Redundancy Optimization for Critical Components in High-Availability Capital Goods

Kurtulus Baris Öner, Alan Scheller-Wolf, Geert-Jan van Houtum

Beta Working Paper series 341

<table>
<thead>
<tr>
<th>BETA publication</th>
<th>WP 341 (working paper)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISBN</td>
<td>978-90-386-2457-0</td>
</tr>
<tr>
<td>ISSN</td>
<td>804</td>
</tr>
<tr>
<td>NUR</td>
<td></td>
</tr>
<tr>
<td>Eindhoven</td>
<td>February 2011</td>
</tr>
</tbody>
</table>
We consider a user who buys a number of identical capital goods systems (e.g., medical, manufacturing, or communication systems) for which she must have very high availability. In such a situation, there are typically several options that can be used to facilitate this availability. Often, the user can choose to build in cold standby redundancy for critical components. She may also typically buy spare parts with the systems so that during their exploitation phase, when a part in a system fails, the failed part can be replaced by a ready-for-use part from inventory. In addition, an emergency procedure is usually available by which a part is shipped from a distant central warehouse (at an additional cost) to be applied when there is a stock out. To these options we introduce another: The possibility of initiating an emergency shipment when stock is one. Thus, the user may choose one of three policies per component: The different combinations of the redundancy decision and the timing of applications of the emergency procedure. (In addition, she must decide how much spare parts inventory to purchase, for any policy.) Each policy provides different total uptime against different total costs incurred.

We formulate the problem as the minimization of the total costs incurred for the systems over their lifetimes, under a constraint for the total uptime of all systems. These total costs consist of acquisition costs, spare parts costs, and repair costs. We optimally solve the problem by decomposing the multi-component problem into single-component problems, and then conducting exact analysis on these single-component problems, we derive results on when each of the three policies is optimal. Using these, we construct an efficient frontier which reflects the trade-off between the uptime and the total costs of the systems. In addition, we provide a method to rank the components by the relative value of investing in redundancy. We illustrate these results through numerical examples.

Key words: Reliability optimization, Total Cost of Ownership, spare parts
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, power generators) form the backbone of much of our society: They are essential for operational continuity in hospitals, airports, factories, banks, computer networks, power plants, etc. Interruptions of these systems lead to significant losses; for example, downtime costs of computer systems of large e-commerce companies and brokerage companies can be $100,000-$1,000,000 per hour (Patterson, 2002, cnet news, 2001); in general, the opportunity costs due to downtime of bottleneck machines in factories are very high, like downtime of step&scan systems in semiconductor companies resulting in millions of Euros (Kranenburg and van Houtum, 2009). Because of these very high costs, extensive maintenance activities are carried out for these systems; taken together, downtime costs and maintenance costs of such systems may account for 70-80% of their Total Cost of Ownership (TCO - the total costs incurred throughout their lifetime); see Öner et al. (2007), Saranga and Dinesh Kumar (2006).

As a result, after-sales service has evolved into an important 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). While approaches have been proposed for improving profitability in the after-sales market (see Cohen et al., 2006) and quantitative models have been developed for service supply chain management (see Muckstadt, 2005), a large portion of a system’s TCO is determined upstream from these decisions, during the system design phase. For example, redundancy (i.e., having a number of identical components in parallel instead of a single component) is a primary design decision affecting the availability and TCO of a system. Surprisingly, such redundancy decisions have never been studied for capital goods while also taking the spare parts inventory and service operations procedures into account. This represents a significant unutilized opportunity for capital goods companies, and provides the setting for this paper.

In general, a high level of system availability (uptime) can be provided by infrequent system failures, and/or rapid system repair (service) activities (yielding a short downtime per system failure). In practice, designing capital goods with redundancy and using the repair-by-replacement concept during the exploitation phase are common to achieve high availability. If using redundancy, once a part fails, its functionality is taken over by a redundant part and the failed part is replaced with a ready-for-use one in a short time; hence, the probability of a system failure becomes negligible. Without redundancy, the repair-by-replacement concept - utilizing spare parts kept on stock
close-by to the installed system - increases the speed of system repair activities (and decreases the
downtime per system failure).

Under the repair-by-replacement concept, the activities that are executed upon a failure of a part
depend primarily on the status of the spare parts supply. If there is a ready-for-use part available
from 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 such a replacement with a part
from inventory as an ordinary procedure. If the system fails when no spare parts are in inventory,
an emergency procedure is carried out to replace the failed part; that is, other means of supply
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. Such an emergency procedure could also be
applied pro-actively: When a failure occurs and inventory is one.

This emergency procedure can be an important tool in controlling downtime costs. 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.

In this work, we focus on the redundancy optimization for critical components in a capital good:
if one of the critical components in a system fails, the system also fails (when there is no redundant
component). We investigate a situation in which a user buys a number of identical systems seeking
to minimize TCO subject to an overall availability constraint. The user may choose to have at
most two identical parts in a cold standby redundancy setting per component; that is, one part
of a component is active and if there is a redundant part, it immediately becomes active when
the active part fails. In addition, the user may buy spare parts together with the systems. Upon
failures, she applies the ordinary procedure and the emergency procedure in the exploitation phase.
Should she choose to execute the emergency procedure when stock-on-hand is one, we refer to this
as the provisional procedure.

We compare the costs of the following three policies per component (assuming each is executed
by an optimal spare parts inventory which we determine):

1. Policy $(0,0)$ - Do not choose redundancy and apply the emergency procedure when a failure
occurs and there is a stock out.

2. Policy $(0,1)$ - Do not choose redundancy and apply the provisional 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 a stock out.

These policies provide different total uptime against different TCO. If Policy (1,0) is chosen for a component, it is highly likely that 100% availability of the component is attained. As a result, choosing redundancy and applying the provisional procedure for a component - Policy (1,1) - does not provide any further advantage in terms of availability.

The contributions of this paper are:

- First, we develop a model to minimize the TCO (acquisition costs, spare parts costs, and repair costs) of a number of identical systems subject to an availability constraint. This model includes the three different policies per component explained above. Our model explicitly relates the redundancy decision and spare parts inventory level of each component to the costs and the downtime of the systems.

- Second, we decompose the multi-component problem into single-component problems by using Lagrangian relaxation. The multi-component problem has a combinatorial nature, as any of the three policies can be applied per component. Our decomposition enables the generation of optimal solutions for the multi-component problem efficiently, without considering all possible combinations. We then develop an efficient optimization procedure which can be used to solve any of the single-component problems for varying levels of the downtime constraint.

- Third, we compare the three policies at the single-component level and provide conditions under which one policy outperforms the others. When the downtime constraint is loose, Policy (0,0) is optimal. As the constraint becomes tight (i.e., the level of the downtime constraint is decreased), one of the following two cases occurs depending on the values of the problem parameters:
  
  — Policy (1,0) - redundancy - becomes optimal after a certain level, say $D_1$, of the constraint and remains optimal for all smaller levels afterwards.
  
  — Policy (0,1) becomes optimal after a certain level, say $D_2$, of the constraint. As the constraint is further tightened, it remains optimal until another level, say $D_3 < D_2$, at which Policy (1,0) becomes optimal. Policy (1,0) remains optimal for all smaller levels afterwards.

Furthermore, we show that the values when the optimal policy changes from one to the other can be easily computed.

- Fourth, we provide the following multi-component results:

  — We introduce a method to construct an efficient frontier which reflects the trade-off between the uptime and the TCO. These show that when the provisional procedure is allowed, similar availability levels can be attained at lower TCO values than when it is excluded.
— We provide a method for ranking of the components in a capital good to reflect the benefits of investing in redundancy. In many cases, one is interested in finding the optimal order to implement redundancy for components, while making one-by-one decisions during the design due to other considerations (e.g. a budget limit).

This paper is organized as follows. We summarize the literature related to our model in Section 2. In Section 3, we present our model assumptions. Next, we formulate our problem, derive the cost functions and decompose the multi-component problem into single-component problems by using Lagrangian relaxation, in Section 4. In Section 5, we compare the three policies on the single-component level and then provide the multi-component results. We finalize the paper by drawing conclusions in Section 6.

2. Literature

Our model fits within the area of system reliability optimization. There exists a large number of papers in this area (see review papers by Kuo and Prasad, 2000, and Kuo and Wan, 2007). In this paper, we introduce a model for a reliability optimization problem which has not been studied in the literature, as we jointly consider the related but different decisions of redundancy and spare parts jointly under a more comprehensive and realistic replenishment and cost model than previously seen.

Cold standby redundancy can be considered a special strategy of keeping spare parts inventory. When a single system is considered, a redundancy allocation model is equivalent to a spare parts inventory model under certain assumptions; 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 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. Sharma and Misra (1988) consider redundancy and spare parts jointly for a single system with subsystems in a serial structure. The decision variables are redundancy level (the number of parts in parallel), the number of spare parts to be bought per component 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.
The models introduced in Sharma and Misra (1988), and Misra and Sharma (1991) have several limitations: First, these models do not involve any emergency procedures; when a part fails and there is no ready-for-use part available, the system is down until a part is available from a repair facility. Second, the cost factors in these papers are limited to acquisition costs or design and production costs, similarly to all redundancy optimization models in the literature. However, maintenance costs (repair and spare parts costs) account for a large portion of TCO of many capital goods; see Öner et al. (2007) for two real-life cases in which maintenance costs are at least as large as acquisition costs. Third, these models are for a single system. In practice, spare parts are usually stocked for multiple systems at a central location leading to a pooling effect not captured in these models.

Quantitative models that incorporate maintenance costs, repair costs in particular, exist primarily in the warranty literature (see the review in Murthy and Blischke, 1992). This research primarily focuses on the optimal length of the warranty period, but there are also papers on reliability optimization: Nguyen and Murthy (1988), Hussain and Murthy (2003), Huang et al. (2007), Hussain and Murthy (1998), and Monga and Zuo (1998). But in all these models, it is assumed that ready-for-use replacement parts are always available and spare parts inventory is not incorporated.

Kim et al. (2007, 2010) do both study spare parts inventory and reliability; but only of a single-component system and in game-theoretic settings. As the authors’ objective is to derive high level managerial insights about different service contract types, they develop stylized models which do not explicitly incorporate redundancy. Likewise, Öner et al. (2010) introduce a model that jointly optimizes component reliability and spare parts inventory, but again for a single component system. They also include an emergency procedure, but in their work reliability enhancement is achieved by improving the reliability of the component itself rather than by using redundancy. Thus our work generalizes theirs in modeling multi-component systems, including the option of redundancy, and incorporation of the provisional procedure.

We provide a comparison of the model in this paper to the models existing in the literature in Table 1. To the best of our knowledge, we introduce and study the provisional procedure in the literature (including the spare parts inventory literature) for the first time.

3. Model

3.1. Terminology

Terms used during system design represent abstract concepts as a physical system does not exist yet. Often, the same terms are also used for the concrete counterparts of those concepts after the design.
Table 1  Comparison of papers

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Redundancy</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Maintenance costs</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Multi-component</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Multiple systems</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Spare parts</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Emergency Procedure</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Provisional Procedure</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Our problem includes both abstract concepts and their concrete counterparts. To differentiate these, in the remainder of this paper, we use *capital good* as an abstract term and *system* as its concrete counterpart. The terms *component* and *part* may be abstract or concrete, depending on the context. The relation between a component and a part is as follows: a component consists of either a single part or two identical parts, depending on the redundancy decision.

### 3.2. Model Description

A capital good is being designed by an Original Equipment Manufacturer (OEM) for a user. The user will buy $N$ ($N \in \mathbb{N} = \{1, 2, \ldots \}$) systems. We assume that the $N$ systems will start operating at the same time, be operating $7 \times 24$ (without any breaks); 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).

The capital good includes $m$ ($m \in \mathbb{N}$) critical components. We index components with $i$, $i \in M = \{1, 2, \ldots, m\}$ and refer to a part of component $i$ as a *component-i part*. Each component includes one or two parts, the latter case corresponding to a cold standby redundancy setting.

The user may keep spare parts inventory for each component at a single stock point. She buys $s_i$, component-$i$ spare parts together with the systems at time 0; i.e., $s_i$ is the initial supply amount for component-$i$. There is a single repair facility for defective parts. The user also contracts with the OEM for replenishment of ready-for-use parts as soon as possible via an emergency transportation mode (e.g. by plane) in case of need. We assume that the OEM has ample supply of the parts.
3.2.1. Failure and Repair Processes  We denote the Mean Time Between Failure (MTBF) of a component-$i$ part by $\tau_i$. The $\tau_i$, $i \in M$, are typically in the order of 1-10 years, and known by the user. When a system is up, the failures of parts in the system are mutually independent. When the system is down due to a failure of one of its parts, the system is shut-down and other parts in the system do not fail. We make the simplifying assumption that the stream of failures of component-$i$ parts follows a Poisson process with the constant rate $N\tau_i^{-1}$ throughout $[0,T]$, even though when a system is down, the failure rates of all parts decreases as there is one system less contributing for the total stream of failures. However, owing to the short downtimes (and possibly large number of systems), we neglect this effect, as is standard in the spare parts inventory literature. This simplifies the analysis considerably and has been demonstrated to be a benign assumption; see Muckstadt (2005) and Sherbrooke (2004).

Upon the failure of a component-$i$ part at time $t \in [0,T]$, either the ordinary procedure, the provisional procedure, or the emergency procedure will be applied. Regardless of the redundancy decision of the component, the procedure that is applied will depend only on a predetermined threshold value, $z_i \geq 0$, and the actual stock-on-hand $H_i(t)$:

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. **Provisional procedure**: If $0 < H_i(t) \leq z_i$, the failed part is replaced with a ready-for-use part from the inventory. An as-good-as-new component-$i$ part is replenished directly from the OEM and added to the inventory. The defective part is returned to the OEM.

3. **Emergency Procedure**: If $H_i(t) = 0$, an as-good-as-new component-$i$ part is replenished directly from the OEM and transported to the location of the failure. The failed part is replaced with the replenished part and returned to the OEM.

We assume that when a part fails, it will be diagnosed with 100% accuracy in a negligibly short time. If the failure occurs in a component with a single part, the component becomes inoperable until just after the replacement of the failed part; the downtime is equal to the replacement time of the part. We call a replacement by either the ordinary procedure or the provisional procedure an *inventory-replacement* and a replacement by the emergency procedure an *OEM-replacement*. 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 $\mu_{2,i}$. Typically, $\mu_{1,i}$ 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 close to the systems (possibly even at the same site). The emergency
procedure assures that downtimes are short even in stock outs. The provisional procedure serves as a preventive action to avoid the longer downtimes that would arise in stock out in which OEM-replacements take place.

We assume that repair times of component-$i$ parts, which include transporting the part to and from the repair facility, are independent and identically distributed with mean $U_i > 0$ (typically in the order of 1-4 months). This independence in some sense implies ample repair capacity, or a repair lead time per part subject to a specified agreement. The orders of magnitude imply that $\mu_{2,i}$ is very small compared to $U_i$, reflecting the user’s incentive to apply the emergency procedure. Under the provisional procedure, the part which is replenished is added to inventory after a random lead time with mean $\mu_{3,i}$. Again, $\mu_{3,i}$ is in the order of 1-48 hours (the distance between the OEM’s site and the stock point is comparable to the distance between the OEM’s site and of the systems).

Upon a failure, the spare parts inventory of component $i$ is affected as follows: If the ordinary procedure or the provisional procedure is applied, a part is removed from the inventory, and added back 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 back. Thus the spare parts inventory is controlled by a continuous-review basestock policy with basestock level $s_i$ and lost sales, implying the inventory position (sum of pipeline stock and stock-on-hand) of component $i$ is always equal to $s_i$.

3.2.2. Policies & Downtimes For each component $i$, $i \in M$, the user can implement one of the three policies defined as couples $(y_i, z_i)$, as follows:

1. Policy $(0,0)$ - Do not choose redundancy ($y_i = 0$) and apply the emergency procedure when a failure occurs and there is a stock out ($z_i = 0$). Under Policy $(0,0)$, downtimes arise during both inventory-replacements and OEM-replacements.

2. Policy $(0,1)$ - Do not choose redundancy ($y_i = 0$) and apply the provisional procedure when a failure occurs and stock-on-hand is 1 ($z_i = 1$). As we assume that $\mu_{3,i} < \frac{\tau_i}{N}$, under Policy $(0,1)$, if a component-$i$ part fails at time $t$ and the provisional procedure is applied, the probability that a component-$i$ part in another system would fail during the replenishment lead time is negligibly small. Hence, we assume that no failure of component-$i$ parts occurs during the replenishment lead time. As a result, the emergency procedure is never applied under Policy $(0,1)$ and downtimes arise only during inventory-replacements.

3. Policy $(1,0)$ - Choose redundancy ($y_i = 1$) and apply the emergency procedure when a failure occurs and there is a stock out ($z_i = 0$). Under Policy $(1,0)$, we assume that 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 $\mu_{1,i} \leq \mu_{2,i} \ll \tau_i$, the failed part will be replaced in a negligibly short time compared to the MTBF of the standby part. Hence, we assume that redundancy yields 100% availability of the component, and limit the redundancy setting to two parts.

Notice that Policy $(0,1)$ requires $s_i \geq 1$, and the number of policies on the system level is $3^m$.

**3.2.3. Cost Factors** Our objective is the minimization of the portions of TCO affected by redundancy, replenishment and inventory decisions: The acquisition costs, the spare parts costs, and the repair costs stemming from the applications of the three procedures. We assume the acquisition costs of the $N$ systems and spare parts costs 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, using a constant discount rate by $\alpha > 0$. We use the following notation for our cost parameters:

- $c_{0,i}$: Unit acquisition cost per component-$i$ spare part during the initial supply.
- $c_{1,i}$: Extra cost incurred for building in redundancy for component $i$ in a system
- $h_i$: The storage cost rate per spare part of component $i$ ($h_i > 0 \forall i \in M$)
- $r_{1,i}$: Expected costs incurred per application of the ordinary procedure for a component-$i$ part ($r_{1,i} > 0 \forall i \in M$).
- $r_{2,i}$: Expected costs incurred per application of the emergency procedure or the provisional procedure for a component-$i$ part, discussed further 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 provisional 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 from 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 (essentially) equal as the stock point is at a close distance to the systems.

The factors $r_{1,i}$ and $r_{2,i}$ include all costs originating from an application of the ordinary or emergency/provisional procedure for a component-$i$ part: administrative costs, costs of a visit of a service engineer, transportation costs, repair costs of a failed component-$i$ part (for ordinary) and replenishment costs of a component-$i$ part from the OEM (for emergency/provisional). We assume that $r_{2,i} \geq r_{1,i}$; generally, $r_{2,i}$ will be much larger than $r_{1,i}$.

The effect of redundancy on acquisition cost of a component appears as an extra cost ($c_{1,i}$) above the acquisition cost without redundancy which we treat as a fixed cost. This is also why $c_{0,i}$ is defined as the unit acquisition cost of a component-$i$ spare part. The spare parts costs includes this
spare parts investment costs, $c_{0,i}$, and the spare parts storage costs, which we assume depend on the initial supply amount $(s_i)$, not on actual stock-on-hand (and thus are constant during $[0,T]$). We also assume that the spare parts stock-on-hand process of each component is in steady state from time 0.

4. Problem Formulation

In this section we first give the multi-component problem formulation and several preliminary results. Then we derive expressions for the cost and downtime functions which appear in the problem formulation. We finalize the section by decomposing the multi-component problem into single-component problems by Lagrangian relaxation.

Our problem formulation is as follows:

$$(Q0) \quad \min \ \pi(y, z, s)$$

s.t. $D(y, z, s) \leq D_0$

$$(y_i, z_i) \in \{(0, 0), (0, 1), (1, 0)\} \text{ for all } i \in M$$

$s_i \in \{z_i, z_i + 1, z_i + 2, \ldots\} \text{ for all } i \in M,$

where $y$, $z$, and $s$ are the vectors of $y_i$, $z_i$, and $s_i$, respectively, $i \in M$. Above, $\pi(y, z, s)$ is the expected NPV of TCO of the $N$ systems, $D(y, z, s)$ is the expected downtime of all $N$ systems throughout $[0,T]$ (in system-years), and $D_0 = (1 - p)NT$ is the maximum downtime that can be tolerated to ensure a percentage uptime of at least $p \in (0, 1]$.

The notation for the explicit cost functions and expressions for downtime per stage is:

$P_i(y_i)$: The expected NPV of additional acquisition costs of the $N$ systems stemming from building in redundancy for component $i$

$S_{1,i}(s_i)$: The expected NPV of spare parts investment costs for component $i$

$S_{2,i}(s_i)$: The expected NPV of spare parts storage costs for 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 component-$i$ parts throughout $[0,T]$.

$\pi_i(y_i, z_i, s_i)$: The expected NPV of total costs of component $i$.

$\pi_i(y_i, z_i, s_i) = P_i(y_i) + S_i(s_i) + R_i(z_i, s_i)$

$D_i(y_i, z_i, s_i)$: The expected total downtime of the $N$ systems stemming from failures of component-$i$ parts throughout $[0,T]$. 

Obviously, $\pi(y, z, s) = \sum_{i=1}^{m} \pi_i(y_i, z_i, s_i)$. Furthermore, as we assume downtimes of different components in the system do not overlap, $D(y, z, s) = \sum_{i=1}^{m} D_i(y_i, z_i, s_i)$.

4.1. Preliminary Results

The acquisition costs of component-$i$ parts only depend on the redundancy decision of component $i$. Furthermore, the mutual independence of failures of different components leads to independence of spare parts usage of different components. 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 $i \in M$.

Moreover, since the objective function and the constraint of problem (Q0) are linear combinations of objectives and constraints per component, (Q0) is separable. We make use of this separability to decompose the multi-component problem into single-component problems by the Lagrangian relaxation method in the next subsection. We continue this section with the derivation of the single-component cost and downtime functions.

Storage Costs. As we assume that the acquisition costs of the $N$ systems and spare parts are incurred at time $0$, $P_i(y_i) = N c_{1,i} y_i$, and $S_{1,i}(s_i) = c_{0,i} s_i$. The expected NPV of spare parts storage costs throughout $[0, T]$ is

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

Hence,

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

Repair Costs and Downtime. To derive the repair costs and downtime stemming from failures of component-$i$ parts, we need to identify how often each of the repair procedures (ordinary, provisional, emergency) is applied. These 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 $N/\tau_i$. 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 $U_i$. Otherwise it is “lost”. Therefore:

- Under Policy $(0,0)$ and Policy $(1,0)$, the emergency procedure is applied if $H_i(t) = 0$. Hence, the stock-on-hand process is stochastically equivalent 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 $N/\tau_i$, mean service time $U_i$, and $s_i$ servers.
• Under Policy (0,1), the provisional procedure is applied if $H_i(t) = 1$. As we assume that $\mu_{a,i}$ is very small compared to $\tau_i/N$, no failures of component-$i$ parts occur during this replenishment lead time. Thus $H_i(u) \geq 1$ for all $u \in [0,T]$. Hence, $H_i(u) = 1 + \tilde{H}_i(u)$ where $\{\tilde{H}_i(u) : u \in [0,T]\}$ is a process stochastically equivalent to the process for the number of free servers in an Erlang loss system with an arrival rate $N/\tau_i$, mean service time $U_i$, and $s_i - 1$ servers.

Therefore, the emergency (provisional) procedures are applied if $H_i(t) = 0$ ($\tilde{H}_i(t) = 0$), which are events with probabilities equal to the Erlang loss probabilities $B_i(s_i)$ and $B_i(s_i - 1)$:

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

By using this result, we first derive the distribution of the numbers of applications of the ordinary procedure and of the emergency procedure for component-$i$ parts throughout $[0,T]$. Next, we derive the expected NPV of repair costs incurred for component-$i$ parts, $R_i(z_i,s_i)$, and the expected downtime stemming from failures of the component-$i$ parts throughout $[0,T]$, $D_i(y_i,z_i,s_i)$.

**PROPERTY 1.** For all $i \in M$, the following hold.

(i) Under Policy $(0,0)$ and Policy $(1,0)$, the numbers of applications of the ordinary procedure and 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 $(0,1)$, the numbers of applications of the ordinary procedure and the provisional 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 component-$i$ parts throughout $[0,T]$ is:

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

(iv) The expected downtime due to failures of component-$i$ parts throughout $[0,T]$ is:

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

**Proof.** See Appendix. □

### 4.2. Reformulation and Reinterpretation of Cost Functions

We reconstruct the storage cost function $S_{2,i}(s_i)$ in equation (1) and repair costs function $R_i(z_i,s_i)$ in equation (2) to simplify their representation and interpretation. Let

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

\[
\text{where for each component } i, \text{ we can rewrite the Lagrangian as}
\]

\[
L_i(y_i, z_i, s_i, \lambda) = \sum_{i=1}^{m} \pi_i(y_i, z_i, s_i) + \lambda \left( \sum_{i=1}^{m} D_i(y_i, z_i, s_i) - D_0 \right),
\]

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

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

where for each component \( i \) \( L_i(y_i, z_i, s_i, \lambda) \) is the decentralized Lagrangian

\[
L_i(y_i, z_i, s_i, \lambda) = \pi_i(y_i, z_i, s_i) + \lambda D_i(y_i, z_i, s_i).
\]

Observe that these decentralized Lagrangian functions are connected to each other through a single Lagrange multiplier (\( \lambda \)), as there is only one constraint in our problem.

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 (Q\(_i\)(\(\lambda\))) given as

\[
(Q_i(\lambda)) \quad \min \pi_i(y_i, z_i, s_i) \\
\text{s.t. } D_i(y_i, z_i, s_i) \leq D_i(y^*_i(\lambda), z^*_i(\lambda), s^*_i(\lambda)) \\
(y_i, z_i) \in \{(0, 0), (0, 1), (1, 0)\} \\
s_i \{z_i, z_i + 1, z_i + 2, \ldots \},
\]

and \((y^*_i(\lambda), z^*_i(\lambda), s^*_i(\lambda))\) satisfies the downtime constraint in Problem (Q\(_i\)(\(\lambda\))) with equality. Furthermore, the vectors \((y^*(\lambda), z^*(\lambda), s^*(\lambda))\) will also be a solution for the system-level problem (Q(\(\lambda\))):
and will likewise satisfy that downtime constraint with equality. Thus, by using various values of \( \lambda \), 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 general system-level problem (Q1) given as

\[
(Q1) \quad \min_{y, z, s} \pi(y, z, s) \\
\text{s.t.} \quad D(y, z, s) \leq D(y^*(\lambda), z^*(\lambda), s^*(\lambda)) \\
(y_i, z_i) \in \{(0,0), (0,1), (1,0)\} \quad \text{for all } i \in M \\
s_i \in \{z_i, z_i + 1, z_i + 2, ...\} \quad \text{for all } i \in M.
\]

These solutions comprise an efficient frontier for the TCO vs. total downtime. From this efficient frontier, an appropriate solution for Problem (Q0) may be selected.

By Theorem 2 in Everett (1963), \( D_i(y_i^*(\lambda), z_i^*(\lambda), s_i^*(\lambda)) \) in Problem (Q\( _i \)) is decreasing in \( \lambda \) for all \( i \in M \). A direct result of this property is that \( D_i(y^*(\lambda), z^*(\lambda), s^*(\lambda)) \) in Problem (Q(\( \lambda \)) will also be decreasing in \( \lambda \). We will make use of this property, and the standard interpretation of \( \lambda \) as a downtime penalty rate, to compare our three policies - (0,0), (0,1), and (1,0) - in Subsection 5.2, and to derive results for the system level problem in Subsection 5.3.

5. Analysis

In this section we first provide a number of results for the optimization of the single-component problems. Then we compare the three policies at the single-component level. Next, we introduce a method to construct an efficient frontier for the multi-component problem, and use it to demonstrate the benefit of the provisional procedure by comparing efficient frontiers of cases when it is included and excluded. We conclude by providing an optimal ordering of the components in a capital good with respect to the relative value of investing in redundancy. The proofs of a number of results appear in the section, while the others are given in the Appendix.

5.1. Optimization of the Single-Component Problems

In this subsection, we derive three lemmas: Lemma 1 states that \( L_i(y_i, z_i, s_i, \lambda) \) is strictly convex in \( s_i \) for a given policy \( (y_i, z_i) \) and a given value of \( \lambda \). We introduce a number of properties of \( L_i(y_i, z_i, s_i, \lambda) \) and the optimal \( s_i \) value, for a given policy, in Lemma 2. Then, we detail an
optimization procedure for $L_i(y_i, z_i, s_i, \lambda)$ for a given value of $\lambda$, in Lemma 3. An optimal solution for $L(y, z, s, \lambda)$ can then be found by generating optimal solutions of $L_i(y_i, z_i, s_i, \lambda)$ for all $i \in M$ via Lemma 3.

**Lemma 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 $s_i$.

We define $s_i^*(y_i, z_i, \lambda) = \arg \min_{s_i} \{L_i(y_i, z_i, s_i, \lambda) | s_i \in \{z_i, z_i + 1, z_i + 2, \ldots\}\}$; for a given value of $\lambda$, $s_i^*(0,0,0,\lambda)$, $s_i^*(0,1,0,\lambda)$, and $s_i^*(1,0,0,\lambda)$ 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, respectively. We also define $\Delta B_i(s_i) = B_i(s_i) - B_i(s_i + 1)$; $\Delta B_i(s_i)$ is strictly positive for all $s_i$. Also, as $B_i(s_i)$ is strictly convex in $s_i$, $\Delta B_i(s_i)$ is strictly decreasing in $s_i$.

**Lemma 2.** For all $i \in M$, the optimal spare parts levels and costs satisfy:

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

$$s_i^*(1,0) = \min \left\{ s_i \in \mathbb{N}_0 = \{0,1,2,\ldots\} \mid \Delta B_i(s_i) \leq \frac{(c_{0,i} + \hat{h}_iT) \tau_i}{(\hat{r}_{2,i} - \hat{r}_{1,i}) NT} \right\}. \ (5)$$

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

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

$$s_i^*(0,0,\lambda) = \min \left\{ s_i \in \mathbb{N}_0 \mid \Delta B_i(s_i) \leq \frac{c_{0,i} + \hat{h}_iT}{\hat{r}_{2,i} - \hat{r}_{1,i} + (\mu_{2,i} - \mu_{1,i})\lambda NT} \right\}. \ (6)$$

is increasing in $\lambda$.

We illustrate the properties through an example:

**Example 1.** Consider the redundancy allocation problem for a capital good with two components, $N = 15$ systems purchased, and an expected lifetime of $T = 15$ years. The annual discount rate is $\alpha = 0.05$. The cost, failure and repair parameters for the components are given in Table 2 (we will also refer to this table in later examples).

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


**Table 2** Parameters for Example 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Component 1</th>
<th>Component 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau$ (years)</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>$c_{0,i}$ (Euros)</td>
<td>5000</td>
<td>125000</td>
</tr>
<tr>
<td>$c_{1,i}$ (Euros)</td>
<td>4000</td>
<td>125000</td>
</tr>
<tr>
<td>$h_i$ (Euros per month)</td>
<td>75</td>
<td>1875</td>
</tr>
<tr>
<td>$r_{1,i}$ (Euros)</td>
<td>1000</td>
<td>25000</td>
</tr>
<tr>
<td>$r_{2,i}$ (Euros)</td>
<td>2000</td>
<td>50000</td>
</tr>
<tr>
<td>$\mu_{1,i}$ (hours)</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>$\mu_{2,i}$ (hours)</td>
<td>24</td>
<td>48</td>
</tr>
<tr>
<td>$U_i$ (months)</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Below, we provide intuitions following from Lemma 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 ($\lambda$) does not affect this policy’s optimal number of spare parts or the optimal total costs.

(ii) In Policy (0,1) all failures result in downtimes equal to the inventory-replacement time, independent of $s_i$. Hence, the downtime penalty rate does not affect the optimal number of spare parts. However, it still 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 affected by $s_i$. Hence, the optimal $s_i$ changes with the downtime penalty rate: Increasing the downtime penalty rate results in increasing downtime costs, leading to a higher number of spare parts.

![Lagrangian of component 1 for Policy (0,0) in Example 1](image-url)
COROLLARY 1. For all \( i \in M \), at the optimal spare parts inventory level for each policy:

(i) 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 Policy (0,1) and Policy (1,0), and the provisional procedure and the emergency procedure are applied for the same failures under Policy (0,1) and Policy (1,0), respectively.

(ii) The expected repair costs for component-\( i \) parts under Policy (0,1) and Policy (1,0) are equal.

Proof. (i) This is a direct result of the relationship between the optimal initial supply amounts, \( s^*_i(0,1,\lambda) = s^*_i(1,0) + 1 \) for all \( \lambda \geq 0 \).

(ii) This result directly follows from part (i). □

LEMMA 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 (5).
2. Determine \( s^*_i(0,0,\lambda) \) by equation (6).
3. Find

\[
(y^*_i(\lambda), z^*_i(\lambda), s^*_i(\lambda)) = \arg \min_{(y_i, z_i, s_i)} \left\{ L_i(y_i, z_i, s_i, \lambda) \mid (y_i, z_i, s_i) \in \left\{ (0,0, s^*_i(0,0,\lambda)), (0,1, s^*_i(1,0) + 1), (1,0, s^*_i(1,0)) \right\} \right\}.
\]

We do not elaborate on the proof of Lemma 3 as it is trivial.

5.2. Comparison of the Policies

In this subsection, we will first make pairwise comparisons of the three policies for single-component problems. Then, we will give an overall comparison.

5.2.1. Policy (0,1) versus Policy (1,0)

LEMMA 4. For all \( i \in M \),

(i) If \( N_{c,1,i} \leq c_{0,i} + \hat{h}_i T \), then Policy (1,0) outperforms Policy (0,1) for all \( \lambda \geq 0 \).

(ii) If \( N_{c,1,i} > c_{0,i} + \hat{h}_i T \), then Policy (0,1) outperforms Policy (1,0) for \( 0 \leq \lambda \leq \lambda_{01-10,i} = \frac{T}{N_{c,1,i} - N_{c,1,i} - \hat{h}_i T} \); and Policy (1,0) outperforms Policy (0,1) for \( \lambda > \lambda_{01-10,i} \).

EXAMPLE 2. Consider the redundancy allocation problem introduced in Example 1 with \( N = 15 \), \( T = 15 \) years, \( \alpha = 0.05 \) annually, and the parameters given in Table 2. Figure 2 compares Policy (0,1) and Policy (1,0) for component 1. The dashed line is the Lagrangian function
Lagrangian of component 1 for Policy (0,1) and Policy (1,0) (Euros)

min \{L_{1}(1,0), L_{1}(0,1,\lambda)\} \geq 0.1-10,1

Figure 2 Comparison of Policy (0,1) and Policy (1,0) for component 1 in Example 2

L_{1}(0,1,\lambda) = L_{1}(0,1,s^*_i(1,0) + 1,\lambda), where s^*_i(1,0) = 2. The dash-dot line (-.) is the Lagrangian function L_{1}(1,0,s^*_i(1,0),\lambda) = L_{1}(1,0). The solid line is the minimum of the Lagrangian functions 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 = 60.67 Euros/hour, 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 in Lemma 4 can be interpreted as follows:

(i) \(N_{c_{1,i}}\) is the extra investment in the systems under Policy (1,0) compared to Policy (0,1). Due to the relationship between the optimal initial supply amounts in the two policies \(s^*_i(0,1,\lambda) = s^*_i(1,0) + 1\), Policy (0,1) procures an extra component-\(i\) spare part compared to Policy (1,0) and \(c_{0,i} + \hat{h}_iT\) is the extra investment in spare parts under Policy (0,1). The repair costs are equal under these two policies. Thus, if the extra investment under Policy (1,0) is less than the extra investment under Policy (0,1), Policy (1,0) always outperforms Policy (0,1) as Policy (1,0) provides higher availability than Policy (0,1).

(ii) If extra investment under Policy (1,0) is greater than the extra investment under Policy (0,1), the comparison of these policies depend on the downtime constraint. If the downtime constraint is relatively loose, Policy (0,1) satisfies it with lower TCO. As it gets tighter and tighter, it is met by Policy (1,0) with lower TCO after a certain value of the resource level of the downtime constraint.

In Corollary 2 we provide the sensitivity of \(\lambda_{01-10,1}\) to each model parameter in case \(N_{c_{1,i}} > c_{0,i} + \hat{h}_iT\), which is highly likely if \(N >> 1\). As an explicit, relatively simple formula for \(\lambda_{01-10,1}\) can...
be derived, the proof of Corollary 2 is omitted.

**Corollary 2.** For all \( i \in M \), if \( Nc_{1,i} > c_{0,i} + \hat{h}_i T \), \( \lambda_{01-10,i} \) is

(i) independent of \( \mu_{2,i} \), \( r_{i,1} \), \( r_{i,2} \) and \( U_i \);

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

(iii) increasing as a function of \( \tau_i \), \( c_{1,i} \), and \( N \).

Below, we provide intuitive explanations for Corollary 2:

(i) The result that \( \lambda_{01-10,i} \) is independent of \( \mu_{2,i} \) is expected, as downtime stemming from OEM-replacements does not occur in either of the policies. However, its independence of \( r_{i,1} \), \( r_{i,2} \), and \( U_i \) is not trivial. These parameters affect the repair costs in both policies: \( r_{i,1} \) and \( r_{i,2} \) affect them directly while the effect of \( U_i \) is through the Erlang loss probability. One would expect these parameters to play a role in the comparison of the two policies. Nevertheless, the results given in Corollary 1, “coupling” the repairs under policies (0,1) and (1,0) eliminate their effects in the comparison.

(ii) As Policy (0,1) procures an extra component-\( i \) spare part compared to Policy (1,0) and the procurement cost and the storage cost rate of this extra part is \( c_{0,i} \) and \( h_i \), respectively, decreasing \( c_{0,i} \) and/or \( h_i \) favors Policy (0,1). Under Policy (0,1), each failure leads to an expected downtime of \( \mu_{1,i} \) while there is no downtime under Policy (1,0); thus, decreasing \( \mu_{1,i} \) favors Policy (0,1). Considering costs as a function of \( T \), Policy (0,1) incurs \( (c_{0,i} + \hat{h}_i T) + \frac{NT}{\tau_i} \lambda_i \mu_{1,i} - Nc_1 \) units (e.g. Euros) more than Policy (1,0), meaning a decrease in \( T \) favors Policy (0,1).

(iii) \( \tau_i \) affects only the total expected downtime under Policy (0,1); as the total expected downtime under Policy (0,1) decreases as a function of \( \tau_i \), an increase in \( \tau_i \) favors Policy (0,1). Similarly, an increase in \( c_{1,i} \) favors Policy (0,1) as it only increases the extra costs incurred for redundancy under Policy (1,0). The effect of \( N \) can be attributed to economies of scale: The only cost factor independent of \( N \) and \( \lambda_i \), and different in the two policies is \( (c_{0,i} + \hat{h}_i T) \). This factor acts as a fixed cost under Policy (0,1). An increase in \( N \) favors Policy (0,1) as the fixed cost per system decreases.

**5.2.2. Policy (0,0) versus Policy (1,0)**

**Lemma 5.** For all \( i \in M \), the followings hold:

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

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

(iii) There exists a \( \lambda_{00-10,i} \), \( c_{1,i} \left( \frac{T}{r_{i,1}} \mu_{2,i} \right)^{-1} \leq \lambda_{00-10,i} \leq c_{1,i} \left( \frac{T}{r_{i,1}} \mu_{1,i} \right)^{-1} \), 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 \( \lambda > \lambda_{00-10,i} \).
A closed form expression cannot be derived for $\lambda_{00-10,i}^{-1}$, but it can be found by simple numerical procedures as $\lambda_{00-10,i}^{-1}$ is the value $\lambda$ at which two linear functions intersect.

**Example 3.** Consider the problem introduced in Example 1. Figure 3 compares Policy (0,0) and Policy (1,0) for component 1. The dotted lines are the Lagrangian functions $L_i(0,0,s,\lambda)$ of Policy (0,0) for $s_i^*(0,0,0) = s_i^*(1,0) = 2$ (Lemma 2), 3, and 4. The dash-dot line (-.) is the Lagrangian function $L_i(1,0, s_i^*(1,0), \lambda) = L_i(1,0)$. The solid line is the minimum of the Lagrangian functions for $\lambda \geq 0$. $L_i(1,0)$ and $L_i(0,0,\lambda)$ intersect at $\lambda_{00-10,1} = 45630.35$ Euros/month = 63.38 Euros/hour, 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 5: Given a sample path for the failures of component-$i$ parts, the ordinary procedure and the emergency procedure are coupled. As $\frac{T_{\tau_i}}{\tau_i} \lambda_1 \mu_1^{-1}$ and $\frac{T_{\tau_i}}{\tau_i} \lambda_2 \mu_2^{-1}$ are the total expected downtime costs that would be incurred per system if all failures are handled by the ordinary or emergency procedure, respectively, Lemma 5 states that:

(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, as this is the worst case for Policy (0,0).

(ii) Similarly, 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, as this is the best case for Policy (0,0).
(iii) As the downtime penalty rate increases there exists a point between the worst case and the best case where choosing redundancy becomes optimal. Increasing the downtime penalty rate after that point will not change the redundancy decision as downtime costs are the incentive to choose redundancy. When the downtime constraint is not tight, one can satisfy it with lower TCO by not choosing redundancy. As the constraint grows tighter, one switches to redundancy.

5.2.3. Policy \((0,0)\) versus Policy \((0,1)\)

**Lemma 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 \(\lambda > \lambda_{00-01,i}\).

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 easily be found numerically.

**Example 4.** Consider the redundancy allocation problem introduced in Example 1; Figure 4 compares Policy \((0,0)\) and Policy \((0,1)\) for component 1. The dotted lines are the Lagrangian functions \(L_1(0,0, s, \lambda)\) of Policy \((0,0)\) for \(s = s_1^*(1,0) = 2, 3, \text{ and } 4\). The dashed line is the Lagrangian function \(L_1(0,1, \lambda) = L_1(0,1, s_1^*(1,0) + 1, \lambda)\). \((L_1(0,0, s, \lambda)\) 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 for \(\lambda \geq 0\). \(L_1(0,1, \lambda)\) and \(L_1(0,0, \lambda)\) intersect at \(\lambda_{00-01,1} = 59977.70\) Euros/month = 83.30 Euros/hour, Policy \((0,0)\) is superior for \(\lambda \leq \lambda_{00-01,1}\) and Policy \((0,1)\) superior for \(\lambda > \lambda_{00-01,1}\).

The intuitions for Lemma 6 is similar to those given for Lemma 4 and Lemma 5.

![Figure 4](attachment:figure4.png)  
**Figure 4** Comparison of Policy \((0,0)\) and Policy \((0,1)\) for component 1 in Example 4
5.2.4. Overall Comparison Combining Lemmas 4, 5, and 6, we see that there are three different cases for the optimal policy structure.

**Theorem 1.** For all $i \in M$:

1. If $N_{c_1,i} \leq c_{0,i} \hat{h}_i T$, Policy $(0,0)$ is optimal for $\lambda \leq \lambda_{00-10,i}$ and Policy $(1,0)$ is optimal for $\lambda \geq \lambda_{00-10,i}$.

2. If $N_{c_1,i} > c_{0,i} \hat{h}_i T$, either
   
   (a) Policy $(0,0)$ is optimal for $\lambda \leq \lambda_{00-01,i}$ and Policy $(1,0)$ is optimal for $\lambda \geq \lambda_{00-10,i}$; or,
   
   (b) Policy $(0,0)$ is optimal for $\lambda \leq \lambda_{00-01,i}$, Policy $(0,1)$ is optimal for $\lambda \in \left[\lambda_{00-01,i}, \lambda_{01-10,i}\right]$, and Policy $(1,0)$ is optimal for $\lambda \geq \lambda_{01-10,i}$.

**Proof.**

1. This follows from Lemmas 4 and 5.

2. If $N_{c_1,i} > c_{0,i} \hat{h}_i T$, 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 Lemmas 5 and 6, respectively. By Lemmas 4 and 5, Policy $(1,0)$ is optimal as $\lambda \to \infty$.

   It remains 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 these in Example 5. □

**Example 5.** We continue with the problem introduced in Example 1. Figure 5 and 6 compare the three policies for component 1 and component 2, respectively. In Figure 5, the functions are the same as those in Figures 2, 3, and 4. Figure 6 shows the corresponding functions for component 2. Observe that Policy $(0,1)$ is never optimal for component 1 (Figure 5) while it is optimal for component 2 for $\lambda \in \left[\lambda_{00-01,2}, \lambda_{01-10,2}\right]$, where $\lambda_{00-01,2} = 818238$ Euros/month = 1136.44 Euros/hour and $\lambda_{01-10,2} = 3630156$ Euros/month = 5041.88 Euros/hour (Figure 6).

Theorem 1 shows the optimal sequence of the policies followed for increasing downtime penalty rate ($\lambda$) (or decreasing resource level of the downtime constraint, $D_0$) in (Q0). The sequence can be either $(0,0)$-$(1,0)$ or $(0,0)$-$(0,1)$-$(1,0)$. For a low downtime penalty rate or high downtime level, 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. If $N_{c_1,i} \leq c_{0,i} \hat{h}_i T$, Policy $(0,1)$ can never be optimal and the switch always occurs to Policy $(1,0)$ at $\lambda_{01-10,i}$. If $N_{c_1,i} > c_{0,i} \hat{h}_i T$, 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 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 quickly identified. 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 $N_{c_{1,i}} \leq c_{0,i} + \hat{h}_i T$, $\lambda_{10,i} = \lambda_{00-10,i}$; and for $N_{c_{1,i}} > c_{0,i} + \hat{h}_i T$, $\lambda_{10,i} = \max\{\lambda_{00-10,i}, \lambda_{01-10,i}\}$. Corollary 3 determines upper and lower bounds for $\lambda_{10,i}$; these can be used for ranking components for redundancy, which will be detailed in Subsection 5.3.2.

**Figure 5** Comparison of Policy (0,1), Policy (1,0) and Policy (1,0) for component 1 in Example 5

**Figure 6** Comparison of Policy (0,1), Policy (1,0) and Policy (1,0) for component 2 in Example 5
Corollary 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 4 and Lemma 5. \( \square \)

5.3. Results for the Multi-Component Problem

We provide two system level results in this subsection. The first describes 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. Efficient solutions provide this exploration. We also demonstrate the benefit of the provisional procedure through these efficient solutions.

The second result ranks the components for redundancy: In many cases, there might be other factors which limit redundancy (e.g. a budget limit). In such cases one might be interested in the optimal order in making components redundant.

5.3.1. Finding Efficient Solutions

The values the downtime and the optimal TCO 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 1) for each component 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.

Example 6. Figure 7 shows the efficient frontier for the capital good with two components introduced in Example 1. The x-axis shows \( p \in (0,1] \), the availability measure that reflects the required uptime portion of the \( NT \) system-years. The frontier shows the efficient solutions at which the optimal solution switches from one policy to the other for a component. When Policy (0,0) is chosen for the two components and initial supply amounts are optimized, the expected overall downtime of the \( N = 15 \) systems is 2.64 months over the 15 years, which is equivalent to an availability level of \( p = 0.9990 \), with TCO of 1371004 Euros. The first policy change - yielding the highest increase in availability per unit increase in TCO - is from Policy (0,0) to Policy (1,0) for component 2. The next change is from Policy (0,0) to Policy (0,1) for component 1. Finally, downtime becomes zero (\( p = 1 \)) when Policy (1,0) is chosen for the both components.

One might argue that in Example 6, when Policy (0,0) is chosen, the availability level \( p = 0.9990 \) is already high. However, this availability level is equivalent to 8.64 hours of average downtime per system per year, which is a significant number. The high \( p \) value is attained due to the small number of components in the capital good, the infrequency of failures (MTBF are \( \tau_1 = 3 \) years and
Figure 7  Efficient frontier for Example 6

\(\tau_2 = 6\) years); and the relative swiftness of the ordinary and emergency procedures (\(\mu_{1,1} = 10\) hours, \(\mu_{1,2} = 24\) hours, \(\mu_{2,1} = 8\) hours, and \(\mu_{2,2} = 48\) hours). Total downtime increases (i.e., \(p\) decreases) as the number of components increases. In Figure 8, EF1 shows the efficient frontier for a capital good with 70 components with similar \(\tau_i\), \(\mu_{1,i}\), and \(\mu_{2,i}\) values given for Example 2. As you can observe, the values of \(p\) ranges within \([0.9595, 1]\).

Figure 8 also shows the benefit of the provisional procedure: EF2 is the efficient frontier when only Policy (0,0) and Policy (1,0) are considered (the provisional procedure is excluded. We can see that EF2 can attain similar availability levels to EF1, but at higher TCO values; for instance, an availability level of 99.4452% is attained at a TCO of 30569515 when Policy (0,1) is excluded while an availability level of 99.4451% is attained at a TCO of 29083419 Euros when Policy (0,1) is included. This corresponds to a decrease of 5.11% in TCO for approximately the same availability level.

Notice that the efficient frontiers given in Figures 7 and 8 are convex. We construct the efficient frontiers by the efficient solutions generated with respect to the Lagrangian multiplier (downtime penalty rate) values at which a switch from one policy to another occurs for one of the components. The convexity results from the fact that the ascending order of these values reflect the least cost increase per unit availability increase.

5.3.2. Ranking Components for Redundancy  Next, we rank the components in decreasing order of “bang for the buck”; i.e., largest increase in \(p\) per unit increase in TCO, when making
them redundant. This ranking is achieved by ordering the components in ascending order of the \( \lambda_{10,i} \)'s, \( i \in M \); the values of \( \lambda \) (downtime penalty rate) at which the optimal policy switches to Policy (1,0). Thus, the procedure for ordering the components can be stated as follows:

1. For all \( i \in M \) with \( N c_{1,i} \leq c_{0,i} + \hat{h}_i T \), find \( \lambda_{10,i} = \lambda_{00-10,i} \). For all \( i \in M \) with \( N c_{1,i} > c_{0,i} + \hat{h}_i T \), find \( \lambda_{10,i} = \max\{\lambda_{00-10,i}, \lambda_{01-10,i}\} \).

2. Permute the set \( M \) such that \( i < k \) implies that \( \lambda_{10,i} \leq \lambda_{01,k} \) for all \( i, k \in M \).

Recall that the resource level of the downtime constraint in \( Q(\lambda), D(y^*(\lambda), z^*(\lambda), s^*(\lambda)) \), is decreasing in \( \lambda \). Hence, by choosing redundancy for the components in increasing order of \( \lambda_{10,i} \), the optimal TCO values are provided for decreasing resource level of the downtime constraint (i.e., increasing \( p \) value) at each step.

**Example 7.** We apply the procedure for ranking the components of the capital good given in Example 1.

1. As \( N c_{1,i} > c_{0,i} + \hat{h}_i T \) for \( \forall i \in \{1, 2\} \), \( \lambda_{10,1} = \max\{\lambda_{00-10,1}, \lambda_{01-10,1}\} = \max\{63.38 \text{ Euros/hour}, 60.67 \text{ Euros/hour}\} = 63.38 \text{ Euros/hour} \) and \( \lambda_{10,2} = \max\{\lambda_{00-10,2}, \lambda_{01-10,2}\} = \max\{4174.86 \text{ Euros/hour}, 5041.88 \text{ Euros/hour}\} = 5041.88 \text{ Euros/hour} \).

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

Hence, one should choose first component 1 and then component 2 to implement redundancy.
6. Conclusions

We develop a redundancy optimization model for critical components in a capital good. We define three potential policies per component, each of which also includes the user choosing the amount of spare parts inventory to stock. One of our policies involves the use of what we define as the “provisional procedure”, in which an emergency replenishment is utilized when actual stock-on-hand is positive. We developed the problem formulation as the minimization of the total cost of ownership (acquisition costs, spare parts costs, and repair costs) of a general number of systems under a defined downtime constraint. The multi-component problem has a combinatorial nature due to the three candidate policies per component. Therefore, we decompose the problem into single-component problems, and show that a solution for the multi-component problem could be generated by finding solutions of each of the single-component problems. We then develop an efficient procedure to find the optimal solutions of the single-component problems for varying resource levels of the downtime constraint. This procedure enable us to find the optimal solutions for the multi-component problem for varying resource levels of the downtime constraint efficiently.

Our optimization procedure is supported by analytical results we prove for the single-component problems and the multi-component problem. These results reveals that for the single-component problem, as the value of the downtime constraint grows tighter, choosing redundancy becomes optimal and remains optimal thereafter. We also showed how the values of the TCO and downtime (or uptime) when the optimal policy changes from one policy to the other can be easily computed. This property leads to a simple method to construct an efficient frontier to explore the trade-off between the availability and the TCO for the multi-component problem. The efficient frontiers allow us to illustrate the benefit of the provisional procedure: When the provisional procedure is available, one can attain similar availability levels at lower TCO values than when it is not. We also provided an optimal ranking of the components in a capital good for implementing redundancy, as one might be interested in finding the optimal order to follow for components to make one-by-one redundancy decisions.

In the case that we studied, there is no priority among the systems in terms of availability. That is, downtime costs are not differentiated by system, and there is a single aggregate constraint on total uptime of all systems. We plan to extend the current work for cases in which systems have different availability requirements (constraints). Then, one may choose to implement redundancy per component per system. This would make the current problem formulation no longer decomposable, and a different approach would be required for the analysis.
References


cnet news. 2001. *California power outages suspended - for now.*


**Appendix**

*Proof of Property 1.*

(i) Under Policy (0,0) and Policy (1,0), the instances of the ordinary and emergency procedure are identical to those defined in Subsection 3.2 of Öner et al. (2010); this property is equivalent to Lemma 1 therein with \( \tau = \tau_i \) and \( G(\tau, s) = B_i(s_i) \).

(ii) This can be proved in a similar manner as Lemma 1 in Öner et al. (2010).

(iii) In part (ii) of Lemma 2 in Öner et al. (2010), the NPV of the expected repair costs are derived. By following a similar manner used in that derivation, it can be shown that

\[
R_i(0, s_i) = \frac{N}{\alpha \tau_i} (1 - e^{-\alpha T}) r_{1,i} + (r_{2,i} - r_{1,i}) B_i(s_i),
\]
The two formulations can be represented by equation (2).

(iii) 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 repaired by 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

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

For Policy \((0,1)\), 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 \((1,0)\), the expected downtime throughout \([0,T] \) is

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

\[\square\]

**Proof of Lemma 1.** Let \( i \in M \) and \( \lambda \geq 0 \). For Policy \((0,0)\),

\[
L_i(0,0,s_i,\lambda) = (c_{0,i} + \hat{h}_i T)s_i + \frac{NT}{\tau_i} (\hat{r}_{1,i} + \lambda \mu_{1,i}) + \frac{NT}{\tau_i} (\hat{r}_{2,i} - \hat{r}_{1,i} + \lambda (\mu_{2,i} - \mu_{1,i})]B_i(s_i). \tag{7}
\]

The second and fourth term of (7) 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 Remark
2 of Kranenburg and van Houtum, 2007, for a correction of the proof). 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),
\[
L_i(0,1,s_i,\lambda) = (c_{0,i} + \hat{h}_iT)s_i + \frac{NT}{\tau_i}(\hat{r}_{1,i} + \lambda \mu_{1,i}) + \frac{NT}{\tau_i}(\hat{r}_{2,i} - \hat{r}_{1,i})B_i(s_i - 1),
\]
and for Policy (1,0),
\[
L_i(1,0,s_i,\lambda) = Nc_{1,i} + (c_{0,i} + \hat{h}_iT)s_i + \frac{NT}{\tau_i}\hat{r}_{1,i} + \frac{NT}{\tau_i}(\hat{r}_{2,i} - \hat{r}_{1,i})B_i(s_i).
\]

Notice that \( L_i(1,0,s_i,\lambda) \) is independent of \( \lambda \), as there is no downtime in Policy (1,0). By following similar arguments as were 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 \( \lambda \).

**Proof of Lemma 2.** Let \( \Delta_i L_i(y_i,z_i,s_i,\lambda) = L_i(y_i,z_i,s_i + 1,\lambda) - L_i(y_i,z_i,s_i,\lambda) \).

(i) The property that \( L_i(1,0,s_i,\lambda) \) is independent of \( \lambda \) (see equation (9)), for all \( \lambda \geq 0 \), implies that \( s^*_i(1,0,\lambda) \) is independent of \( \lambda \) and may be denoted as \( s^*_i(1,0) \). \( L_i(1,0,s^*_i(1,0),\lambda) = L_i(1,0,s^*_i(1,0),0) \) is also a constant that we denote by \( L_i(1,0) \). As \( L_i(1,0,s_i,\lambda) \) is strictly convex in \( s_i \), \( s^*_i(1,0) = \min\{ s_i \in \mathbb{N}_0 \mid \Delta_i L_i(1,0,s_i,\lambda) \geq 0 \} \). Then
\[
\Delta_i L_i(1,0,s_i,\lambda) = (c_{0,i} + \hat{h}_iT) - \frac{NT}{\tau_i}(\hat{r}_{2,i} - \hat{r}_{1,i})\Delta B_i(s_i)
\]
implies equation (5).

(ii) As \( L_i(0,1,s_i,\lambda) \) is strictly convex in \( s_i \),
\[
s^*_i(0,1,\lambda) = \min\{ s_i \in \mathbb{N} \mid \Delta_i L_i(0,1,s_i,\lambda) \geq 0 \}.
\]
\[
\Delta_i L_i(0,1,s_i,\lambda) = (c_{0,i} + \hat{h}_iT) - \frac{NT}{\tau_i}(\hat{r}_{2,i} - \hat{r}_{1,i})\Delta B_i(s_i - 1);
\]
thus
\[
s^*_i(0,1,\lambda) = \min\left\{ s_i \in \mathbb{N} \mid \Delta B_i(s_i - 1) \leq \frac{c_{0,i} + \hat{h}_iT}{\hat{r}_{2,i} - \hat{r}_{1,i}} \frac{\tau_i}{NT} \right\} = s^*_i(1,0) + 1.
\]
\[
L_i(0,1,\lambda) = L_i(0,1,s^*_i(0,1),\lambda) = L_i(0,1,s^*_i(1,0) + 1,\lambda)
\]
\[
= (c_0 + \hat{h}_iT)(s^*_i(1,0) + 1)
\]
\[
+ \frac{NT}{\tau} [r_{2,i} + (r_{2,i} - r_{2,i})B_i(s^*_i(1,0))] + \frac{NT}{\tau} \mu_{1,i} \lambda
\]
is an increasing linear function of \( \lambda \) 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 \( \lambda \geq 0 \),

\[
s^*_i(0,0,\lambda) = \min \left\{ s_i \in \mathbb{N}_0 \mid \Delta_i L_i(0,0,s_i,\lambda) \geq 0 \right\}.
\]

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

thus

\[
s^*_i(0,0,\lambda) = \min \left\{ s_i \in \mathbb{N}_0 \mid \Delta B_i(s_i) \leq \frac{c_{0,i} + \hat{h}_iT}{\bar{r}_{2,i} - \bar{r}_{1,i} + (\mu_{2,i} - \mu_{1,i})\lambda} \frac{\tau_i}{NT} \right\}, \quad (10)
\]

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 \( \lambda \); see equation (7). 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 \( \lambda \) and \( s^*_i(0,0,\lambda) \) is increasing in \( \lambda \). \( \square \)

**Proof of Lemma 4.** As \( s^*_i(0,1,\lambda) = s^*_i(1,0) + 1 \) for all \( \lambda \geq 0 \), equations (8) and (9) imply:

\[
L_i(0,1,\lambda) = L_i(1,0) - Nc_{1,i} + c_{0,i} + \hat{h}_iT + \frac{NT}{\tau_i} \mu_{1,i} \lambda.
\]

Observe that \( L_i(1,1,0) = L_i(1,0) - Nc_{1,i} + c_{0,i} + \hat{h}_iT \) and \( L_i(0,1,1) \) is an increasing linear function of \( \lambda \). 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 \( Nc_{1,i} \leq c_{0,i} + \hat{h}_iT, \ L_i(0,1,0) \leq L_i(0,1,\lambda) \) for all \( \lambda \geq 0 \).

(ii) If \( Nc_{1,i} > c_{0,i} + \hat{h}_iT, \ L_i(1,0,1) \) and \( L_i(0,1,1) \) intersects at \( \lambda_{01-10} = \frac{Nc_{1,i} - c_{0,i} - \hat{h}_iT}{NT\tau_{p1,i}} \).

For \( 0 \leq \lambda < \lambda_{01-10}, \ L_i(1,0,1) \leq L_i(1,0,\lambda) \); for \( \lambda > \lambda_{01-10}, \ L_i(0,1,0) \leq L_i(0,1,\lambda) \). \( \square \)

**Proof of Lemma 5.** 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 Lagrangians of Policies (1,0) and (0,0).

(i) If \( 0 \leq \lambda < c_{1,i} \left( \frac{\tau_i}{\tau_{p1,i}} \right)^{-1}, \) 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 \( c_{1,i} \left( \frac{\tau_i}{\tau_{p1,i}} \right)^{-1} < \lambda, \) 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 \( 0 \leq \lambda < c_{1,i} \left( \frac{T}{\chi_{1,i}} \mu_{2,i} \right)^{-1} \). 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 \( \lambda \) and \( L_i(1,0) \) is a constant, there exists a \( \lambda_{00-10,i} \) such that 
\[
L_i(0,0,\lambda) \leq L_i(1,0) \quad \text{for all} \quad \lambda < \lambda_{00-10,i} \quad \text{and} \quad L_i(0,0,\lambda) > L_i(1,0) \quad \text{for all} \quad \lambda > \lambda_{00-10,i}.
\]
By (i) and (ii), 
\[
\lambda_{00-10,i} - 1 \leq \frac{T}{\chi_{1,i}} \mu_{2,i} \leq \lambda_{00-10,i} \leq \frac{T}{\chi_{1,i}} \mu_{1,i} - 1.
\]
\[\Box\]

**Proof of Lemma 6.** For all \( i \in M \), remember that \( L_i(0,0,\lambda) \) is a strictly increasing, concave, piecewise linear function of \( \lambda \) whose pieces are constructed by \( L_i(0,0,s_i,\lambda) \), \( s_i \in \{ s_{0,i}(1,0), s_{0,i}(1,0) + 1, \ldots \} \) as \( s_{0,i}(0,0,0) = s_{0,i}(1,0) \). 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 \( s_i \in \{ s_{0,i}(1,0), s_{0,i}(1,0) + 1, \ldots \} \). 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 \( \lambda > \lambda_{00-01,i} \). \[\Box\]
<table>
<thead>
<tr>
<th>nr.</th>
<th>Year</th>
<th>Title</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>341</td>
<td>2011</td>
<td>Redundancy Optimization for Critical Components in High-Availability Capital Goods</td>
<td>Kurtulus Baris Öner, Alan Scheller-Wolf</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Geert-Jan van Houtum</td>
</tr>
<tr>
<td>339</td>
<td>2010</td>
<td>Analysis of a two-echelon inventory system with two supply modes</td>
<td>Joachim Arts, Gudrun Kiesmüller</td>
</tr>
<tr>
<td>338</td>
<td>2010</td>
<td>Analysis of the dial-a-ride problem of Hunsaker and Savelsbergh</td>
<td>Murat Firat, Gerhard J. Woeginger</td>
</tr>
<tr>
<td>335</td>
<td>2010</td>
<td>Attaining stability in multi-skill workforce scheduling</td>
<td>Murat Firat, Cor Hurkens</td>
</tr>
<tr>
<td>332</td>
<td>2010</td>
<td>Efficiency evaluation for pooling resources in health care</td>
<td>Christian Howard, Ingrid Reijnen, Johan Marklund, Tarkan Tan</td>
</tr>
<tr>
<td>331</td>
<td>2010</td>
<td>The Effect of Workload Constraints in Mathematical Programming Models for Production Planning</td>
<td>M.M. Jansen, A.G. de Kok, I.J.B.F. Adan</td>
</tr>
<tr>
<td>329</td>
<td>2010</td>
<td>Reducing costs of repairable spare parts supply systems via dynamic scheduling</td>
<td>H.G.H. Tiemessen, G.J. van Houtum</td>
</tr>
<tr>
<td>328</td>
<td>2010</td>
<td>Identification of Employment Concentration and Specialization Areas: Theory and Application</td>
<td>Murat Firat, Cor Hurkens</td>
</tr>
<tr>
<td>327</td>
<td>2010</td>
<td>A combinatorial approach to multi-skill workforce scheduling</td>
<td>Murat Firat, Cor Hurkens, Alexandre Laugier</td>
</tr>
<tr>
<td>326</td>
<td>2010</td>
<td>Stability in multi-skill workforce scheduling</td>
<td>Murat Firat, Cor Hurkens, Alexandre Laugier</td>
</tr>
</tbody>
</table>
Maintenance spare parts planning and control: A framework for control and agenda for future research  

Near-optimal heuristics to set base stock levels in a two-echelon distribution network  
R.J.I. Basten, G.J. van Houtum

Inventory reduction in spare part networks by selective throughput time reduction  
M.C. van der Heijden, E.M. Alvarez, J.M.J. Schutten

The selective use of emergency shipments for service-contract differentiation  
E.M. Alvarez, M.C. van der Heijden, W.H. Zijm

Heuristics for Multi-Item Two-Echelon Spare Parts Inventory Control Problem with Batch Ordering in the Central Warehouse  
B. Walrave, K. v. Oorschot, A.G.L. Romme

Preventing or escaping the suppression mechanism: intervention conditions  
Nico Dellaert, Jully Jeunet.

Hospital admission planning to optimize major resources utilization under uncertainty  
R. Seguel, R. Eshuis, P. Grefen.

Minimal Protocol Adaptors for Interacting Services  

Teaching Retail Operations in Business and Engineering Schools  
Lydie P.M. Smets, Geert-Jan van Houtum, Fred Langerak.

Design for Availability: Creating Value for Manufacturers and Customers  
Pieter van Gorp, Rik Eshuis.

Transforming Process Models: executable rewrite rules versus a formalized Java program  
Bob Walrave, Kim E. van Oorschot, A. Georges L. Romme

Getting trapped in the suppression of exploration: A simulation model  
S. Dabia, T. van Woensel, A.G. de Kok


Tales of a So(u)rcerer: Optimal Sourcing Decisions Under Alternative Capacitated Suppliers and General Cost Structures  
Osman Alp, Tarkan Tan
<table>
<thead>
<tr>
<th>Page</th>
<th>Year</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>311</td>
<td>2010</td>
<td>In-store replenishment procedures for perishable inventory in a retail environment with handling costs and storage constraints</td>
<td>R.A.C.M. Broekmeulen, C.H.M. Bakx</td>
</tr>
<tr>
<td>310</td>
<td>2010</td>
<td>The state of the art of innovation-driven business models in the financial services industry</td>
<td>E. Lüftenegger, S. Angelov, E. van der Linden, P. Grefen</td>
</tr>
<tr>
<td>308</td>
<td>2010</td>
<td>Effect of carbon emission regulations on transport mode selection in supply chains</td>
<td>K.M.R. Hoen, T. Tan, J.C. Fransoo, G.J. van Houtum</td>
</tr>
<tr>
<td>307</td>
<td>2010</td>
<td>Interaction between intelligent agent strategies for real-time transportation planning</td>
<td>Martijn Mes, Matthieu van der Heijden, Peter Schuur</td>
</tr>
<tr>
<td>306</td>
<td>2010</td>
<td>Internal Slackening Scoring Methods</td>
<td>Marco Slikker, Peter Borm, René van den Brink</td>
</tr>
<tr>
<td>304</td>
<td>2010</td>
<td>Practical extensions to the level of repair analysis</td>
<td>R.J.I. Basten, M.C. van der Heijden, J.M.J. Schutten</td>
</tr>
<tr>
<td>303</td>
<td>2010</td>
<td>Ocean Container Transport: An Underestimated and Critical Link in Global Supply Chain Performance</td>
<td>Jan C. Fransoo, Chung-Yee Lee</td>
</tr>
<tr>
<td>302</td>
<td>2010</td>
<td>Capacity reservation and utilization for a manufacturer with uncertain capacity and demand</td>
<td>Y. Boulaksil; J.C. Fransoo; T. Tan</td>
</tr>
<tr>
<td>300</td>
<td>2009</td>
<td>Spare parts inventory pooling games</td>
<td>F.J.P. Karsten; M. Slikker; G.J. van Houtum</td>
</tr>
<tr>
<td>299</td>
<td>2009</td>
<td>Capacity flexibility allocation in an outsourced supply chain with reservation</td>
<td>Y. Boulaksil, M. Grunow, J.C. Fransoo</td>
</tr>
<tr>
<td>298</td>
<td>2010</td>
<td>An optimal approach for the joint problem of level of repair analysis and spare parts stocking</td>
<td>R.J.I. Basten, M.C. van der Heijden, J.M.J. Schutten</td>
</tr>
<tr>
<td>297</td>
<td>2009</td>
<td>Responding to the Lehman Wave: Sales Forecasting and Supply Management during the Credit Crisis</td>
<td>Robert Peels, Maximiliano Udenio, Jan C. Fransoo, Marcel Wolfs, Tom Hendrikx</td>
</tr>
<tr>
<td>296</td>
<td>2009</td>
<td>An exact approach for relating recovering surgical patient workload to the master surgical schedule</td>
<td>Peter T. Vanberkel, Richard J. Boucherie, Erwin W. Hans, Johann L. Hurink, Wineke A.M. van Lent, Wim H. van Harten</td>
</tr>
<tr>
<td>294</td>
<td>2009</td>
<td>Fujaba hits the Wall(-e)</td>
<td>Pieter van Gorp, Ruben Jubeh, Bernhard Grusie, Anne Keller</td>
</tr>
<tr>
<td>291</td>
<td>2009</td>
<td>Efficient Optimization of the Dual-Index Policy</td>
<td>Joachim Arts, Marcel van Vuuren,</td>
</tr>
<tr>
<td>Page</td>
<td>Year</td>
<td>Title</td>
<td>Authors</td>
</tr>
<tr>
<td>------</td>
<td>------</td>
<td>----------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>290</td>
<td>2009</td>
<td>Using Markov Chains</td>
<td>Gudrun Kiesmuller</td>
</tr>
<tr>
<td>289</td>
<td>2009</td>
<td>Hierarchical Knowledge-Gradient for Sequential Sampling</td>
<td>Martijn R.K. Mes; Warren B. Powell; Peter I. Frazier</td>
</tr>
<tr>
<td>288</td>
<td>2009</td>
<td>Analyzing combined vehicle routing and break scheduling from a distributed decision making perspective</td>
<td>C.M. Meyer; A.L. Kok; H. Kopfer; J.M.J. Schutten</td>
</tr>
<tr>
<td>287</td>
<td>2009</td>
<td>Anticipation of lead time performance in Supply Chain Operations Planning</td>
<td>Michiel Jansen; Ton G. de Kok; Jan C. Fransoo</td>
</tr>
<tr>
<td>286</td>
<td>2009</td>
<td>Inventory Models with Lateral Transshipments: A Review</td>
<td>Colin Paterson; Gudrun Kiesmuller; Ruud Teunter; Kevin Glazebrook</td>
</tr>
<tr>
<td>285</td>
<td>2009</td>
<td>Efficiency evaluation for pooling resources in health care</td>
<td>P.T. Vanberkel; R.J. Boucherie; E.W. Hans; J.L. Hurink; N. Litvak</td>
</tr>
<tr>
<td>284</td>
<td>2009</td>
<td>A Survey of Health Care Models that Encompass Multiple Departments</td>
<td>P.T. Vanberkel; R.J. Boucherie; E.W. Hans; J.L. Hurink; N. Litvak</td>
</tr>
<tr>
<td>283</td>
<td>2009</td>
<td>Supporting Process Control in Business Collaborations</td>
<td>S. Angelov; K. Vidyasankar; J. Vonk; P. Grefen</td>
</tr>
<tr>
<td>282</td>
<td>2009</td>
<td>Inventory Control with Partial Batch Ordering</td>
<td>O. Alp; W.T. Huh; T. Tan</td>
</tr>
<tr>
<td>281</td>
<td>2009</td>
<td>Translating Safe Petri Nets to Statecharts in a Structure-Preserving Way</td>
<td>R. Eshuis</td>
</tr>
<tr>
<td>280</td>
<td>2009</td>
<td>The link between product data model and process model</td>
<td>J.J.C.L. Vogelaar; H.A. Reijers</td>
</tr>
<tr>
<td>279</td>
<td>2009</td>
<td>Inventory planning for spare parts networks with delivery time requirements</td>
<td>I.C. Reijnen; T. Tan; G.J. van Houtum</td>
</tr>
<tr>
<td>278</td>
<td>2009</td>
<td>Co-Evolution of Demand and Supply under Competition</td>
<td>B. Vermeulen; A.G. de Kok</td>
</tr>
<tr>
<td>277</td>
<td>2009</td>
<td>Toward Meso-level Product-Market Network Indices for Strategic Product Selection and (Re)Design Guidelines over the Product Life-Cycle</td>
<td>B. Vermeulen; A.G. de Kok</td>
</tr>
<tr>
<td>276</td>
<td>2009</td>
<td>An Efficient Method to Construct Minimal Protocol Adaptors</td>
<td>R. Seguel, R. Eshuis, P. Grefen</td>
</tr>
<tr>
<td>275</td>
<td>2009</td>
<td>Coordinating Supply Chains: a Bilevel Programming Approach</td>
<td>Ton G. de Kok, Gabriella Muratore</td>
</tr>
<tr>
<td>274</td>
<td>2009</td>
<td>Inventory redistribution for fashion products under demand parameter update</td>
<td>G.P. Kiesmuller, S. Minner</td>
</tr>
<tr>
<td>273</td>
<td>2009</td>
<td>Comparing Markov chains: Combining aggregation and precedence relations applied to sets of states</td>
<td>A. Busic, I.M.H. Vliegen, A. Scheller-Wolf</td>
</tr>
<tr>
<td>272</td>
<td>2009</td>
<td>Separate tools or tool kits: an exploratory study of engineers’ preferences</td>
<td>I.M.H. Vliegen, P.A.M. Kleingeld, G.J. van Houtum</td>
</tr>
<tr>
<td>271</td>
<td>2009</td>
<td>An Exact Solution Procedure for Multi-Item Two-Echelon Spare Parts Inventory Control Problem with Batch Ordering</td>
<td>Engin Topan, Z. Pelin Bayindir, Tarkan Tan</td>
</tr>
<tr>
<td>270</td>
<td>2009</td>
<td>Distributed Decision Making in Combined</td>
<td>C.M. Meyer, H. Kopfer, A.L. Kok, M.</td>
</tr>
<tr>
<td>Year</td>
<td>Title</td>
<td>Authors</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>----------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>270</td>
<td>Vehicle Routing and Break Scheduling</td>
<td>Schutten</td>
<td></td>
</tr>
<tr>
<td>269</td>
<td>Similarity of Business Process Models: Metics and Evaluation</td>
<td>Remco Dijkman, Marlon Dumas, Boudewijn van Dongen, Reina Kaarik, Jan Mendling</td>
<td></td>
</tr>
<tr>
<td>266</td>
<td>Restricted dynamic programming: a flexible framework for solving realistic VRPs</td>
<td>J. Gromicho; J.J. van Hoorn; A.L. Kok; J.M.J. Schutten;</td>
<td></td>
</tr>
</tbody>
</table>

Working Papers published before 2009 see: [http://beta.ieis.tue.nl](http://beta.ieis.tue.nl)