OPTIMAL MAINTENANCE PLANNING IN NOVEL SETTINGS

by

Kai He

B.S., Shanghai Jiao Tong University, 2011

Submitted to the Graduate Faculty of

the Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2016

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Kai He

It was defended on

October 19, 2016

and approved by

Oleg A. Prokopyev, PhD, Associate Professor, Department of Industrial Engineering

Lisa M. Maillart, PhD, Associate Professor, Department of Industrial Engineering

Jeffrey P. Kharoufeh, PhD, Professor, Department of Industrial Engineering

Mark S. Roberts, MD, MPP, Professor, Department of Health Policy and Management

Dissertation Advisors: Oleg A. Prokopyev, PhD, Associate Professor, Department of Industrial Engineering,

Lisa M. Maillart, PhD, Associate Professor, Department of Industrial Engineering

OPTIMAL MAINTENANCE PLANNING IN NOVEL SETTINGS

Kai He, PhD

University of Pittsburgh, 2016

In this dissertation work, we focus on optimal planning of maintenance activities in several novel settings.

First, we consider a maintenance optimization model for a system with periodic preventive maintenance (PM), and periodic imperfect inspections to detect hidden failures. Our stylized mathematical model is inspired by the increasingly popular remote monitoring practices. We describe, both analytically and numerically, important structural properties of the model, and propose a simple approach to find a globally optimal solution.

In the second chapter, we investigate a maintenance planning scenario in which the implementation of PM is unpunctual. Under the assumption that the degree of the unpunctuality follows a known probability distribution, we formulate cost-rate minimizing models to study the impact of such deviations. We establish both analytical and numerical results for two specific types of maintenance policies common in practice, namely age replacement with and without minimal repair.

Finally, we focus on "maintaining" the health status of a patient with a chronic disease by investigating an optimal medical treatment sequencing problem. We restrict our attention to the two treatment case, and simultaneously balance three tradeoffs inherent to these treatments, i.e., length of effectiveness delay, probability of effectiveness and cost/reward. We provide both theoretical conditions and numerical examples that indicate when, as a function of the model parameters, it is optimal to initiate treatment with one treatment versus the other.

Keywords: Maintenance optimization, medical decision models, stochastic processes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

Proper functionality of a system depends not only on its components' reliability but also on its maintenance [28, 61]. There are two major types of maintenance activities - reactive maintenance (RM) and preventive maintenance (PM) [61, 76]. If the maintenance activity is to correct an existing failure, then it is referred to as RM. Performing RM alone can possibly save money in short term, but it often ends up incurring more cost in the long run [69]. For instance, consider a manufacturing company that relies on some highly automated equipment to perform mass production. If maintenance activities only aim to correct failures when they occur, then breakdown of the production line could cause substantial losses [88]. As a result, a pure RM policy might not be optimal.

The importance of PM is ever increasing because of its imperative role in keeping good condition of the system and reducing or avoiding possible failures [6]. For example, when it comes to airline industry, any failure can possibly cause devastating consequences [65]. Other benefits of PM, such as decreasing equipment downtime and improving equipment efficiency, are widely recognized [83] as well. Indeed, a good maintenance policy is usually a mixture of both RM and PM activities [8].

Optimal maintenance planning remains a very active research domain [88]. In the past several decades, a number of maintenance optimization models have been developed, see review articles [22, 29, 32, 43, 58, 77, 87, 90]. Recently, research efforts have been extended to more complex settings and multi-component or even multi-subsystem models [22, 62, 64].

There are many parallels between maintaining a degrading machine and "maintaining" a degrading human body. Indeed, Dekker [25] observes in his maintenance optimization survey that "maintaining" a human being involves concepts similar to those associated with maintaining machines (e.g., lifetime distribution, disease screening corresponds to inspections, etc). In therapeutic optimization models, the patient's health status is usually assumed to be stochastically degrading. Optimal decisions about therapy initiation and switching are made in order to maximize the patient's quality adjusted life years (QALYs). Therefore, the patient can be viewed as the system of interest to be maintained; prescribing or switching a therapy for the patient is equivalent to a "maintenance" activity; and the outcome after "maintenance" is that the patient's disease level can change stochastically. A body of recent research work focuses on "maintaining" the health status of patients with different diseases. (See examples in [4, 5, 46, 47, 74, 78, 79, 81, 82].)

Motivated by the connections between maintenance optimization concepts and models in medical decision making, this dissertation is focused on developing optimal maintenance policies under several novel settings inspired by healthcare problems.

To be more specific, Chapter 2 is inspired by remote monitoring (i.e., telehealth) practices that have become prevalent in recent years. We consider a maintenance scenario in which imperfect periodic inspections (IPIs) occur at a chosen interval to detect hidden failures with a certain probability less than one[1]. Both reactive maintenance (RM), performed when a hidden failure is detected by an IPI, and PM, performed after a multiple of the IPI intervals, renew the system. The objective is to determine the optimal frequency and quantity of IPIs between PM actions such that the expected cost (which includes the costs of undetected failures, IPIs, PM and RM) per unit time is minimized over an infinite horizon. We analytically establish conditions for the existence of a finite optimal IPI interval for a given quantity of IPIs between PM actions, and discuss asymptotic behavior of the objective function. These results are further exploited to describe convergence properties of a proposed approach for finding a globally optimal solution. Also, for the special case of a Weibull time-to-failure distribution, we derive conditions that guarantee uniqueness of a locally optimal solution for a given quantity of IPIs between PM actions.

In Chapter 3, we tackle the maintenance planning scenario in which the implementation of PM is unpunctual[2]. In traditional maintenance decision-making, maintenance planners

[1] 2015 IEEE. Reprinted, with permission, from He, K. , L. M. Maillart, O. A. Prokopyev, Scheduling Preventive Maintenance as a Function of an Imperfect Inspection Interval. *IEEE Transactions on Reliability*, Vol. 64/3 (2015), pp. 983-997.

[2] He, K., L. M. Maillart, O. A. Prokopyev. Optimal Planning of Unpunctual Preventive Maintenance. *IIE Transactions*, to appear.

assume that their prescribed PM policies will be implemented without error. In practice, however, the individuals responsible for implementing such plans often deviate from the intended PM policy resulting in unpunctual PM actions. In a healthcare context, doctors usually recommend screening policies for disease prevention (e.g., American Cancer Society suggests women with age 45 to 54 should get mammograms every year [1]), but patients may not adhere to the prescribed schedule [56]. In either scenario, the punctuality or inadherence to maintenance (screening) policy could potentially leave the degrading system at risk. We formulate cost-rate minimizing models to investigate the impact of such deviations, assuming that the actual PM time deviates from the scheduled PM time in a probabilistic manner. We establish both analytical and numerical results for two specific types of maintenance policies common in practice, namely age replacement with and without minimal repair.

Chapter 4 studies the best treatment sequence for a chronic disease by formulating a stylized mathematical model with two treatment options. Our model simultaneously captures three characteristics of these two available treatments, namely, length of effectiveness delay, probability of effectiveness and cost/reward. Both numerical and analytical results are established to illustrate how to balance the trade-offs inherent to these three characteristics. In particular, we provide conditions under which a specific treatment should be prescribed first.

2.0 SCHEDULING PREVENTIVE MAINTENANCE AS A FUNCTION OF AN IMPERFECT INSPECTION INTERVAL

2.1 INTRODUCTION

Consider a system for which both imperfect periodic inspections and perfect preventive maintenance may be performed to detect so-called hidden or silent failures. Each failure incurs some fixed, instantaneous, nonnegative cost ζ and positive cost $\int_0^\tau c_3(u)du$, where τ is the length of time between the failure and its detection and $c_3(u)$ is the corresponding (possibly, non-constant) cost rate associated with the failure. Preventive maintenance (PM) is assumed to be perfect in that it detects existing failures with probability one and instantaneously renews the system by addressing any underlying problems. The imperfect periodic inspections (IPIs) are less expensive, but less reliable in that they detect existing failures with probability $p \in (0, 1)$. We assume a fixed PM interval, within which some number of IPIs (possibly, none) are equally spaced. That is, given a positive $t > 0$ and nonnegative integer n, IPIs are performed at times t, $2t$, $\ldots$, nt, and planned PM occurs every $(n + 1)t$ units of time. Any time a failure is detected by an IPI, reactive maintenance (RM) is performed and instantaneously renews the system. The costs of performing PM (or RM) and an IPI are given by c_1 and c_2, respectively, where $c_1 > c_2 > 0$. The overall objective is to select t and n, i.e., a policy (t, n), such that the long-run average cost rate (which combines the costs associated with undetected failures, IPIs, PM and RM) is minimized.

Our analysis is inspired by various remote monitoring practices that have become prevalent in recent years [3]. One interesting example is the Care Coordination/Home Telehealth (CCHT) program supported by the U.S. Veterans Health Administration (VHA), which currently serves more than 30,000 mostly elderly patients [24]. The focus of the

CCHT program is to provide chronic care services to veterans with various conditions, e.g., diabetes mellitus, congestive heart failure, hypertension, postraumatic stress disorder [2]. In addition to regularly scheduled hospital visits, e.g., every 6 months, CCHT involves periodic remote monitoring activities via telehealth technologies that transmit (e.g., over a phone line or wirelessly) a patient's health information such as vital signs (e.g., weight, oxygen, blood pressure, pulse, blood glucose), and answers to a set of scripted questions about the patient's symptoms and health status.

The transmitted data are processed upon arrival to determine whether the patient has developed a problem that may require action. If necessary, the nurse responsible for monitoring the patient is alerted and determines the appropriate course of action (e.g., none, schedule an in-office visit, advise the patient to go to the emergency room). In our stylized model, each instance of a patient's remote data collection corresponds to an IPI, each scheduled checkup corresponds to PM and each unscheduled visit corresponds to RM. In the CCHT context [59], a hidden failure corresponds to an asymptomatic change in the patient's condition (e.g., abnormal hemoglobin values, high blood pressure) that results in some type of cumulative damage to the patient's health while it remains undetected (e.g., narrowing and hardening of the arteries, thickening of the heart walls, accumulation of fluid in the kidneys). The costs associated with these progressive conditions are often measured in years of life lost. The renewal actions correspond to changes in patient care that address the underlying problem (e.g., medication adjustment that stabilizes the patient's hemoglobin values or blood pressure) and effectively reset the time until the next hidden problem (e.g., episode of high blood pressure) develops, prompting another adjustment in therapy.

The maintenance optimization literature to which this chapter contributes is vast; see surveys in [58, 70, 90] as well as references therein. Within the maintenance optimization literature, determining a periodic inspection interval to detect hidden failures has received much attention; the majority of this literature, however, assumes error-free inspections, see, e.g., [12, 37, 44, 55, 66, 86, 89]. More specifically, for example, Badía and Berrade [12] analyze the problem of optimally determining the inspection interval for a system subject to imperfect repairs after a failure detection and perfect repairs after the nth detected failure. Taghipour and Banjevic [86] consider periodic inspection optimization models for a multi-component system with a cost structure similar to ours.

Examples of papers that optimize the timing of imperfect inspections can be found in [11, 14, 17, 68, 91, 92]. More specifically, Parmigiani [68] studies the problem of designing inspection schedules for both perfect and imperfect, time-consuming inspections where imperfect inspections are less expensive, but may result in both false positive or false negative outcomes. Badía and Berrade [11] consider a maintenance model for a system with two types of failures, namely, hidden ones that are costly, and obvious ones, which are minor and can be removed by a minimal repair. Periodic inspections detect hidden failures imperfectly and the system is renewed either after the nth obvious minor failure, or after a hidden failure is detected.

Most closely related to our work, Zequeira and Bérenguer [92] consider optimal inspection policies for a system subject to three types of inspections and three types of failures. The inspection types include partial (which detect type I failures only), imperfect (which detect only type II failures with some non-zero probability) and perfect (which detect all failures with certainty). When our parameters $\zeta = \eta = 0$ (i.e., the cost rate associated with undetected failures is constant), our model reduces to a special case of the model in [92] (namely that with no type I failures and no partial inspections). Badía and Berrade [13] consider the same special case, but with false positives. Compared to these papers, we contribute by:

(i) establishing both a necessary condition and a sufficient condition for the existence of an optimal solution under a more general, possibly nonlinear cost rate function associated with undetected failures;

(ii) establishing that the sufficient condition in both [92] &[13] is also necessary under a constant cost rate for the special case of [92] &[13] considered here;

(iii) establishing sufficient conditions for the uniqueness of a locally optimal solution for limited numbers of inspections;

(iv) developing a solution procedure to identify globally optimal solutions as opposed to the locally optimal solutions obtained in [92] &[13] as well as accompanying theoretical results regarding the asymptotic behavior of the objective function.

More specifically with regard to (i), in addition to some fixed, instantaneous, nonnegative failure cost ζ, we consider a penalty cost rate $c_3(u)$ associated with undetected failures of the form

$$c_3(u) = \lambda + \theta(u), \tag{2.1}$$

where $\lambda > 0$ and $\theta(u) \geq 0$ represent the constant and variable components of the cost rate, respectively. Furthermore, we assume that

A1: $\theta(0) = 0$, $\int_0^{+\infty} \theta(u)du = \int_0^{\Delta} \theta(u)du = \eta < +\infty$, where $0 \leq \Delta < +\infty$ and $\theta(u)$ is continuous.

In other words, λ is the long-run constant rate incurred if a hidden failure goes undetected, and the term $\theta(u)$ captures any initial nonlinearities in the cost rate. That is, assumption **A1** implies that after Δ units of time the cost rate of a hidden failure stabilizes at value λ, i.e., $\theta(u) = 0$ for $u \geq \Delta$. Assumptions similar to **A1** can be found in various disease progression models, see, e.g., [35] & [36]. Setting $\zeta = \eta = 0$ (note that if $\eta = 0$ then $\theta(u) = 0$ for $u \geq 0$) yields a constant penalty cost rate as in [92] & [13], which is commonly used in the majority of existing relevant literature (see, e.g., [13, 11, 12, 14, 17, 68, 86, 92]).

The remainder of the chapter is organized as follows. Section 2.2 presents our mathematical model and proposed solution procedure. In Section 2.3, some properties of the cost function are illustrated with several insightful numerical examples; these properties are formally considered in Section 2.4. In particular, we analytically establish conditions that guarantee the existence of a finite optimal solution for a given value of n (i.e., number of IPIs between PM), and discuss asymptotic properties of the objective function for large n and t. These results are further exploited to derive convergence properties of the proposed solution approach. Moreover, for the case of a Weibull TTF distribution, we discuss conditions that guarantee the existence of a unique optimal solution for a given value of n. Finally, Section 2.5 concludes by summarizing the results. All proofs are included in the appendix.

2.2 MODEL FORMULATION

Let the time until a hidden failure develops (following the most recent system renewal) be given by the random variable X with cdf $F_X(x)$ and pdf $f_X(x)$. In the remainder of the chapter we assume that

A2: $\lim_{t\to+\infty} tf_X(t) = 0$;
A3: $\lim_{t\to+0} F_X(t) = F_X(0) = 0$.
A4: The failure rate, i.e., $f_X(t)/(1 - F_X(t))$, is strictly increasing over $t > 0$.

Assumption **A2** guarantees that the expected time until a hidden failure develops, $\mathbb{E}[X]$, is finite. Assumption **A3** implies there are no instantaneous failures. Assumption **A4** captures the intuitive notion that the longer the time since the last PM (or RM), the more likely a hidden failure is to develop.

As mentioned in Section 2.1, the objective is to determine an optimal policy (t^*, n^*) such that the long-run average cost rate is minimized. Because PM and RM renew the system, we take a renewal-reward approach ([72], p. 52) and refer to the time between system renewals as a cycle. Note that depending on the problem parameters, n^* may be zero, i.e., it may be optimal to not perform IPIs at all. System downtime due to inspection and maintenance (including IPIs, PM and RM) is assumed to be negligible.

To illustrate the overall problem dynamics, Figure 2.1 depicts three possible cycles:

(a) no hidden failure develops and the cycle ends after a planned PM,

(b) a hidden failure occurs when there are i IPIs remaining before the next planned PM, but IPIs do not detect the failure, and

(c) a hidden failure occurs when there are i IPIs remaining before the next planned PM and one of these remaining IPIs detects the failure after which RM is performed.

Let γ_i be the probability that following PM (or RM), a hidden failure occurs when i IPIs remain prior to the next scheduled PM, $i = 0, \ldots, n$, i.e.,

$$\gamma_i = F_X\left((n - i + 1)t\right) - F_X\left((n - i)t\right).$$

Furthermore, let L_{ik} (respectively, C_{ik}) denote the cycle time (respectively, cycle cost) if a hidden failure occurs when i IPIs remain prior to the next scheduled PM and the hidden

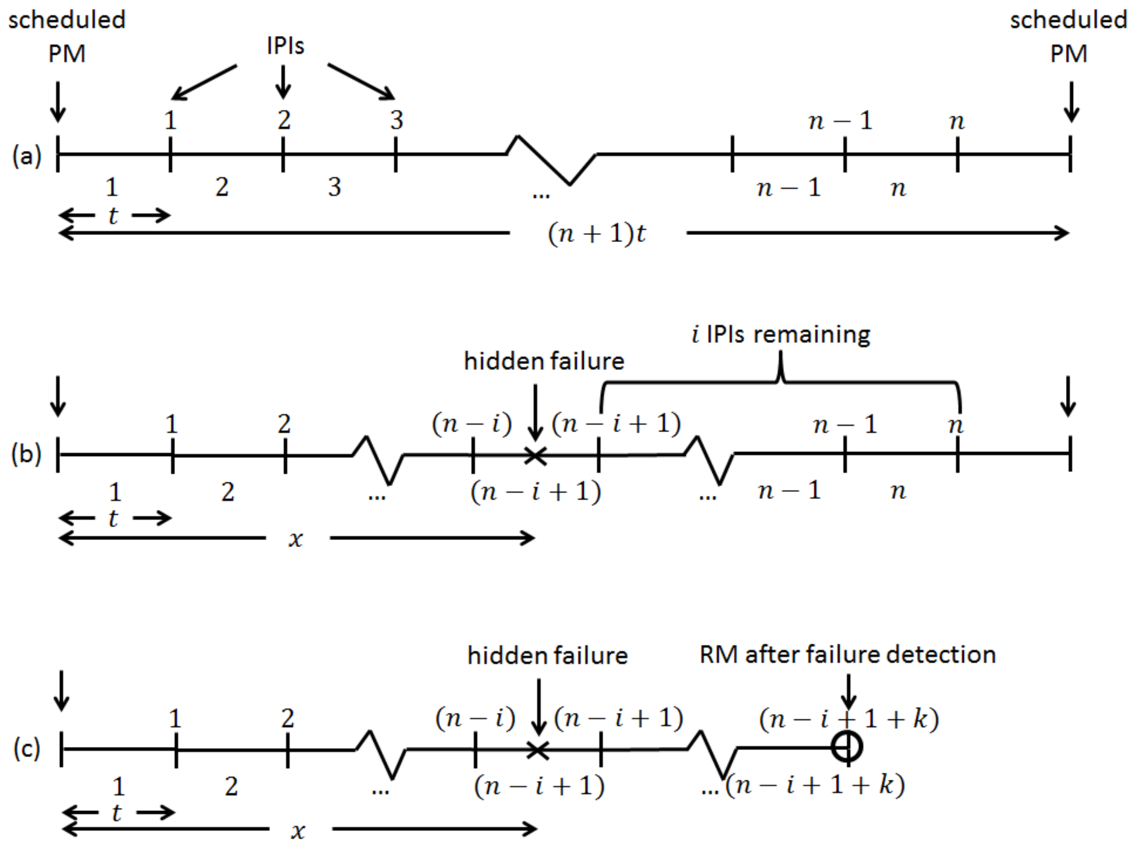

Figure 2.1: Three possible cases: (a) a cycle ends after planned PM and no hidden failure develops; (b) a cycle ends after planned PM because no IPIs detect an existing hidden failure; (c) a cycle ends after one of the remaining IPIs (i.e., $0 \leq k \leq i-1$) detects a hidden failure and RM is performed.

failure is detected after k IPIs fail to detect it, i.e., on the $(k+1)^{\text{st}}$ IPI $(k < i)$ or at the next PM $(k = i)$. Then for each case illustrated in Figure 2.1, the corresponding expected cycle length and cost can be expressed as follows.

Case (a): No hidden failure occurs within the cycle; the cycle length and cycle cost are $(n+1)t$ and $c_1 + nc_2$, respectively.

Case (b): A hidden failure occurs when there are i IPIs remaining before the next planned PM, but none of these IPIs detect the failure. In this case,

$$\mathbb{E}[L_{ii}] = (n+1)t,$$

and

$$\mathbb{E}[C_{ii}] \;=\; c_1 + nc_2 + \zeta + \int_{(n-i)t}^{(n-i+1)t} \Big(\int_0^{(n+1)t-x} (\lambda + \theta(u))du \Big) \frac{f_X(x)}{\gamma_i} dx.$$

Case (c): A hidden failure occurs when there are i IPIs remaining before the next planned PM and the $(k+1)^{\text{st}}$ detects the failure after which RM is performed, where $0 \le k \le i-1$. In this case,

$$\mathbb{E}[L_{ik}] = (n-i+1)t + kt,$$

and

$$\mathbb{E}[C_{ik}] \;=\; c_1 + c_2\,(n-i+1+k) + \zeta + \int_{(n-i)t}^{(n-i+1)t} \Big(\int_0^{(n-i+1)t-x+kt} (\lambda + \theta(u))du \Big) \frac{f_X(x)}{\gamma_i} dx.$$

Let L and C denote the cycle length and cost of an arbitrary cycle. Combining the terms above with the corresponding probabilities yields

$$\mathbb{E}[L] = \sum_{i=0}^{n} \gamma_i t \left(n - i + 1 + \sum_{j=1}^{i} (1-p)^j \right) + \quad (n+1)t\bar{F}_X\left((n+1)t\right), \tag{2.2}$$

$$\begin{aligned} \mathbb{E}[C] \;&=\; c_2 \left[\sum_{i=0}^{n} \gamma_i \left(n - i + \sum_{j=0}^{i-1} (1-p)^j \right) + n\bar{F}_X((n+1)t) \right] \\ &\quad + \sum_{i=0}^{n} \gamma_i \mathbb{E}\left[\int_0^{D_i} c_3(u)du \right] + \zeta \sum_{i=0}^{n} \gamma_i + c_1, \end{aligned} \tag{2.3}$$

where D_i is the time until the hidden failure is detected given that $(n-i)t < X < (n-i+1)t$, i.e.,

$$\mathbb{E}\left[\int_0^{D_i} c_3(u)du\right] = \lambda\mathbb{E}[D_i] + \int_{(n-i)t}^{(n-i+1)t}\left(\sum_{k=0}^{i} p_{ik}\int_0^{(n-i+1)t-x+kt}\theta(u)du\right)\frac{f_X(x)}{\gamma_i}dx,$$

and

$$\mathbb{E}[D_i] = \int_{(n-i)t}^{(n-i+1)t}((n-i+1)t-x)\frac{f_X(x)}{\gamma_i}dx + t\sum_{j=1}^{i}(1-p)^j,$$

where p_{ik} is the probability that a hidden failure is detected on the $(k+1)^{\text{st}}$ IPI ($k < i$) or at the next PM ($k = i$), i.e.,

$$p_{ik} = \begin{cases}(1-p)^k \cdot p, & \text{if } \ k < i;\\ (1-p)^i, & \text{if } \ k = i.\end{cases}$$

Next, defining

$$M(t,n) = \sum_{i=1}^{n} F_X(it) - \sum_{i=1}^{n} F_X(it)(1-p)^{n-i+1} \tag{2.4}$$

$$N(t,n) = \sum_{i=1}^{n} F_X(it) - \sum_{i=1}^{n} F_X(it)(1-p)^{n-i}, \tag{2.5}$$

$$Z(t,n) = \sum_{i=0}^{n}\int_{(n-i)t}^{(n-i+1)t}\left(\sum_{k=0}^{i} p_{ik} \times \int_0^{(n-i+1)t-x+kt}\theta(u)du\right)f_X(x)dx, \tag{2.6}$$

Equations (2.2) and (2.3) can be simplified as follows (detailed derivations are provided in the appendix)

$$\mathbb{E}[L] = (n+1)t - tM(t,n), \tag{2.7}$$

$$\begin{aligned}\mathbb{E}[C] &= c_1 + c_2(n - N(t,n)) + \lambda\left(\int_0^{(n+1)t} F_X(x)dx - tM(t,n)\right)\\ &+ \zeta F_X((n+1)t) + Z(t,n).\end{aligned} \tag{2.8}$$

The intuition behind equations (2.7)-(2.8) is that the value of $tM(t,n)$ can be interpreted as the expected decrease in the length of a failure-free cycle (given by $(n+1)t$, which is the first term in (2.7)) because of a hidden failure detection by an IPI. Similarly, $N(t,n)$ corresponds to the expected number of IPIs that are not performed during a cycle due to a hidden failure detection by an IPI. Note that $M(t,0) = N(t,0) = 0$ because $n = 0$ corresponds to not performing any IPIs. Lastly, $Z(t,n)$ represents the expected cumulative cost caused by the non-constant component $\theta(u)$ of the penalty cost rate $c_3(u)$ given by (2.1).

Let $\Omega(t,n) = \mathbb{E}[C]/\mathbb{E}[L]$ be the cost rate incurred under policy (t,n). Then our main optimization problem is formulated as

$$\min_{t,n} \Omega(t,n) = \min_{t,n} \left\{ \frac{\mathbb{E}[C]}{\mathbb{E}[L]} \;\middle|\; t > 0, n \in \mathbb{Z}^1_+ \right\}. \tag{2.9}$$

We define (t^*, n^*) to be the global optimal solution of (2.9). Next, observe that for any fixed $n \in \mathbb{Z}^1_+$, problem (2.9) reduces to a continuous optimization problem given by

$$[\mathbf{CP}_n] \quad \min_t \{\Omega(t,n) \mid t > 0\}, \tag{2.10}$$

and let t_n be the corresponding optimal solution of $\mathbf{CP}_n$, i.e.,

$$t_n \in \underset{t>0}{\operatorname{argmin}} \ \Omega(t,n).$$

Note that neither problem (2.9) nor problem (2.10) necessarily has a finite optimal solution (see additional discussion in Sections 2.3 and 2.4). For example, for a fixed $n \in \mathbb{Z}^1_+$ it is possible that the function $\Omega(t,n)$ monotonically decreases in t and $\lim_{t\to+\infty} \Omega(t,n) = \lambda$. That is, performing inspections and maintenance at any frequency may result in a higher long-run average cost rate than doing nothing. We assume for this case that $t_n = +\infty$ and $\Omega(t_n, n) = \lambda$.

Summarizing the discussion in this section, we conclude that the main problem (2.9) can be solved as a sequence of continuous optimization problems $\mathbf{CP}_n$ with the final solution given by:

$$n^* \in \underset{n}{\operatorname{argmin}} \ \Omega(t_n, n) \quad \text{and} \quad t^* = t_{n^*}. \tag{2.11}$$

Therefore, if there exists a known finite upper bound $\bar{n} \in \mathbb{Z}^1_+$ for n^*, then (t^*, n^*) can be obtained by:

(i) solving $\bar{n}+1$ continuous optimization problems $\mathbf{CP}_n$ given by (2.10) for each $n \in \{0, 1, \ldots, \bar{n}\}$, and

(ii) choosing a pair (t_n, n) that returns the minimum objective function value.

Note that in [92], the authors focus on the identification of locally optimal solutions, whereas our approach ensures global optimality as long as a finite $\bar{n}$ exists and problems $\mathbf{CP}_n$ are solved to global optimality for each integer $n \in [0, \bar{n}]$. For a detailed discussion on obtaining such a bound $\bar{n} \in \mathbb{Z}^1_+$, see Section 2.4.

2.3 MOTIVATING NUMERICAL EXAMPLES

In this section, we demonstrate that for any fixed $n \in \mathbb{Z}^1_+$ there are three possible cases for $\mathbf{CP}_n$:

Case 1 : there exists a unique local and global optimum,

Case 2 : there are multiple local minima, and

Case 3 : there is no finite optimal solution.

Analytical conditions under which each of these cases holds are formally derived in Section 2.4. For the examples demonstrated in this section, we let the random variable X follow a Weibull distribution with shape and scale parameters α and β, respectively.

Case 1: unique local and global minimum. Define

$$\theta_1(u) = \begin{cases} u/200, & \text{if } \ u \leq 50; \\ 0.25 - (u-50)/200, & \text{if } \ 50 < u \leq 100; \\ 0, & \text{otherwise}, \end{cases}$$

and consider the problem instance given by Table 2.1.

Figure 2.2 illustrates that when $n = 4$, there exists a unique optimal solution of $\mathbf{CP}_n$ and $\Omega(t, n)$ is quasiconvex. That is, if four IPIs are scheduled between two consecutive PMs, then after each renewal, IPIs and PMs should be performed every 16.34 and 81.7 units of time, respectively.

Due to the existence of a unique local minimum, a continuous optimization solver can be applied to locate the optimal solution of (2.10).

Table 2.1: Parameter values for the Case 1 example.

α	β	$\mathbb{E}[X]$	c_1	c_2	λ	p	ζ	$\theta(u)$
2	100	88.62	10	1	1	0.8	5	$\theta_1(u)$

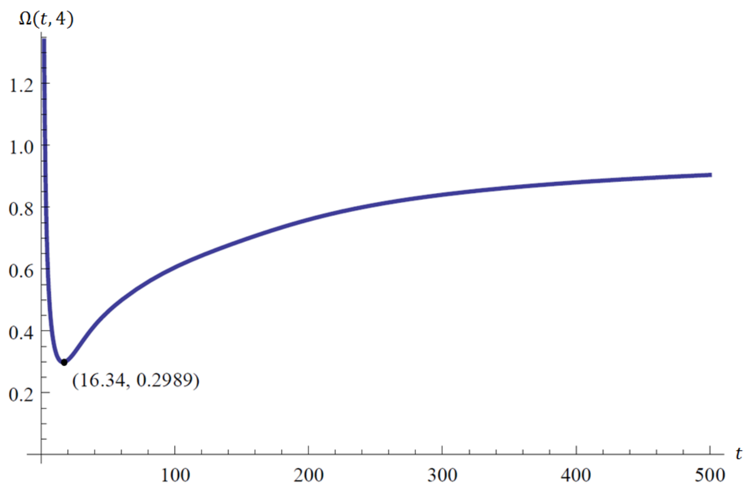

Figure 2.2: Function $\Omega(t, 4)$ for the example in Case 1; $t_4 = 16.34$ and $\Omega(t_4, 4) = 0.2989$.

Moreover, in this example, for every fixed $n \in \mathbb{Z}_+^1$ there exists a finite t_n. Enumerating all these solutions yields Figure 2.3, which makes it clear that the global optimal solution is given by $n^* = 2$, $t^* = 22.76$ and $\Omega(t^*, n^*) = 0.2953$.

However, it is also possible that $\mathbf{CP}_n$ has a unique local and global minimum, but $\Omega(t, n)$ is not quasiconvex. For example, let

$$\theta_2(u) = \begin{cases} u/20, & \text{if } \ u \leq 50; \\ 2.5 - (u - 50)/20, & \text{if } \ 50 < u \leq 100; \\ 0, & \text{otherwise.} \end{cases}$$

When compared to Figure 2.2, Figure 2.4 illustrates that if the portion of the penalty cost attributable to either ζ or $\theta(u)$ is sufficiently large, then $\Omega(t, n)$ may not be quasiconvex. The intuition behind this observation is that for larger values of t the penalty cost of an undetected failure is dominated by the limiting behavior of $\lim_{u \to +\infty} c_3(u) = \lambda$ (recall Assumption **A1**). However, there may exist a range of values of t for which either ζ or $\theta(u)$ makes a sufficiently large contribution to the objective function value resulting in a local maximum of $\Omega(t, 4)$. Regardless, for this numerical example, $\mathbf{CP}_n$ still has a unique local and global minimum.

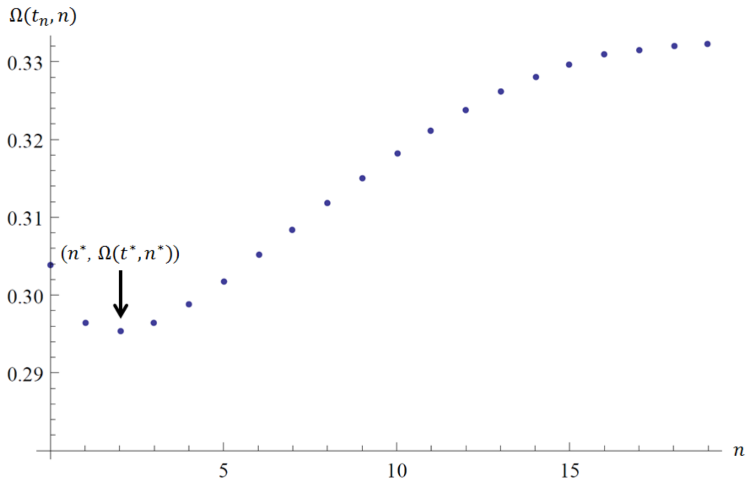

Figure 2.3: Function $\Omega(t_n, n)$ for the example in Case 1; $n^* = 2$, $t^* = 22.76$ and $\Omega(t^*, n^*) = 0.2953$.

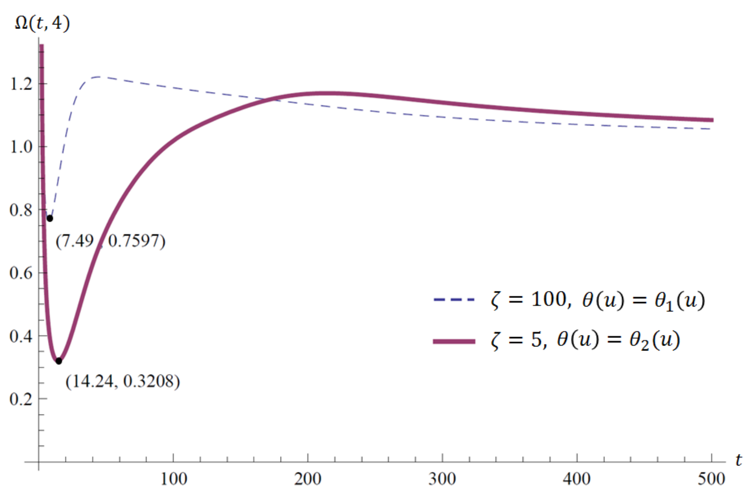

Figure 2.4: Potential impacts of ζ and $\theta(u)$ on $\Omega(t, 4)$; the quasiconvexity property is lost.

Case 2: multiple local minima. Table 2.2 summarizes parameter values for the example illustrating this case. Figure 2.5 depicts $\Omega(t, 4)$, which has multiple local minima. This type of behavior arises in situations where the probability of detecting a hidden failure is sensitive to the choice of t (see additional discussion in Section 2.4.2). Thus, here $\mathbf{CP}_n$ is a more challenging optimization problem than the example considered in Case 1. In general, it requires application of a global optimization method [39]. On the other hand, $\mathbf{CP}_n$ has only one variable, which substantially simplifies the solution procedure. For the example in Table 2.2, enumerating $\Omega(t_n, n)$ results in Figure 2.6 with $n^* = 5$, $t^* = 12.54$ and $\Omega(t^*, n^*) = 0.1665$. Therefore, the optimal maintenance policy is to schedule 5 IPIs between PMs with IPIs performed every 12.54 units of time.

Table 2.2: Parameter values for the Case 2 example.

α	β	$\mathbb{E}[X]$	c_1	c_2	λ	p	ζ	$\theta(u)$
6	100	92.77	10	0.1	1	0.8	5	$\theta_1(u)$

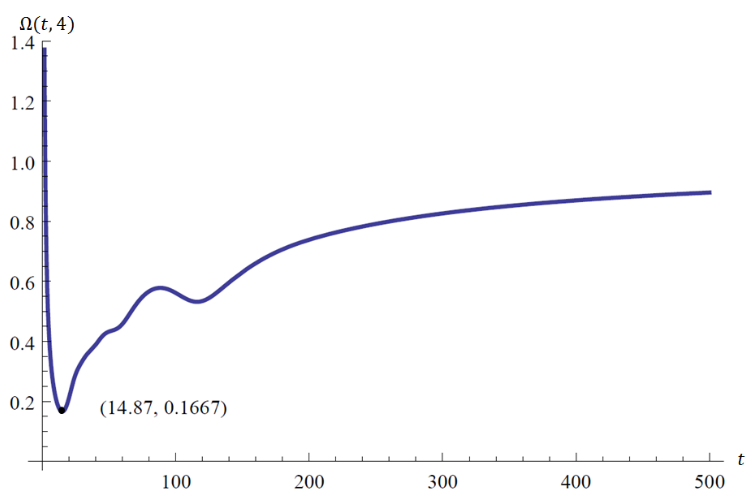

Figure 2.5: Function $\Omega(t, 4)$ for the example in Case 2; $t_4 = 14.87$ and $\Omega(t_4, 4) = 0.1667$.

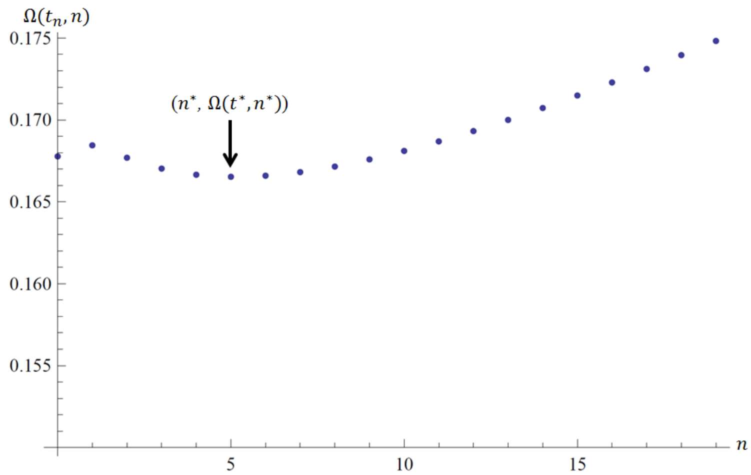

Figure 2.6: Function $\Omega(t_n, n)$ for the example in Case 2; $n^* = 5$, $t^* = 12.54$ and $\Omega(t^*, n^*) = 0.1665$.

Case 3: no finite optimal solution. As shown in Figure 2.7, under certain conditions for a fixed n (see example parameter values in Table 2.3 and the formal result established in Proposition 4 in Section 2.4), it is possible that the function $\Omega(t, n)$ is monotonically decreasing in t, i.e., there is no finite optimal solution for $\mathbf{CP}_n$, i.e., $t_n = +\infty$. In this case, no IPIs or PM should be performed as they are too costly and it is favorable to simply let the system fail. The intuition behind the irregular shape of $\Omega(t, 4)$ for non-zero values of ζ and $\theta(u)$ in Figure 2.7 is the same as that given for Figure 2.4.

Table 2.3: Parameter values for the Case 3 example.

α	β	$\mathbb{E}[X]$	c_1	c_2	λ	p
2	100	88.62	10	1	0.1	0.8

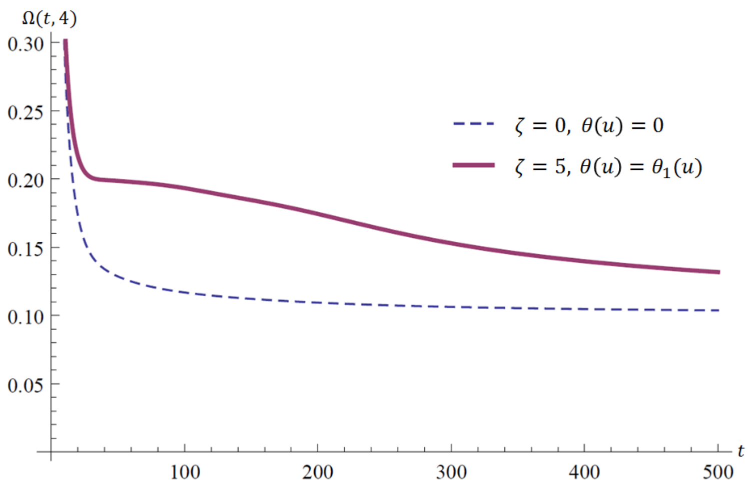

Figure 2.7: Function $\Omega(t, 4)$ is monotonically decreasing for the example in Case 3, i.e., $t_4 = +\infty$.

2.4 ANALYTICAL RESULTS

This section focuses on establishing theoretical results for optimization problems (2.10) and (2.11). Our discussion is motivated by the illustrative examples provided in Section 2.3. First, for a general distribution of the random variable X, i.e., the time until a hidden failure develops, or time-to-failure (TTF), we establish conditions on the problem parameters, specifically, c_1, c_2, λ, ζ, η, p and $\mathbb{E}[X]$, that ensure the existence of a finite optimal solution for $\mathbf{CP}_n$, $n \in \mathbb{Z}^1_+$. Furthermore, we provide some results on the existence of a finite bound $\bar{n} \in \mathbb{Z}^1_+$ that guarantees the convergence of the solution procedure (presented in Section 2.2) to global optimality. Finally, we discuss the intuition behind problem instances of (2.10) with multiple local minima and derive some sufficient conditions for uniqueness of a locally optimal solution under a Weibull TTF distribution.

2.4.1 General TTF Distribution

First, we consider the limiting behavior of $\Omega(t, n)$ for $t \to +\infty$ and any fixed $n \in \mathbb{Z}^1_+$.

Proposition 1. For any fixed $n \in \mathbb{Z}_+^1$,

$$\lim_{t \to +\infty} \Omega(t, n) = \lambda.$$

Proposition 1 is quite intuitive as it corresponds to the scenario in which no inspections, and hence no maintenance, are performed. In this case, the objective function value converges to λ because in this limit, a hidden failure eventually occurs and remains undetected. Next, we focus on obtaining sufficient conditions for the existence of a finite optimal solution to $\mathbf{CP}_n$. Initially, we consider the case without IPIs, i.e., $n = 0$.

Proposition 2. If $c_1 + \zeta + \eta < \lambda \mathbb{E}[X]$, then problem $\mathbf{CP}_0$ has a finite optimal solution, i.e., there exists t_0, such that $0 < t_0 < +\infty$, and

$$t_0 \in \operatorname*{argmin}_{t>0} \ \Omega(t, 0).$$

Moreover, if $\zeta = \eta = 0$, then $\Omega(t, 0)$ is quasiconvex and t_0 is unique.

The intuition behind Proposition 2 follows from the fact that if failures occur ideally, i.e., such that they are detected by PM almost immediately, then the expected cost rate is bounded above by $(c_1 + \zeta + \eta)/\mathbb{E}[X]$. On the other hand, if no PMs are performed, then after a hidden failure develops (recall that $\mathbb{E}[X]$ is finite under assumption **A3**) the cost rate for sufficiently large values of t is equal to λ. Thus, if $\lambda > (c_1 + \zeta + \eta)/\mathbb{E}[X]$, then it is beneficial to perform PMs.

Proposition 3 extends Proposition 2 to any $n \in \mathbb{Z}_+^1$. Unfortunately, the uniqueness of a locally optimal solution is not guaranteed even if $\zeta = \eta = 0$ in this general case. Furthermore, the next result ensures only a sufficient condition.

Proposition 3. For any fixed $n \in \mathbb{Z}_+^1$, a sufficient condition for problem $\mathbf{CP}_n$ to have a finite optimal solution $t_n \in \operatorname*{argmin}_{t>0} \ \Omega(t, n)$ such that $0 < t_n < +\infty$, is

$$c_1 + c_2 \sum_{i=1}^{n} (1 - p)^{n-i} + \zeta + \eta < \lambda \mathbb{E}[X]. \tag{2.12}$$

The intuition behind Proposition 3 is rather similar to that of Proposition 2. Specifically, the term $c_2 \sum_{i=1}^{n}(1-p)^{n-i}$ represents an additional expected cost incurred by performing IPIs (until they detect a hidden failure if one occurs). However, in contrast to the previous result for $\zeta = \eta = 0$, the function $\Omega(t, n)$ may have multiple local minima if $n > 0$. We provide additional discussion on this issue in Section 2.4.2. Note that (2.12) can easily be verified for the examples in Figures 2.2 and 2.5, for which the term on the left-hand side of (2.12) is equal to 28.75 and 27.62, respectively, which are smaller than the values of $\lambda\mathbb{E}[X]$ equal to 88.62 and 92.77, respectively.

Next, Proposition 4 shows that a relaxed version of (2.12) can be used to derive a necessary condition for the existence of a finite optimal solution.

Proposition 4. For any fixed $n \in \mathbb{Z}_{+}^{1}$, a necessary condition for problem $\mathbf{CP}_n$ to have a finite optimal solution $t_n \in \underset{t>0}{\operatorname{argmin}}\ \Omega(t, n)$ such that $0 < t_n < +\infty$, is

$$c_1 + c_2 \sum_{i=1}^{n}(1-p)^{n-i} < \lambda\mathbb{E}[X]. \tag{2.13}$$

Note that for the example in Figure 2.7, $c_1 + c_2 \sum_{i=1}^{n}(1-p)^{n-i} = 11.25$, while $\lambda\mathbb{E}[X] = 8.86$. Thus, (2.13) does not hold, and $\mathbf{CP}_n$ does not have a finite optimal solution.

Note that in both [92] and [13], the authors establish a sufficient condition for the existence of a finite optimal solution, which reduces to (2.12) when assuming $\zeta = \eta = 0$. We not only extend their result to our more general penalty cost setting, but also prove in Proposition 4 that (2.13) is a necessary condition. Thus, if (2.13) does not hold, then the total cost of performing IPIs and PM is too costly, and one should simply let the system fail. Moreover, if $\zeta = \eta = 0$ then $\mathbf{CP}_n$ has a finite optimal solution if and only if (2.13) holds, as stated in Corollary 1.

Corollary 1. *If $\zeta = \eta = 0$, i.e., $c_3(u) = \lambda$, then (2.13) is a necessary and sufficient condition for problem $\mathbf{CP}_n$ to have a finite optimal solution.*

Furthermore, note that $\lim_{n\to+\infty} \sum_{i=1}^{n}(1-p)^{n-i} = 1/p$. Thus, if

$$c_1 + \frac{c_2}{p} > \lambda\mathbb{E}[X], \tag{2.14}$$

then it is straightforward to show that there exists a finite $\bar{n} \in \mathbb{Z}^1_+$ (which is also easy to find, e.g., using binary search) such that for all integers $n > \bar{n}$ constraint (2.13) is not satisfied. Therefore, to solve the overall optimization problem (2.9) it is enough to consider $\mathbf{CP}_n$ only for $n \in \{0, 1, \ldots, \bar{n}-1, \bar{n}\}$.

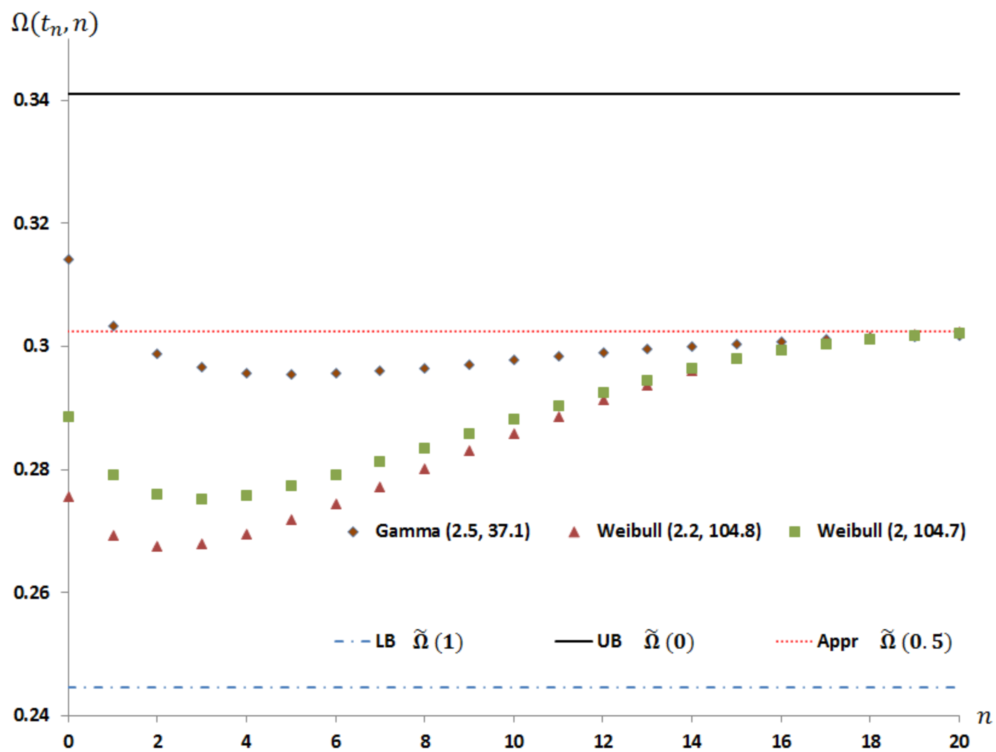

Figure 2.8: Comparison of $\Omega(t_n, n)$, $\widetilde{\Omega}(0)$, $\widetilde{\Omega}(0.5)$ and $\widetilde{\Omega}(1)$ for $c_1 = 10$, $c_2 = 0.8, \lambda = 1$, $\zeta = 5$, $\eta = 0$ and $p = 0.8$.

Next, consider the situation in which (2.12) holds for all $n \in \mathbb{Z}^1_+$, which is clearly the case if

$$c_1 + \frac{c_2}{p} + \zeta + \eta \leq \lambda \mathbb{E}[X],$$

which then implies that for any positive integer n, optimization problem $\mathbf{CP}_n$ has a finite optimal solution. This observation poses the question as to whether there exists a finite $\bar{n}$ such that all subproblems $\mathbf{CP}_n$ for $n \geq \bar{n}$ can be discarded, which ensures the finiteness of the procedure outlined in Section 2.2 for solving (2.11). To explore this issue, we study asymptotic properties of $\Omega(t,n)$ as $n \to +\infty$ in Propositions 5 and 6.

First, we consider the simpler case of $\eta = 0$.

Proposition 5. If $\eta = 0$, then for any fixed $t > 0$,

$$\lim_{n\to+\infty} \Omega(t,n) = \frac{c_1 + \zeta - \lambda \mathbb{E}[X]}{\lim\limits_{n\to+\infty} \mathbb{E}[L]} + \frac{c_2}{t} + \lambda. \tag{2.15}$$

Estimation of the term $\lim\limits_{n\to+\infty} \mathbb{E}[L]$ is difficult in the general case. However, observe that for large enough n, the value of $\mathbb{E}[L]$ can be approximately lower- and upper-bounded by

$$\mathbb{E}[X] + \left(\frac{1}{p} - 1\right) t \leq \mathbb{E}[L] \leq \mathbb{E}[X] + \frac{1}{p} t, \tag{2.16}$$

where the term $\frac{1}{p}$ corresponds to the expected number of IPIs necessary for detecting a failure after it occurs. The approximate lower (and upper) bound is obtained assuming that a failure occurs immediately before (or after) an IPI.

Combining (2.15) and (2.16) we define

$$\widetilde{\Omega}(y) = \min_{t>0} \left\{ \frac{c_1 + \zeta - \lambda \mathbb{E}[X]}{\mathbb{E}[X] + \left(\frac{1}{p} - y\right) t} + \frac{c_2}{t} + \lambda \right\}, \tag{2.17}$$

so that $\widetilde{\Omega}(1)$ and $\widetilde{\Omega}(0)$ correspond to the lower and upper bounds in (2.16), respectively. Figure 2.8 provides an illustrative comparison of $\Omega(t_n, n)$ with $\widetilde{\Omega}(0)$, $\widetilde{\Omega}(0.5)$ and $\widetilde{\Omega}(1)$ for several examples of F_X.

Unfortunately, it turns out to be difficult to establish any analytical relationship between $\lim\limits_{n\to+\infty} \Omega(t_n, n)$ and $\widetilde{\Omega}(y)$ in the general case. However, while setting $y = 0$ (or $y = 1$) overestimates (or underestimates, respectively) the expected length of a cycle, a reasonable choice could be $y = 0.5$. In fact, for all of our test instances, $\widetilde{\Omega}(0.5)$ serves as a reasonably good approximation for characterizing the limiting behavior of $\lim\limits_{n\to+\infty} \Omega(t_n, n)$ (see Figure 2.8 and Table 2.4).

Let $\widetilde{y}$ be such that $\widetilde{\Omega}(\widetilde{y}) = \lim\limits_{n\to+\infty} \Omega(t_n, n)$. Then for any given $\epsilon > 0$ there exists a finite $\bar{n}$ to satisfy

$$\left| \Omega(t_{\bar{n}}, \bar{n}) - \widetilde{\Omega}(\widetilde{y}) \right| \leq \epsilon.$$

Table 2.4: Example for $c_1 = 10$, $c_2 = 0.8$, $\lambda = 1$, $\zeta = 5$, $\eta = 0$, $p = 0.8$ and $X \sim$ Weibull$(2, 104.7)$. The optimal solution corresponds to the row in bold, i.e., $n^* = 3$ and $t^* = 19.43$.

n	$\lvert\Omega(t_n, n) - \widetilde{\Omega}(0.5)\rvert$	$\Omega(t_n, n)$	t_n
0	0.01397	0.28856	56.15
1	0.02332	0.27921	32.15
2	0.02653	0.27599	23.75
3	**0.02724**	**0.27529**	**19.43**
4	0.02662	0.27591	16.79
⋮	⋮	⋮	⋮
10	0.01439	0.28814	11.03
⋮	⋮	⋮	⋮
20	0.00043	0.30209	11.47
⋮	⋮	⋮	⋮
30	$< 10^{-7}$	0.30253	11.91

Table 2.5: Example for $c_1 = 10$, $c_2 = 0.7, \lambda = 1$, $p = 0.8$, $\zeta = 5$, $\Delta = 4$, $\eta = 2$, $\theta(u) = \theta_3(u)$, $\widetilde{n} = 75$ and $X \sim \text{Weibull}(2, 100)$. The optimal solution corresponds to the row in bold, i.e., $n^* = 3$ and $t^* = 18.20$.

n	$\lvert\Omega(t_n, n) - \widetilde{\Omega}(0.5)\rvert$	$\Omega(t_n, n)$	t_n
0	0.01463	0.30663	53.70
1	0.02469	0.29657	30.51
2	0.02843	0.29282	22.38
3	**0.02954**	**0.29171**	**18.20**
4	0.02923	0.29203	15.63
⋮	⋮	⋮	⋮
10	0.01763	0.31998	9.95
⋮	⋮	⋮	⋮
20	0.00094	0.32032	9.90
⋮	⋮	⋮	⋮
30	$< 10^{-6}$	0.32126	10.75

Thus, it is enough to solve $\bar{n}+1$ optimization problems $\mathbf{CP}_n$ for $n \in \{0, 1, \ldots, \bar{n}\}$ to guarantee that the obtained solution to (2.11) is at least ϵ-approximate. To illustrate the overall solution procedure, consider the numerical example in Table 2.4 with $\epsilon = 10^{-7}$ under the assumption that $\widetilde{y} = 0.5$. At $n = 30$, $|\Omega(t_n, n) - \widetilde{\Omega}(0.5)| \leq 10^{-7}$, hence the procedure terminates at $\bar{n} = 30$ with $n^* = 3$, $t^* = 19.43$ and $\Omega(t^*, n^*) = 0.2753$. In general, if $|\widetilde{\Omega}(\widetilde{y}) - \Omega(t_n, n)|$ does not appear to be converging for the selected value of $\widetilde{y}$, then one could simply generate a finite sequence $y_1, \ldots, y_K$, where $y_1 = 0$ and $y_K = 1$, such that $|\widetilde{\Omega}(y_k) - \widetilde{\Omega}(y_{k+1})| < \epsilon$ for $k = 1, \ldots, K-1$ and find $\bar{n}$.

Next, we focus on the general case when $\eta > 0$. We first consider the convergence of $Z(t, n)$ for $n \to \infty$ in Lemma 1.

Lemma 1. For any $t > \Delta$, $\lim\limits_{n \to +\infty} Z(t, n)$ *exists and is finite.*

Proposition 6 can be viewed as a generalization of Proposition 5.

Proposition 6. For any $t > \Delta$,

$$\lim_{n \to +\infty} \Omega(t, n) = \frac{c_1 + \zeta + \lim\limits_{n \to +\infty} Z(t, n) - \lambda \mathbb{E}[X]}{\lim\limits_{n \to +\infty} \mathbb{E}[L]} + \frac{c_2}{t} + \lambda. \tag{2.18}$$

Proposition 6 suggests that if $t > \Delta$, then one can use (2.18) to approximate the limiting behavior of $\lim\limits_{n \to +\infty} \Omega(t, n)$ as follows. Let

$$\widetilde{\Omega}(y) = \min_{t > 0} \left\{ \frac{c_1 - \lambda \mathbb{E}[X] + \zeta + \widetilde{Z}(t)}{\mathbb{E}[X] + \left(\frac{1}{p} - y\right) t} + \frac{c_2}{t} + \lambda \right\}, \tag{2.19}$$

where

$$\widetilde{Z}(t) = (1-p)\eta + p \int_0^{\Delta} \theta(u) \sum_{i=1}^{\widetilde{n}} \Big(F_X(it - u) - F_X((i-1)t) \Big) du \tag{2.20}$$

is an approximation of $\lim_{n\to+\infty} Z(t,n)$ for sufficiently large values of $\widetilde{n} \in \mathbb{Z}_+^1$. Additional derivation details can be found in the appendix. Finally, Table 2.5 provides a comparison of $\Omega(t_n, n)$ and $\widetilde{\Omega}(0.5)$ using

$$\theta_3(u) = \begin{cases} u/2, & \text{if } \; u \le 2; \\ 1-(u-2)/2, & \text{if } \; 2 < u \le 4; \\ 0, & \text{otherwise,} \end{cases}$$

which demonstrates the applicability of our solution approach for problems with non-zero values of ζ and η.

2.4.2 Weibull TTF Distribution

In this section, we assume that X follows a Weibull distribution and seek conditions that guarantee the existence of a unique locally optimal solution for $\mathbf{CP}_n$ (recall the example in Figure 2.2), which simplifies the solution of $\mathbf{CP}_n$. Moreover, we show the presence of multiple local minima implies that the system can be rather sensitive to the choice of t.

Special Case: $n = 1$ and $\zeta = \eta = 0$. Observe that when $n = 1$ and $\zeta = \eta = 0$, equations (2.4) and (2.5) reduce to $M(t,n) = pF_X(t)$ and $N(t,n) = 0$, respectively. Therefore, our objective function simplifies to

$$\Omega(t,1) \quad = \quad \frac{c_1 + c_2 - \lambda \int_0^{2t} \bar{F}_X(x)dx}{2t - tpF_X(t)} + \lambda, \tag{2.21}$$

and Proposition 7 result follows.

Proposition 7. Let X follow a Weibull distribution with shape parameter $\alpha > 1$ and $\zeta = \eta = 0$. If

$$c_1 + c_2 < \lambda \mathbb{E}[X], \tag{2.22}$$

and

$$2 - p - \alpha p \frac{1}{e} > 0, \tag{2.23}$$

then problem $\mathbf{CP}_1$ has a unique local minimum.

Note that by Proposition 3, constraint (2.22) guarantees the existence of a finite optimal solution for $n = 1$ and $\zeta = \eta = 0$; furthermore, (2.23) describes a sufficient condition on α and p to ensure that the solution is unique. The condition $\alpha > 1$ is necessary to satisfy Assumption **A4**.

Define $\hat{\alpha}_{n,p}$ to be the minimal value of α for fixed n and p, such that $\mathbf{CP}_n$ has a unique locally optimal solution (if such finite solution exists). Figure 2.9 presents an example that demonstrates relationship between $\hat{\alpha}_{1,p}$ and p for $n = 1$ based on the sufficient condition (2.23) from Proposition 7 and the results obtained through our computational observations.

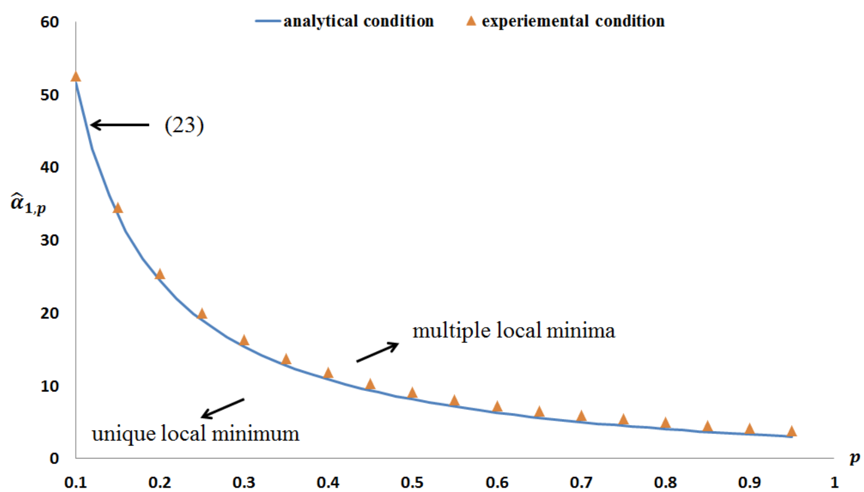

Figure 2.9: Comparison of conditions for unique local optimality of $\mathbf{CP}_1$ established analytically and experimentally with $\beta = 100$, $c_2 = \lambda = 1$, $c_1 = 10$ and $\zeta = \theta(u) \equiv 0$.

To explain the intuition behind the analytical results and experimental observations, consider the schematic in Figure 2.10. For the Weibull distribution, the shape parameter α completely determines the coefficient of variation of X. For larger values of α, $f_X(x)$ becomes less flat, i.e., a hidden failure develops with a very high probability within a small interval of time. Therefore, scheduling an IPI around the mode of $f_X(x)$ substantially influences the outcome of the optimization.

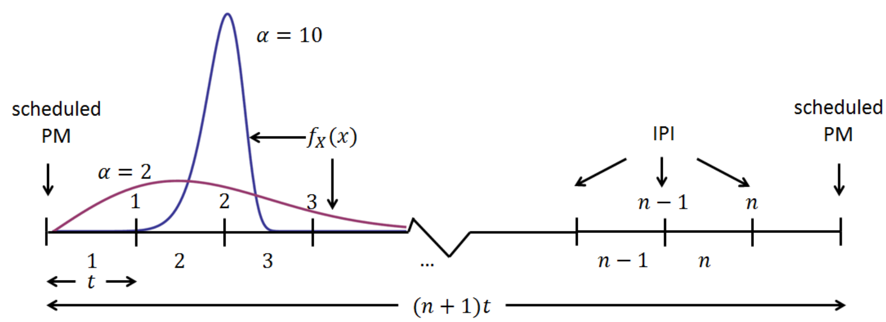

Figure 2.10: The influence of the shape parameter, α, on $f_X(x)$.

To be more specific, Figure 2.11 demonstrates an instance of $\mathbf{CP}_1$ with multiple local minima and maxima given by $t^{(1)} = 41$, $t^{(2)} = 88$ and $t^{(3)} = 110$, where $\alpha = 10$ and $p = 0.8$. (Recall that γ_i is the probability that following a PM, a hidden failure occurs when i IPIs remain prior to the next scheduled PM, $i = 0, \dots, n$.) As t increases from 0, $\Omega(t, 1)$ decreases and reaches its first local minimum at $t = t^{(1)}$. The benefit of this policy is very intuitive as it corresponds to scheduling PM close to the mode of $f_X(x)$, whereas policies corresponding to smaller values of t require inspection and PM too early. For $t = t^{(2)}$ a hidden failure occurs with high probability after the IPI ($\gamma_0 = 0.75$), thus, the performed inspection is essentially wasted. Therefore, the increased penalty cost makes policy $t = t^{(2)}$ inferior, and this point defines a local maximum of function $\Omega(t, 1)$. For policy $t = t^{(3)}$ a hidden failure develops before the IPI with probability close to 1 ($\gamma_1 = 0.92$). Note that each IPI is rather reliable (recall that $p = 0.8$). Thus, the failure (if it occurs) is detected with high probability. As a result, the cost advantage of IPIs (recall that $c_2 < c_1$) and reduced penalty cost in expectation makes this policy favorable, and $\Omega(t, 1)$ has a second local minimum at $t = t^{(3)}$.

Based on the discussion of the example illustrated in Figures 2.10–2.11, we conclude that instances of $\mathbf{CP}_1$ with smaller values of α and p result in optimization problems with unique local minimum. Moreover, for every fixed α, if p is small enough, then $\mathbf{CP}_1$ has a unique local minimum. This inverse relationship between α and p is represented by sufficient condition (2.23).

General Case. For general case of $n \geq 2$ as well as nonzero values of ζ and η, it turns out to be rather difficult to establish a closed form condition for uniqueness of a locally optimal solution to $\mathbf{CP}_n$. However, motivated by the intuition behind Proposition 7, we numerically explore and illustrate the relationship between $\hat{\alpha}_{n,p}$ and p for $n \geq 2$.

First, we focus on the simpler case of $\zeta = \eta = 0$. Consider the example in Figure 2.12 for which $\beta = 100$, $c_1 = 10$, $c_2 = \lambda = 1$ and $\zeta = \eta = 0$. Given $p \in (0, 1)$ and $n \in \mathbb{Z}^1_+$, we numerically identify $\hat{\alpha}_{n,p}$ such that problem $\mathbf{CP}_n$ has a unique local and global minimizer. For each value of $n \geq 2$, the shape of the function $\hat{\alpha}_{n,p}$ is similar to that of $\hat{\alpha}_{1,p}$. However, for any fixed p, as n increases, $\hat{\alpha}_{n,p}$ decreases. The reason behind this behavior is that stationary points of $\mathbf{CP}_n$ are often located around t such that $kt \sim \mathbb{E}[X]$, where k varies from 1 to n (recall the intuition behind the example illustrated in Figure 2.11), which implies that for larger values of n there are more opportunities for local optimality. Thus, conditions under which a unique local minimum exists are more strict.

Another interesting experimental observation is that as n increases, $\hat{\alpha}_{n,p}$ seems to converge to some finite value $\hat{\alpha}_p$. We interpret this observation as follows. For a fixed value of t and a large enough n, a hidden failure occurs with high probability. Also, as IPIs can be regarded as Bernoulli trials, the expected number of IPIs to detect a hidden failure is $1/p$, which implies that most of the subsequent IPIs and PM are typically not performed. Thus, for large enough n, as it can be observed from (2.15), solutions of $\mathbf{CP}_n$ (and, subsequently the uniqueness conditions for all n) should coincide.

Next, we present an example with nonzero values of ζ and η for which a similar relationship between $\hat{\alpha}_{n,p}$ and p can be observed. Let $\beta = 100$, $c_1 = 10$, $c_2 = \lambda = 1$, $\zeta = 5$ and $\theta = \theta_4(u)$ given by

$$\theta_4(u) = \begin{cases} u/10, & \text{if } \ 0 < u \leq 25; \\ 2.5 - (u - 25)/10, & \text{if } \ 25 < u \leq 50; \\ 0, & \text{otherwise.} \end{cases}$$

Figure 2.13 reports the same behavior as seen in Figure 2.12.

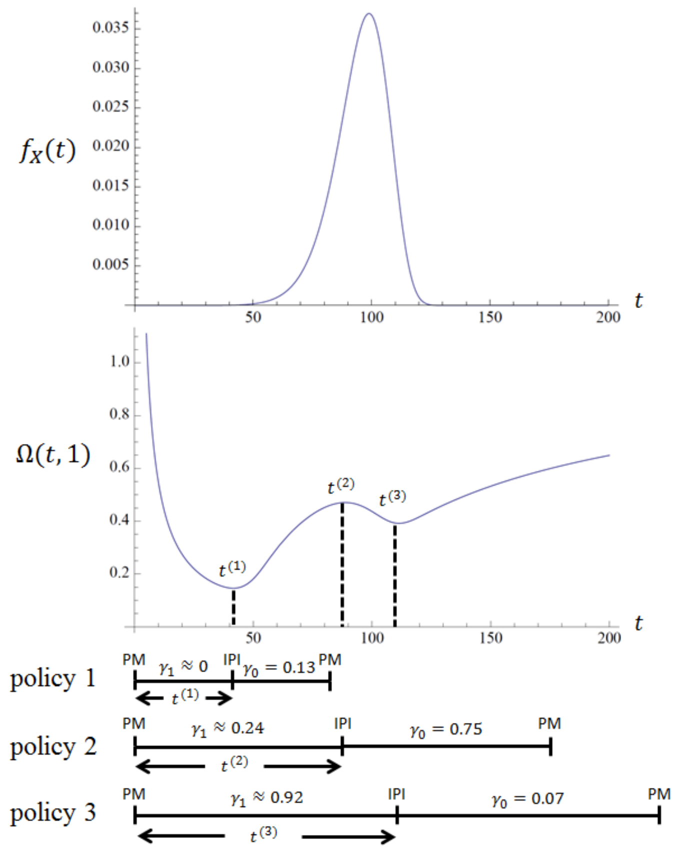

Figure 2.11: Comparison of three policies for $\mathbf{CP}_1$ with multiple local minima.

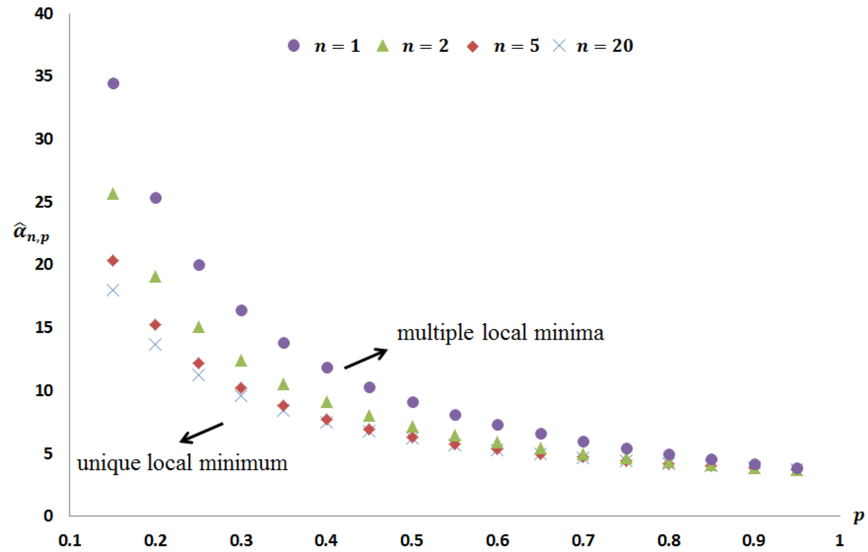

Figure 2.12: Relationship between $\hat{\alpha}_{n,p}$ and p for $n \in \{1, 2, 5, 20\}$ obtained experimentally for $\zeta = \eta = 0$.

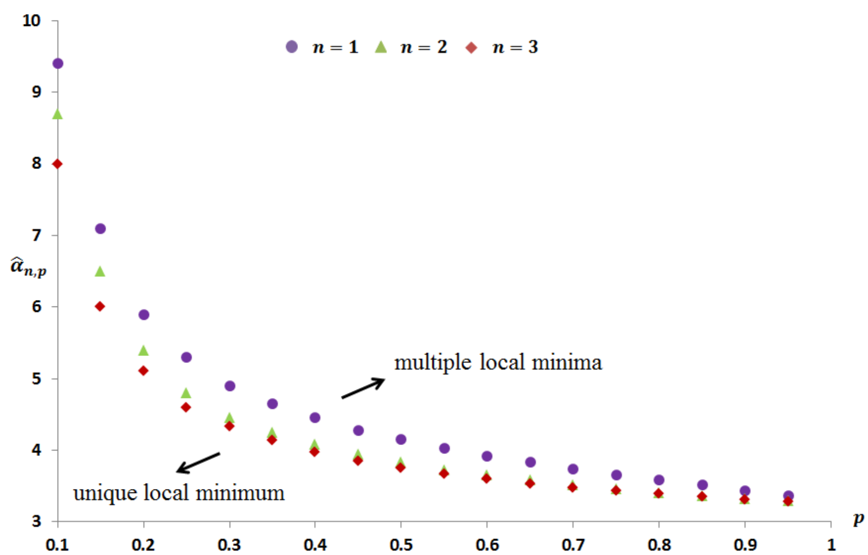

Figure 2.13: Relationship between $\hat{\alpha}_{n,p}$ and p for $n \in \{1, 2, 3\}$ obtained experimentally when $\zeta = 5$ and $\theta(u) = \theta_4(u)$.

2.5 CONCLUSION

In this chapter, we consider a maintenance optimization model for a system with periodic preventive maintenance and periodic imperfect inspections to detect hidden failures. The objective is to determine the optimal frequency and quantity of imperfect inspections between PM such that the total expected cost rate is minimized over an infinite horizon. We describe, both analytically and numerically, important structural properties of the model and propose a simple approach for finding a globally optimal solution.

3.0 OPTIMAL PLANNING OF UNPUNCTUAL PREVENTIVE MAINTENANCE

3.1 INTRODUCTION

The success of a policy in any context depends not only on how well it is constructed, but also on how well it is implemented. Moreover, in many settings, policy specification and policy implementation are carried out by separate parties [45]. For example, in a supply chain context, manufacturers set inventory replenishment schedules, but the suppliers' deliveries may not be on time. In a healthcare setting, doctors recommend screening policies to detect early-stage cancers, but patients may not adhere to the schedule ([54, 56, 67, 85]). In situations like these, upstream decision makers can benefit by adjusting their prescribed policies in anticipation of downstream deviations (e.g., [56]). Indeed, as [10] state in their study on the effects of personality on punctuality, "there are even cases when we adjust to someone's assumed (un)punctuality: for example, we make an appointment for 7 p.m. if we want to meet that person at 8 p.m."

To explore the gains (losses) associated with anticipating (or not anticipating) unpunctual policy implementation, we focus on yet another setting, namely preventive maintenance (PM) planning. More specifically, we consider a maintenance planner who prescribes a maintenance policy for a degrading system, but relies on a maintenance worker to implement the policy. The maintenance worker, however, may be unpunctual, i.e., the maintenance activities may be performed earlier or later than intended. We focus on preventive maintenance planning because of its well-established literature [22, 29, 32, 43, 58, 77, 87, 90], and its practical importance. Indeed, maintenance spending is well-known to comprise a large portion of operating budgets for organizations with heavy machinery and significant equipment investments [73].

Reasons behind unpunctual behavior have been well established in the psychology literature. Lau et al. [52] conduct a qualitative and quantitative review of possible factors affecting counterproductive behaviors (CPBs), including lateness. Predictors of CPBs are classified into four categories: personal, organizational, work and contextual. They find that employees with low job satisfaction engage in more CPBs. There also exist a number of studies that examine the influence of personality on behavioral indicators of punctuality. Back et al. [10], for example, find that punctuality may be predicted by the Big Five personality factors (openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism); link conscientiousness to punctuality; and link agreeableness and neuroticism to earliness. Koslowsky [50] also discusses the role of conscientiousness in lateness behavior. Others (e.g., [30, 84]) investigate the relationship between task type and procrastination. Lastly, additional examples of quantitative analysis of unpunctual behavior can be found in the medical appointment literature (e.g., [20, 48]).

In a maintenance context, the mistiming of activities can be especially costly. In general, delayed PM could result in a more deteriorated system, thereby increasing the likelihood of failure. On the other hand, if PM is performed earlier than scheduled, the useful life of the component is unnecessarily truncated. A more specific example is that of machine bearings, which often “run hot” due to either too little grease because PM was performed late, or due to overgreasing because PM was performed early ([9]); both of these situations can lead to high operating temperatures, which shorten the bearings’ lifetime and incur additional costs.

The literature on *imperfect maintenance* (i.e., maintenance activities that result in an outcome other than certain, as-good-as-new status) is vast (see early work [16] and survey papers [71, 90]). In contrast to perfect repairs/replacements (i.e., those that render the system as-good-as-new) and minimal repairs (i.e., those that render the system as-bad-as-old), the outcome of imperfect maintenance lies somewhere between these two extremes and may be stochastic. Here, we restrict our attention to perfect preventive maintenance activities, but allow their implementation to be unpunctual.

The class of maintenance optimization problems involving *random replacement policies* ([16, 21, 63, 93, 94]) is perhaps most closely related to our work. In random replacement scenarios, as in ours, the times at which PM is performed are random. However, the literature on random replacement policies assumes that these times are stochastic because opportunities for performing PM arise randomly due to the variable work cycle of the system, and determine the optimal parameter values of the distribution that governs the time between replacements. In contrast, we assume that the potential unpunctuality of the maintenance worker is what causes the actual PM times to deviate from the intended times in a stochastic way, and determine the optimal planned PM time in anticipation of this unpunctuality. That said, our formulation can be viewed as a random replacement problem, but with a different motivation.

However, the particular random replacement problem on which we focus has not been explored previously. For this novel case, we provide in-depth analysis of the impact of unpunctual behavior on maintenance planning. In particular, we obtain insights as to how the maintenance planner should optimally prescribe PM in anticipation of the maintenance worker's unpunctual behavior, characterized by a given distribution. We also provide bounds on the percent increase in the cost-rate caused by (i) the possibility of unpunctual PM versus certain, punctual PM and (ii) ignoring the possibility of unpunctual PM when it is, in fact, possible.

Lastly, we note that an alternative to the anticipatory planning approach examined here might be the use of incentives, which has proven to be a popular tool in behavioral interventions. A review article by Bucklin and Dickinson [18] summarizes studies that examine the relationship between monetary incentives and employee performance. For example, Hermann et al. [38] study the effects of incentives on improving workers' punctuality in a manufacturing company, and conclude that a small daily bonus is effective in changing workers' chronic tardiness. However, opponents of incentive-based tools worry that extrinsic incentives may "crowd out" intrinsic motivations that are important to producing the desired behavior [34], and hence once the incentives are removed, the eroded intrinsic motivation may result in even poorer performance. We leave such considerations for future work.

The remainder of the chapter is organized as follows. In Section 3.2, we formally state the problem and present a general mathematical framework that can be used in determining any type of anticipatory cost-rate-minimizing maintenance policy. Section 3.3 focuses on age replacement with minimal repair, including the special case of Weibull time to failure; Section 3.4 focuses on classic age replacement without minimal repair. Section 3.5 compares the performance of these two maintenance strategies in the presence of unpunctual PM. In Section 3.6, we summarize our findings.

3.2 MODEL FORMULATION

Consider a failure-prone system, for which the time to failure is denoted by the continuous random variable X, with known c.d.f. $F_X(x)$, p.d.f. $f_X(x)$ and mean μ_X. Failures are assumed to be self-announcing and require immediate corrective maintenance (e.g., reactive replacement or minimal repair). Let $h_X(t)$ be the corresponding hazard rate function, i.e., $h_X(t) = f_X(t)/\bar{F}_X(t)$, where $\bar{F}_X(t)$ is the survival function. We impose the following assumptions on $h_X(t)$:

A1: $h_X(0) = 0$;
A2: $h_X(t)$ is strictly increasing to $+\infty$.

Assumption **A1** implies that there are no instantaneous failures at the time of renewal. Assumption **A2** assumes a strictly increasing hazard rate function $h_X(t)$. Both of these assumptions are commonly used in the maintenance and reliability literature [19, 60].

Consider a maintenance policy $\pi(\theta)$ that determines when to preventively replace the system based on a vector of parameters θ (e.g., time, age, usage, or deterioration level). However, possibly unpunctual behavior of the maintenance worker leads to unpunctual PM actions. Let the continuous random variable Y with known c.d.f. $F_Y(y)$ and p.d.f. $f_Y(y)$ be the deviation between the actual time of implementation and that prescribed by $\pi(\theta)$.

The overall objective of the maintenance planner is to minimize the long-run cost-rate by identifying an optimal policy $\pi^*(\theta)$ that anticipates the unpunctual PM implementation. Because we assume that PM outcomes are perfect, i.e., PM returns the system to as-good-as-new, we take a renewal-reward approach and formulate the long-run average cost-rate as the ratio of the expected renewal cycle cost to the expected renewal cycle length. More specifically, let $C^\pi(\theta)$ (respectively, $L^\pi(\theta)$) be the cycle cost (respectively, cycle length) associated with policy $\pi(\theta)$ and

$$\Omega^\pi(\theta) = \frac{\mathbb{E}_{X,Y}[C^\pi(\theta)]}{\mathbb{E}_{X,Y}[L^\pi(\theta)]}$$

be the corresponding long-run cost rate. The main decision-making problem for the maintenance planner is then

$$\min_\theta \ \Omega^\pi(\theta). \tag{3.1}$$

In this chapter, we focus on the case in which θ is a scalar, T, corresponding to an age threshold. Thus, the maintenance policy $\pi(T)$ is an age replacement policy. We drop dependence on T for notational convenience, and in a slight abuse of notation, let $\pi \in \{\mathcal{A}, \mathcal{B}\}$, where "$\mathcal{A}$" denotes age replacement policy with minimal repair (Section 3.3) and "$\mathcal{B}$" denotes an age replacement policy without minimal repair (Section 3.4).

Correspondingly, the actual PM implementation time is at age $T + Y$. If $Y < 0$, then PM is performed earlier than scheduled, and vice versa if $Y > 0$. We impose the following assumptions on Y:

A3: Y has support $[a, b]$, where $-\infty < a \leq b < \infty$;
A4: Y is independent of X and T.

Assumption **A3** implies that scheduled PM is never delayed indefinitely, which is reasonable for most practical settings. Assumption **A4** states that the unpunctual behavior of the maintenance worker is not affected by either the time to failure distribution or the scheduled PM time. The independence of Y and X is intuitive, as the failure time of the system in the absence of any interventions depends only on its characteristics. To justify the assumption that Y and T are independent, we note that in practice for sufficiently large values of T (e.g., one year) the maintenance worker may not be tasked with performing the maintenance until nearer the scheduled PM time (e.g, several weeks or a month ahead of time). Consequently, the dependence of Y on T can be ignored.

Let μ_Y and σ_Y^2 be the mean and variance of Y, respectively. Because Y can assume negative values, our formulation requires the feasible set of T to be $\{T \mid T \ > \max\{-a, 0\}\}$. Let $\widetilde{C}^\pi(T)$ (respectively, $\widetilde{L}^\pi(T)$) be the cycle cost (respectively, cycle length) under punctual implementation (i.e., $Y \equiv 0$) and $\widetilde{\Omega}^\pi(T)$ be the corresponding long-run cost-rate. Then optimization problem (3.1) becomes the well studied classic model (see [16]):

$$\min_{T>0} \widetilde{\Omega}^\pi(T) = \frac{\mathbb{E}_X[\widetilde{C}^\pi(T)]}{\mathbb{E}_X[\widetilde{L}^\pi(T)]}. \tag{3.2}$$

Let T^* and $\widetilde{T}^*$ be the optimal solutions to the optimization problems given by equations (3.1) and (3.2), respectively. In [16], an intuitive fact is established that when the timing of PM is uncertain, the long-run cost-rate is greater than it would be under deterministically timed PM. Theorem 1 restates this result in the context of our problem.

Theorem 1. *If under policy $\pi \in \{\mathcal{A}, \mathcal{B}\}$ both T^* and $\widetilde{T}^*$ exist and are unique, then*

$$\Omega^\pi(T^*) \geq \widetilde{\Omega}^\pi(\widetilde{T}^*). \tag{3.3}$$

Next, we analyze the impact of unpunctual maintenance under policies $\mathcal{A}$ and $\mathcal{B}$ in Sections 3.3 and 3.4, respectively. The theoretical results in these sections can be divided into two broad categories. First (Propositions 9, 10, 11 and 12 as well as Theorems 2 and 3), we establish results on the relative magnitudes of T^* and $\widetilde{T}^*$, which can help maintenance planners understand how the optimal PM schedule is influenced by maintenance workers' unpunctual behavior. Second (Theorems 1, 4, 5, 6 and 7), we examine how unpunctual implementation affects the long-run average cost rate. These results include bounds on the percent increase in the long-run average cost-rate caused by unpunctual PM, which can be particularly useful when fully characterizing the distribution of Y is challenging.

3.3 AGE REPLACEMENT WITH MINIMAL REPAIR

In this section, we consider a maintenance planner who prescribes an age replacement policy with minimal repair, i.e., $\pi = \mathcal{A}$. That is, perfect PM is scheduled to be performed after the system has been operating for a total of T units of time regardless of any failures; failures that occur before age T are minimally repaired (see Figure 3.1). (As mentioned in Section 3.2, if the maintenance worker is punctual, then the PM actions under policy $\pi = \mathcal{A}$

would be performed periodically; such a policy is referred to in the maintenance literature as periodic replacement with minimal repair. However, because in our setting the maintenance worker is unpunctual, PM actions are not necessarily periodic, hence we label this section age replacement with minimal repair so as not to abuse the term “periodic.”)

Let c_p and c_m denote the PM cost and the minimal repair cost, respectively.

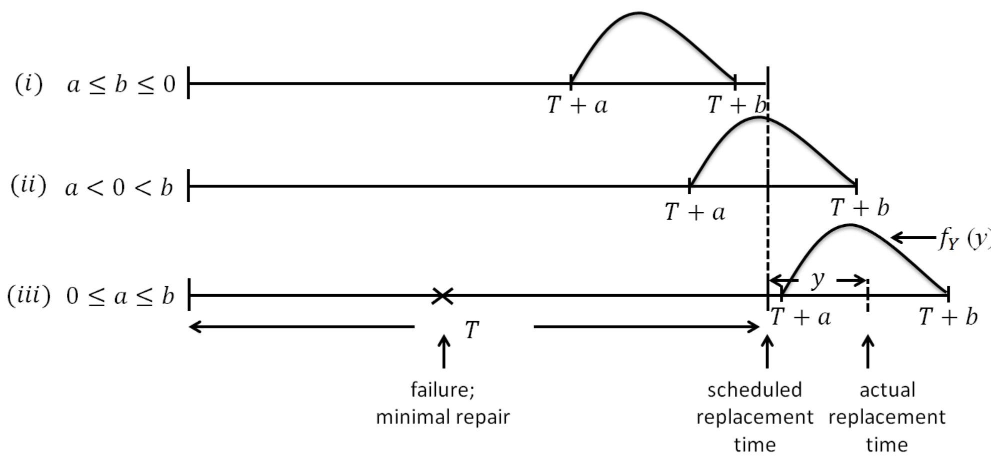

Figure 3.1: Possible cycle dynamics under age replacement policy with minimal repair for different ranges of Y: (i) $a \leq b \leq 0$, i.e., the unpunctual PM actions are never performed later than scheduled; (ii) $a < 0 < b$, the unpunctual PM actions may be performed either earlier or later than scheduled; (iii) $0 \leq a \leq b$, the unpunctual PM actions are never performed earlier than scheduled.

When PM is always performed on time,

$$\mathbb{E}_X[\widetilde{L}^{\mathcal{A}}(T)] = T, \quad \text{and} \quad \mathbb{E}_X[\widetilde{C}^{\mathcal{A}}(T)] = c_m \int_0^T h_X(x)dx + c_p,$$

and problem (3.2) reduces to finding $\widetilde{T}^*$ that is optimal for

$$\min_{T>0} \widetilde{\Omega}^{\mathcal{A}}(T) = \frac{c_m \int_0^T h_X(x)dx + c_p}{T}, \tag{3.4}$$

where $\int_0^T h_X(x)dx$ represents the total expected number of failures (equivalently, minimal repairs) during a renewal cycle [16]. Note that $c_m \int_0^T h_X(x)dx/T$ and c_p/T represent the long-run minimal repair cost-rate and long-run PM cost-rate, respectively. On the other hand, if the timing of PM is unpunctual, then renewal occurs every $T+Y$ units of time and the expected cycle length and the corresponding expected cycle cost are given by

$$\mathbb{E}_{X,Y}[L^{\mathcal{A}}(T)] = T + \mu_Y \quad \text{and} \quad \mathbb{E}_{X,Y}[C^{\mathcal{A}}(T)] = \int_a^b \left(c_m \int_0^{T+y} h_X(x)dx + c_p \right) dF_Y(y),$$

respectively. Therefore, problem (3.1) corresponds to finding T^* that is optimal for

$$\min_{T>0} \Omega^{\mathcal{A}}(T) = \frac{\int_a^b \left(c_m \int_0^{T+y} h_X(x)dx + c_p \right) dF_Y(y)}{T + \mu_Y}, \tag{3.5}$$

where $\int_a^b \left(c_m \int_0^{T+y} h_X(x)dx \right) dF_Y(y)/(T+\mu_Y)$ and $c_p/(T+\mu_Y)$ represent the long-run minimal repair cost-rate and long-run PM cost-rate, respectively.

In Section 3.3.1, we establish analytical conditions on the distribution of Y that characterize the relationship between T^* and $\widetilde{T}^*$ for general time to failure distributions. We also provide bounds on the percent increase in the cost-rate caused by (i) the possibility of unpunctual PM versus certain, punctual PM and (ii) ignoring the possibility of unpunctual PM when it is, in fact, possible.

3.3.1 General Results

By setting the first derivative of the objective function in (3.4) equal to zero and letting $k_1 = c_p/c_m > 1$, i.e., letting minimal repair be less expensive than PM, the optimal solution $\widetilde{T}^*$ to (3.4) satisfies

$$h_X(\widetilde{T}^*)\widetilde{T}^* - \int_0^{\widetilde{T}^*} h_X(x)dx = k_1. \tag{3.6}$$

Similarly, for problem (3.5), the optimal solution T^* satisfies

$$(T^* + \mu_Y) \int_a^b h_X(T^*+y) f_Y(y)dy - \int_a^b \int_0^{T^*+y} h_X(x) f_Y(y) dx dy = k_1. \tag{3.7}$$

The uniqueness of $\widetilde{T}^*$ follows directly from **A1** and **A2**. Furthermore, $\widetilde{\Omega}^{\mathcal{A}}(\widetilde{T}^*) = c_m h_X(\widetilde{T}^*)$ [16]. Analogously, Proposition 8 establishes a sufficient condition for the existence of a unique optimal solution T^* to problem (3.5), i.e., problem (3.1) for $\pi \in \mathcal{A}$. To facilitate the statement of Proposition 8 and the results that follow, we define the functions

$$\widetilde{m}(T) = h_X(T)T - \int_0^T h_X(x)dx \text{ and} \tag{3.8}$$

$$m(T) = (T + \mu_Y) \int_a^b h_X(T+y) f_Y(y)dy - \int_a^b \int_0^{T+y} h_X(x) f_Y(y)dxdy, \tag{3.9}$$

which can be interpreted as follows. For (3.8), consider two consecutive cycles over time intervals $[0, T]$ and $[T, 2T]$. The subtrahend in (3.8) represents the expected number of failures for a new system over T units of time. If the system is not replaced at time T, then the expected minimal repair cost for the system during $[T, 2T]$ is at least $c_m h_X(T)T$ (the minuend in (3.8) times c_m) by **A2**. However, if the system is preventively replaced at time T, then the corresponding cost (including replacement and minimal repair) during $[T, 2T]$ is $c_p + c_m \int_0^T h_X(x)dx$. When the costs of these two possible scenarios are equal during $[0, 2T]$, i.e., $\widetilde{m}(\widetilde{T}^*) = k_1$ and equation (3.6) holds, the objective function (3.4) achieves its global minimum. Equation (3.9) can be interpreted similarly for the unpunctual implementation case.

Proposition 8. *If*

$$\lim_{T \to + \max\{-a, 0\}} m(T) < k_1, \tag{3.10}$$

then $\Omega^{\mathcal{A}}(T)$ is quasi-convex and there exists a unique solution T^* to (3.7), and

$$\Omega^{\mathcal{A}}(T^*) = c_m \int_a^b h_X(T^* + y) f_Y(y)dy.$$

Otherwise,

$$\inf \Omega^{\mathcal{A}}(T) = \lim_{T \to + \max\{-a, 0\}} \Omega^{\mathcal{A}}(T).$$

The key idea behind Proposition 8 is as follows. Based on the interpretation of (3.9), if (3.10) does not hold, then for any choice of T, the expected cost during $[0, 2T]$ for the scenario without replacement at time T is larger than that of the scenario with replacement at time T, i.e., (3.7) can never be achieved. Note that the ratio of c_p and c_m must be relatively large for (3.10) to hold, which implies that the minimal repair cost should be small compared to the PM cost. We assume that condition (3.10) holds throughout the remainder of Section 3.3.

Next, Proposition 9, Theorems 2 and 3 establish how unpunctual policy implementation can affect the optimal solution. More specifically, we consider the relative values of $\widetilde{T}^*$ and T^* under different conditions on the distribution of the time to failure, X, and the deviation from the scheduled PM time, Y. First, Proposition 9 shows that when μ_Y is zero and the rate of increase in the hazard rate decreases over time, the maintenance planner should schedule PM later than he would under a punctual implementation scenario.

Proposition 9. *If $\mu_Y = 0$ and $h_X(t)$ is concave, then $T^* \geq \widetilde{T}^*$.*

Note that a Weibull distribution with shape parameter between 1 and 2 has a concave hazard function that satisfies both **A1** and **A2**. For shape parameter values greater than or equal to 2, however, $h_X(t)$ is convex. Theorems 2 and 3 address the general case of convex hazard, which applies to the majority of commonly used time to failure distributions.

Theorem 2. *If $0 \leq a < b$, $h_X(t)$ is convex and*

$$\lim_{T \to +0} m(T) \geq 0, \tag{3.11}$$

then $0 < T^* < \widetilde{T}^*$.

Corollary 2. *If $0 \leq a < b$, $h_X(t)$ is convex, and $\mu_Y \geq \frac{b}{2}$, then $T^* < \widetilde{T}^*$.*

Both Theorem 2 and Corollary 2 imply that if the maintenance worker never performs PM earlier than intended, then the maintenance planner should schedule PM earlier than he would under a punctual implementation scenario; surprisingly, however, both results depend on the distribution of the delay time, Y (Corollary 2 requiring a stronger condition than that in Theorem 2). More specifically, they require that the mean delay time μ_Y be relatively large. The following numerical example illustrates that if this condition is violated, then it may in fact be optimal to shift the PM time later (i.e., $T^* > \widetilde{T}^*$) even when the worker is never early (i.e., $Y \geq 0$).

Example 1. Consider the problem instance given in Table 3.1. Assume X follows a Weibull distribution with shape and scale parameter α and β, respectively, and let Y follow a Gamma(κ, ϕ) distribution, truncated on the range $[a, b]$ with corresponding shape and scale parameters κ and ϕ, respectively. Note that condition (3.11) does not hold for this example, as $\lim_{T \to +0} m(T) = -18.45$. Figure 3.2 depicts the functions $\widetilde{\Omega}^{\mathcal{A}}(T)$ and $\Omega^{\mathcal{A}}(T)$ for which the corresponding optimal solutions are $\widetilde{T}^* = 1.26$, $\widetilde{\Omega}^{\mathcal{A}}(\widetilde{T}^*) = 4.76$, and $T^* = 1.41$, $\Omega^{\mathcal{A}}(T^*) = 26.82$, see Figure 3.2(a).

Table 3.1: Parameter values for the counter-intuitive Example 1.

α	β	μ_X	c_m	c_p	κ	ϕ	a	b	μ_Y	σ_Y
3	1	0.89	1	4	0.1	100	0	10	0.87	1.93

To explain this counterintuitive behavior, consider the fact that $Y \sim$ Gamma$(0.1, 100)$, truncated on the range $[0, 10]$. The variance of Y, $\sigma_Y^2 = 3.72$, is much larger than the mean μ_Y, indicating that the delay has considerable variation. The large variation of Y results in the majority of the cost-rate being attributable to minimal repairs. In particular, under the optimal solution $T^* = 1.41$, the long-run minimal repair cost-rate is 25.07 compared to the long-run PM cost-rate of 1.75; please see Figure 3.2(b), where the cost-rate functions in (a) are further decomposed into their PM and minimal repair components. Furthermore, the minimal repair cost-rate function for the unpunctual case decreases first ($\lim_{T \to +0} m(T) < 0$), and achieves its minimum at $1.28 > \widetilde{T}^*$; please see the red dashed line and red arrow in Figure 3.2(b). Therefore, it is optimal to prescribe PM at an age greater than $\widetilde{T}^*$.

Next, Theorem 3 shows that in the opposite case (i.e., when the worker is never late), the relative values of the age replacement times are intuitively ordered, i.e., the adjusted PM age is greater than the non-adjusted PM age, regardless of the distribution of Y. Unlike the counter-intuitive result in Example 1, when the maintenance worker never implements PM later than intended, the minimal repairs do not dominate the objective function regardless of the distribution of Y; thus, T^* is never smaller than $\widetilde{T}^*$.

Theorem 3. *If $a < b \leq 0$, $h_X(t)$ is convex, and (i) $\widetilde{m}(-a) \geq k_1$, then $\widetilde{T}^* \leq -a < T^*$; (ii) $\widetilde{m}(-a) < k_1$, then $-a < \widetilde{T}^* < T^*$.*

Theorems 4 and 5 examine how the unpunctual implementation of PM affects the long-run cost-rate. More specifically, these theorems establish bounds on the percent increase in long-run cost-rate under unpunctual implementation.

Theorem 4. *If $\widetilde{T}^*$ is a unique solution to (3.6), and $\widetilde{T}^* - \mu_Y$ is feasible to problem (3.5), then*

$$1 \leq \frac{\Omega^{\mathcal{A}}(T^*)}{\widetilde{\Omega}^{\mathcal{A}}(\widetilde{T}^*)} \leq U_Y^{\mathcal{A}1}(\widetilde{T}^*) \leq U^{\mathcal{A}1}(\widetilde{T}^*). \tag{3.12}$$

where

$$U_Y^{\mathcal{A}1}(\widetilde{T}^*) = \frac{c_m \int_a^b \int_0^{\widetilde{T}^*-\mu_Y+y} h_X(x) f_Y(y) dx dy + c_p}{c_m h_X(\widetilde{T}^*)\widetilde{T}^*}, \text{ and}$$

$$U^{\mathcal{A}1}(\widetilde{T}^*) = \frac{c_m M(\widetilde{T}^* - \mu_Y) + c_p}{c_m h_X(\widetilde{T}^*)\widetilde{T}^*}, \text{ where}$$

$$M(T) = \frac{\int_{T+a}^{T+b} h_X(x)dx}{b-a}(\mu_Y - a) + \int_0^{T+a} h_X(x)dx.$$

Theorem 4 provides upper bounds on the ratio of the optimal long-run cost-rates for problems (3.4) and (3.5). This ratio measures the percent increase in the cost-rate caused by the possibility of unpunctual PM versus certain, punctual PM.

We use a subscript Y in $U_Y^{\mathcal{A}1}(\widetilde{T}^*)$ to emphasize the fact that this bound depends on the functional form of the distribution of Y, whereas $U^{\mathcal{A}1}(\widetilde{T}^*)$ depends only on its mean. Thus, $U^{\mathcal{A}1}(\widetilde{T}^*)$ can be computed with minimal knowledge of Y (only a, b and μ_Y are required). Table 3.2 provides a numerical illustration of the bounds' performance. For this particular example, the presence of unpunctual PM with a fully specified distribution of Y could cost the maintenance planner upwards of 12.4% (see column II). With only μ_Y specified, this bound is as large as of 39.1% (see column III).

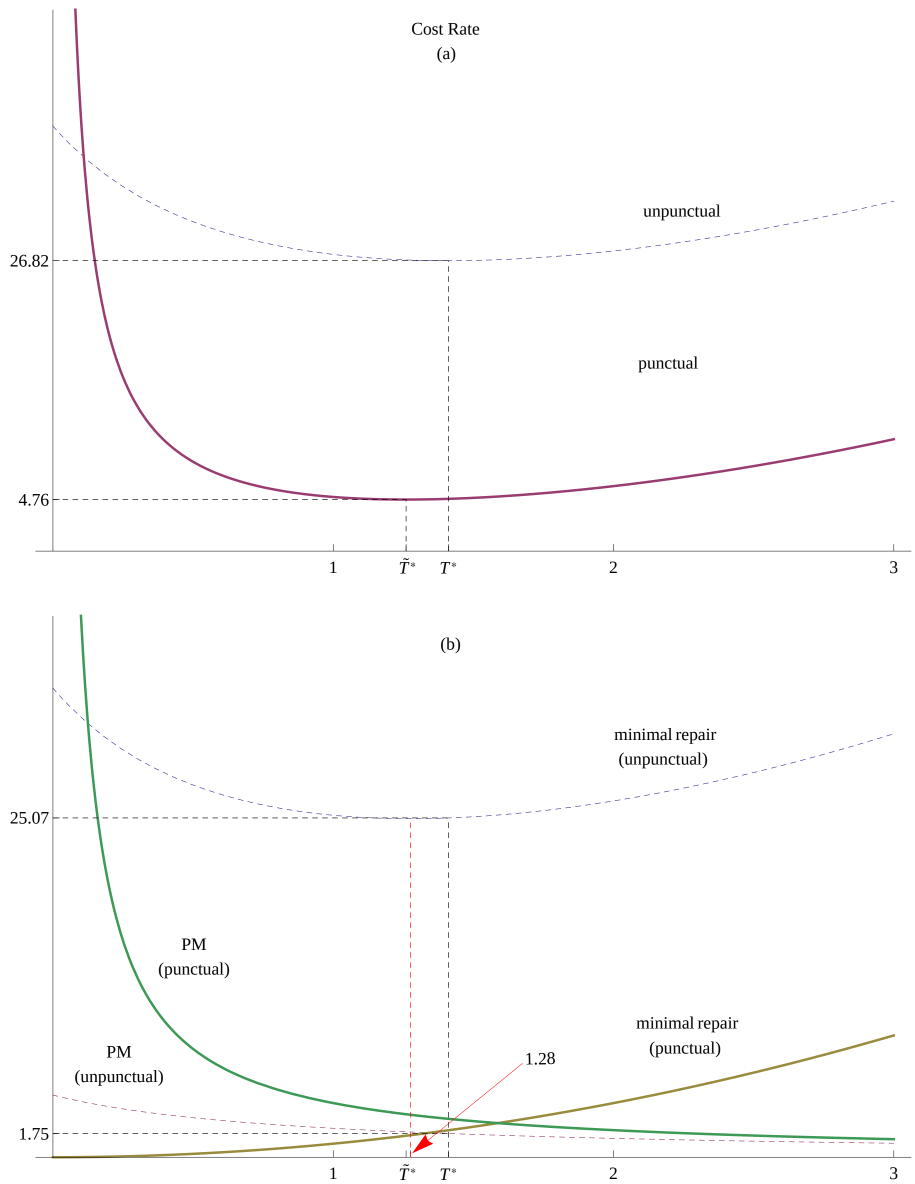

Figure 3.2: (a) Cost-rate functions under punctual and unpunctual PM actions for Example 1, $\widetilde{T}^* = 1.26$, $\widetilde{\Omega}^{\mathcal{A}}(\widetilde{T}^*) = 4.76$, and $T^* = 1.41$, $\Omega^{\mathcal{A}}(T^*) = 26.82$; (b) cost-rate functions in (a) are further decomposed into PM and minimal repair components (under the optimal solution $T^* = 1.41$, the long-run minimal repair cost-rate is 25.07 compared to the long-run PM cost-rate of 1.75), the minimal repair cost-rate function for the unpunctual case achieves its minimum at $1.28 > \widetilde{T}^* = 1.26$. Please see the corresponding discussion in Example 1.

Theorem 5 provides both lower and upper bounds for the ratio of the long-run cost-rates obtained by scheduling PM at $\widetilde{T}^*$ and T^* under unpunctual implementation. These bounds assess the loss associated with ignoring the possibility of unpunctual PM.

Theorem 5. *If $\widetilde{T}^*$ is a unique solution to (3.6), and $\widetilde{T}^* - \mu_Y$ is feasible to problem (3.5), then*

$$L_Y^{\mathcal{A}2}(\widetilde{T}^*) \leq \frac{\Omega^{\mathcal{A}}(\widetilde{T}^*)}{\Omega^{\mathcal{A}}(T^*)} \leq U_Y^{\mathcal{A}2}(\widetilde{T}^*) \leq U^{\mathcal{A}2}(\widetilde{T}^*), \tag{3.13}$$

where

$$\begin{aligned}
L_Y^{\mathcal{A}2}(\widetilde{T}^*) &= \max\left\{ \frac{c_m \int_a^b \int_0^{\widetilde{T}^*+y} h_X(x) f_Y(y) dx dy + c_p}{c_m \int_a^b \int_0^{\widetilde{T}^*-\mu_Y+y} h_X(x) f_Y(y) dx dy + c_p} \cdot \frac{\widetilde{T}^*}{\widetilde{T}^* + \mu_Y}, 1 \right\}, \text{ and} \\
U_Y^{\mathcal{A}2}(\widetilde{T}^*) &= \frac{c_m \int_a^b \int_0^{\widetilde{T}^*+y} h_X(x) f_Y(y) dx dy + c_p}{c_m h_X(\widetilde{T}^*)(\widetilde{T}^* + \mu_Y)}, \text{ and} \\
U^{\mathcal{A}2}(\widetilde{T}^*) &= \frac{c_m M(\widetilde{T}^*) + c_p}{c_m h_X(\widetilde{T}^*)(\widetilde{T}^* + \mu_Y)}.
\end{aligned}$$

As in Theorem 4, the subscript Y in $U_Y^{\mathcal{A}2}(\widetilde{T}^*)$ and $L_Y^{\mathcal{A}2}(\widetilde{T}^*)$ denotes these values' dependency on the full distribution of Y. Table 3.2 also provides a numerical illustration of the bounds' performance. It is not surprising that the upper bounds in Theorem 5 are greater than those in Theorem 4 given that the former values compare suboptimal behavior to optimal behavior whereas the latter considers optimal behavior is both cases. However, this ordering may not always hold. The value of $U^{\mathcal{A}2}(\widetilde{T}^*)$ can be useful in assessing how important it is to fully characterize the function f_Y (e.g., via analysis of maintenance log data) and solve for T^*; if this bound is "close enough" to one, then it may not be worth the effort.

If the maintenance planner acknowledges the possibility of unpunctual maintenance, then three options exist: (i) ignore the maintenance worker's unpunctual behavior and prescribe PM at $\widetilde{T}^*$ (recall that $\widetilde{T}^*$ is an optimal solution when the implementation is punctual), which is obviously suboptimal in the unpunctual case; (ii) solve optimization problem (3.5), which involves full characterization of the distribution of Y; (iii) adopt a heuristic solution, e.g., $\widetilde{T}^* - \mu_Y$, which requires only knowledge of μ_Y.

For the instances included in Table 3.2, $\widetilde{T}^* - \mu_Y$ provides a reasonable estimate of T^* (see the reported ratio $\Omega^{\mathcal{A}}(\widetilde{T}^* - \mu_Y)/\Omega^{\mathcal{A}}(T^*)$ in column VIII). However, as we illustrate in the following example, this heuristic may not always perform well, especially if the mean μ_Y is small and the variance σ_Y^2 is sufficiently large.

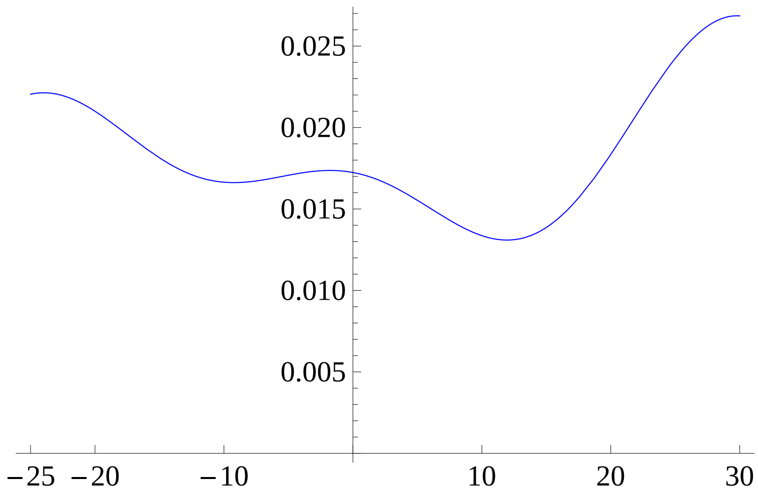

Figure 3.3: The probability density function of Y in Example 2.

Example 2. Figure 3.3 depicts the p.d.f. of Y, which is a mixture of three normal distributions ($N(-25, 10^2)$, $N(0, 10^2)$ and $N(30, 10^2)$ with weights 4, 3 and 5) truncated on $[-25, 30]$, for which $\mu_Y = 2.86$ and $\sigma_Y = 17.06$. If $X \sim$ Weibull$(4, 12)$ with location parameter 80, $c_m = 1$ and $c_p = 50$, then $T^* = 77.83$ and $\widetilde{T}^* = 94.19$. For this example, using the heuristic solution $\widetilde{T}^* - \mu_Y$ results in $\Omega^{\mathcal{A}}(\widetilde{T}^* - \mu_Y)/\Omega^{\mathcal{A}}(T^*) = 1.21$, i.e., a 21% increase in the long-run cost-rate.

3.3.2 Results for $X \sim$ Weibull(α, β)

In this section, we consider several special cases of the optimization problem given by equation (3.5) for which we can derive closed form solutions to (3.7). In particular, we assume a Weibull time to failure distribution, which is widely used to characterize survival data in reliability engineering due to its simplicity and versatility ([23]). For $X \sim$ Weibull(α, β), $h_X(t) = \frac{\alpha}{\beta^\alpha} t^{\alpha-1}$, and the unique solution to (3.6) (i.e., the optimal solution to problem (3.4)) is given by [16] to be

$$\widetilde{T}^* = \left(\frac{k_1}{\alpha - 1}\right)^{\frac{1}{\alpha}} \beta. \tag{3.14}$$

If condition (3.10) in Proposition 8 holds, then the optimality criterion (3.7) for problem (3.5) reduces to

$$\int_a^b \left(\alpha(T^* + y)^{\alpha-1}(T^* + \mu_Y) - (T^* + y)^\alpha\right) f_Y(y)dy = k_1 \beta^\alpha. \tag{3.15}$$

Furthermore, for the case when the maintenance worker never implements PM earlier than intended, Proposition 10 provides an equivalent condition to (3.11).

Proposition 10. *If $X \sim$ Weibull(α, β), $\alpha \geq 2$, $0 \leq a < b$, and*

$$(\alpha - 1)\frac{\mu_Y \mathbb{E}[Y^{\alpha-1}]}{\sigma_Y \sigma_{Y^{\alpha-1}}} \geq \rho_{Y, Y^{\alpha-1}}, \tag{3.16}$$

where $\rho_{Y,Y^{\alpha-1}}$ is the correlation coefficient of random variables Y and $Y^{\alpha-1}$, then condition (3.11) holds. In particular, if $\alpha = 2$, then (3.16) reduces to

$$\frac{\sigma_Y^2}{\mu_Y^2} \leq 1.$$

Note that $\rho_{Y,Y^{\alpha-1}}$ is bounded above by 1. As a result, condition (3.16) is more likely to hold for large values of μ_Y and small values of σ_Y, i.e., the mean PM delay should be sufficiently large and the probability distribution of the delay should have small variability, which is consistent with the insights generated by Corollary 2 and Example 1. For example, inequality (3.16) holds if Y is uniformly distributed and $a \geq 0$. If $\alpha = 2$, (3.16) requires that the distribution of Y have a small coefficient of variation. This observation also matches the intuition of Example 1, where large variation of Y may result in counter-intuitive solutions.

Next, Proposition 11 establishes a closed form solution T^* when $X \sim$ Weibull$(2, \beta)$.

Table 3.2: Numerical example of the bounds in Theorems 4 and 5 for $X \sim$ Weibull (α, β) and $Y \sim$ Uniform (a, b) with $\mu_Y = \frac{a+b}{2}$, $c_m = 1$, $c_p = 16$.

								I	II	III	IV	V	VI	VII	VIII
α	β	a	b	μ_Y	$\widetilde{T}^*$	T^*	$\widetilde{T}^* - \mu_Y$	$\frac{\Omega^{\mathcal{A}}(T^*)}{\widetilde{\Omega}^{\mathcal{A}}(\widetilde{T}^*)}$	$U_Y^{\mathcal{A}1}(\widetilde{T}^*)$	$U^{\mathcal{A}1}(\widetilde{T}^*)$	$\frac{\Omega^{\mathcal{A}}(\widetilde{T}^*)}{\Omega^{\mathcal{A}}(T^*)}$	$L_Y^{\mathcal{A}2}(\widetilde{T}^*)$	$U_Y^{\mathcal{A}2}(\widetilde{T}^*)$	$U^{\mathcal{A}2}(\widetilde{T}^*)$	$\frac{\Omega^{\mathcal{A}}(\widetilde{T}^* - \mu_Y)}{\Omega^{\mathcal{A}}(T^*)}$
5	10	0	5	2.5	13.20	10.54	10.70	1.02391	1.02419	1.07308	1.08119	1.08089	1.10703	1.17578	1.00028
6	10	0	5	2.5	12.14	9.39	9.64	1.03514	1.03625	1.11054	1.13916	1.13794	1.17919	1.30750	1.00107
5	10	-4	0	-2	13.20	15.09	15.20	1.01531	1.01542	1.04648	1.04174	1.04162	1.05768	1.08015	1.00011
6	10	-4	0	-2	12.14	13.98	14.14	1.02253	1.02299	1.06971	1.05517	1.05470	1.07895	1.10654	1.00045
5	10	-4	5	0.5	13.20	12.20	12.70	1.07735	1.08024	1.24614	1.01091	1.00820	1.08910	1.26697	1.00268
6	10	-4	5	0.5	12.14	10.87	11.64	1.11286	1.12402	1.39118	1.02749	1.01729	1.14345	1.44164	1.01003

Table 3.3: Numerical example of the bounds in Theorems 6 and 7 for $X \sim$ Weibull (α, β) and $Y \sim$ Uniform (a, b) with $\mu_Y = \frac{a+b}{2}$, $c_r = 6$, $c_p = 1$.

								I	II	III	IV	V	VI	VII	VIII	IX
α	β	a	b	μ_Y	$\widetilde{T}^*$	T^*	$\widetilde{T}^* - \mu_Y$	$\frac{\Omega^{\mathcal{B}}(T^*)}{\widetilde{\Omega}^{\mathcal{B}}(\widetilde{T}^*)}$	$U_Y^{\mathcal{B}1}(\widetilde{T}^*)$	$U^{\mathcal{B}1}(\widetilde{T}^*)$	$\hat{U}^{\mathcal{B}1}(\widetilde{T}^*)$	$\frac{\Omega^{\mathcal{B}}(\widetilde{T}^*)}{\Omega^{\mathcal{B}}(T^*)}$	$L_Y^{\mathcal{B}2}(\widetilde{T}^*)$	$U_Y^{\mathcal{B}2}(\widetilde{T}^*)$	$U^{\mathcal{B}2}(\widetilde{T}^*)$	$\hat{U}^{\mathcal{B}2}(\widetilde{T}^*)$
3	20	0	4	2	9.32	7.36	7.32	1.01415	1.01417	1.24299	1.04252	1.03405	1.03403	1.04868	1.30085	1.07318
4	20	0	4	2	10.18	8.15	8.18	1.01824	1.01825	1.24622	1.05481	1.05051	1.05050	1.06967	1.36274	1.10782
3	20	-5	0	-2.5	9.32	11.88	11.82	1.02205	1.02209	1.31810	1.06633	1.08599	1.08595	1.10994	1.35674	1.15925
4	20	-5	0	-2.5	10.18	12.63	12.68	1.02846	1.02849	1.32846	1.08560	1.09096	1.09093	1.12201	1.32000	1.16947
3	20	-2	5	1.5	9.32	7.94	7.82	1.04292	1.04307	1.48633	1.12946	1.01645	1.01630	1.06008	1.52879	1.13787
4	20	-2	5	1.5	10.18	8.59	8.68	1.05556	1.05567	1.52064	1.16748	1.02964	1.02953	1.08685	1.63380	1.20231

Proposition 11. *If $X \sim$ Weibull$(2, \beta)$, then the optimal solution to problem (3.5) is*

$$T^* = \left(k_1\beta^2 + \sigma_Y^2\right)^{\frac{1}{2}} - \mu_Y.$$

Moreover, if $\mu_Y = 0$, then

$$T^* = \left(k_1\beta^2 + \sigma_Y^2\right)^{\frac{1}{2}} > \widetilde{T}^* = k_1^{\frac{1}{2}}\beta.$$

Observe that Proposition 11 does not depend on the distributional form of Y. The intuition behind Proposition 11 is that when $\alpha = 2$, the failure rate $h_X(t)$ increases linearly in t, i.e., more slowly than when $\alpha > 2$. Therefore, if the maintenance planner prescribes PM at an appropriately greater age, then the increase in the expected minimal repair cost-rate is smaller than the decrease in the PM cost-rate, which results in a lower total long-run cost-rate. Moreover, if $\mu_Y = 0$, then the difference between T^* and $\widetilde{T}^*$ depends only on the variance σ_Y^2.

Lastly, Proposition 12 characterizes the relationship between T^* and $\widetilde{T}^*$ when $f_Y(y)$ is symmetric w.r.t. $y = 0$ and α attains an integer value larger than 2.

Proposition 12. *If $X \sim$ Weibull (α, β), $\alpha \in \mathbb{Z}_+$, $\alpha > 2$, and $f_Y(y)$ is symmetric w.r.t. $y = 0$, then*

$$T^* \leq \widetilde{T}^* = \left(\frac{k_1}{\alpha - 1}\right)^{\frac{1}{\alpha}}\beta.$$

Moreover, if $\alpha = 3$, then

$$T^* = \widetilde{T}^* = \left(\frac{k_1}{2}\right)^{\frac{1}{3}}\beta. \tag{3.17}$$

Proposition 12 shows that if the hazard rate increases quickly (i.e., $\alpha > 2$), then a conservative PM schedule is preferred, i.e., it is optimal to prescribe PM at an age that is earlier than that for the punctual implementation case. Interestingly, the result in (3.17) establishes that the optimal solutions for both unpunctual and punctual implementation coincide under a symmetric $f_Y(y)$ and $\alpha = 3$.

When the p.d.f. of Y is not symmetric about $y = 0$, it is difficult to characterize the relationship between T^* and $\widetilde{T}^*$ in general. Example 3 illustrates the type of results that can be established for specific asymmetrical p.d.f.s of Y.

Example 3. If $X \sim$ Weibull$(3, \beta)$, $Y \sim$ Triangular(a, c, b) with lower limit a, upper limit b and mode c and $\mu_Y = 0$, then (3.15) reduces to

$$2(T^*)^3 + 2(a+b+c)(T^*)^2 + \frac{2}{3}(a+b+c)^2 T^* + z = k_1\beta^3,$$

where $z = \frac{(a+b+c)^3}{15} + \frac{(a+b+c)(a^2+b^2+c^2)}{30} - \frac{a^3+b^3+c^3}{30}$. If $c > 0$ and $a+b < 0$, then $T^* < \widetilde{T}^* = (\frac{k_1}{2})^{\frac{1}{3}}\beta$; furthermore, if $c < 0$ and $a+b > 0$, then $T^* > \widetilde{T}^* = (\frac{k_1}{2})^{\frac{1}{3}}\beta$. That is, when the mean deviation $\mu_Y = \frac{a+b+c}{3} = 0$ and the distribution of Y is right-skewed, then it is optimal to schedule PM earlier, i.e., $T^* < \widetilde{T}^*$. In contrast, when the distribution is left-skewed, the opposite holds. (Note that for $a = -b$ and $c = 0$ (i.e., if $f_Y(y)$ is symmetric about $y = 0$), then $T^* = \widetilde{T}^* = (\frac{k_1}{2})^{\frac{1}{3}}\beta$, which is consistent with Proposition 12.)

3.4 AGE REPLACEMENT WITHOUT MINIMAL REPAIR

In this section, we consider a maintenance planner who prescribes an age replacement policy without minimal repair, i.e., $\pi = \mathcal{B}$. That is, perfect PM is scheduled to be performed when the system attains a specified age T, or reactively repaired (also perfectly) at failure, whichever occurs first. (As mentioned in Section 3.2, such policies may be more simply referred to as age replacement policies. Our naming convention is motivated by the need to differentiate this class of policies from those in Section 3.3; please see the discussion at the beginning of Section 3.3.)

Let c_p and c_r denote the PM cost and the reactive repair cost, respectively. If PM is always performed on time, then

$$\mathbb{E}_X[\widetilde{C}^{\mathcal{B}}(T)] = c_r \int_0^T f_X(x)dx + c_p \int_T^\infty f_X(x)dx, \quad \text{and} \quad \mathbb{E}_X[\widetilde{L}^{\mathcal{B}}(T)] = \int_0^T x f_X(x)dx + T \int_T^\infty f_X(x)dx,$$

and problem (3.2) reduces to finding $\widetilde{T}^*$ that is optimal for

$$\min_{T>0} \widetilde{\Omega}^{\mathcal{B}}(T) = \frac{c_r F_X(T) + c_p \bar{F}_X(T)}{\int_0^T \bar{F}_X(x)dx} \tag{3.18}$$

(see [16] and [33]). On the other hand, if the timing of PM is unpunctual, then the expected cycle length and the corresponding expected cycle cost are given by

$$\mathbb{E}_{X,Y}[C^{\mathcal{B}}(T)] = \int_a^b \left(c_r \int_0^{T+y} f_X(x)dx + c_p \int_{T+y}^\infty f_X(x)dx \right) dF_Y(y), \quad \text{and}$$

$$\mathbb{E}_{X,Y}[L^{\mathcal{B}}(T)] = \int_a^b \left(\int_0^{T+y} x f_X(x)dx + (T+y) \int_{T+y}^\infty f_X(x)dx \right) dF_Y(y),$$

respectively, and problem (3.1) corresponds to finding T^* that is optimal for

$$\min_{T>0} \Omega^{\mathcal{B}}(T) = \frac{\int_a^b \left(c_r F_X(T+y) + c_p \bar{F}_X(T+y) \right) dF_Y(y)}{\int_a^b \int_0^{T+y} \bar{F}_X(x)dx dF_Y(y)}. \tag{3.19}$$

Similar to the derivations in Section 3.3, by setting the first derivative of the objective function in (3.18) equal to zero and letting $k_2 = c_r/c_p > 1$, i.e., letting the reactive repair be more expensive than PM, the optimal solution $\widetilde{T}^*$ to (3.18) satisfies

$$h_X(\widetilde{T}^*)G(\widetilde{T}^*) - F_X(\widetilde{T}^*) = 1/(k_2 - 1), \tag{3.20}$$

where $G(T) = \int_0^T \bar{F}_X(x)dx$. Analogously, for problem (3.19), the optimal solution T^* satisfies

$$\mathcal{H}(T^*)\mathcal{G}(T^*) - \mathcal{F}_X(T^*) = 1/(k_2 - 1), \tag{3.21}$$

where

$$\mathcal{G}(T) = \int_a^b \int_0^{T+y} \bar{F}_X(x)dx dF_Y(y), \qquad \mathcal{F}_X(t) = \int_a^b F_X(T+y)dF_Y(y),$$

and

$$\mathcal{H}(T) = \frac{\int_a^b f_X(T+y)dF_Y(y)}{\int_a^b \bar{F}_X(T+y)dF_Y(y)}.$$

The uniqueness of $\widetilde{T}^*$ follows directly from **A2**. Furthermore, $\widetilde{\Omega}^{\mathcal{B}}(\widetilde{T}^*) = c_r(1-1/k_2)h_X(\widetilde{T}^*)$ ([16] and [33]). Proposition 13 establishes conditions on the denominator of $\mathcal{H}$ and the function

$$n(T) \equiv \mathcal{H}(T)\mathcal{G}(T) - \mathcal{F}_X(T),$$

that ensure the existence of a unique optimal solution to problem (3.19).

Proposition 13. *If $1/\int_a^b \bar{F}_X(T+y)dF_Y(y)$ is strictly logarithmically convex, and*

$$\lim_{T\to+\max\{-a,0\}} n(T) < 1/(k_2-1), \tag{3.22}$$

then there exists a unique solution T^* to (3.21), and the minimal long-run cost-rate is

$$\Omega^{\mathcal{B}}(T^*) = c_r(1-1/k_2)\mathcal{H}(T^*).$$

Note that many natural forms of the distributions of X and Y satisfy the first condition in Proposition 13. For example, it is easy to verify that if $X \sim$ Weibull(2,β) and $Y \sim$ Uniform $(0, b)$, where $b > 0$, then $1/\int_a^b \bar{F}_X(T+y)dF_Y(y)$ is strictly logarithmically convex.

In the absence of minimal repair, it is more difficult to characterize how unpunctual policy implementation affects the optimal solution. However, using approaches similar to those in Theorems 4 and 5 we can bound the percent increase in the cost-rate caused by (i) the possibility of unpunctual PM (Theorem 6) and (ii) ignoring the possibility of unpunctual PM when it is, in fact, possible (Theorem 7).

Theorem 6. *If $\widetilde{T}^*$ is a unique solution to (3.20), and $\widetilde{T}^* - \mu_Y$ is a feasible solution to problem (3.19), then*

$$1 \leq \frac{\Omega^{\mathcal{B}}(T^*)}{\widetilde{\Omega}^{\mathcal{B}}(\widetilde{T}^*)} \leq U_Y^{\mathcal{B}1}(\widetilde{T}^*) \leq U^{\mathcal{B}1}(\widetilde{T}^*). \tag{3.23}$$

where

$$U_Y^{\mathcal{B}1}(\widetilde{T}^*) = \frac{\Omega^{\mathcal{B}}(\widetilde{T}^* - \mu_Y)}{c_r(1-1/k_2)h_X(\widetilde{T}^*)}, \text{ and}$$

$$U^{\mathcal{B}1}(\widetilde{T}^*) = \frac{c_r + (c_p - c_r)\bar{F}_X(\widetilde{T}^* - \mu_Y + b)}{c_r(1-1/k_2)h_X(\widetilde{T}^*)N(\widetilde{T}^* - \mu_Y)}, \text{ where}$$

$$N(T) = \frac{\int_{T+a}^{T+b}\bar{F}_X(x)dx}{b-a}(\mu_Y - a) + \int_0^{T+a}\bar{F}_X(x)dx.$$

Moreover, if $X \sim$ *Weibull* (α, β)*, and*

$$\widetilde{T}^* - \mu_Y + b < t^0 = \left(\frac{\alpha - 1}{\alpha}\right)^{\frac{1}{\alpha}} \beta, \tag{3.24}$$

then $U^{\mathcal{B}1}(\widetilde{T}^*)$ *becomes*

$$\hat{U}^{\mathcal{B}1}(\widetilde{T}^*) = \frac{c_r + (c_p - c_r)\hat{N}(\widetilde{T}^* - \mu_Y)}{c_r(1 - 1/k_2)h_X(\widetilde{T}^*)N(\widetilde{T}^* - \mu_Y)}, \text{ where}$$
$$\hat{N}(T) = \frac{\bar{F}_X(T + b) - \bar{F}_X(T + a)}{b - a}(\mu_Y - a) + \bar{F}_X(T + a).$$

Theorem 6 provides upper bounds on the ratio of the optimal long-run cost-rates for problems (3.18) and (3.19). As before, the subscript Y in $U_Y^{\mathcal{B}1}(\widetilde{T}^*)$ is included to emphasize the fact that this bound depends on the functional form of the distribution of Y, whereas $U^{\mathcal{B}1}(\widetilde{T}^*)$ requires minimal knowledge of Y (only a, b and μ_Y are needed). In addition, note that $\hat{U}^{\mathcal{B}1}(\widetilde{T}^*)$ is a tighter bound than $U^{\mathcal{B}1}(\widetilde{T}^*)$ if (3.24) holds.

Table 3.3 provides a numerical illustration of the performance of the bounds established in Theorem 6. For this particular example, the presence of unpunctual PM with a fully specified distribution of Y could cost the maintenance planner upwards of 5.6% (see column II). With only μ_Y specified, the bound $U^{\mathcal{B}1}(\widetilde{T}^*)$ is rather loose (see column III). However, because (3.24) holds for each problem instance in the table, the tighter bound $\hat{U}^{\mathcal{B}1}(\widetilde{T}^*)$ also applies (see column IV).

Theorem 7 provides both lower and upper bounds for the ratio of the long-run cost-rates obtained by scheduling PM at $\widetilde{T}^*$ and T^* under unpunctual implementation. These bounds assess the loss associated with ignoring the possibility of unpunctual PM.

Theorem 7. *If* $\widetilde{T}^*$ *is the unique solution to* (3.20)*, and* $\widetilde{T}^* - \mu_Y$ *is feasible to problem* (3.19)*, then*

$$L_Y^{\mathcal{B}2}(\widetilde{T}^*) \leq \frac{\Omega^{\mathcal{B}}(\widetilde{T}^*)}{\Omega^{\mathcal{B}}(T^*)} \leq U_Y^{\mathcal{B}2}(\widetilde{T}^*) \leq U^{\mathcal{B}2}(\widetilde{T}^*) \tag{3.25}$$

where

$$L_Y^{\mathcal{B}2}(\widetilde{T}^*) = \max\left\{\frac{\Omega^{\mathcal{B}}(\widetilde{T}^*)}{\Omega^{\mathcal{B}}(\widetilde{T}^* - \mu_Y)}, 1\right\}, \text{ and}$$

$$U_Y^{\mathcal{B}2}(\widetilde{T}^*) = \frac{c_r + (c_p - c_r)\int_a^b \bar{F}_X(\widetilde{T}^* + y) f_Y(y)dy}{c_r(1 - 1/k_2)h_X(\widetilde{T}^*)\int_a^b \int_0^{\widetilde{T}^*+y} \bar{F}_X(x) f_Y(y)dxdy}, \text{ and}$$

$$U^{\mathcal{B}2}(\widetilde{T}^*) = \frac{c_r + (c_p - c_r)\bar{F}_X(\widetilde{T}^* + b)}{c_r(1 - 1/k_2)h_X(\widetilde{T}^*)N(\widetilde{T}^*)}.$$

Moreover, if $X \sim$ Weibull (α, β), and

$$\widetilde{T}^* + b < t^0 = \left(\frac{\alpha - 1}{\alpha}\right)^{\frac{1}{\alpha}} \beta, \tag{3.26}$$

then $U^{\mathcal{B}2}(\widetilde{T}^*)$ becomes

$$\hat{U}^{\mathcal{B}2}(\widetilde{T}^*) = \frac{c_r + (c_p - c_r)\hat{N}(\widetilde{T}^*)}{c_r(1 - 1/k_2)h_X(\widetilde{T}^*)N(\widetilde{T}^*)}.$$

As in Theorem 6, the subscript Y in $U_Y^{\mathcal{B}2}(\widetilde{T}^*)$ and $L_Y^{\mathcal{B}2}(\widetilde{T}^*)$ denotes dependency on the full distribution of Y and $\hat{U}^{\mathcal{B}2}(\widetilde{T}^*)$ provides a tighter bound than $U^{\mathcal{B}2}(\widetilde{T}^*)$ if (3.26) holds. Table 3.3 also provides a numerical illustration of the performance of the bounds established in Theorem 7. It is not surprising that the upper bounds in Theorem 7 are greater than those in Theorem 6, given that the former values compare suboptimal behavior to optimal behavior whereas the latter considers optimal behavior in both cases; however, this ordering may not always hold.

Comparing columns $\widetilde{T}^* - \mu_Y$ and T^* in Table 3.3, the heuristic solution $\widetilde{T}^* - \mu_Y$ appears to provide reasonably good estimate of T^*. However, similar to Example 2 for $\pi = \mathcal{A}$, we can generate examples under $\pi = \mathcal{B}$ such that the heuristic solution performs poorly.

3.5 IMPLICATIONS FOR POLICY COMPARISONS

For a system that is preventively maintained based on its age, Sections 3.3 and 3.4 establish properties of optimal policies that anticipate the possibility of unpunctual PM implementation. The difference between the maintenance policies in these two sections is in the type of repair performed at failure. For less complex (e.g., single-unit) systems, repair at failure can be regarded as an overhaul, thus an age replacement policy without minimal repair (Section 3.4) is often adopted. For complex systems, such as computers and televisions, after repairing a failed component (e.g., a single tube), the system is as likely to breakdown as it was before repair because the other components are also deteriorating ([15]); hence, an age replacement policy with minimal repair (Section 3.3) is usually adopted. In this section, we compare these two types of maintenance polices, and explore how the optimal maintenance planning under each policy responds to the possibility of unpunctual PM implementation.

For age replacement without minimal repair, i.e., $\pi = \mathcal{B}$, the complexity of equation (3.21) makes it difficult to infer the relationship between T^* and $\widetilde{T}^*$ analytically. Therefore, we numerically explore three cases:

Case (i) PM is never implemented late,

Case (ii) PM may be implemented early or late, and

Case (iii) PM is never implemented early.

For each case, we compute the long-run cost-rate function for $\pi = \mathcal{B}$, and compare it to that for $\pi = \mathcal{A}$; see Figure 3.4. The problem instances considered are summarized in Table 3.4; the corresponding optimal solutions are in Table 3.5. We assume $X \sim$ Weibull(α, β), $Y \sim$ Uniform(a, b) and that, although Policies $\mathcal{A}$ and $\mathcal{B}$ have different cost structures, $k_1 = k_2$. Note that the conditions in Propositions 8 and 13 hold for all of these problem instances, and therefore a unique optimal solution exists under each maintenance policy. Similarly, it is straightforward to verify that Theorem 1 holds for all problem instances (see Table 3.5 and Figure 3.4) and that Theorems 2 and 3 are illustrated by Cases (iii) and (i), respectively. More specifically, Cases (i) and (iii) result in $\widetilde{m}(-a) = 0.0048 < k_1 = 6$ and $\lim_{T \to +0} m(T) = 0.0005 > 0$, respectively.

Table 3.4: Parameter values for Cases (i)-(iii) in Figure 3.4.

Case	α	β	μ_X	c_p	c_r	c_m	a	b	μ_Y	σ_Y
(i)	4	50	45.32	1	6	1/6	-10	0	-5	8.33
(ii)	4	50	45.32	1	2	1/2	-20	20	0	11.55
(iii)	4	50	45.32	1	6	1/6	0	10	5	8.33

Table 3.5: Optimal solutions for Cases (i)-(iii) in Figure 3.4.

	$\pi = \mathcal{A}$				$\pi = \mathcal{B}$			
Case	T^*	$\widetilde{T}^*$	$\Omega^{\mathcal{A}}(T^*)$	$\widetilde{\Omega}^{\mathcal{A}}(\widetilde{T}^*)$	T^*	$\widetilde{T}^*$	$\Omega^{\mathcal{B}}(T^*)$	$\widetilde{\Omega}^{\mathcal{B}}(\widetilde{T}^*)$
(i)	64.39	59.46	0.0225	0.0224	30.37	25.45	0.0537	0.0527
(ii)	43.76	45.18	0.0324	0.0295	42.65	38.31	0.0390	0.0360
(iii)	54.39	59.46	0.0225	0.0224	20.37	25.45	0.0537	0.0527

In Cases (i) and (iii) in Figure 3.4 the adjustment of the scheduled PM age is in the same direction under both policies, i.e., the adjusted PM age is greater than the non-adjusted PM age if the maintenance worker never implements PM later than intended, and vice versa if the maintenance worker never implements PM earlier than intended. However, as seen in Case (ii) when PM may be performed early or late, the adjustments of the PM age are in opposing directions under Policies $\mathcal{A}$ and $\mathcal{B}$.

Recalling (3.5) and (3.19), the denominator in the latter objective function requires the fully specified distribution of Y, which makes it more difficult to bound. A natural conjecture might be that the upper bounds for age replacement without minimal repair are not as tight as those for age replacement with minimal repair when the distribution of Y is not fully specified (recall Theorems 4 and 5 in Section 3.3 and Theorems 6 and 7 in Section 3.4). This conjecture is supported by the numerical results in Tables 3.2 and 3.3 (see columns III and

VIII in Table 3.3 vs. columns III and VII in Table 3.2). Hence, these results suggest that the distribution information of Y is more important for age replacement without minimal repair if the maintenance planner is seeking a satisfactory estimate of the long-run cost-rate.

Next, as originally presented in [15], we provide an example comparison of the two policy types by fixing the common cost term $c_p = 1$ and separating the (c_m, c_r) plane into two regions (Figure 3.5). Under punctual PM and $X \sim \text{Weibull}(4, 10)$, an optimal age replacement policy without (with) minimal repair achieves a lower long-run cost-rate if (c_m, c_r) falls below (above) the solid curve in Figure 3.5. (The shaded region corresponds to the cost combinations that satisfy the assumptions that $c_r > c_p$ and $c_m < c_p$.) We add two additional curves to perform the same comparison under unpunctual PM implementation assuming $Y \sim \text{Uniform}(-1, 4)$. The dashed line (labeled $\Omega^{\mathcal{A}}(T^*)$ vs. $\Omega^{\mathcal{B}}(T^*)$) represents the boundary between the two policies under the optimally adjusted PM time. The dash-dot line (labeled $\Omega^{\mathcal{A}}(\widetilde{T}^*)$ vs. $\Omega^{\mathcal{B}}(\widetilde{T}^*)$) delineates the boundary between the two policies when the maintenance planner ignores the possibility of unpunctual implementation; doing so causes the region of cost combinations when minimal repair is preferred to shrink as compared to when scheduling optimally under unpunctual PM implementation. However, because the dashed curves depend on the distributions of both X and Y, the relative order may be reversed in other instances.

For this particular example and all other examples we examined, the age replacement policy without minimal repair is more robust to unpunctual PM (i.e., both the dashed line and the dash-dot line lie above the solid line). The intuition behind this observation is that under such a policy (in contrast to a policy with minimal repair) not every cycle ends with unpunctual PM because renewal after failure may occur before the preventive replacement age. We are, however, unable to prove this result for the general case.

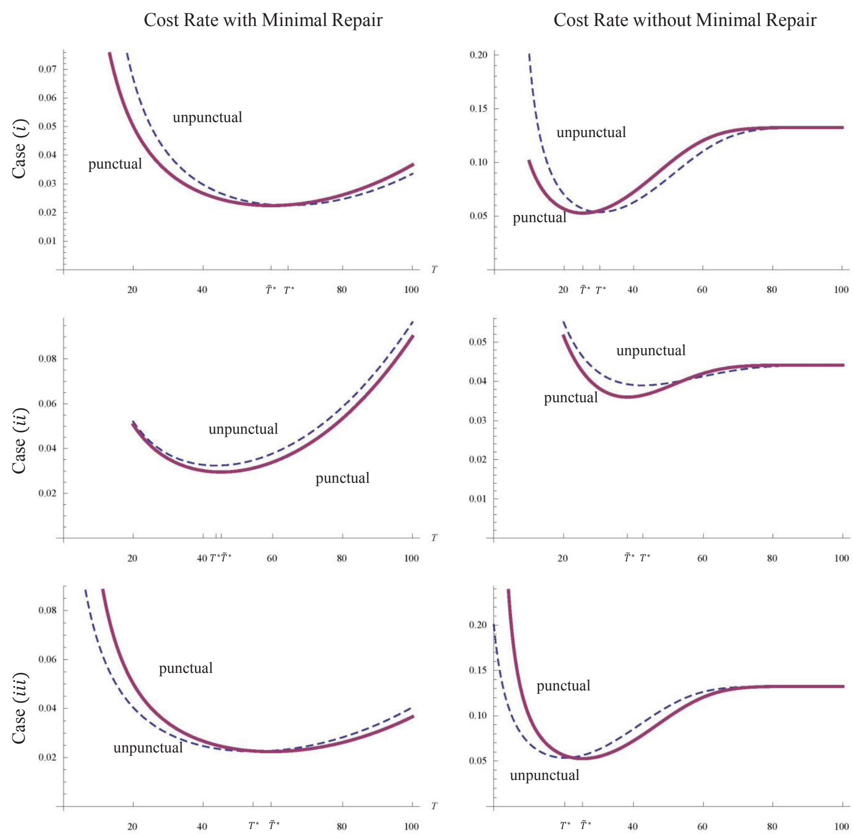

Figure 3.4: Plots of the cost-rate functions for Case (i) (top) to (iii) (bottom) under two Policies $\mathcal{A}$ and $\mathcal{B}$.

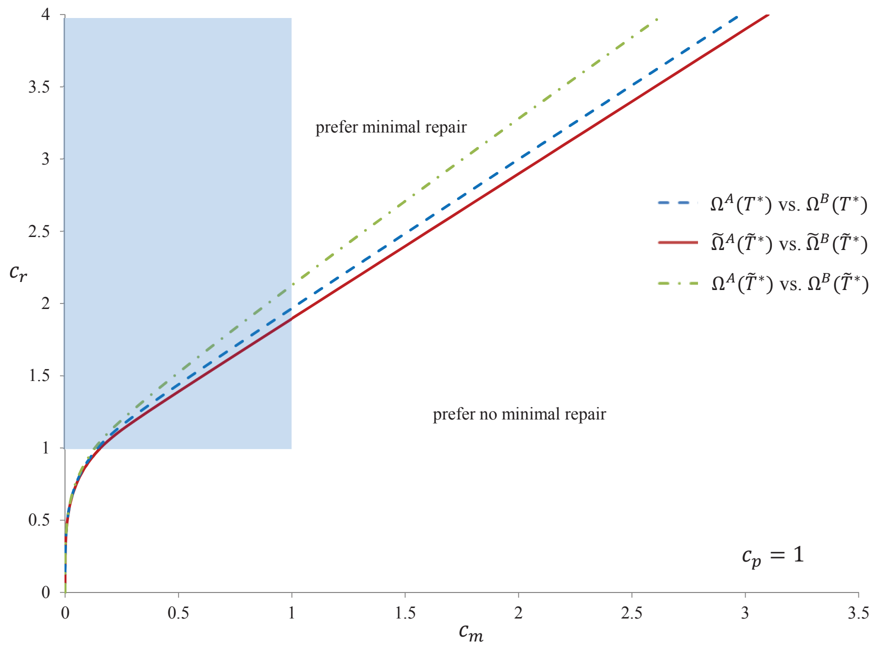

Figure 3.5: Preferences between age replacement with minimal repair and without minimal repair as a function of cost parameter combinations for $X \sim \text{Weibull}(4, 10)$.

3.6 CONCLUSION

In this chapter, we study how the possibly unpunctual behavior of a maintenance worker affects a maintenance planner's decision-making by formulating cost-rate minimization age replacement models. Both analytical and numerical results are provided on how the maintenance planner should adjust the maintenance policy in anticipation of unpunctual PM actions. Furthermore, we compare age replacement policies with and without minimal repair, and explore how the optimal maintenance planning under each policy responds to the possibility of unpunctual PM implementation. We also establish bounds on the percent increase in the long-run cost-rate which are useful in assessing how important it is to fully characterize the probability density function of the unpunctual behavior.

4.0 OPTIMAL SEQUENCING OF TWO MEDICAL TREATMENTS

4.1 INTRODUCTION

This chapter is inspired by current treatment practices for chronic diseases (e.g., rheumatoid arthritis (RA)). For many of these diseases, it is difficult for physicians to navigate the treatment process because the multitude of available treatments from which to choose can be unique in terms of potential effectiveness, length of effectiveness delay (treatment-specific time before revealing its effectiveness), price, etc. [31, 41] For example, a less expensive treatment may take longer to reveal its effectiveness; if the patient finds it is ineffective and must switch treatments, then the disease may have progressed to a more advanced level. In contrast, a more expensive treatment may take less time to reveal its effectiveness and have a greater chance of being effective, but a greater potential for side effects. Therefore, a key issue in managing many chronic diseases is to identify an optimal sequence of treatments that balances the multiple trade-offs inherent to the different treatment options.

In this chapter, we propose a stylized model inspired by this problem and restrict our focus to the special case of two treatments. More specifically, we aim to balance three treatment-specific trade-offs: probability of effectiveness, length of effectiveness delay, and reward/cost. The decision maker is the healthcare provider, whose overall objective is to identify which treatment to prescribe first, such that the total expected QALYs gained by the patient are maximized.

More formally, consider a patient with a chronic disease that has several observable levels of severity. In the absence of an effective treatment, the disease progresses stochastically to more advanced levels in which the quality of life for the patient is lower. Each of the two treatment options that we consider has a known probability of effectiveness, a known delay period before the treatment reveals whether or not it is effective, and a known cost per unit time. The treatment cost includes both price and side effects, measured in QALYs. If a treatment reveals itself to be effective at the end of its delay period, then the patient remains on the treatment indefinitely and receives a terminal reward that depends on patient age and treatment type. If both treatments have been attempted and found to be ineffective, then the patient is placed on palliative care and receives an associated terminal reward.

In Chapter 1 of this dissertation, we draw a parallel between maintaining degrading equipment and degrading human body. Scheduling "maintenance actions" is equivalent to prescribing possible treatments. A proper maintenance action can bring the equipment from a more deteriorated state to a less deteriorated state, which is equivalent in our setting to an effective treatment that moves patient to a healthier state. However, in most traditional maintenance optimization models [22, 29, 32, 43, 58, 77, 87, 90], the supply of spare parts for maintenance/replacement is usually assumed to be infinite. As a result, the maintenance activities can be scheduled periodically or performed whenever necessary. Our model considers a limited number (two) of available treatments, and once a treatment is attempted and found to be ineffective, it cannot be used again.

There exists research work similar to our setting in terms of the limited number of maintenance actions [27, 42, 51, 80]. For example, Icten et al. [42] develop an Markov decision process (MDP) model to adaptively schedule a fixed number of identical replacements of a vital component. Failure of the component is assumed to cause the system's breakdown, and the objective is to maximize the total expected lifetime of the system. However, the replacements considered in [42] are assumed to be identical, which is not the case in our setting. Shechter et al. [80] focus on the optimal sequencing of nonidentical components to maximize the expected system survival time. They use stochastic orderings to compare components' general lifetime distribution, and find a counterexample that the strongest ordering does not guarantee the optimal sequencing. The objective for their model is to extend the system's survival time as much as possible, while we aim to maximize the total expected QALYs a patient can gain from a particular treatment sequence.

Furthermore, there exist therapeutic optimization studies that consider the optimal time to initiate treatment for different diseases (see examples in [26, 53, 57, 79, 82].) Generally speaking, most of these works either consider only a single treatment or a predetermined sequence of treatments. We, in contrast, assume treatments for the patient start as soon as the patient is symptomatic, so the time to initiate therapy is predetermined. Shechter et al. [79] develop an MDP modeling framework with application to HIV therapy switching problem. They assume a finite number of therapies, and consider the trade-offs between the decrease in viral level by taking HIV therapy and the increase in viral resistance over time. The available therapies, however, are assumed to be identical. Thus, there is no balancing between different treatment-specific characteristics as in our problem.

In terms of determining optimal treatment pathways, there is a large body of health economics literature focused on the application of three approaches: decision trees, simulation and Markov models. For example, Aloia and Fahy [7] conduct an analysis of optimal treatment combinations for patients with colorectal cancer and resectable liver metastases. They consider all possible treatment combinations in a simulation model and predict the optimal treatment pathway based on the estimated 5-year survival rate. Another example is Kobelt et al. [49] who perform a cost-effectiveness study of an early biologic treatment for RA by considering both dose reduction and treatment switches. They apply a Markov model with five states and analyze the model with simulation. Generally speaking, such studies focus on careful calibration of model parameters and computations in order to evaluate important health economics decisions. They, however, deliver little analytical results or structural properties of the problem.

The main contribution of our work is to provide theoretical analysis for the optimal two treatment sequencing problem. To the best of our knowledge, there are no analytical studies that consider balancing the trade-offs between three treatment-specific characteristics (probability of effectiveness, length of delay and reward/cost) simultaneously. We first provide theoretical conditions that indicate when, as a function of the model parameters, it is optimal to initiate treatment with one treatment versus the other. Then, we illustrate those results by insightful numerical examples.

The remainder of this chapter is organized as follows. Section 4.2 formulates our discrete time mathematical model. In Section 4.3. we establish our analytical results. Numerical examples are provided in Section 4.4.

4.2 MODEL FORMULATION

Before formally stating our mathematical model, we first introduce the following notation.

t_0	age of the patient when he/she starts treatment;
Δ	set of ordered disease levels, $\Delta = \{0, 1, 2, \dots, D\}$ with '0' denoting the disease-free stage and 'D' the most severe stage; $\delta \in \Delta$ represents the current disease level;
δ_{t_0}	disease level of the patient when he/she starts treatment;
Θ	set of all possible treatments excluding palliative care, $\Theta = \{A, B\}$;
θ^P	palliative care (patient is placed on palliative care indefinitely if both treatments have been exhausted and proven ineffective);
d^θ	the length of delay after which treatment θ reveals whether it is effective or not, $\theta \in \Theta$; we assume d^θ is integer;
ρ^θ	probability that treatment θ reveals itself to be effective after the delay period d^θ, $\theta \in \Theta$;
$P(\delta'\vert\delta)$	probability that patient transitions from disease level δ to level δ' at the beginning of each unit time during the delay period of any treatment;
$q^\theta(\delta)$	net QALYs gained per unit time by the patient during the delay period of treatment θ, given that the disease level is δ;
$Q_E^\theta(t)$	net QALYs gained if the patient continues an effective treatment θ indefinitely from age t;
$Q^{\theta^P}(\delta, t)$	net QALYs gained if the patient continues palliative care θ^P indefinitely from age t with disease level δ.

Remark 1. The definitions of net QALYs in $q^\theta(\delta)$, $Q_E^\theta(t)$ and $Q^{\theta^P}(\delta, t)$ implicitly consider the QALYs loss due to disease level as well as the cost of taking treatment θ. The treatment cost includes both the economic price of θ and the side effects for being on θ.

We introduce the following assumptions, which are essential for our mathematical model:
A1: The effectiveness of a treatment $\theta \in \Theta$ can only be revealed after a treatment-specific time d^θ; during the delay period d^θ, the treatment is assumed to be *ineffective*, and the disease level of the patient transitions at the beginning of each unit time.

A2: Each treatment reveals itself to be effective at the end of its delay period with probability ρ^θ. If it is effective, then the disease level of the patient moves to $\delta = 0$ and the patient continues with treatment θ indefinitely, getting an age-dependent lump sum QALYs that also depend on the treatment type. Furthermore, $Q_E^\theta(t)$ is nonincreasing in patient age t, $\theta \in \{A, B\}$.

A3: $q^\theta(\delta)$ is nonincreasing in disease level $\delta \in \Delta$; furthermore, net QALYs gained per unit time when the patient is on an effective treatment θ indefinitely is not smaller than that during delay period, i.e.,

$$Q_E^\theta(t) - Q_E^\theta(t+1) \geq \max_{\delta, a \in \Theta} q^a(\delta), \text{for any } t \in [t_0, t_0 + d^A + d^B] \text{ and } \theta.$$

A4: Transition matrix P is totally positive of order 2 (TP2) and upper triangular, hence 'D' is an absorbing state. This assumption implies that the j-step transition matrix $P^{(j)}$ ($j \in \mathbb{Z}_+$) is IFR (Increasing Failure Rate).

A5: $Q_E^{\theta^P}(\delta, t)$ is nonincreasing in δ for any t, i.e., if the patient starts the palliative care in a more severe disease level, then the expected QALYs he/she gains are lower. We also assume $Q_E^{\theta^P}(\delta, t)$ is nonincreasing in t for any δ.

The problem dynamics are illustrated in Figure 4.1. If the patient starts treatment θ at disease level δ, then the QALYs gained during the delay period are denoted as

$$r(\theta, \delta) = \sum_{j=1}^{d^\theta} \sum_{\delta'=1}^{D} P^{(j)}(\delta'|\delta) q^\theta(\delta'). \tag{4.1}$$

Given a treatment sequence $\langle AB \rangle$, let $f_1^{\langle AB \rangle}(\delta_{t_0})$ be the total QALYs gained if the first treatment A reveals itself to be effective, i.e.,

$$f_1^{\langle AB \rangle}(\delta_{t_0}) = r(A, \delta_{t_0}) + Q_E^A(t_0 + d^A). \tag{4.2}$$

Similarly, the total QALYs gained if the first treatment A fails but the second treatment B reveals itself to be effective are given by

$$f_2^{\langle AB \rangle}(\delta_{t_0}) = r(A, \delta_{t_0}) + \sum_{\delta=1}^{D} P^{(d^A)}(\delta|\delta_{t_0}) \cdot r(B, \delta) + Q_E^B(t_0 + d^A + d^B), \tag{4.3}$$

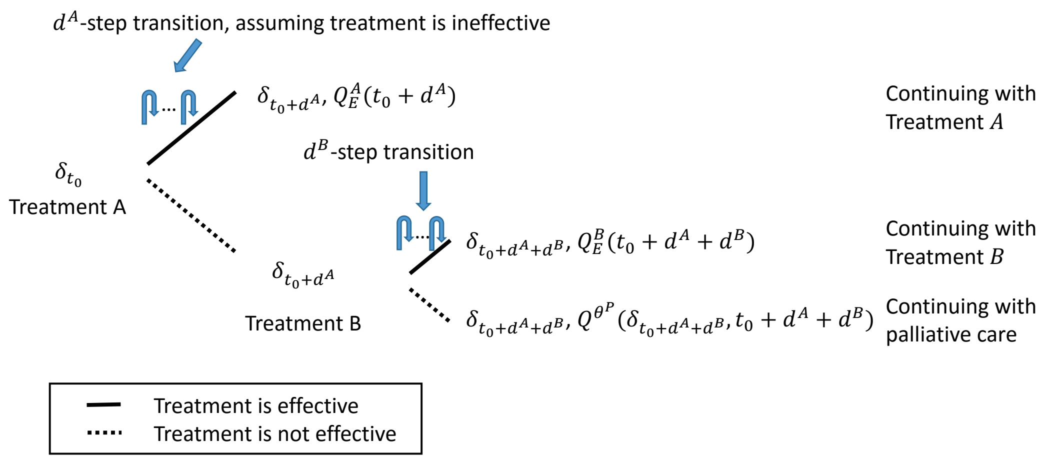

Figure 4.1: Problem dynamics with two treatments available

and the total QALYs gained if both treatments fail are given by

$$f_3^{\langle AB\rangle}(\delta_{t_0}) = r(A,\delta_{t_0}) + \sum_{\delta=1}^{D} P^{(d^A)}(\delta|\delta_{t_0}) \cdot r(B,\delta) + \sum_{\delta=1}^{D} P^{(d^A+d^B)}(\delta|\delta_{t_0}) Q^{\theta^P}(\delta, t_0 + d^A + d^B). \tag{4.4}$$

To identify which treatment to prescribe first in order to maximize the total QALYs gained by the patient, we need to solve the following optimization problem

$$V(\delta_{t_0}) = \max \begin{cases} \rho^A f_1^{\langle AB\rangle}(\delta_{t_0}) + (1-\rho^A)\rho^B f_2^{\langle AB\rangle}(\delta_{t_0}) + (1-\rho^A)(1-\rho^B) f_3^{\langle AB\rangle}(\delta_{t_0}), \\ \rho^B f_1^{\langle BA\rangle}(\delta_{t_0}) + (1-\rho^B)\rho^A f_2^{\langle BA\rangle}(\delta_{t_0}) + (1-\rho^A)(1-\rho^B) f_3^{\langle BA\rangle}(\delta_{t_0}). \end{cases} \tag{4.5}$$

4.3 ANALYTICAL RESULTS

In this section, we establish analytical conditions under which a particular treatment should be prescribed first. Lemma 2 and Theorem 8 provide intuitive results in terms of the monotonicity property of the reward function $r(\theta,\delta)$ and value function $V(\delta)$ over disease level δ.

Lemma 2. *$r(\theta,\delta)$ is nonincreasing in δ for any $\theta \in \Theta$.*

Lemma 2 states that the net QALYs gained during the delay period are a nonincreasing function of disease level δ. The proof of Lemma 2 is straightforward based on Assumption **A3**.

Theorem 8. *$V(\delta)$ is nonincreasing in δ.*

Theorem 8 states that for patients starting with a more severe disease level, the total expected QALYs gained are lower. Next, Lemma 3-5 are technical results necessary for proving Theorem 9-11.

Lemma 3. *If $d^A \le d^B$, $q^A(\delta) \ge q^B(\delta)$ for all $\delta \in \Delta$, and $Q_E^A(t) \ge Q_E^B(t)$ for all t, then*

$$f_1^{\langle AB\rangle}(\delta_{t_0}) \ge f_1^{\langle BA\rangle}(\delta_{t_0}) \ge f_2^{\langle AB\rangle}(\delta_{t_0}). \tag{4.6}$$

Lemma 3 considers the scenario when treatment A works at least as fast and is at most as costly as treatment B. We can order the value functions in three cases, (a) treatment sequence $\langle AB\rangle$ is prescribed and treatment A is effective; (b) sequence $\langle BA\rangle$ is prescribed and treatment B is effective; and (c) treatment sequence $\langle AB\rangle$ is prescribed, treatment A is ineffective while treatment B is effective. Lemma 3 suggests that the QALYs gained in case (a) are the highest, while and the QALYs gained in case (c) are lowest.

Lemma 4. *If $d^A \leq d^B$, $q^A(\delta) \geq q^B(\delta)$ for all $\delta \in \Delta$, and $Q_E^A(t) \geq Q_E^B(t)$ for all t, and $q^A(\delta) - q^B(\delta)$ is nonincreasing in δ, then*

$$r(A, \delta_{t_0}) + \sum_{\delta=1}^{D} P^{(d^A)}(\delta|\delta_{t_0}) \cdot r(B, \delta) \geq r(B, \delta_{t_0}) + \sum_{\delta=1}^{D} P^{(d^B)}(\delta|\delta_{t_0}) \cdot r(A, \delta). \tag{4.7}$$

Lemma 4 considers the case when treatment A works at least as fast and is at most as costly as treatment B, and the cost advantage of A over B is nonincreasing in the disease level. If treatment sequence $\langle AB\rangle$ is prescribed, then the total expected QALYs gained during the delay periods are higher than those gained by prescribing sequence $\langle BA\rangle$ instead.

Remark 2. To justify the requirement that $q^A(\delta) - q^B(\delta)$ is nonincreasing in δ in Lemma 4, consider the following example. Let $q^\theta(\delta)$ be given as

$$q^\theta(\delta) = q - s(\delta) - c^\theta(\delta),$$

i.e., the QALYs gained per unit time depend on the maximal QALYs that can be gained, subtract the QALYs loss $s(\delta)$ due to being in disease level δ, and the cost of taking treatment $c^\theta(\delta)$. If $c^A(\delta) \leq c^B(\delta)$, $\forall\delta$, and $c^B(\delta) - c^A(\delta)$ is nonincreasing in δ (the cost advantage of A over B is nonincreasing in disease level), then $q^A(\delta) - q^B(\delta)$ is nonincreasing in δ. We can construct counterexamples for which if $q^A(\delta) - q^B(\delta)$ is not nonincreasing in δ, then (4.7) does not hold.

Lemma 5. *If $d^A \leq d^B$, $q^A(\delta) \geq q^B(\delta)$ for all $\delta \in \Delta$, and $Q_E^A(t) \geq Q_E^B(t)$ for all t, and $q^A(\delta) - q^B(\delta)$ is nonincreasing in δ, then*

$$f_1^{\langle AB\rangle}(\delta_{t_0}) \geq f_2^{\langle BA\rangle}(\delta_{t_0}). \tag{4.8}$$

Lemma 5 considers the case when treatment A works at least as fast and is at most as costly as treatment B, and the cost advantage of A over B is nonincreasing in the disease level. Then, we compare the following two scenarios, (i) sequence $\langle AB\rangle$ is prescribed, treatment A turns out to be effective; and (ii) sequence $\langle BA\rangle$ is prescribed, treatment B turns out to be ineffective, while A is effective. The total expected QALYs gained by the patient in the latter scenario are no larger than the former one.

Theorem 9. *If $d^A \leq d^B$, $\rho^A \geq \rho^B$, $q^A(\delta) \geq q^B(\delta)$ for all $\delta \in \Delta$, $Q_E^A(t) \geq Q_E^B(t)$, $Q_E^A(t) - Q_E^A(t+k) \geq Q_E^B(t) - Q_E^B(t+k)$ for all t and $k > 0$, and $q^A(\delta) - q^B(\delta)$ is nonincreasing in δ, then it is optimal to prescribe treatment A first.*

Theorem 9 establishes the intuitive fact that if one of the two treatments dominates the other, i.e., it works faster, it has higher probability of effectiveness and it costs less, then it is optimal to prescribe this dominating treatment first under some mild assumptions on the terminal reward $Q_E^\theta(t)$, $\theta \in \{A, B\}$.

However, it is usually rather difficult to decide which treatment to prescribe first, because neither of the treatments may be superior in all three characteristics we examine. Recall the motivating example in Section 4.1, where a less expensive treatment needs more time to reveal its effectiveness. In contrast, a fast-working and more effective treatment can be very costly. Next, Theorems 10 and 11 consider the interesting questions that if treatment A is not dominating, i.e., at least one of its three characteristics are inferior than that of treatment B, under which condition should we prefer prescribing sequence $\langle AB\rangle$.

Theorem 10. *Given $d^A \leq d^B$, $q^A(\delta) \geq q^B(\delta)$ for all $\delta \in \Delta$, and $Q_E^A(t) \geq Q_E^B(t)$, $Q_E^A(t) - Q_E^A(t+k) \geq Q_E^B(t) - Q_E^B(t+k)$ for all t and $k > 0$, and $q^A(\delta) - q^B(\delta)$ is nonincreasing in δ, if*

$$\rho^A \geq \bar{\rho} = \frac{\rho^B \cdot (f_1^{\langle BA\rangle}(\delta_{t_0}) - f_2^{\langle AB\rangle}(\delta_{t_0}))}{f_1^{\langle AB\rangle}(\delta_{t_0}) - \rho^B f_2^{\langle AB\rangle}(\delta_{t_0}) - (1-\rho^B) f_2^{\langle BA\rangle}(\delta_{t_0})}, \tag{4.9}$$

then it is optimal to prescribe treatment A first.

Theorem 10 considers the case when treatment A works at least as fast and is at most as costly as treatment B, and the cost advantage of A over B is nonincreasing in the disease level. The obtained results state that even if A has a lower probability of effectiveness, then it is still possible to have treatment A to be the optimal treatment to prescribe first as long as (4.9) holds, where $\bar{\rho}$ serves as a lower bound for ρ^A.

Theorem 11. *Given $d^A \le d^B$, $\rho^A \ge \rho^B$, if $q^A(\delta) = \mu q^B(\delta)$ for all $\delta \in \Delta$, and $Q_E^A(t) = \mu Q_E^B(t)$ for all t, a necessary and sufficient condition for A to be the first treatment to prescribe is*

$$\mu \ge \mu^*,$$

where

$$\mu^* = \frac{\rho^B Q_E^B(t_0+d^B) - (1-\rho^A)\rho^B Q_E^B(t_1) + r(B,\delta_{t_0}) - (1-\rho^A)\sum_{\delta=1}^{D} P^{(d^A)}(\delta|\delta_{t_0})\cdot r(B,\delta)}{\rho^A Q_E^B(t_0+d^A) + \sum_{j=1}^{d^A}\sum_{\delta'=1}^{D} P^{(j)}(\delta'|\delta_{t_0}) q^B(\delta') - (1-\rho^B)\left(\rho^A Q_E^B(t_1) + \sum_{j=d^B+1}^{d^A+d^B}\sum_{\delta'=1}^{D} P^{(j)}(\delta'|\delta_{t_0}) q^B(\delta')\right)}, \tag{4.10}$$

and $t_1 = t_0 + d^A + d^B$.

Theorem 11 considers the case when treatment A is at least as effective and works at least as fast as treatment B. If A has a higher cost (or equivalently, a lower reward), then it is still possible to have treatment A to be the optimal treatment to be prescribed first as long as (4.10) holds. Note that the assumption $Q_E^A(t) = \mu Q_E^B(t)$ holds if the terminal reward is a linear function of the remaining life time, and the QALYs gained per unit time when on treatment A are μ times that of B.

The remaining measure not examined so far is the length of effectiveness delay. Because it affects the reward gained during delay periods, the terminal reward as well as the distribution of diseases level after the delay period, it is somewhat challenging to establish a closed form result as in Theorems 10 and 11. To address this difficulty, we first show there exists at most one $\bar{d^A}$, such that we prefer sequence $\langle AB \rangle$ if $d^A \le \bar{d^A}$, and prefer sequence $\langle BA \rangle$ otherwise in Lemma 6. We then provide a numerical example in Section 4.4 (see Example 3) to illustrate when our choice of the first treatment changes as a function of treatments' lengths of delay.

Lemma 6. *Given $\rho^A \geq \rho^B$, $q^A(\delta) \geq q^B(\delta)$ for all $\delta \in \Delta$, $Q_E^A(t) \geq Q_E^B(t)$, and $Q_E^A(t) - Q_E^A(t+k) \geq Q_E^B(t) - Q_E^B(t+k)$ for all t and $k > 0$, if*

$$\rho^B \left(Q_E^B(t) - Q_E^B(t+1) \right) \geq \max_{\delta} q^A(\delta), \textit{for all } t \in [t_0, t_0 + d^A + d^B] \textit{ and} \tag{4.11}$$

$$Q_E^A(t) - Q_E^A(t+1) \geq Q_E^A(t+k) - Q_E^A(t+k+1), \textit{for all } t \textit{ and } k > 0, \tag{4.12}$$

then there exists at most one $\bar{d^A} \in \mathbb{Z}_+$, such that if $d^A \leq \bar{d^A}$, then it is optimal to prescribe treatment A first; if $d^A > \bar{d^A}$, then it is optimal to prescribe treatment B first.

Condition (4.11) is slightly stronger than **A3**, and (4.12) implies that if treatment A is effective one unit of time earlier, then the additional QALYs gain are larger for younger patients than elder ones, which is a reasonable assumption.

4.4 NUMERICAL EXAMPLES

In this section, we provide three numerical examples that illustrate when, as a function of the model parameters, it is optimal to initiate treatment with one treatment versus the other. Recall that the three treatment-specific measures we examine are, probability of effectiveness, length of delay and reward/cost. In each of the examples below, we assume treatment A has dominating advantages in two out of these characteristics, and observe how the total expected QALYs gained for treatment sequences $\langle AB \rangle$ and $\langle BA \rangle$ change as a function of the remaining characteristic. We use the following transition matrix P in all three examples:

$$\begin{bmatrix} 0.9 & 0.08 & 0.02 \\ 0.0 & 0.8 & 0.2 \\ 0.0 & 0.0 & 1.0 \end{bmatrix} \tag{4.13}$$

We also set $t_0 = 400$, $D = \{1, 2, 3\}$, $\delta_{t_0} = 1$.

Example 1. In this example, we assume treatment A dominates treatment B in terms of the length of delay, and reward/cost (see model parameters in Table 4.1 and functions (4.2)-(4.4) in Table 4.2). We fix ρ^B to be 0.9, and increase ρ^A from 0.02 to 1. Figure 4.2 depicts the total expected QALYs gained by the patient if on treatment sequence $\langle AB \rangle$ or $\langle BA \rangle$

as a function of ρ^A. To explain the linear relationship between the expected QALYs and ρ^A, recall that in (4.5), $V(\delta_{t_0})$ is a linear function of ρ^A if all other parameters are fixed. Observe that as ρ^A increases, the expected QALYs gained for both sequences increase as well. If $\rho^A \leq 0.68$, then prescribing sequence $\langle BA \rangle$ results in higher expected QALYs gain. This is intuitive because the lower probability of effectiveness for treatment A outweighs its advantages in the length of delay and reward/cost. The value functions intersect at $\rho^A = 0.68$. If $\rho^A > 0.68$, then sequence $\langle AB \rangle$ is preferred, and the difference in QALYs between the two sequences increases as ρ^A becomes larger.

Recall that in Theorem 10, we establish a sufficient condition (4.9) under which it is optimal to prescribe treatment A first. Given the model parameters in this example, the lower bound for ρ^A is $\bar{\rho} = 0.69$. It suggests that as long as $\rho^A \geq 0.69$, treatment sequence $\langle AB \rangle$ is preferred, which is consistent with our numerical observations.

Table 4.1: Parameters for Example 1 in Figure 4.2

Parameter	$\theta = A$	$\theta = B$
ρ^θ	(0,1)	0.9
d^θ	15	16
$Q_E^\theta(t)$	$\frac{50}{1+\frac{t^2}{400^2}}$	$\frac{50}{2^{0.01}(1+\frac{t^2}{400^2})}$
$q^\theta(\cdot)$	[0.03,0.01,0.01]	[0.02,0.01,0.01]
$Q^{\theta^P}(\delta, t)$	$\frac{200}{\delta\sqrt{t}}$	

Example 2. In this example, we assume treatment A dominates treatment B in terms of probability of effectiveness, and length of delay (see model parameters in Table 4.3). We assume $q^A(\delta) = \mu q^B(\delta)$ for all $\delta \in \Delta$, and $Q_E^A(t) = \mu Q_E^B(t)$ for all t. Figure 4.3 depicts the total expected QALYs gained by the patient if on treatment sequence $\langle AB \rangle$ or $\langle BA \rangle$ as a function of μ. The linear relationship between the expected QALYs and μ can be observed from (4.5) given all other parameters fixed. If $\mu \leq 0.874$, then prescribing sequence $\langle BA \rangle$

Table 4.2: Results for Example 1 in Figure 4.2

Function	$X = AB$	$X = BA$
$f_1^{\langle X \rangle}$	24.37	24.17
$f_2^{\langle X \rangle}$	23.53	23.55
$f_3^{\langle X \rangle}$	3.97	3.91

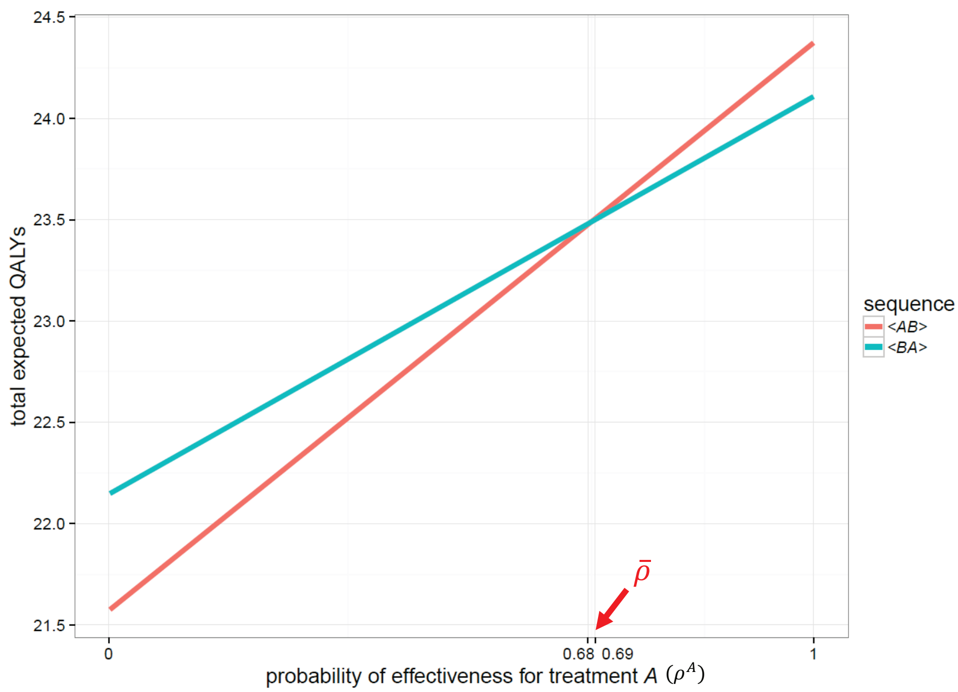

Figure 4.2: Preference of treatment sequence vs. ρ^A. Thus, if $\rho^A \leq 0.68$, then sequence $\langle BA \rangle$ is preferred. Similarly, if $\rho^A > 0.68$, then sequence $\langle AB \rangle$ is preferred. The sufficient condition provided in Theorem 10 is labelled as $\bar{\rho} = 0.69$.

results in a larger expected QALYs gain. The reason is that the higher cost for treatment A outweighs its advantages in the length of delay and the probability of effectiveness. The value functions intersect at $\mu = 0.874$. If $\mu > 0.874$, then sequence $\langle AB \rangle$ is preferred, and the difference in QALYs between the two sequences increases as μ becomes larger.

Recall that in Theorem 11, we derive a necessary and sufficient condition under which it is optimal to prescribe treatment A first. Given model parameters in this example, the value obtained by (4.10) is $\mu^* = 0.874$, which coincides with our numerical observation.

Table 4.3: Parameters for Example 2 in Figure 4.3

Parameter	$\theta = A$	$\theta = B$
ρ^θ	0.9	0.3
d^θ	4	6
$Q_E^\theta(t)$	$\mu \in (0, 1.2]$	$\frac{2500}{1+\frac{t^6}{200^6}}$
$q^\theta(\cdot)$	$\mu \in (0, 1.2]$	[0.2,0.1,0.1]
$Q^{\theta^P}(\delta, t)$	$\frac{1000}{\delta(1+\frac{t^6}{200^6})}$	

Example 3. In this example, we assume treatment A dominates treatment B in terms of probability of effectiveness, and reward/cost (see model parameters in Table 4.4). We fix d^B to be 5, and increase d^A from 1 to 40. Figure 4.4 depicts the total expected QALYs gained by the patient if on treatment sequence $\langle AB \rangle$ or $\langle BA \rangle$ as a function of d^A. Observe that both value functions are non-linear functions of d^A, which is different from Examples 1 and 2. As d^A increases, the expected QALYs gained for both sequences decrease. If $d^A \leq 28$, then prescribing sequence $\langle AB \rangle$ results in a larger expected QALYs gain. The value functions intersect after $d^A = 28$. If $d^A > 28$, then sequence $\langle BA \rangle$ is preferred to $\langle AB \rangle$. This is intuitive because the longer length of delay for treatment A outweighs its advantages in the reward/cost and the probability of effectiveness. In this example, $\bar{d^A} = 28$, which is consistent with Lemma 6.

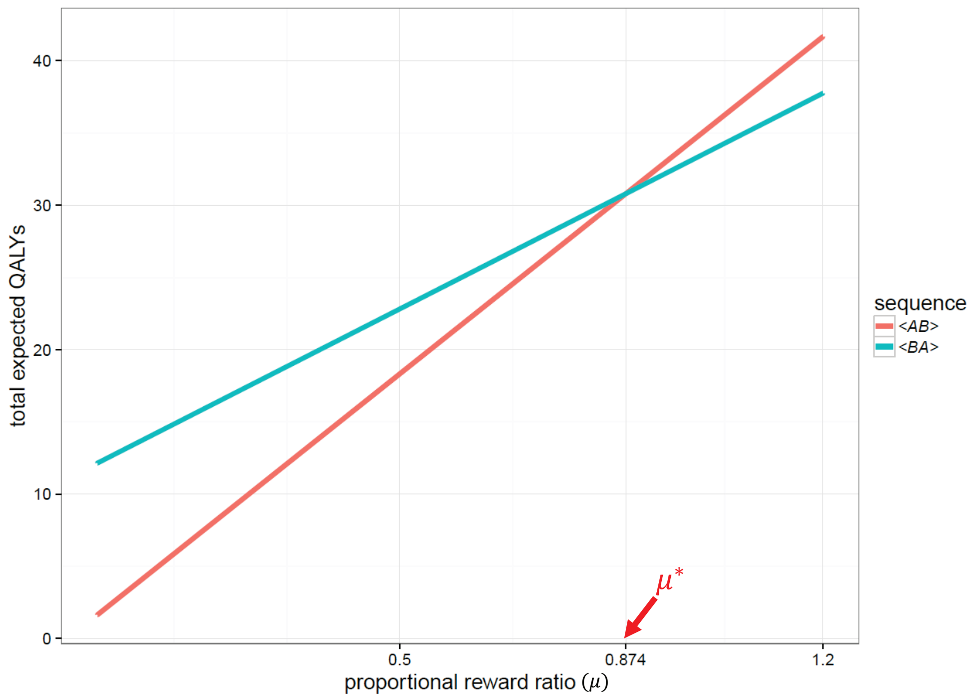

Figure 4.3: Preference of treatment sequence vs. μ. Thus, if $\mu \leq 0.874$, then sequence $\langle BA \rangle$ is preferred. Similarly, if $\mu > 0.874$, then sequence $\langle AB \rangle$ is preferred. The necessary and sufficient condition provided in Theorem 11 coincides $\mu^* = 0.874$.

Table 4.4: Parameters for Example 3 in Figure 4.4

Parameter	$\theta = A$	$\theta = B$
ρ^θ	0.8	0.7
d^θ	$\{1,\ldots,40\}$	5
$Q_E^\theta(t)$	$\frac{1000}{1+\frac{t^6}{200^6}}$	$\frac{1000}{\sqrt{2}(1+\frac{t^6}{200^6})}$
$q^\theta(\cdot)$	[0.02,0.01,0.01]	[0.01,0.01,0.01]
$Q^{\theta^P}(\delta, t)$	$\frac{200}{\delta t}$	

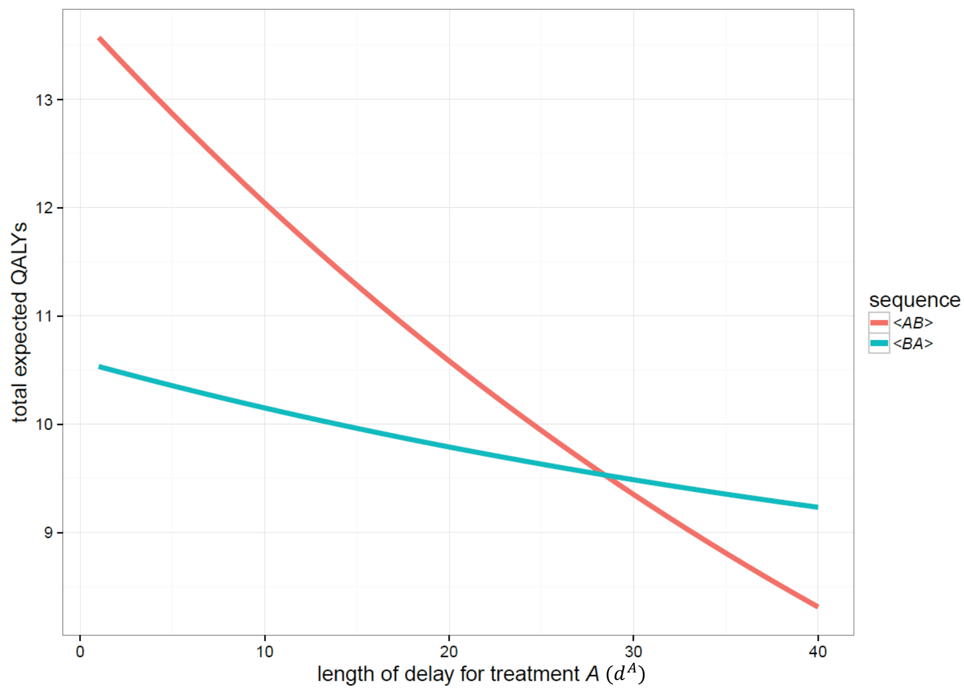

Figure 4.4: Preference of treatment sequence vs. d^A. Thus, if $d^A \leq 28$, then sequence $\langle AB \rangle$ is preferred. If $d^A > 28$, then sequence $\langle BA \rangle$ is preferred.

5.0 CONCLUDING REMARKS AND FUTURE WORK

In this dissertation, we address three problems concerning optimal planning of maintenance activities in different scenarios, motivated by current healthcare practices. We formulate three maintenance optimization models, and each considers a novel setting which has not been examined before in the literature.

Chapter 2 is inspired by various remote monitoring applications that have become prevalent in recent years. We develop a stylized model, in which the patients' data collected remotely corresponds to an imperfect inspection, each scheduled check up for the patient corresponds to a PM action and each unscheduled visit corresponds to RM. We focus on determining the optimal frequency and quantity of imperfect inspections between PM actions so that the total expected cost rate incurred for the patient is minimized over an infinite horizon. One direction for future research would be to investigate modeling extensions that incorporate various practical considerations (e.g., resource constraints, imperfect maintenance outcome, and adverse events) and more complex types of cost functions arising in these real-life situations.

Chapter 3 studies optimal maintenance scheduling in anticipation of possible unpunctual PM actions. We are motivated by a healthcare problem in which patients may not adhere to the prescribed screening to detect early stage cancer (e.g., American Cancer Society suggests women with age 45 to 54 should get mammograms every year [1], but women's compliance with mammography guidelines is low [56]). The consideration of unpunctuality is novel, as in majority of the maintenance optimization literature, PM actions are assumed to be always timely. We provide a thorough analysis of the optimal adjustment of prescribed PM actions for age replacement policies. Future work could first extend the current model to more complex systems and other types of maintenance policies. More general models could also consider dependence between the scheduled PM interval and the unpunctual behavior of the maintenance worker; for example, scheduling a longer PM interval could increase the likelihood and/or magnitude of deviation from the prescribed PM time.

Finally, in Chapter 4 we are motivated by the current practice for treating chronic diseases. We aim to determine the optimal sequence of two treatments and develop a stylized model. Our model considers multiple trade-offs of three treatment-specific characteristics. Future research could focus on more treatment choices, along with the development of more sophisticated sequential decision-making models that relax some of our assumptions.

APPENDIX

PROOFS

A.1 PROOFS FOR CHAPTER 2

Derivation of (2.7) and (2.8). First, we derive the following four equalities:

$$\begin{aligned}\sum_{i=1}^{n}\gamma_i &= F_X(nt) - F_X((n-1)t) + F_X((n-1)t) \\ &- F_X((n-2)t) + \cdots + F_X(t) - F_X(0) \\ &= F_X(nt); \qquad \text{(A.1.1)}\end{aligned}$$

$$\begin{aligned}\sum_{i=1}^{n} i\cdot\gamma_i &= F_X(nt) - F_X((n-1)t) + 2F_X((n-1)t) \\ &- 2F_X((n-2)t) + \cdots + nF_X(t) - nF_X(0) \\ &= \sum_{i=1}^{n} F_X(it); \qquad \text{(A.1.2)}\end{aligned}$$

$$\begin{aligned}\sum_{i=1}^{n}\gamma_i\sum_{j=1}^{i}(1-p)^j = &\Big(F_X(nt) - F_X((n-1)t)\Big)\sum_{j=1}^{1}(1-p)^j \\ + &\Big(F_X((n-1)t) - F_X((n-2)t)\Big)\sum_{j=1}^{2}(1-p)^j \\ + &\ \ldots + \Big(F_X(t) - F_X(0)\Big)\sum_{j=1}^{n}(1-p)^j\end{aligned}$$

$$= \sum_{i=1}^{n} F_X(it)(1-p)^{n-i+1}; \tag{A.1.3}$$

$$\begin{aligned} \gamma_0 + \bar{F}_X((n+1)t) &= F_X((n+1)t) - F_X(nt) + \bar{F}_X((n+1)t) \\ &= 1 - F_X(nt). \end{aligned} \tag{A.1.4}$$

Using (A.1.1)-(A.1.4) to replace the corresponding terms in $\mathbb{E}[L]$, we obtain that:

$$\begin{aligned} \mathbb{E}[L] &= t\Bigg((n+1)\sum_{i=1}^{n}\gamma_i - \sum_{i=1}^{n} i\cdot\gamma_i + \sum_{i=1}^{n}\gamma_i\sum_{j=1}^{i}(1-p)^j\Bigg) + (n+1)t\Bigg(\gamma_0 + \bar{F}_X((n+1)t)\Bigg) \\ &= t\Bigg((n+1)F_X(nt) - \sum_{i=1}^{n}F_X(it) + \sum_{i=1}^{n}F_X(it)(1-p)^{n-i+1}\Bigg) + (n+1)t\Bigg(1 - F_X(nt)\Bigg) \\ &= (n+1)t - t\Bigg(\sum_{i=1}^{n}F_X(it) - \sum_{i=1}^{n}F_X(it)(1-p)^{n-i+1}\Bigg), \end{aligned}$$

which completes derivation of (2.7).

Next, we need the following two equalities:

$$\begin{aligned} &\sum_{i=0}^{n}\int_{(n-i)t}^{(n-i+1)t}(n-i+1)tf_X(x)dx = \\ &\quad t\sum_{i=0}^{n}(n-i+1)\Bigg(F_X((n-i+1)t) - F_X((n-i)t)\Bigg) \\ &= t\Bigg((n+1)F_X((n+1)t) - (n+1)F_X(nt) \\ &\quad + nF_X(nt) - nF_X((n-1)t) + \cdots + F_X(t) - F_X(0)\Bigg) \\ &= (n+1)tF_X((n+1)t) - t\sum_{i=1}^{n}F_X(it), \end{aligned} \tag{A.1.5}$$

and

$$\begin{aligned} &\sum_{i=0}^{n}\int_{(n-i)t}^{(n-i+1)t}xf_X(x)dx = \int_0^{(n+1)t}xf_X(x)dx \\ &= (n+1)tF_X((n+1)t) - \int_0^{(n+1)t}F_X(x)dx. \end{aligned} \tag{A.1.6}$$

Based on (A.1.5) and (A.1.6), we derive the following equality:

$$\begin{aligned} &\sum_{i=0}^{n}\int_{(n-i)t}^{(n-i+1)t}\Big((n-i+1)t-x\Big)f_X(x)dx \\ = &-t\sum_{i=1}^{n}F_X(it)+\int_{0}^{(n+1)t}F_X(x)dx. \end{aligned} \tag{A.1.7}$$

Finally, using (A.1.1)-(A.1.7) to replace the corresponding terms in $\mathbb{E}[C]$, we obtain that:

$$\begin{aligned} \mathbb{E}[C] &= c_1+c_2\Bigg[n\sum_{i=1}^{n}\gamma_i-\sum_{i=1}^{n}i\cdot\gamma_i+\sum_{i=1}^{n}\gamma_i\sum_{j=0}^{i-1}(1-p)^j+n\Big(\gamma_0+\bar{F}_X((n+1)t)\Big)\Bigg] \\ &+\lambda\sum_{i=0}^{n}\Bigg(\int_{(n-i)t}^{(n-i+1)t}((n-i+1)t-x)f_X(x)dx+\gamma_i t\sum_{j=1}^{i}(1-p)^j\Bigg)+\zeta\sum_{i=0}^{n}\gamma_i \\ &+\sum_{i=0}^{n}\int_{(n-i)t}^{(n-i+1)t}\Bigg(\sum_{k=0}^{i}p_{ik}\times\int_0^{(n-i+1)t-x+kt}\theta(u)du\Bigg)f_X(x)dx \\ &= c_1+c_2\Bigg[nF_X(nt)-\sum_{i=1}^{n}F_X(it)+\sum_{i=1}^{n}F_X(it)(1-p)^{n-i}+n(1-F_X(nt))\Bigg] \\ &+\lambda\Bigg(-t\sum_{i=1}^{n}F_X(it)+\int_0^{(n+1)t}F_X(x)dx+t\sum_{i=1}^{n}F_X(it)(1-p)^{n-i+1}\Bigg)+\zeta F_X((n+1)t) \\ &+\sum_{i=0}^{n}\gamma_i\mathbb{E}\Bigg[\int_0^{D_i}\theta(u)du\Bigg] \\ &= c_1+c_2\Bigg(n-\Bigg(\sum_{i=1}^{n}F_X(it)-\sum_{i=1}^{n}F_X(it)(1-p)^{n-i}\Bigg)\Bigg) \\ &+\lambda\Bigg(\int_0^{(n+1)t}F_X(x)dx-t\Bigg(\sum_{i=1}^{n}F_X(it)-\sum_{i=1}^{n}F_X(it)(1-p)^{n-i+1}\Bigg)\Bigg)+\zeta F_X((n+1)t) \\ &+\sum_{i=0}^{n}\gamma_i\mathbb{E}\Bigg[\int_0^{D_i}\theta(u)du\Bigg], \end{aligned}$$

where

$$p_{ik}=\begin{cases}(1-p)^k\cdot p, & \text{if } \ k<i; \\ (1-p)^i\cdot 1, & \text{if } \ k=i.\end{cases}$$

and

$$\mathbb{E}\Bigg[\int_0^{D_i}\theta(u)du\Bigg]=\int_{(n-i)t}^{(n-i+1)t}\Bigg(\sum_{k=0}^{i}p_{ik}\times\int_0^{(n-i+1)t-x+kt}\theta(u)du\Bigg)\frac{f_X(x)}{\gamma_i}dx,$$

and

$$\mathbb{E}[D_i] = \int_{(n-i)t}^{(n-i+1)t} ((n-i+1)t - x)\frac{f_X(x)}{\gamma_i}dx + t\sum_{j=1}^{i}(1-p)^j.$$

which completes derivation of (2.8).

Proof of Proposition 1. From $\lim_{t\to+\infty} F(it) = 1, \forall i = 1, 2, ..., n$, it follows that

$$\lim_{t\to+\infty} N(t,n) = n - \sum_{i=1}^{n}(1-p)^{n-i}, \tag{A.1.8}$$

$$\lim_{t\to+\infty} M(t,n) = n - \sum_{i=1}^{n}(1-p)^{n-i+1}. \tag{A.1.9}$$

Next to show that

$$\lim_{t\to+\infty} Z(t,n) = \eta, \tag{A.1.10}$$

we consider two cases.

Case 1 ($i < n$):

For $t \to +\infty$, $\gamma_i \to 0$. If $k = 0$, then

$$\begin{aligned} & \lim_{t\to+\infty} \int_{(n-i)t}^{(n-i+1)t} p_{ik} \int_0^{(n-i+1)t-x+kt} \theta(u) f_X(x) du dx \\ \leq \quad & \lim_{t\to+\infty} \int_{(n-i)t}^{(n-i+1)t} p_{i0} \int_0^{t} \theta(u) f_X(x) du dx \leq p_{i0}\eta\gamma_i = 0. \end{aligned}$$

If $0 < k \leq i$, then

$$\lim_{t\to+\infty} \int_{(n-i)t}^{(n-i+1)t} p_{ik} \int_0^{(n-i+1)t-x+kt} \theta(u) f_X(x) du dx = p_{ik}\eta\gamma_i = 0.$$

Thus, we conclude that

$$\lim_{t\to+\infty} \sum_{i=0}^{n-1} \int_{(n-i)t}^{(n-i+1)t} \left(\sum_{k=0}^{i} p_{ik} \times \int_0^{(n-i+1)t-x+kt} \theta(u) du \right) f_X(x) dx = 0. \tag{A.1.11}$$

Case 2 ($i = n$):

For $t \to +\infty$, $\gamma_n \to 1$, i.e., the probability that a hidden failure occurs within the first IPI interval is 1. If $k = 0$, then

$$\begin{aligned}
& \lim_{t\to+\infty} \int_{(n-i)t}^{(n-i+1)t} p_{ik} \int_0^{(n-i+1)t-x+kt} \theta(u) f_X(x) du dx \\
= & \lim_{t\to+\infty} \int_0^t p_{n0} \int_0^{t-x} f_X(x)\theta(u) du dx \\
= & \lim_{t\to+\infty} \int_0^t p_{n0} \int_0^{t-u} f_X(x)\theta(u) dx du \\
= & \lim_{t\to+\infty} \Big[\int_0^\Delta p_{n0} F_X(t-u)\theta(u) du + \int_\Delta^t p_{n0} F_X(t-u)\theta(u) du \Big].
\end{aligned}$$

From Assumption **A1**,

$$\lim_{t\to+\infty} \int_0^\Delta p_{n0} F_X(t-u)\theta(u) du = p_{n0}\eta, \tag{A.1.12}$$

and

$$\lim_{t\to+\infty} \int_\Delta^t p_{n0} F_X(t-u)\theta(u) du \le \lim_{t\to+\infty} \int_\Delta^t p_{n0}\theta(u) du = 0. \tag{A.1.13}$$

If $0 < k \le n$, then

$$\begin{aligned}
& \lim_{t\to+\infty} \int_{(n-i)t}^{(n-i+1)t} p_{ik} \int_0^{(n-i+1)t-x+kt} \theta(u) f_X(x) du dx \\
= & \lim_{t\to+\infty} \int_0^t p_{nk} \int_0^{(k+1)t-x} \theta(u) f_X(x) du dx \\
= & \; p_{nk}\eta\gamma_n = p_{nk}\eta. \qquad\qquad (A.1.14)
\end{aligned}$$

From (A.1.12)-(A.1.14), we obtain

$$\lim_{t\to+\infty} \int_0^t \sum_{k=0}^n p_{nk} \int_0^{(k+1)t-x} \theta(u) f_X(x) du dx = \eta. \tag{A.1.15}$$

Combining the results of (A.1.11) and (A.1.15), we obtain (A.1.10). Hence, for large enough t, we have

$$\mathbb{E}[L] \;\approx\; \left(1 + \sum_{i=1}^n (1-p)^{n-i+1}\right) t, \tag{A.1.16}$$

$$\begin{aligned}\mathbb{E}[C] &\approx c_1 + c_2\sum_{i=1}^{n}(1-p)^{n-i} - \lambda\int_0^{(n+1)t}\bar{F}_X(x)dx \\ &\quad +\lambda\left(1+\sum_{i=1}^{n}(1-p)^{n-i+1}\right)t+\zeta+\eta. \qquad \text{(A.1.17)}\end{aligned}$$

The result follows from the fact that $\lim_{t\to+\infty}\int_0^t \bar{F}_X(x)dx = \mathbb{E}[X] < +\infty$, $\zeta < +\infty$ and $\eta < +\infty$.

■

Proof of Proposition 2. From (A.1.16) and (A.1.17) in the proof of Proposition 1, we know that for sufficiently large t,

$$\begin{aligned}\Omega(t,0) &\approx \frac{c_1 - \lambda\int_0^t \bar{F}_X(x)dx + \lambda t + \zeta + \eta}{t} \\ &= \lambda + \frac{c_1 - \lambda\mathbb{E}[X] + \zeta + \eta}{t}. \qquad \text{(A.1.18)}\end{aligned}$$

When $c_1 + \zeta + \eta < \lambda\mathbb{E}[X]$, the function $\Omega(t,0)$ is increasing for sufficiently large values of t and converges to λ by Proposition 1. Note also that $\lim_{t\to+0}\Omega(t,0) = +\infty$. Therefore, $\mathbf{CP}_0$ has a finite optimal solution by the continuity property of $\Omega(t,0)$ for $t > 0$.

Next, we show that if $\zeta = \eta = 0$ and $c_1 < \lambda\mathbb{E}[X]$, then $\Omega(t,0)$ is quasiconvex. Observe that

$$\begin{aligned}\frac{\partial\Omega(t,0)}{\partial t} &= \frac{-c_1 + \lambda\left(F_X(t)t - \int_0^t F_X(x)dx\right)}{t^2} \\ &= \frac{-c_1 + \lambda\int_0^t xf_X(x)dx}{t^2}.\end{aligned}$$

Define

$$g(t) = \frac{-c_1 + \lambda\int_0^t xf_X(x)dx}{t^2}.$$

Then

$$\lim_{t\to+0} g(t) = -\infty \quad \text{and} \quad \lim_{t\to+\infty} g(t) = 0.$$

Consider the numerator of $g(t)$ and define

$$h(t) = -c_1 + \lambda\int_0^t xf_X(x)dx.$$

Observe that $h(t)$ is monotonically increasing, $h(0) < 0$ and

$$\lim_{t\to+\infty}\int_0^t xf_X(x)dx \quad = \quad \mathbb{E}[X].$$

If $c_1 < \lambda\mathbb{E}[X]$, then

$$\lim_{t\to+\infty} h(t) = -c_1 + \lambda\mathbb{E}[X] > 0.$$

Therefore, as $h(t)$ is a continuous function, then there exists τ, such that $h(\tau) = 0$ and $g(\tau) = 0$. Moreover, if $t \in (\tau, +\infty)$, then $g(t) > 0$; if $t \in (0, \tau)$, then $g(t) < 0$. These observations imply the result.

■

Proof of Proposition 3. From (A.1.16) and (A.1.17) in the proof of Proposition 1, we know that for sufficiently large t,

$$\Omega(t,n) \quad \approx \quad \lambda + \frac{c_1 + c_2\sum_{i=1}^{n}(1-p)^{n-i} - \lambda\mathbb{E}[X] + \zeta + \eta}{\left(1 + \sum_{i=1}^{n}(1-p)^{n-i+1}\right)t}.$$

When $c_1 + c_2\sum_{i=1}^{n}(1-p)^{n-i} + \zeta + \eta < \lambda\mathbb{E}[X]$, the function $\Omega(t,n)$ is increasing for sufficiently large t. Thus, the result can be shown using arguments similar to those used in the proof of Proposition 2.

■

Proof of Proposition 4. To show necessity of (2.13), observe from (2.8) that if (2.13) does not hold, then

$$\begin{aligned}
\mathbb{E}[C] \quad &\geq \quad c_1 + c_2\Bigg(n - N(t,n)\Bigg) - \lambda\int_0^{(n+1)t}\bar{F}_X(x)dx + \lambda\Bigg((n+1)t - tM(t,n)\Bigg) &\text{(A.1.19)}\\
&> \quad \Bigg(c_1 + c_2\sum_{i=1}^{n}(1-p)^{n-i} - \lambda\mathbb{E}[X]\Bigg) + \lambda\Bigg((n+1)t - tM(t,n)\Bigg) &\text{(A.1.20)}\\
&\geq \quad \lambda\Bigg((n+1)t - tM(t,n)\Bigg) = \lambda\mathbb{E}[L].
\end{aligned}$$

Where (A.1.19) is true because ζ and $\theta(u)$ are both non-negative, and (A.1.20) holds due to the fact that $N(t,n) < n - \sum_{i=1}^{n}(1-p)^{n-i}$ and $\int_0^{(n+1)t} \bar{F}_X(x)dx < \int_0^{+\infty} \bar{F}_X(x)dx = \mathbb{E}[X]$ by Assumption **A4**. Note that $\mathbb{E}[L] > 0$. Then the necessary result follows from Proposition 1.

■

Proof of Corollary 1. The results follows directly from Propositions 3 and 4. ■

Proof of Proposition 5. Observe from (2.8) that $\mathbb{E}[C]$ can be re-written as follows:

$$\begin{aligned}\mathbb{E}[C] &= c_1 + c_2(n - N(t,n)) + \zeta F_X((n+1)t) \\ &+ \lambda\left((n+1)t - \int_0^{(n+1)t} \bar{F}_X(x)dx - tM(t,n)\right) \\ &= c_1 + c_2(n - N(t,n)) - \lambda \int_0^{(n+1)t} \bar{F}_X(x)dx + \lambda\mathbb{E}[L] + \zeta F_X((n+1)t),\end{aligned}$$

which implies that $\Omega(t,n)$ is given by:

$$\Omega(t,n) = \lambda + c_2\frac{n - N(t,n)}{\mathbb{E}[L]} + \frac{c_1 - \lambda \int_0^{(n+1)t} \bar{F}_X(x)dx + \zeta F_X((n+1)t)}{\mathbb{E}[L]}. \tag{A.1.21}$$

Consider the second term in (A.1.21). Using (2.7) we obtain:

$$c_2\frac{n - N(t,n)}{\mathbb{E}[L]} = c_2\frac{n - N(t,n)}{(n+1)t - tM(t,n)}. \tag{A.1.22}$$

Because t is fixed, then $nt \to +\infty$ as $n \to +\infty$. Thus, $F_X(nt) \to 1$. Then $M(t,n)$, given by (2.4), can be approximated for large enough n as follows:

$$\begin{aligned}M(t,n) &\approx \sum_{i=1}^{\tilde{N}} F_X(it) + (n - \tilde{N}) \\ &\quad - \sum_{i=1}^{\tilde{N}} F_X(it)(1-p)^{n-i+1} - \sum_{i=\tilde{N}+1}^{n} (1-p)^{n-i+1} \\ &\approx \sum_{i=1}^{\tilde{N}} F_X(it) + (n - \tilde{N}) - \frac{1-p}{p},\end{aligned} \tag{A.1.23}$$

where $\widetilde{N}$ is a large enough constant. Similarly, we derive that:

$$N(t,n) \approx \sum_{i=1}^{\widetilde{N}} F_X(it) + (n - \widetilde{N}) - \frac{1}{p}. \tag{A.1.24}$$

After substituting (A.1.23)-(A.1.24) into (A.1.22) we conclude that if $n \to +\infty$, then

$$c_2 \frac{n - N(t,n)}{\mathbb{E}[L]} \to \frac{c_2}{t}. \tag{A.1.25}$$

Finally, the required result follows from (A.1.21) using (A.1.25) and the fact that $\int_0^{(n+1)t} \bar{F}_X(x)dx \to \mathbb{E}[X]$ as $n \to +\infty$.

■

Proof of Lemma 1. Note that (2.6) can be re-written as:

$$Z(t,n) = \sum_{i=0}^{n} \sum_{k=0}^{i} p_{ik} \times \int_{(n-i)t}^{(n-i+1)t} \int_0^{(n-i+1)t-x+kt} \theta(u) f_X(x) du dx.$$

Given $t > \Delta$ and using Assumption **A1**, we have

$$\begin{aligned}
&\sum_{i=0}^{n} \sum_{k=1}^{i} p_{ik} \int_{(n-i)t}^{(n-i+1)t} \int_0^{(n-i+1)t-x+kt} \theta(u) f_X(x) du dx \\
=&\sum_{i=0}^{n} \sum_{k=1}^{i} p_{ik} \int_{(n-i)t}^{(n-i+1)t} \int_0^{\Delta} \theta(u) f_X(x) du dx \\
=&\sum_{i=0}^{n} \sum_{k=1}^{i} p_{ik} \eta \int_{(n-i)t}^{(n-i+1)t} f_X(x) dx \\
=&\sum_{i=0}^{n} \sum_{k=1}^{i} p_{ik} \eta \gamma_i = (1-p)\eta.
\end{aligned} \tag{A.1.26}$$

If $k = 0$ and $t > \Delta$, then

$$\begin{aligned}
&\sum_{i=0}^{n} p_{ik} \int_{(n-i)t}^{(n-i+1)t} \int_0^{(n-i+1)t-x+kt} \theta(u) f_X(x) du dx \\
=&p \sum_{i=0}^{n} \int_{(n-i)t}^{(n-i+1)t} \int_0^{(n-i+1)t-x} \theta(u) f_X(x) du dx
\end{aligned}$$

$$=p\sum_{i=0}^{n}\int_0^t \theta(u)\int_{(n-i)t}^{(n-i+1)t-u} f_X(x)dxdu$$

$$=p\sum_{i=1}^{n+1}\int_0^t \theta(u)\Bigg(F_X(it-u)-F_X((i-1)t)\Bigg)du \tag{A.1.27}$$

Note that

$$\begin{aligned}
& \int_0^t \theta(u)\Bigg(F_X(it-u)-F_X((i-1)t)\Bigg)du \\
< & \int_0^t \theta(u)\Bigg(F_X(it)-F_X((i-1)t)\Bigg)du \\
= & \ \eta\Bigg(F_X(it)-F_X((i-1)t)\Bigg),
\end{aligned}$$

and

$$\lim_{n\to+\infty}\sum_{i=1}^{n}\eta\Bigg(F_X(it)-F_X((i-1)t)\Bigg)=\eta. \tag{A.1.28}$$

Therefore, by (A.1.27)-(A.1.28) we have that

$$\lim_{n\to+\infty} p\sum_{i=1}^{n+1}\int_0^t \theta(u)\Big(F_X(it-u)-F_X((i-1)t)\Big)du$$

also exists and is finite. Finally, the required result follows from (A.1.26).

■

Proof of Proposition 6. The result follows directly from Lemma 1 using the arguments similar to those used in the proof of Proposition 5.

■

Derivation of (2.20). Equality (2.20) follows from (A.1.26) and (A.1.27) derived in the proof of Lemma 1.

Proof of Proposition 7. For $n = 1$ function $\Omega(t, n)$ simplifies to

$$\Omega(t,1) \quad = \quad \frac{c_1+c_2+\lambda\left(-\int_0^{2t}\bar{F}_X(x)dx+2t-tpF_X(t)\right)}{2t-tpF_X(t)}$$

$$= \quad \frac{c_1 + c_2 - \lambda \int_0^{2t} \bar{F}_X(x)dx}{2t - tpF_X(t)} + \lambda,$$

and its first derivative is given by:

$$\frac{\partial \Omega(t,1)}{\partial t} = \frac{-2\lambda \bar{F}_X(2t)(2t - tpF_X(t))}{(2t - tpF_X(t))^2} - \frac{(c_1 + c_2 - \lambda \int_0^{2t} \bar{F}_X(x)dx)}{(2t - tpF_X(t))^2} \times \left(2 - pF_X(t) - tpf_X(t)\right). \tag{A.1.29}$$

Denote the numerator part of $\partial\Omega(t,1)/\partial t$ by $\Gamma^{(1)}(t)$ (note that both terms in (A.1.29) have the same denominator). Then it is easy to show $\lim_{t\to+0} \Gamma^{(1)}(t) < 0$. Moreover, $\lim_{t\to+\infty} \Gamma^{(1)}(t) = -(c_1 + c_2 - \lambda\mathbb{E}(x))(2-p) > 0$ by (2.22). Thus, it is enough to show that $\Gamma^{(1)}(t) = 0$ has a unique solution for $t \in (0 + \infty)$.

Recall that for the Weibull distribution given the shape and scale parameters α and β, respectively, we have:

$$f'_X(t) = \alpha(\alpha-1)\frac{t^{\alpha-2}}{\beta^\alpha}e^{-(\frac{t}{\beta})^\alpha} - (\alpha\frac{t^{\alpha-1}}{\beta^\alpha})^2 e^{-(\frac{t}{\beta})^\alpha}. \tag{A.1.30}$$

Consider the term $2 - pF_X(t) - tpf_X(t)$ in (A.1.29). Specifically,

$$\begin{aligned} 2 - pF_X(t) - tpf_X(t) &\geq 2 - p - \alpha p \frac{(\frac{t}{\beta})^\alpha}{e^{(\frac{t}{\beta})^\alpha}} \\ &\geq 2 - p - \alpha p \frac{1}{e} > 0, \end{aligned} \tag{A.1.31}$$

which follows from (2.23) and the observation that function $(t/\beta)^\alpha / e^{(t/\beta)^\alpha}$ achieves its maximum for $(t/\beta)^\alpha = 1$.

Next, $\Gamma^{(1)}(t)$ can be re-written as:

$$\begin{aligned} &\Gamma^{(1)}(t) \\ = \; &(2 - pF_X(t) - tpf_X(t))\left(\left[-\frac{2\bar{F}_X(2t)(2t - tpF_X(t))}{2 - pF_X(t) - tpf_X(t)} \right.\right. \\ &\left.\left. + \int_0^{2t} \bar{F}_X(x)dx\right]\lambda - (c_1 + c_2)\right). \end{aligned}$$

Define

$$\Gamma^{(2)}(t) = \int_0^{2t} \bar{F}_X(x)dx - \frac{2\bar{F}_X(2t)(2t - tpF_X(t))}{2 - pF_X(t) - tpf_X(t)}$$

From (A.1.31) we conclude that $\Gamma^{(1)}(t) = 0$ has a unique solution if $\Gamma^{(2)}(t)$ is increasing, i.e., its first derivative is positive for $t > 0$.

Let $D(t) = 2t - tpF_X(t)$. Then $D'(t) = 2 - pF_X(t) - tpf_X(t)$ and $D''(t) = -2pf_X(t) - tpf'_X(t)$. Taking the first derivative of $\Gamma^{(2)}(t)$, we obtain:

$$\begin{aligned}\left(\Gamma^{(2)}(t)\right)' &= 2\bar{F}_X(2t) + \frac{2\bar{F}_X(2t)D(t)D''(t)}{(D'(t))^2} \\ &\quad - \frac{-4f_X(2t)D(t) + 2\bar{F}_X(2t)D'(t)}{D'(t)} \\ &= \frac{4f_X(2t)D(t)D'(t) + 2\bar{F}_X(2t)D(t)D''(t)}{(D'(t))^2}.\end{aligned}$$

Note that $D(t) > 0$ for $t > 0$. Thus, it is sufficient to show that the following term:

$$\begin{aligned}\Lambda^{(1)}(t) &= 4f_X(2t)D'(t) + 2\bar{F}_X(2t)D''(t) \\ &= 4f_X(2t)(2 - pF_X(t) - tpf_X(t)) \\ &+ 2\bar{F}_X(2t)(-2pf_X(t) - tpf'_X(t)), \qquad \text{(A.1.32)}\end{aligned}$$

is strictly positive.

Substituting (A.1.30) into (A.1.32) we obtain:

$$\Lambda^{(1)}(t) = 2\alpha e^{-(\frac{2t}{\beta})^\alpha} \cdot \frac{t^{\alpha-1}}{\beta^\alpha} \cdot \left(2^{\alpha+1} - 2^\alpha p + e^{-(\frac{t}{\beta})^\alpha} \times \left(2^\alpha p - 2^\alpha p\alpha\frac{t^\alpha}{\beta^\alpha} - p - \alpha p + p\alpha\frac{t^\alpha}{\beta^\alpha}\right)\right).$$

Then

$$\begin{aligned}\Lambda^{(2)}(t) &= 2^{\alpha+1} - 2^\alpha p + e^{-(\frac{t}{\beta})^\alpha} \times \left(2^\alpha p - 2^\alpha p\alpha\frac{t^\alpha}{\beta^\alpha} - p - \alpha p + p\alpha\frac{t^\alpha}{\beta^\alpha}\right) \\ &= 2^{\alpha+1} - 2^\alpha p + \frac{2^\alpha p - p - p\alpha}{e^{(\frac{t}{\beta})^\alpha}} - (2^\alpha - 1)\alpha p\frac{(\frac{t}{\beta})^\alpha}{e^{(\frac{t}{\beta})^\alpha}} \\ &\geq 2^{\alpha+1} - 2^\alpha p - (2^\alpha - 1)\alpha p\frac{1}{e}, \qquad \text{(A.1.33)} \\ &\geq 2^\alpha(2-p) - (2^\alpha - 1)(2-p) \\ &= 2 - p > 0,\end{aligned}$$

where (A.1.33) holds because $2^\alpha - 1 - \alpha \geq 0$ for $\alpha \geq 1$ and the fact that function $(t/\beta)^\alpha / e^{(t/\beta)^\alpha}$ obtains its maximum for $(t/\beta)^\alpha = 1$. The last inequality follows from (2.23). Therefore, $\Lambda^{(1)}(t) > 0$ and the result follows. ■

A.2 PROOFS FOR CHAPTER 3

Proof of Proposition 8. We follow the same argument as in [16] (see Page 97). First, observe that $m(T)$ is continuous on $[0, +\infty)$. Let $\{T_i\}$ be an infinite sequence such that $\max\{0, -\mu_Y\} < T_1 < T_2 < \cdots$, with $h_X(T_{i-1}) < h_X(T_i)$ for $i = 2, 3, \ldots$, and $\lim_{i \to +\infty} h_X(T_i) = +\infty$. Such a sequence exists because $h_X(t)$ is strictly increasing and unbounded by **A2**. Then it is true that

$$\begin{aligned} m(T_i) &= (T_i + \mu_Y) \int_a^b h_X(T_i + y) f_Y(y) dy - \int_a^b \int_0^{T_i + y} h_X(x) dx f_Y(y) dy \\ &\geq (T_1 + \mu_Y) \int_a^b h_X(T_i + y) f_Y(y) dy - \int_a^b \int_0^{T_1 + y} h_X(x) dx f_Y(y) dy \to +\infty, \end{aligned}$$

because $\int_{T_1+y}^{T_i+y} h_X(x)dx \leq (T_i - T_1) h_X(T_i + y)$ by **A2**. Thus, $m(T)$ is unbounded as T approaches $+\infty$. It is straightforward to verify that $m(T)$ is increasing in T, since the first derivative of $m(T)$ is

$$\begin{aligned} m'(T) &= (T + \mu_Y) \int_a^b h'_X(T + y) f_Y(y) dy + \int_a^b h_X(T + y) f_Y(y) dy - \int_a^b \frac{d}{dT} \int_0^{T+y} h_X(x) f_Y(y) dx dy \\ &= (T + \mu_Y) \int_a^b h'_X(T + y) f_Y(y) dy + \int_a^b h_X(T + y) f_Y(y) dy - \int_a^b h_X(T + y) f_Y(y) dy \\ &= (T + \mu_Y) \int_a^b h'_X(T + y) f_Y(y) dy > 0, \qquad \text{(A.2.1)} \end{aligned}$$

and the first derivative of the hazard rate function $h'_X(x) > 0$ by **A2**. The uniqueness of the solution to (3.7) is guaranteed when the limiting value of $\lim\limits_{T \to \max\{-a,0\}} m(T)$ is less than k_1. Otherwise, $\inf \Omega^{\mathcal{A}}(T) = \lim_{T \to + \max\{-a,0\}} \Omega^{\mathcal{A}}(T)$.

Given the unique solution T^* to (3.7), rearranging terms yields

$$c_m \int_a^b h_X(T^* + y) f_Y(y) dy = \frac{c_m \int_a^b \int_0^{T^*+y} h_X(x) f_Y(y) dx dy + c_p}{T^* + \mu_Y}.$$

It can be easily seen that $\Omega^{\mathcal{A}}(T^*) = c_m \int_a^b h_X(T^* + y) f_Y(y) dy$. ∎

Proof of Proposition 9. Based on Jensen's Inequality, we know that if $\mu_Y = 0$, then

$$\int_a^b \int_0^{T+y} h_X(x) dx f_Y(y) dy \geq \int_0^{T+\mu_Y} h_X(x) dx = \int_0^T h_X(x) dx.$$

Again, applying Jensen's Inequality, the concavity of $h_X(t)$ implies

$$\int_a^b h_X(T + y) f_Y(y) dy \leq h_X(T + \mu_Y) = h_X(T).$$

Therefore $m(T) \leq \widetilde{m}(T)$ by definitions (3.8) and (3.9) for any T that is feasible for both $m(T)$ and $\widetilde{m}(T)$. The result follows directly. ∎

Proof of Theorem 2. First, recall that $\widetilde{m}(T)$ and $m(T)$ are increasing. From (3.6)-(3.9), if we can show that for all T, $m(T) > \widetilde{m}(T)$, then $T^* < \widetilde{T}^*$. The first derivatives of $m(T)$ and $\widetilde{m}(T)$ are given by

$$m'(T) = (T + \mu_Y) \int_a^b h'_X(T + y) f_Y(y) dy \quad \text{and} \quad \widetilde{m}'(T) = h'_X(T) T,$$

respectively. If $0 \leq a < b$, then $\mu_Y > 0$, with $h_X(t)$ convex which implies

$$m'(T) = (T + \mu_Y) \int_a^b h'_X(T + y) f_Y(y) dy > T \int_a^b h'_X(T) f_Y(y) dy = h'_X(T) T = \widetilde{m}'(T).$$

Thus, $m(T)$ is increasing faster than $\widetilde{m}(T)$, and both problems have the same feasible region.

Next, we compare the limiting values of $\lim_{T \to +0} m(T)$ and $\lim_{T \to +0} \widetilde{m}(T)$. By **A1**, $\lim_{T \to +0} \widetilde{m}(T) = 0$. Hence, by (3.11), $\lim_{T \to +0} m(T) \geq \lim_{T \to +0} \widetilde{m}(T)$.

Coupled with the fact that $m'(T) > \widetilde{m}'(T)$, we conclude that $m(T) > \widetilde{m}(T)$ for all $T > 0$, which implies $T^* < \widetilde{T}^*$. ■

Proof of Corollary 2. Note that (3.11) implies

$$\mu_Y \int_a^b h_X(y) f_Y(y) dy - \int_a^b \int_0^y h_X(x) f_Y(y) dx dy = \int_a^b \left(\mu_Y h_X(y) - \int_0^y h_X(x) dx \right) f_Y(y) dy \geq 0, \tag{A.2.2}$$

and a sufficient condition for (A.2.2) is

$$\mu_Y h_X(y) - \int_0^y h_X(x) dx \geq 0, \text{ for any } y \in [a, b]. \tag{A.2.3}$$

By **A1** and the assumption that $h_X(x)$ is convex, it is obvious that

$$\int_0^y h_X(x) dx \leq \frac{1}{2} h_X(y) y, \forall\ y \geq 0.$$

If $\mu_Y \geq \frac{b}{2}$, then $\mu_Y \geq \frac{y}{2}$ for any $y \in [a, b]$, and therefore (A.2.3) holds. The result follows directly. ■

Proof of Theorem 3. First, we prove $\lim_{T \to -a} m(T) < \widetilde{m}(-a)$. Note that

$$\lim_{T \to -a} m(T) = (-a + \mu_Y) \int_a^b h_X(-a + y) f_Y(y) dy - \int_a^b \int_0^{-a+y} h_X(x) f_Y(y) dx dy, \tag{A.2.4}$$

and

$$\begin{aligned} \widetilde{m}(-a) =& h_X(-a)(-a) - \int_0^{-a} h_X(x) dx \\ =& (-a + \mu_Y) h_X(-a) - \Big(\int_0^{-a} h_X(x) dx + h_X(-a) \mu_Y \Big). \end{aligned} \tag{A.2.5}$$

Observe that $h_X(-a) > \int_a^b h_X(-a + y) f_Y(y) dy$ for $a < b \leq 0$. Thus, comparing (A.2.4) and (A.2.5), if we can show that

$$\int_0^{-a} h_X(x) dx + h_X(-a) \mu_Y \leq \int_a^b \int_0^{-a+y} h_X(x) f_Y(y) dx dy, \tag{A.2.6}$$

then the result follows directly. Recall that $\int_0^T h_X(x)dx$ is convex ($\frac{d^2}{dT^2}\int_0^T h_X(x)dx = h'_X(T) > 0$ by **A2**), hence, based on Jensen's Inequality,

$$\int_a^b \int_0^{-a+y} h_X(x)dx f_Y(y)dy \geq \int_0^{-a+\mu_Y} h_X(x)dx$$
$$= \int_0^{-a} h_X(x)dx - \int_{-a+\mu_Y}^{-a} h_X(x)dx. \qquad \text{(A.2.7)}$$

Clearly,

$$-\int_{-a+\mu_Y}^{-a} h_X(x)dx \geq -h_X(-a)(-\mu_Y) = h_X(-a)\mu_Y. \qquad \text{(A.2.8)}$$

Inequality (A.2.8) is by **A2** and the fact that $h_X(x)$ is convex. Thus, (A.2.7) and (A.2.8) imply (A.2.6), and we have $\lim_{T\to -a} m(T) \leq \widetilde{m}(-a)$.

The remainder of the proof is similar to that of Theorem 2. Specifically, using (A.2.1) we have

$$m'(T) = (T+\mu_Y)\int_a^b h'_X(T+y)f_Y(y)dy < \int_a^b h'_X(T)Tf_Y(y)dy = h'_X(T)T = \widetilde{m}'(T),$$

where the inequality follows by the negativity of a and b. Thus, $m(T)$ is increasing slower than $\widetilde{m}(T)$. As $a < b \leq 0$, we have $T > \max\{-a, 0\}$, and there are two possible cases:

(i) if $\widetilde{m}(-a) \geq k_1$, then $\widetilde{T}^* \leq -a$, which implies that $\widetilde{T}^* < T^*$.

(ii) if $\widetilde{m}(-a) < k_1$, then both $\widetilde{T}^*$ and T^* are greater than $-a$. We already know that $\lim_{T\to -a} m(T) < \widetilde{m}(-a)$. From $m'(T) < \widetilde{m}'(T)$, we conclude that $m(T) < \widetilde{m}(T)$ for any given $T > -a$, and $T^* > \widetilde{T}^*$ follows directly. ■

Lemma 7. *(Edmundson-Madansky inequality ([40])) Let $Y \in [a,b]$ have a c.d.f. $F_Y(y)$ and finite mean μ_Y. Suppose ϕ is a bounded convex function of $y \in [a,b]$. An upper bound for $\mathbb{E}[\phi(Y)]$ is*

$$\mathbb{E}[\phi(Y)] \leq \frac{\phi(b)-\phi(a)}{b-a}(\mu_Y - a) + \phi(a).$$

Proof of Theorem 4. The first inequality in (3.12) is by Theorem 1. To prove the second equality, recall that $\widetilde{\Omega}^{\mathcal{A}}(\widetilde{T}^*) = c_m h_X(\widetilde{T}^*)$ ([16]). If $\widetilde{T}^* - \mu_Y$ is feasible to optimization problem (3.1), then it is obvious that $\Omega^{\mathcal{A}}(T^*) \leq \Omega^{\mathcal{A}}(\widetilde{T}^* - \mu_Y)$. Therefore, using (3.5) we have the second inequality

$$\frac{\Omega^{\mathcal{A}}(T^*)}{\widetilde{\Omega}^{\mathcal{A}}(\widetilde{T}^*)} \leq \frac{c_m \int_a^b \int_0^{\widetilde{T}^*-\mu_Y+y} h_X(x) f_Y(y) dx dy + c_p}{c_m h_X(\widetilde{T}^*)\widetilde{T}^*}.$$

To obtain the third inequality, we apply the Edmundson-Madansky upper bound for the convex function $\phi(y) = \int_0^{T+y} h_X(x)dx$ (Lemma 7). The result follows directly. ■

Proof of Theorem 5. The first inequality in (3.13) follows from the fact that $\Omega^{\mathcal{A}}(T^*) \leq \Omega^{\mathcal{A}}(\widetilde{T}^* - \mu_Y)$ if $\widetilde{T}^* - \mu_Y$ is feasible to optimization problem (3.1). The second equality is based on Theorem 1 and $\widetilde{\Omega}^{\mathcal{A}}(\widetilde{T}^*) = c_m h_X(\widetilde{T}^*)$. The third inequality is obtained by applying the Edmundson-Madansky upper bound for the convex function $\phi(y) = \int_0^{T+y} h_X(x)dx$ (Lemma 7). The result follows directly. ■

Proof of Proposition 10. If $\alpha \geq 2$, multiply both sides of (3.16) by the standard deviation of Y and $Y^{\alpha-1}$ (denoted by σ_Y and $\sigma_{Y^{\alpha-1}}$, respectively), we have

$$(\alpha - 1)\mu_Y \mathbb{E}[Y^{\alpha-1}] \geq \mathrm{Cov}(Y, Y^{\alpha-1}), \tag{A.2.9}$$

where $\mathrm{Cov}(Y, Y^{\alpha-1})$ is the covariance of random variables Y and $Y^{\alpha-1}$. If we add $\mu_Y \mathbb{E}[Y^{\alpha-1}]$ to both sides of (A.2.9), we have

$$\alpha \mu_Y \mathbb{E}[Y^{\alpha-1}] \geq \mathbb{E}[Y^{\alpha}], \tag{A.2.10}$$

which is equivalent to (3.11) if $X \sim \mathrm{Weibull}(\alpha, \beta)$. ■

Proof of Proposition 11. By setting $\alpha = 2$ in equation (3.15), we obtain

$$\int_a^b \left(2(T^* + y)(T^* + \mu_Y) - (T^* + y)^2\right) f_Y(y)dy = k_1\beta^2. \tag{A.2.11}$$

The terms of (A.2.11) can be expanded as follows

$$\begin{aligned}\int_a^b \left(2(T^* + y)(T^* + \mu_Y) - (T^* + y)^2\right) f_Y(y)dy &= T^{*2} + 2\mu_Y T^* - \int_a^b (y^2 - 2\mu_Y y) f_Y(y)dy \\ &= (T^* + \mu_Y)^2 - \int_a^b (y - \mu_Y)^2 f_Y(y)dy \\ &= (T^* + \mu_Y)^2 - \sigma_Y^2 = k_1\beta^2. \qquad \text{(A.2.12)}\end{aligned}$$

The results follows from (A.2.12). ■

Proof of Proposition 12. Because $f_Y(y)$ is symmetric w.r.t. $y = 0$, $\mu_Y = 0$. Expanding the terms on the left hand side of (3.15), we have

$$\begin{aligned} & \textstyle\int_a^b \left(\alpha(T^* + y)^{\alpha-1}T^* - (T^* + y)^\alpha\right) f_Y(y)dy \\ = & \textstyle\int_a^b \Bigg(\alpha T^* \left((T^*)^{\alpha-1} + \sum_{i=1}^{\alpha-1} \binom{\alpha-1}{i} y^i (T^*)^{\alpha-1-i}\right) \\ & \textstyle -\left((T^*)^\alpha + \sum_{i=1}^{\alpha-1} \binom{\alpha}{i} y^i (T^*)^{\alpha-i} + y^\alpha\right)\Bigg) f_Y(y)dy \\ = & \textstyle(\alpha - 1)(T^*)^\alpha + Q(T) + W(T) - \int_a^b y^\alpha f_Y(y)dy,\end{aligned}$$

where

$$\begin{aligned} Q(T) &= \sum_{k=1}^{\lceil\frac{\alpha-1}{2}\rceil} \left(\alpha\binom{\alpha-1}{2k-1} - \binom{\alpha}{2k-1}\right)(T^*)^{\alpha-(2k-1)} \int_a^b y^{2k-1} f_Y(y)dy \\ W(T) &= \sum_{k=1}^{\lfloor\frac{\alpha-1}{2}\rfloor} \left(\alpha\binom{\alpha-1}{2k} - \binom{\alpha}{2k}\right)(T^*)^{\alpha-(2k)} \int_a^b y^{2k} f_Y(y)dy.\end{aligned}$$

The term $Q(T)$ can be discarded as it is equal to 0 by the symmetry of $f_Y(y)$ w.r.t. $y = 0$. Note that $W(T) \geq 0$, because $\alpha\binom{\alpha-1}{p} - \binom{\alpha}{p} \geq 0 \ \forall p \in \mathbb{Z}_+$. In the two cases to be discussed next, we show that $W(T) - \int_a^b y^\alpha f_Y(y)dy \geq 0$.

Case 1: α is odd. We have $\int_a^b y^\alpha f_Y(y)dy = 0$.

Case 2: α is even. The symmetry of $f_Y(y)$ w.r.t. $y = 0$ also implies that $b = -a > 0$. Thus, $T^* > -a = b$. Then the following inequalities hold:

$$\begin{aligned}(T^*)^{\alpha-(2k)} \int_a^b y^{2k} f_Y(y)dy &\geq (b)^{\alpha-(2k)} \int_a^b y^{2k} f_Y(y)dy \\ &\geq \int_a^b |y|^{\alpha-(2k)} \cdot |y|^{2k} f_Y(y)dy = \int_a^b y^\alpha f_Y(y)dy.\end{aligned}$$

Thus, for both cases, we have $W(T) - \int_a^b y^\alpha f_Y(y)dy \geq 0$. From (3.15), we obtain $T^* \leq (\frac{k_1}{\alpha-1})^{\frac{1}{\alpha}} \beta = \widetilde{T}^*$.

■

Proof of Proposition 13. First, we show $n(T)$ is monotone increasing if $1/\int_a^b \bar{F}_X(T + y)dF_Y(y)$ is logarithmically convex. Taking the first derivative of $n(T)$, we have

$$\begin{aligned}n'(T) &= \mathcal{H}'(T)\mathcal{G}(T) + \mathcal{H}(T)\mathcal{G}'(T) - \mathcal{F}_X'(T) \\ &= \mathcal{H}'(T)\mathcal{G}(T) + \mathcal{H}(T) \int_a^b \bar{F}_X(T + y)dF_Y(y) - \int_a^b f_X(T + y)dF_Y(y) \\ &= \mathcal{H}'(T)\mathcal{G}(T) + \frac{\int_a^b f_X(T + y)dF_Y(y)}{\int_a^b \bar{F}_X(T + y)dF_Y(y)} \int_a^b \bar{F}_X(T + y)dF_Y(y) - \int_a^b f_X(T + y)dF_Y(y) \\ &= \mathcal{H}'(T)\mathcal{G}(T) + \int_a^b f_X(T + y)dF_Y(y) - \int_a^b f_X(T + y)dF_Y(y) \\ &= \mathcal{H}'(T)\mathcal{G}(T).\end{aligned}$$

Clearly, $\mathcal{G}(T) > 0$ for any finite T by **A2**. Thus, if $\mathcal{H}'(T)$ is positive, then $n(T)$ is strictly increasing.

Note that if $1/\int_a^b \bar{F}_X(T + y)dF_Y(y)$ is strictly logarithmically convex, then

$$\frac{d^2}{dT^2}\left(-\log \int_a^b \bar{F}_X(T + y)dF_Y(y)\right) = \frac{d}{dT}\left(-\frac{\int_a^b -f_X(T + y)dF_Y(y)}{\int_a^b \bar{F}_X(T + y)dF_Y(y)}\right) = \mathcal{H}'(T) > 0.$$

Therefore, $\mathcal{H}'(T)$ is positive, and $n(T)$ is strictly increasing. Next, it is easy to verify that

$$\lim_{T\to+\infty} \mathcal{H}(T) = \frac{f_X(T)}{\bar{F}_X(T)} = h_X(T) \to +\infty,$$

$$\lim_{T\to+\infty} \mathcal{G}(T) = \mu_X, \quad \text{and} \lim_{T\to+\infty} \mathcal{F}_X(T) = 1.$$

If (3.22) holds, and $n(T)$ is strictly increasing with $\lim_{T\to+\infty} n(T) \to +\infty$, then there exists a unique solution to (3.21). Rearranging the terms of (3.21) yields

$$\mathcal{H}(T^*) = \frac{\mathcal{F}_X(T^*) + 1/(k_2-1)}{\mathcal{G}(T^*)},$$

in which case

$$c_r(1-1/k_2)\mathcal{H}(T^*) = \frac{(c_r-c_p)\mathcal{F}_X(T^*) + c_p}{\mathcal{G}(T^*)} = \frac{\int_a^b \left(c_r F_X(T^*+y) + c_p \bar{F}_X(T^*+y)\right) dF_Y(y)}{\int_a^b \int_0^{T^*+y} \bar{F}_X(x)dxdF_Y(y)} = \Omega^{\mathcal{B}}(T^*).$$

■

Proof of Theorem 6. The first inequality in (3.23) follows from Theorem 1. To prove the second equality, first recall that $\widetilde{\Omega}^{\mathcal{B}}(\widetilde{T}^*) = c_r(1-1/k_2)h_X(\widetilde{T}^*)$ ([16]). If $\widetilde{T}^* - \mu_Y$ is feasible to optimization problem (3.1), then $\Omega^{\mathcal{B}}(T^*) \le \Omega^{\mathcal{B}}(\widetilde{T}^* - \mu_Y)$. Therefore,

$$\frac{\Omega^{\mathcal{B}}(T^*)}{\widetilde{\Omega}^{\mathcal{B}}(\widetilde{T}^*)} \le \frac{\Omega^{\mathcal{B}}(\widetilde{T}^* - \mu_Y)}{c_r(1-1/k_2)h_X(\widetilde{T}^*)}.$$

To obtain the third inequality, first apply the Edmundson-Madansky upper bound for the convex function $\phi(y) = -\int_0^{\widetilde{T}^*-\mu_Y+y} \bar{F}_X(x)dx$ (Lemma 7), i.e.,

$$-\int_a^b \int_0^{\widetilde{T}^*-\mu_Y+y} \bar{F}_X(x)dxdF_Y(y) \le -N(\widetilde{T}^* - \mu_Y) < 0,$$

where

$$N(T) = \frac{\int_{T+a}^{T+b} \bar{F}_X(x)dx}{b-a}(\mu_Y - a) + \int_0^{T+a} \bar{F}_X(x)dx.$$

Therefore,

$$\begin{aligned}\frac{\Omega^{\mathcal{B}}(T^*)}{\widetilde{\Omega}^{\mathcal{B}}(\widetilde{T}^*)} &\le \frac{c_r + (c_p - c_r)\int_a^b \bar{F}_X(\widetilde{T}^* - \mu_Y + y)dF_Y(y)}{c_r(1-1/k_2)h_X(\widetilde{T}^*)\int_a^b \int_0^{\widetilde{T}^*-\mu_Y+y} \bar{F}_X(x)dxdF_Y(y)} \\ &\le \frac{c_r + (c_p - c_r)\int_a^b \bar{F}_X(\widetilde{T}^* - \mu_Y + y)dF_Y(y)}{c_r(1-1/k_2)h_X(\widetilde{T}^*)N(\widetilde{T}^* - \mu_Y)}.\end{aligned}$$

Observe that if $c_p - c_r < 0$, then a lower bound on $\int_a^b \bar{F}_X(\widetilde{T}^* - \mu_Y + y) f_Y(y) dy$ provides an upper bound on $(c_p - c_r) \int_a^b \bar{F}_X(\widetilde{T}^* - \mu_Y + y) f_Y(y) dy$. Note that $\bar{F}_X(x)$ is monotone decreasing, thus,

$$\int_a^b \bar{F}_X(\widetilde{T}^* - \mu_Y + y) f_Y(y) dy \geq \bar{F}_X(\widetilde{T}^* - \mu_Y + b), \tag{A.2.13}$$

which completes the proof of the third equality.

However, the bound by applying (A.2.13) is loose in general. We can improve it by exploiting the local concavity of $\bar{F}_X(\widetilde{T}^* - \mu_Y + y)$ w.r.t. $y \in [a, b]$. If $X \sim$ Weibull (α, β), then

$$\begin{aligned}
&\bar{F}_X(t) = e^{-(\frac{t}{\beta})^\alpha}, \\
&\bar{F}'_X(t) = -f_X(t) = -\frac{\alpha t^{\alpha-1}}{\beta^\alpha} e^{-(\frac{t}{\beta})^\alpha}, \text{and} \\
&\bar{F}''_X(t) = -f'_X(t) = -\frac{\alpha t^{\alpha-2}}{\beta^\alpha} e^{-(\frac{t}{\beta})^\alpha} (\alpha - 1 - \frac{\alpha t^\alpha}{\beta^\alpha}).
\end{aligned}$$

For $t^0 = (\frac{\alpha-1}{\alpha})^{\frac{1}{\alpha}} \beta$, $\bar{F}''_X(t^0) = 0$, then $\bar{F}''_X(t) \leq 0$ if $t \leq t^0$, and $\bar{F}''_X(t) \geq 0$ if $t \geq t^0$, i.e., $\bar{F}_X(t)$ is concave for $t \in (0, t^0]$ and convex in $[t^0, \infty)$. Therefore, if $\widetilde{T}^* - \mu_Y + b \leq t^0$, then $\bar{F}_X(\widetilde{T}^* - \mu_Y + y)$ is concave w.r.t. $y \in [a, b]$. We can apply the Edmundson-Madansky upper bound for the convex function $\phi(y) = -\bar{F}_X(\widetilde{T}^* - \mu_Y + y)$, see Lemma 7,

$$\begin{aligned}
&\int_a^b \bar{F}_X(\widetilde{T}^* - \mu_Y + y) f_Y(y) dy \geq \hat{N}(\widetilde{T}^* - \mu_Y), \text{ where} \\
&\hat{N}(T) = \frac{\bar{F}_X(T + b) - \bar{F}_X(T + a)}{b - a} (\mu_Y - a) + \bar{F}_X(T + a).
\end{aligned}$$

The result follows directly. ∎

Proof of Theorem 7. The first inequality in (3.25) follows from the fact that $\Omega^{\mathcal{B}}(T^*) \leq \Omega^{\mathcal{B}}(\widetilde{T}^* - \mu_Y)$ if $\widetilde{T}^* - \mu_Y$ is feasible to the main optimization problem (3.1). The second equality is based on Theorem 1 and $\widetilde{\Omega}^{\mathcal{B}}(\widetilde{T}^*) = c_m h_X(\widetilde{T}^*)$. The third inequality is obtained by applying the Edmundson-Madansky upper bound for the convex function $\phi(y) = -\int_0^{T+y} \bar{F}_X(x) dx$ (Lemma 7) together with the fact that $\bar{F}_X(x)$ is monotone decreasing. To improve the upper bound $U^{\mathcal{B}1}(\widetilde{T}^*)$, a similar approach can be used as in proof of Theorem 6.

∎

A.3 PROOFS FOR CHAPTER 4

First, we introduce an important result for some of the proofs in this chapter.

Lemma 8. *(Puterman 1994) Let $\{x_j\}$, $\{x'_j\}$ be real-valued non-negative sequences satisfying*

$$\sum_{j=k}^{\infty} x_j \geq \sum_{j=k}^{\infty} x'_j \tag{A.3.1}$$

for all k, with equality holding in (A.3.1) for $k = 0$. Suppose $v_{j+1} \geq v_j$ for $j = 0, 1, \ldots,$ then

$$\sum_{j=0}^{\infty} v_j x_j \geq \sum_{j=0}^{\infty} v_j x'_j, \tag{A.3.2}$$

where limits in (A.3.2) exist but may be infinite.

Proof of Lemma 2. Assume $\delta_1 \geq \delta_2$. By Assumption **A4**, we have

$$\sum_{\delta'=k}^{D} P^{(d)}(\delta'|\delta_1) \geq \sum_{\delta'=k}^{D} P^{(d)}(\delta'|\delta_2)$$

for all $k \in \Delta$ and $d \in \mathbb{Z}_+$. Recall that by Assumption **A3**, $q^\theta(\delta)$ is nonincreasing in δ. Then, by Lemma 8, it follows that

$$\sum_{\delta'=1}^{D} P^{(j)}(\delta'|\delta_1)(-q^\theta(\delta')) \geq \sum_{\delta'=1}^{D} P^{(j)}(\delta'|\delta_2)(-q^\theta(\delta')), \forall j \in \{1, 2, \ldots, d^\theta\}.$$

Therefore, $r(\theta, \delta_1) \leq r(\theta, \delta_2)$. ■

Proof of Theorem 8. Without loss of generality, assume that the optimal sequence to prescribe is $\langle AB \rangle$. By Lemma 2, $f_1^{\langle AB \rangle}(\delta)$ is nonincreasing in δ.

Applying Lemma 8 and mimicking the proof of Lemma 2, we can also prove that $f_2^{\langle AB \rangle}(\delta)$ is nonincreasing in δ. Similarly, we can show that $f_3^{\langle AB \rangle}(\delta)$ is nonincreasing in δ by exploiting Assumption **A5**.

Since $V(\delta)$ is a linear combination of $f_1^{\langle AB\rangle}(\delta)$, $f_2^{\langle AB\rangle}(\delta)$ and $f_3^{\langle AB\rangle}(\delta)$ with nonnegative coefficients, the nonincreasing property of $V(\delta)$ in δ follows.

■

Proof of Lemma 3. First, we prove

$$f_1^{\langle AB\rangle}(\delta_{t_0}) \geq f_1^{\langle BA\rangle}(\delta_{t_0}).$$

Based on Assumption **A3** and the fact that $q^A(\delta) \geq q^B(\delta)$ for all $\delta \in \Delta$, and $Q_E^A(t) \geq Q_E^B(t)$ for all t, we observe

$$Q_E^A(t_0 + d^A) - Q_E^A(t_0 + d^B) \geq \max_\delta q^A(\delta) \cdot (d^B - d^A) \geq \max_\delta q^B(\delta) \cdot (d^B - d^A). \qquad \text{(A.3.3)}$$

As $d^B \geq d^A$ and $q^A(\delta) \geq q^B(\delta)$ for all $\delta \in \Delta$, from (4.1) we have

$$\begin{aligned}
r(B, \delta_{t_0}) =& \sum_{j=1}^{d^B} \sum_{\delta'=1}^{D} P^{(j)}(\delta'|\delta_{t_0}) q^B(\delta') && \text{(A.3.4)}\\
=& \sum_{j=1}^{d^A} \sum_{\delta'=1}^{D} P^{(j)}(\delta'|\delta_{t_0}) q^B(\delta') + \sum_{j=d^A+1}^{d^B} \sum_{\delta'=1}^{D} P^{(j)}(\delta'|\delta_{t_0}) q^B(\delta')\\
\leq& \sum_{j=1}^{d^A} \sum_{\delta'=1}^{D} P^{(j)}(\delta'|\delta_{t_0}) q^A(\delta') + \max_\delta q^B(\delta) \cdot \sum_{j=d^A+1}^{d^B} \sum_{\delta'=1}^{D} P^{(j)}(\delta'|\delta_{t_0})\\
=& \sum_{j=1}^{d^A} \sum_{\delta'=1}^{D} P^{(j)}(\delta'|\delta_{t_0}) q^A(\delta') + \max_\delta q^B(\delta) \cdot (d^B - d^A)\\
\leq& r(A, \delta_{t_0}) + Q_E^A(t_0 + d^A) - Q_E^A(t_0 + d^B), && \text{(A.3.5)}
\end{aligned}$$

where the last inequality follows from (A.3.3). As $Q_E^A(t_0 + d^B) \geq Q_E^B(t_0 + d^B)$, then (A.3.5) implies that

$$\begin{aligned}
r(A, \delta_0) + Q_E^A(t_0 + d^A) =& r(A, \delta_0) + Q_E^A(t_0 + d^A) - Q_E^A(t_0 + d^B) + Q_E^A(t_0 + d^B)\\
\geq& r(B, \delta_{t_0}) + Q_E^B(t_0 + d^B).
\end{aligned}$$

Next, we show

$$f_1^{\langle BA\rangle}(\delta_{t_0}) \geq f_2^{\langle AB\rangle}(\delta_{t_0}).$$

From Lemma 2 and **A4**, we observe that

$$r(B,\delta_{t_0}) \geq \sum_{\delta=1}^{D} P^{(d^A)}(\delta|\delta_{t_0}) \cdot r(B,\delta). \tag{A.3.6}$$

Furthermore, we have

$$\begin{aligned} & Q_E^B(t_0+d^B) - Q_E^B(t_0+d^A+d^B) - r(A,\delta_{t_0}) \\ \geq & \max_{\delta} q^A(\delta)\cdot d^A - r(A,\delta_{t_0}) \geq 0, \end{aligned} \tag{A.3.7}$$

where the second inequality is by **A3**. Therefore,

$$\begin{aligned} & f_1^{\langle BA\rangle}(\delta_{t_0}) - f_2^{\langle AB\rangle}(\delta_{t_0}) \\ =& \Big(r(B,\delta_{t_0}) - \sum_{\delta=1}^{D} P^{(d^A)}(\delta|\delta_{t_0})\cdot r(B,\delta)\Big) + \Big(Q_E^B(t_0+d^B) - Q_E^B(t_0+d^A+d^B) - r(A,\delta_{t_0})\Big) \geq 0, \end{aligned}$$

which completes the proof. ■

Proof of Lemma 4. Define

$$L(u) = \sum_{j=1}^{d^A}\sum_{\delta'=1}^{D} P^{(j)}(\delta'|\delta_{t_0})q^A(\delta') + \sum_{\delta'=1}^{D} P^{(d^A)}(\delta'|\delta_{t_0})\sum_{j=1}^{u}\sum_{\delta''=1}^{D} P^{(j)}(\delta''|\delta')q^B(\delta''). \tag{A.3.8}$$

Similarly, define

$$R(u) = \sum_{j=1}^{u}\sum_{\delta'=1}^{D} P^{(j)}(\delta'|\delta_{t_0})q^B(\delta') + \sum_{\delta'=1}^{D} P^{(u)}(\delta'|\delta_{t_0})\sum_{j=1}^{d^A}\sum_{\delta''=1}^{D} P^{(j)}(\delta''|\delta')q^A(\delta''). \tag{A.3.9}$$

For any $k, j \in \mathbb{Z}_+$ and $\delta'' \in \Delta$, we have (e.g., see [75])

$$\sum_{\delta'=1}^{D} P^{(k)}(\delta'|\delta_{t_0})P^{(j)}(\delta''|\delta') = P^{(k+j)}(\delta''|\delta_{t_0}),$$

which implies that

$$\sum_{\delta'=1}^{D} P^{(d^A)}(\delta'|\delta_{t_0})\sum_{j=1}^{u}\sum_{\delta''=1}^{D} P^{(j)}(\delta''|\delta')q^B(\delta'') = \sum_{j=d^A+1}^{d^A+u}\sum_{\delta=1}^{D} P^{(j)}(\delta|\delta_{t_0})q^B(\delta),$$

and

$$\sum_{\delta'=1}^{D} P^{(u)}(\delta'|\delta_{t_0}) \sum_{j=1}^{d^A} \sum_{\delta''=1}^{D} P^{(j)}(\delta''|\delta') q^A(\delta'') = \sum_{j=u+1}^{d^A+u} \sum_{\delta=1}^{D} P^{(j)}(\delta|\delta_{t_0}) q^A(\delta),$$

Therefore,

$$L(d^A) - R(d^A) = \sum_{j=1}^{d^A} \sum_{\delta=1}^{D} P^{(j)}(\delta|\delta_{t_0})(q^A(\delta) - q^B(\delta)) - \sum_{j=d^A+1}^{d^A+d^A} \sum_{\delta=1}^{D} P^{(j)}(\delta|\delta_{t_0})(q^A(\delta) - q^B(\delta)) \geq 0, \tag{A.3.10}$$

where the inequality holds by Assumption **A4**. Because both d^A and d^B are integers, and $d^B \geq d^A$, we can now increase u from d_A to d_B and check if a similar inequality still holds as in (A.3.10). If $u = d^A + 1$, then the increase in (A.3.8) is

$$\begin{aligned} L(d^A+1) - L(d^A) =& \sum_{\delta=1}^{D} \sum_{j=d^A+1}^{d^A+d^A+1} P^{(j)}(\delta|\delta_{t_0}) q^B(\delta) - \sum_{\delta=1}^{D} \sum_{j=d^A+1}^{d^A+d^A} P^{(j)}(\delta|\delta_{t_0}) q^B(\delta) \\ =& \sum_{\delta=1}^{D} P^{(d^A+d^A+1)}(\delta|\delta_{t_0}) q^B(\delta), \end{aligned}$$

and the increase in (A.3.9) is

$$\begin{aligned} & R(d^A+1) - R(d^A) \\ =& \sum_{\delta=1}^{D} P^{(d^A+1)}(\delta|\delta_{t_0}) q^B(\delta) + \sum_{j=d^A+2}^{d^A+d^A+1} \sum_{\delta=1}^{D} P^{(j)}(\delta|\delta_{t_0}) q^A(\delta) - \sum_{j=d^A+1}^{d^A+d^A} \sum_{\delta=1}^{D} P^{(j)}(\delta|\delta_{t_0}) q^A(\delta) \\ =& \sum_{\delta=1}^{D} P^{(d^A+1)}(\delta|\delta_{t_0}) q^B(\delta) + \sum_{\delta=1}^{D} P^{(d^A+d^A+1)}(\delta|\delta_{t_0}) q^A(\delta) - \sum_{\delta=1}^{D} P^{(d^A+1)}(\delta|\delta_{t_0}) q^A(\delta). \end{aligned}$$

Then by Lemma 8

$$\begin{aligned} & L(d^A+1) - L(d^A) - \Big(R(d^A+1) - R(d^A)\Big) \\ =& \sum_{\delta=1}^{D} P^{(d^A+d^A+1)}(\delta|\delta_{t_0})(q^B(\delta) - q^A(\delta)) - \sum_{\delta=1}^{D} P^{(d^A+1)}(\delta|\delta_{t_0})(q^B(\delta) - q^A(\delta)) \geq 0. \end{aligned} \tag{A.3.11}$$

Therefore, we have

$$
\begin{aligned}
&L(d^A+1) - R(d^A+1)\\
=&\Big(L(d^A+1) - L(d^A) + L(d^A)\Big) - \Big(R(d^A+1) - R(d^A) + R(d^A)\Big)\\
=&\Big(L(d^A+1) - L(d^A) - R(d^A+1) + R(d^A)\Big) + \Big(L(d^A) - R(d^A)\Big) \geq 0.
\end{aligned}
$$

The last inequality is by (A.3.10) and (A.3.11). It implies that increasing one unit of u in both (A.3.8) and (A.3.9) does not change the inequality in (A.3.10). We can apply the same approach to u until it reaches d^B, and the same result holds. ■

Proof of Lemma 5. By taking the difference between the left-hand and right-hand sides of (4.8), we have

$$
\begin{aligned}
&r(A,\delta_{t_0}) + Q_E^A(t_0+d^A) - \Big(r(B,\delta_{t_0}) + \sum_{\delta=1}^{D} P^{(d^B)}(\delta|\delta_{t_0})\cdot r(A,\delta) + Q_E^A(t_0+d^A+d^B)\Big)\\
=&\Big(r(A,\delta_{t_0}) - \sum_{\delta=1}^{D} P^{(d^B)}(\delta|\delta_{t_0})\cdot r(A,\delta)\Big) + \Big(Q_E^A(t_0+d^A) - Q_E^A(t_0+d^A+d^B) - r(B,\delta_{t_0})\Big).
\end{aligned}
\tag{A.3.12}
$$

Observe that $r(A,\delta_{t_0}) \geq \sum_{\delta=1}^{D} P^{(d^B)}(\delta|\delta_{t_0})\cdot r(A,\delta)$, because P is upper triangular by Assumption **A4**. In addition, by Assumption **A3**

$$Q_E^A(t_0+d^A) - Q_E^A(t_0+d^A+d^B) \geq \max_{\delta} q^A(\delta)d^B \geq \max_{\delta} q^B(\delta)d^B \geq r(B,\delta_{t_0}).$$

Therefore, (A.3.12) is nonnegative, which completes the proof. ■

Before proving Theorem 9, we first show the following technical result.

Lemma 9. *If $X+x \geq Y+y$, $X \geq Y \geq y$, $0 < \rho^B \leq \rho^A \leq 1$, and $\rho = \rho^A + \rho^B - \rho^A\rho^B$, then*

$$\rho^B Y + (\rho - \rho^B)y \leq \rho^A X + (\rho - \rho^A)x. \tag{A.3.13}$$

Proof. From $Y \geq y$ and $\rho^A \geq \rho^B$, we have

$$(\rho^A - \rho^B)Y \geq (\rho^A - \rho^B)y,$$

which can be re-written as

$$\rho^B Y + (\rho - \rho^B)y \leq \rho^A Y + (\rho - \rho^A)y,$$

Next, we show that

$$\rho^A Y + (\rho - \rho^A)y \leq \rho^A X + (\rho - \rho^A)x, \tag{A.3.14}$$

by considering the following two cases:

Case 1: $x \geq y$ This case is trivial because $X \geq Y$ and $x \geq y$.

Case 2: $x \leq y$ In this case, we have

$$\rho^A(X - Y) \geq \rho^A(y - x) \geq (\rho - \rho^A)(y - x). \tag{A.3.15}$$

The first inequality in (A.3.15) holds because $X + x \geq Y + y$ and ρ^A is positive. The second inequality in (A.3.15) holds because $\rho - \rho^A = \rho^B - \rho^A\rho^B = \rho^B(1 - \rho^A) \leq \rho^B \leq \rho^A$. Note (A.3.15) is equivalent to (A.3.14), which completes the proof of Lemma 9. ■

Proof of Theorem 9. First, from Assumption **A3** and similar to (A.3.3) we observe that

$$\begin{aligned}
&r(A, \delta_{t_0}) + \Big(Q_E^B(t_0 + d^A) - Q_E^B(t_0 + d^B)\Big) \\
\geq& r(A, \delta_{t_0}) + \max_{\delta} q^B(\delta)(d^B - d^A) \\
\geq& \sum_{j=1}^{d^A} \sum_{\delta'=1}^{D} P^{(j)}(\delta'|\delta) q^B(\delta') + \max_{\delta} q^B(\delta)(d^B - d^A) \\
\geq& r(B, \delta_{t_0}).
\end{aligned} \tag{A.3.16}$$

Next, we want to show that

$$r(A, \delta_{t_0}) + Q_E^A(t_0 + d^A) + Q_E^B(t_0 + d^A + d^B) \geq r(B, \delta_{t_0}) + Q_E^B(t_0 + d^B) + Q_E^A(t_0 + d^A + d^B). \tag{A.3.17}$$

Note that

$$
\begin{aligned}
& r(A,\delta_{t_0}) + Q_E^A(t_0+d^A) - Q_E^A(t_0+d^A+d^B) \\
=& r(A,\delta_{t_0}) + \Big(Q_E^A(t_0+d^A) - Q_E^A(t_0+d^A+d^B)\Big) + \Big(Q_E^B(t_0+d^B) - Q_E^B(t_0+d^A+d^B)\Big) \\
& - \Big(Q_E^B(t_0+d^B) - Q_E^B(t_0+d^A+d^B)\Big) \\
\geq& r(A,\delta_{t_0}) + \Big(Q_E^B(t_0+d^A) - Q_E^B(t_0+d^A+d^B)\Big) + \Big(Q_E^B(t_0+d^B) - Q_E^B(t_0+d^A+d^B)\Big) \\
& - \Big(Q_E^B(t_0+d^B) - Q_E^B(t_0+d^A+d^B)\Big) \\
=& r(A,\delta_{t_0}) + \Big(Q_E^B(t_0+d^A) - Q_E^B(t_0+d^B)\Big) + \Big(Q_E^B(t_0+d^B) - Q_E^B(t_0+d^A+d^B)\Big) \\
\geq& r(B,\delta_{t_0}) + Q_E^B(t_0+d^B) - Q_E^B(t_0+d^A+d^B), \qquad \text{(A.3.18)}
\end{aligned}
$$

where the first inequality holds by the assumption that $Q_E^A(t) - Q_E^A(t+k) \geq Q_E^B(t) - Q_E^B(t+k)$ for all t, k, and the second inequality holds by (A.3.16). Next, define

$$
\begin{aligned}
X =& f_1^{\langle AB \rangle}(\delta_{t_0}) \text{ and } x = f_2^{\langle AB \rangle}(\delta_{t_0}), \\
Y =& \max\Big(f_1^{\langle BA \rangle}(\delta_{t_0}), f_2^{\langle BA \rangle}(\delta_{t_0})\Big) \text{ and } y = \min\Big(f_1^{\langle BA \rangle}(\delta_{t_0}), f_2^{\langle BA \rangle}(\delta_{t_0})\Big). \qquad \text{(A.3.19)}
\end{aligned}
$$

We can verify that $X \geq x$ and $X \geq Y \geq y$ by Lemmas 3-4. Next, we show that $X+x \geq Y+y$ based on definitions in A.3.19.

$$
\begin{aligned}
& X + x - (Y+y) \\
=& r(A,\delta_{t_0}) + Q_E^A(t_0+d^A) + r(A,\delta_{t_0}) + \sum_{\delta=1}^{D} P^{(d^A)}(\delta|\delta_{t_0}) \cdot r(B,\delta) + Q_E^B(t_0+d^A+d^B) \\
& - \Big(r(B,\delta_{t_0}) + Q_E^B(t_0+d^B) + r(B,\delta_{t_0}) + \sum_{\delta=1}^{D} P^{(d^B)}(\delta|\delta_{t_0}) \cdot r(A,\delta) + Q_E^A(t_0+d^A+d^B)\Big) \\
=& \Big(r(A,\delta_{t_0}) + Q_E^A(t_0+d^A) + Q_E^B(t_0+d^A+d^B) - r(B,\delta_{t_0}) - Q_E^B(t_0+d^B) - Q_E^A(t_0+d^A+d^B)\Big) \\
& + \Big(r(A,\delta_{t_0}) + \sum_{\delta=1}^{D} P^{(d^A)}(\delta|\delta_{t_0}) \cdot r(B,\delta) - r(B,\delta_{t_0}) - \sum_{\delta=1}^{D} P^{(d^B)}(\delta|\delta_{t_0}) \cdot r(A,\delta)\Big) \geq 0,
\end{aligned}
$$

where the last inequality is by (A.3.18) and Lemma 4. Therefore, we have

$$
\rho^A f_1^{\langle AB \rangle}(\delta_{t_0}) + (1-\rho^A)\rho^B f_2^{\langle AB \rangle}(\delta_{t_0}) \geq \rho^B f_1^{\langle BA \rangle}(\delta_{t_0}) + (1-\rho^B)\rho^A f_2^{\langle BA \rangle}(\delta_{t_0}).
$$

From Lemma 4, we have

$$\begin{aligned}
f_3^{\langle AB\rangle}(\delta_{t_0}) =& r(A,\delta_{t_0}) + \sum_{\delta=1}^{D} P^{(d^A)}(\delta|\delta_{t_0})\cdot r(B,\delta) + \sum_{\delta=1}^{D} P^{(d^A+d^B)}(\delta|\delta_{t_0}) Q^{\theta P}(\delta, t_0+d^A+d^B) \\
\geq& r(B,\delta_{t_0}) + \sum_{\delta=1}^{D} P^{(d^B)}(\delta|\delta_{t_0})\cdot r(A,\delta) + \sum_{\delta=1}^{D} P^{(d^A+d^B)}(\delta|\delta_{t_0}) Q^{\theta P}(\delta, t_0+d^A+d^B) \\
=& f_3^{\langle BA\rangle}(\delta_{t_0}).
\end{aligned}$$

Based on (4.5), we know it is optimal to try treatment A first. ■

Proof of Theorem 10. Given $d^A \leq d^B$, $q^A(\delta) \geq q^B(\delta)$ for all $\delta \in \Delta$, and $Q_E^A(t) \geq Q_E^B(t)$ for all t, and $q^A(\delta) - q^B(\delta)$ is nonincreasing in δ. In proof of Theorem 9, we observe

$$f_3^{\langle AB\rangle}(\delta_{t_0}) \geq f_3^{\langle BA\rangle}(\delta_{t_0}).$$

Therefore, from (4.5), we observe that a sufficient condition under which it is optimal to prescribe treatment A first is

$$\rho^A f_1^{\langle AB\rangle}(\delta_{t_0}) + (1-\rho^A)\rho^B f_2^{\langle AB\rangle}(\delta_{t_0}) \geq \rho^B f_1^{\langle BA\rangle}(\delta_{t_0}) + (1-\rho^B)\rho^A f_2^{\langle BA\rangle}(\delta_{t_0}),$$

which can be re-written as

$$\rho^A \geq \bar{\rho} = \frac{\rho^B\cdot(f_1^{\langle BA\rangle}(\delta_{t_0}) - f_2^{\langle AB\rangle}(\delta_{t_0}))}{f_1^{\langle AB\rangle}(\delta_{t_0}) - \rho^B f_2^{\langle AB\rangle}(\delta_{t_0}) - (1-\rho^B) f_2^{\langle BA\rangle}(\delta_{t_0})}. \tag{A.3.20}$$

Now, let us verify that $\bar{\rho} \leq \rho^B$ (otherwise this result is equivalent to Theorem 9). The difference between $f_1^{\langle BA\rangle}(\delta_{t_0}) - f_2^{\langle AB\rangle}(\delta_{t_0})$ and $f_1^{\langle AB\rangle}(\delta_{t_0}) - \rho^B f_2^{\langle AB\rangle}(\delta_{t_0}) - (1-\rho^B) f_2^{\langle BA\rangle}(\delta_{t_0})$ in (A.3.20) is

$$f_1^{\langle BA\rangle}(\delta_{t_0}) + (1-\rho^B) f_2^{\langle BA\rangle}(\delta_{t_0}) - \left(f_1^{\langle AB\rangle}(\delta_{t_0}) + (1-\rho^B) f_2^{\langle AB\rangle}(\delta_{t_0})\right) \tag{A.3.21}$$

By the proof of Theorem 9, we have

$$f_1^{\langle BA\rangle}(\delta_{t_0}) + f_2^{\langle BA\rangle}(\delta_{t_0}) \leq f_1^{\langle AB\rangle}(\delta_{t_0}) + f_2^{\langle AB\rangle}(\delta_{t_0}),$$

and by multiplying the both sides by $(1-\rho^B)$, we have

$$(1-\rho^B)(f_1^{\langle BA\rangle}(\delta_{t_0})+f_2^{\langle BA\rangle}(\delta_{t_0})) \leq (1-\rho^B)(f_1^{\langle AB\rangle}(\delta_{t_0})+f_2^{\langle AB\rangle}(\delta_{t_0})). \tag{A.3.22}$$

From Lemma 3, we have

$$f_1^{\langle BA\rangle}(\delta_{t_0}) \leq f_1^{\langle AB\rangle}(\delta_{t_0}). \tag{A.3.23}$$

Multiplying (A.3.23) by ρ^B and adding to (A.3.22), we have

$$f_1^{\langle BA\rangle}(\delta_{t_0})+(1-\rho^B)f_2^{\langle BA\rangle}(\delta_{t_0}) \leq f_1^{\langle AB\rangle}(\delta_{t_0})+(1-\rho^B)f_2^{\langle AB\rangle}(\delta_{t_0}),$$

which implies that $\bar{\rho} \leq \rho^B \cdot 1 = \rho^B$. This completes the proof.

■

Proof of Theorem 11. The necessary and sufficient condition to prefer treatment sequence $\langle AB\rangle$ rather than $\langle BA\rangle$ is given by

$$\rho^A Q_E^A(t_0+d^A)+(1-\rho^A)\rho^B Q_E^B(t_0+d^A+d^B)+r(A,\delta_{t_0})+(1-\rho^A)\sum_{\delta'=1}^{D} P^{(d^A)}(\delta'|\delta_{t_0})r(B,\delta')$$
$$\geq \rho^B Q_E^B(t_0+d^B)+(1-\rho^B)\rho^A Q_E^A(t_0+d^A+d^B)+r(B,\delta_{t_0})+(1-\rho^B)\sum_{\delta'=1}^{D} P^{(d^B)}(\delta'|\delta_{t_0})r(A,\delta'). \tag{A.3.24}$$

Since $q^A(\delta)=\mu q^B(\delta)$ for all $\delta\in\Delta$, and $Q_E^A(t)=\mu Q_E^B(t)$ for all t, we can rearrange the terms in (A.3.24), and have

$$\mu \geq \frac{\rho^B Q_E^B(t_0+d^B)-(1-\rho^A)\rho^B Q_E^B(t_1)+r(B,\delta_{t_0})-(1-\rho^A)\sum_{\delta=1}^{D} P^{(d^A)}(\delta|\delta_{t_0})\cdot r(B,\delta)}{\rho^A Q_E^B(t_0+d^A)+\sum_{j=1}^{d^A}\sum_{\delta'=1}^{D} P^{(j)}(\delta'|\delta_{t_0})q^B(\delta')-(1-\rho^B)\left(\rho^A Q_E^B(t_1)+\sum_{j=d^B+1}^{d^A+d^B}\sum_{\delta'=1}^{D} P^{(j)}(\delta'|\delta_{t_0})q^B(\delta')\right)}, \tag{A.3.25}$$

where $t_1=t_0+d^A+d^B$.

Next, we show that $\mu \le 1$ (otherwise, this result is equivalent to Theorem 9). First, we can verify that both numerator and denominator in (A.3.25) are positive. First define

$$
\begin{aligned}
g_1 &= \rho_A Q_E^B(t_0 + d^A) - \rho_B Q_E^B(t_0 + d^B) - (\rho^A - \rho^B) Q_E^B(t_0 + d^A + d^B), \\
g_2 &= -\rho^A \sum_{j=d^A+1}^{d^A+d^B} \sum_{\delta'=1}^{D} P^{(j)}(\delta'|\delta_{t_0}) q^B(\delta') + \rho^B \sum_{j=d^B+1}^{d^A+d^B} \sum_{\delta'=1}^{D} P^{(j)}(\delta'|\delta_{t_0}) q^B(\delta').
\end{aligned}
$$

We can see $g_1 + g_2$ is the difference between the denominator and numerator of (A.3.25). To prove $g_1 + g_2 \ge 0$, note first

$$
Q_E^B(t_0 + d^A) - Q_E^B(t_0 + d^B) \ge \max_{\delta} q^B(\delta)(d^B - d^A) \ge \sum_{j=d^A+1}^{d^B} \sum_{\delta'=1}^{D} P^{(j)}(\delta'|\delta_{t_0}) q^B(\delta'), \quad \text{(A.3.26)}
$$

$$
Q_E^B(t_0 + d^B) - Q_E^B(t_0 + d^A + d^B) \ge \max_{\delta} q^B(\delta) d^A \ge \sum_{j=d^B+1}^{d^A+d^B} \sum_{\delta'=1}^{D} P^{(j)}(\delta'|\delta_{t_0}) q^B(\delta'). \quad \text{(A.3.27)}
$$

By rearranging terms, we have

$$
\begin{aligned}
g_1 =& \rho_A Q_E^B(t_0 + d^A) - \rho_B Q_E^B(t_0 + d^B) - (\rho^A - \rho^B) Q_E^B(t_0 + d^A + d^B) \\
=& \rho_A \Big(Q_E^B(t_0 + d^A) - Q_E^B(t_0 + d^B) \Big) + (\rho^A - \rho^B) \Big(Q_E^B(t_0 + d^B) - Q_E^B(t_0 + d^A + d^B) \Big).
\end{aligned}
$$

We also have

$$
g_2 = -\rho^A \sum_{j=d^A+1}^{d^B} \sum_{\delta'=1}^{D} P^{(j)}(\delta'|\delta_{t_0}) q^B(\delta') - (\rho^A - \rho^B) \sum_{j=d^B+1}^{d^A+d^B} \sum_{\delta'=1}^{D} P^{(j)}(\delta'|\delta_{t_0}) q^B(\delta').
$$

From (A.3.26) and (A.3.27), we can verify that $g_1 + g_2 \ge 0$. Therefore, $\mu \le 1$. ■

Proof of Lemma 6. Define $W^{\langle AB \rangle}(d^A)$ and $W^{\langle BA \rangle}(d^A)$ as the value function if taking treatment sequence $\langle AB \rangle$ and $\langle BA \rangle$, respectively. We first show both value functions are nonincreasing in d^A. Note that

$$
\begin{aligned}
& W^{\langle AB \rangle}(d^A) \\
=& r(A, \delta_{t_0}) + \rho^A Q_E^A(t_0 + d^A) + (1 - \rho^A) \sum_{\delta=1}^{D} P^{(d^A)}(\delta|\delta_{t_0}) \cdot r(B, \delta)
\end{aligned}
$$

$$+ (1-\rho^A)\rho^B Q_E^B(t_0+d^A+d^B) + (1-\rho^A)(1-\rho^B)\sum_{\delta=1}^{D} P^{(d^A+d^B)}(\delta|\delta_{t_0})\cdot Q^{\theta^P}(\delta, t_0+d^A+d^B)$$

$$=\sum_{j=1}^{d^A}\sum_{\delta=1}^{D} P^{(j)}(\delta|\delta_{t_0})q^A(\delta) + \rho^A Q_E^A(t_0+d^A) + (1-\rho^A)\sum_{j=d^A+1}^{d^A+d^B}\sum_{\delta=1}^{D} P^{(j)}(\delta|\delta_{t_0})q^B(\delta)$$

$$+ (1-\rho^A)\rho^B Q_E^B(t_0+d^A+d^B) + (1-\rho^A)(1-\rho^B)\sum_{\delta=1}^{D} P^{(d^A+d^B)}(\delta|\delta_{t_0})\cdot Q^{\theta^P}(\delta, t_0+d^A+d^B).$$

Hence,

$$W^{\langle AB\rangle}(d^A) - W^{\langle AB\rangle}(d^A+1)$$

$$= -\sum_{\delta=1}^{D} P^{(d^A+1)}(\delta|\delta_{t_0})q^A(\delta) + \rho^A\Big(Q_E^A(t_0+d^A) - Q_E^A(t_0+d^A+1)\Big)$$

$$+ (1-\rho^A)\Big(\sum_{\delta=1}^{D} P^{(d^A+1)}(\delta|\delta_{t_0})q^B(\delta) - \sum_{\delta=1}^{D} P^{(d^A+d^B+1)}(\delta|\delta_{t_0})q^B(\delta)\Big)$$

$$+ (1-\rho^A)\rho^B\Big(Q_E^B(t_0+d^A+d^B) - Q_E^B(t_0+d^A+d^B+1)\Big) + (1-\rho^A)(1-\rho^B)$$

$$\sum_{\delta=1}^{D} P^{(d^A+d^B)}(\delta|\delta_{t_0})\cdot Q^{\theta^P}(\delta, t_0+d^A+d^B) - \sum_{\delta=1}^{D} P^{(d^A+d^B+1)}(\delta|\delta_{t_0})\cdot Q^{\theta^P}(\delta, t_0+d^A+d^B+1).$$

Because $\rho^A \geq \rho^B$ and $Q_E^A(t) - Q_E^A(t+k) \geq Q_E^B(t) - Q_E^B(t+k)$, we have

$$\rho^A\Big(Q_E^A(t_0+d^A) - Q_E^A(t_0+d^A+1)\Big) \geq \rho^B\Big(Q_E^B(t_0+d^A) - Q_E^B(t_0+d^A+1)\Big). \quad (A.3.28)$$

From (4.11), **A3** and (A.3.28), we can show that

$$-\sum_{\delta=1}^{D} P^{(d^A+1)}(\delta|\delta_{t_0})q^A(\delta) + \rho^A\Big(Q_E^A(t_0+d^A) - Q_E^A(t_0+d^A+1)\Big) \geq 0. \quad (A.3.29)$$

From **A3-A4** and Lemma 8, we have

$$\sum_{\delta=1}^{D} P^{(d^A+1)}(\delta|\delta_{t_0})q^B(\delta) - \sum_{\delta=1}^{D} P^{(d^A+d^B+1)}(\delta|\delta_{t_0})q^B(\delta) \geq 0. \quad (A.3.30)$$

Based on **A2**, we have $Q_E^B(t_0+d^A+d^B) - Q_E^B(t_0+d^A+d^B+1) \geq 0$. Furthermore, **A4-A5** and Lemma 8 imply that

$$\sum_{\delta=1}^{D} P^{(d^A+d^B)}(\delta|\delta_{t_0})\cdot Q^{\theta^P}(\delta, t_0+d^A+d^B) - \sum_{\delta=1}^{D} P^{(d^A+d^B+1)}(\delta|\delta_{t_0})\cdot Q^{\theta^P}(\delta, t_0+d^A+d^B+1) \geq 0. \quad (A.3.31)$$

We conclude that $W^{\langle AB\rangle}(d^A) - W^{\langle AB\rangle}(d^A+1)$ is nonnegative from (A.3.29)-(A.3.31). Similarly, for treatment sequence $\langle BA\rangle$,

$$\begin{aligned}
&W^{\langle BA\rangle}(d^A)\\
=&r(B,\delta_{t_0}) + \rho^B Q_E^B(t_0+d^B) + (1-\rho^B)\sum_{\delta=1}^{D} P^{(d^B)}(\delta|\delta_{t_0})\cdot r(A,\delta)\\
&+(1-\rho^B)\rho^A Q_E^A(t_0+d^A+d^B) + (1-\rho^A)(1-\rho^B)\sum_{\delta=1}^{D} P^{(d^A+d^B)}(\delta|\delta_{t_0})\cdot Q^{\theta^P}(\delta, t_0+d^A+d^B)\\
=&\sum_{j=1}^{d^B}\sum_{\delta=1}^{D} P^{(j)}(\delta|\delta_{t_0})q^B(\delta) + \rho^B Q_E^B(t_0+d^B) + (1-\rho^B)\sum_{j=d^B+1}^{d^A+d^B}\sum_{\delta=1}^{D} P^{(j)}(\delta|\delta_{t_0})q^A(\delta)\\
&+(1-\rho^B)\rho^A Q_E^A(t_0+d^A+d^B) + (1-\rho^A)(1-\rho^B)\sum_{\delta=1}^{D} P^{(d^A+d^B)}(\delta|\delta_{t_0})\cdot Q^{\theta^P}(\delta, t_0+d^A+d^B).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&W^{\langle BA\rangle}(d^A) - W^{\langle BA\rangle}(d^A+1)\\
=&-(1-\rho^B)\sum_{\delta=1}^{D} P^{(d^A+d^B+1)}(\delta|\delta_{t_0})q^B(\delta)\\
&+(1-\rho^B)\rho^A\Big(Q_E^A(t_0+d^A+d^B) - Q_E^A(t_0+d^A+d^B+1)\Big) + (1-\rho^A)(1-\rho^B)\\
&\sum_{\delta=1}^{D} P^{(d^A+d^B)}(\delta|\delta_{t_0})\cdot Q^{\theta^P}(\delta, t_0+d^A+d^B) - \sum_{\delta=1}^{D} P^{(d^A+d^B+1)}(\delta|\delta_{t_0})\cdot Q^{\theta^P}(\delta, t_0+d^A+d^B+1).
\end{aligned}$$

Similar to the way as we show (A.3.29), we have

$$-\sum_{\delta=1}^{D} P^{(d^A+d^B+1)}(\delta|\delta_{t_0})q^B(\delta) + \rho^A\Big(Q_E^A(t_0+d^A+d^B) - Q_E^A(t_0+d^A+d^B+1)\Big) \geq 0, \tag{A.3.32}$$

because $q^A(\delta) \geq q^B(\delta)$ for any δ. Based on (A.3.31), we conclude that $W^{\langle BA\rangle}(d^A) - W^{\langle BA\rangle}(d^A+1)$ is also nonnegative. Therefore, both value functions are nonincreasing in d^A. Next, we prove that for any d^A,

$$W^{\langle AB\rangle}(d^A) - W^{\langle AB\rangle}(d^A+1) \geq W^{\langle BA\rangle}(d^A) - W^{\langle BA\rangle}(d^A+1).$$

Note that

$$
\begin{aligned}
&W^{\langle AB\rangle}(d^A) - W^{\langle AB\rangle}(d^A+1) - \Big(W^{\langle BA\rangle}(d^A) - W^{\langle BA\rangle}(d^A+1)\Big)\\
=& -\sum_{\delta=1}^{D} P^{(d^A+1)}(\delta|\delta_{t_0})q^A(\delta) + \rho^A\Big(Q_E^A(t_0+d^A) - Q_E^A(t_0+d^A+1)\Big)\\
&+ (1-\rho^A)\Big(\sum_{\delta=1}^{D} P^{(d^A+1)}(\delta|\delta_{t_0})q^B(\delta) - \sum_{\delta=1}^{D} P^{(d^A+d^B+1)}(\delta|\delta_{t_0})q^B(\delta)\Big)\\
&+ (1-\rho^A)\rho^B\Big(Q_E^B(t_0+d^A+d^B) - Q_E^B(t_0+d^A+d^B+1)\Big)\\
&+ (1-\rho^B)\sum_{\delta=1}^{D} P^{(d^A+d^B+1)}(\delta|\delta_{t_0})q^B(\delta)\\
&- (1-\rho^B)\rho^A\Big(Q_E^A(t_0+d^A+d^B) - Q_E^A(t_0+d^A+d^B+1)\Big)\\
\geq& -\sum_{\delta=1}^{D} P^{(d^A+1)}(\delta|\delta_{t_0})q^A(\delta) + \rho^A\Big(Q_E^A(t_0+d^A) - Q_E^A(t_0+d^A+1)\Big)\\
&+ (1-\rho^A)\rho^B\Big(Q_E^B(t_0+d^A+d^B) - Q_E^B(t_0+d^A+d^B+1)\Big)\\
&- (1-\rho^B)\rho^A\Big(Q_E^A(t_0+d^A+d^B) - Q_E^A(t_0+d^A+d^B+1)\Big)\\
\geq& -\sum_{\delta=1}^{D} P^{(d^A+1)}(\delta|\delta_{t_0})q^A(\delta) + \rho^A\rho^B\Big(Q_E^A(t_0+d^A+d^B) - Q_E^A(t_0+d^A+d^B+1)\Big)\\
&+ (1-\rho^A)\rho^B\Big(Q_E^B(t_0+d^A+d^B) - Q_E^B(t_0+d^A+d^B+1)\Big)\\
\geq& -\sum_{\delta=1}^{D} P^{(d^A+1)}(\delta|\delta_{t_0})q^A(\delta) + \rho^B\Big(Q_E^B(t_0+d^A+d^B) - Q_E^B(t_0+d^A+d^B+1)\Big) \geq 0.
\end{aligned}
$$

The first inequality follows by (A.3.30) and the nonnegativity of $q^B(\delta)$. The second inequality is based on (4.12). The third inequality is due to our assumption $Q_E^A(t) - Q_E^A(t+k) \geq Q_E^B(t) - Q_E^B(t+k)$ for all t and $k > 0$ as well as **A3**. Finally, the last inequality is by (4.11).

As a result, if d^A increases by one unit, then the decrease in value function $W^{\langle AB\rangle}(d^A)$ is no smaller than that in $W^{\langle BA\rangle}(d^A)$. Recall Theorem 9 shows that $W^{\langle AB\rangle}(1) \geq W^{\langle BA\rangle}(1)$. Therefore, there exists at most one $\bar{d^A}$, such that we prefer sequence $\langle AB\rangle$ if $d^A \leq \bar{d^A}$, and prefer sequence $\langle BA\rangle$ otherwise. ■

BIBLIOGRAPHY

[1] American Cancer Society recommendations for early breast cancer detection in women without breast symptoms. Available: http://www.cancer.org/cancer/breastcancer/moreinformation/breastcancerearlydetection/breast-cancer-early-detection-acs-recs. Accessed: September 2016.

[2] VHA Office of Telehealth Services, Care Coordination/Home Telehealth. Available: http://www.telehealth.va.gov/ccht/index.asp. Accessed: January 2014.

[3] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: A survey. *Computer Networks*, 38(4):393–422, 2002.

[4] O. Alagoz, L.M. Maillart, A.J. Schaefer, and M.S. Roberts. The optimal timing of living-donor liver transplantation. *Management Science*, 50(10):1420–1430, 2004.

[5] O. Alagoz, L.M. Maillart, A.J. Schaefer, and M.S. Roberts. Determining the acceptance of cadaveric livers using an implicit model of the waiting list. *Operations Research*, 55(1):24–36, 2007.

[6] C. Alexopoulos, D. Goldsman, J. Fontanesi, D. Kopald, and J.R. Wilson. Modeling patient arrivals in community clinics. *Omega*, 36(1):33–43, 2008.

[7] T.A. Aloia and B.N. Fahy. A decision analysis model predicts the optimal treatment pathway for patients with colorectal cancer and resectable synchronous liver metastases. *Clinical Colorectal Cancer*, 7(3):197–201, 2008.

[8] W.M. Arrasmith. *Systems Engineering and Analysis of Electro-optical and Infrared Systems*. CRC Press, 2015.

[9] R. Atkins. The fundamentals: How to begin measuring maintenance effectiveness Part iii, Sep 2009. http://www.maintenancetechnology.com/2009/09/, last accessed on March 25, 2016.

[10] M.D. Back, S.C. Schmukle, and B. Egloff. Who is late and who is early? Big five personality factors and punctuality in attending psychological experiments. *Journal of Research in Personality*, 40(5):841–848, 2006.

[11] F.G. Badía and M.D. Berrade. Optimum maintenance of systems under two types of failure. *International Journal of Material and Structural Reliability*, 4(1):27–37, 2006.

[12] F.G. Badía and M.D. Berrade. Optimum maintenance policy of a periodically inspected system under imperfect repair. *Advances in Operations Research*, 2009.

[13] F.G. Badía, M.D. Berrade, and C.A. Campos. Maintenance and monitoring policy under unrevealed failure. *Abstract Book of the* 10^{th} *International Symposium on Applied Stochastic Models and Data Analysis*, 1:142–147, 2001.

[14] F.G. Badía, M.D. Berrade, and C.A. Campos. Optimum inspection under competing risks with proportional hazards. *Applied Stochastic Models in Business and Industry*, 18(4):369–379, 2002.

[15] R.E. Barlow and L. Hunter. Optimum preventive maintenance policies. *Operations Research*, 8(1):90–100, 1960.

[16] R.E. Barlow and F. Proschan. *Mathematical Theory of Reliability.* John Wiley and Sons, New York, 1965.

[17] M.D. Berrade, C.A.V. Cavalcante, and P.A. Scarf. Maintenance scheduling of a protection system subject to imperfect inspection and replacement. *European Journal of Operational Research*, 218(3):716–725, 2012.

[18] B.R. Bucklin and A.M. Dickinson. Individual monetary incentives: A review of different types of arrangements between performance and pay. *Journal of Organizational Behavior Management*, 21(3):45–137, 2001.

[19] R.V. Canfield. Cost optimization of periodic preventive maintenance. *IEEE Transactions on Reliability*, 35(1):78–81, 1986.

[20] T. Cayirli, E. Veral, and H. Rosen. Designing appointment scheduling systems for ambulatory care services. *Health Care Management Science*, 9(1):47–58, 2006.

[21] M. Chen, S. Mizutani, and T. Nakagawa. Random and age replacement policies. *International Journal of Reliability, Quality and Safety Engineering*, 17(01):27–39, 2010.

[22] D.I. Cho and M. Parlar. A survey of maintenance models for multi-unit systems. *European Journal of Operational Research*, 51(1):1–23, 1991.

[23] D. Cousineau. The fallacy of large shape parameters when using the two-parameter Weibull distribution. *IEEE Transactions on Dielectrics and Electrical Insulation*, 18(6):2095–2102, 2011.

[24] A. Darkins, P. Ryan, R. Kobb, L. Foster, E. Edmonson, B. Wakefield, and A.E. Lancaster. Care Coordination/Home Telehealth: The systematic implementation of health informatics, home telehealth, and disease management to support the care of veteran patients with chronic conditions. *Telemedicine and e–Health*, 14(10):1118–1126, 2009.

[25] R. Dekker. Applications of maintenance optimization models: A review and analysis. *Reliability Engineering & System Safety*, 51(3):229–240, 1996.

[26] B.T. Denton, M. Kurt, N.D. Shah, S.C. Bryant, and S.A. Smith. Optimizing the start time of statin therapy for patients with diabetes. *Medical Decision Making*, 29(3):351–367, 2009.

[27] C. Derman, G.J. Lieberman, and S.M. Ross. On the use of replacements to extend system life. *Operations Research*, 32(3):616–627, 1984.

[28] S. Ding and S. Kamaruddin. Maintenance policy optimizationliterature review and directions. *The International Journal of Advanced Manufacturing Technology*, 76(5-8):1263–1283, 2015.

[29] T. Dohi and T. Nakagawa, editors. *Stochastic Reliability and Maintenance Modelling*. Springer, London, 2013.

[30] A.D. Froese, S.J. Nisly, and R.M. May. The effects of task interest and difficulty on procrastination. *Transactions of the Kansas Academy of Science (1903)*, pages 119–128, 1984.

[31] D.L. Frosch and R.M. Kaplan. Shared decision making in clinical medicine: Past research and future directions. *American Journal of Preventive Medicine*, 17(4):285–294, 1999.

[32] A. Garg and S.G. Deshmukh. Maintenance management: Literature review and directions. *Journal of Quality in Maintenance Engineering*, 12(3):205–238, 2006.

[33] G.J. Glasser. The age replacement problem. *Technometrics*, 9(1):83–91, 1967.

[34] U. Gneezy, S. Meier, and P. Rey-Biel. When and why incentives (don't) work to modify behavior. *The Journal of Economic Perspectives*, pages 191–209, 2011.

[35] R. Gomeni, M. Simeoni, M. Zvartau-Hind, M.C. Irizarry, D. Austin, and M. Gold. Modeling Alzheimer's disease progression using the disease system analysis approach. *Alzheimer's & Dementia*, 8(1):39–50, 2012.

[36] P. Guimaraes, K. Kieburtz, C.G. Goetz, J.J. Elm, Y.Y. Palesch, P. Huang, B. Ravina, C.M. Tanner, and B.C. Tilley. Non-linearity of Parkinson's disease progression: Implications for sample size calculations in clinical trials. *Clinical trials*, 2(6):509–518, 2005.

[37] M.A. Hariga. A maintenance inspection model for a single machine with general failure distribution. *Microelectronics and Reliability*, 36(3):353–358, 1996.

[38] J.A. Hermann, A.I. Montes, B. Domínguez, F. Montes, and B.L. Hopkins. Effects of bonuses for punctuality on the tardiness of industrial workers. *Journal of Applied Behavior Analysis*, 6(4):563–570, 1973.

[39] R. Horst, P.M. Pardalos, and N. Van Thoai. *Introduction to Global Optimization.* Springer, 1995.

[40] C. Huang, W.T. Ziemba, and A. Ben-Tal. Bounds on the expectation of a convex function of a random variable: With applications to stochastic programming. *Operations Research*, 25(2):315–325, 1977.

[41] M.G.M. Hunink, M.C. Weinstein, and et al. *Decision making in health and medicine: integrating evidence and values.* Cambridge University Press, 2014.

[42] Z.G. Icten, S.M. Shechter, L.M. Maillart, and M. Nagarajan. Optimal management of a limited number of replacements under Markovian deterioration. *IIE Transactions*, 45(2):206–214, 2013.

[43] A.K.S. Jardine, D. Lin, and D. Banjevic. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7):1483–1510, 2006.

[44] R. Jiang and A.K.S. Jardine. Two optimization models of the optimum inspection problem. *Journal of the Operational Research Society*, 56(10):1176–1183, 2005.

[45] J.M. Keisler and P.S. Noonan. Communicating analytic results: A tutorial for decision consultants. *Decision Analysis*, 9(3):274–292, 2012.

[46] A. Khademi, D.R. Saure, A.J. Schaefer, R.S. Braithwaite, and M.S. Roberts. The price of nonabandonment: HIV in resource-limited settings. *Manufacturing & Service Operations Management*, 17(4):554–570, 2015.

[47] A. Khojandi, L.M. Maillart, O.A. Prokopyev, M.S. Roberts, Roberts, T. Brown, and W. Barrington. Optimal implantable cardioverter defibrillator (ICD) generator replacement. *INFORMS Journal on Computing*, 26(3):599–615, 2014.

[48] K.J. Klassen and R. Yoogalingam. Appointment system design with interruptions and physician lateness. *International Journal of Operations & Production Management*, 33(4):394–414, 2013.

[49] G. Kobelt, I. Lekander, and A. Lang et al. Cost-effectiveness of etanercept treatment in early active rheumatoid arthritis followed by dose adjustment. *International Journal of Technology Assessment in health care*, 27(3):193–200, 2011.

[50] M. Koslowsky. A new perspective on employee lateness. *Applied Psychology*, 49(3):390–407, 2000.

[51] M. Kurt and J.P. Kharoufeh. Optimally maintaining a markovian deteriorating system with limited imperfect repairs. *European Journal of Operational Research*, 205(2):368–380, 2010.

[52] V.C.S. Lau, W.T. Au, and J.M.C. Ho. A qualitative and quantitative review of antecedents of counterproductive behavior in organizations. *Journal of Business and Psychology*, 18(1):73–99, 2003.

[53] C.P. Lee, G.M. Chertow, and S.A. Zenios. Optimal initiation and management of dialysis therapy. *Operations Research*, 56(6):1428–1449, 2008.

[54] C. Lerman and M. Schwartz. Adherence and psychological adjustment among women at high risk for breast cancer. *Breast Cancer Research and Treatment*, 28(2):145–155, 1993.

[55] B. Lienhardt, E. Hugues, C. Bes, and D. Noll. Failure-finding frequency for a repairable system subject to hidden failures. *Journal of Aircraft*, 45(5):1804–1809, 2008.

[56] M. Madadi, S. Zhang, and L.M. Henderson. Evaluation of breast cancer mammography screening policies considering adherence behavior. *European Journal of Operational Research*, 247(2):630–640, 2015.

[57] J.E. Mason, D.A. England, B.T. Denton, S.A. Smith, M. Kurt, and N.D. Shah. Optimizing statin treatment decisions for diabetes patients in the presence of uncertain future adherence. *Medical Decision Making*, 32(1):154–166, 2012.

[58] J. J. McCall. Maintenance policies for stochastically failing equipment: A survey. *Management Science*, 11(5):493–524, 1965.

[59] M. McFarland, K. Davis, J. Wallace, J. Wan, R. Cassidy, T. Morgan, and D. Venugopal. Use of home telehealth monitoring with active medication therapy management by clinical pharmacists in veterans with poorly controlled type 2 diabetes mellitus. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*, 32(5):420–426, 2012.

[60] T. Nakagawa. Sequential imperfect preventive maintenance policies. *IEEE Transactions on Reliability*, 37(3):295–298, 1988.

[61] T. Nakagawa. *Maintenance Theory of Reliability*. Springer Science & Business Media, 2006.

[62] T. Nakagawa. *Advanced Reliability Models and Maintenance Policies*. Springer Science & Business Media, 2008.

[63] T. Nakagawa. *Random Maintenance Polices*. Springer, London, 2014.

[64] R.P. Nicolai and R. Dekker. Optimal maintenance of multi-component systems: A review. In *Complex System Maintenance Handbook*, pages 263–286. Springer, 2008.

[65] Nirjharini. What are preventative maintenance practices in aviation?, JUl 2011. http://www.brighthub.com/science/aviation/articles/122479.aspx, last accessed on Sep 22, 2016.

[66] S. Okumura. Determination of inspection schedules of equipment by variational method. *Mathematical Problems in Engineering*, 2006.

[67] A.S. O'Malley, C.B. Forrest, and J. Mandelblatt. Adherence of low-income women to cancer screening recommendations. *Journal of General Internal Medicine*, 17(2):144–154, 2002.

[68] G. Parmigiani. Optimal scheduling of fallible inspections. *Operations Res.*, 44(2):360–367, 1996.

[69] S. Peters. Why a reactive maintenance plan is a bad idea, JUN 2016. https://blog.craneengineering.net/why-a-reactive-maintenance-plan-is-a-bad-idea, last accessed on Sep 22, 2016.

[70] H. Pham and H. Wang. Imperfect maintenance. *European Journal of Operational Research*, 94(3):425–438, 1996.

[71] H. Pham and H. Wang. Imperfect maintenance. *European Journal of Operational Research*, 94(3):425–438, 1996.

[72] S.M. Ross. *Applied Probability Models With Optimization Applications*. Holden-Day, San Francisco, 1970.

[73] A. Salonen and M. Deleryd. Cost of poor maintenance: A concept for maintenance performance improvement. *Journal of Quality in Maintenance Engineering*, 17(1):63–73, 2011.

[74] B. Sandikci, L.M. Maillart, A.J. Schaefer, O. Alagoz, and M.S. Roberts. Estimating the patient's price of privacy in liver transplantation. *Operations Research*, 56(6):1393–1410, 2008.

[75] R. Serfozo. *Basics of Applied Stochastic Processes*. Springer Science & Business Media, 2009.

[76] M. Shafiee and S. Chukova. Maintenance models in warranty: A literature review. *European Journal of Operational Research*, 229(3):561–572, 2013.

[77] A. Sharma, G.S. Yadava, and S.G. Deshmukh. A literature review and future perspectives on maintenance optimization. *Journal of Quality in Maintenance Engineering*, 17(1):5–25, 2011.

[78] S.M. Shechter. *When to initiate, when to switch, and how to sequence HIV therapies: A Markov decision process approach*. PhD thesis, University of Pittsburgh, 2006.

[79] S.M. Shechter, M.D. Bailey, and A.J. Schaefer. A modeling framework for replacing medical therapies. *IIE Transactions*, 40(9):861–869, 2008.

[80] S.M. Shechter, M.D. Bailey, and A.J. Schaefer. Replacing nonidentical vital components to extend system life. *Naval Research Logistics (NRL)*, 55(7):700–703, 2008.

[81] S.M. Shechter, M.D. Bailey, A.J. Schaefer, and M.S. Roberts. Are threshold policies sufficient for the optimal timing of HIV treatment? 2008.

[82] S.M. Shechter, M.D. Bailey, A.J. Schaefer, and M.S. Roberts. The optimal time to initiate HIV therapy under ordered health states. *Operations Research*, 56(1):20–33, 2008.

[83] S. Smith. The advantages of preventive maintenance, SEP 2012. http://www.mintek.com/blog/eam-cmms/advantages-preventive-maintenance/, last accessed on Sep 22, 2016.

[84] P. Steel. The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin*, 133(1):65, 2007.

[85] S. Subramanian, M. Klosterman, M.M. Amonkar, and T.L. Hunt. Adherence with colorectal cancer screening guidelines: A review. *Preventive Medicine*, 38(5):536–550, 2004.

[86] S. Taghipour and D. Banjevic. Periodic inspection optimization models for a repairable system subject to hidden failures. *IEEE Transactions on Reliability.*, 60(1):275 –285, 2011.

[87] C. Valdez-Flores and R.M. Feldman. A survey of preventive maintenance models for stochastically deteriorating single-unit systems. *Naval Research Logistics*, 36(4):419–446, 1989.

[88] A. Van Horenbeek, L. Pintelon, and P. Muchiri. Maintenance optimization models and criteria. *International Journal of System Assurance Engineering and Management*, 1(3):189–200, 2010.

[89] G.J. Wang and Y.L. Zhang. Optimal periodic preventive repair and replacement policy assuming geometric process repair. *IEEE Transactions on Reliability*, 55(1):118–122, 2006.

[90] H. Wang. A survey of maintenance policies of deteriorating systems. *European Journal of Operational Research*, 139(3):469–489, 2002.

[91] W. Wang. Delay time modelling. In *Complex Syst. Maintenance Handbook*, pages 345–370. Springer, 2008.

[92] R.I. Zequeira and C. Bérenguer. Optimal scheduling of non-perfect inspections. *IMA Journal of Management Mathematics*, 17(2):187–207, 2006.

[93] X. Zhao and T. Nakagawa. Optimization problems of replacement first or last in reliability theory. *European Journal of Operational Research*, 223(1):141–149, 2012.

[94] X. Zhao and T. Nakagawa. Comparisons of periodic and random replacement policies. In I.B. Frenkel, A. Karagrigoriou, A. Lisnianski, and A. Kleyner, editors, *Applied Reliability Engineering and Risk Analysis: Probabilistic Models and Statistical Inference*, pages 193–204. Wiley, 2014.