

$$\begin{array}{lll} (\mathrm{Q}(\lambda)) & \min & \pi(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{s}) \\ & \text{s.t.} & D(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{s}) \leq D(\boldsymbol{y}^*(\lambda), \boldsymbol{z}^*(\lambda), \boldsymbol{s}^*(\lambda)) \\ & & s_i \in \mathbb{N}_0 = \{0, 1, 2, ...\} \text{ for all } i \in M \\ & & (y_i, z_i) \in \{(0,0), (0,1), (1,0)\} \text{ for all } i \in M, \end{array}$$

and they will satisfy the downtime constraint with equality. Thus, by using various values of λ, we can generate optimal solutions of Problem (Q0) for specific values of D_0 (equivalently, specific values of the availability measure p). As a direct result of Theorem 1 in Fox (1966), such solutions are also so-called *efficient solutions* for the problem (Q1) given as

$$\begin{array}{lll} (\mathrm{Q1}) & \min & \pi(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{s}) \\ & \min & D(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{s}) \\ & & s_i \in \mathbb{N}_0 = \{0, 1, 2, ...\} \text{ for all } i \in M \\ & & (y_i, z_i) \in \{(0,0), (0,1), (1,0)\} \text{ for all } i \in M. \end{array}$$

A solution $(\boldsymbol{y}_e, \boldsymbol{z}_e, \boldsymbol{s}_e)$ is efficient for Problem (Q1) if and only if $\pi(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{s}) > \pi(\boldsymbol{y}_e, \boldsymbol{z}_e, \boldsymbol{s}_e)$, or $D(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{s}) > D(\boldsymbol{y}_e, \boldsymbol{z}_e, \boldsymbol{s}_e)$, or $(\pi(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{s}), D(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{s})) = (\pi(\boldsymbol{y}_e, \boldsymbol{z}_e, \boldsymbol{s}_e), D(\boldsymbol{y}_e, \boldsymbol{z}_e, \boldsymbol{s}_e))$, for all $\boldsymbol{y} = \{y_1, ..., y_m\}$, $\boldsymbol{z} = \{z_1, ..., z_m\}$ and $\boldsymbol{s} = \{s_1, ..., s_m\}$ where $(y_i, z_i) \in \{(0,0), (0,1), (1,0)\}$ and $s_i \in \mathbb{N}_0 = \{0, 1, 2, ...\}$ for all $i \in M$. Let ϵ denote the set of all efficient solutions of problem (Q1). Then, the points $(\pi(\boldsymbol{y}_e, \boldsymbol{z}_e, \boldsymbol{s}_e), D(\boldsymbol{y}_e, \boldsymbol{z}_e, , \boldsymbol{s}_e))$, $(\boldsymbol{y}_e, \boldsymbol{z}_e, \boldsymbol{s}_e) \in \epsilon$, constitute an *efficient frontier* for the total costs vs. total downtime. From this efficient frontier, an appropriate solution for Problem (Q0) may be selected.

In Subsection 3.4.1, we develop a procedure for finding an optimal solution $(y_i^*(\lambda), z_i^*(\lambda), s_i^*(\lambda))$ for given values of $\lambda \geq 0$ and $i \in M$ (see Lemma 3.3). Once this procedure is applied for all $i \in M$, it suffices to obtain an optimal solution $(\boldsymbol{y}^*(\lambda), \boldsymbol{z}^*(\lambda), \boldsymbol{s}^*(\lambda))$ for $\mathrm{Q}(\lambda)$, without considering all 3^m combinations of policies for the multi-stage problem. Then, a set of efficient solutions can be generated by varying values of λ and repeating this procedure.

An intuitive explanation for the Lagrangian multiplier λ in $L_i(y_i, z_i, s_i, \lambda)$ and $L(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{s}, \lambda)$ is as follows. Remember the interpretations of $\hat{h}_i$, $\hat{r}_{1,i}$ and $\hat{r}_{2,i}$ as the parameters that include the discounting effect. As $D_i(y_i, z_i, s_i)$ is the expected downtime stemming from failures of stage-i subsystems throughout $[0, T]$, $\lambda D_i(y_i, z_i, s_i)$ becomes the expected NPV of downtime costs stemming from failures of stage-i subsystems once λ is considered as a downtime penalty rate that includes the discounting effect throughout $[0, T]$. Then, $L_i(y_i, z_i, s_i, \lambda)$ and $L(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{s}, \lambda)$ can be interpreted as pure cost formulations which also include downtime costs. Also remember that the effect of downtime is incorporated into the model through downtime costs in Chapter 2. So, this interpretation also shows the consistency between the approaches followed to incorporate the downtime costs into the models in this chapter and in Chapter 2.

The effect of downtime is incorporated differently in the two models. In the reliability optimization model for critical components (Chapter 2), downtime costs are included in the model, while there is a constraint on the total uptime (or downtime) throughout the lifetime of a number of systems in the redundancy allocation model (Chapter 3). These two approaches are consistent with each other.

By Theorem 2 in Everett (1963), $D_i(y_i^*(\lambda), z_i^*(\lambda), s_i^*(\lambda))$ in Problem ($Q_i(\lambda)$) is decreasing in λ for all $i \in M$. A direct result of this property is that $D(\boldsymbol{y}^*(\lambda), \boldsymbol{z}^*(\lambda), \boldsymbol{s}^*(\lambda))$ in Problem ($Q(\lambda)$) will also be decreasing in λ. That is, if we start with a λ value and find solutions for each single-stage problem ($Q_i(\lambda)$) by increasing λ, the resource levels of the downtime constraint in Problem $Q_i(\lambda)$ in those solutions will be decreasing . Likewise, the corresponding resource levels of the downtime constraint in Problem $Q(\lambda)$ will be decreasing (equivalently, corresponding availability levels will be increasing), too.

We will make use of this property and the interpretation of λ as a downtime penalty rate to compare the three policies in subsection 3.4.2 and to derive results for the system level problem in subsection 3.4.3.

3.4. Analysis

In this section, we first provide a number of results for the optimization of the single-stage problems. Then, we compare the three policies on stage-level. We finalize the section by introducing a method to construct an efficient frontier and an ordering of the stages in a capital good for redundancy, which are results for the multi-stage problem.

3.4.1 Optimization of the Single-Stage Problems

In this subsection, we derive three lemmas. Lemma 3.1 states that $L_i(y_i, z_i, s_i, \lambda)$ is strictly convex in s_i for a given policy and given values of λ. We introduce a number of properties of the optimal s_i values for a given policy and $L_i(y_i, z_i, s_i, \lambda)$ in Lemma 3.2. Lemma 3.3 details an optimization procedure for $L_i(y_i, z_i, s_i, \lambda)$ for a given value of λ. An optimal solution for $L(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{s}, \lambda)$ can be found by generating optimal solutions of $L_i(y_i, z_i, s_i, \lambda)$ for all $i \in M$ by this procedure.

Lemma 3.1 For all $i \in M$, for a given policy $(y_i, z_i) \in \{(0,0),(0,1),(1,0)\}$ and for a given value of $\lambda \geq 0$, $L_i(y_i, z_i, s_i, \lambda)$ is convex in s_i .

Proof: Let $i \in M$ and $\lambda \geq 0$. For Policy (0,0),

$$L_i(0,0,s_i,\lambda) = (c_{0,i}+\hat{h}_i T)s_i + \frac{NT}{\tau_i}(\hat{r}_{1,i}+\lambda\mu_{1,i}) + \frac{NT}{\tau_i}[(\hat{r}_{2,i}-\hat{r}_{1,i})+\lambda(\mu_{2,i}-\mu_{1,i})]B_i(s_i) \tag{3.5}$$

The second and fourth term of (3.5) is constant and the first term is linear in s_i. The Erlang Loss probability $B_i(s_i)$ is strictly decreasing and strictly convex in s_i (see Karush (1957) and also Kranenburg and van Houtum (2007)). It holds that $r_{2,i} \geq r_{1,i}$, $\hat{r}_{2,i} \geq \hat{r}_{1,i}$, and $\mu_{2,i} \geq \mu_{1,i}$. The last term is convex in s_i as $\mu_{2,i} \geq \mu_{1,i}$, and thus $L_i(0,0,s_i\lambda)$ is convex.

For Policy (0,1),

$$L_i(0,1,s_i,\lambda) = (c_{0,i}+\hat{h}_i T)s_i + \frac{NT}{\tau_i}(\hat{r}_{1,i}+\lambda\mu_{1,i}) + \frac{NT}{\tau_i}(\hat{r}_{2,i}-\hat{r}_{1,i})B_i(s_i-1), \tag{3.6}$$

and for Policy (1,0),

$$L_i(1,0,s_i,\lambda) = Nc_{1,i} + (c_{0,i}+\hat{h}_i T)s_i + \frac{NT}{\tau_i}\hat{r}_{1,i} + \frac{NT}{\tau_i}(\hat{r}_{2,i}-\hat{r}_{1,i})B_i(s_i) \tag{3.7}$$

Notice that $L_i(1,0,s_i,\lambda)$ is independent of λ, as there is no downtime in Policy (1,0). By following similar arguments that we used for $L_i(0,0,s_i,\lambda)$, it can be shown that $L_i(0,1,s_i,\lambda)$ and $L_i(1,0,s_i,\lambda)$ are convex in s_i for a given value of λ. $\square$

We define

$$s_i^*(y_i,z_i,\lambda) = \arg\min_{s_i}\{L_i(y_i,z_i,s_i,\lambda) | s_i \in \mathbb{N}_0\}.$$

That is, for a given value of λ, $s_i^*(0,0,\lambda)$, $s_i^*(0,1,\lambda)$, and $s_i^*(1,0,\lambda)$ are the smallest values of s_i under which $L_i(0,0,s_i,\lambda)$, $L_i(0,1,s_i,\lambda)$ and $L_i(1,0,s_i,\lambda)$ are minimized (optimal values of initial supply in Policy (0,0), Policy (0,1), and Policy (1,0)), respectively. We also define

$$\Delta B_i(s_i) = B_i(s_i) - B_i(s_i+1).$$

$\Delta B_i(s_i)$ is positive for all s_i. As $B_i(s_i)$ is strictly convex in s_i, $\Delta B_i(s_i)$ is strictly decreasing in s_i.

Lemma 3.2 For all $i \in M$, it holds that:

(i) $s_i^*(1,0,\lambda) = s_i^*(1,0)$ and $L_i(1,0,s_i^*(1,0,\lambda),\lambda) = L_i(1,0,s_i^*(1,0),0) = L_i(1,0)$ (constant) for all $\lambda \geq 0$, where

$$s_i^*(1,0) = \min\left\{ s_i \in \mathbb{N}_0 \mid \Delta B_i(s_i) \leq \frac{(c_{0,i} + \hat{h}_i T)}{(\hat{r}_{2,i} - \hat{r}_{1,i})} \frac{\tau_i}{NT} \right\}. \tag{3.8}$$

(ii) $s_i^*(0,1,\lambda) = s_i^*(1,0)+1$ for all $\lambda \geq 0$ and $L_i(0,1,s_i^*(0,1,\lambda),\lambda) = L_i(0,1,s_i^*(1,0)+1,\lambda) = L_i(0,1,\lambda)$ is an increasing linear function of λ.

(iii) $s_i^*(0,0,0) = s_i^*(1,0)$, $L_i(0,0,s_i^*(0,0,\lambda),\lambda) = L_i(0,0,\lambda)$ is a strictly increasing, concave, piecewise linear function of λ, and $s_i^*(0,0,\lambda)$ is increasing in λ.

Proof: Let $\Delta_s L_i(y_i,z_i,s_i,\lambda) = L_i(y_i,z_i,s_i+1,\lambda) - L_i(y_i,z_i,s_i,\lambda)$.

(i) The property that $L_i(1,0,s_i,\lambda)$ is independent of λ (see equation (3.7)), for all $\lambda \geq 0$, implies that $s_i^*(1,0,\lambda)$ is independent of λ and may be denoted as $s_i^*(1,0)$. $L_i(1,0,s_i^*(1,0,\lambda),\lambda) = L_i(1,0,s_i^*(1,0),0)$ is also a constant that we denote by $L_i(1,0)$. As $L_i(1,0,s_i,\lambda)$ is strictly convex in s_i, $s_i^*(1,0) = \min\{s_i \in \mathbb{N}_0 \mid \Delta_\mathrm{s}\mathrm{L}_\mathrm{i}(1,0,\mathrm{s}_\mathrm{i},\lambda) \geq 0\}$.

$$\Delta_s L_i(1,0,s_i,\lambda) = (c_{0,i} + \hat{h}_i T) - \frac{NT}{\tau_i}(\hat{r}_{2,i} - \hat{r}_{1,i})\Delta B_i(s_i)$$

implies equation (3.8).

(ii) As $L_i(0,1,s_i,\lambda)$ is strictly convex in s_i,

$$s_i^*(0,1,\lambda) = \min\{s_i \in \mathbb{N}_0 \mid \Delta_s L_i(0,1,s_i,\lambda) \geq 0\}.$$

$$\Delta_s L_i(0,1,s_i,\lambda) = (c_{0,i} + \hat{h}_i T) - \frac{NT}{\tau_i}(\hat{r}_{2,i} - \hat{r}_{1,i})\Delta B_i(s_i - 1);$$

thus

$$s_i^*(0,1,\lambda) = \min\left\{ s_i \in \mathbb{N}_0 \mid \Delta B_i(s_i - 1) \leq \frac{c_{0,i} + \hat{h}_i T}{\hat{r}_{2,i} - \hat{r}_{1,i}} \frac{\tau_i}{NT} \right\} = s_i^*(1,0) + 1.$$

$$\begin{aligned} L_i(0,1,\lambda) &= L_i(0,1,s_i^*(0,1,\lambda),\lambda) = L_i(0,1,s_i^*(1,0)+1,\lambda) \\ &= (c_0 + \hat{h}_i T)(s_i^*(1,0)+1) \\ &\quad + \frac{NT}{\tau}\left[\hat{r_{1,i}} + (\hat{r_{2,i}} - \hat{r_{1,i}})B_i(s_i^*(1,0))\right] + \frac{NT}{\tau}\mu_{1,i}\lambda \end{aligned}$$

is an increasing linear function of λ as N, T, and $\mu_{1,i} \geq 0$.

(iii) As $L_i(0,0,s_i,\lambda)$ is strictly convex in s_i for $\lambda \geq 0$,

$$s_i^*(0,0,\lambda) = \min\left\{s_i \in \mathbb{N}_0 \mid \Delta_s L_i(0,0,s_i,\lambda) \geq 0\right\}.$$

$$\Delta_s L_i(0,0,s_i,\lambda) = (c_{0,i} + \hat{h}_i T) - \frac{NT}{\tau_i}[(\hat{r}_{2,i} - \hat{r}_{1,i}) + (\mu_{2,i} - \mu_{1,i})\lambda]\Delta B_i(s_i);$$

thus

$$s_i^*(0,0,\lambda) = \min\left\{s_i \in \mathbb{N}_0 \mid \Delta B_i(s_i) \leq \frac{c_{0,i} + \hat{h}_i T}{\hat{r}_{2,i} - \hat{r}_{1,i} + (\mu_{2,i} - \mu_{1,i})\lambda}\frac{\tau_i}{NT}\right\}. \tag{3.9}$$

Obviously, $s_i(0,0,0) = s_i^*(1,0)$.

For a given $s_i \in \mathbb{N}_0$, $L_i(0,0,s_i,\lambda)$ is a linear function of λ; see equation (3.5). The function $L_i(0,0,s_i,\lambda)$ starts with $L_i(0,0,s_i,0)$, which is increasing in s_i. The slope of $L_i(0,0,s_i,\lambda)$ is positive for each s_i, and the slope decreases as s_i increases. $L_i(0,0,s_i^*(0,0,\lambda),\lambda) = \min\{L_i(0,0,s_i,\lambda), s_i \in \mathbb{N}_0\}$; hence, $L_i(0,0,\lambda) = L_i(0,0,s_i^*(0,0,\lambda),\lambda)$ is strictly increasing, concave, and piecewise linear function of λ and $s_i^*(0,0,\lambda)$ is increasing in λ. In Figure 3.1, we illustrate $L_i(0,0,\lambda)$ for a given stage in Example 3.1.

□

Example 3.1 Consider the redundancy allocation problem for a capital good with two stages. $N = 15$ systems are purchased with an expected lifetime of $T = 15$ years. The annual discount rate is $\alpha = 0.05$. The parameters for the stages are given in Table 3.2 (we will also refer to this table in the other examples in this chapter later).

Table 3.2 Parameters for Example 3.1

	Stage 1	Stage 2
τ (years)	3	6
$c_{0,i}$ (Euros)	5000	125000
$c_{1,i}$ (Euros)	4000	125000
h_i (Euros per month)	75	1875
$r_{1,i}$ (Euros)	1000	25000
$r_{2,i}$ (Euros)	2000	50000
$\mu_{1,i}$ (hours)	10	8
$\mu_{2,i}$ (hours)	24	48
U_i (months)	3	3

In Figure 3.1, the solid line depicts the minimum of the Lagrangian functions under Policy (0,0) for stage 1, $L_1(0,0,\lambda) = L_1(0,0,s_i^*(0,0,\lambda),\lambda)$. The dotted lines are the Lagrangian functions $L_1(0,0,s,\lambda)$ of Policy (0,0) for three different values of s. By part (iii) of Lemma 3.2, we know that $s_1^*(0,0,0) = s_1^*(1,0)$, hence we start plotting $L_1(0,0,s,\lambda)$ with $s = s_1^*(1,0) = 2$. Observe that $L_1(0,0,\lambda)$ is a strictly increasing, concave, piecewise linear function of λ.

We provide the following intuitions for Lemma 3.2:

(i) As there is no downtime in Policy (1,0) ($D_i(1,0,s_i) = 0$ for all s_i), the downtime penalty rate (λ) does not affect the optimal number of spare parts and the optimal total costs.

(ii) In Policy (0,1), all failures result in downtimes equal to the respective inventory-replacement times, independent of inventory-on-hand or inventory positions. Hence, the downtime penalty rate (λ) does not affect the optimal number of spare parts. However, it affects optimal total costs due to its effect on downtime costs.

(iii) In Policy (0,0), downtime per failure is equal to either inventory-replacement time or OEM-replacement time depending on the stock-on-hand, which is

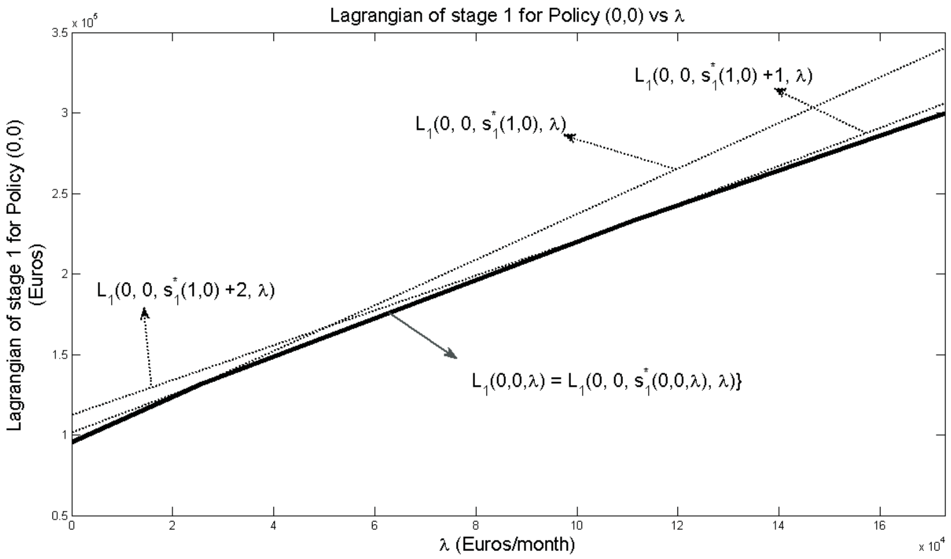

Figure 3.1 Lagrangian of stage 1 for Policy (0,0) in Example 3.1

affected by the initial supply. Hence, the optimal initial supply amount changes with the downtime penalty rate (λ). Increasing the downtime penalty rate results in increasing downtime costs, unless downtime is decreased. Thus, one has to keep a higher number of spare parts to compensate for the increase in downtime penalty rate by a decrease in downtime. However, this does not fully compensate for the increase in downtime costs, and total costs increases.

Lemma 3.3 For all $i \in M$, for a given value of $\lambda \geq 0$, the following procedure determines a solution $(y_i^*(\lambda), z_i^*(\lambda), s_i^*(\lambda))$ which minimizes $L_i(y_i, z_i, s_i, \lambda)$.

1. Determine $s_i^*(1,0)$ by equation (3.8).
2. Determine $s_i^*(0,0,\lambda)$ by equation (3.9).
3. Find

$$\begin{aligned}(y_i^*(\lambda), z_i^*(\lambda), s_i^*(\lambda)) \quad &= \quad \underset{(y_i,z_i,s_i)}{\arg\min}\Big\{ L_i(y_i, z_i, s_i, \lambda) | (\mathrm{y_i}, \mathrm{z_i}, \mathrm{s_i}) \in \\ &\Big\{(0,0,s_i^*(0,0,\lambda)), (0,1,s_i^*(1,0)+1), (1,0,s_i^*(1,0))\Big\}\Big\}.\end{aligned}$$

3.4.2 Comparison of the Policies

In this subsection, we will first make pairwise comparisons of the three policies for single-stage problems. Then, we will give an overall comparison. We will use $L_i(0,0,\lambda)$, $L_i(0,1,\lambda)$, and $L_i(1,0)$ (see Lemma 3.2) for the comparisons as they constitute the optimal values of the Lagrangian function as a function of λ in Policy (0,0), Policy (0,1), and Policy (1,0), respectively. We will illustrate the comparisons by examples.

Policy (0,1) and Policy (1,0)

Lemma 3.4 For all $i \in M$,

(i) If $Nc_{1,i} \leq c_{0,i} + \frac{h_i}{\alpha}(1 - e^{-\alpha T})$, then Policy (1,0) outperforms Policy (0,1) for all $\lambda \geq 0$.

(ii) If $Nc_{1,i} > c_{0,i} + \frac{h_i}{\alpha}(1 - e^{-\alpha T})$, then Policy (0,1) outperforms Policy (1,0) for $0 \leq \lambda < \lambda_{01-10,i} = \frac{\tau_i}{NT\mu_{1,i}}[Nc_{1,i} - c_{0,i} - \frac{h_i}{\alpha}(1 - e^{-\alpha T})]$; and Policy (1,0) outperforms Policy (0,1) for $\lambda > \lambda_{01-10,i}$.

Proof: As $s_i^*(0,1,\lambda) = s_1^*(1,0) + 1$ for all $\lambda \geq 0$, the following can be deduced by equations (3.6) and (3.7):

$$L_i(0,1,\lambda) = L_i(1,0) - Nc_{1,i} + c_{0,i} + \hat{h}_i T + \frac{NT}{\tau_i}\mu_{1,i}\lambda.$$

Observe that $L_i(0,1,0) = L_i(1,0) - Nc_{1,i} + c_{0,i} + \hat{h}_i T$ and $L_i(0,1,\lambda)$ is an increasing linear function of λ. Remember that $\hat{h}_i = \frac{h_i}{\alpha T}(1 - e^{-\alpha T})$ to see the relation to h_i and T.

(i) If $Nc_{1,i} \leq c_{0,i} + \frac{h_i}{\alpha}(1 - e^{-\alpha T})$, $Ł_i(1,0) \leq L_i(0,1,\lambda)$ for all $\lambda \geq 0$.

(ii) If $Nc_{1,i} > c_{0,i} + \frac{h_i}{\alpha}(1 - e^{-\alpha T})$, $Ł_i(1,0)$ and $L_i(0,1,\lambda)$ intersects at $\lambda_{01-10,i} = \frac{\tau_i}{NT\mu_{1,i}}[Nc_{1,i} - c_{0,i} - \frac{h_i}{\alpha}(1 - e^{-\alpha T})]$. For $0 \leq \lambda < \lambda_{01-10,i}$, $L_i(0,1,\lambda) \leq Ł_i(1,0)$; for $\lambda > \lambda_{01-10,i}$, $Ł_i(1,0) \leq L_i(0,1,\lambda)$.

□

Example 3.2 Consider the redundancy allocation problem introduced in Example 3.1 with $N = 15$, $T = 15$ years, $\alpha = 0.05$ annually, and the parameters given in Table

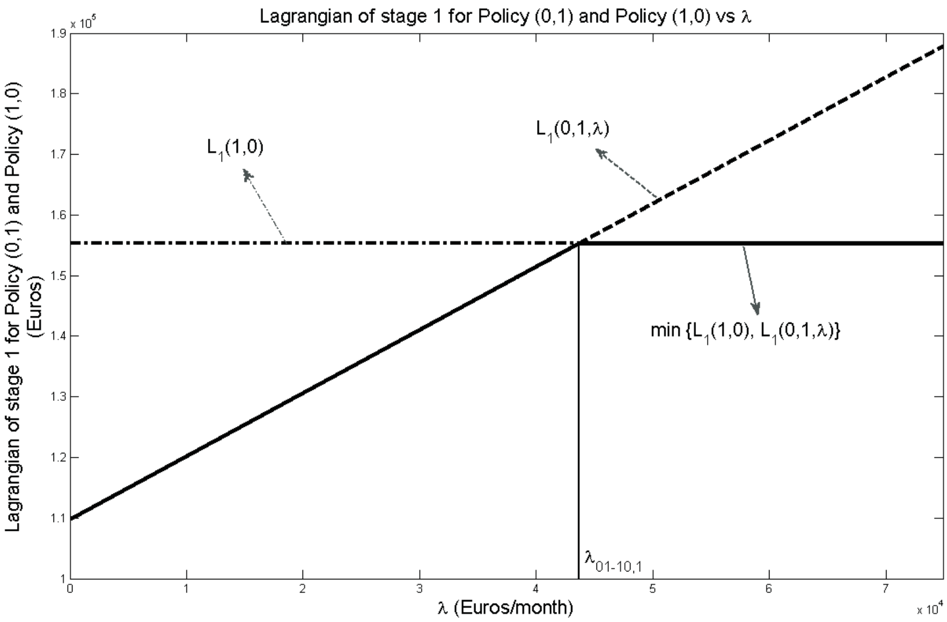

Figure 3.2 Comparison of Policy (0,1) and Policy (1,0) for stage 1 in Example 3.2

3.2. In Figure 3.2, you can see the comparison of Policy (0,1) and Policy (1,0) for stage 1. The dashed line is the Lagrangian function $L_1(0,1,\lambda) = L_1(0,1,s_i^*(1,0)+1,\lambda)$, where $s_1^*(1,0) = 2$. The dash-dot line (-.) is the Lagrangian function $L_1(1,0,s_1^*(1,0),\lambda) = L_1(1,0)$. The solid line is the minimum of the Lagrangian functions of Policy (0,1) and Policy (1,0) for $\lambda \geq 0$. $L_1(0,0,s,\lambda)$ and $L_{01,i}(\lambda)$ intersect at $\lambda_{01-10,1} = 43682.49$ Euros/month, Policy (0,1) outperforms Policy (1,0) for $\lambda \leq \lambda_{01-10,1}$ and Policy (1,0) outperforms Policy (0,1) for $\lambda > \lambda_{01-10,1}$.

The result given in Lemma 3.6 can be interpreted in two ways:

- For small values of the downtime penalty rate λ, Policy (0,1) results in less TCO. After a certain value, Policy (1,0) provides less TCO.
- When the downtime constraint is relatively loose, Policy (0,1) satisfies it with less TCO. As it gets tighter and tighter, it is met by Policy (1,0) with less TCO after a certain value of the resource level of the downtime constraint.

In Corollary 3.1 we provide the relation of $\lambda_{01-10,i}$ to each model parameter. Corollary 3.2 specifies an upper bound for $\lambda_{01-10,i}$. As an explicit formula for $\lambda_{01-10,i}$ can be derived and the formula is not complex, the proofs of these corollaries are trivial and we will not elaborate on them here.

Corollary 3.1 For all $i \in M$, $\lambda_{01-10,i}$ is

(i) decreasing as a function of T, $\mu_{1,i}$, and $c_{0,i}$;

(ii) increasing as a function of N, τ_i, and $c_{1,i}$; and,

(iii) independent of $\mu_{2,i}$, $r_{i,1}$, $r_{i,2}$ and U_i.

Corollary 3.2

$$\lambda_{01-10,i} < c_{1,i}\left(\frac{T}{\tau_i}\mu_{1,i}\right)^{-1}.$$

Parts (i) and (ii) in Corollary 3.1 are intuitively easy to understand: The increases in the parameters in (i) favors Policy (1,0) and it outperforms Policy (0,1) at a smaller value of λ (or availability level), while the increases in the parameters in (ii) favors Policy (0,1). The result that $\lambda_{01-10,i}$ is independent of $\mu_{2,i}$ is also expected as downtime stemming from OEM-replacements does not occur in either of the policies. However, independence of $r_{i,1}$, $r_{i,2}$ and U_i are not trivial, as given a sample path for the failures of component-i parts, the provision procedure and the emergency procedure are applied at different points in time in different ways under Policy (0,1) and Policy

(1,0), respectively. These parameters affect the repair costs in both policies: $r_{i,1}$, $r_{i,2}$ affects them directly while the effect of U_i is through the Erlang loss probability. This independence stem form the relationship between the optimal initial supply amounts, $s_i^*(0,1,\lambda) = s_i^*(1,0)+1$ for all $\lambda \geq 0$. By the comparison of equations (3.6) and (3.7), we see that the effects of these parameters depend on $B_i(s_i-1)$ and $B_i(s_i)$ under Policy (0,1) and Policy (1,0), respectively. These effects are equal in the two policies due to the relationship of the optimal values of s_i.

Corollary 3.2 can be explained as follows: Due to the relationship between the optimal initial supply amounts in the two policies, given a sample path for the failures of component-i parts, the ordinary procedure is applied for the same failures (at the same time points) under both policies. The provision procedure and the emergency procedure are applied under Policy (0,1) and Policy (1,0), respectively, for the failures which are not handled by the ordinary procedure. As the expected costs incurred per application of the provision procedure and the expected costs incurred per the application of the emergency procedure are equal, the repair costs are equal under the two policies. This leaves $\frac{T}{\tau_i}\mu_{1,i}^{-1}\lambda$, the total expected downtime costs that would be incurred per system under Policy (0,1) for a given downtime penalty rate $\lambda \geq 0$, as all failures lead to inventory replacements. There is no downtime cost under Policy (1,0). Hence, Policy (0,1) can outperform Policy (1,0) only if the extra cost for redundancy ($c_{1,i}$) is less than the expected total downtime costs that would be incurred per system.

Policy (0,0) and Policy (1,0)

Lemma 3.5 For all $i \in M$, the followings hold:

(i) For $0 \leq \lambda < c_{1,i}\left(\frac{T}{\tau_i}\mu_{2,i}\right)^{-1}$, Policy (0,0) outperforms Policy (1,0).

(ii) For $c_{1,i}\left(\frac{T}{\tau_i}\mu_{1,i}\right)^{-1} < \lambda$, Policy (1,0) outperforms Policy (0,0).

(iii) There exists a $\lambda_{00-10,i}$, $c_{1,i}\left(\frac{T}{\tau_i}\mu_{2,i}\right)^{-1} \leq \lambda_{00-10,i} \leq c_{1,i}\left(\frac{T}{\tau_i}\mu_{1,i}\right)^{-1}$, such that Policy (0,0) outperforms Policy (1,0) for all $\lambda \leq \lambda_{00-10,i}$ and Policy (1,0) outperforms Policy (0,0) for all $\lambda > \lambda_{00-10,i}$.

Proof: For a given s_i, define

$$\Delta_y L_i(z_i,s_i,\lambda) = L_i(1,z_i,s_i,\lambda) - L_i(0,z_i,s_i,\lambda) = Nc_{1,i} - \frac{NT}{\tau_i}[\mu_{1,i} + (\mu_{2,i}-\mu_{1,i})B_i(s_i)]\lambda.$$

Then $\Delta_y L_i(0,s_i,\lambda)$ is the difference between the decentralized Lagrangian of Policy (1,0) and the decentralized Lagrangian of Policy (0,0).

(i) If $0 \leq \lambda < c_{1,i}\left(\frac{T}{\tau_i}\mu_{2,i}\right)^{-1}$, then $\Delta_y L_i(0, s_i, \lambda) > Nc_{1,i}(\mu_{2,i} - \mu_{1,i})[1 - B_i(s_i)]$. As $Nc_{1,i}(\mu_{2,i} - \mu_{1,i})[1 - B_i(s_i)] \geq 0$ for all $s_i \in \mathbb{N}_0$, Policy (0,0) outperforms Policy (1,0).

(ii) If $c_{1,i}\left(\frac{T}{\tau_i}\mu_{1,i}\right)^{-1} < \lambda$, then $\Delta_y L_i(0, s_i, \lambda) < -Nc_{1,i}(\mu_{2,i} - \mu_{1,i})B_i(s_i)$. As $-Nc_{1,i}(\mu_{2,i} - \mu_{1,i})B_i(s_i) \leq 0$ for all $s_i \in \mathbb{N}_0$, Policy (1,0) outperforms Policy (0,0).

(iii) By (i), $L_i(0, 0, \lambda) < L_i(1, 0)$ for $0 \leq \lambda < c_{1,i}\left(\frac{T}{\tau_i}\mu_{2,i}\right)^{-1}$. It is trivial to show that $\lim_{\lambda\to\infty} L_i(0, 0, \lambda) = \infty$. As $L_i(0, 0, \lambda)$ is a strictly increasing function of λ and $L_i(1, 0)$ is a constant, there exists a $\lambda_{00-10,i}$ such that $L_i(0, 0, \lambda) \leq L_i(1, 0)$ for all $\lambda \leq \lambda_{00-10,i}$ and $L_i(0, 0, \lambda) > L_i(1, 0)$ for all $\lambda > \lambda_{00-10,i}$. By (i) and (ii), $c_{1,i}\left(\frac{T}{\tau_i}\mu_{2,i}\right)^{-1} \leq \lambda_{00-10,i} \leq c_{1,i}\left(\frac{T}{\tau_i}\mu_{1,i}\right)^{-1}$.

□

A closed form expression cannot be derived for $\lambda_{00-10,i}$, but it can be found by simple numerical procedures as these are the values λ at which two linear functions intersect.

Example 3.3 Consider the redundancy allocation problem introduced in Example 3.1. In Figure 3.3, you can see the comparison of Policy (0,0) and Policy (1,0) for stage 1. The dotted lines are the Lagrangian functions $L_1(0, 0, s, \lambda)$ of Policy (0,0) for three different values of s. As we know that $s_1^*(0, 0, 0) = s_1^*(1, 0)$, we start plotting $L_1(0, 0, s, \lambda)$ with $s = s_1^*(1, 0) = 2$. The dash-dot line (-.) is the Lagrangian function $L_1(1, 0, s_1^*(1, 0), \lambda) = L_1(1, 0)$. The solid line is the minimum of the Lagrangian functions of Policy (0,0) and Policy (1,0) for $\lambda \geq 0$. $L_1(0, 0, s, \lambda)$ cannot be the minimum of the Lagrangian functions of Policy (0,0) and Policy (1,0) at $\lambda > 0$ for $s \geq s_1^*(1, 0) + 2$. $L_{10,1}$ and $L_1(0, 0, \lambda)$ intersects at $\lambda_{00-10,1} = 45630.35$ Euros/month, Policy (0,0) outperforms Policy (1,0) for $\lambda \leq \lambda_{00-10,1}$ and Policy (1,0) outperforms Policy (0,0) for $\lambda > \lambda_{00-10,1}$.

We now provide some intuition for Lemma 3.5 with interpretation of λ as downtime penalty rate. Notice that given a sample path for the failures of component-i parts, the ordinary procedure and the emergency procedure are applied in the same way under both policies at the failure times. $\frac{T}{\tau_i}\lambda\mu_{1,i}^{-1}$ is the total expected downtime costs that would be incurred per system if all failures are handled by the ordinary procedure and $\frac{T}{\tau_i}\lambda\mu_{2,i}^{-1}$ is the total expected downtime costs that would be incurred per system if all failures are handled by the emergency procedure. Then, Lemma 3.5 states that

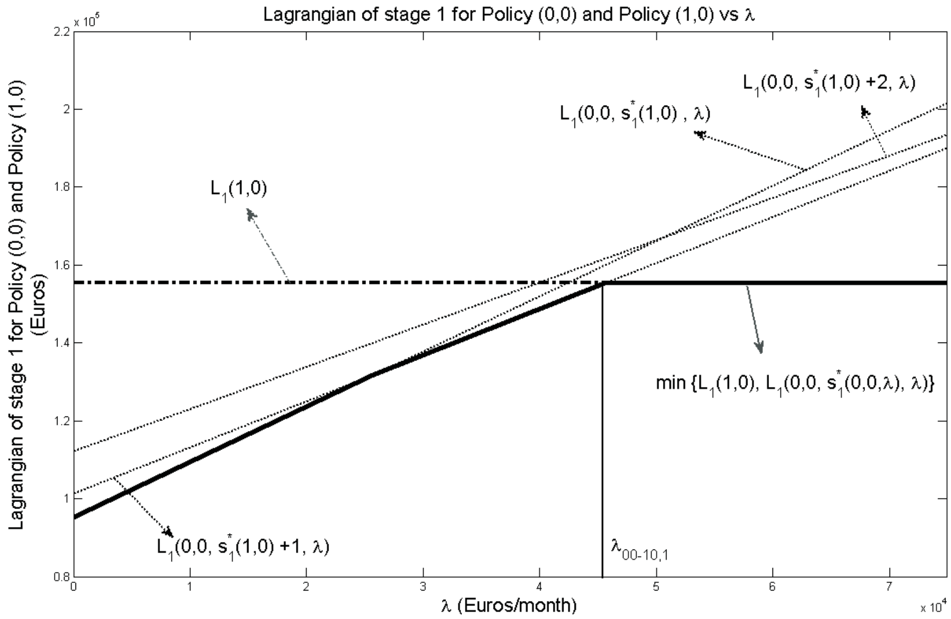

Figure 3.3 Comparison of Policy (0,0) and Policy (1,0) for stage 1 in Example 3.3

(i) If the extra cost for redundancy ($c_{1,i}$) is greater than or equal to the total expected downtime costs that would be incurred per system if all failures are handled by the emergency procedure, not choosing redundancy is optimal. Without redundancy, handling all failures by the emergency procedure represents the *worst case* in terms of the expected downtime costs. Not choosing redundancy is optimal as the downtime costs that would be incurred in the worst case is less than the cost of the extra unit for redundancy.

(ii) If the extra cost for redundancy is less than or equal to the expected total downtime costs that would be incurred per system if all failures are handled by the ordinary procedure, choosing redundancy is optimal. Without redundancy, handling all failures by the ordinary procedure represents the *best case* in terms of the expected downtime costs. Choosing redundancy is optimal as no downtime cost is incurred and the extra unit for redundancy is less costly than the best case for not choosing redundancy.

(iii) There exists a point between the worst case and the best case where choosing redundancy becomes optimal as the downtime penalty rate increases. Increasing the downtime penalty rate after that point will not change the redundancy decision as downtime costs are the incentive to choose redundancy.

An interpretation of part (iii) of Lemma 3.5 with respect to the downtime constraint, similar to the one given for Lemma 3.4, can also be given: When the downtime constraint is not tight, one can satisfy it with less TCO by not choosing redundancy. As the constraint gets tighter and tighter, one has to choose redundancy to satisfy it with less TCO after a certain value of the resource level of the downtime constraint.

Policy (0,0) and Policy (0,1)

Lemma 3.6 For all $i \in M$, there exists a $\lambda_{00-01,i} > 0$ such that Policy (0,0) outperforms Policy (0,1) for $\lambda \leq \lambda_{00-01,i}$ and Policy (0,1) outperforms Policy (0,0) for $\lambda > \lambda_{00-01,i}$.

Proof: For all $i \in M$, remember that $L_i(0,0,\lambda)$ is a strictly increasing, concave, piecewise linear function of λ whose pieces are constructed by $L_i(0,0,s_i,\lambda)$, $s_i \in \{s_i^*(1,0), s_i^*(1,0)+1, ...\}$ as $s_i^*(0,0,0) = s_i^*(1,0)$. It is trivial to show that $L_i(0,0,0) < L_i(0,1,0)$ and the slope of $L_i(0,0,s_i,\lambda)$ is greater than the slope of $L_i(0,1,\lambda)$ for each $s_i \in \{s_i^*(1,0), s_i^*(1,0)+1, ...\}$. Hence, there exists a $\lambda_{00-01,i} > 0$ such that Policy (0,0) is optimal for $\lambda \leq \lambda_{00-01,i}$ and Policy (0,0) is optimal for $\lambda > \lambda_{00-01,i}$. □

As it was the case for $\lambda_{00-10,i}$, a closed form expression cannot be derived for $\lambda_{00-01,i}$ but it can be found numerically.

Example 3.4 Consider the redundancy allocation problem introduced in Example 3.1. In Figure 3.4, you can see the comparison of Policy (0,0) and Policy (0,1) for stage 1. The dotted lines are the Lagrangian functions $L_1(0,0,s,\lambda)$ of Policy (0,0) for three different values of s. We start plotting $L_1(0,0,s,\lambda)$ with $s = s_1^*(1,0) = 2$. The dashed line is the Lagrangian function $L_1(0,1,\lambda) = L_1(0,1,s_i^*(1,0)+1,\lambda)$, where $s_1^*(1,0)+1 = 3$. $L_1(0,0,s,\lambda)$ and $L_1(0,1,\lambda)$ do not intersect at $\lambda > 0$ for $s \geq s_1^*(1,0)+2$. The solid line is the minimum of the Lagrangian functions of Policy (0,0) and Policy (0,1) for $\lambda \geq 0$. $L_1(0,1,\lambda)$ and $L_1(0,0,\lambda)$ intersects at $\lambda_{00-01,1} = 59977.70$ Euros/month, Policy (0,0) outperforms Policy (0,1) for $\lambda \leq \lambda_{00-01,1}$ and Policy (0,1) outperforms Policy (0,0) for $\lambda > \lambda_{00-01,1}$.

Intuitions similar to those given for Lemma 3.4 and Lemma 3.5 can also be provided for Lemma 3.6.

Overall Comparison

Theorem 3.1 For all $i \in M$:

- If $Nc_{1,i} \leq c_{0,i} + \frac{h_i}{\alpha}(1 - e^{-\alpha T})$, Policy (0,0) is optimal for $\lambda \in [0, \lambda_{00-10,i}]$ and Policy (1,0) is optimal for $\lambda \in [\lambda_{00-10,i}, \infty)$.
- If $Nc_{1,i} > c_{0,i} + \frac{h_i}{\alpha}(1 - e^{-\alpha T})$, either
  - Policy (0,0) is optimal for $\lambda \in [0, \lambda_{00-10,i}]$ and Policy (1,0) is optimal for $\lambda \in [\lambda_{00-10,i}, \infty)$; or,
  - Policy (0,0) is optimal for $\lambda \in [0, \lambda_{00-01,i}]$, Policy (0,1) is optimal for $\lambda \in [\lambda_{00-01,i}, \lambda_{01-10,i}]$, and Policy (1,0) is optimal for $\lambda \in [\lambda_{01-10,i}, \infty)$.

Proof:

- If $Nc_{1,i} \leq c_{0,i} + \frac{h_i}{\alpha}(1 - e^{-\alpha T})$, Policy (1,0) outperforms Policy (0,1) for all $\lambda \geq 0$ by Lemma 3.4. Then, by Lemma 3.5, Policy (0,0) is optimal for $\lambda \in [0, \lambda_{00-10,i}]$ and Policy (1,0) is optimal for $\lambda \in [\lambda_{00-10,i}, \infty)$.

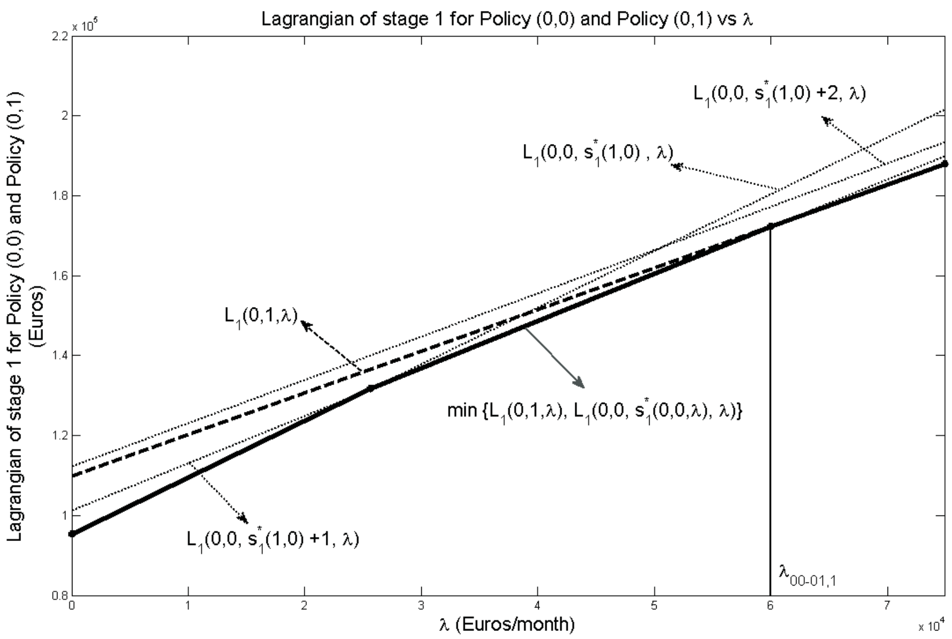

Figure 3.4 Comparison of Policy (0,0) and Policy (0,1) for stage 1 in Example 3.4

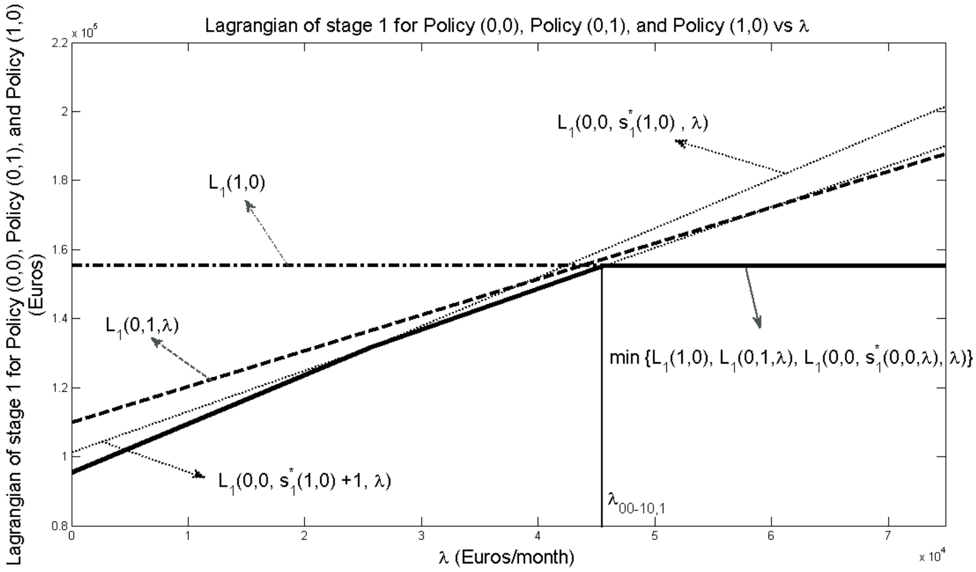

Figure 3.5 Comparison of Policy (0,1), Policy (1,0) and Policy (1,0) for stage 1 in Example 3.5

- If $Nc_{1,i} > c_{0,i} + \frac{h_i}{\alpha}(1 - e^{-\alpha T})$, Policy (0,0) is the optimal policy for either $\lambda \in [0, \lambda_{00-10,i}]$ or $\lambda \in [0, \lambda_{00-01,i}]$ by Lemma 3.5 and Lemma 3.6, respectively. By Lemma 3.4 and Lemma 3.5, it is obvious that Policy (1,0) is optimal as $\lambda \to \infty$. We need to show that there exists cases in which Policy (0,1) is optimal for some $\lambda > 0$ and there exists cases in which it is never optimal. We show that both situations may occur in Example 3.5, which completes the proof.

□

Example 3.5 We continue with the redundancy allocation problem introduced in Example 3.1. In Figure 3.5 and Figure 3.6, we give the comparison of the three policies for stage 1 and stage 2. In Figure 3.5, the functions are the same as those in Figure 3.2, Figure 3.3, and Figure 3.4. In Figure 3.6, the functions are represented for stage 2 as they were represented for stage 1 in those figures. Observe that Policy (0,1) is never optimal for stage 1 (Figure 3.5) while it is optimal for $\lambda \in [\lambda_{00-01,2}, \lambda_{01-10,2}]$, where $\lambda_{00-01,2} = 818238$ Euros/month and $\lambda_{01-10,2} = 3630156$ Euros/month (Figure 3.6).

Theorem 3.1 shows the optimal sequence of the policies followed for stage i for increasing downtime penalty rate (λ) or decreasing resource level of the downtime

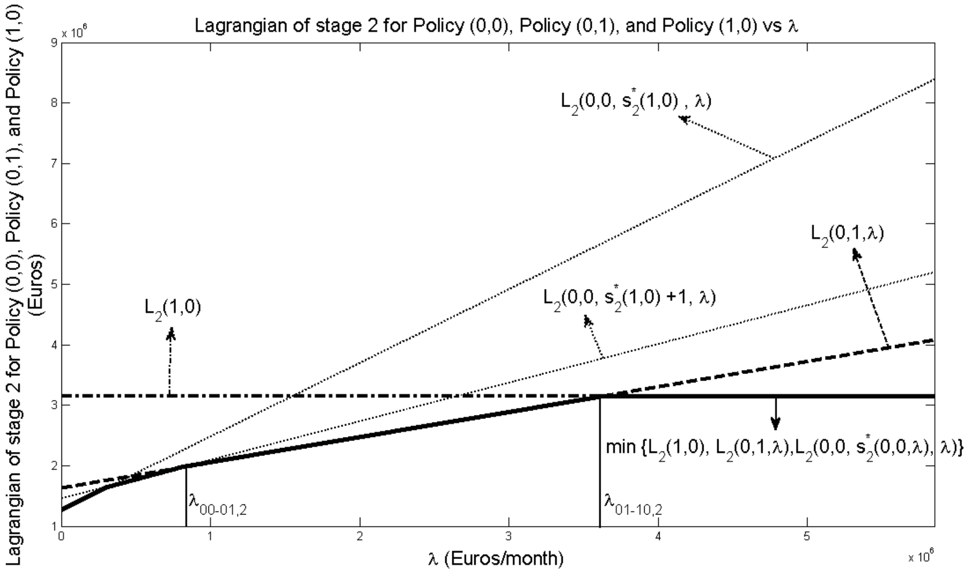

Figure 3.6 Comparison of Policy (0,1), Policy (1,0) and Policy (1,0) for stage 2 in Example 3.5

constraint (D_0) in (Q0). The sequence can be either (0,0)-(1,0) or (0,0)-(0,1)-(1,0). For low downtime penalty rate or high resource level of the downtime constraint in (Q0), Policy (0,0) is optimal. As the penalty rate increases or the resource level decreases, a switch from Policy (0,0) to Policy (0,1) or Policy (1,0) occurs to compensate the increase in the penalty rate by decreasing the downtime or to cope with the tighter downtime constraint. If $Nc_{1,i} \leq c_{0,i} + \frac{h_i}{\alpha}(1 - e^{-\alpha T})$, Policy (0,1) can never be optimal and the switch always occurs to Policy (1,0) at $\lambda_{01-10,i}$. If $Nc_{1,i} > c_{0,i} + \frac{h_i}{\alpha}(1 - e^{-\alpha T})$, the switch is to Policy (0,1) if $\lambda_{00-01,i} < \lambda_{01-10,i}$ (or $\lambda_{00-10,i} < \lambda_{01-10,i}$) and it is to Policy (1,0) otherwise. If the switch occurs to Policy (0,1), Policy (0,1) remains optimal for $\lambda_{00-01,i} < \lambda \leq \lambda_{01-10,i}$ and Policy (1,0) becomes optimal for $\lambda \geq \lambda_{01-10,i}$.

Remember that $\lambda_{01-10,i}$ has a closed form expression, and $\lambda_{00-01,i}$ and $\lambda_{00-10,i}$ can be found by simple numerical procedures. So, the optimal sequence of the policies can be identified for given instances. As our focus is on the redundancy decision, our major interest is in the switching point to Policy (1,0). We denote this point by $\lambda_{10,i}$. Obviously, for $Nc_{1,i} \leq c_{0,i} + \frac{h_i}{\alpha}(1 - e^{-\alpha T})$, $\lambda_{10,i} = \lambda_{00-10,i}$; and for $Nc_{1,i} > c_{0,i} + \frac{h_i}{\alpha}(1 - e^{-\alpha T})$, $\lambda_{10,i} = \max\{\lambda_{00-10,i}, \lambda_{01-10,i}\}$. Corollary 3.3 determines an upper bound and a lower bound for $\lambda_{10,i}$.

Corollary 3.3

$$c_{1,i}\left(\frac{T}{\tau_i}\mu_{2,i}\right)^{-1} \leq \lambda_{10,i} \leq c_{1,i}\left(\frac{T}{\tau_i}\mu_{1,i}\right)^{-1}.$$

Proof: These bounds are immediate results of Corollary 3.2 and Lemma 3.5. □

3.4.3 Results for the Multi-Stage Problem

We will provide two system level results in this subsection. The first one is for the generation of efficient solutions. In our original problem formulation (Q0), we define a specific resource level for the downtime constraint. However, in general, one is interested in exploring the trade-off between the downtime (availability) and minimum TCO rather than in finding the optimum LCC for some given resource level. Efficient solutions provide this exploration.

The second result is for ordering the stages for choosing redundancy (Policy (1,0)). In many cases, there might be other factors which affects the redundancy decision (e.g. a budget limit) and one might be interested in the optimal order to follow for the stages to choose redundancy one-by-one.

Finding Efficient Solutions

The most interesting efficient solutions are those at which a decision changes. That is, the values the downtime and the optimal LCC assume when the optimal decision changes from Policy (0,0) to Policy (0,1) or Policy (1,0), and from Policy (0,1) to Policy (1,0) (see Theorem 3.1) for each stage will constitute the most important elements of the efficient frontier. These points are generated by $\lambda_{00-01,i}$, $\lambda_{00-10,i}$, and $\lambda_{01-10,i}$, which can be determined either directly or by simple numerical procedures. Hence, the most interesting points for the efficient frontier can easily be generated for given instances.

Example 3.6 In Figure 3.7, we give an efficient frontier for the capital good with two stages introduced in Example 3.1. This efficient frontier has the downtime on its x-axis. Figure 3.8 depicts the same efficient frontier with p values on the x-axis. Remember that $p \in (0, 1]$ is the availability measure that reflects the required uptime portion of the NT system-years. The frontier includes the efficient solutions at which switches from one policy to the other occurs for a stage. When Policy (0,0) is chosen for the two stages and initial supply amounts are optimized, expected downtime of the $N = 15$ systems is 2.64 months (Figure 3.7), which is equivalent to an availability

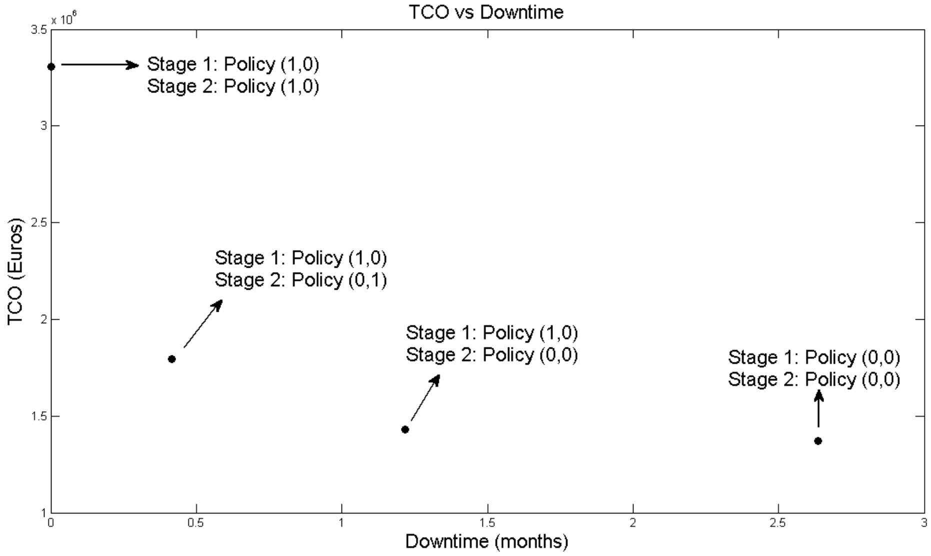

Figure 3.7 Efficient frontier for Example 3.6 with downtime

level of $p = 0.999$ (Figure 3.8), with TCO of 1371004 Euros. The first policy change - the one that brings the most benefit (highest increase in availability per unit increase in TCO) - is the one from Policy (0,0) to Policy (1,0) for stage 2. The next change is from Policy (0,0) to Policy (0,1) for stage 1. Finally, downtime becomes zero ($p = 1$) when Policy (1,0) is chosen for the both stages, as we assume that a subsystem is 100% available when redundancy is implemented.

Notice that in Figure 3.8, the value of p is already significantly high without choosing redundancy for either of the stages. This is due to the small number of stages in the capital good. The total downtime of the systems is equal to the sum of the downtimes of the subsystems which are short as failures occur rarely (MTBF of the subsystems are $\tau_1 = 3$ years and $\tau_2 = 6$ years) and the ordinary and emergency procedures are applied. Total downtime increases (i.e., p decreases) as the number of stages increases. In Figure 3.9, you can see the efficient frontier for a capital good with 60 stages. In this figure, the point displayed as a circle depicts the $(TCO\ , p)$ couple when Policy (0,0) is chosen for all the stages, the points displayed as squares depict $(TCO\ , p)$ couples when a switch from Policy (0,0) to Policy (0,1) occurs for a stage, and the points displayed as stars depict $(TCO\ , p)$ couples when a switch from Policy (0,1) to Policy (1,0) occurs for a stage. As you can observe, the values of p ranges within [0.961 1].

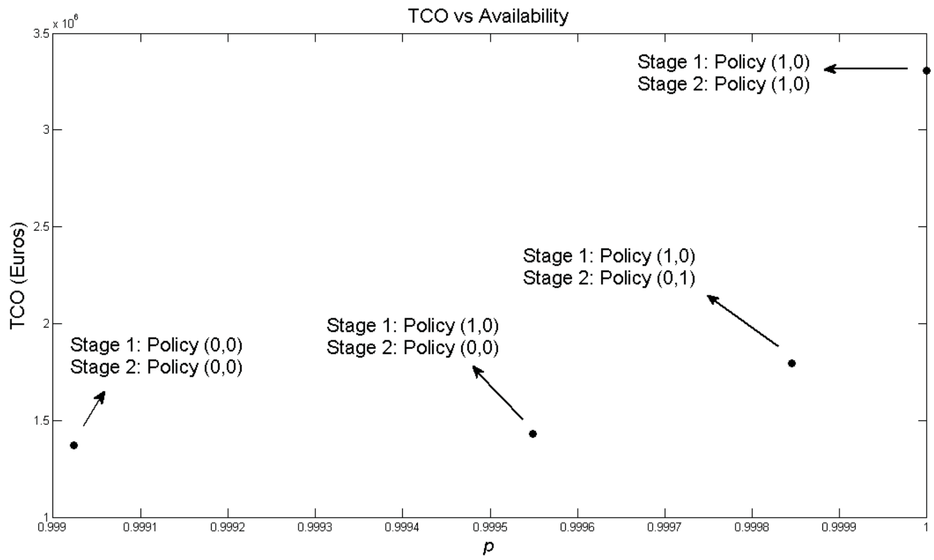

Figure 3.8 Efficient frontier for Example 3.6 with availability

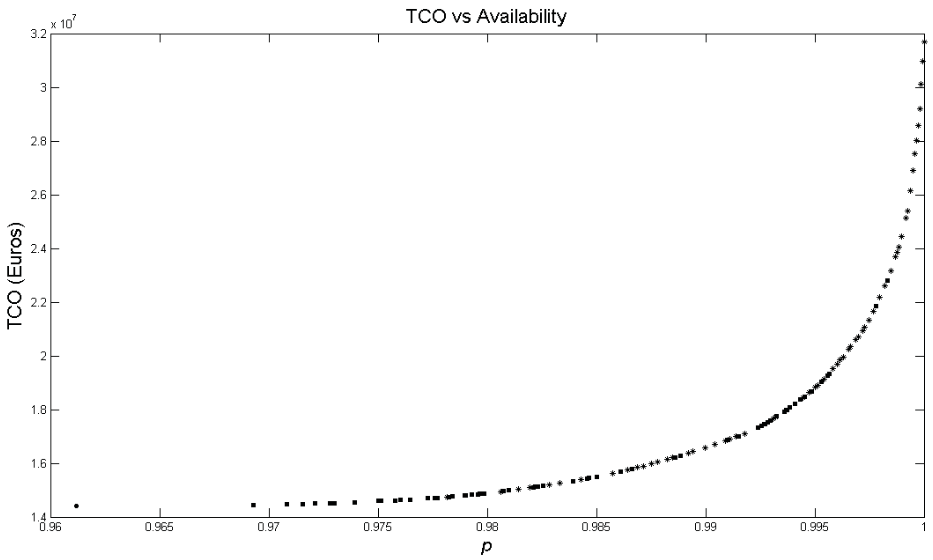

Figure 3.9 Efficient frontier for a capital good with 60 stages

Ordering Stages for Redundancy

We aim at generating an order of the stages such that if one follows this order while choosing redundancy one-by-one for stages, she will pay the least amount (lowest increase in TCO) per unit decrease in the resource level of the downtime constraint (or unit increase in p value) at each step. This ordering can be achieved by ordering the stages with respect to $\lambda_{10,i}$'s, $i \in M$, the values of λ (downtime penalty rate) at which optimal Policy switches to Policy (0,1). The ascending order of $\lambda_{10,i}$'s provides the intended order for the stages.

The procedure for ordering the stages can be stated as follows:

(1) For all $i \in M$ with $Nc_{1,i} \leq c_{0,i} + \frac{h_i}{\alpha}(1 - e^{-\alpha T})$, find $\lambda_i = \lambda_{00-10,i}$. For all $i \in M$ with $Nc_{1,i} > c_{0,i} + \frac{h_i}{\alpha}(1 - e^{-\alpha T})$, find $\lambda_{10,i} = \max\{\lambda_{00-10,i}, \lambda_{01-10,i}\}$.

(2) Generate the permutation $J = (j_1, ..., j_m)$ of set M such that $i < k$ implies that $\lambda_{j_i} \leq \lambda_{j_k}$ for all $i, k \in M$.

For each $i \in M$, λ_i found in step (1) is the value of λ for which Policy (1,0) becomes optimal. By permutation $J = (j_1, ..., j_m)$, an ascending order $\{\lambda_{10,j_1}, \lambda_{10,j_2} ..., \lambda_{10,j_m}\}$ for the set $\{\lambda_{10,1}, \lambda_{10,2} ..., \lambda_{10,m}\}$ can be generated. Remember that the resource level of the downtime constraint in Q(λ), $D(\boldsymbol{y}^*(\lambda), \boldsymbol{z}^*(\lambda), \boldsymbol{s}^*(\lambda))$, is decreasing in λ. Hence, by choosing redundancy for the stages one-by-one with the order in permutation J (i.e., in the order (stage j_1, stage j_2,..., stage j_m)), the optimal TCO values are provided for decreasing resource level of the downtime constraint (i.e., increasing p value) at each step.

Example 3.7 We apply the procedure for ordering the stages of the capital good given in Example 3.1. The relevant λ values are $\lambda_{00-10,1} = 45630.35$ Euros/month, $\lambda_{01-10,1} = 43682.49$ Euros/month, $\lambda_{00-10,2} = 3005896$ Euros/month, and $\lambda_{01-10,2} = 3630156$ Euros/month.

(1) $\lambda_{10,1} = 45630.35$ Euros/month and $\lambda_{10,2} = 3630156$ Euros/month.

(2) $J = (1, 2)$ as $\lambda_{10,1} < \lambda_{10,2}$.

Hence, one should choose first stage 1 and then stage 2 to implement redundancy.

3.5. Conclusions

In this chapter, we developed a redundancy allocation model for capital goods. In the problem that we studied, three policies per stage, namely Policy (0,0), Policy (0,1), and Policy (1,0), were defined. Under Policy (0,0) and Policy (1,0), a defined emergency procedure was applied when a failure occurred and there was an out-of-stock situation. Under Policy (0,1), a defined provision procedure was applied when a failure occurred and the actual stock-on-hand of the relevant component was 1, which prevented out-of-stock situations. Each of these policies provided different levels of uptime. We developed the problem formulation as the minimization of the TCO of a general number of systems under a defined downtime constraint. TCO included acquisition costs, spare parts costs, and repair costs. The multi-stage problem had a combinatorial nature due to the three candidate policies per stage.

We decomposed the problem into single-stage problems and showed that a solution for the multi-stage problem could be generated by finding solutions of each of the single-stage problems. We developed an efficient procedure to find the optimal solutions of the single-stage problems for varying resource levels of the downtime constraint. This procedure enabled us to find the optimal solutions for the multi-stage problem for varying resource levels of the downtime constraint efficiently, without considering all possible combinations of the three policies for all stages.

We derived results for the single-stage problems and the multi-stage problem. The results for the single-stage problems revealed that when the value of the resource level of the downtime constraint was varied by starting from a high value and decreased down to zero; i.e., the constraint was initially loose and got tighter and tighter, Policy (1,0), which corresponded to choosing redundancy, became optimal at a certain value of the resource level and remained optimal for all its smaller values afterwards. We also showed that the values of the TCO and downtime (or uptime) when the optimal policy changed from one to the other could be easily computed. This property lead to a simple method to construct an efficient frontier to explore the trade-off between the downtime and the TCO for the multi-stage problem. We also provided an ordering of the stages in a capital good for redundancy as one might be interested in finding the optimal order to follow for stages to make one-by-one decisions.

In Chapter 2 and this chapter, we investigated optimal reliability of a component and the redundancy allocation for a capital good, respectively. In the next chapter, we will study the upgrading problem for an improved component. In this problem, the reliability level of the component is already fixed and redundancy is not considered.

Appendix

Proof of Property 3.1:

(i) Under Policy (0,0) and Policy (1,0), the processes of instances of the ordinary procedure and the emergency procedure are identical to those defined in Subsection 2.2.1 and this property is equivalent to Lemma 2.1 with $\tau = \tau_i$ and $G(\tau, s) = B_i(s_i)$).

(ii) The property for Policy (0,1) can be proved in a similar manner given for Lemma 2.1.

(iii) In part (ii) of Lemma 2.2, we derive NPV of the expected repair costs. By following a similar manner used in that derivation, it can be shown that

$$R_i(0, s_i) = \frac{N}{\alpha\tau_i}(1 - e^{-\alpha T})r_{1,i} + (r_{2,i} - r_{1,i})B_i(s_i),$$

and

$$R_i(1, s_i) = \frac{N}{\alpha\tau_i}(1 - e^{-\alpha T})r_{1,i} + (r_{2,i} - r_{1,i})B_i(s_i - 1.)]$$

The two formulations can be represented by equation (3.2).

(iv) For a given $i \in M$, suppose that Policy (0,0) is chosen $((y_i, z_i) = (0, 0))$. Let Z_1 be the random variable denoting the downtime after a failure of a component-i part that leads to an instance of the ordinary procedure. Remember that this downtime is equivalent to the inventory-replacement time of the failed part, and inventory-replacement times have independent and identical distributions with mean $\mu_{1,i}$. Let F_1 denote the random variable representing the number of failures of component-i parts that lead to applications of the ordinary procedure throughout $[0, T]$. Then, $E[F_1] = \frac{NT}{\tau_i}[1 - B_i(s_i)]$ by part (i). Let $D_{1,i}$ be the expected downtime stemming from those failures. Then,

$$D_{1,i} = E[Z_1]E[F_1] = \frac{NT}{\tau_i}\mu_{1,i}[1 - B_i(s_i)]$$

Similarly, the expected downtime stemming from failures of component-i parts that lead to applications of the emergency procedure throughout $[0, T]$ can be derived as

$$D_{2,i} = \frac{NT}{\tau_i}\mu_{2,i}B_i(s_i)$$

Hence, for Policy (0,0), the expected downtime throughout $[0, T]$ is

$$D_i(0, 0, s_i) = \frac{NT}{\tau_i}[\mu_{1,i} + (\mu_{2,i} - \mu_{1,i})B_i(s_i)].$$

For Policy (0,1) ($(y_i, z_i) = (0,1)$), downtime per failure is equal to the inventory-replacement time and the expected downtime can be derived similarly as

$$D_i(0, 1, s_i) = \frac{NT}{\tau_i}\mu_{1,i}.$$

As there is no downtime in Policy (1,0) ($(y_i, z_i) = (1,0)$), the expected downtime throughout $[0, T]$ can be written as

$$D_i(y_i, z_i, s_i) = \frac{NT}{\tau_i}(1 - y_i)[\mu_{1,i} + (\mu_{2,i} - \mu_{1,i})(1 - z_i)B_i(s_i)].$$

Chapter 4

Upgrading Policy After Redesign of a Component

4.1. Introduction

In Chapter 2 and 3, we studied reliability decisions during the design phase of a capital good. In practice, reliability improvement activities are also performed during the exploitation phase. In this chapter, we study the upgrading policy problem, which is encountered during the exploitation phase.

We consider a situation in which an OEM is responsible for the availability of a general number N of systems in the field through service contracts. She has already decided that it is economical to improve the reliability of a certain critical component by redesign and upgrade the systems in the field by replacing the parts in the field (old parts) with the redesigned parts (new parts). The question to be investigated is when the upgrading should take place: immediately when the component has been redesigned or old parts are replaced only when they fail. Contrary to the cases that we studied in Chapter 2 and 3, repair-on-site is applied for the critical component and no spare parts are kept on stock.

In general, an OEM and a supplier of the new parts might agree on different terms for the supply of the new parts, such as one-for-one replenishment, replenishment in batches, unit price, etc. We assume the following setting: The OEM can buy any number of new parts just after the redesign (at time 0) and she can replenish new parts only in batches after time 0. The OEM and the supplier agree on a fixed batch size and unit price(s) of the new parts through negotiations. The unit price after time

0 is greater than or equal to the unit price at time 0. This is a very likely situation as the production facility of the supplier might undergo some changes after time 0 (e.g., the production line or the technology might change) and an extra effort might be necessary to produce the new parts.

The OEM considers the following two upgrading policies for the N systems in the field:

- Policy 1 - Upgrade all systems preventively at time 0: N new parts are bought at time 0 and all the old parts in the field are preventively replaced with the new parts at time 0.

- Policy 2 - Upgrade systems one-by-one correctively: A number of new parts is bought at time 0 (initial supply) and is kept on stock. When an old component in the field fails, it is correctively replaced with a new one from the inventory. The OEM replenishes new parts in batches whenever a new part is needed and there is an out-of-stock situation after time 0.

Under Policy 1, the OEM faces less failures and less downtime as all old parts are replaced with the new ones immediately after the redesign. However, she forfeits the remaining lifetimes of the old parts. Under Policy 2, OEM benefits from the remaining lifetimes; however, she faces more failures and downtime. An increase in the unit price after time 0 (which is probable as we stated above) favors Policy 1. All factors that play a role in Policy 1 are predetermined. The initial supply quantity is a decision that the OEM has to make and it affects the costs incurred under Policy 2. All other factors in Policy 2 are predetermined.

In this chapter, we develop a quantitative model for the upgrading problem with Policy 1 and Policy 2. We formulate the effect of the initial supply quantity on procurement costs, inventory storage costs, replenishment costs and salvage value under Policy 2 explicitly. Although these policies were previously studied in the literature (see Mercier and Labeau (2004), Mercier (2008), and Clavareau and Labeau (2009b,a)), the existing models did not incorporate these effects which is fundamental in our case.

Mercier and Labeau (2004) investigate a situation in which the failure rate and the energy consumption rate of a unit (part or system) are improved. They define a so-called K *strategy* for a general number, N, of identical and independent units (parts or systems) on some finite time interval $[0, T]$. The upgrading period is separated into two phases under this strategy. Until the K^{th} failure of the old units, $K \in \{0, 1, ..., N\}$, failed old units (including the K^{th} unit) are replaced with the new ones correctively. Afterwards, the remaining $N-K$ old units are replaced preventively. They investigate situations in which failure rates for both old units and new units are constant. $K = 0$

represents the strategy under which all old units are replaced preventively at time 0 while $K = N$ represents the strategy under which each old unit is replaced correctively (no preventive replacement). Notice that $K = 0$ corresponds to the same as Policy 1 in our model; and $K = N$ results in the same policy as Policy 2 with an initial supply quantity of zero and a batch size of 1 for replenishment after time 0. When a new unit fails, it is replaced with another new unit with zero lead time. They formulate the mean discounted total cost over $[0, T]$ at time 0. The mean discounted total cost includes replacement costs and energy consumption costs. They show that only three strategies can be optimal: the strategies with $K = 0$, $K = 1$, and $K = N$, respectively.

Mercier (2008) incorporates general failure rates (e.g., increasing failure rate due to degradation) into the model developed by Mercier and Labeau (2004). They show that in this case the optimal strategy can be different than $K = 0$, $K = 1$, and $K = N$; and it depends on the time horizon T.

The inventory of new parts, which is a crucial aspect in our case, is not incorporated in the models by Mercier and Labeau (2004) and Mercier (2008). Clavareau and Labeau (2009b,a) incorporate the inventory into a Petri net model and a simulation model, respectively, to study the K *strategy*. In their model, an order quantity is determined and new parts can be procured at any time with this order quantity. However, the order quantity is fixed by certain inventory control methods, such as point command method and Economic Order Quantity rather than being optimized with respect to the total costs incurred for upgrading the systems in these papers. These models also incorporate other aspects (e.g., different types of maintenance actions, limited maintenance capacity, priority rules for different actions, effectiveness of a repair, etc.) which makes it difficult to realize the interaction between the inventory decisions and the optimal strategy. In our model, the initial supply quantity under Policy 2 is a decision variable. We formulate the relevant costs as a function of the initial supply quantity and optimize the initial supply quantity with respect to total costs.

We conducted a numerical study to derive insights about conditions which favor each policy. We used the percentage difference in the MTBF of the old parts and the MTBF of the new parts as a measure of the reliability improvement. In our numerical study, we found out that Policy 1 is favored by low values of the number of systems, long lifetime of the systems, low values of the MTBF of the old parts (for fixed percentage improvement in MTBF), high values of the percentage improvement in MTBF, high values of the increase in the unit price of the new parts after time 0, large batch sizes, and high values of the downtime costs per failures. The reverse of each of these conditions favors Policy 2. Our numerical study showed that the optimal policy may change by varying any of the mentioned factors.

The contribution of this chapter can be stated as follows:

- First, we introduce a model for the upgrading problem with Policy 1 and Policy 2 for a general number of systems. We formulate total costs incurred under Policy 1 and Policy 2. These costs include procurement costs of the new parts, costs incurred for upgrading the systems and costs incurred during repairs of the new parts under Policy 1; and costs of the initial supply, costs incurred for upgrading the systems, repair costs incurred during repairs of the new parts, replenishment costs after time 0 and inventory storage costs under Policy 2. Downtime costs are incorporated into the costs incurred for upgrading the systems and the costs incurred during repairs of the new parts. We develop a problem formulation in which the relationship between the initial supply quantity and the costs affected by the initial supply quantity under Policy 2 is explicitly established.
- Second, we perform an exact analysis on the total costs under Policy 2 and we derive several analytical properties.
- Third, we develop an efficient solution procedure for the initial supply quantity in Policy 2.
- Fourth, we perform a numerical study and provide insights about conditions which favor each policy. We use the percentage difference in the MTBF of the old parts and the MTBF of the new parts as a measure of the reliability improvement. Policy 1 is advantageous for

  - low values of the number of systems,
  - long lifetime of the systems,
  - low values of the MTBF of the old parts (for fixed percentage improvement in MTBF),
  - high values of the percentage improvement in MTBF,
  - high values of the increase in the unit price of the new parts after time 0,
  - large batch sizes,
  - high values of the downtime costs per failures.

  Policy 2 is advantageous for

  - high values of the number of systems,
  - short lifetime of the systems,
  - high values of the MTBF of the old parts (for fixed percentage improvement in MTBF),
  - low values of the percentage improvement in MTBF,

– low values of the increase in the unit price of the new parts after time 0,
– small batch sizes,
– low values of the downtime costs per failures.

Our numerical study shows that varying any of the mentioned factors may lead to a change in the optimal policy.

The outline of this chapter is as follows. In Section 4.2, we detail our model and develop a problem formulation. We derive the total cost function per policy and provide a number of analytical properties and an optimization procedure for the total cost function of Policy 2 in Section 4.3. We give the setting and the results of our numerical study in Section 4.4. We finalize the chapter by drawing conclusions in Section 4.5.

4.2. Model

An OEM provides service for N identical systems that she produced and sold with a service contract. The contract covers the systems' lifetime and we assume that the lifetime of all systems will end at the same time.

The OEM redesigns one of the critical repairable components as she realizes that it is economical to improve its reliability and upgrade the systems. Each system has a single unit of the critical component. The OEM starts to upgrade the systems at time 0 (just after the redesign) and the lifetime of the systems ends at time T; i.e., the new parts can be used throughout the time interval $[0, T]$. The remaining lifetime of the systems (T) is in the order of 1-30 years.

We denote the Mean Time Between Failures (MTBF) of the old parts by τ_{old}, which is in the order of 1-10 years. After the redesign, an MTBF of $\tau_{new} > \tau_{old}$ is achieved for the new parts. We assume that the time to failure of each old part and time to failure of each new part in the field have exponential distributions with mean τ_{old} and τ_{new}, respectively, throughout $[0, T]$. We further assume that τ_{old} and τ_{new} are fixed throughout $[0, T]$.

The OEM has two options for the procurement of new parts: First, she can buy any number of new parts with a unit price of c_0 at time 0. Second, she can replenish batches of q_1 new parts throughout $(0, T]$ with a unit price of c_1. The batch size q_1 ($q_1 \in \{1, 2, ...N\}$) is a fixed value that the OEM and the supplier of the new parts agree on. The OEM may choose to keep an inventory of the new parts for upgrading the systems. The storage cost rate per part is $h > 0$.

There are two types of upgrading per system:

- Preventive Upgrading: Planning and executing the replacement of the old part in a system before the old part fails.
- Corrective Upgrading: Replacing the old part when it fails.

In both cases the following actions are taken. A new part is transported to the customer site; a service engineer visits the customer site and replaces the old part with the new one; the old part is transported to a disposal site and discarded. The expected costs that are incurred for a preventive upgrading and a corrective upgrading are u_1 and u_2, respectively. These costs include administrative costs, costs of a visit of a service engineer, the downtime costs stemming from the interruption of operation during the upgrading, transportation costs of the new part and the old part. As preventive upgrading is planned beforehand, the downtime costs and the costs of the visit of a service engineer incurred during a preventive upgrading are less than that incurred during a corrective upgrading. All the actions that are taken in both cases are the same; thus, the difference between the upgrading costs stem from the difference between the downtime costs and the service engineer costs; and $u_1 \leq u_2$.

When a new part fails in the field, a service engineer visits the customer site and repairs the part. That is, the failed part is not replaced with a ready-for-use one but repaired on site. If the OEM keeps an inventory of the new parts, the parts in stock are used only for upgrading the systems. We denote the expected costs incurred during such an on-site repair by r. r includes administrative costs, costs of a visit of a service engineer, downtime costs, and the repair costs of a new part.

The salvage value of an old part and that of a new part are s_{old} and s_{new}, respectively. s_{old} and s_{new} can be positive or negative. Positive values mean that revenue is generated and negative values mean that discarding costs are incurred by salvaging parts. We assume that $s_{new} \leq c_0$; that is, the new parts cannot be salvaged at a higher value than their unit price. All new parts are salvaged at time T. An old part is salvaged immediately after it is replaced with a new one.

4.2.1 Policies

The OEM chooses one of the following two upgrading policies after the redesign:

1. Policy 1 - Upgrade all systems preventively at time 0: N new parts are replenished and all old parts in the field are replaced with new parts at time 0.

2. Policy 2 - Upgrade systems one-by-one correctively: q_0 ($q_0 \in \{0, 1, 2, ...N\}$) new parts are replenished at time 0. q_0 is a decision variable. After their replenishment, the parts are kept in an inventory. When an old part fails, it is replaced by a new part from the inventory as long as there is a ready-for-use part available. After the initial supply, whenever a failure of an old part occurs and there is an out-of-stock situation in the inventory of the new parts, a batch of new parts is replenished. After each replenishment, one of the q_1 new parts is used to replace the failed old part which triggered the replenishment. The remaining $q_1 - 1$ new parts are kept in the inventory for later use to replace the old parts when they fail.

The OEM's decision about the implementation policy depends on the total costs that would be incurred throughout $[0, T]$. Under Policy 1, the total costs would be the sum of the investment costs of N new parts, costs incurred for upgrading the systems (including downtime costs), costs incurred during repairs of the new parts (including downtime costs), and salvage value. Under Policy 2, the total costs would be the sum of the costs of initial supply, costs incurred for upgrading the systems, inventory costs, replenishment costs that would be incurred after the usage of the first q_0 new parts, costs incurred during repairs of the new parts, and salvage value. We assume that the production costs, costs incurred for upgrading the systems, and salvage value of the old parts under Policy 1 and initial replenishment costs of q_0 parts under Policy 2 are incurred at time 0. As the other costs are incurred throughout $[0, T]$, we formulate the Net Present Values (NPVs) of these costs at time 0. We denote the discount rate by $\alpha > 0$.

We assume that the replenishment lead time, which plays a role only under Policy 2, is zero. In case of a positive replenishment lead time (in the order of 1-2 weeks), the OEM would keep a small safety stock; that is, she would order a new batch whenever a failure occurred and the actual inventory-on-hand decreased to a certain positive level. The safety stock would serve to replace old parts that failed during the lead time; so, the OEM would decide on the safety stock level with respect to the probability of having failures during the lead time. As the lead time would be small compared to the MTBF of the old parts (τ_{old}), the mean number of failures during the lead time would also be small, which would impose a small safety stock level and safety stock costs. Hence, the effect of a positive lead time on the optimal initial order quantity (q_0) and total costs would be very small. Ignoring a possible small replenishment lead time has no or a very limited effect on the costs under Policy 1 and 2, and it keeps the analysis clean.

4.2.2 Problem Formulation

We use the following notation for the costs incurred under the two policies:

π_1:	The expected NPV of the total costs under Policy 1.
$\pi_2(q_0)$:	The expected NPV of the total costs under Policy 2.
P_1:	The expected NPV of the procurement costs of N new parts under Policy 1.
V_1:	The expected NPV of the salvage value under Policy 1.
U_1:	The expected NPV of the costs incurred for upgrading the systems under Policy 1.
R_1:	The expected NPV of the costs incurred during repairs of the new parts under Policy 1.
$P_2(q_0)$:	The expected NPV of the procurement costs of the initial supply under Policy 2.
$S_2(q_0)$:	The expected NPV of the inventory storage costs under Policy 2.
$K_2(q_0)$:	The expected NPV of the replenishment costs incurred after time 0 under Policy 2.
$V_2(q_0)$:	The expected NPV of the salvage value under Policy 1.
U_2:	The expected NPV of the costs incurred for upgrading the systems under Policy 2.
R_2:	The expected NPV of the costs incurred during repairs of the new parts under Policy 2.

There is no decision variable in Policy 1, so the costs do not have a functional form. $\pi_1 = P_1 - V_1 + U_1 + R_1$, where $P_1 = c_0 N$, $V_1 = s_{old} N + s_{new} N e^{-\alpha T}$ and $U_1 = u_1 N$. The derivation of R_1 will be given in Section 4.3.

Under Policy 2, the old parts are replaced with the new ones whenever they fail, thus the upgrading costs U_2 is independent of q_0. The failure process of the new parts do not depend on q_0; therefore, the repair costs R_2 are not a function of q_0 either. The other costs are functions of q_0. For Policy 2, the following optimization problem has to be solved:

$$
\text{(Q)} \qquad \begin{array}{ll} \min & \pi_2(q_0) \\ \text{s.t.} & q_0 \in M = \{0, 1, 2, ..., N\}, \end{array}
$$

where $\pi_2(q_0) = P_2(q_0) + S_2(q_0) + K_2(q_0) - V_2(q_0) + U_2 + R_2$ and $P_2(q_0) = c_0 q_0$. The formulations of the other costs will be derived in Section 4.3.

Let q_0^* be an optimal solution of problem (Q) and $\pi_2^* = \pi_2(q_0^*)$; i.e., π_2^* is the minimum cost found for problem (Q). The OEM compares π_1 and π_2^*, and selects the policy with the minimum cost. That is, if $\pi_1 \leq \pi_2^*$, she will implement Policy 1; otherwise, she will implement Policy 2 with initial supply amount of q_0^*.

4.3. Analysis

In this section, we first derive the costs incurred during repairs of the new parts under Policy 1. Then, we provide a number of preliminary results for Policy 2. Next, we derive inventory storage costs, replenishment costs, salvage value, costs incurred for upgrading the systems, and costs incurred during repairs of the new parts under Policy 2. We finalize the section by developing a solution procedure for Problem (Q).

Lemma 4.1 The costs incurred during repairs of the new parts under Policy 1 has the following formulation:

$$R_1 = \frac{N}{\tau_{new}} \frac{r}{\alpha}(1 - e^{-\alpha T})$$

Proof: See Appendix at the end of this chapter. □

So

$$\begin{aligned}
\pi_1 &= P_1 - V_1 + U_1 + R_1 \\
&= c_0 N - (s_{old} N + s_{new} N e^{-\alpha T}) + u_1 N + \frac{N}{\tau_{new}} \frac{r}{\alpha}(1 - e^{-\alpha T}).
\end{aligned}$$

4.3.1 Preliminary Results for Policy 2

Let $I(t)$ represent the actual inventory-on-hand at time $t > 0$ under Policy 2. For a given initial order quantity (q_0), $I(t)$ follows a certain cyclic pattern. In Figure 4.1, we illustrate this pattern for $q_0 = 4$ and $q_1 = 3$. In this figure, $(t_n)_{n \in \{1,2,\ldots,N\}}$, represents the times of an arbitrary realization of the failures of the old parts. We consider the procurement of the q_0 new parts at time 0 as batch replenishment and refer to this batch as the 0^{th} batch. Observe that after the procurement of any batch, $I(t)$ is depleted in the same manner, which depends on the times of the failures of the old parts after the procurement.

In the remainder of this section, we first construct a mini-model as an analogue of an arbitrary cycle of $I(t)$. We derive generic results for the mini-model, which we will use to characterize failure times of the old parts and derive storage costs under Policy 2. Next, we give the probability density and the probability distribution functions of

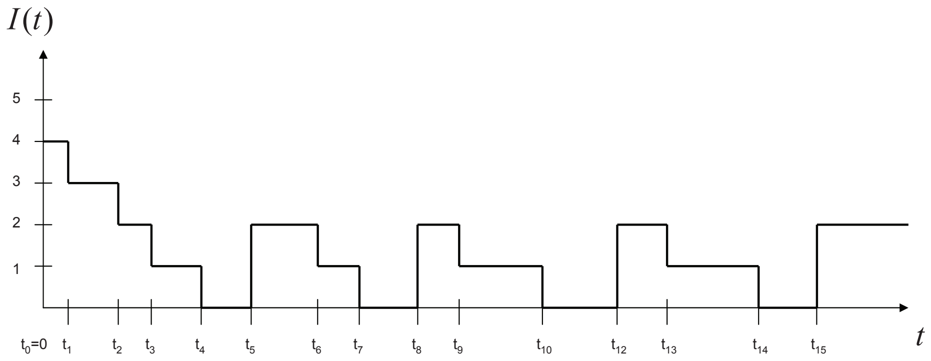

Figure 4.1 Visual representation for of the inventory-on-hand

the failure times of the old parts and two properties regarding replenishment of the new parts.

Mini-model and Generic Results

Suppose that there are $\hat{N}$ systems with old parts in the field at time 0. The systems will be used until time $\hat{T}$. $\hat{q} \in \{0, 1, 2, ..., \hat{N}\}$ new parts are procured and added to stock at time 0. These new parts will be used to upgrade the systems correctively. We denote the random variable that represents the time of the n^{th} failure of the old parts under infinite horizon by $\hat{T}_n$, $n \in \{0, 1, 2, ..., \hat{N}\}$, $\hat{T}_0 = 0$. That is, $\hat{T}_n$ is the time of the n^{th} event in the process of the total stream of failures of the old parts. Under finite horizon ($\hat{T} < \infty$), the old part which fails at time $\hat{T}_n \leq \hat{T}$ is replaced with a new part. If $\hat{T}_n > \hat{T}$, the old part is not replaced. All the other parameters are the same as the ones in the original model. We do not consider replenishment of new batches if the number of the failures of the old parts is larger than q until time $\hat{T}$, as we focus on the storage costs of the batch which is procured at time 0.

In Property 4.1, we derive the probability density and probability distribution functions of $\hat{T}_n$, which we denote by $f_n(t, \hat{N})$ and $F_n(t, \hat{N})$. Including $\hat{N}$ as an independent variable in the definitions of these functions enables us to use these functions to formulate the probability density and probability distribution functions of the times of failures of the old parts in the original model.

Property 4.1 $\hat{T}_n$, $n \in \{1, 2, ..., \hat{N}\}$, has the probability density function

$$f_n(t, \hat{N}) = \sum_{i=1}^{n} C_{i,n}(\hat{N}) \lambda_i(\hat{N}) e^{-\lambda_i(\hat{N})t}, \; t \geq 0, \tag{4.1}$$

and the probability distribution function

$$F_n(t,\hat{N}) = \sum_{i=1}^{n} C_{i,n}(\hat{N})(1-e^{-\lambda_i(\hat{N})t}),\ t \geq 0,$$

where $\lambda_i(\hat{N}) = \frac{\hat{N}-i+1}{\tau_0}$, $C_{1,1}(\hat{N}) = 1$, and $C_{i,n}(\hat{N}) = \prod_{\substack{j=1 \\ j\neq i}}^{n} \frac{\lambda_j(\hat{N})}{\lambda_j(\hat{N})-\lambda_i(\hat{N})}$ for $n \in \{2,3,...,\hat{N}\}$.

Proof: Let $\hat{X}_i$ be the random variable that represents the time between the $(i-1)^{st}$ failure and i^{th} failure, for $i \in \{1,2,3,...,\hat{N}\}$. Then, $\hat{X}_i$ is exponentially distributed with rate $\hat{\lambda}_i(\hat{N}) = \frac{\hat{N}-i+1}{\tau_0}$ and $\hat{T}_n = \sum_{i=1}^{n} \hat{X}_i$, $n \in \{1,2,3,...,\hat{N}\}$; that is, $\hat{T}_n$ is the sum of n exponential random variables with different rates. Such a random variable is said to be a hypoexponential random variable and its probability density function is given by equation (4.1) (see equation (5.8) in Ross (2007)). □

Let $\hat{S}(\hat{N},\hat{T},\hat{q})$ denote the expected NPV of the storage costs of the batch at time 0. We derive $\hat{S}(\hat{N},\hat{T},\hat{q})$ in Lemma 4.2. The storage costs of each cycle (batch) in the original model can be derived by varying $\hat{N}$, $\hat{T}$, and $\hat{q}$ with respect to the parameters of the cycle.

Lemma 4.2 It holds that

$$\hat{S}(\hat{N},\hat{T},\hat{q}) = h \sum_{n=1}^{\hat{q}} \sum_{i=1}^{n} C_{i,n}(\hat{N}) \frac{1}{\alpha+\lambda_i(\hat{N})} \left[1-e^{-\left(\alpha+\lambda_i(\hat{N})\right)\hat{T}}\right].$$

Proof: See Appendix at the end of this chapter. □

Failure Times of the Old Parts and Replenishment After Time 0

As we mentioned in Subsection 4.2.2, the costs under Policy 2 depend on the times of the failures of the old parts. Hence, for their derivation, we need to characterize these failure times. Let T_n, $n \in M$, denote the random variable that represents the time of the n^{th} failure of the old parts under infinite horizon under Policy 2, where $T_0 = 0$. Under finite horizon, the old part which fails at time $T_n \leq T$ is replaced with a new part. We refer to the failed old part as the n^{th} *old part* and the new part which is used to replace the old part as the n^{th} *new part*. We denote the probability density and the probability distribution functions of T_n by $f_{T_n}(t)$ and $F_{T_n}(t)$, respectively. We derive these functions as as direct results of Property 4.1.

Corollary 4.1 For $n \in \{1, 2, ..., N\}$, it holds that $f_{T_n}(t) = f_n(t, N)$ and $F_{T_n}(t) = F_n(t, N)$.

We give two straightforward properties about batch replenishment after time 0 in Property 4.2. We make use of these properties in the derivation of the cost functions.

Property 4.2 (i) The maximum number of batches that can be replenished is $\bar{k}(q_0) = \left\lceil \frac{N-q_0}{q_1} \right\rceil$.

(ii) For each $k \in \{1, 2, ..., \bar{k}(q_0)\}$, if $T_{q_0+(k-1)q_1+1} \leq T$, the k^{th} batch is replenished at time $T_{q_0+(k-1)q_1+1}$ and $N_k(q_0) = N - [q_0 + (k-1)q_1 + 1]$ old parts remain in the field just after the replenishment.

4.3.2 Analysis of the Cost Functions Under Policy 2

We derive the cost functions under Policy 2 in Lemma 4.3. We define the following notation for the storage costs and the salvage value under Policy 2, which helps us in the derivation of these costs:

$S_{2,k}(q_0)$:	The expected NPV of the storage costs of the k^{th} batch, $k \in \{0, 1, 2, ..., \bar{k}(q_0)\}$, under Policy 2.
$V_{new,k}(q_0)$:	The expected NPV of the salvage value received for the k^{th} batch of the new parts under Policy 2.
V_{old}:	The expected NPV of the salvage value of the old parts under Policy 2.

Lemma 4.3 It holds that:

(i)

$$S_2(q_0) = \sum_{k=0}^{\bar{k}(q_0)} S_{2,k}(q_0),$$

where

$$S_{2,0}(q_0) = \hat{S}(N, T, q_0),$$

and for $k \in \{1, 2, ..., \bar{k}(q_0)\}$,

$$\begin{aligned} S_{2,k}(q_0) &= \sum_{n=1}^{q_1-1}\sum_{i=1}^{n} C_{i,n}(N_k(q_0))\frac{1}{\alpha+\lambda_i(N_k(q_0))}\times \\ &\quad \sum_{j=1}^{q_0+(k-1)q_1+1} C_{j,q_0+(k-1)q_1+1}(N)\lambda_j(N)\times \\ &\quad \left[\frac{1-e^{-\left(\alpha+\lambda_j(N)\right)T}}{\alpha+\lambda_j(N)} + \frac{e^{-\left(\alpha+\lambda_j(N)\right)T} - e^{-\left(\alpha+\lambda_i(N_k(q_0))\right)T}}{\lambda_j(N)-\lambda_i(N_k(q_0))}\right]. \end{aligned}$$

(ii)

$$\begin{aligned} K_2(q_0) &= c_1 q_1 \sum_{k=1}^{\bar{k}(q_0)} \sum_{i=1}^{q_0+(k-1)q_1+1} C_{i,q_0+(k-1)q_1+1}(N)\times \\ &\quad \frac{\lambda_i(N)}{\alpha+\lambda_i(N)}\left[1-e^{-\left(\alpha+\lambda_i(N)\right)T}\right]. \end{aligned}$$

(iii)

$$V_2(q_0) = V_{old} + \sum_{k=0}^{\bar{k}(q_0)} V_{new,k}(q_0),$$

where

$$V_{old} = s_{old}\sum_{n=1}^{N}\sum_{i=1}^{n} C_{i,n}(N)\frac{1}{\alpha+\lambda_i(N)}\left[\lambda_i(N)+\alpha e^{-\left(\alpha+\lambda_i(N)\right)T}\right], \tag{4.2}$$

$$V_{new,0}(q_0) = s_{new}e^{-\alpha T}q_0,$$

and for $k \in \{1, 2, ..., \bar{k}(q_0)\}$,

$$V_{new,k}(q_0) = s_{new}q_1 \sum_{i=1}^{q_0+(k-1)q_1+1} C_{i,q_0+(k-1)q_1+1}(N)\left[e^{-\alpha T} - e^{-\left(\alpha+\lambda_i(N)\right)T}\right].$$

(iv)

$$U_2 = u_2\sum_{n=1}^{N}\sum_{i=1}^{n} C_{i,n}(N)\frac{\lambda_i(N)}{\alpha+\lambda_i(N)}\left[1-e^{-\left(\alpha+\lambda_i(N)\right)T}\right].$$

(v)

$$
\begin{aligned}
R_2 &= \frac{r}{\alpha\tau_1}\sum_{n=1}^{N}\sum_{i=1}^{n} C_{i,n}(N) \times \\
&\left\{\frac{\lambda_i(N)}{\alpha+\lambda_i(N)}\left[1-e^{-\left(\alpha+\lambda_i(N)\right)T}\right]-e^{-\alpha T}(1-e^{-\lambda_i(N)T})\right\}.
\end{aligned}
$$

Proof: See Appendix at the end of this chapter. □

4.3.3 Solution Procedure for Problem (Q)

We derive a convexity-like property for $S_2(q_0)$ and $K_2(q_0) - V_2(q_0)$ on the elements of a certain partition of $\{1, 2, ..., N\}$ in Lemma 4.4. Lemma 4.5 makes use of these properties to derive a similar property for $\pi_2(q_0)$. Finally, we develop a solution procedure for Problem (Q) in Theorem 4.1 by exploiting the result given in Lemma 4.5.

Remember that $q_0 \in M = \{0, 1, 2, ..., N\}$. For a given $q_1 \in \{1, ..., N\}$ and $x \in \{0, 1, ..., q_1 - 1\}$, let

$$M_x = \left\{x + zq_1 \mid z = 0, 1, ..., \left\lfloor \frac{N-x}{q_1} \right\rfloor \right\},$$

and

$$\bar{P} = \{M_x \mid x = 0, 1, ..., q_1 - 1\}.$$

Observe that $M_x \cap M_y = \emptyset$ for any $x,\, y \in \{0, 1, ..., q_1 - 1\}$, $x \neq y$ and $\bigcup_{x=0}^{q_1-1} M_x = M$; that is, $\bar{P}$ a partition of M.

Let

$$\theta_2(q_0) = K_2(q_0) - V_2(q_0)$$

We define the operator Δ_{q_1} for an arbitrary function $g(q_0)$ as

$$\Delta_{q_1} g(q_0) = g(q_0 + q_1) - g(q_0).$$

Lemma 4.4 For each $M_x \in \bar{P}$,

(i)

$$\Delta_{q_1} S_2(q_0 + q_1) \geq \Delta_{q_1} S_2(q_0),\ q_0 \in M_x, \tag{4.3}$$

(ii)

$$\Delta_{q_1}\theta_2(q_0+q_1) \geq \Delta_{q_1}\theta(q_0),\ q_0 \in M_x. \tag{4.4}$$

Proof: See Appendix at the end of this chapter. □

Lemma 4.4 states a convexity-like property for $S_2(q_0)$ and $\theta_2(q_0)$ on each element of $\bar{P}$. For any $M_x \in \bar{P}$, $x \in \{0, 1, ..., q_1 - 1\}$, when the elements of M_x are put in an ascending order, two consecutive elements can be represented by q_0 and $q_0 + q_1$. Inequality (4.3) indicates that the difference between the values of $S_2(q_0)$ at two consecutive elements is increasing, which is equivalent to the definition of convexity when the two consecutive elements differ by one. The same argument also holds for inequality (4.4).

Lemma 4.5 For each $M_x \in \bar{P}$,

$$\Delta_{q_1}\pi_2(q_0+q_1) \geq \Delta_{q_1}\pi_2(q_0) \tag{4.5}$$

on M_x:

Proof: It directly follows from Lemma 4.4 as $\pi_2(q_0) = P_2(q_0) + S_2(q_0) + K_2(q_0) - V_2(q_0) + U_2 + R_2$, $P_2(q_0) = c_0 q_0$, and U_2 and R_2 are constant. □

The relation of Lemma 4.4 to convexity also holds for Lemma 4.5.

Corollary 4.2 If $q_1 = 1$ (one-for-one replenishment), $\pi_2(q_0)$ is convex on M.

Proof: If $q_1 = 1$, $\bar{P}$ has only one element which is M itself. Then, Lemma 4.5 is equivalent to stating that $\pi_2(q_0)$ is convex on M. □

In Figure 1.1, we illustrate the property given in Lemma 4.5 for $q_1 = 2$ and $N = 13$ (i.e., $M = \{0, 1, 2, ..., 13\}$). As you can see in Figure 4.2(a), $\pi_2(q_0)$ is not convex on M. The partition $\bar{P}$ is given as $\bar{P} = \{M_0, M_1\}$, where

$$M_0 = \{0, 2, 4, 6, 8, 10, 12\},$$

$$M_1 = \{1, 3, 5, 7, 9, 11, 13\}.$$

In Figure 4.2(b), you can observe that $\pi_2(q_0)$ has a convex shape on M_0 and M_1.

For a given $x \in \{0, 1, ..., q_1 - 1\}$, we define

$$q_{0,x} = \min\{q_0 \in M_x | \pi_2(q_0) \leq \pi_2(q) \text{ for all } q \in M_x\}.$$

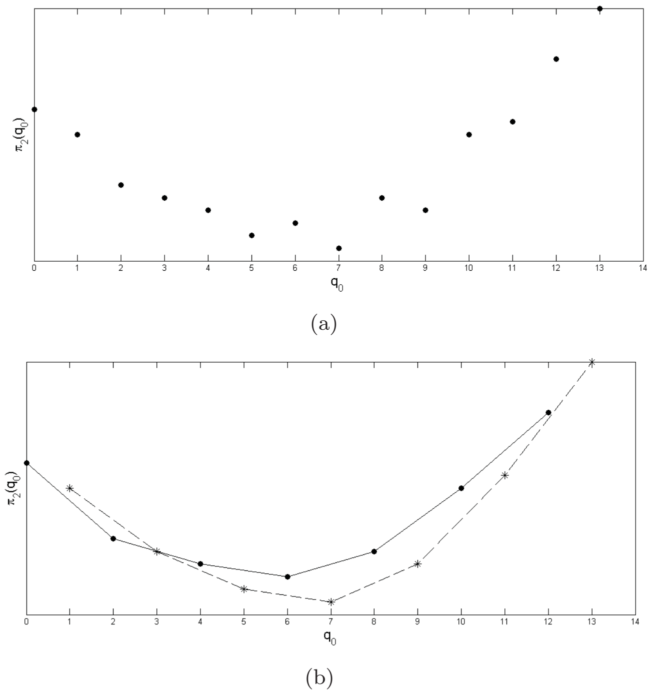

Figure 4.2 Visual Representation of Lemma 4.5

That is, $q_{0,x}$ is the smallest value of q_0 which minimizes $\pi_2(q_0)$ over M_x. By the convexity-like property, we may take

$$q_{0,x} = \begin{cases} \min\{q_0 | \Delta_{q_1}\pi_2(q_0) \geq 0\} & \text{if there exists } q_0 \in M_x - \left\{x + \left\lfloor \frac{N-x}{q_1} \right\rfloor q_1\right\} \\ & \qquad\qquad \text{such that } \Delta_{q_1}\pi_2(q_0) \geq 0 \\ x + \left\lceil \frac{N-x}{q_1} \right\rceil q_1 & \text{otherwise.} \end{cases} \tag{4.6}$$

As a result, an optimal solution q_0^* can be found for Policy 2 by enumerating all optimal solutions $q_{0,x}$ for all $M_x \in \bar{P}$.

Theorem 4.1 The following procedure determines an optimal solution of problem (Q):

1. For all $M_x \in \bar{P}$, find $q_{0,x}$ by equation (4.6).
2. $q_0^* = \underset{q_{0,x}}{\arg\min}\big\{\pi(q_{0,x}) \mid x \in \{0, 1, ..., q_1 - 1\}\big\}$ is an optimal solution.

4.4. Numerical Study

In this section, we present a numerical study for the investigation of the effect of 7 factors on optimality of the two policies. One of these factors is the percentage improvement in MTBF defined as

$$\Delta\tau = \frac{\tau_{new} - \tau_{old}}{\tau_{old}}(100\%).$$

The other 6 factors are MTBF of old parts (τ_{old}) (when percentage improvement $\Delta\tau$ is fixed), downtime costs (included in u_1, u_2, and r), the increase in the unit price of the new parts after time 0 ($c_1 - c_0$) (when initial unit price c_0 is fixed), batch size (q_1), number of systems (N), and remaining lifetime of the systems (T). In our study, we first define a base case with certain factor/parameter values. Then we vary a single factor of interest at a time to extract its effect. Each factor has two choices other than its value in the base case. That is, there are three instances per factor, one of which is common for all (there are $7 \times 2 + 1 = 15$ instances in total). We compute total cost function of Policy 1 (π_1), optimize q_0 under Policy 2, and compute the optimal total cost function of Policy 2 (π_2^*) for each instance. We compare values of π_1 and π_2^* per factor.

4.4.1 Base Case and Choices of Factors

We define the base case with values of factors/parameters given in Table 4.1. We consider a situation in which costs of activities and downtime costs during upgrading a system in Policy 2 and repairing a new part are comparable; so, u_2 and r has equal values. We vary u_2 and r simultaneously while studying the effect of downtime costs.

The choices of the factors other than their values in the base case are given in Table 4.2. For the effect of downtime, we vary u_2 and r, but not u_1, as upgrading is performed preventively (i.e., it is planned) and the effect of downtime can be kept under control under Policy 1.

4.4.2 Results

We define the *relative cost difference* between Policy 1 and Policy 2 as

$$\Delta\pi = \frac{\pi_2^* - \pi_1}{\pi_1}(100\%).$$

We use it as a measure for the effects of the factors on optimality of the policies. We report the optimal policy, values of $\Delta\pi$, and the optimal initial supply quantities q_0^*, for the instances we generate in Table 4.3.

Table 4.1 Values of factors/parameters for the base case

Factor / Parameter	Value
N	50
T (years)	10
τ_{old} (years)	3
$\Delta\tau$	50%
c_0 (Euros per part)	25000
$c_1 - c_0$ (Euros per part)	5000
q_1	4
h (Euros per month per part)	400
s_{old} (Euros per part)	0
s_{new} (Euros per part)	0
u_1 (Euros)	9000
$u_2 = r$ (Euros)	25000
α (yearly)	0.05

We derive the following managerial insights by the results observed in our numerical study:

- High number of systems favors Policy 2. As the number of systems increases, one has to buy a higher number of new parts initially under Policy 1 and replace all old parts without making use of their remaining lifetime. Benefiting from the remaining lifetimes makes Policy 2 favorable. The optimal initial supply quantity (q_0^*) is higher for higher number of systems as the failure rate of the total stream of failures is higher initially.
- Short lifetime favors Policy 2. Once Policy 1 is optimal for a certain lifetime,

Table 4.2 Choices of the factors

Factor	Values
N	40, 60
T (years)	5, 15
τ_{old} (years)	1, 5
$\Delta\tau$	20%, 100%
$c_1 - c_0$ (Euros per part)	0, 10000
q_1	2, 6
u_2 (Euros)	12500, 50000
r (Euros)	12500, 50000

Table 4.3 Results of the numerical study for the effects of the factors

		π_1	π_2^*	$\Delta\pi$	Optimal Policy	q_0^*
N	40	3,108,753	3,116,587	0.25%	Policy 1	12
	50	3,885,941	3,883,587	-0.06%	Policy 2	14
	60	4,663,129	4,648,567	-0.31%	Policy 2	16
T	5	2,928,885	2,704,236	-7.67%	Policy 2	14
	10	3,885,941	3,883,587	-0.06%	Policy 2	14
	15	4,631,297	4,642,833	0.25%	Policy 1	14
τ_{old}	1	8,257,822	8,328,512	0.86%	Policy 1	30
	3	3,885,941	3,883,587	-0.06%	Policy 2	14
	5	3,011,564	2,820,218	-6.35%	Policy 2	10
$\Delta\tau$	20%	4,432,426	4,252,833	-4.05%	Policy 2	14
	50%	3,885,941	3,883,587	-0.06%	Policy 2	14
	100%	3,339,456	3,514,341	5.24%	Policy 1	14
$c_1 - c_0$	0	3,885,941	3,705,901	-4.63%	Policy 2	6
	5000	3,885,941	3,883,587	-0.06%	Policy 2	14
	10000	3,885,941	4,014,705	3.31%	Policy 1	22
q_1	2	3,885,941	3,834,851	-1.31%	Policy 2	12
	4	3,885,941	3,883,587	-0.06%	Policy 2	14
	6	3,885,941	3,910,380	0.63%	Policy 1	14
$u_2 = r$	12500	2,792,970	2,613,377	-6.43%	Policy 2	14
	25000	3,885,941	3,883,587	-0.06%	Policy 2	14
	50000	6,071,882	6,424,007	5.80%	Policy 1	14

it remains optimal for longer lifetimes. Under Policy 1, the replenishment costs and the upgrading costs are incurred at time 0. Only the repair costs are incurred throughout $(0, T]$. Under Policy 2, all costs except the costs of initial supply are incurred throughout $(0, T]$. Under Policy 1, the failures stem only from the new parts and costs are incurred for the repair of the new parts. Under Policy 2, the failures stem from the old parts and the new parts. The old parts are replaced with new ones when they fail, while the new parts are repaired. The expected number of failures of the new parts throughout $(0, T]$ under Policy 1 is smaller than the total expected number of failures of the new parts and the old parts throughout $(0, T]$ under Policy 2. As u_2 and r are equal, after any instant $t \in (0, T]$, the upgrading costs and the repair costs under Policy 2 will be greater than the repair costs under Policy 1. As a result, as T increases, the difference between the total costs incurred under Policy 1 and the total costs incurred under Policy 2 will get smaller.

- High values of the MTBF of old parts favor Policy 2 as Policy 2 benefits from the remaining lifetime of old parts while Policy 1 forfeits from it. The optimal initial supply quantity is lower for higher values of the MTBF of the old parts as the failure rate of the total stream of failures is lower.

- Low values of the percentage improvement in MTBF ($\Delta\tau$) favor Policy 2. Once Policy 1 is optimal for a certain value of the percentage improvement, it remains optimal for its higher values. When the percentage improvement is higher, the total costs decreases under Policy 1 and Policy 2 as the repair costs of the new parts decreases. However, the relative effect of decrease under Policy 1 is higher as Policy 2 still suffers from the failures of the old parts. The percentage improvement does not affect the optimal initial supply quantity as the costs that depend on the initial supply quantity (the costs of the initial supply, the replenishment costs and the storage costs) are not affected by the MTBF of the new parts.

- Low values of the increase in the unit price of the new parts after time 0 favors Policy 2 as one can make use of the remaining lifetime of the old parts and replenish the new parts at a relatively low price when they are needed after time 0. Higher values of the increase in the unit price is an incentive to replenish parts at time 0. This effect is reflected by the decrease in the relative cost difference and the increase in the optimal initial order quantity.

- A small batch size provides the opportunity to fine-tune replenishment timings and actual inventory-on-hand, which favors Policy 2. This advantage declines when the batch size is increased as some parts might be replenished unnecessarily and the average actual inventory-on-hand increases. A high batch size result in high replenishment cost per batch after time 0 and leads to higher optimal initial order quantity.

- Low values of the downtime costs per failure favors Policy 2. This is due to the fact that the expected number of failures of the new parts throughout $(0, T]$ under Policy 1 is smaller than the total expected number of failures of the new parts and the old parts throughout $(0, T]$ under Policy 2. As the total number of failures under Policy 2 is independent of the initial supply quantity, the optimal initial supply quantity is not affected by the increase in the downtime costs.

Our numerical study reveals that a change in any factor may lead to a change in the optimal policy. All factors except the number of the systems and the batch size lead to significant changes in the difference between the total costs under Policy 1 and the total costs under Policy 2 when they are varied.

Another interesting observation is that the difference between the number of the systems and the optimal initial supply quantity ($N - q_0^*$) is a multiple of the fixed batch size (q_1) in all instances. So, no new parts would be procured unnecessarily due to the fixed batch size in the optimal solutions.

4.5. Conclusions

In this chapter, we introduced a model for studying the following two upgrading policies that an OEM may implement after the redesign of a component:

- Policy 1 - Upgrade all systems preventively at time 0: N new parts are bought at time 0 and all the old parts in the field are preventively replaced with the new parts at time 0.

- Policy 2 - Upgrade systems one-by-one correctively: A number of new parts is bought at time 0 (initial supply) and is kept on stock. When an old component in the field fails, it is correctively replaced with a new one from the inventory. The OEM replenishes new parts in batches whenever a new part is needed and there is an out-of-stock situation after time 0.

Lower number of failures and less downtime are the advantages of Policy 1 compared to Policy 2. However, one forfeits the remaining lifetimes of the old parts under Policy 1, while Policy 2 makes use of them. The unit price of the new parts might increase after time 0, which favors Policy 1.

Initial supply quantity is a major factor affecting the costs that would be incurred under Policy 2. We developed a problem formulation which includes the relationship between the initial supply quantity and the relevant costs under Policy 2. We performed exact analysis on our problem formulation and derived a convexity-like property for the total costs under Policy 2. This property enabled us to develop a solution procedure for the optimal initial supply quantity under Policy 2. We conducted a numerical study to derive insights about conditions which favor each policy. We used the percentage difference in the MTBF of the old parts and the MTBF of the new parts as a measure of the reliability improvement. In our numerical study, we found out that Policy 1 is favored by low values of the number of systems, long lifetime of the systems, low values of the MTBF of the old parts (for fixed percentage improvement in MTBF), high values of the percentage improvement in MTBF, high values of the increase in the unit price of the new parts after time 0, large batch sizes, and high values of the downtime costs per failures. The reverse of each of these conditions favors Policy 2. Our numerical study showed that the optimal policy may change by varying any of the mentioned factors.

Appendix

[**Proof of Lemma 4.1**]: As we assume that failures of each new part in the field occur according to Poisson processes with rate $1/\tau_{new}$, total stream of failures of the new parts occur according to a Poisson process with rate N/τ_{new} under Policy 1. Then, the repair costs can be derived in a similar way followed in part (ii) of Lemma 2.2 for the derivation of the repair costs in Chapter 2.

[**Proof of Lemma 4.2**]: In the mini-model, the $\hat{q}$ parts added to the stock at time 0 are the 1^{st}, 2^{nd},..., $\hat{q}^{th}$ new parts. If $\hat{T}_n \leq \hat{T}$, $n \in \{1, 2, ..., \hat{q}\}$, the n^{th} new part is kept on stock throughout $[0, \hat{T}_n]$; otherwise, it is kept throughout $[0, \hat{T}]$. Let $S_n(\hat{N}, \hat{T})$ be the random variable denoting the storage costs incurred for the n^{th} new part. Then,

$$E[S_n(\hat{N}, \hat{T})|\hat{T}_n = t] = \begin{cases} \int_0^t he^{-\alpha u} du = \frac{h}{\alpha}(1 - e^{-\alpha t}) & \text{if } t \leq \hat{T} \\ \int_0^{\hat{T}} he^{-\alpha u} du = \frac{h}{\alpha}(1 - e^{-\alpha \hat{T}}) & \text{if } t > \hat{T}, \end{cases}$$

and

$$\begin{aligned}
E[S_n(\hat{N}, \hat{T})] &= \int_0^{\hat{T}} \frac{h}{\alpha}(1 - e^{-\alpha t}) f_n(t, \hat{N}) dt + \int_{\hat{T}}^{\infty} \frac{h}{\alpha}(1 - e^{-\alpha \hat{T}}) f_n(t, \hat{N}) dt \\
&= \int_0^{\hat{T}} \frac{h}{\alpha}(1 - e^{-\alpha t}) \sum_{i=1}^{n} C_{i,n}(\hat{N}) \lambda_i(\hat{N}) e^{-\lambda_i(\hat{N})t} dt \\
&\quad + \int_{\hat{T}}^{\infty} \frac{h}{\alpha}(1 - e^{-\alpha \hat{T}}) \sum_{i=1}^{n} C_{i,n}(\hat{N}) \lambda_i(\hat{N}) e^{-\lambda_i(\hat{N})t} dt \\
&= \frac{h}{\alpha} \sum_{i=1}^{n} C_{i,n}(\hat{N}) \left[\int_0^{\hat{T}} \lambda_i(\hat{N}) e^{-\lambda_i(\hat{N})t} dt - \int_0^{\hat{T}} \lambda_i(\hat{N}) e^{-\left(\alpha + \lambda_i(\hat{N})\right)t} dt \right. \\
&\quad \left. + (1 - e^{-\alpha \hat{T}}) \int_{\hat{T}}^{\infty} \lambda_i(\hat{N}) e^{-\lambda_i(\hat{N})t} dt \right]
\end{aligned}$$

$$\begin{aligned}
&= \frac{h}{\alpha}\sum_{i=1}^{n} C_{i,n}(\hat{N})\Bigg\{(1-e^{-\lambda_i(\hat{N})\hat{T}}) - \frac{\lambda_i(\hat{N})}{\alpha+\lambda_i(\hat{N})}\left[1-e^{-\left(\alpha+\lambda_i(\hat{N})\right)\hat{T}}\right] \\
&\quad +(1-e^{-\alpha\hat{T}})e^{-\lambda_i(\hat{N})\hat{T}}\Bigg\} \\
&= \frac{h}{\alpha}\sum_{i=1}^{n} C_{i,n}(\hat{N})\Bigg\{1-e^{-\left(\alpha+\lambda_i(\hat{N})\right)\hat{T}} \\
&\quad -\frac{\lambda_i(\hat{N})}{\alpha+\lambda_i(\hat{N})}\left[1-e^{-\left(\alpha+\lambda_i(\hat{N})\right)\hat{T}}\right]\Bigg\} \\
&= \frac{h}{\alpha}\sum_{i=1}^{n} C_{i,n}(\hat{N})\left(1-\frac{\lambda_i(\hat{N})}{\alpha+\lambda_i(\hat{N})}\right)\left[1-e^{-\left(\alpha+\lambda_i(\hat{N})\right)\hat{T}}\right] \\
&= h\sum_{i=1}^{n} C_{i,n}(\hat{N})\frac{1}{\alpha+\lambda_i(\hat{N})}\left[1-e^{-\left(\alpha+\lambda_i(\hat{N})\right)\hat{T}}\right].
\end{aligned}$$

Hence,

$$\hat{S}(\hat{N},\hat{T},\hat{q}) = h\sum_{n=1}^{\hat{q}}\sum_{i=1}^{n} C_{i,n}(\hat{N})\frac{1}{\alpha+\lambda_i(\hat{N})}\left[1-e^{-\left(\alpha+\lambda_i(\hat{N})\right)\hat{T}}\right].$$

[**Proof of Lemma 4.3**]: *(i)* As the 0^{th} batch is procured at time 0, it is equivalent to the batch considered in the mini-model with $\hat{N} = N$, $\hat{T} = T$, and $\hat{q} = q_0$ and $S_{2,0}(q_0) = \hat{S}(N, T, q_0)$.

Remember that if $T_{q_0+(k-1)q_1+1} \leq T$, k^{th} batch, $k \in \{1, 2, ..., \bar{k}(q_0)\}$, is replenished at time $T_{q_0+(k-1)q_1+1}$ and its storage costs are incurred until either time T or the batch is depleted. Otherwise, replenishment of the k^{th} batch does not occur and no storage costs are incurred. Also, remember that the number of the old parts remaining in the field just after time $T_{q_0+(k-1)q_1+1} \leq T$ is $N_k(q_0) = N-[q_0+(k-1)q_1+1]$. Then, the k^{th} batch can be represented by the mini-model with $\hat{N} = N_k(q_0)$, $\hat{T} = T-T_{q_0+(k-1)q_1+1}$, and $\hat{q} = q_1 - 1$ (remember that one new part in the batch is used to replace the old part failed at time $T_{q_0+(k-1)q_1+1}$ and remaining $q_1 - 1$ new parts are added to the inventory). So, $\hat{S}\big(N_k(q_0), T - T_{q_0+(k-1)q_1+1}, q_1 - 1\big)$ is equivalent to the discounted expected storage costs of the k^{th} batch at time $T_{q_0+(k-1)q_1+1}$. Then,

$$\begin{aligned}
S_{2,k}(q_0) &= \int_0^T e^{-\alpha t}\hat{S}\big(N_k(q_0), T - T_{q_0+(k-1)q_1+1}, q_1 - 1\big) f_{T_{q_0+(k-1)q_1+1}}(t)dt \\
&= \int_0^T \sum_{n=1}^{q_1-1}\sum_{i=1}^{n} C_{i,n}(N_k(q_0))\frac{1}{\alpha+\lambda_i(N_k(q_0))}\left[1 - e^{-\left(\alpha+\lambda_i(N_k(q_0))\right)(T-t)}\right] \times \\
&\qquad \sum_{j=1}^{q_0+(k-1)q_1+1} C_{j,q_0+(k-1)q_1+1}(N)\lambda_j(N)e^{-\left(\alpha+\lambda_j(N)\right)t}dt \\
&= \sum_{n=1}^{q_1-1}\sum_{i=1}^{n} C_{i,n}(N_k(q_0))\frac{1}{\alpha+\lambda_i(N_k(q_0))} \times \\
&\qquad \sum_{j=1}^{q_0+(k-1)q_1+1} C_{j,q_0+(k-1)q_1+1}(N)\lambda_j(N) \times \\
&\qquad \left[\int_0^T e^{-\left(\alpha+\lambda_j(N)\right)t}dt - e^{-\left(\alpha+\lambda_i(N_k(q_0))\right)T}\int_0^T e^{\left(\lambda_i(N_k(q_0))-\lambda_j(N)\right)t}dt\right] \\
&= \sum_{n=1}^{q_1-1}\sum_{i=1}^{n} C_{i,n}(N_k(q_0))\frac{1}{\alpha+\lambda_i(N_k(q_0))} \times \\
&\qquad \sum_{j=1}^{q_0+(k-1)q_1+1} C_{j,q_0+(k-1)q_1+1}(N)\lambda_j(N) \times \\
&\qquad \left[\frac{1-e^{-\left(\alpha+\lambda_j(N)\right)T}}{\alpha+\lambda_j(N)} + \frac{e^{-\left(\alpha+\lambda_j(N)\right)T} - e^{-\left(\alpha+\lambda_i(N_k(q_0))\right)T}}{\lambda_j(N)-\lambda_i(N_k(q_0))}\right].
\end{aligned}$$

(ii) Let $K_{2,k}(q_0)$ be the random variable denoting the NPV of the replenishment costs of the k^{th} batch, $k = \{1, 2, ..., \bar{k}(q_0)\}$, of new parts. As k^{th} batch is replenished at time $T_{q_0+(k-1)q_1+1}$,

$$E[K_{2,k}(q_0)|T_{q_0+(k-1)q_1+1} = t] = \begin{cases} c_1 q_1 e^{-\alpha t} & \text{if } t \le T \\ 0 & \text{if } t > T, \end{cases}$$

and,

$$E[K_{2,k}(q_0)] = \int_0^T c_1 q_1 e^{-\alpha t} \sum_{i=1}^{q_0+(k-1)q_1+1} C_{i,q_0+(k-1)q_1+1}(N)\lambda_i(N)e^{-\lambda_i(N)t}dt$$

$$\begin{aligned}
&= c_1 q_1 \sum_{i=1}^{q_0+(k-1)q_1+1} C_{i,q_0+(k-1)q_1+1}(N)\lambda_i(N) \int_0^T e^{-\left(\alpha+\lambda_i(N)\right)T} dt \\
&= c_1 q_1 \sum_{i=1}^{q_0+(k-1)q_1+1} C_{i,q_0+(k-1)q_1+1}(N) \times \\
&\quad \frac{\lambda_i(N)}{\alpha+\lambda_i(N)} \left[1 - e^{-\left(\alpha+\lambda_i(N)\right)T}\right].
\end{aligned}$$

Hence,

$$\begin{aligned}
K_2(q_0) &= \sum_{k=1}^{\bar{k}(q_0)} E[K_{2,k}] \\
&= c_1 q_1 \sum_{k=1}^{\bar{k}(q_0)} \sum_{i=1}^{q_0+(k-1)q_1+1} C_{i,q_0+(k-1)q_1+1}(N) \times \\
&\quad \frac{\lambda_i(N)}{\alpha+\lambda_i(N)} \left[1 - e^{-\left(\alpha+\lambda_i(N)\right)T}\right].
\end{aligned}$$

(iii) In this part, we first derive V_{old}. Then, we derive $V_{new,k}$ for $k = 0$ and $k \in \{1, 2, ..., \bar{k}(q_0)\}$ separately.

Let $V_{old,n}$ be the random variable denoting the NPV of the salvage value of the n^{th} old part under Policy 2.

$$E[V_{old,n}|T_n = t] = \begin{cases} s_{old}e^{-\alpha t} & \text{if } t \le T \\ s_{old}e^{-\alpha T} & \text{if } t > T. \end{cases}$$

Then,

$$\begin{aligned}
E[V_{old,n}] &= \int_0^T s_{old}e^{-\alpha t} f_{T_n}(t)dt + \int_T^\infty s_{old}e^{-\alpha T} f_{T_n}(t)dt \\
&= \int_0^T s_{old}e^{-\alpha t} \sum_{i=1}^{n} C_{i,n}(N)\lambda_i(N)e^{-\lambda_i(N)t}dt \\
&\quad + \int_T^\infty s_{old}e^{-\alpha T} \sum_{i=1}^{n} C_{i,n}(N)\lambda_i(N)e^{-\lambda_i(N)t}dt
\end{aligned}$$

$$\begin{aligned} &= s_{old}\left[\sum_{i=1}^{n} C_{i,n}(N)\lambda_i(N)\int_0^T e^{-\left(\alpha+\lambda_i(N)\right)t}dt\right. \\ &\left. +e^{-\alpha T}\sum_{i=1}^{n} C_{i,n}(N)\lambda_i(N)\int_T^{\infty} e^{-\lambda_i(N)t}dt\right] \\ &= s_{old}\left\{\sum_{i=1}^{n} C_{i,n}(N)\frac{\lambda_i(N)}{\alpha+\lambda_i(N)}\left[1-e^{-\left(\alpha+\lambda_i(N)\right)T}\right]\right. \\ &\left. +\sum_{i=1}^{n} C_{i,n}(N)e^{-\left(\alpha+\lambda_i(N)\right)T}\right\} \\ &= s_{old}\sum_{i=1}^{n} C_{i,n}(N)\frac{1}{\alpha+\lambda_i(N)}\left[\lambda_i(N)+\alpha e^{-\left(\alpha+\lambda_i(N)\right)T}\right], \end{aligned}$$

and

$$V_{old} = s_{old}\sum_{n=1}^{N}\sum_{i=1}^{n} C_{i,n}(N)\frac{1}{\alpha+\lambda_i(N)}\left[\lambda_i(N)+\alpha e^{-\left(\alpha+\lambda_i(N)\right)T}\right].$$

As q_0 new parts procured by the initial supply and these parts are salvaged at time T,

$$V_{new,0}(q_0) = s_{new}e^{-\alpha T}q_0.$$

q_1 new parts are replenished in k^{th} batch, $k \in \{1, 2, ..., \bar{k}(q_0)\}$, at time $T_{q_0+(k-1)q_1+1} < T$ are salvaged at time T. Hence,

$$\begin{aligned} V_{new,k}(q_0) &= \int_0^T s_{new}q_1e^{-\alpha T}f_{T_{q_0+(k-1)q_1+1}}(t)dt \\ &= s_{new}q_1e^{-\alpha T}\int_0^T \sum_{i=1}^{q_0+(k-1)q_1+1} C_{i,q_0+(k-1)q_1+1}(N)\lambda_i(N)e^{-\lambda_i(N)t}dt \\ &= s_{new}q_1 \sum_{i=1}^{q_0+(k-1)q_1+1} C_{i,q_0+(k-1)q_1+1}(N)\left[e^{-\alpha T}-e^{-\left(\alpha+\lambda_i(N)\right)T}\right]. \end{aligned}$$

(iv) Let D_n be the random variable denoting the NPV of the costs that are incurred for upgrading the system at which the n^{th} failure of the old parts occurs under Policy 2.

$$E[D_n|T_n = t] = \begin{cases} u_2e^{-\alpha t} & \text{if } t \le T \\ 0 & \text{if } t > T. \end{cases}$$

Then, the following can be derived by similar operations done for $E[K_{2,k}(q_0)]$:

$$E[D_n] = u_2 \sum_{i=1}^{n} C_{i,n}(N) \frac{\lambda_i(N)}{\alpha + \lambda_i(N)} \left[1 - e^{-\left(\alpha+\lambda_i(N)\right)T}\right].$$

Hence,

$$U_2 = \sum_{n=1}^{N} E\left[D_n\right] = u_2 \sum_{n=1}^{N} \sum_{i=1}^{n} C_{i,n} \frac{\lambda_i}{\alpha + \lambda_i} \left[1 - e^{-\left(\alpha+\lambda_i\right)T}\right].$$

(v) For $n \in \{1, 2, ..., N\}$ such that $T_n < T$, n^{th} new part is in use throughout $[T_n, T]$. Let $H_n(t)$ be the random variable that denotes the discounted costs that are incurred during repairs of the n^{th} new part throughout $[T_n, T]$ at time t. Then, the following can be derived in a similar way followed in part (ii) of Lemma 2.2 for the derivation of the repair costs in Chapter 2:

$$E[H_n(t)|T_n = t] = \begin{cases} \frac{r}{\alpha\tau_{new}} \left[1 - e^{-\alpha(T-t)}\right] & \text{if } t \leq T \\ 0 & \text{if } t > T. \end{cases}$$

Then,

$$\begin{aligned} E[H_n(0)] &= \int_0^T e^{-\alpha t} \frac{r}{\alpha\tau_{new}} \left[1 - e^{-\alpha(T-t)}\right] \sum_{i=1}^{n} C_{i,n}(N) \lambda_i(N) e^{-\lambda_i(N)t} dt \\ &= \frac{r}{\alpha\tau_{new}} \left[\sum_{i=1}^{n} C_{i,n}(N)\lambda_i(N) \int_0^T e^{-\left(\alpha+\lambda_i(N)\right)t} dt\right. \\ &\quad \left. - e^{-\alpha T} \sum_{i=1}^{n} C_{i,n}(N)\lambda_i(N) \int_0^T e^{-\lambda_i(N)t} dt\right] \\ &= \frac{r}{\alpha\tau_{new}} \sum_{n=1}^{N} \sum_{i=1}^{n} C_{i,n}(N) \times \\ &\quad \left\{\frac{\lambda_i(N)}{\alpha + \lambda_i(N)} \left[1 - e^{-\left(\alpha+\lambda_i(N)\right)T}\right] - e^{-\alpha T}(1 - e^{-\lambda_i(N)T})\right\}. \end{aligned}$$

Hence,

$$\begin{aligned} R_2 &= \sum_{n=1}^{N} E[H_n(0)] \\ &= \frac{r}{\alpha\tau_1} \sum_{n=1}^{N} \sum_{i=1}^{n} C_{i,n}(N) \times \\ &\quad \left\{\frac{\lambda_i(N)}{\alpha + \lambda_i(N)} \left[1 - e^{-\left(\alpha+\lambda_i(N)\right)T}\right] - e^{-\alpha T}(1 - e^{-\lambda_i(N)T})\right\}. \end{aligned}$$

Proof of Lemma 4.4: We will prove parts (i) and (ii) via a sample path approach. For a given $x \in \{0, 1, ..., q_1 - 1\}$

$$M_x = \left\{ x, x + q_1, x + 2q_1, ..., x + \left\lceil \frac{N - x}{q_1} \right\rceil q_1 \right\}.$$

Three consecutive elements of M_x can be represented by x, $x + q_1$, and $x + 2q_1$ when its elements are put in an ascending order. Consider three cases in which Policy 2 is implemented with initial supply amounts $q_0^0 = x + zq_1$, $q_0^1 = q_0^0 + q_1$, and $q_0^2 = q_0^0 + 2q_1$, where z is an arbitrary element of $\left\{0, 1, ..., \left\lceil \frac{N-x}{q_1} \right\rceil - 2\right\}$ (i.e. $q_0^2 \leq N$). We refer to these cases as Case 0, Case 1, and Case 2, respectively; i.e., the initial supply amount is q_0^m in Case m, $m \in \{0, 1, 2\}$.

We couple the failure times of the old parts in the three cases. Let y_n, $n \in \{1, 2, ..., N\}$, be an arbitrary realization of a random variable distributed exponentially with rate $\lambda_n = \frac{N-n+1}{\tau_{old}}$. We define the sequence $(t_n)_{n \in \{0,1,2,...,N\}}$ with $t_0 = 0$ and $t_n = t_{n-1} + y_n$. Then, $(t_n)_{n \in \{0,1,2,...,N\}}$ corresponds to an arbitrary sequence of failure times of the old parts under infinite horizon. Notice that $0 = t_0 < t_1 < t_2 < ... < t_N$. Let $\bar{n} = \max\{n \mid t_n \leq T\}$. Then, $(t_n)_{n \in \{1,2,...,\bar{n}\}}$ corresponds to an arbitrary sequence of failure times of the old parts throughout $[0, T]$ (the last failure is the $\bar{n}^{th}$ one). The inequalities $0 = t_0 < t_1 < t_2 < ... < t_N$ assures that there is one failure at a time. We start with the proof of part (i).

(i) Let $I^m(t)$ be the actual inventory-on-hand at time $t \in [0, T]$ in Case m. We will show that

$$I^2(t) - I^1(t) \geq I^1(t) - I^0(t) \ , \ t \in [0, T] \tag{4.7}$$

Inequality (4.7) implies inequality (4.3) as

$$S_2^k = \int_0^T he^{-\alpha t} I^k(t) dt,$$

where S_2^k denotes the storage costs in Case m and we show the inequality for an arbitrary realization of the failure times of the old parts.

If $T \to \infty$, then the following equalities hold:

$$I^1(t) = \begin{cases} I^0(t) + q_1 & \text{for } 0 \leq t < t_{q_0^0+1} \\ I^0(t) & \text{for } t \geq t_{q_0^0+1}, \end{cases} \tag{4.8}$$

$$I^2(t) = \begin{cases} I^1(t) + q_1 & \text{for } 0 \leq t < t_{q_0^0+q_1+1} \\ I^1(t) = I^0(t) & \text{for } t \geq t_{q_0^0+q_1+1}. \end{cases} \tag{4.9}$$

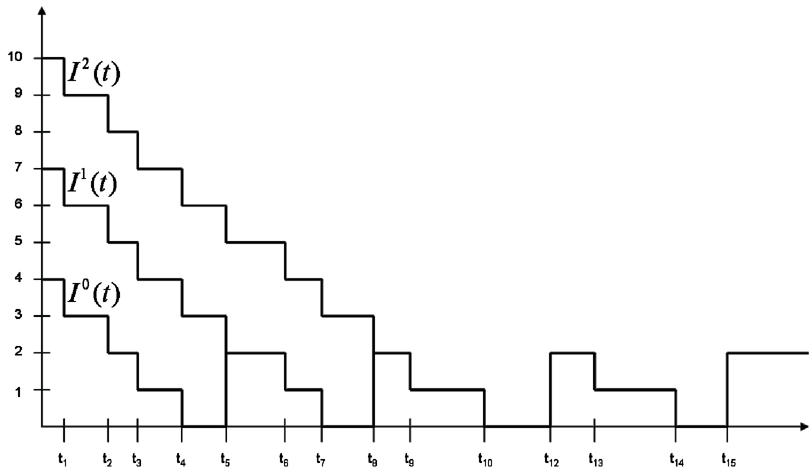

Figure 4.3 Visual representation for equations (4.8) and (4.9)

In Figure 4.3, you can see $I^0(t)$, $I^1(t)$, and $I^2(t)$ graphically for a case with $q_0 = 4$ and $q_1 = 3$. Observe that $I^1(t) = I^0(t) + 3$ for $0 \le t < t_5$ and $I^1(t) = I^0(t)$ for $t \ge t_5$, $I^2(t) = I^1(t) + 3$ for $0 \le t < t_8$ and $I^2(t) = I^1(t) = I^0(t)$ for $t \ge t_8$.

Under a finite horizon T for the inventory, the following cases can be distinguished:

- If $0 \le T < t_{q_0^0+1}$, then:

$$\begin{aligned} I^1(t) &= I^0(t) + q_1 \quad \text{for } 0 \le t \le T, \\ I^2(t) &= I^1(t) + q_1 \quad \text{for } 0 \le t \le T. \end{aligned}$$

  Then $I^2(t) - I^1(t) = I^1(t) - I^0(t) = q_1$, which implies (4.7).

- If $t_{q_0^0+1} \le T < t_{q_0^0+q_1+1}$, then:

$$I^1(t) = \begin{cases} I^0(t) + q_1 & \text{for } 0 \le t < t_{q_0^0+1} \\ I^0(t) & \text{for } t_{q_0^0+1} \le t \le T, \end{cases}$$

$$I^2(t) = I^1(t) + q_1 \quad \text{for } 0 \le t \le T.$$

  Then,

$$I^2(t) - I^1(t) = I^1(t) - I^0(t) = q_1 \quad \text{for } 0 \le t < t_{q_0^0+1},$$

$$I^2(t) - I^1(t) = q_1 > I^1(t) - I^0(t) = 0 \quad \text{for } t_{q_0^0+1} \le t \le T,$$

  which implies (4.7).

- If $T \geq t_{q_0^0+q_1+1}$, then:

$$I^1(t) = \begin{cases} I^0(t) + q_1 & \text{for } 0 \leq t < t_{q_0^0+1} \\ I^0(t) & \text{for } t_{q_0^0+1} \leq t \leq T, \end{cases}$$

$$I^2(t) = \begin{cases} I^1(t) + q_1 & \text{for } 0 \leq t < t_{q_0^0+q_1+1} \\ I^1(t) = I^0(t) & \text{for } t_{q_0^0+q_1+1} \leq t \leq T. \end{cases}$$

Then,

$$I^2(t) - I^1(t) = I^1(t) - I^0(t) = q_1 \quad \text{for } 0 \leq t < t_{q_0^0+1},$$

$$I^2(t) - I^1(t) = q_1 > I^1(t) - I^0(t) = 0 \quad \text{for } t_{q_0^0+1} \leq t < t_{q_0^0+q_1 1},$$

$$I^2(t) - I^1(t) = I^1(t) - I^0(t) = 0 \quad \text{for } t_{q_0^0+q_1+1} \leq t \leq t_T,$$

which implies (4.7).

(ii) Now, we will prove that inequality (4.4) holds. Let K_2^m, and V_2^m be replenishment costs of the new parts throughout $(0, T]$ and the salvage value of the new parts in Case m. Let $\theta_2^m = K_2^m - V_2^m$. We will show that inequality

$$\theta_2^2 - \theta_2^1 \geq \theta_2^1 - \theta_2^0 \tag{4.10}$$

holds, which implies inequality (4.4) as the salvage value of the old parts are fixed and the same for the three cases; see equation (4.2).

K_2^m and V_2^m are dependent on T. Below, we distinguish the different cases that may occur with respect to the value of T.

- If $0 \leq T < t_{q_0^0+1}$, no replenishment occurs throughout $[0, T]$ in the three cases and $K_2^0 = K_2^1 = K_2^2 = 0$. The salvage value of the new parts are as follows:

$$V_2^0 = s_{new} e^{-\alpha T} q_0^0,$$

$$V_2^1 = s_{new} e^{-\alpha T} (q_0^0 + q_1),$$

$$V_2^2 = s_{new} e^{-\alpha T} (q_0^0 + 2q_1).$$

  Then, $\theta_2^2 - \theta_2^1 = \theta_2^1 - \theta_2^0 = -s_{new} e^{-\alpha T} q_1$, which implies (4.10).

- If $t_{q_0^0+1} \leq T < t_{q_0^1+1}$, a batch is replenished at time $t_{q_0^0+1}$ in Case 0 while no replenishment occurs in Case 1 and Case 2 throughout $[0, T]$. Thus,

$$K_2^0 = c_1 q_1 \exp(-\alpha t_{q_0^0+1}),$$

$$K_2^1 = K_2^2 = 0.$$

The salvage value of the new parts are as follows:

$$V_2^0 = s_{new}e^{-\alpha T}(q_0^0 + q_1),$$

$$V_2^1 = s_{new}e^{-\alpha T}(q_0^0 + q_1),$$

$$V_2^2 = s_{new}e^{-\alpha T}(q_0^0 + 2q_1).$$

Then,

$$\theta_2^2 - \theta_2^1 = -s_{new}e^{-\alpha T}q_1,$$

$$\theta_2^1 - \theta_2^0 = -c_1 \exp(-\alpha t_{q_0^0+1})q_1,$$

which implies (4.10) as $s_{new} \le c_1$ and $t_{q_0^0+1} \le T$.

- If $t_{q_0^1+1} \le T < t_{q_0^2+1}$, two batches are replenished at time $t_{q_0^0+1}$ and $t_{q_0^1+1}$, respectively, in Case 0; and one batch is replenished at time $t_{q_0^1+1}$ in Case 1. Thus,

$$K_2^0 = c_1 q_1 \big(\exp(-\alpha t_{q_0^0+1}) + \exp(-\alpha t_{q_0^1+1})\big),$$

$$K_2^1 = c_1 q_1 \exp(-\alpha t_{q_0^1+1}),$$

$$K_2^2 = 0.$$

The salvage value of the new parts are as follows:

$$V_2^0 = s_{new}e^{-\alpha T}(q_0^0 + 2q_1),$$

$$V_2^1 = s_{new}e^{-\alpha T}(q_0^0 + 2q_1),$$

$$V_2^2 = s_{new}e^{-\alpha T}(q_0^0 + 2q_1).$$

Then,

$$\theta_2^2 - \theta_2^1 = -c_1 \exp(-\alpha t_{q_0^1+1})q_1,$$

$$\theta_2^1 - \theta_2^0 = -c_1 \exp(-\alpha t_{q_0^0+1})q_1,$$

which implies (4.10) as $t_{q_0^0+1} \le t_{q_0^1+1}$.

- If $t_{q_{2,0}+1} \le T$, the number of batches replenished in Case 0, Case 1 and Case 2 throughout $[0, T]$ are (remember that $\bar{n} = \max\{n \mid t_n \le T\}$)

  * $\bar{k}^0 = \left\lceil \frac{\bar{n}-q_0}{q_1} \right\rceil$,
  * $\left\lceil \frac{\bar{n}-(q_0+q_1)}{q_1} \right\rceil = \bar{k}^0 - 1$, and
  * $\left\lceil \frac{\bar{n}-(q_0+2q_1)}{q_1} \right\rceil = \bar{k}^0 - 2$,

respectively. The k^{th} batch is replenished at time

* $t_{q_0^0+(k-1)q_1+1}$,
* $t_{q_0^0+q_1+(k-1)q_1+1} = t_{q_0^0+kq_1+1}$, and
* $t_{q_0^0+2q_1+(k-1)q_1+1} = t_{q_0^0+(k+1)q_1+1}$,

in Case 0, Case 1 and Case 2, respectively. Then,

$$
\begin{aligned}
K_2^0 &= c_1 q_1 \sum_{k=1}^{\bar{k}^0} \exp(-\alpha t_{q_0^0+(k-1)q_1+1}) \\
&= c_1 q_1 \left(\exp(-\alpha t_{q_0^0+1}) + \exp(-\alpha t_{q_0^0+q_1+1}) + \sum_{k=3}^{\bar{k}^0} \exp(-\alpha t_{q_0^0+(k-1)q_1+1}) \right), \\
K_2^1 &= c_1 q_1 \sum_{k=1}^{\bar{k}^0-1} \exp(-\alpha t_{q_0^0} + kq_1 + 1) \\
&= c_1 q_1 \left(\exp(-\alpha t_{q_0^0+q_1+1}) + \sum_{k=3}^{\bar{k}^0} \exp(-\alpha t_{q_0^0+(k-1)q_1+1}) \right), \\
K_2^2 &= c_1 q_1 \sum_{k=1}^{\bar{k}^0-2} \exp(-\alpha t_{q_0^0} + (k+1)q_1 + 1) = c_1 q_1 \sum_{k=3}^{\bar{k}^0} \exp(-\alpha t_{q_0^0+(k-1)q_1+1}).
\end{aligned}
$$

In Case 0, Case 1, and Case 2, the total number of the new parts procured is $q_0 + \bar{k}^0 q_1$. As all these parts are salvaged at time T in the three cases, $V_2^0 = V_2^1 = V_2^2$. Then,

$$\theta_2^2 - \theta_2^1 = -c_1 \exp(-\alpha t_{q_0^1+1}) q_1,$$

$$\theta_2^1 - \theta_2^0 = -c_1 \exp(-\alpha t_{q_0^0+1}) q_1,$$

which implies (4.10) as $t_{q_0^0+1} \leq t_{q_0^1+1}$.

Chapter 5

Conclusions

In this thesis, we studied the optimal reliability decisions and upgrading policies for advanced capital goods. Our primary goal was to develop quantitative models and methods for reliability optimization in the design phase. Advanced capital goods require the following attributes to be included in the models which distinguish them from the existing models in the literature:

1. maintenance costs
2. downtime costs or availability (or downtime) constraints

Furthermore, if the repair-by-replacement concept is used for the components whose reliability decisions are considered, the following attributes also become fundamental for the models:

3. spare parts inventory
4. emergency procedure

We started our investigation with the optimization of the reliability of a critical component, which was a single-stage problem. Next, we focused on a redundancy allocation problem, which was a multi-stage one. In both cases, the repair-by-replacement concept was used for components and we developed models which included the 4 attributes listed above. Spare parts inventory levels were also incorporated into these models as decision variables to represent their interaction with the reliability decisions and effect on availability, maintenance activities and relevant costs properly.

We also studied the upgrading policy problem as it is one of the major reliability related issues that OEMs face during the exploitation phase of capital goods. We developed a model for a component which is repaired on site. We included two major policies that are common in both practice and the literature. In one of the policies, an inventory of the improved parts is kept. The initial order quantity for the inventory was one of the main factors that affected the costs incurred for upgrading the systems. We established an explicit relation between the initial order quantity and the relevant costs in our model, which distinguished it from the existing models. We compared the two policies and derived results about conditions which favor each policy.

Below, we shortly summarize our findings per chapter.

In Chapter 2, we developed a quantitative model for the optimization of the reliability level of a critical component. In this model, we formulate the portions of the LCC of a general number of systems that were affected by component reliability and spare parts inventory level. We developed an efficient solution procedure for the problem. By conducting a numerical experiment, we showed that the joint optimization of the component reliability and spare parts inventory level lead to significant cost reductions compared to solutions generated by sequential consideration of these decisions (i.e., component reliability is optimized first; then spare parts inventory level is optimized with respect to the fixed component reliability). The results of the experiment also revealed that the optimal component reliability is much higher for a cheap component than for an expensive component and increases as

- the number of the systems increases,
- the downtime penalty rate increases; and,
- the exploitation phase gets longer.

We also showed that the optimal LCC have negligible or limited sensitivity to the most of the major parameters in our model.

The approach that we followed in Chapter 2 forms the background to tackle single-stage problems that can serve to analyze and solve a related multi-stage problem. We show how to establish such a relationship between a multi-stage problem and a single-stage problem in Chapter 3 for a redundancy allocation problem.

In Chapter 3, we introduced a redundancy allocation model for capital goods. In the problem that we studied, three policies per stage were defined. Redundancy was included by only one of the policies. Each of the three policies provided different levels of uptime. We formulated the problem as the minimization of the TCO of a general number of systems under a defined constraint on the expected downtime of the systems throughout their life cycle. We decomposed the problem into single-stage problems and showed that a solution for the multi-stage problem could be generated by finding solutions of each of the single-stage problems. We developed an efficient procedure to find optimal solutions of the single-stage problems for varying resource levels of the downtime constraint. Solutions for the multi-stage problem for varying resource levels of the downtime constraint can also be generated efficiently by repeating this procedure for each stage.

We derived the following major results through the analysis of the single-stage and multi-stage problem formulations:

- Single-stage: When the value of the resource level of the downtime constraint was varied by starting from a high value and decreased down to zero; i.e., the constraint was initially loose and got tighter and tighter, the policy which included choosing redundancy, became optimal at a certain value of the resource level and remained optimal for all its smaller values afterwards.
- Multi-stage:
  - The values of the TCO and downtime (or uptime) when the optimal policy changed from one to the other could be easily computed. An efficient frontier which reflects the trade-off between the uptime and the TCO can easily be generated by these values.
  - An optimal ordering of the stages to follow for choosing redundancy one-by-one could be generated.

In Chapter 4, we developed a model for studying the following two upgrading policies that an OEM may implement after the redesign of a component (we denote the time just after the redesign by time 0):

- Policy 1 - Upgrade all systems preventively at time 0: N new parts are bought at time 0 and all the old parts in the field are preventively replaced with with the new ones at time 0.
- Policy 2 - Upgrade systems one-by-one correctively: A number of new parts are bought at time 0 (initial supply) and are kept on stock. As an old component in the field fails, it is correctively replaced with a new one from the inventory. The OEM replenishes new parts in batches whenever a new part is needed and there is an out-of-stock situation after time 0.

The two policies can be compared as follows: There would be less number of failures and less downtime under Policy 1; however, one forfeits the remaining lifetimes of the old parts under Policy 1, while Policy 2 benefits from them. Furthermore, an increase in the unit price after time 0, which is probable in the case that we consider, favors Policy 1.

As the initial supply quantity is a major factor affecting the costs that would be incurred under Policy 2, we developed a problem formulation which includes the relationship between the initial supply quantity and the relevant costs under Policy 2. We performed exact analysis on our problem formulation and derived a convexity-like property for the total costs under Policy 2. This property enabled us to develop a solution procedure for the optimal initial supply quantity under Policy 2. We conducted a numerical study to derive insights about the conditions under which each policy is optimal. We used the percentage difference in the MTBF of the old parts and the MTBF of the new parts as a measure of the reliability improvement. We found out that Policy 1 is favored by low values of the number of systems, long lifetime of the systems, low values of the MTBF of the old parts (for fixed percentage improvement in MTBF), high values of the percentage improvement in MTBF, high values of the increase in the unit price of the new parts after time 0, large batch sizes, and high values of the downtime costs per failures. The reverse of each of these conditions favors Policy 2. Our numerical study showed that each of the mentioned factors affect the optimal policy.

Appendices

Appendix A

Monotonicity and Supermodularity Results for the Erlang Loss System

A.1. Introduction

Consider the Erlang loss system, also denoted as $M/G/s/s$ queue, with arrival rate $\lambda > 0$, mean service time μ^{-1}, ($\mu > 0$), and s parallel servers ($s \in \mathbb{N}_0 := \{0\} \cup \mathbb{N}$). Its steady-state probability that all servers are busy is equal to

$$B(s, a) = \frac{\frac{a^s}{s!}}{\sum\limits_{i=0}^{s} \frac{a^i}{i!}}, \tag{A.1}$$

where $a = \lambda\mu^{-1}$ (> 0) is the offered load. The formula in (A.1) is called the *Erlang loss formula* or *Erlang B formula*, and it was first derived by Erlang (1918) for deterministic service times. Later, Sevastyanov (1957) showed that $B(s, a)$ is insensitive to the service time distribution; that is, equation (A.1) is valid for any service time distribution with mean μ^{-1}. The Erlang loss formula occurs in many different applications and its analytical properties are useful for e.g. solving design problems; see Cooper (1982).

In the literature, the following properties are known for $B(s, a)$ and related quantities. Karush (1957) showed that $B(s, a)$ is strictly convex and decreasing as a function of $s \in \mathbb{N}_0$ (see also Remark 2 in Kranenburg and van Houtum (2007)). Harel (1990) investigated $B(s,a)$ as a function of the traffic intensity $\rho = \frac{\lambda}{s\mu}$, service rate μ, and arrival rate λ. He showed that, for each fixed $s \in \mathbb{N}$, there exists a ρ^* such that $B(s,a)$ is strictly convex and increasing in ρ for all $\rho < \rho^*$ and strictly concave and increasing in ρ for all $\rho > \rho^*$. Hence, equivalently, for each fixed $s \in \mathbb{N}$, there exists an a^* such that $B(s,a)$ is strictly convex and increasing in a for all $a < a^*$ and strictly concave and increasing in a for all $a > a^*$. For $s = 0$, $B(s,a) = 1$ for all a, i.e., then $B(s,a)$ is a constant function of a (or ρ). Harel also showed that $B(s,a)$ is strictly convex and decreasing in μ for a fixed λ and $s \in \mathbb{N}$.

The *carried load* $A(s,a)$ is defined as the time-average amount of work carried out by the Erlang loss system, and is equal to

$$A(s,a) = a\left[1 - B(s,a)\right], \qquad s \in \mathbb{N}_0. \tag{A.2}$$

By the above result of Karush for $B(s,a)$, $A(s,a)$ is strictly concave and increasing in s. Yao and Shanthikumar (1987) showed that, the throughput $\lambda[1 - B(s,a)]$ is concave and increasing in λ for a fixed μ. Hence, equivalently, $A(s,a)$ is concave and increasing in a.

The *load carried by the last server* of a system with s servers is defined as the extra load that can be handled in comparison to a system with $s-1$ servers. This load carried by the last server is denoted by $F_B(s,a)$, and it holds that

$$\begin{aligned} F_B(s,a) &= A(s,a) - A(s-1,a) \\ &= a\left[B(s-1,a) - B(s,a)\right], \quad s \in \mathbb{N}. \end{aligned} \tag{A.3}$$

Because of the strict concavity of $A(s,a)$ as a function of s, $F_B(s,a)$ is strictly decreasing in s. In this technical note, we prove that $F_B(s,a)$ is strictly increasing as a function of the offered load a and that $A(s,a)$ is strictly supermodular on $X := \{(s,a) | s \in \mathbb{N}_0 \text{ and } a \in (0,\infty)\}$. We use the regular '$\leq$' ordering for X; i.e., for elements $(s^-,a^-),(s^+,a^+) \in X$, we say that $(s^-,a^-) \leq (s^+,a^+)$ if and only if $s^- \leq s^+$ and $a^- \leq a^+$. Then the set X is a so-called lattice, and thus the definitions of supermodular and submodular functions apply; see p. 43 of **?**. These definitions imply that the function $A(s,a)$ is strictly supermodular on X if and only if

$$A(s^+,a^-) + A(s^-,a^+) < A(s^-,a^-) + A(s^+,a^+) \tag{A.4}$$

for all $(s^-,a^-),(s^+,a^+) \in X$ with $s^- < s^+$ and $a^- < a^+$.

Theorem A.1

(i) For each $s \in \mathbb{N}$, $F_B(s,a)$ is strictly increasing as a function of $a \in (0,\infty)$.

(ii) $A(s,a)$ is strictly supermodular on X.

The proof of Theorem A.1 is given in Section A.2, where we first prove part (i) via a kind of sample path method (the actual result that we prove is even stronger than part (i) and also holds under generalized assumptions; see Remark A.1). Next, we show that part(i) and part (ii) are equivalent. An alternative, algebraic proof of Theorem A.1 is given in Öner et al. (2008).

As $A(s,a) = \sum_{i=1}^{s} F_B(i,a)$, part (i) of Theorem A.1 implies that, for each fixed $s \in \mathbb{N}$, $A(s,a)$ is strictly increasing in a. (For $s = 0$, $A(s,a) = 0$ for all a, i.e., then $B(s,a)$ is a constant function of a.)

Theorem A.1 may be relevant for design problems with the offered load a (or the arrival rate λ when μ is fixed) and the number of servers s as decision variables. To demonstrate this relevance, we exploit part (ii) of Theorem A.1 in a simple optimization problem for an Erlang loss system in Section A.3.

The main motivation for deriving Theorem A.1 came from the component reliability problem that we study in Chapter 2. Remember that in Chapter 2, we develop a model for the effect of the reliability level of a single component of a capital good on the life cycle costs for the whole installed base of that capital good. In the resulting optimization problem, one has the reliability level and the spare parts stock as decision variables. These variables play a similar role as the arrival rate λ and the number of servers s of the Erlang loss system. Part (i) of Theorem A.1 is used in the derivation of an efficient optimization procedure.

A.2. Proof of Theorem A.1

The Erlang loss system with s servers ($s \in \mathbb{N}$) can be viewed as an *ordered-entry* system with all servers having the same service rate but rank-ordered; that is, the servers are rank-ordered from 1 to s and each arriving customer will be served by the first available server under this rank order. Then, $F_B(s,a)$ is equal to the steady state probability that the last (s^{th}) server is busy at an arbitrary instant (see Cooper (1982)). We make use of this equality in a sample path based approach and prove part (i) of Theorem A.1 for the special case with exponential service times. As $B(s,a)$ is insensitive to the service time distribution (so is $F_B(s,a)$), this result implies that part (i) of Theorem A.1 also holds for generally distributed service times. Next, we prove part (ii) by showing that it is equivalent to part (i).

Consider two $M/M/s/s$ systems with rank-ordered servers, with service rates μ, and arrival rates $\lambda_1 = \lambda$ and $\lambda_2 = \lambda + \epsilon$, $\epsilon > 0$, respectively. We refer to these systems as System 1 and System 2. We denote the server with rank order j in System i by C_j^i, $j \in S = \{1, 2, \ldots, s\}$, $i \in \{1, 2\}$. We denote the offered load to System i by a_i;

it holds that $a_1 < a_2$. Letting P_s^i be the steady-state probability that the last server in System i is busy at an arbitrary instant, we need to show that $P_s^1 < P_s^2$. We first show $P_s^1 \leq P_s^2$ through the sample path based approach, and after that we show that the inequality is strict.

Since $\lambda_1 < \lambda_2$, we cannot couple all arrivals of the two systems; however, coupling all arrivals at System 1 with a subset of arrivals at System 2 is possible. To enable this coupling, we first split the Poisson arrival process with rate $\lambda_2 = \lambda + \epsilon$ for System 2 into two independent Poisson arrival processes with rates $\lambda_1 = \lambda$ and ϵ, respectively. We denote the time of the n^{th} arrival to System 2 by t_n, $n \in \mathbb{N}$. It holds that $0 < t_1 < t_2 < \ldots$ (only one arrival at a time) and $\lim_{n\to\infty} t_n = \infty$. Then, $(t_n)_{n\in\mathbb{N}}$ is the sequence of arrival times for all customers at System 2. Let $(\tilde{t}_n)$ be a subsequence of (t_n) that corresponds to arrival times originating from the first decomposed Poisson process with rate λ_1. The times $(\tilde{t}_n)$ are coupled with the arrivals at System 1. Thus, the n^{th} arrival to System 2 can be of two types which we denote by r_n:

$$r_n = \begin{cases} 1 & \text{if the } n^{th} \text{ arrival is common at both systems;} \\ 2 & \text{if the } n^{th} \text{ arrival is only at System 2.} \end{cases}$$

We use $I_j^i(t)$ to indicate whether server C_j^i is busy at time t:

$$I_j^i(t) = \begin{cases} 1 & \text{if server } C_j^i \text{ is busy at time;} \\ 0 & \text{otherwise.} \end{cases}$$

Exponential service times allow us to sample new remaining service times at any point in time; see Enders et al. (2008) for a similar approach. In both System 1 and System 2, we resample the service times at the times t_n, i.e., the arrival times at System 2. Let $u_{j,n}^i$ be the remaining service time sampled for server C_j^i at time t_n if $I_j^i(t_n) = 1$. If servers C_j^1 and C_j^2 are both busy at time t_n, we couple their remaining service times by sampling a single service time for both; that is, $u_{j,n}^1 = u_{j,n}^2$ for each j for which $I_j^1(t_n) = I_j^2(t_n) = 1$.

We will show that for all $j \in S$: if server C_j^1 is busy at any time $t \in [0, \infty)$, then server C_j^2 is also busy at t. We state this claim mathematically as

$$I_j^1(t) \leq I_j^2(t) \quad \text{for all } j \in S \text{ and } t \in [0, \infty). \tag{A.5}$$

This result implies that the probability that the last server in System 1 is busy is less than or equal to the probability that the last server in System 2 is busy at any time; i.e., that $P_s^1 \leq P_s^2$.

We prove inequality (A.5) by induction. There are no arrivals or departures in the time interval $[0, t_1)$. So, all servers are idle in both systems and inequality (A.5) holds in this period. Assume that inequality (A.5) is valid for $t \in [0, t_n)$, for some $n \in \mathbb{N}$. Below, we prove that inequality (A.5) also holds for $t \in [t_n, t_{n+1})$, and thus, for $[0, t_{n+1})$.

At time t_n, a new customer arrives. Let t_n^- denote the time just before t_n. We define

$$p = \begin{cases} \min\{j | I_j^1(t_n^-) = 0\} & \text{if there exists a } j \text{ such that } I_j^1(t_n^-) = 0 \\ \infty & \text{otherwise,} \end{cases}$$

and

$$q = \begin{cases} \min\{j | I_j^2(t_n^-) = 0\} & \text{if there exists a } j \text{ such that } I_j^2(t_n^-) = 0 \\ \infty & \text{otherwise.} \end{cases}$$

That is, if $p < \infty$, then C_p^1 is the idle server with the lowest rank-order in System 1 just before t_n, and all servers are occupied otherwise; and, similarly, q denotes the 'lowest' idle server for System 2. By inequality (A.5), $p \leq q$.

- If $r_n = 1$ (i.e., we have an arrival at both systems), then we distinguish 4 cases:
  - (i) $p = q < \infty$: The customer arriving at time t_n is served by servers C_p^1 (in System 1) and C_p^2 (in System 2). That is, $I_p^1(t_n) = I_p^2(t_n) = 1$ and $I_j^1(t_n) = I_j^1(t_n^-) \leq I_j^2(t_n^-) = I_j^2(t_n)$ for all $j \in S$, $j \neq p$.
  - (ii) $p < q < \infty$: $I_p^1(t_n^-) = 0$, $I_p^2(t_n^-) = 1$, $I_q^1(t_n^-) = I_q^2(t_n^-) = 0$, and the customer arriving at time t_n is served by C_p^1 in System 1 and by C_q^2 in System 2. Then, $I_p^1(t_n) = I_p^2(t_n) = 1$, $I_q^1(t_n) = 0 < 1 = I_q^2(t_n)$, and $I_j^1(t_n) = I_j^1(t_n^-) \leq I_j^2(t_n^-) = I_j^2(t_n)$ for all $j \in S$, $j \neq p$, $j \neq q$.
  - (iii) $p < q = \infty$: The customer arriving at time t_n is served by C_p^1 in System 1, while she/he is lost by System 2. Then, $I_p^1(t_n) = I_p^2(t_n) = 1$ and $I_j^1(t_n) = I_j^1(t_n^-) \leq I_j^2(t_n^-) = I_j^2(t_n) = 1$ for all $j \in S$, $j \neq p$.
  - (iv) $p = q = \infty$: The customer arriving at time t_n is lost by both systems. Thus $I_j^1(t_n) = I_j^1(t_n^-) = I_j^2(t_n^-) = I_j^2(t_n) = 1$ for all $j \in S$.
- If $r_n = 2$ (i.e., we have an arrival at System 2 only), we distinguish 2 cases:
  - (i) $q < \infty$: $I_q^1(t_n^-) = I_q^2(t_n^-) = 0$ and the customer arriving at time t_n is served by C_q^2. Then, $I_q^1(t_n) = I_q^1(t_n^-) = 0 < 1 = I_q^2(t_n)$ and $I_j^1(t_n) = I_j^1(t_n^-) \leq I_j^2(t_n^-) = I_j^2(t_n)$ for all $j \in S$, $j \neq q$.
  - (ii) $q = \infty$: The customer arriving at time t_n is lost; that is, $I_j^1(t_n) = I_j^1(t_n^-) \leq I_j^2(t_n^-) = I_j^2(t_n) = 1$ for all $j \in S$.

Hence, inequality (A.5) holds at t_n in all cases.

Now, we will show that inequalities (A.5) hold on (t_n, t_{n+1}). We will do this by comparing the status of C_j^1 and C_j^2 (as denoted by $I_j^1(t)$ and $I_j^2(t)$) for each $j \in S$

on (t_n, t_{n+1}), which depends on their status and remaining service times at time t_n ($I_j^1(t_n)$, $I_j^2(t_n)$, $u_{j,n}^1$ and $u_{j,n}^2$). For each $j \in S$, the following cases may occur:

- $I_j^1(t_n) = I_j^2(t_n) = 1$ (both servers are busy at time t_n): It holds that $u_{j,n}^1 = u_{j,n}^2$. If $t_n + u_{j,n}^1 < t_{n+1}$, then $I_j^1(t) = I_j^2(t) = 1$ for $t \in (t_n, t_n + u_{j,n}^1)$ and $I_j^1(t) = I_j^2(t) = 0$ for $t \in (t_n + u_{j,n}^1, t_{n+1})$. If $t_n + u_{j,n}^1 > t_{n+1}$, then $I_j^1(t) = I_j^2(t) = 1$ for $t \in (t_n, t_{n+1})$.
- $I_j^1(t_n) = 0, I_j^2(t_n) = 1$ (the server in System 1 is idle and the server in System 2 is busy at time t_n): If $t_n + u_{j,n}^2 < t_{n+1}$, then $0 = I_j^1(t) < I_j^2(t) = 1$ for $t \in (t_n, t_n + u_{j,n}^2)$ and $I_j^1(t) = I_j^2(t) = 0$ for $t \in (t_n + u_{j,n}^2, t_{n+1})$. If $t_n + u_{j,n}^2 > t_{n+1}$, then $0 = I_j^1(t) < I_j^2(t) = 1$ for $t \in (t_n, t_{n+1})$.
- $I_j^1(t_n) = I_j^2(t_n) = 0$ (both servers are idle at time t_n): $I_j^1(t) = I_j^2(t) = 0$ for $t \in (t_n, t_{n+1})$.

Hence, $I_j^1(t) \le I_j^2(t)$ holds for $t \in (t_n, t_n + 1)$ in all cases and the proof of $P_s^1 \le P_s^2$ is completed. Next, we will show that this inequality is strict; i.e. $P_s^1 < P_s^2$.

Let Systems 1 and 2 be coupled in the same way as above. System 2 regenerates at the time points that all servers become idle. P_s^i, $i = 1, 2$, is equal to the fraction of time that the last server of System i is busy in the first renewal interval of System 2, where one has to take the average over all possible sample paths (notice that a sample path is described by the arrival times t_n, the variables r_n, and the remaining service times $u_{j,n}^i$). Under the sample paths with $r_s = 2$ and no service completions before time t_s, the s-th server of System 2 is busy for a positive fraction of time, while the s-th server of System 1 is idle. As these sample paths have a positive probability mass, it holds that $P_s^1 < P_s^2$.

Now, we prove that part (i) and part (ii) are equivalent. Let $(s^-, a^-), (s^+, a^+) \in X$ with $s^- < s^+$ and $a^- < a^+$. For the strict supermodularity of $A(s, a)$ on X, we must show that inequality (A.4) holds. By the strictly increasing behavior of $F_B(s, a)$ in a, we find that

$$\begin{aligned} A(s^+, a^-) - A(s^-, a^-) &= \sum_{s=s^-+1}^{s^+} [A(s, a^-) - A(s-1, a^-)] = \sum_{s=s^-+1}^{s^+} F_B(s, a^-) \\ &< \sum_{s=s^-+1}^{s^+} F_B(s, a^+) = \sum_{s=s^-+1}^{s^+} [A(s, a^+) - A(s-1, a^+)] \\ &= A(s^+, a^+) - A(s^-, a^+), \end{aligned} \tag{A.6}$$

which implies (A.4). As a final step, it is trivial to show that inequality (A.4) implies the strictly increasing behavior of $F_B(s, a)$ in a.

Remark A.1 Via the sample path proof, we proved inequality (A.5). This result does not only imply that the load of the last server increases in a. In fact, it implies that, in an ordered-entry system, the load of each of the servers increases in a. Further, the proof of inequality (A.5) is easily extended to situations with generalized assumptions for the service times and arrival streams at Systems 1 and 2:

- As long as we assume an ordered-entry system, we may allow different service rates at the s parallel servers. However, we need to keep the assumption of exponentially distributed service times, so that we can still resample service times at the arrival times t_n.

- We may relax the assumption of Poisson arrivals. The only property that we need is that the arrival times at System 1 form a subset of the arrival times at System 2, so that all arrivals at System 1 can be coupled with arrivals at System 2. This allows for the following generalizations of the arrival processes at the Systems 1 and 2: (i) Potential arrivals are generated by a renewal process, and each potential arrival becomes an actual arrival with a given probability, where the given probability is larger for System 2 than for System 1; (ii) Compound Poisson arrival processes with the same distribution for the compounds for both systems but with a higher arrival rate of compounds for System 2 than for System 1; (iii) Compound Poisson arrival processes with the same arrival rate of compounds for both systems but with stochastically larger compounds for System 2 than for System 1.

A.3. Application

Consider an Erlang loss system (e.g., a call center), with arrival rate λ, average service time μ^{-1} (> 0), and $s \in \mathbb{N}_0$ parallel servers. The arrival rate depends on the intensity of advertisements activities; $\lambda \in [\lambda_l, \lambda_u]$, where $0 < \lambda_l < \lambda_u$. One earns a fixed revenue r (> 0) for each served customer, and costs consist of advertisement costs and costs for the servers. The advertisement costs to obtain an arrival rate λ are given by a function $K(\lambda)$, which is assumed to be increasing and convex on $[\lambda_l, \lambda_u]$. These costs are made per time unit. The cost per server per time unit is c (> 0). The average profit per time unit is denoted by the function $P(s, \lambda)$, and is equal to

$$P(s, \lambda) = rA(s, a) - K(\lambda) - cs, \qquad s \in \mathbb{N}_0,\ \lambda \in [\lambda_l, \lambda_u], \tag{A.7}$$

where $a = \lambda\mu^{-1}$ is the offered load and $A(s, a)$ is the carried load by the system (cf. the definitions in Section A.1).

By part (ii) of Theorem A.1, we know that, for a fixed μ, $A(s,a) = A(s,\lambda/\mu)$ is strictly supermodular in (s,λ), where $(s,\lambda) \in X' = \{(s,\lambda)|s \in \mathbb{N}_0 \text{ and } \lambda \in [\lambda_l,\lambda_u]\}$. As the second and third term on the righthand side of equation (A.7) only depend on λ and s, respectively, it immediately follows that also $P(s,\lambda)$ is strictly supermodular on X'. Therefore we obtain the following monotonicity results for optimal solutions.

Suppose that $s \in \mathbb{N}_0$ is fixed and that we are interested in the optimization of λ. By Yao and Shanthikumar (1987), $A(s,a) = A(s,\lambda/\mu)$ is concave in λ. Further, $K(\lambda)$ is convex, and thus $P(s,\lambda)$ is concave in λ. Therefore $P(s,\lambda)$ is maximized by

$$\lambda^*(s) := \begin{cases} \lambda_l \text{ if } P(s,\lambda) \text{ is stricly decreasing on } [\lambda_l,\lambda_u]; \\ \lambda_u \text{ if } P(s,\lambda) \text{ is stricly increasing on } [\lambda_l,\lambda_u]; \\ \text{the smallest } \lambda \text{ for which } \frac{\mathrm{d}}{\mathrm{d}\lambda}P(s,\lambda) = 0 \text{ otherwise.} \end{cases}$$

Because of the supermodularity of $P(s,\lambda)$, it holds that $\lambda^*(s)$ is increasing as a function of s. Similarly, we may assume that $\lambda \in [\lambda_l,\lambda_u]$ is fixed and that we want to optimize s. $P(s,\lambda)$ is strictly concave in s, and hence $P(s,\lambda)$ is maximized by

$$s^*(\lambda) := \text{the smallest } s \text{ for which } P(s+1,\lambda) - P(s,\lambda) \leq 0.$$

Because of the supermodularity of $P(s,\lambda)$, $s^*(\lambda)$ is increasing as a function of λ.

Finally, suppose that we want to optimize both s and λ. Then the above properties can be exploited to obtain the following efficient optimization procedure. First, determine $s_l = s^*(\lambda_l)$ and $s_u = s^*(\lambda_u)$. Notice that there is an optimal solution (s^*,λ^*) with $s^* \in \{s|s_l \leq s \leq s_u\}$. Next, determine $\lambda^*(s)$ for each $s = s_l, s_l+1, \ldots, s_u$. Finally, an optimal solution (s^*,λ^*) is found as a best solution among the set $\{(s,\lambda^*(s))|s_l \leq s \leq s_u\}$.

Bibliography

AberdeenGroup. 2003. *Service parts management, Unlocking value and profits in the service chain*. AberdeenGroup, Boston.

AberdeenGroup. 2005. *The Service parts management solution selection report - SPM Strategy and Technology Selection Handbook*. AberdeenGroup, Boston.

AberdeenGroup. 2006. *The convergence of people and parts in the service chain: Benchmarking the alignment of service labor and inventory management*. AberdeenGroup, Boston.

Asiedu, Y., P. Gu. 1998. Life cycle cost analysis: State of the art review. *International Journal of Production Research* **36** 883–908.

Barringer, H.P., D.P. Weber. 1996. *Life cycle cost tutorial*. Fifth International Conference of Process Plant Reliability.

Basten, R.J.I. 2006. *Baseline Study Thales Nederland*. University of Twente, Enschede (the Netherlands). Confidential report.

Birolini, A. 2007. *Reliability Engineering: Theory and Practice*. Springer, Berlin.

Black, G., F. Proschan. 1959. On optimal redundancy. *Operations Research* **7** 582–588.

Blischke, W.R. 1990. Mathematical models for analysis of warranty policies. *Mathematical and computer modelling* **13** 1–16.

Blischke, W.R., D.N.P. Murthy. 1992. Product warranty management - i: A taxonomy for warranty policies. *European Journal of Operational Research* **62** 127–148.

Blischke, W.R., D.N.P. Murthy. 1996. *Product Warranty Handbook*. Dekker, New York.

Blischke, W.R., D.N.P. Murthy. 2000. *Reliability: Modeling, Prediction, and Optimization*. Wiley-Interscience, New York.

Bryant, J.L., R.A. Murphy. 1983. Stocking repair kits for systems with limited life. *Management Science* **29** 546–558.

Chand, S., S. Sethi. 1982. Planning horizon procedures for machine replacement models with several possible replacement alternatives. *Naval Research Logistics Quartterly* **33** 483–493.

Chattopadhyay, G., A. Rahman. 2008. Development of lifetime warranty policies and models for estimating costs. *Reliability Engineering and System Safety* **93** 522–529.

Chen, M. S. 1992. On the computational complexity of reliability redundancy allocation in a series system. *Operations Research Letters* **11** 309–315.

Christensen, G.A., P.N. Sparks, K.J. Kostuk. 2005. A method-based survey of life cycle costing literature pertinent to infrastructure design and renewal. *Canadian Journal of Civil Engineering* **32** 250–259.

Clavareau, J., P.E. Labeau. 2009a. Maintenance and replacement policies under technological obsolescence. *Reliability Engineering and System Safety* **94** 370–381.

Clavareau, J., P.E. Labeau. 2009b. A petri net-based modelling of replacement strategies under technological obsolescence. *Reliability Engineering and System Safety* **94** 357–369.

cnet news. 2001. *California power outages suspended - for now*. http://news.cnet.com/2100-1017-251167.html. Last checked on December 21, 2009.

Coetzee. 2004. *Maintenance*. Maintenance Publishers, USA.

Cohen, M., N. Agrawal, V. Agrawal. 2006. Winning in the aftermarket. *Harvard Business Review (May)* 129–138.

Cooper, R.B. 1982. *Introduction to Queueing Theory*. 2nd ed. North Holland, NY.

Deloitte. 2006. *The service revolution in gloabl manufacturing industries*. Deloitte Research.

Downtime-Central. 2009. *TDC - Real Examples of Downtime Cost*. http://www.downtimecentral.com/Examples.shtml. Last checked on December 21, 2009.

Elegbede, C., K. Adjallah. 2003. Availability allocation to repairable systems with genetic algorithms: a multi-objective formulation. *Reliability Engineering and System Safety* **82** 319–330.

Ellram, L. 1994. A taxonomy of total cost of ownership models. *Journal of Business Logistics* **15** 171–191.

Enders, P., I. Adan, A. Scheller-Wolf, G.J. van Houtum. 2008. *Inventory Rationing for a System with Heterogeneous Customer Classes*. Working paper 2008-E2, Tepper School of Business, Carnegie Mellon University.

Erlang, A.K. 1918. Solutions of some problems in the theory of probabilities of significance in automatic telephone exchanges. *The Post Office Electrical Engineers' Journal* **10** 189–197.

Everett, H. 1963. Generalized lagrange multiplier method for solving problems of optimum allocation of resources. *Operations Research* **11** 399–417.

Fox, B. 1966. Discrete optimization via marginal analysis. *Managament Science* **13** 210–216.

Gupta, Y., W.S. Chow. 1985. Twenty-five years of life cycle costing - theory and applications: A survey. *The International Journal of Quality and Reliability Management* **2** 551–576.

Gupta, Y.P. 1983. *Life Cycle Cost models and associated uncertainties*. Volume F3 of NATO ASI Series, 535–549.

Harel, A. 1990. Convexity properties of the Erlang loss formula. *Operations Research* **38** 499–505.

Huang, H.Z., H.J. Liu, D.N.P. Murthy. 2007. Optimal reliability, warranty and price for new products. *IIE Transactions* **39** 819–827.

Hussain, A.Z.M.O., D.N.P. Murthy. 1998. Warranty and redundancy design with uncertain quality. *IIE Transactions* **30** 1191–1199.

Hussain, A.Z.M.O., D.N.P. Murthy. 2003. Warranty and optimal reliability improvement through product development. *Mathematical and Computer Modelling* **38** 1211–1217.

Karush, W. 1957. A queueing model for an inventory problem. *Operations Research* **5** 693–703.

Kim, S.H., M.A. Cohen, S. Netessine. 2007a. Performance contracting in after-sales service supply chains. *Management Science* **53** 1843–1858.

Kim, S.H., M.A. Cohen, S. Netessine. 2007b. *Reliability or inventory? Contracting strategies for after-sales product support*. Working paper, The Wharton School.

Kranenburg, A.A., G.J. van Houtum. 2007. Cost optimization in the (S-1,S) lost sales inventory model with multiple demand classes. *Operations Research Letters* **35** 493–502.

Kumar, U. Dinesh, J. Kroker, J. Knezevic, M. El-Haram. 2000. *Reliability, Maintenance and Logistic Support: A Life Cycle Approach*. Kluwer Academic Publishers, Indian Institute of Management, Calcuta.

Kuo, W., V.R. Prasad. 2000. An annotated overview of system-reliability optimization. *IEEE Transactions on Reliability* **49** 176–187.

Kuo, W., R. Wan. 2007. Recent advances in optimal reliability allocation. *IEEE Transactions on Systems, Man, and Cybernetics* **37** 143–156.

Lewis, E.E. 1996. *Introduction to Reliability Engineering*. 2nd ed. Wiley, New York.

Mercier, S. 2008. Optimal replacement policy for obsolete components with general failure rates. *Applied Stochastic Models in Business and Industry* **24** 221–235.

Mercier, S., P.E. Labeau. 2004. Optimal replacement policy for a series system with obsolescence. *Applied Stochastic Models in Business and Industry* **20** 73–91.

Mettas, A. 2000. Reliability allocation and optimization for complex systems. *Proceedings of the Annual Reliability and Maintainability Symposium* 216–221.

Meutstege, G.J.H. 2007. *Life Cycle Cost Analysis at PANalytical. Analyzing and modelling the cost components throughout a product's life cycle*. Master's thesis, University of Twente, Enchede (the Netherlands).

Misra, K.B., U. Sharma. 1991. An efficient algorithm to solve integer-programming problems arising in s y stem-reliability design. *IEEE Transactions on Reliability* **40** 81–91.

Mizukami, K. 1968. Optimum redundancy for maximum system reliability by the method of convex and integer programming. *Operations Research* **16** 392–406.

Monga, A., M.J. Zuo. 1998. Optimal system design considering maintenance and warranty. *Computers and Operations Research* **25** 691–705.

Moss, M. 1985. *Designing for minimal expense: the practical application of reliability and maintainability*. Marcel Dekker, New York.

Muckstadt, J.A. 2005. *Analysis and Algorithms for Service Parts Supply Chains*. Springer, New York.

Murthy, D.N.P., W.R. Blischke. 1992a. Product warranty management - ii: An integrated framework for study. *European Journal of Operational Research* **62** 261–281.

Murthy, D.N.P., W.R. Blischke. 1992b. Product warranty management - iii: A review of mathematical models. *European Journal of Operational Research* **62** 1–34.

Murthy, D.N.P., I. Djamaludin. 2002. New product warranty: A literature review. *International Journal of Production Economics* **79** 231–260.

Murthy, D.N.P., O. Solem, T. Roren. 2004. Product warranty logistics: Issues and challanges. *European Journal of Operational Research* **156** 110–126.

Nair, S.K. 1995. Modeling strategic investment decisions under sequential technological change. *Management Science* **41** 282–297.

Nair, S.K., W.J. Hopp. 1992. A model for equipment replacement due to technological obsolescence. *European Journal of Operational Research* **63** 207–221.

Nguyen, D.G., D.N.P. Murthy. 1988. Optimal reliability allocation for products sold under warranty. *Engineering Optimization* **13** 34–45.

Oliva, R., R. Kallenberg. 2003. Managing the transition from products to services. *International Journal of Service Industry Management* **14** 160–172.

Öner, K.B., R. Franssen, G.P. Kiesmüller, G.J. Van Houtum. 2007. Life cycle costs measurement of complex systems manufactured by an engineer-to-order company. *In R.G. Qui, D.W. Russell, W.G. Sullivan, Ahmad, M. (Eds.), The 17th International Conference on Flexible Automation and Intelligent Manufacturing* 569–589.

Öner, K.B., A. Scheller-Wolf, G.J. van Houtum. 2010a. *Redundancy Allocation for Capital Goods*. Beta Working Paper, Eindhoven University of Technology, (in preparation).

Öner, K.B., G.P. Kiesmüller, G.J. Van Houtum. 2008. *Monotonicity and Supermodularity Results for the Erlang Loss System*. Working paper 255, Beta Research School, Eindhoven University of Technology, 2008. Available at www.tm.tue.nl/beta.

Öner, K.B., G.P. Kiesmüller, G.J. van Houtum. 2009. Monotonicity and supermodularity results for the Erlang loss system. *Operations Research Letters* **37** 265–268.

Öner, K.B., G.P. Kiesmüller, G.J. van Houtum. 2010b. Optimization of component reliability in the design phase of capital goods. *European Journal of Operational Research, to appear* .

Öner, K.B., G.P. Kiesmüller, G.J. van Houtum. 2010c. *Upgrading Policy After the Redesign of a Component for Reliability Improvement*. Beta Working Paper, Eindhoven University of Technology, (in preparation).

Patterson, D. 2002. *A simple way to estimate the cost of downtime*. Last checked on

December 21, 2009.

Rahman, A., G. Chattopadhyay. 2006. Review of long term warranty policies. *Asia-Pacific Journal of Operational Research* **23** 453–472.

Rajagopalan, S., T.E. Morton M.R. Singh. 1998. Capacity expansion and replacement in growing markets with uncertain technological breakthroughs. *Management Science* **44** 12–30.

Ross, S. M. 2007. *Introduction to Probability Models*. Elsevier, USA.

Saranga, H., U. Dinesh Kumar. 2006. Optimization of aircraft maintenance/support infrastructure using genetic algorithms level of repair analysis. *Annals of Operations Research* **143** 91–106.

Sethi, S., S. Chand. 1979. Planning horizon procedures for machine replacement models. *Management Science* **25** 140–151.

Sevastyanov, B.A. 1957. An ergodic theorem for markov processes and its applications to telephone systems with refusals. *Theoretical Probability Applications* **2** 104–112.

Sharma, U., K.B. Misra. 1988. Optimal availability design of a maintained system. *Reliability Engineering and System Safety* **20** 147–159.

Sherbrooke, C.C. 2004. *Optimal Inventory Modeling of Systems: Multi-Echelon Techniques - INTL Series in Operations Research & Management Science*. 2nd ed. Kluwer, Dordrecht (the Netherlands).

Thompson, G. 1999. *Improving maintainability and reliability through design*. Professional Engineering Publishing Limited, London and Bury St Edmunds, UK.

Vintr, Z., R. Holub. 2001. R&M requirements allocation in upgrading a system. *Proceedings Annual Reliability and Maintainability Symposium* **37** 258–263.

Wells, C.E., J.L. Bryant. 1985. On the replenishment of repair kit inventories for systems with life distributions of phase type. *Management Science* **31** 1440–1450.

Yao, D.D., J.G. Shanthikumar. 1987. The optimal input rates to a system of manufacturing cells. *IFOR* **25** 57–65.

About the author

Kurtuluş Barış Öner was born in Tarsus, a city on the southern coast of Turkey, on December 11, 1977. In 2001 and 2004, he received his Bachelor's degree and Master's degree, respectively, in Industrial Engineering from the Industrial Engineering Department at Boğaziçi University, Istanbul. In 2006, he started his PhD research on optimal reliability and upgrading decisions for capital good, under the supervision of Geert-Jan van Houtum and Gudrun Kiesmüller. He cooperated with a number of companies, in particular with Philips Healthcare and Vanderlande Industries, during his research. He carried out a part of his research at Carnegie Mellon University in Pittsburgh (PA, USA) under the supervision of Alan Scheller-Wolf throughout March-June, 2009. On August 30, 2010, Kurtuluş defends his PhD thesis at Eindhoven University of Technology. From September 1 on, he is working on a temporary basis as an assistant professor within the group Operations, Planning, Accounting and Control of the same university.