OPTIMAL BURN-IN UNDER COMPLEX FAILURE PROCESSES: SOME NEW PERSPECTIVES

YE ZHISHENG

(B.ENG. & B.ECO., TSINGHUA UNIVERSITY)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF INDUSTRIAL AND SYSTEMS ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2011

ACKNOWLEDGEMENT

The 3.5 years study in the ISE department has been an unforgettable experience of my life. First of all, I need to express my sincere thanks to my main advisor – Professor Min Xie – for his great care, selfless help and invaluable guidance on every aspect of my life that allow me to grope my way in an unknown academic world. His talent and enthusiasm in research and his positive attitude towards everyday life have influenced me a lot, and will definitely become a fabulous legacy to my future life. I am also deeply indebted to my co-advisor, Professor Loon-Ching Tang for encouraging me to link the theory to practice. His wisdoms in research and his experience in solving practical problems have enlightened me a lot in my research directions. I am also grateful to Professor Pra Murthy for teaching me how to organize an academic paper, and Professor Melvyn Sim for encouraging high-quality research.

I am fortunate to have worked closely with many colleagues, such as Chan Ping-Shing, Chen Nan, Huang Boray, Ng Tony, Shen Yan and Xu Haiyan, etc. Discussions with them enable me to learn a lot from them. I am also grateful to the Chinese University of Hong Kong for involving me in the "Global Scholarship Programme for Research Excellence" so that I got the opportunity to visit the department of Statistics and knew many good friends there, including Chen Pengcheng, Hai Yizhen, Liu Pengfei, Tang Yanlin, and Zhao Honghao, etc.

More importantly, I shall thank the National University of Singapore for providing me the prestigious President Graduate Fellowship. This fellowship provides me with physical support so that I can devote to my research. I also wish to express my gratitude to my labmates and tennismates, including Chen Liangpeng, Deng Peipei, Faghih-Roohi Shahrzad, Jiang Jun, Jiang Xinjia, Li Xiang, Peng Rui, Wang Chaoxu, Wu Jun, Wu Zhengxiao, Xie Yujuan, Yuan Jun, Zhao Gongyun and Yao Zhishuang, etc.

TABLE OF CONTENTS

SUMMARY

It is well-known that most semi-conductor devices suffer from infant mortality, resulting in billions of warranty losses due to early field failures. Burn-in is an important engineering procedure used to identify defective units by subjecting all units to a screening test with a certain duration. Optimal determination of the burn-in settings is of particular importance, as it enhances field performance of a product and saves field operation costs up to the hilt. Motivated by some practical problems with complex failure processes, this thesis is aimed at developing some practical burn-in models to help determine the optimal burn-in settings.

We first propose a burn-in scheme based on change points of the *p*-percentile function of the residual life function. This scheme is able to simultaneously yield the optimal burn-in duration and the optimal warranty period, which is important for products whose warranty coverage is yet to be determined. We also identify severe infant mortality faced by products sold with two-dimensional warranties, and subsequently propose two novel burn-in models. In view of the fact that modern manufacturing technique has led to what is commonly known as highly reliable products, this thesis advocates degradation-based burn-in approaches that base the screening decision on a product's degradation level after burn-in. We first develop two degradation-based joint burn-in and preventive maintenance models for products whose degradation is measurable. Then, we recognize the fact that product failures are much more complex, and thus propose a degradation-based burn-in framework under competing risks. In addition, we propose a bi-objective burn-in framework that simultaneously takes the cost and field performance of a burnt-in unit into consideration. These proposed models are successfully applied to solve a number of real problems, which shows the significant practical contributions of this thesis.

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

1.1 Background

To meet a sequence of performance specifications, the design reliability of most products is often very high. During the post-development stage, however, actual reliability usually differs from the designed one due to quality variations. These variations include non-conforming components/materials, design defects and manufacturing defects, which result in a subpopulation of defective units in the product population. These defectives lead to significant number of early failures in field use, commonly known as infant mortalities. Infant mortalities are not uncommon in practice. Analyses of automobile warranty data by Attardi *et al.* (2005), Majeske (2003, 2007) and Rai and Singh (2006) indicated that automobiles suffer from infant mortality, accounting for as high as 5.6% of the total failures. Kececioglu and Sun (1994) analyzed some CMOS data and find that the infant mortality rate reaches 6.7%. More remarkably, the infant mortalities can be even as high as 10% in a new and unproven technology (Kececioglu and Sun 1997). These undesirable early failures degrade performance of the product and significantly increase field operation costs.

To alleviate the impact of early failures, engineers often resort to burn-in testing. In fact, burn-in has been an important manufacturing screening operation. It is often conducted under harsh environments that simulate the severest working conditions, such as a combination of random vibration, thermal cycling and shock, for a certain duration. After a moderate duration of burn-in, defective units can be identified and eliminated, and reliability of the product can be greatly enhanced. However, excessive burn-in will shorten the useful lifetimes of the normal units and prolong time-to-market of the product. Therefore, efficient burn-in models are imperative in deciding on optimal burn-in settings.

Burn-in models are often built jointly with warranty decisions or preventive maintenance decisions. A bulk of literature focuses on the benefits of burn-in on products sold with warranties. Previous burn-in models were often built given a specific warranty policy. But considering the fact that more and more warranty policies, e.g., the two-dimensional warranty provided by most car suppliers and airplane manufacturers, have received successful applications recently, it calls for the development of new burn-in models for products sold with these warrantees. On the other hand, some researchers proposed burn-in models for products under preventive maintenance. These models are based on traditional burn-in approaches that intend to fail short-lived units (Nelson 1990, Chap. 5.5). In a traditional burn-in test, only units that survive the test are put into field use. However, there are many modern products that are so well designed and manufactured that they are highly reliable. It may take a very long time for a defective unit to fail even under highly accelerated stress. In addition, failure mechanisms for modern reliable products are increasingly more complex. In practice, some quality characteristic of a product closely related to lifetime of the product usually degrades over time and causes a product failure when the degradation level of such quality characteristic exceeds a certain threshold, which is often stipulated by the industrial standards. If the quality characteristic of a defective unit degrades faster than a normal one, this unit can be effectively identified through degradation-based burn-in. Nevertheless, degradation-based burn-in models are rare in the literature and need further investigation.

The remainder of this chapter briefly introduces how burn-in is implemented in practice, after which a review of burn-in literature will be presented. This review will address issues of current research and highlight necessity of work in this thesis.

1.2 Burn-In Modeling

Burn-in is beneficial as it is able to decrease the possibility of field failures. Therefore, burn-in models are often built jointly with certain field operation. Based on two kinds of field operations, i.e., warranty and preventive maintenance, this thesis thus classifies existing burn-in models into two categories for ease of exposition. The first category focuses on effects of burn-in on performance of products sold with warranties, while the second category models the joint decisions of burn-in and preventive maintenance. These two categories of models are briefly reviewed in this section.

1.2.1 Joint Burn-In and Warranty Models

In our modern commercial society, products are becoming more and more complex with each new generation to meet the growing needs and expectations of customers. Due to this complexity, more defects may be introduced into such products, leading to a significant number of early failures. Therefore, customers need assurance that the product will perform satisfactorily. A warranty provides such an assurance. Nevertheless, offering warranty implies additional costs to a manufacturer due to servicing of claims. These claims also impact sales as well as the manufacturer's reputation. The effect of early failures on warranty costs provides strong motivation for conducting a burn-in test, as burn-in has been proven to be effective in removing early failures before the products are sold to customers.

Burn-in models for products sold with warranties can be broadly classified into two classes based on the objective function of burn-in. The first class tries to optimize certain performance index given a warranty period. A widely accepted burn-in criterion is the probability of failure within the warranty period, as this probability represents the proportion of field returns, which, in turn, is related to warranty cost. Typically, the optimal burn-in duration is determined by minimizing the probability of failure within a given warranty period, if there is an established norm for this period. This criterion has been investigated by many studies, including Mi (1994b , 2003), Kim and Kuo (2003, 2005, 2009) and Cha and Finkelstein (2010b).

The second class is the cost-based burn-in models that seek to achieve an optimal balance between burn-in costs and warranty costs. Different warranty policies and different lifetime distributions of products lead to various cost models. Optimal burn-in decisions under different warranty policies have been discussed by Mi (1999), Sheu and Chien (2005), Wu *et al.* (2007), etc. As for the lifetime distributions, Chou and Tang (1992), Mi (1997) and Cha *et al.* (2008) formulated cost-based burn-in warranty models under a mixture of Weibull distributions, distributions with BTFRs, and distributions with eventually increasing FRs, respectively.

Although many burn-in warranty models have been proposed and studied, there are still a number of deficiencies. First of all, all models assumed a predetermined warranty period, notwithstanding the fact that manufacturers may also be interested in determining an optimal warranty period, especially when there is not a norm for this period. In addition, all models assumed single failure mode while this mode is subject to infant mortality. However, most products can fail due to one of a series of failure modes, or competing risks. Last but not least, no model has dealt with burn-in for products sold with two-dimensional warranty, a very important warranty policy for expensive products such as cars.

1.2.2 Joint Burn-In and Maintenance Models

Obviously, not all products are sold with warranty, especially when a product is cheap, or is a component to a complex system. For this kind of product, preventive maintenance is often applied to improve field performance. Therefore, the other category of burn-in models deals with joint burn-in and preventive maintenance decisions. Mi (1994a) systematically studied this joint decision problem under the bathtub failure rate assumption. The models in Mi (1994a) were further extended by Cha (2001, 2003, 2005) and Cha and Mi (2007). A detailed review of these models will be provided in Chapter 2. But as noted by Cha and Finkelstein (2010a), the bathtub failure rate describes only up to 15% of applications. As opposed to the bathtub failure rate assumption, some authors investigated this problem by assuming a bimodal distribution, e.g., Drapella and Kosznik (2002) and Jiang and Jardine (2007) to name a few. For excellent overviews on joint burn-in and maintenance modeling, readers are referred to Liu and Mazzuchi (2008) and Cha (2011).

Basically, the above-mentioned models considered binary system states, either failed or working. It is generally believed that this kind of models is not efficient for reliable products. If the degradation of some quality characteristic is observable, degradation-based burn-in would be much more efficient. However, degradation-based burn-in maintenance models are not found in literature.

1.3 Research Objectives

The comprehensive review above has revealed that current research on burn-in models is still far from perfection. As products are becoming more sophisticated, their failure mechanisms are much more complex. The purpose of this thesis is to develop practical burn-in models for these products from some new perspectives. More specifically, this thesis is to:

- Investigate a performance-based burn-in scheme for achieving the maximum allowable warranty period with a prespecified field-return probability. This is done by exploiting properties of the *p*-percentile function of the residual life when a distribution exhibits a bathtub shape failure rate.
- Develop a burn-in planning framework for products with independent multiple failure modes. This framework would be potentially very important, as a complex product often has a couple of failure modes.
- Develop both performance and cost-based burn-in models for products sold with two-dimensional warranty.

- Develop degradation-based burn-in models for products with preventive maintenance. Two maintenance options, i.e., age-based maintenance and block-based maintenance are considered.

All the models in this thesis are motivated by practical problems, and thus may be potentially very useful for burn-in practitioners to achieve more cost-effective burn-in decisions. These practical problems will also shed some light on the focuses of future research on burn-in problems. On the other hand, the modeling methodologies developed in this thesis, e.g., isolating defect failures from normal failures, taking into account the parameter uncertainty on the optimal burn-in decision, etc. may open up a new avenue for burn-in analysis. This thesis advocates uses of degradation signals for burn-in decision-making, and thus has a direct link to the important area of prognostics and health management, in which the future performance of a product is predicted by detection of early signs of wear and aging.

The remainder of this thesis is organized as follows. Chapter 2 provides a comprehensive review of the burn-in literature. Chapter 3 proposes a burn-in scheme based on percentile of the residual life. Chapter 4 builds two burn-in models for products sold with two-dimensional warranties. Chapter 5 investigates degradation-based joint burn-in and preventive maintenance problems. Chapter 6 proposes a burn-in framework for products with competing risks. Chapter 7 deals with a bi-objective burn-in framework. Chapter 8 concludes the whole thesis and points out possible topics for future research.

CHAPTER 2 LITERATURE REVIEW

Burn-in testing has become an important engineering practice to deal with infant mortalities. To help decide on the optimal burn-in settings, many burn-in models have been proposed in the literature. Based on the objective functions of burn-in, most of these models can be classified into two categories, i.e., joint burn-in and warranty models and joint burn-in and maintenance models. Some models do not belong to these two categories. They are classified into the third category, and will be reviewed in Section 2.3.

2.1 Joint Burn-In and Warranty Models

Lifetimes of many commercial products exhibit a bathtub-shaped failure rate (BTFR) which consists of a short infant mortality period with a decreasing failure rate (FR), followed by a useful life period with relatively constant and low FR, and then a wear-out period that exhibits an increasing FR. Customers need assurance, most often in the form of a warranty contract, to protect against possible early failures. However, the warranty obligation engenders additional cost to the manufacturers due to service of warranty claims. Burn-in is an effective method used to reduce the number of early failures and cut down the warranty costs. A number of burn-in models for products sold with warranties have been developed in the literature. These models can be classified into two classes, i.e., cost-based and performance-based.

The first work on cost-based burn-in modeling for products sold with a warranty dates back to Nguyen and Murthy (1982), who examined the optimal burn-in time to achieve a trade-off between reduction in the warranty cost and increase in the manufacturing cost, as burn-in is viewed as part of the manufacturing process. Following this ground-breaking work, many burn-in models for products sold with warranty have been proposed and studied subsequently.

Chou and Tang (1992) extended the models of Nguyen and Murthy (1982) by using a mixture of exponential distribution and a Weibull hyperexponential distribution with shape parameter less than one. On the other hand, Mi (1997) extended the models of Nguyen and Murthy (1982) under the assumption of bathtub failure rate. He then showed that the optimal burn-in duration that minimizes the total burn-in warranty cost function never exceeds the first change point of the failure-rate function. Mi (1997)'s model was extended by Cha *et al.* (2008) to the eventually increasing failure rate case. On the other hand, Chang (2000) examined the optimal burn-in problem under the unimodal failure rate assumption. Mi (1999) examined the expected burn-in warranty cost under different warranty policies, including replacement-free warranty, renewable warranty, and pro-rata warranty. Yun *et al.* (2002) studied the burn-in problems under the cumulative free replacement warranty policy. Sheu and Chien (2005) considered burn-in tests for general repairable products and examined different warranty policies. Ulusoy *et al.* (2011) proposed a Bayes method to compute the burn-in warranty cost and to determine the optimal burn-in time.

The second class of models is performance-based burn-in models. Most models in this class aimed at minimizing the probability of failure within a given warranty period, as this probability represents the proportion of field returns, which, in turn, is related to warranty cost. An early study dating back to Mi (1994b) considered this criterion under the assumption of bathtub failure rate. Mi (2003) extended the model to the case of eventually increasing failure rate. Instead of considering the overall failure rate, Kim and Kuo (2003) built a model by analyzing a system at component level to trace back sources of the assembly defects. Their model was further extended to different types of time-to-defect-failure distributions by Kim and Kuo (2005), and Kim and Kuo (2009). Besides the probability of failure, another performance index that is closely related to the warranty cost is the mean number of failures within warranty. This criterion was investigated by Mi (1994b) who assumed minimal repair and bathtub failure rate. Cha and Finkelstein (2010b) revisited this criterion under the assumption of minimal repair and bimodal distribution.

The above models were built by assuming a pre-specified warranty period. However, in addition to the burn-in duration, a manufacturer may also be interested in determining an optimal warranty period. This is especially true for newly developed products for which a norm for the warranty period has yet to be determined, or for second-hand products for which the warranty coverage is quite flexible and depends partly on other considerations of the manufacturer. Another scenario is when the designed-in reliability of a product has been greatly improved to the extent that a longer warranty period may be considered to provide a competitive advantage. In the literature, Kar and Nachlas (1997) treated burn-in and warranty strategies together and examined the possible benefits from this coordinated strategy. Wu *et al.* (2007) developed a cost model to determine the optimal burn-in time and warranty length for non-repairable products under the fully renewing combination free replacement and pro-rata warranty policy. These two models are all cost-based. A problem associated with these models is that it is hard to determine the marginal benefit of prolonging the warranty period. Therefore, performance-based models that can be used to simultaneously determine the optimal burn-in duration and the optimal warranty period are desired.

Moreover, all the above models have been restricted to the case of one-dimensional warranties, under which only the age is restricted. In contrast, not any effort has been found with regard to burn-in modeling under two-dimensional warranties (Ye et al. 2012b). Unlike the one-dimensional warranty, a two-dimensional warranty is characterized by a region in two dimensions with one axis representing age and the other usage. Such a warranty policy has received successful applications in many industries including automobile, locomotive traction motor, aircrafts and printers. Many products sold with two-dimensional warranties do have infant mortality, as well. For example, analyses of automobile warranty data by Attardi *et al.* (2005), Rai and Singh (2006) and Majeske (2007) revealed that automobiles suffer from infant mortality, accounting for as high as 5.6% of the total failures. These reports justify the necessities and importance of burn-in for products sold with two-dimensional warranties.

In addition, existing burn-in models commonly assumes single failure mode while this mode is subject to infant mortality. However, most products can fail due to one of a series of failure modes, or competing risks. A good real life example can be found in Meeker *et al.* (2009), where a newly designed product has 12 failure modes. Making use of the failure mode information is able to improve accuracy of both estimation and prediction (Hong and Meeker 2010, pp. 150), and thus is important. In fact, there has been a bulk of literature on the research of competing risks. Suzuki *et al.* (2010) reported two competing failure modes, i.e., internal and surface cracks, in a load bending test for brittle materials. Liu and Tang (2010) developed accelerate life test (ALT) plans for products with independent competing risks. Ye *et al.* (2011b) proposed a system reliability model under dependent competing risks. Crowder (2001) provided a book-length treatment on competing risk modeling and estimation.

2.2 Joint Burn-In and Preventive Maintenance Models

When items with burn-in are put into field use, some rational preventive replacement (PM) strategies are often adopted to further improve the system availability and cut down field failure costs (Chen *et al.* 2011). For reviews of the literature on maintenance, see Wang (2002) and van Noortwijk (2009). Compared with making isolated burn-in and maintenance decisions, joint modeling of the burn-in procedure and the PM decision would be more cost-effective, and thus has attracted many attentions. The first study on joint burn-in and maintenance modeling dates back to Mi (1994a). More specifically, he considered two scenarios, i.e., age replacement with complete repair upon failure, and block replacement with minimal repair upon failure. Under these two scenarios, he proved that with the BTFR assumption, the optimal burn-in time is smaller than the first change point of the BTFR, while the optimal maintenance time is larger than the second change point of the BTFR. Cha (2000) showed that under age replacement, minimal repair is always economical than complete repair. Moreover, with age replacement and minimal repair, the optimal burn-in time is again smaller than the first change point while the optimal maintenance is again larger than the second change point of the BTFR. Cha (2001, 2003) further extended the joint burn-in and age replacement model to include two types of failures, i.e., Type I failure that can be minimally repaired and Type II failure that has to be completely repaired. Cha and Mi (2007) extended the models to the assumption of eventually increasing failure rate.

All the above models focused on the overall failure rate of a product. On the other hand, some researchers built joint burn-in maintenance models based on the assumption that the population consists of two sub-populations, i.e., normal and weak. Drapella and Kosznik (2002) developed a software package to solve the optimal burn-in duration and maintenance intervals. Jiang and Jardine (2007) assumed two subpopulations and perfect repair, i.e., replacement upon failure, and derived the total costs of joint burn-in and maintenance under age replacement policy.

It is noted that a common feature of these models is that they are all distribution-based in the sense that only the binary system state, i.e. working and failure, is observable. On the other hand, the rapid development of modern technology and the increasing efforts on process quality management have led to what is commonly referred to as highly reliable products. In fact, many devices are so well designed that it may need a long burn-in duration to fail a freaky unit even under highly accelerated environment, e.g., a light emitting diode (LED) product (Tseng and Peng 2004) and electronic devices (Ye et al. 2012e). The traditional failure-based burn-in approach is thus not effective. Compounded by the need to shorten the time-to-market, engineers are faced with a difficult task of making the screening decision for a reliable product within an acceptable time frame. For these reliable products, there is often some quality characteristic that degrades over time and causes product failure when the degradation level exceeds some threshold. Cumulative degradations can often be measured through modern real-time diagnostic techniques. Moreover, the quality characteristic of a defective unit often degrades faster than a normal one, and thus can be effectively identified through degradation-based burn-in. These are some degradation-based burn-in models in the literature, e.g., Tseng and Tang (2001), Tseng *et al.* (2003), Tseng and Peng (2004) and Tsai *et al.* (2011). These models will be reviewed in the next section. However, degradation-based joint burn-in and preventive maintenance models are not found in literature and needs further investigation (Ye et al. 2012c).

2.3 Other Burn-In Models

Some other burn-in models do not belong to the above two categories. Objective functions of these models include maximizing the mean residual life (MRL) of a burnt-in unit, and minimizing the misclassification cost.

The MRL is an important index in the literature of reliability engineering. The first paper on burn-in modeling owes to Watson and Wells (1961). This ground-breaking paper examined a couple of well-known distributions and found the conditions under which the MRL is greater than the original mean life. Following this idea, Lawrence (1966) derived sharp upper and lower bounds on the burn-in time to achieve a specified MRL, given that the product has decreasing failure rate. Mi (1993) considered the case of discrete failure. Mi (1995) showed that when a distribution has BTFR, the resultant MRL has an upside-down BTFR. Xie *et al.* (2004) further examined the associated relationship between change points of the failure rate function and the MRL function. But none of these studies took the cost into account, notwithstanding the fact that burn-in is an expensive procedure. This motivates Weiss and Dishon (1971) to introduce a cost-based burn-in model by considering burn-in cost and gain due to improvement in the MRL. Mi (1996) further develop a similar model under the assumption of bathtub failure rate. Mi (1996)'s model was extended by a number of studies, including Cha (2005), Sheu and Chien (2004) and Cha and Finkelstein (2010b).

Some other burn-in studies try to minimize the misclassification cost of a burn-in test. Tseng and Tang (2001) introduced a degradation-based burn-in model with the purpose of minimizing the burn-in cost plus the misspecification cost. The optimal cut-off degradation levels were determined. This model is further extended by Tseng *et al.* (2003) and Tseng and Peng (2004) based on variants of the Wiener process. On the other hand, based on the gamma process, Tsai *et al.* (2011) developed similar cost models and determined the optimal cut-off levels. Wu and Xie (2007) also developed a burn-in model to minimize the misspecification cost by means of the ROC curve.

CHAPTER 3 A BURN-IN SCHEME BASED ON PERCENTILES OF THE RESIDUAL LIFE

3.1 Introduction

As reviewed in Section 2.1, the optimal burn-in duration is often determined by minimizing the probability of failure within a given warranty period if there is an established norm for this period. On the other hand, a manufacturer might prefer to control the percentage of warranty return to make full use of the service facilities and personnel, as the investment in the after-sales service department has high fixed costs (e.g. purchase of the repair machine and employment of the after-sales personnel). But this percentage should not be so large that it exceeds the capacity of the after-sales service department. These examples emphasize the need for burn-in models that are able to simultaneously determine the optimal burn-in duration and an optimal warranty period during which the probability of failure can be controlled.

In this chapter, we investigate a performance-based burn-in scheme for achieving the maximum allowable warranty period with a pre-specified field return probability. This is done by exploiting properties of the p-percentile function of the residual life (PRL-p function) when distributions exhibit a bathtub shape failure rate (BTFR). Block *et al.* (1999) showed that the PRL-p function may be an upside-down bathtub shape when the distribution has a BTFR. Therefore, given the percentile $100p$, the change point at which the PRL-p function attains its maximum can be adopted as the burn-in duration. The associated PRL-p is the longest warranty period that fulfills a pre-specified reliability target.

It is also noted that another commonly used measure in describing the lifetime of items is the mean residual life (MRL). However, this measure, although closely related to other performance measures, should not be directly used in a burn-in warranty problem. This is because it is possible for a product to have a high MRL, while having some substantial subset of the population fail very early (cf. Coit and Smith 2002), e.g. if the lifetime distribution has heavy-tail property. Therefore, a high MRL does not necessarily imply low warranty claims and may not be a good criterion for risk-averse users.

In the following, we define the PRL-*p* and outline some salient features of the change point of the PRL-*p* function in Section 3.2. The maximum likelihood estimators (MLEs) of the change points and the corresponding PRL-*p*, as well as their asymptotic distributions, are then derived in Section 3.3. Section 3.4 focuses on PRL-*p* functions of some generalized Weibull distributions with BTFRs and present the inference procedure for the modified Weibull extension proposed by Xie *et al.* (2002). Section 3.5 provides a numerical example to illustrate the estimation of jointly optimal burn-in duration and warranty period proposed in this study. The last section concludes the chapter.

3.2 The *p*-Percentile Function of the Residual Life

According to Joe and Proschan (1984), the *p*-percentile of the residual life (PRL-*p*) $T_p(t)$ is the $100p^{\text{th}}$ percentile of the residual life given survival up to time *t*, which can be expressed as

$$T_p(t) = \inf\left\{u \geq 0 : R(t+u) \leq (1-p)R(t)\right\}, \tag{3.1}$$

where $R(\cdot)$ is the reliability function. When $p = 0.5$, the PRL-0.5 is the median residual life; while $T_p(0)$ is exactly the percentile life of the product.

Properties of the PRL-*p* have been explored in a number of studies, e.g. see Csörgö and Csörgö (1987), Jung *et al.* (2009) and Franco-Pereira *et al.* (2011) among others. The median residual life as a special case of the PRL-*p* has also received much attention, e.g. see Jeong *et al.* (2008) and Ma and Yin (2011), to name a few.

Consider a continuous random variable. Denote by $h(\cdot)$ and $H(\cdot)$ the FR and cumulative failure rate (CFR) functions, respectively. The reliability function $R(\cdot)$ has a simple relation with the FR and CFR:

$$R(t) = \exp\left(-H(t)\right) = \exp\left(-\int_0^t h(u)\,du\right). \tag{3.2}$$

When a distribution has a positive FR within its support, u in (3.1) achieves its minimum when the expression in the brackets is an equality. It follows that (3.1) can be simplified as

$$H\left(T_p(t)+t\right) - H(t) = -\ln(1-p). \tag{3.3}$$

When a distribution has BTFR, its PRL-p function turns out to be upside-down bathtub shaped in many cases. In the following, we focus on distributions with BTFR and discuss the conditions for the existence and the uniqueness of the change point of the PRL-p function.

Proposition 3.1. Consider a distribution with continuous BTFR function $h(t)$ and a PRL-p function $T_p(t)$. Let $c_1 = h(0)$ and $c_2 = h(\infty)$.

(a) If $c_1 \geq c_2$, then for any $0 < p < 1$, $\left(t_p^*, T_p\left(t_p^*\right)\right)$ is the unique solution of the following system of equations.

$$\begin{aligned} &H\left(T_p(t)+t\right) - H(t) = -\ln(1-p) \\ &h\left(T_p(t)+t\right) = h(t) \end{aligned}. \tag{3.4}$$

(b) If $c_1 < c_2 \leq \infty$, denote $t_1 = \inf\left\{t : h(t) > c_1\right\}$. When $0 < p < 1-\exp\left(-H(t_1)\right)$, $\left(t_p^*, T_p\left(t_p^*\right)\right)$ is the unique solution of Equation (3.4); when $1-\exp\left(-H(t_1)\right) \leq p < 1$, $T_p(t)$ is decreasing over $(0,\infty)$.

Proof: We shall prove (b), as (a) can be deduced in a similar way.

Suppose $c_1 < c_2 \leq \infty$. When $1-\exp\left(-H(t_1)\right) \leq p < 1$, if we can prove that $h\left(T_p(t)+t\right) \geq h(t)$

for any $t \ge 0$, then substituting this relation into Equation (3.8) yields

$$\frac{d}{dt}T_p(t) = \frac{h(t)}{h\left(t+T_p(t)\right)} - 1 \le 0.$$

That is, $T_p(t)$ is decreasing over $(0,\infty)$, and thus the change point is 0. In the following, we shall show that $1-\exp\left(-H(t_1)\right) \le p < 1$ implies $h\left(T_p(t)+t\right) \ge h(t)$.

(i) Let $\upsilon = \arg\min h(t)$. From the definition of t_1, we have $t_1 > \upsilon$. This means that $h(t)$ is increasing over $[t_1,\infty)$. Therefore, $h\left(T_p(t)+t\right) \ge h(t)$ when $t > t_1$.

(ii) When $t < t_1$, the definition of t_1 implies $h(t_1) \ge h(t)$. The following relation

$$\begin{aligned} &1-\exp\left(-H(t_1)\right) \le p < 1 \Rightarrow -\log(1-p) \ge H(t_1) \\ &\Rightarrow H\left(T_p(t)+t\right) - H(t) \ge H(t_1) \Rightarrow H\left(T_p(t)+t\right) \ge H(t_1) \end{aligned} \tag{3.5}$$

implies that $T_p(t)+t \ge t_1$. Since $t_1 > \upsilon$, it follows that $h\left(T_p(t)+t\right) \ge h(t_1)$. This inequality, along with $h(t_1) \ge h(t)$ when $t < t_1$, yields the result that $h\left(T_p(t)+t\right) \ge h(t)$ when $t < t_1$.

Combining (i) and (ii) yields the desired result.

When $0 < p < 1-\exp\left(-H(t_1)\right)$, by setting $\left.\frac{dT_p}{dt}\right|_{t=t_p^*} = 0$, we find that, $\left(t_p^*, T_p\left(t_p^*\right)\right)$ also satisfies

$$h\left(T_p\left(t_p^*\right)+t_p^*\right) = h\left(t_p^*\right). \tag{3.6}$$

Therefore, $\left(t_p^*, T_p\left(t_p^*\right)\right)$ is the solution of Equation (3.4).

Uniqueness: Suppose $\left(b, T_p(b)\right)$ is another solution for this system of equations that differs from $\left(t_p^*, T_p\left(t_p^*\right)\right)$. Without loss of generality, suppose $b < t_p^*$. Since $t_p^* < \upsilon$ (Block *et al.* 1999), it follows from the definition of τ that $h(b) > h(t_p^*)$. It then follows from the second equation of Equation (3.4) that $h(T_p(b)+b) > h(T_p(t_p^*)+t_p^*)$. Due to the fact that $h(t)$ is decreasing over $(0,\upsilon)$, it is not possible to find an u, $u \in (b,\upsilon)$, such that $h(b) = h(u)$. Therefore, the second equation of Equation (3.4) also implies that both $T_p(b)+b$ and $T_p(t_p^*)+t_p^*$ should be greater than υ. Therefore, $T_p(b)+b > T_p(t_p^*)+t_p^*$. However,

$$\begin{aligned}
&H(T_p(b)+b) - H(b) \\
&= \left\{H(T_p(b)+b) - H(T_p(t_p^*)+t_p^*)\right\} + \left\{H(T_p(t_p^*)+t_p^*) - H(t_p^*)\right\} + \left\{H(t_p^*) - H(b)\right\}. \\
&> H(T_p(t_p^*)+t_p^*) - H(t_p^*) = -\ln(1-p)
\end{aligned}$$

This results in a contradiction. A similar argument applies if we assume $b > t_p^*$. Thus, $(t_p^*, T_p(t_p^*))$ is the unique solution of Equation (3.4).

■

Essentially, Proposition 3.1 implies that when the FR function has a bathtub shape, there must be some *p* such that the corresponding PRL-*p* function has an upside-down bathtub shape. If the manufacturer allows for few warranty claims, i.e., a small *p*, a moderate duration of burn-in would lead to a longer maximum allowable warranty period, as the PRL-*p* function is of upside-down bathtub shape. On the other hand if the manufacturer allows for more field return, burn-in is effective and economical only when the failure rate during infant mortality period is higher than that of the wear-out period.

Proposition 3.2. Continue with Proposition 3.1. Both t_p^* and $T_p(t_p^*)$ are continuous in *p*. Moreover, t_p^* is decreasing in *p* while $T_p(t_p^*)$ is increasing in *p*.

Proof: Consider the case when $c_1 < c_2 \le \infty$. From the definition of t_2 we can see that $t_2 = 0$. Suppose $\left(b_1^*, T_{p_1}\left(b_1^*\right)\right)$ and $\left(b_2^*, T_{p_2}\left(b_2^*\right)\right)$ are the change points and the corresponding PRL-*p* for p_1 and p_2, where $p_1 < p_2$.

(i) When $1-\exp\left(-H\left(t_1\right)\right) \le p_1 < p_2 < 1$, Proposition 3.1(b) states that $b_1^* = b_2^* = 0$ and $T_{p_1}\left(0\right) < T_{p_2}\left(0\right)$.

(ii) When $p_1 < 1-\exp\left(-H\left(t_1\right)\right) < p_2 < 1$, $0 = b_2^* < b_1^*$. Moreover, Equation (3.6) implies $T_{p_1}\left(b_1^*\right) + b_1^* < t_1$, while Equation (3.5) implies $T_{p_2}\left(0\right) > t_1$. Therefore, $T_{p_1}\left(b_1^*\right) < T_{p_2}\left(0\right)$.

(iii) When $0 < p_1 < p_2 < 1-\exp\left(-H\left(t_1\right)\right)$, we have

$$-\ln\left(1-p_1\right) = H\left(T_{p_1}\left(b_1^*\right) + b_1^*\right) - H\left(b_1^*\right) < H\left(T_{p_2}\left(b_2^*\right) + b_2^*\right) - H\left(b_2^*\right) = -\ln\left(1-p_2\right),$$

$$h\left(T_{p_2}\left(b_2^*\right) + b_2^*\right) = h\left(b_2^*\right) > h\left(b_1^*\right) = h\left(T_{p_1}\left(b_1^*\right) + b_1^*\right).$$

Because $b_1^*, b_2^* < \upsilon$ and $T_{p_2}\left(b_2^*\right) + b_2^*, T_{p_1}\left(b_1^*\right) + b_1^* > \upsilon$, it follows that $b_1^* > b_2^*$ and $T_{p_2}\left(b_2^*\right) + b_2^* > T_{p_1}\left(b_1^*\right) + b_1^*$. Thus, $T_{p_1}\left(b_1^*\right) < T_{p_2}\left(b_2^*\right)$.

For all the above cases, the continuity of $h(\cdot)$ implies that when $p_1 \to p_2$, $b_1^* \to b_2^*$ and $T_{p_1}\left(b_1^*\right) \to T_{p_2}\left(b_2^*\right)$. The scenario where $c_1 \ge c_2$ can be proven in a similar way. This establishes the proposition.

■

Proposition 3.2 describes how the maximum PRL-*p* value $T_p\left(t_p^*\right)$ and the corresponding change point t_p^* behave when *p* varies. Since t_p^* is decreasing in *p* while $T_p\left(t_p^*\right)$ is increasing in *p*, it implies that as *p* decreases, the optimal burn-in duration increases. This is because more latent defects must be precipitated during burn-in to achieve a higher reliability; so a longer burn-in duration is needed to improve the screening strength. On the other hand, if one allows for more field returns, the maximum allowable warranty period increases; i.e. $T_p\left(t_p^*\right)$ is an increasing function of *p*. This presents the trade-off between choice of the reliability target, the burn-in duration and the duration of warranty period. Further development of a product may be considered if, after some testing, the result indicates that the maximum allowable warranty period is too short for a pre-specified risk level. This motivates the next section that deals with parametric estimation.

3.3 Parametric Inference and the Limiting Distribution

Statistical inferences for this change point were first presented by Launer (1993), in which he proposed a graphical technique to compute the change point of the PRL-*p* function based on nonparametric estimate of the FR function. However, the proposed graphical method in Launer (1993), though intuitive, has some drawbacks:

- There is no well-developed statistical theory for determining the small sample or asymptotic properties (Murthy *et al.* 2004).
- For moderate data size, the nonparametric estimate of the FR function may not be stable enough to exhibit a bathtub shape, under which the graphical approach would fail.

In adopting PRL-*p* as a burn-in criterion in the face of limited burn-in data, parametric inferences for change points of their PRL-*p* functions are essential to provide better estimates and to quantify the associated sampling risk. In view of the availability of a wide range of distributions with BTFRs, e.g. some generalized Weibull distributions reviewed by Murthy *et al.* (2004) and Pham and Lai (2007), we give the point and interval estimates for the change point and the associated maximum PRL-*p*, assuming that some of these distributions provide a good fit to the test data.

3.3.1 MLE for the Change Points

Denote $\Omega \subset \Re^k$ as the parameter space of the parametric BTFR distribution family *F*. The expression for the PRL-*p* function can be obtained by solving (3.3). We shall begin with the MLE $\hat{\Theta}_n$ of the parameter vector $\Theta \in \Omega$, where *n* is the sample size. By the invariance property of MLEs, we can substitute the estimated parameters $\hat{\Theta}_n$ into the corresponding PRL-*p* function and then optimize the PRL-*p* function to obtain the MLE of t_p^*.

$$\hat{t}_n^* \equiv \hat{t}_{p,n}^* = \arg\max_t T_p\left(\hat{\Theta}_n, t\right). \tag{3.7}$$

Substituting $\hat{t}_n^*$ and $\hat{\Theta}_n$ into the PRL-*p* function yields the MLE of $T_p\left(\hat{\Theta}_n, \hat{t}_n^*\right)$.

3.3.2 Asymptotic Distributions

After obtaining the MLEs $\hat{t}_n^*$ and $T_p\left(\hat{\Theta}_n, \hat{t}_n^*\right)$ for the change point and the associated maximum PRL-*p*, it is a common practice to construct a confidence interval for t_p^* and $T_p\left(t_p^*\right)$. When the sample size is relatively large, the desired confidence intervals can be constructed via the asymptotic distribution of $\hat{t}_n^* - t_p^*$ and $T_p\left(\hat{\Theta}_n, \hat{t}_n^*\right) - T_p\left(t_p^*\right)$. The asymptotic results require that $T_p(t)$ is differentiable. Checking the differentiability of $T_p(t)$ is somewhat complicated. We find that this problem can be simplified to that of checking the continuity and differentiability of the FR function. Previously, some authors assumed both the continuity of $h(t)$ and the differentiability of $T_p(t)$, e.g. Theorem 3 in Joe and Proschan (1984) and Theorem 1 in Launer (1993). Here, we show in the following lemma that continuity of $h(t)$

implies differentially of the PRL-p function $T_p(t)$.

Lemma 3.1. Consider a distribution with support $[0,W)$, $0<W\le\infty$. When its FR $h(\cdot)$ exists and greater than 0 $\forall t\in[0,W)$, the PRL-p function $T_p(\cdot)$ is continuous within the support. Moreover, if $h(\cdot)$ is continuous over $[0,W)$, $T_p(\cdot)$ is differentiable with derivative

$$\frac{d}{dt}T_p(t)=\frac{h(t)}{h\left(t+T_p(t)\right)}-1. \tag{3.8}$$

Proof: When $h(t)$ exists and is greater than 0, $R(\cdot)$ is continuous and is strictly decreasing over $[0,W)$. Therefore, $R^{-1}(\cdot)$ is also continuous and strictly decreasing over $[0,W)$. By applying these results and re-writing Equation (3.3) as

$$T_p(t)=R^{-1}\left((1-p)R(t)\right)-t, \tag{3.9}$$

the continuity of $T_p(\cdot)$ can be readily obtained.

If, in addition, $h(t)$ is continuous, then $R(t)$ is differentiable and its derivative $-f(t)$ is not equal to 0. Therefore, $R^{-1}(t)$ is differentiable at t, and thus $T_p(t)$. Differentiating Equation (3.9) w.r.t. t yields

$$\frac{d}{dt}T_p(t)=\frac{(1-p)f(t)}{f\left(R^{-1}\left((1-p)R(t)\right)\right)}-1=\frac{(1-p)h(t)R(t)}{h\left[R^{-1}\left((1-p)R(t)\right)\right]\cdot R\left[R^{-1}\left((1-p)R(t)\right)\right]}-1.$$

Since $R\left(R^{-1}(u)\right)=u$ and $(1-p)R(t)=R\left(T_p(t)+t\right)$, the above equation can be simplified to Equation (3.8).

■

Equation (3.8) has also been derived by Joe and Proschan (1984), but we derive it in a different way and under milder conditions. More importantly, Our results also indicate that the differentiability of $T_p(t)$ w.r.t. t depends only on the continuity of $h(t)$, while the existence of its $(n+1)$st order derivative depends on the nth order differentiability of $h(t)$. This result is very useful in checking the conditions in the asymptotic theorem. The conditions for Theorem 3.1 are stated as follows.

Conditions

(1) The MLE $\hat{\Theta}_n$ is a best asymptotically normal (BAN) estimator of Θ, i.e. $\sqrt{n}\left(\hat{\Theta}_n - \Theta\right) \xrightarrow{L} N\left(0, I^{-1}(\Theta)\right)$, where $I(\Theta)$ is the Fisher information matrix.

(2) $T_p(\Theta, t)$ is upside-down bathtub-shaped and $t_p^* \equiv t_p^*(\Theta) > 0$.

(3) The first three order partial derivatives of $T_p(\Theta, t)$ w.r.t. t, denoted by $g(\Theta,t) \equiv \frac{\partial}{\partial t} T_p(\Theta,t), g'(\Theta,t) \equiv \frac{\partial^2}{\partial t^2} T_p(\Theta,t)$ and $g''(\Theta,t) \equiv \frac{\partial^3}{\partial t^3} T_p(\Theta,t)$ exist. In addition, $g'(\Theta,t)$ is continuous in Θ.

(4) The partial derivative of $g(\Theta,t)$ w.r.t. Θ, denoted by $C(\Theta,t) = \frac{\partial}{\partial \Theta} g(\Theta,t)$, is continuous in t.

BAN properties of $\hat{\Theta}_n$ for BTFR distributions, although not the focus of this chapter, have been well established in theory or by simulation, e.g. see Gupta and Kundu (2001) for the exponentiated Weibull, Tang *et al.* (2003) for the modified Weibull extension and Bebbington *et al.* (2008) for the modified Weibull distribution. For the BTFR distributions reviewed by Pham and Lai (2007), we can verify that the differentiability of $g(\Theta,t)$ is also reasonable. Moreover, by applying Lemma 3.1, checking the differentiability of $g(\Theta,t)$ simplifies to verifying differentiability of the FR function. These results justify Conditions (1), (3) and (4). The asymptotic distributions are given in the following theorem.

Theorem 3.1. Suppose that conditions (1)-(4) hold.

- The asymptotic distribution of $\sqrt{n}\left(\hat{t}_n^{*}-t_p^{*}\right)$, as $n\to\infty$, is normal with mean zero and variance $\dfrac{C\left(\Theta,t_p^{*}\right)I^{-1}(\Theta)C^{T}\left(\Theta,t_p^{*}\right)}{\left[g'\left(\Theta,t_p^{*}\right)\right]^{2}}$, i.e.

$$\sqrt{n}\left(\hat{t}_n^{*}-t_p^{*}\right)\xrightarrow{L}N\left(0,\frac{C\left(\Theta,t_p^{*}\right)I^{-1}(\Theta)C^{T}\left(\Theta,t_p^{*}\right)}{\left[g'\left(\Theta,t_p^{*}\right)\right]^{2}}\right). \tag{3.10}$$

- Denote $B(\Theta,t)=\dfrac{\partial T_p(\Theta,t)}{\partial\Theta}$. The asymptotic distribution of $\sqrt{n}\left[T_p\left(\Theta,t_p^{*}\right)-T_p\left(\hat{\Theta}_n,\hat{t}_n^{*}\right)\right]$ is normal with mean zero and variance $B\left(\Theta,t_p^{*}\right)I^{-1}(\Theta)B^{T}\left(\Theta,t_p^{*}\right)$, i.e.

$$\sqrt{n}\left[T_p\left(\hat{\Theta}_n,\hat{t}_n^{*}\right)-T_p\left(\Theta,t_p^{*}\right)\right]\xrightarrow{L}N\left(0,B\left(\Theta,t_p^{*}\right)I^{-1}(\Theta)B^{T}\left(\Theta,t_p^{*}\right)\right) \tag{3.11}$$

Proof: Note that

$$t_p^{*}=\max\left\{T_p(\Theta,t),t\geq 0\right\} \text{ and } \hat{t}_n^{*}=\max\left\{T_p\left(\hat{\Theta}_n,t\right),t\geq 0\right\}.$$

Based on Condition (2), we know that for sufficiently large *n*,

$$g\left(\Theta,t_p^{*}\right)=0,\ t_p^{*}>0 \text{ and } g\left(\hat{\Theta}_n,\hat{t}_n^{*}\right)=0,\ \hat{t}_n^{*}>0.$$

The first order Taylor series expansion with Cauchy form for the remainder is as follows.

$$0=g\left(\hat{\Theta}_n,\hat{t}_n^{*}\right)=g\left(\hat{\Theta}_n,t_p^{*}\right)+\left[g\left(\hat{\Theta}_n,\hat{t}_n^{*}\right)-g\left(\hat{\Theta}_n,t_p^{*}\right)\right]=g\left(\hat{\Theta}_n,t_p^{*}\right)+g'\left(\hat{\Theta}_n,t_p^{*}\right)\cdot\left(\hat{t}_n^{*}-t_p^{*}\right)+R_2\cdot\left(\hat{t}_n^{*}-t_p^{*}\right)$$

Note that for some ξ, $\left|\xi-t_p^{*}\right|\leq\left|\hat{t}_n^{*}-t_p^{*}\right|$, R_2 satisfies

$$R_2=\frac{1}{2}\left(\hat{t}_n^{*}-t_p^{*}\right)g''\left(\hat{\Theta}_n,\xi\right)\leq\frac{1}{2}\sup\left\{\left|g'\left(\hat{\Theta}_n,u\right)-g'\left(\hat{\Theta}_n,t_p^{*}\right)\right|:\left|u-\tau^{*}\right|\leq\left|\hat{t}_n^{*}-t_p^{*}\right|\right\}.$$

Re-arranging this equation yields

$$-\sqrt{n}g\left(\hat{\Theta}_n, t_p^*\right) = g'\left(\hat{\Theta}_n, t_p^*\right)\cdot\sqrt{n}\left(\hat{t}_n^* - t_p^*\right) + R_2\cdot\sqrt{n}\left(\hat{t}_n^* - t_p^*\right).$$

We shall look at each term in this equation.

The existence of the partial derivative of $g\left(\Theta, t\right)$ w.r.t. Θ indicates that $g\left(\Theta, t\right)$ is continuous in Θ. Therefore,

$$\sqrt{n}g\left(\hat{\Theta}_n, t^*\right) = \sqrt{n}\left[g\left(\hat{\Theta}_n, t_p^*\right) - g\left(\Theta, t_p^*\right)\right] \xrightarrow{L} N\left(0, C\left(\Theta, t_p^*\right)I^{-1}(\Theta)C^T\left(\Theta, t_p^*\right)\right).$$

Condition (3) implies that

$$g'\left(\hat{\Theta}_n, t_p^*\right) = g'\left(\Theta, t_p^*\right) + \left(g'\left(\hat{\Theta}_n, t_p^*\right) - g'\left(\Theta, t_p^*\right)\right) \xrightarrow{p} g'\left(\Theta, t_p^*\right).$$

Because $g''\left(\Theta, t\right)$ exists, $g'\left(\Theta, t\right)$ is a continuous function of *t*. We have

$$\sup\left\{\left|g'\left(\hat{\Theta}_n, u\right) - g'\left(\hat{\Theta}_n, t_p^*\right)\right| : \left|u - t_p^*\right| \le \left|\hat{t}_n^* - t_p^*\right|\right\} \xrightarrow{p} 0.$$

By Slutsky's theorem, we therefore have the following asymptotic property

$$g'\left(\Theta, t_p^*\right)\cdot\sqrt{n}\left(\hat{t}_n^* - t_p^*\right) \xrightarrow{L} N\left(0, C\left(\Theta, t_p^*\right)I^{-1}(\Theta)C^T\left(\Theta, t_p^*\right)\right).$$

Re-arranging this asymptotic equation yields Equation (3.10).

To prove Equation (3.11), rewrite $\sqrt{n}\left[T_p\left(\Theta, t_p^*\right) - T_p\left(\hat{\Theta}_n, \hat{t}_n^*\right)\right]$ as follows.

$$\sqrt{n}\left[T_p\left(\Theta, t_p^*\right) - T_p\left(\hat{\Theta}_n, \hat{t}_n^*\right)\right] = \sqrt{n}\left[T_p\left(\Theta, t_p^*\right) - T_p\left(\Theta, \hat{t}_n^*\right)\right] + \sqrt{n}\left[T_p\left(\Theta, \hat{t}_n^*\right) - T_p\left(\hat{\Theta}_n, \hat{t}_n^*\right)\right] \quad \text{(a6)}$$

We shall examine each term in the right-hand side. Condition (4) ensures that $B\left(\Theta, t\right)$ exists and is continuous in *t*. Writing in the first order Taylor series expansion with Peano's remainder term and using the fact that $\hat{t}_n^*$ is a consistent estimator, we have

$$\sqrt{n}\left[T_p\left(\Theta,\hat{t}_n^*\right)-T_p\left(\hat{\Theta}_n,\hat{t}_n^*\right)\right]=B\left(\Theta,\hat{t}_n^*\right)\sqrt{n}\left(\Theta-\hat{\Theta}_n\right)+o\left(\sqrt{n}\left(\Theta-\hat{\Theta}_n\right)\right)$$
$$\xrightarrow{L} N\left(0, B\left(\Theta,t_p^*\right)I^{-1}\left(\Theta\right)B^T\left(\Theta,t_p^*\right)\right).$$

On the other hand, from Equation (3.10) and the Slutsky's theorem, we have

$$\sqrt{n}\left[T_p\left(\Theta,t_p^*\right)-T_p\left(\Theta,\hat{t}_n^*\right)\right]=g\left(\Theta,t_p^*\right)\cdot\sqrt{n}\left(t_p^*-\hat{t}_n^*\right)+o\left[\sqrt{n}\left(t_p^*-\hat{t}_n^*\right)\right]\xrightarrow{p}0.$$

Substituting these results into Equation (a6) yields Equation (3.11).

■

Since $\hat{t}_n^*$, $\hat{\Theta}_n$ and the observed information matrix $I_{obs}\left(\hat{\Theta}_n\right)$ are respectively consistent estimators of t_p^*, Θ and $n\cdot I\left(\Theta\right)$, we can substitute them into the asymptotic variances to obtain the "observed variances". The asymptotic results based on these observed variances still hold because based on Conditions (3) and (4), g' and *C* are continuous in neighborhoods of their supports, respectively. The observed variances simplify computation of the confidence intervals, and often there are also theoretical reasons to prefer them (e.g. Lindsay and Li 1997).

Given Θ and *p*, the exact values of $T_p\left(t_p^*\right)$ and t_p^* can be computed and used as the warranty period and the burn-in duration, respectively. In practice, Θ is estimated from test data. To account for the sampling risk, the one-sided lower confidence limit of $T_p\left(t_p^*\right)$ can be used for the warranty period to ensure that the average warranty return is not higher than a given probability, while the upper confidence limit of t_p^* is suggested as the burn-in duration to protect against insufficient screening strength. When the sample size is extremely small, confidence intervals based on large sample normal approximation may not work well. Under this scenario, the bootstrap method is recommended by some authors (e.g., Efron and Tibshirani 1993).

The expressions of the asymptotic distributions (3.10) and (3.11) depend on specific forms of the BTFR distributions. In the next section, we investigate the PRL-*p* functions of some generalized Weibull distributions, and derive the corresponding asymptotic distributions for the modified Weibull extension based on Theorem 3.1.

3.4 Application to Some Generalized Weibull Models with BTFR

In the last two decades, many simple but yet useful distributions with BTFR functions have been proposed and studied in-depth in the literature, e.g. the exponentiated Weibull (Mudholkar *et al.* 1995), the modified Weibull extension (Xie *et al.* 2002) and the modified Weibull distribution (Lai *et al.* 2003). The PRL-*p* functions for these distributions have closed form expressions.

3.4.1 PRL-*p* Functions for Some Generalized Weibull Models

Table 3.1 lists the PRL-*p* functions of some generalized Weibull distributions. The conditions under which these distributions have BTFR can be found from Pham and Lai (2007) or from papers cited in the first column of Table 3.1. The change points of these PRL-*p* functions can be readily computed through simple search algorithms. Note that $\Psi(\cdot)$ for the Modified Weibull distribution denotes the Lambert-W function, which is the inverse function of

$$g(x) = x\exp(x).$$

Based on Equation (3.8), it is straightforward, although tedious, to verify that the required conditions (2)-(4) are satisfied for these distributions when their PRL-*p* functions have upside-down bathtub shapes. Nevertheless, the methods developed in this study are also applicable to other BTFR distributions if we can verify that Condition (1)-(4) are met. To demonstrate the applicability of Theorem 3.1, we analyze the modified Weibull extension in-depth. Other distributions can be analyzed in a similar way.

Table 3.1. PRL-*p* functions for some generalized Weibull models

Distributions	Reliability function	PRL-*p* function
Exponential power model (Smith and Bain 1975)	$\exp\left\{1-\exp\left[\lambda t^{\alpha}\right]\right\}$	$\left\{\frac{\ln\left[\exp\left(\lambda t^{\alpha}\right)-\ln\left(1-p\right)\right]}{\lambda}\right\}^{1/\alpha}-t$
Modified Weibull distribution (Lai *et al.* 2003)	$\exp\left\{-at^{b}\exp\left(\lambda t\right)\right\}$	$\frac{b}{\lambda}\Psi\left(\frac{b}{\lambda}\left[t^{b}\exp(\lambda t)-\frac{\ln(1-p)}{a}\right]^{1/b}\right)-t$
Two-parameter model (Chen 2000)	$\exp\left\{-\lambda\left[\exp\left(t^{\beta}\right)-1\right]\right\}$	$\left\{\ln\left[\exp\left(t^{\beta}\right)-\frac{\ln\left(1-p\right)}{\lambda}\right]\right\}^{1/\beta}-t$
Modified Weibull extension (Xie *et al.* 2002)	$\exp\left\{-\lambda\alpha\left[\exp\left(\left(\frac{t}{\alpha}\right)^{\beta}\right)-1\right]\right\}$	$\alpha\left\{\ln\left[\exp\left(\left(\frac{t}{\alpha}\right)^{\beta}\right)-\frac{\ln\left(1-p\right)}{\lambda\alpha}\right]\right\}^{1/\beta}-t$
Generalized power Weibull (Bagdonavicius and Nikulin 2002)	$\exp\left\{-\left[\left(1+\left(t/\alpha\right)^{\beta}\right)^{\theta}-1\right]\right\}$	$\alpha\left\{\left[\left(1+\left(t/\alpha\right)^{\beta}\right)^{\theta}-\ln\left(1-p\right)\right]^{1/\theta}-1\right\}^{1/\beta}-t$
Exponentiated Weibull (Mudholkar *et al.* 1995)	$1-\left\{1-\exp\left[-\left(t/\alpha\right)^{\beta}\right]\right\}^{\theta}$	$\alpha\left\{-\ln\left\{1-\left[\left(1-p\right)\left(1-\exp\left(-\left(t/\alpha\right)^{\beta}\right)\right)^{\theta}+p\right]^{1/\theta}\right\}\right\}^{1/\beta}-t$

3.4.2 The Case of The Modified Weibull Extension

The PDF and the FR function of the modified Weibull extension are

$$f\left(t\right)=\beta\lambda\left(t/\alpha\right)^{\beta-1}\exp\left[\left(t/\alpha\right)^{\beta}\right]\exp\left\{-\lambda\alpha\left[\exp\left(\left(\frac{t}{\alpha}\right)^{\beta}\right)-1\right]\right\}, \qquad (3.12)$$

$$h\left(t\right)=\beta\lambda\left(\frac{t}{\alpha}\right)^{\beta-1}\exp\left[\left(\frac{t}{\alpha}\right)^{\beta}\right].$$

When β < 1, this distribution has a BTFR (Xie *et al.* 2002). Therefore, $\Omega=\left\{\left(\alpha,\beta,\lambda\right):\alpha,\lambda>0;\ 0<\beta<1\right\}$. Figure 3.1 depicts some FR functions and the corresponding PRL-0.01 functions for $\lambda = 0.05$, $\beta = 0.5$ and $\alpha = 1, 2, 3$.

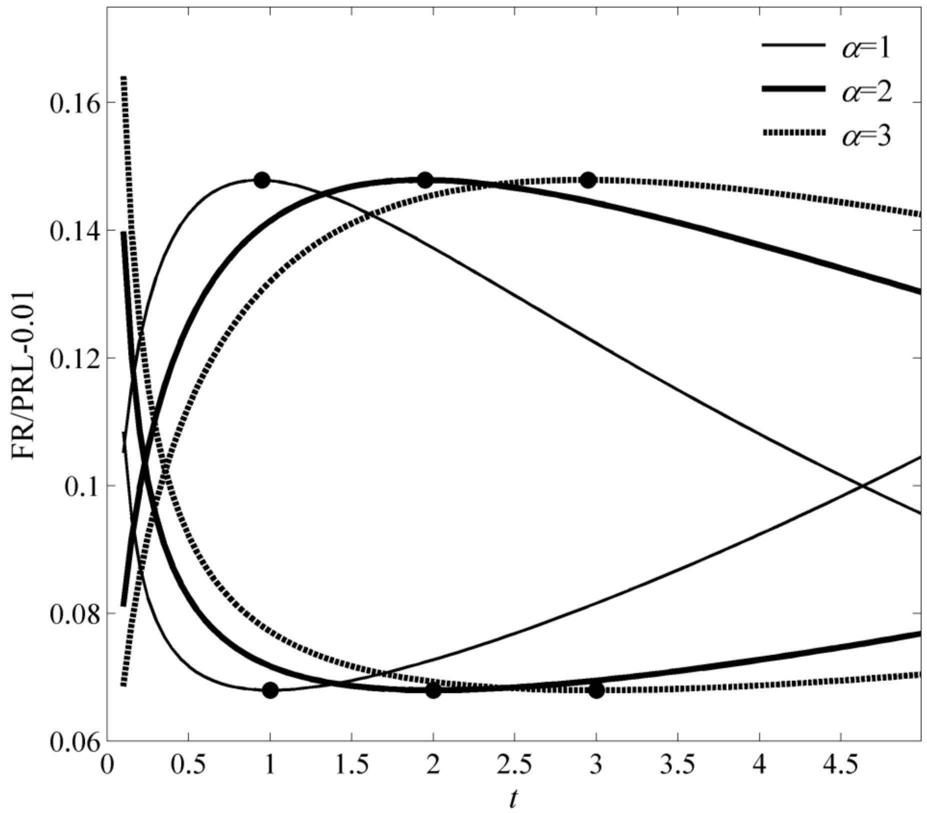


Figure 3.1. Typical PRL-*p* (upside-down bathtub) and FR (bathtub) curves with extreme points for modified Weibull extension distribution (λ = 0.05, β = 0.5, p = 0.01)

Define

$$\Lambda = \exp\left(\left(\frac{t}{\alpha}\right)^{\beta}\right) - \frac{\ln(1-p)}{\lambda\alpha}.$$

From Table 3.1, the PRL-*p* function is given by

$$T_p(\Theta, t) = \alpha(\ln\Lambda)^{1/\beta} - t.$$

In order to obtain the asymptotic variances of $\sqrt{n}\left[T_p\left(\hat{\Theta}_n, \hat{t}_n^*\right) - T_p\left(\Theta, t_p^*\right)\right]$ and $\sqrt{n}\left(\hat{t}_n^* - t_p^*\right)$, we need to derive explicit expressions for $B(\Theta,t)$, $C(\Theta,t)$, $g(\Theta,t)$, $g'(\Theta,t)$ and the Fisher information matrix $I(\Theta)$ in Equations (3.10) and (3.11). The Fisher information matrix for this distribution has been derived in Tang *et al.* (2003) under general Type-II censoring. Define $\Lambda_x = \frac{\partial \Lambda}{\partial x}$ and $\Lambda_{xy} = \frac{\partial^2 \Lambda}{\partial x \partial y}$, where *x* and *y* can be either *α, β, λ* or *t*. The formulas for Λ_x and Λ_{xy} can be directly computed. Expressions for the first order derivatives are given by

$$
\begin{aligned}
\Lambda_t &= \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1}\exp\left(\left(\frac{t}{\alpha}\right)^{\beta}\right) \\
\Lambda_\alpha &= \frac{\ln(1-p)}{\lambda\alpha^2} - \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta}\exp\left(\left(\frac{t}{\alpha}\right)^{\beta}\right) \\
\Lambda_\beta &= \left(\frac{t}{\alpha}\right)^{\beta}\exp\left(\left(\frac{t}{\alpha}\right)^{\beta}\right)\ln\left(\frac{t}{\alpha}\right) \\
\Lambda_\lambda &= \frac{\ln(1-p)}{\lambda^2\alpha}
\end{aligned}.
$$

Based on the first order derivatives, the second order derivatives can be readily obtained.

$$
\begin{aligned}
\Lambda_{tt} &= \left(\frac{\beta t^{\beta-1}}{\alpha^{\beta}}\right)^2\exp\left(\left(\frac{t}{\alpha}\right)^{\beta}\right) + \frac{\beta(\beta-1)}{\alpha^2}\left(\frac{t}{\alpha}\right)^{\beta-2}\exp\left(\left(\frac{t}{\alpha}\right)^{\beta}\right) \\
\Lambda_{t\alpha} &= -\frac{\beta^2 t^{\beta-1}}{\alpha^{\beta+1}}\exp\left(\left(\frac{t}{\alpha}\right)^{\beta}\right) - \frac{\beta^2}{\alpha^2}\left(\frac{t}{\alpha}\right)^{2\beta-1}\exp\left(\left(\frac{t}{\alpha}\right)^{\beta}\right) \\
\Lambda_{t\beta} &= \frac{t^{\beta-1} + \beta t^{\beta-1}\ln(t/\alpha)}{\alpha^{\beta}}\exp\left(\left(\frac{t}{\alpha}\right)^{\beta}\right) + \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{2\beta-1}\exp\left(\left(\frac{t}{\alpha}\right)^{\beta}\right)\ln\left(\frac{t}{\alpha}\right) \\
\Lambda_{t\lambda} &= 0
\end{aligned}
$$

Then, $g(\Theta,t)$ and $g'(\Theta,t)$ can be obtained as

$$
g(\Theta,t) = \frac{\alpha\Lambda_t}{\beta\Lambda}(\ln\Lambda)^{1/\beta-1} - 1,
$$

$$
g'(\Theta,t) = \frac{\alpha}{\beta}\frac{\Lambda_{tt}\Lambda - \Lambda_t^2}{\Lambda^2}(\ln\Lambda)^{1/\beta-1} + \frac{\alpha(1-\beta)}{\beta^2}\frac{\Lambda_t^2}{\Lambda^2}(\ln\Lambda)^{1/\beta-2}.
$$

We write $B(\Theta,t) = (B_\alpha, B_\beta, B_\lambda)$, where the entries are given by

$$B_\alpha = (\ln\Lambda)^{1/\beta} + \frac{\alpha\Lambda_\alpha}{\beta\Lambda}(\ln\Lambda)^{1/\beta-1},$$

$$B_\beta = \frac{\alpha\Lambda_\beta}{\beta\Lambda}(\ln\Lambda)^{1/\beta-1} - \frac{\alpha\ln(\ln\Lambda)}{\beta^2}(\ln\Lambda)^{1/\beta},$$

$$B_\lambda = \frac{\alpha\Lambda_\lambda}{\beta\Lambda}(\ln\Lambda)^{1/\beta-1}.$$

Similarly, we can derive the expressions for $C(\Theta,t) = (C_\alpha, C_\beta, C_\lambda)$ as

$$C_\alpha = \frac{\alpha\Lambda_{t\alpha}\Lambda + \Lambda_t\Lambda - \alpha\Lambda_t\Lambda_\alpha}{\beta\Lambda^2}(\ln\Lambda)^{1/\beta-1} + \frac{\alpha\Lambda_t\Lambda_\alpha(1-\beta)}{\beta^2\Lambda^2}(\ln\Lambda)^{1/\beta-2},$$

$$C_\beta = \frac{\alpha\beta\Lambda\Lambda_{t\beta} - \alpha\Lambda_t(\Lambda+\beta\Lambda_\beta)}{\beta^2\Lambda^2}(\ln\Lambda)^{1/\beta-1} + \frac{\alpha\Lambda_t}{\beta^2\Lambda}(\ln\Lambda)^{1/\beta-2}\left[\frac{(1-\beta)\Lambda_\beta}{\Lambda} - \frac{(\ln\Lambda)\cdot\ln(\ln\Lambda)}{\beta}\right],$$

$$C_\lambda = \frac{\alpha}{\beta}\frac{\Lambda_{t\lambda}\Lambda - \Lambda_\lambda\Lambda_t}{\Lambda^2}(\ln\Lambda)^{1/\beta-1} + \frac{\alpha(1-\beta)\Lambda_t\Lambda_\lambda}{\beta^2\Lambda^2}(\ln\Lambda)^{1/\beta-2}.$$

After $\hat{\Theta}_n$ and $\hat{t}_n^*$ are estimated, substituting the above results into Equations (3.10) and (3.11) yields the observed variances. Note that $B(\Theta,t)$ and $g(\Theta,t)$ are gradients of $T_p(\Theta,t)$, while $g'(\Theta,t)$ and $C(\Theta,t)$ are gradients of $g(\Theta,t)$. Alternatively, their numerical values can be easily computed from some software packages, e.g. the GRADIENT function in Matlab®.

3.5 An Illustrative Example

To illustrate some of the above results, the car engine failure data from Xie and Lai (1996) are used for demonstration. A motor company developed a new engine for its cars. To assess the reliability of this engine, 311 units are subject to test during a unit testing phase. Failures of the units are intermittently inspected until all units are failed. The time scale used throughout this example is the accumulated mileage in thousands of kilometers ($\times 10^3$ km). For confidentiality reasons, the data have been scaled. The group censoring data are given in

Table 3.2.

Table 3.2. Lifetime data for a group of 311 units of a new engine

Time interval	1	2	3	4	5	6	7	8	9
No. of failures	53	29	29	36	13	25	22	16	18
Time interval	10	11	12	13	14	15	16	17	18
No. of failures	8	22	11	13	5	5	4	1	1

As can be seen from Table 3.2, there are a substantial number of early failures for this product. This is why in practice, most engines should go through a diagnostic test after assembly. Xie and Lai (1996) had shown that the estimated FR is bathtub-shaped. However, due to the problem of limited data, standard errors for the estimated FRs are rather large. Analysis based on the estimated FRs should be tentative. Therefore, large estimation error is expected when applying the graphical method proposed by Launer (1993). In this section, we shall use parametric methods to analyze this dataset and estimate the change points of the PRL-*p* functions.

In order to specify a best-fit distribution for this dataset, we apply the Kaplan-Meier (KM) method, the modified Weibull distribution (MWD), the modified Weibull extension (MWE) and the generalized power Weibull (GPW) to fit the data. The reliability functions of these distributions can be found in Table 3.1. The estimated CDF by using these methods are compared in Figure 3.2.

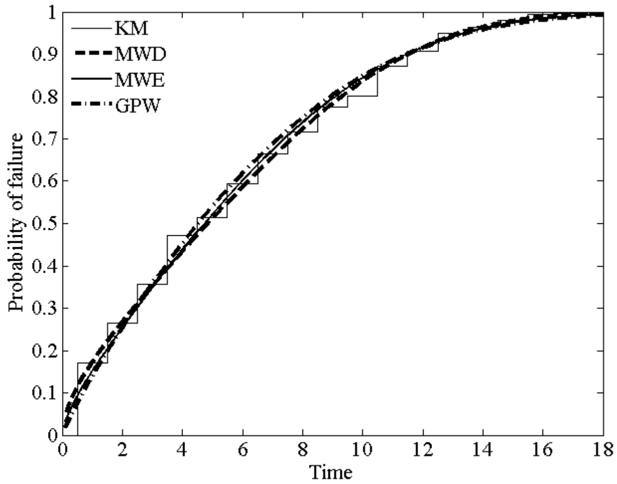


Figure 3.2. The estimated CDF by the Kaplan-Meier (KM) method, the modified Weibull distribution (MWD), the modified Weibull extension (MWE) and the generalized power Weibull (GPW)

It is clear from Figure 3.2 that all three models fit the data well. For illustrative purpose, the modified Weibull extension is chosen in the following analysis. The distribution of the modified Weibull extension is given by (3.12). MLEs of the parameters and the associated standard error are respectively given by

$$\hat{\alpha} = 2.412\ (2.48),\ \hat{\beta} = 0.593\ (0.16),\ \hat{\lambda} = 0.0835\ (0.009).$$

To demonstrate the goodness of fit, the empirical FR and the estimated FR are depicted in Figure 3.3. It can be seen that these two FR curves tally reasonably well.

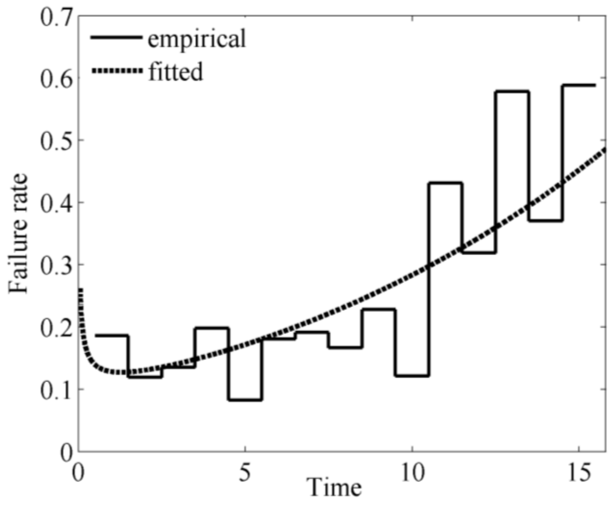


Figure 3.3. The empirical failure rate and the estimated failure rate by using the modified Weibull extension

MLEs for change points of the PRL-p function and the corresponding PRL-p can be readily computed. The confidence intervals for the change points and the corresponding PRL-p can be constructed based on the asymptotic distributions (3.10) and (3.11) or using the parametric/nonparametric bootstrap methods. The asymptotic variances can be computed based on the results in the previous section. Numerical value of the hessian matrix, which is the inverse of the fisher information matrix for these parameters, is directly provided by the software, and is used for computing these asymptotic variances. For the parametric bootstrap, 1,000 bootstrap samples are generated. The results are visualized in Figure 3.4.

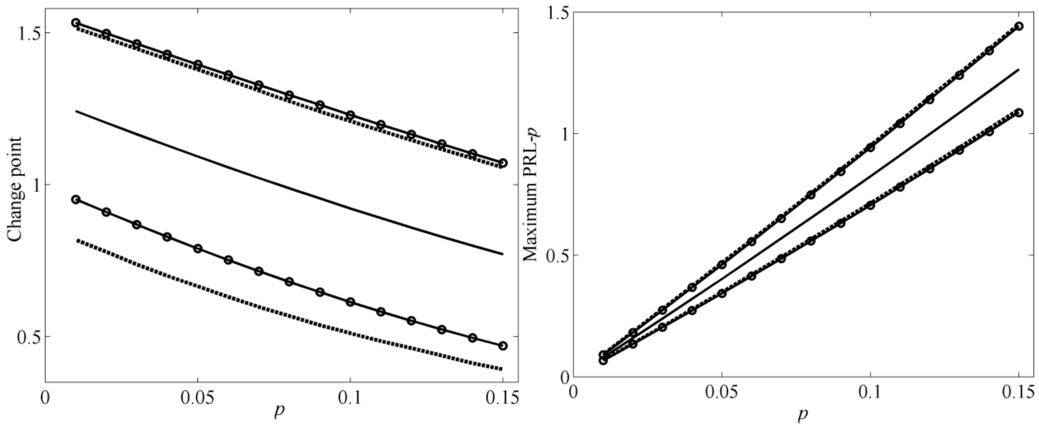


Figure 3.4. The MLE and 90% confidence bands for t_p^* and $T_p\left(\Theta, t_p^*\right)$ under different values of p: The bold line is for point estimation, the dash lines are for the parametric bootstrap confidence band and the lines with circles are for confidence band based on asymptotic distributions.

Figure 3.4 shows the estimated change points of the PRL-p function for different values of p and the 90% confidence bands using the asymptotic result given by Equations (3.10) and (3.11), as well as the confidence bands using parametric bootstrapping method. We can see that the estimated $\hat{t}_{p,n}^*$ is decreasing in p, while the corresponding maximum PRL-p is increasing in p, which is concordant with Proposition 3.2. The confidence bands based on these two approaches agree quite well, although the asymptotic method tends to yield a slightly narrower band. This is reasonable as the asymptotic variances are expected to be lower bounds for variances of the estimators. This example indicates that the asymptotic distribution in Theorem 3.1 is accurate enough for constructing the confidence intervals. Construction of confidence intervals based on the asymptotic method requires much less computation time compared with the bootstrapping method, and is thus recommended.

With the previous results, various combinations of $\left(\overline{t_p^*}, \underline{T_p}\left(t_p^*\right)\right)$ versus different values of p are calculated and presented in Table 3.3. Here we choose $\overline{t_p^*}$ to be the 90% upper confidence limit of $\hat{t}_{p,n}^*$ and $\underline{T_p}\left(t_p^*\right)$ be the 90% lower limit of $T_p\left(\hat{\Theta}_n, \hat{t}_{p,n}^*\right)$.

Table 3.3. The estimated change points and the corresponding RPL-p for different values of p

p	0.01	0.05	0.07	0.08	0.09	0.1	0.11	0.12	0.13	0.14	0.15	0.2
$\underline{T_p}\left(t_p^*\right)$	0.070	0.36	0.51	0.58	0.66	0.73	0.81	0.89	0.97	1.05	1.23	1.54
$\overline{t_p^*}$	1.46	1.32	1.25	1.22	1.19	1.16	1.12	1.09	1.06	1.03	1.00	0.85

The combinations of $\left(\overline{t_p^*}, \underline{T_p}\left(t_p^*\right)\right)$ in Table 3.3 provide the manufacturers with a number of choices, which is not possible if the burn-in criteria were solely based on probability of failure or mean number of failures. When the estimated maximum PRL-*p* is used as the maximum allowable warranty period, the proportion of field returns is known, and thus the users' risk is controllable. After the related costs are estimated, the manufacturer may choose the preferred combination of the conservative burn-in duration $\overline{t_p^*}$ and the warranty period $\underline{T_p}\left(t_p^*\right)$ from Table 3.3. Note that even though the burn-in duration provided by Table 3.3 appears to be quite long, in practice, the operational burn-in duration could be shortened under accelerated test environments.

3.6 Conclusions

Motivated by the need for simultaneous determination of burn-in duration and warranty period, this chapter investigated a burn-in criterion based on the PRL-*p* function. It was shown that the PRL-*p* function is either decreasing or upside-down bathtub shaped for distributions having BTFRs. This gives rise to a natural way to control the proportion of warranty claims when the burn-in duration is set to the change point of the PRL-*p* function, as it attains the maximum value corresponding to the maximum allowable warranty period. Such an approach has advantages over the conventional ones as it provides a set of choices for the warranty period and the associated burn-in duration by varying the pre-specified proportion of warranty returns. The conditions under which the change point exists and the properties of this change point were investigated in depth. Notably, the change point is decreasing in p while the associated PRL-*p* is increasing in p, which indicates a trade-off between screening strength of the burn-in process, the reliability target and the warranty duration. Asymptotic distributions for MLEs of the change point of the PRL-*p* function and its corresponding maximum were established. The results were applied to some generalized Weibull models with BTFRs, and a numerical example was given for the modified Weibull extension.

CHAPTER 4 BURN-IN FOR PRODUCTS WITH A TWO-DIMENSIONAL WARRANTY

4.1 Introduction

As reviewed in the Chapter 2, the literature of burn-in modeling has focused on failures indexed by a single time scale. Generally speaking, these models can be classified into three classes. The first one is due to Kim and Kuo (2003, 2005, 2009, 2011), who analyzed a system at the component level and traced back all defects, i.e., component defects, component connection defects and series connection defects. The second class models the overall failure process with a bathtub failure rate assumption (e.g., Mi 1996, Cha 2001, Sheu and Chien 2005) or an eventually increasing failure rate assumption (e.g., Mi 2003, Cha 2006). The third class uses the bimodal distribution (e.g., Tseng and Peng 2004, Jiang and Jardine 2007, Cha and Finkelstein 2010b), which is more appropriate for non-repairable systems. The first class may not be suitable for complex systems sold with two-dimensional warranties, because a complex system, e.g., a car, contains thousands of different components. The defect rate for each component is almost impossible to estimate. The second class is also not appropriate for complex systems, because a complex system with infant mortality has at least two failure modes, i.e., aging failures and defect failures, each having different accelerated coefficients (c.f. Hong *et al.* 2009). In addition, all the existing burn-in models did not take the heterogeneous customer usage behavior into account. It is not uncommon to observe that for products sold with a one-dimensional warranty, e.g., laptops, cell phones, washing machines, etc., different users would have different usage rates. Ignorance of this fact will lead to an inferior burn-in decision and inaccurate cost estimation. More importantly, all these models did not deal with failures indexed by two time scales. For products sold with a two-dimensional warranty, failures need to be modeled as points in a two-dimensional plane with the two-axis representing age and usage. In terms of the ways that failure rate of an item is modeled, three different approaches have been used to model the two-dimensional failure processes (Jack *et al.* 2009), i.e., one-dimensional approach, two-dimensional approach and composite scale approach. These approaches differ from the current burn-in modeling techniques that focused on a single time scale.

This chapter thus introduces a new modeling approach by focusing on two types of failures. The normal failures are inherent due to product aging. By assuming minimal repair with repair time insignificant, normal failures are modeled as a non-homogeneous Poisson process (NHPP). The second type is defect failures due to latent defects such as cavities in the welding spots, electrostatic discharge in some components, component defects due to assembly errors, faults in the software, incompatibility issues, etc. When these kinds of defects surface as failures, it is reasonable to assume that they would be perfectly removed with some rectification efforts. Parallel to the research in software reliability (Xie 1991), the number of latent defects is modeled as a random variable. This modeling approach can be regarded as a compromise between the approach developed by Kim and Kuo (2003, 2005, 2009, 2011) and the bathtub failure rate approach, and has a number of merits.

- Many organizations adopt a modular approach to reuse proven modules in subsequent products (Turner 2010). Infant mortality rates for these modules, rather than a single component or the new product, are available from past experiences and data.
- Normal failures and defect failures have different accelerated relationship with respect to the usage rate, which can be easily reflected in our modeling approach.
- It can be extended to distinguish between different sources of latent defects, e.g. defects from software and hardware.
- Most previous models cannot yield the screening strength, which is a widely-used

index in industrial burn-in procedure. However, this index can be easily determined in our formulation.

Effects of usage rate on both failures are modeled by the accelerated failure time (AFT) model with different accelerated relationships. Based on these settings, the mean number of warranty claims and the expected total cost per burnt-in item can be computed by taking the consumer usage heterogeneity into consideration, after which both performance and cost-based burn-in models can be developed. In both models, there are two decision variables, i.e., burn-in duration and burn-in usage rate. It should be noted that the burn-in usage rate need not be always as high as possible, because the usage rate affects the burn-in costs, and it has different effects on the normal failure process and the defect failures.

In this chapter, necessary conditions under which the optimal burn-in usage rate should be as high as possible are investigated. The remainder of the chapter is organized as follows. Section 4.2 states the problem and presents the notations and model formulation. In Section 4.3, two burn-in models are developed and analyzed in detail. A numerical example is used to demonstrate our results in Section 4.4. Section 4.5 concludes the chapter.

4.2 Model Formulation

4.2.1 Modeling Customer Usage Rates and Product Failures

To model the failure, we use the one-dimensional approach where the usage rate is treated as a covariate, conditioning on which the two-dimensional failure process is effectively reduced to a one-dimensional one. To apply this approach, an assumption adopted here is that usage rate over time is constant for each customer and it varies across the customer population. Let Z be the random usage rate, $\Psi(z)$ be the cumulative distribution function (CDF) of Z, and z be a realization of Z. Further assume the manufacturer knows this distribution, either through information of previous vintages of products, or from a customer survey.

Two types of failures are considered. When there are no defects or faults, failures in a system are inherent, occurring over time due to aging, and are determined by the design decisions, which are called normal failures here. The other type is defect failures due to defects in hardware and faults in software. The number of defects in the system is modeled by a random variable M with probability mass function $\pi(\cdot)$. All defect failure times are assumed to be statistically independent and identical. For an item designed for some nominal usage rate z_0, the normal failure has a non-decreasing failure rate $\upsilon_0^n(t)$; while for a specific defect, the time till it surfaces is denoted by T_0^d with CDF $F_0^d(t)$. It is assumed that during both burn-in and field use, all normal failures are minimally repaired and all defect failures are perfectly repaired. As such, normal failures occur according to a non-homogeneous Poisson process (NHPP) with rate of occurrence of failure (ROCOF) $\upsilon_0^n(t)$; while a defect failure is perfectly removed through rectification action, resulting in the number of defects remaining getting reduced by one.

When the item is sold to a customer with usage rate $Z = z$, the normal failure has ROCOF $\upsilon_z^n(t)$ whilst the time till a defect surfaces is T_z^d. The effects of usage rate on both types of failures are modeled through the accelerated failure time (AFT) approach (Nelson 1990, Lawless et al. 2009) as:

$$\upsilon_z^n(t) = (z/z_0)^\gamma \cdot \upsilon_0^n\left(t(z/z_0)^\gamma\right) \text{ and } T_z^d = T_0^d(z_0/z)^\eta, \tag{4.1}$$

where γ and η are two accelerated coefficients.

4.2.2 Burn-In

In practice, burn-in should be carried out at some usage rate z_b. In addition, the duration τ of burn-in is another decision variable. Both these two decision variables would affect costs, normal failures and defect failures, and thus should be included in the models and optimally determined.

Total burn-in costs per unit $C_b(\tau, z_b)$ include fixed set-up cost c_f, burn-in operational cost $c_{op}(\tau, z_b)$ and cost of repair. The operational cost $c_{op}(\tau, z_b)$ should be proportional to the burn-in duration τ whilst increasing in z_b. This cost can be expressed as

$$c_{op}(\tau, z_b) = c_0 \cdot \tau \xi(z_b), \tag{4.2}$$

where c_0 is a coefficient and $\xi(z_b)$ is an increasing function of z_b. The costs of repairing defect and usage related failures are often different. However, comparing with the additional costs of servicing a warranty claim, the difference may be ignored. We assume that the repair cost for each failure during burn-in is c_r regardless of the types of failures.

Burn-in has significant effects on both types of failures. For an item that just finishes burn-in, we introduce the virtual age τ_0^n under the nominal rate z_0 for the normal failure process and the virtual age τ_0^d for a remaining defect (see Huang and Yen 2009 for this concept). It means that running the normal failure process under rate z_b with duration b is equivalent to running it under rate z_0 with duration τ_0^n, and the same for the defect failures. The number of remaining defects after burn-in is denoted by K. The purpose of burn-in is to make K as small as possible.

If a burnt-in item is used at nominal rate z_0, to simplify the notation, we use $G_0^n(t)$ for the CDF of first normal failure during field use and $\omega_0^n(t)$ the failure rate associated with $G_0^n(t)$, where t = 0 at the outset of field operation. If it is sold to a user with rate z, the CDF and failure rate for the first field failure are given by $G_z^n(t)$ and $\omega_z^n(t)$, respectively. Because minimal repair has been assumed for normal failures, normal failures under field operation occur according to an NHPP with ROCOF given by $\omega_z^n(t)$. Similarly, distribution for the time to detect a remaining defect is given by $G_0^d(t)$ if used under z_0. If used at usage rate z the distribution is given by $G_z^d(t)$.

4.2.3 Warranty Policy and Burn-In Criteria

There are some different types of two-dimensional warranty policies and different shapes for the warranty region in the literature, e.g. the triangle, the rectangle, the L-shape, etc. (Murthy and Blischke 2006). The non-renewing free repair warranty (FRW) policy is considered here, as it is suitable for expensive items, which is usually indeed so for items sold with two-dimensional warranty. In addition, we confine to the rectangular warranty region, as most two-dimensional warranties are of this shape. Let W and U be respectively the age and usage limits. Then the warranty period W_z conditional on rate z is given by

$$W_z = \begin{cases} W, & z \le U/W; \\ U/z, & z > U/W. \end{cases}$$

Under this warranty policy, the number of warranty claims $N_w(\tau, z_b)$ for a burnt-in unit sold to a random user is a random variable. It depends on the number of remaining defects, the normal failures and the usage rate distribution of customers. It is perceived as a signal of product quality by consumers, and thus the first criterion is to minimize the expected warranty claims given by

$$J_1(\tau, z_b) = E\left[N_w(\tau, z_b)\right], \tag{4.3}$$

subjected to the constraints

$$z_b \le \bar{z}_b \text{ and } \tau \le \bar{\tau}\ . \tag{4.4}$$

$\overline{z}_b$ is the upper limit for the burn-in usage rate while $\overline{\tau}$ is the possible time constraint imposed on the burn-in duration. An upper limit for the burn-in usage rate is necessary, as extremely high usage rate would cause some extraneous failure modes that would never occur at use levels (Nelson 1990, pp.38). An upper limit for τ is necessary since long burn-in duration prolongs the time-to-market.

Offering such a warranty also incurs warranty costs, which depend on $N_w(\tau, z_b)$ and the warranty repair cost. When a warranty failure occurs, the cost includes the regular repair cost c_r, and an additional cost c_a due to additional handling and administrative costs. Therefore, the total cost to rectify a warranty failure is c_r+c_a. The expected total costs $C_t(\tau, z_b)$ for a unit include both burn-in costs and warranty costs. The expected costs tell the manufacturer how much money that should be put into reserve to meet the burn-in expense and the future claims, and is extremely important for accounting purpose. Therefore, the second criterion is to minimize the per unit expected total cost

$$J_2(\tau, z_b) = E\left[C_t(\tau, z_b)\right], \tag{4.5}$$

subjected to the constraints given by (4.4).

4.3 Model Analysis and Optimization

4.3.1 Failures during Burn-In

Under burn-in usage rate z_b and the assumption of minimal repair, the normal failures constitute an NHPP with ROCOF

$$\upsilon_b^n(t) = (z_b / z_0)^\gamma \cdot \upsilon_0^n\left(t(z_b / z_0)^\gamma\right).$$

Therefore, the number of normal failures during burn-in is Poisson distributed with mean number of occurrence

$$E\left[N_b^n(\tau, z_b)\right] = \int_0^\tau \upsilon_b^n(t)\,du \,. \tag{4.6}$$

Under z_b, the distribution of the defect failure time T_b^d is linked to $F_0^d(t)$ via (4.1) as

$$F_b^d(t) = F_0^d\left(t\left(z_b / z_0\right)^\eta\right).$$

Consider a specific defect. The probability that it is detected within burn-in is

$$F_b^d(\tau) = F_0^d\left(\tau_0^d\right). \tag{4.7}$$

There are M identical and independent defects in the item. Failures due to defects can be described by a binomial process with M trials, each with success probability given by (4.7). Conditional on M, The number of defects detected during burn-in follows a binomial distribution with mean value

$$E\left[N_b^d(\tau, z_b \mid M)\right] = M \cdot F_0^d\left(\tau_0^d\right). \tag{4.8}$$

Here M is a random variable. Integrating it out of (4.8) yields

$$E\left[N_b^d(\tau, z_b)\right] = F_0^d\left(\tau_0^d\right) \cdot E[M]. \tag{4.9}$$

The screening strength (SS) is defined to be the proportion of defects detected during the burn-in test as

$$SS(\tau, z_b) = E\left\{E\left[\frac{N_b^d(\tau, z_b)}{M} \mid M\right]\right\} = F_0^d\left(\tau_0^d\right). \tag{4.10}$$

This is an important index in industrial practice. A practical burn-in plan usually requires the screening strength to exceed some bottom line. However, it cannot be determined through existing burn-in models based on bathtub characterization. Our method is able to give a simple expression for the screening strength.

The number of burn-in failures includes both normal and defect failures, its expected value is given by

$$E\left[N_b\left(\tau,z_b\right)\right]=E\left[N_b^d\left(\tau,z_b\right)\right]+E\left[N_b^n\left(\tau,z_b\right)\right]. \quad (4.11)$$

4.3.2 Failures under Warranty

Normal failures

According to (4.1), after burn-in, the virtual age of the normal failure process under z_0 is

$$\tau_0^n=\tau\cdot\left(z_b/z_0\right)^{\gamma}. \quad (4.12)$$

Under z_0, normal failures during field use constitute an NHPP due to the minimal repair assumption. The associated failure rate is $\upsilon_0^n\left(t+\tau_0^n\right)$. In practice, the usage rate is uncertain. To obtain the expected number of normal failures within warranty, we use conditional approach by first conditional on the usage rate and then obtaining unconditioned results. Conditional on Z = z, the normal failure rate is given by (Nelson 1990) $\left(z/z_0\right)^{\gamma}\omega_0^n\left(t\left(z/z_0\right)^{\gamma}+\tau_0^n\right)$. The number of normal failures within the warranty region is a Poisson variable with mean number of occurrence

$$E\left[N_w^n\left(W_z\mid\tau,z_b\right)\right]=\int_0^{W_z}\left(z/z_0\right)^{\gamma}\omega_0^n\left(t\left(z/z_0\right)^{\gamma}+\tau_0^n\right)du. \quad (4.13)$$

The final mean number of normal failures is obtained by un-conditioning, i.e., by taking expectation of (4.13) with respect to the usage rate:

$$E\left[N_w^n\left(\tau,z_b\right)\right]=\int_0^{\infty}E\left[N_w^n\left(W_z\mid\tau,z_b\right)\right]d\Psi\left(z\right). \quad (4.14)$$

Defect failures

Consider a specific remaining defect under nominal rate z_0. According to (4.1), virtual age of this remaining defect right after burn-in is $\tau_0^d=\tau\cdot\left(z_b/z_0\right)^{\eta}$. The distribution for time to detect this remaining defect is given by

$$G_0^d(t) = \frac{F_0^d\left(t+\tau_0^d\right) - F_0^d\left(\tau_0^d\right)}{1 - F_0^d\left(\tau_0^d\right)}.$$

If used at rate *z*, the distribution function for this defect failure is given by $G_z^d(t)$. This distribution is linked to $G_0^d(t)$ through

$$G_z^d(t) = G_0^d\left(t\left(z/z_0\right)^\eta\right).$$

This defect either gets detected during the warranty period or after warranty expiry. The probability that it would be detected during the warranty period is

$$G_z^d(W_z) = G_0^d\left(W_z\left(z/z_0\right)^\eta\right). \quad (4.15)$$

Totally *K* remaining defects are released to the user and *K* is a random variable. Given *M*, *K* follows a binomial distribution as

$$(K \mid M) \sim Bi\left(M, 1 - F_0^d\left(\tau_0^d\right)\right). \quad (4.16)$$

The unconditional distribution of *K* can be obtained by taking expectation of (4.16) with respect to *M*. Conditional on *K*, the numbers of defect failures during the warranty period and after warranty expiry respectively follow binomial distributions as

$$Bi\left(K, G_z^d(W_z)\right) \text{ and } Bi\left(K, 1 - G_z^d(W_z)\right). \quad (4.17)$$

The conditional mean number of defect failures during warranty is thus given by

$$E\left[N_w^d\left(W_z \mid K, \tau, z_b\right)\right] = K \cdot G_0^d\left(W_z\left(z/z_0\right)^\eta\right). \quad (4.18)$$

Further taking expectation of (4.18) with respect to *K* yields

$$E\left[N_w^d\left(W_z \mid \tau, z_b\right)\right] = \left[F_0^d\left(W_z\left(z/z_0\right)^\eta + \tau_0^d\right) - F_0^d\left(\tau_0^d\right)\right] E[M]. \quad (4.19)$$

The unconditional expected number of defect failures under warranty is then given by

$$E\left[N_w^d(\tau,z_b)\right]=\int_0^\infty E\left[N_w^d(W_z\mid\tau,z_b)\right]d\Psi(z). \tag{4.20}$$

Mean number of warranty claims

The expected total number of warranty claims per burnt-in unit includes both normal failures and failures due to defects, which is given by

$$E\left[N_w(\tau,z_b)\right]=E\left[N_w^n(\tau,z_b)\right]+E\left[N_w^d(\tau,z_b)\right], \tag{4.21}$$

where the two terms on the right-hand side are given by (4.14) and (4.20), respectively.

4.3.3 Cost Analysis

The burn-in cost includes fixed cost, operational cost and repair cost. There are $N_b(\tau,z_b)$ burn-in failures for an item, and thus the burn-in repair cost is $c_r N_b(\tau,z_b)$. The expected burn-in cost can then be written down as

$$E\left[C_b(\tau,z_b)\right]=c_f+c_{op}(\tau,z_b)+c_r\cdot E\left[N_b(\tau,z_b)\right], \tag{4.22}$$

where $E\left[N_b(\tau,z_b)\right]$ is given by (4.11).

When a burn-in product is sold to a user with rate *z*, there are $N_w(W_z\mid\tau,z_b)$ failures during the warranty period. The expected warranty cost associated with this user is

$$E\left[C_w(W_z\mid\tau,z_b)\right]=(c_a+c_r)\cdot E\left[N_w(W_z\mid\tau,z_b)\right].$$

The unconditional expected warranty cost is obtained by integrating *z* out of the above expression, which yields

$$E\left[C_w(\tau,z_b)\right]=(c_a+c_r)\cdot E\left[N_w(\tau,z_b)\right], \tag{4.23}$$

where $E\left[N_w(\tau,z_b)\right]$ is given by (4.21).

The total cost includes both burn-in cost and warranty cost. The expected total cost is given by

$$E\left[C_t\left(\tau, z_b\right)\right] = E\left[C_b\left(\tau, z_b\right)\right] + E\left[C_w\left(\tau, z_b\right)\right], \tag{4.24}$$

where the two terms on the right-hand side are given by (4.22) and (4.23), respectively.

4.3.4 Two Optimization Problems

Performance-based burn-in models

The performance-based burn-in decision involves simultaneously selecting τ^* and z_b^* to minimize the objective function (4.3) subject to the constraint (4.4). The model can be expressed as follows.

$$\begin{aligned} \left[\tau^*, z_b^*\right] &= \arg\min J_1\left(\tau, z_b\right) = E\left[N_w\left(\tau, z_b\right)\right] \\ s.t. \qquad & z_b \le \overline{z}_b, \\ & \tau \le \overline{\tau}. \end{aligned}$$

The following theorem gives a characterization of the optimal burn-in usage rate under the performance-based model.

THEOREM 4.1. Let $\left(\tau^*, z_b^*\right)$ be the optimal burn-in setting for the performance-based burn-in model. Suppose the normal failure rate $\lambda_0^n\left(t\right)$ is non-decreasing in *t*. When the accelerated coefficients follow the relationship $\eta \ge \gamma$, if $\tau^* > 0$, then the optimal burn-in usage rate is $z_b^* = \overline{z}_b$.

Proof: Suppose that $\left(\tau^*, z_b^*\right)$ is the optimal burn-in setting and $\tau^* > 0$. To prove $z_b^* = \overline{z}_b$, we use proof by contradiction. Suppose $z_b^* < \overline{z}_b$, because the distribution of T_b^d is a continuous function of τ and z_b. we can always find $(\tau', \overline{z}_b)$ with $\tau' < \tau^*$ such that

$$\Pr\left(T_b^d < \tau^* \mid z_b^*\right) = \Pr\left(T_b^d < \tau' \mid \overline{z}_b\right).$$

Based on the concept of virtual age, this means that

$$\tau' \left(\overline{z}_b / z_0\right)^{\eta} = \tau^* \left(z_b^* / z_0\right)^{\eta}. \tag{25}$$

Therefore, and $(\tau', \overline{z}_b)$, the number of defect failures within warranty under the burn-in setting $\left(\tau^*, z_b^*\right)$is the same as that under $(\tau', \overline{z}_b)$, i.e.,

$$E\left[N_w^d\left(\tau^*, z_b^*\right)\right] = E\left[N_w^d\left(\tau', \overline{z}_b\right)\right]. \tag{26}$$

In addition, since we have assumed that $\eta \geq \gamma$, so we have

$$\tau' \left(\overline{z}_b / z_0\right)^{\gamma} = \tau^* \left(\overline{z}_b / z_0\right)^{\gamma} \left(z_b^* / z_0\right)^{\eta} / \left(\overline{z}_b / z_0\right)^{\eta} = \tau^* \left(z_b^* / \overline{z}_b\right)^{\eta-\gamma} \left(z_b^* / z_0\right)^{\gamma} \leq \tau^* \left(z_b^* / z_0\right)^{\gamma},$$

where the first equality follows from a simple rearrangement of Eqn. (25), while the inequality follows from the fact that $z_b^* / \overline{z}_b < 1$ and $\eta - \gamma > 0$. This inequality means that for the normal failure process, the virtual age after burn-in with setting $(\tau', \overline{z}_b)$ is less than the virtual age with setting $\left(\tau^*, z_b^*\right)$, i.e., $\tau_0'^n \leq \tau_0^{*n}$. We have assumed that $\lambda_0^n(t)$ is an non-decreasing function of t Therefore, for any t > 0, it can be easily verified that $E\left[N_w^n\left(W_z \mid \tau', \overline{z}_b\right)\right] \leq E\left[N_w^n\left(W_z \mid \tau^*, z_b^*\right)\right]$. Integrating z out yields

$$E\left[N_w^n\left(\tau', \overline{z}_b\right)\right] \leq E\left[N_w^n\left(\tau^*, z_b^*\right)\right] \tag{27}$$

Combining the results of Eqns. (25) and (27) yields that the mean number of warranty claims under the burn-in setting $\left(\tau^*, z_b^*\right)$ is greater than that under $(\tau', \overline{z}_b)$, which contradicts the fact that $\left(\tau^*, z_b^*\right)$ is the optimal burn-in setting. Therefore, we should have $z_b^* = \overline{z}_b$. This completes the proof.

■

There are some practical insights from this theorem. Generally speaking, latent defects are much more sensitive to harsher conditions (Yan and English 1997). This theorem indicates that under this circumstance, burn-in should be conducted at conditions as severe as possible, as long as no extraneous failure modes are introduced. This is because to achieve the same virtual age τ_0^d for the defects, the virtual age τ_0^n for normal failures decreases if the burn-in usage rate increases. This means less damage to the system as long as the designed failure rate $\lambda_0^n(t)$ is increasing. Therefore, this theorem justifies the practice of conducting burn-in under harsh environment.

Cost-based burn-in model

Given the objective function (4.5) and the constraint (4.4), the performance-based burn-in model can be expressed as follows.

$$\begin{aligned} &\left[\tau^*, z_b^*\right] = \arg\min J_2(\tau, z_b) = E\left[C_t(\tau, z_b)\right] \\ &s.t. \qquad z_b \le \overline{z}_b, \\ &\qquad\qquad \tau \le \overline{\tau}. \end{aligned}$$

Similar to Theorem 4.1, we investigate the optimal burn-in usage rate for this model. The result is given in Theorem 4.2.

THEOREM 4.2. Let $\left(\tau^*, z_b^*\right)$ be the optimal burn-in setting for the cost-based burn-in model. Suppose the normal failure rate $\lambda_0^n(t)$ is a non-decreasing function of t and $\eta \ge \gamma$. When $\xi(z_b)/z_b^\eta$ is non-increasing in z_b, if $\tau^* > 0$, then the optimal burn-in usage rate is $z_b^* = \overline{z}_b$.

Proof: Theorem 4.1 implies that for any burn-in setting (τ, z_b) with $\tau > 0$ and $z_b < \overline{z}_b$, we can always find another burn-in setting $(\tau', \overline{z}_b)$ such that $E\left[N_w^n(\tau', \overline{z}_b)\right] \le E\left[N_w^n(\tau, z_b)\right]$. This means that the warranty cost with burn-in setting $(\tau', \overline{z}_b)$ is lower than that with (τ, z_b).

Furthermore, if we assume $c_{op}\left(\tau, z_b\right)/z_b^{\eta}$ is decreasing in z_b, then

$$C_{op}\left(\tau', \bar{z}_b\right) = C_0\tau \frac{z_b^{\eta}}{\bar{z}_b^{\eta}} \xi\left(\bar{z}_b\right) = C_0\tau z_b^{\eta} \frac{\xi\left(\bar{z}_b\right)}{\bar{z}_b^{\eta}} \le C_0\tau z_b^{\eta} \frac{\xi\left(z_b\right)}{z_b^{\eta}} = C_0\tau\xi\left(z_b\right) = C_{op}\left(\tau, z_b\right).$$

This relation implies that for any combination $\left(\tau, z_b\right)$ with $\tau > 0$ and $z_b < \bar{z}_b$, we can find $\left(\tau', \bar{z}_b\right)$ that yields the same screening strength while cutting down the burn-in cost. Taken the warranty cost and burn-in cost together, we have that for any combination $\left(\tau, z_b\right)$ with $\tau > 0$ and $z_b < \bar{z}_b$, we can always find a burn-in setting $\left(\tau', \bar{z}_b\right)$ that yields a lower total cost, indicating that the optimal burn-in usage rate should be $\bar{z}_b$. Therefore, Theorem 4.2 follows.

■

Note that the conditions in this theorem are slightly different from those in Theorem 4.1. If the objective is to minimize the total cost, the optimal burn-in usage rate depends on the trade-off between the additional burn-in costs versus the reduction in warranty costs. When the cost incurred by usage rate increment is not so significant, which is represented by the condition that $\xi\left(z_b\right)/z_b^{\eta}$ is decreasing in z_b, the optimal decision is again to set the test condition as harsh as possible.

When the conditions in the theorems are met, there is only one decision variable left. It is very difficult, if not impossible, to obtain analytical solutions for the optimal burn-in durations. But the optimum can be easily located by some efficient one-dimensional search methods such as the bisectional approach.

Comparison

To get more insights into these two models, the optimal burn-in durations derived from these two models are compared.

THEOREM 4.3. Suppose that the optimal usage rates for both the performance and the cost based burn-in models are $\overline{z}_b$. Then the optimal burn-in duration for the cost-based burn-in model is not greater than that for the performance-based model.

Proof: Denote by $\left(\tau_p^*, z_p^*\right)$ and $\left(\tau_c^*, z_c^*\right)$ the optimal burn-in settings for the performance and the cost-based models, respectively. If $\tau_c^* = 0$, then the result is obvious. So we consider the case $\tau_c^* > 0$ below. The condition in this theorem states that

$$z_p^* = z_c^* = \overline{z}_b .$$

Because the burn-in cost is an increasing function of time, we have that for any $\tau \leq \tau_c^*$,

$$E\left[C_b\left(\tau, \overline{z}_b\right)\right] < E\left[C_b\left(\tau_c^*, \overline{z}_b\right)\right]. \tag{28}$$

In addition, for any $\tau \leq \tau_c^*$, we must have

$$E\left[N_w\left(\tau, \overline{z}_b\right)\right] > E\left[N_w\left(\tau_c^*, \overline{z}_b\right)\right] \tag{29}$$

This is because if $E\left[N_w\left(\tau, \overline{z}_b\right)\right] \leq E\left[N_w\left(\tau_c^*, \overline{z}_b\right)\right]$, then in conjunction with (28) we will have

$$E\left[C_b\left(\tau, \overline{z}_b\right)\right] + E\left[C_w\left(\tau, \overline{z}_b\right)\right] < E\left[C_b\left(\tau_c^*, \overline{z}_b\right)\right] + E\left[C_w\left(\tau_c^*, \overline{z}_b\right)\right],$$

which contradicts with the fact that τ_c^* is the optimal burn-in duration.

Because $\left(\tau_p^*, \overline{z}_b\right)$ is the optimal burn-in setting for the performance-based model, we have $\left(\tau_p^*, \overline{z}_b\right) = \arg\min E\left[N_w\left(\tau, \overline{z}_b\right)\right]$. Therefore,

$$E\left[N_w\left(\tau_p^*, \overline{z}_b\right)\right] \leq E\left[N_w\left(\tau_c^*, \overline{z}_b\right)\right] \tag{30}$$

Eqns. (29) and (30) implies that $\tau_c^* \leq \tau_p^*$

■

Results in Theorem 4.3 can be justified as follows. Performance-based burn-in only needs to strike an optimal balance between the defect failures and the normal failures. Put it another way, the performance-based burn-in seeks to minimize the expected warranty cost, as this cost is linked to the expected number of warranty claims through a linear relationship (4.23). On the other hand, the cost-based burn-in considers the trade-off between burn-in costs and warranty costs due to defect and normal failures within the warranty period. It can be expected that the costlier the burn-in test is, the shorter the optimal burn-in duration for the cost-based model will be.

4.4 A Numerical Example

4.4.1 Model Structure and Parameters

The following settings are adopted to demonstrate our burn-in models.

- The nominal usage rate is $z_0 = 1$.
- Normal failures under z_0 follow a power law process with

$$\upsilon_0^n(t) = \beta_1^{-\beta_2}\beta_2 t^{\beta_2 - 1}.$$

- The number of defects M follows Poisson (μ). Defect failure time T_0^d follows EXP(θ), and the customer usage rate follows $\text{Uniform}(0, \bar{z}_b)$.
- The burn-in operational cost $C_{op}(\tau, z_b)$ is concave in z_b and can be approximated by $C_{op}(\tau, z_b) = c_0 \tau z_b^{0.8}$.

Parameter values for the above specific models and for the cost parameters are given in Table 4.1.

Table 4.1. Parameter settings

para.	$\bar{z}_b$	$\bar{\tau}$	W	U	η	γ	θ	μ	β_1	β_2	c_f	c_r	c_a	c_0

Value	4	0.25	2	2	0.85	0.27	0.3	0.05	6	1.5	0.2	3	40	0.1

4.4.2 Model Analysis

Using (4.11), the mean number of burn-in failures is

$$E\left[N_b\left(\tau, z_b\right)\right] = \mu\left[1-\exp\left(-\tau_0^d / \theta\right)\right] + \left(\tau_0^n / \beta_1\right)^{\beta_2}. \tag{4.31}$$

The screening strength in (4.10) boils down to

$$SS\left(\tau, z_b\right) = 1-\exp\left(-\tau_0^d / \theta\right). \tag{4.32}$$

The unconditional distribution of K is readily obtained by taking expectation of (4.16) with respect to M, which yields

$$K \sim \text{Poisson}\left[\mu\left(1-F_0^d\left(\tau_0^d\right)\right)\right].$$

Given z, the distributions for the numbers of defects detected during the warranty period and remaining after warranty expiry are also obtained by taking expectation of (4.17) with respect to M, which leads to two Poisson variables as

$$\text{Poisson}\left[\mu\left(F_0^d\left(\tau_0^d + W_z\left(z / z_0\right)^\eta\right) - F_0^d\left(\tau_0^d\right)\right)\right] \text{ and } \text{Poisson}\left[\mu\left(1-F_0^d\left(\tau_0^d + W_z\left(z / z_0\right)^\eta\right)\right)\right].$$

The expected number of warranty claims in (4.21) is thus given by

$$E\left[N_w\left(W_z \mid \tau, z_b\right)\right] = \left[\left(\tau_0^n + W_z\right) / \beta_1\right]^{\beta_2} - \left(\tau_0^n / \beta_1\right)^{\beta_2} + \mu\left[F_0^d\left(\tau_0^d + W_z\left(z / z_0\right)^\eta\right) - F_0^d\left(\tau_0^d\right)\right].$$

4.4.3 Optimal Solutions

Performance-based model

According to Theorem 4.1, the optimal burn-in usage rate is $z_b^* = 4$. A simple bisectional search shows that the optimal burn-in duration is 0.170. Figure 4.1 illustrates the effect of burn-in on the mean number of warranty claims. The optimal burn-in settings are summarized in the second row of Table 4.2.

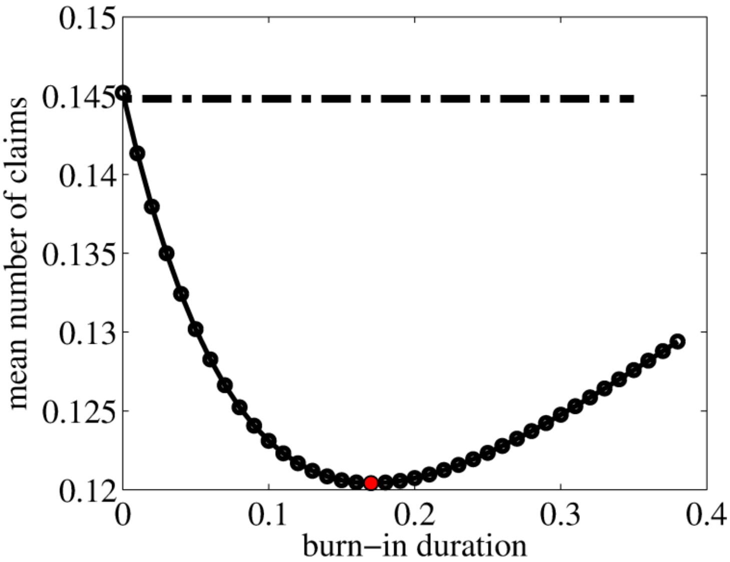


Figure 4.1. The expected number of warranty claims versus the burn-in duration: The horizontal line represents mean number of warranty claims without burn-in

Cost-based model

It is readily checked that $\xi(z_b)/z_b^{\eta} = z_b^{-0.05}$ is decreasing in z_b. Therefore, the conditions in Theorem 4.2 are satisfied. The optimal burn-in usage rate is thus $z_b^* = 4$. The optimal burn-in duration is 0.149. Figure 4.2 demonstrates the effect of burn-in on the total costs. The optimal burn-in settings are given in the last row of Table 4.2.

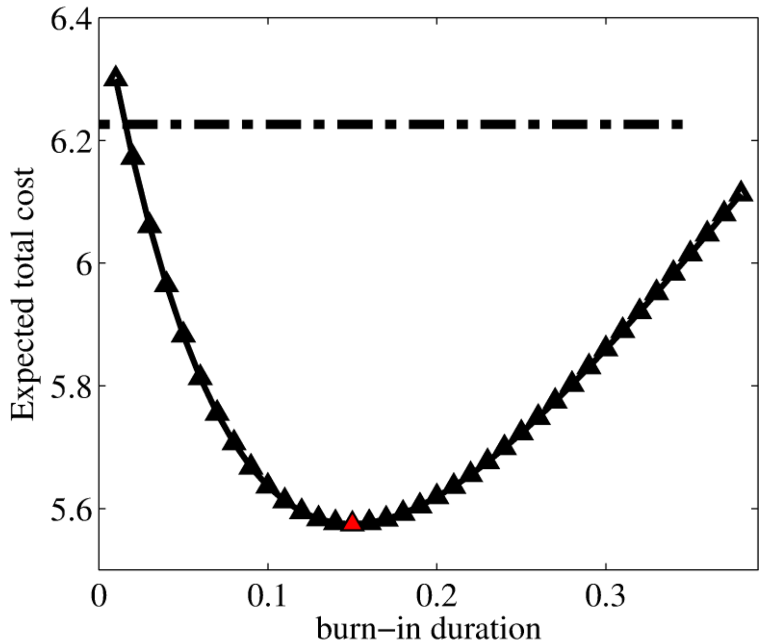


Figure 4.2. The expected total cost versus the burn-in duration: The horizontal line represents the total costs without burn-in

From Figure 4.2, we can see that when the burn-in duration is short, the total cost is higher than that without burn-in. This is due to the fixed set-up cost of burn-in. However when we prolong the burn-in duration, the benefits of burn-in from defect reduction quickly dominate the set-up cost, and thus leading to significant cost reductions.

Table 4.2. Optimal settings under the two burn-in models

Objective	z_b^*	τ^*	SS	$J_i\left(\tau^*, z_b^*\right)$	$J_i\left(0,0\right)$	Reduction (%)
J_1	4	0.170	0.8414	0.1204	0.1448	16.9
J_2	4	0.149	0.8008	5.5746	6.2264	10.5

Comparison

Table 4.2 suggests that both cost and performance based burn-ins are able to achieve high screening strengths and effectively improve the system performance in terms of mean number of warranty claims and expected total costs, respectively. Optimal burn-in durations from these two models do not have significant discrepancy. But compared with the cost-based model, the performance-based burn-in requires a relatively longer burn-in duration so as to achieve a higher screening strength. This result tallies with Theorem 4.3.

4.4.4 Defect Failures after Burn-In

One of the most important benefits of burn-in is that it effectively reduces defects detected within warranty period and those remaining after warranty expiry. The first index is directly related to warranty cost while the second is related to the preventive maintenance costs after expiry of warranty, e.g. Huang and Yen (2009). Both of these two indices are directly related to customer satisfactions. Figure 4.3 illustrates the effects of burn-in on these two indices.

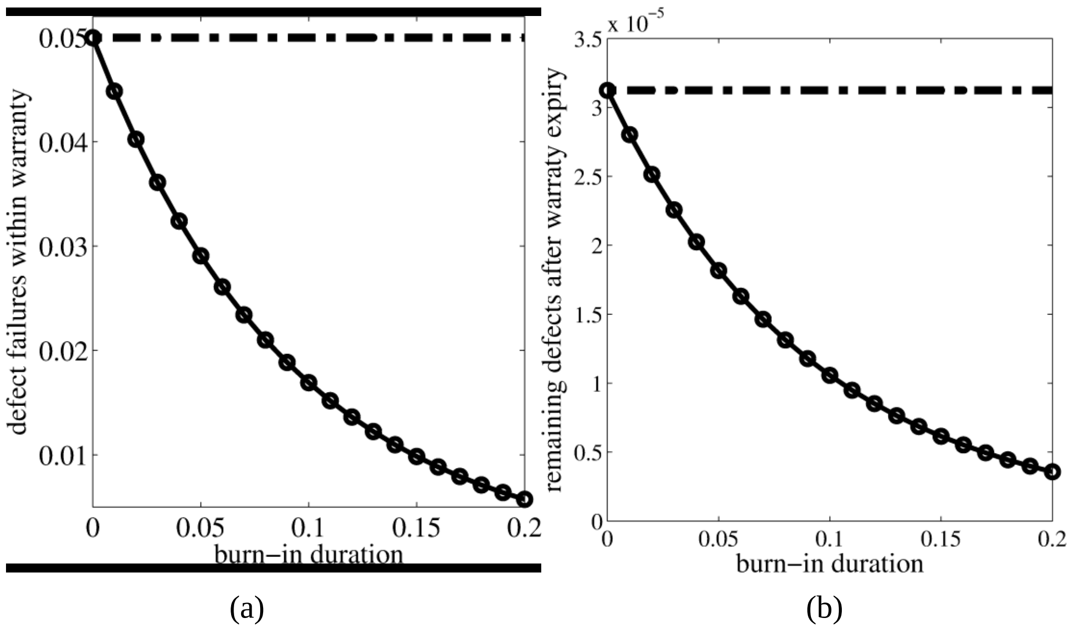


Figure 4.3. (a) The mean number of defect failures within warranty; and (b) The mean number of defects remaining after warranty expiry.

Figure 4.3 indicates that most defects that do not get detected during burn-in would lead to failures during warranty. This is due to the fact that most latent defects will easily lead to item failures in a relatively short time period. As can be seen from this figure, the mean number of defects detected during warranty and those remaining after warranty expiry are significantly reduced with burn-in. These results justify the importance of burn-in.

4.4.5 Sensitivity Analysis

The optimal burn-in schemes mainly depend on the number of defects, the distribution for defect failure times, and the design reliability, i.e. the normal failure process. In the case of cost-based burn-in, the optimal solutions also depend on the cost parameters. This subsection examines effect of the distribution parameters of M, T_0^d, and T_0^n on the optimal burn-in decisions. Typically, the sensitivity analysis is carried out by varying μ, β_1 and θ, respectively. The analysis is done by changing one parameter and keeping other parameters fixed. The results are depicted in Figures 4.4-4.6.

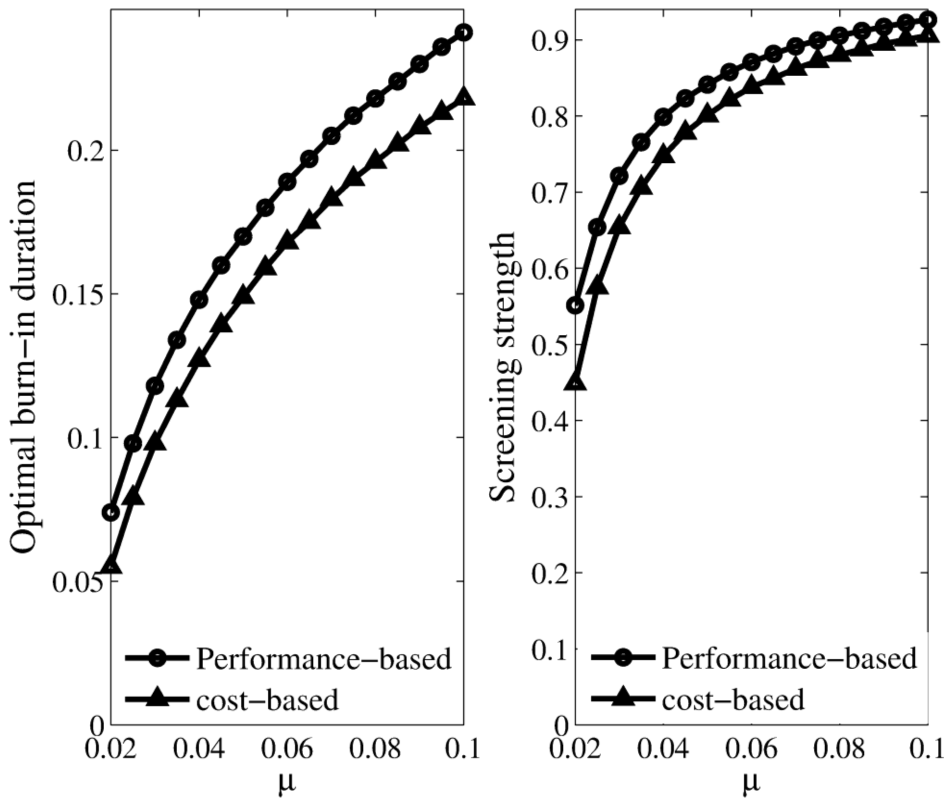


Figure 4.4. Sensitivity analysis: Optimal burn-in durations and the corresponding screening strength for the performance and cost-based burn-in models with μ varying.

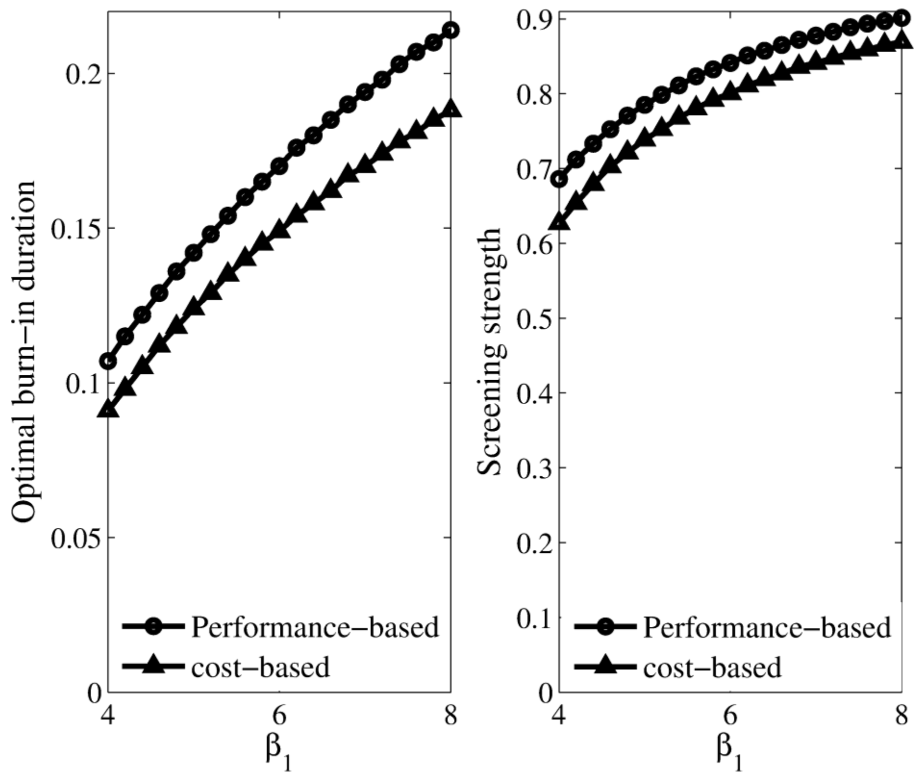


Figure 4.5. Sensitivity analysis: Optimal burn-in durations and the corresponding screening strength for the performance and cost-based burn-in models with β_1 varying.

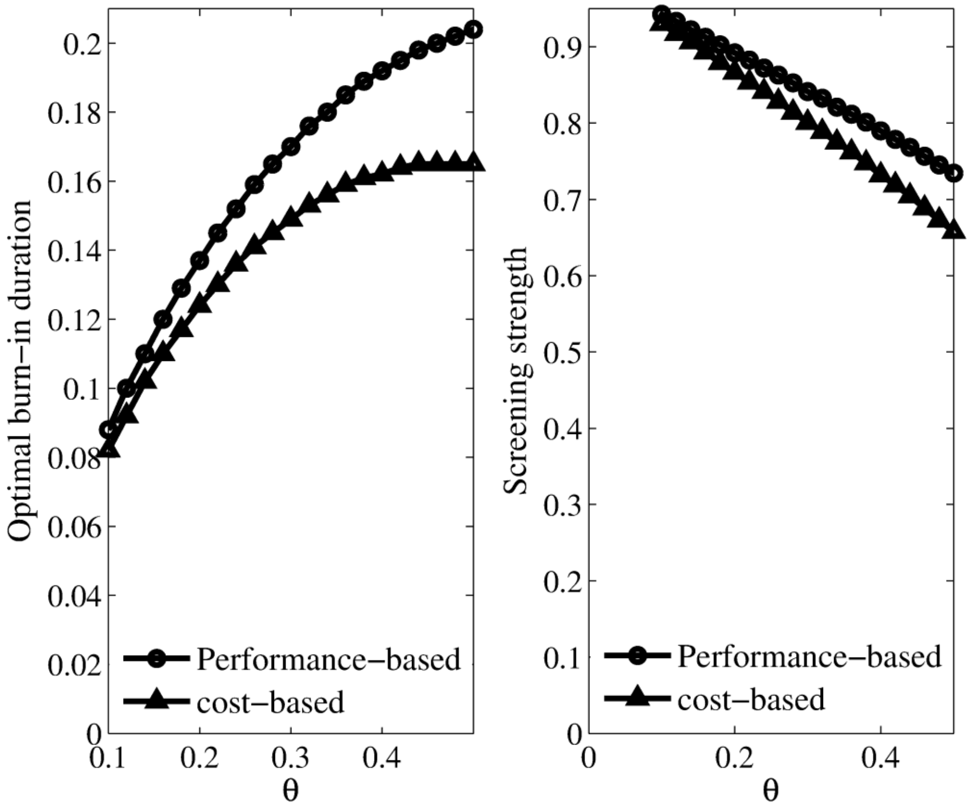


Figure 4.6. Sensitivity analysis: Optimal burn-in durations and the corresponding screening strength for the performance and cost-based burn-in models with θ varying.

From these three figures, we make the following observations.

Optimal burn-in durations for the performance-based burn-in model are always higher than those for the cost-based model. This result is concordant with Theorem 4.3.

As the mean time to detect a defect increases, the optimal burn-in durations increase while the screening strengths decrease. When it gets harder to detect a defect, defect detection takes a longer time. But a longer burn-in duration would result in more damage to an item. The rate with which the optimal burn-in duration increases should be lower than the rate of increment for the mean time to detect a defect. This leads to a decreasing screening strength.

When the mean number of defects increases, the burn-in duration should also increase to get more defects removed.

When the design reliability gets higher, which is represented by a larger β_1, the optimal burn-in duration prolongs. This is because burn-in would also introduce damage to the items. A high design reliability ensures that this damage has less effect on the system. This result implies the importance of design reliability, as a high design reliability not only ensures low normal failure rate, but also tolerates longer tests so as to identify more defects.

4.5 Conclusions

This chapter has developed two burn-in models for products sold with a two-dimensional non-renewing FRW policy by introducing a modeling technique that differentiates normal and defect failures, each having a different accelerated coefficient. A number of contributions have been achieved. We identified the important problem of infant mortality issue faced by expensive products sold with two-dimensional warranties and built two burn-in models which are able to significantly improve the product performance and cut down the warranty costs. As noted by Chukova and Johnston (2006), even a very small proportion of warranty cost reduction for cars would lead to billions of savings for a firm! We also pointed out that the normal and the defect failure modes have different accelerated relationships. This is a very important practical issue which is ignored in previous burn-in studies. By taking the difference in accelerated relationships into account, Theorems 4.1 and 4.2 showed that under realistic assumptions, burn-in should be conducted at the harshest environments, provided that no extraneous failure modes are introduced. These results support the engineering practice of accelerated burn-in testing. We also conducted a sensitivity analysis and revealed the importance of designed reliability from the perspective of defect detection. In addition, it should be noted that for most products sold with one-dimensional warranties, the customer usage behavior is also heterogeneous. For instance, it is not difficult to imagine that a businessman and a student would use their mobile phones differently. Our models subsume the burn-in models under a one-dimensional warranty as special cases by simply setting the usage limit of the two-dimensional warranty to be infinite. Under this special one-dimensional context, our models are expected to lead to better burn-in decisions compared with the existing models, because variation in customer usage rates is taken into consideration.

CHAPTER 5 DEGRADATION-BASED BURN-IN WITH PREVENTIVE MAINTENANCE

5.1 Introduction

As reviewed in Chapter 2, degradation-based joint burn-in and maintenance models are not found, regardless of their potential importance. For example, this chapter is motivated by the problems in modern Micro-Electro-Mechanical Systems (MEMSs). MEMS that emerged in the late 1980s have the ability to sense, actuate and control on the micro scale, and generate effects on the macro scale. Generally speaking, a MEMS device consists of (a) mechanical micro-structures, (b) micro-sensors, (c) micro-electronics and (d) micro-actuators, all integrated onto the same silicon chip. They have shown great potentials in many applications including medical, aeronautical, space and military industries. But the greatest challenge to the successful commercialization of this new technology is how to improve the reliability in a cost-effective way. As pointed out by Arney (2001), infant mortalities in MEMS devices are not uncommon due to the short history and the extremely small size. This leads to significant amount of early failures. To eliminate these early failures and improve MEMS reliability, packaging engineers usually resort to burn-in (Lee *et al.* 2003). Success of a burn-in testing depends on correct analysis of the failure mechanism. According to Tanner (2009), MEMSs are classified into four groups as follows

- Class I – No moving parts.
- Class II – Moving parts with no rubbing or impacting surfaces.
- Class III – Moving parts with impacting surfaces.
- Class IV – Moving parts with impacting and rubbing surfaces.

The first two classes are susceptible to traumatic failures while devices in Classes III and IV are more reliable and are prone to measureable wear-related failures. As such, degradation-based burn-in for devices in Classes III and IV is more cost-effective. In effect, Hogan *et al.* (2003) have addressed the importance of degradation-based burn-in for the production of Digital Micro-mirror Devices.

After a burnt-in MEMS device is put into field use, it is often preventively maintained. In the literature, Peng *et al.* (2009) developed a rudimentary maintenance model for MEMS devices. It is potentially very useful for improving the reliability of MEMS devices and reducing the field operational costs. They also mentioned the importance of burn-in for MEMS devices. However, the effects of burn-in and the determination of optimal burn-in settings are not considered by them. Obviously, simultaneous determination of both optimal burn-in duration and the PM interval would lead to lower costs. To address this problem, this study proposes two general degradation-based burn-in maintenance models under the age and block based maintenance policies, respectively.

The rest of this chapter is organized as follows. In Section 5.2, we briefly introduce the Wiener process with linear drift and use it to model degradation of the product. The problem is then stated and the some assumptions are made. Section 5.3 builds two burn-in maintenance models and derives the corresponding average cost functions. An example is provided to elaborate on the benefits of our models in Section 5.4. Section 5.5 concludes the chapter.

5.2 Problem Statement

5.2.1 The Wiener Process for Degradation Modeling

The Wiener process has received lots of applications in the practice of reliability engineering and survival analysis (Nikulin *et al.* 2009). In this study, we confine our attention to the

Wiener process with linear shift as it is sufficient to describe quality characteristics of many items, probably with a proper time-scale transformation (Tseng *et al.* 2003). A typical Wiener process with linear drift $\{Y(t); t \geq 0\}$ can be expressed as

$$Y(t) = \beta t + \sigma B(t)$$

where β is the drift parameter, σ is the variance coefficient and $B(t)$ is the standard Brownian motion. This process has independent and normally distributed increments, i.e., for $0 \leq u < t$, $Y(t) - Y(u)$ follows the normal distribution $N(\beta(t-u), \sigma^2(t-u))$. When the degradation level reaches a prespecified threshold Y_f, which is often defined by the industrial standard, the product fails. For example, when the wear exceeds Y_f, some quality losses may occur due to inaccuracy of the device. Or when the required minimal thickness to sustain a certain known load is reached, a sudden shock may lead to a sudden breakdown. This minimal thickness can be regarded as the failure threshold to prevent a catastrophic failure. The first passage time of this process, i.e., the time-to-failure, conforms to the famous inverse Gaussian distribution with probability density function (PDF) and the cumulative distribution function (CDF) as follows (Folks and Chhikara 1978).

$$f_{IG}(t, \lambda, \mu) = \left(\frac{\lambda}{2\pi t^3}\right)^{1/2} \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 t}\right), \tag{5.1}$$

$$F_{IG}(t; \mu, \lambda) = \Phi\left(\sqrt{\frac{\lambda}{t}}\left(\frac{t}{\mu} - 1\right)\right) + \exp\left(\frac{2\lambda}{\mu}\right) \cdot \Phi\left(-\sqrt{\frac{\lambda}{t}}\left(\frac{t}{\mu} + 1\right)\right). \tag{5.2}$$

where $\lambda = Y_f^2 / \sigma^2$ and $\mu = Y_f / \beta$.

5.2.2 Model Assumptions

Consider a non-repairable device whose degradation $\{L(t); t \geq 0\}$ is measurable and follows the Wiener process with linear drift. Cha and Finkelstein (2010a) has pointed out that mixed population composed of several ordered subpopulations is not uncommon. In this study, we follow the convention of Jiang and Jardine (2007), Cha and Finkelstein (2010a, 2010b, 2011) and Tsai *et al.* (2011) and assume the product population consists of two subpopulations, i.e., the weak and the normal classes. This assumption is appropriate when assembly defects are possible during manufacturing. Under this circumstance, most units are normal while a small proportion of the product would be deteriorated due to existence of the defects, which contributes to the weak class. The purpose of burn-in is to identify these weak units and eliminate them from the main population. To make the mathematics tractable, some assumptions regarding burn-in are made as follows.

(a) Degradation of units in the normal class follows a Wiener process with drift parameter β_1 and variance coefficient σ while degradation of units in the weak class follows a Wiener process with drift parameter β_2 and variance coefficient σ, where $\beta_2 > \beta_1 > 0$. A unit fails when its wear level exceeds a fixed threshold Y_f.

(b) The burn-in duration is b. Degradation of each unit is not inspected during burn-in. After burn-in, the degradation level of each unit is nondestructively measured. Units with degradations exceeding ξ_b are discarded. ξ_b is called the cut-off level (e.g. Tseng *et al.* 2003).

(c) The per-unit burn-in costs include the manufacturing cost c_s. the fixed burn-in cost plus the measurement cost for each burn-in unit c_m, and the unit time burn-in cost c_0.

It is noted that we consider single inspection right after burn-in above. Compared with continuous or intermittent inspection, this method generally requires less measuring effort and is operationally more convenient.

If a unit survives burn-in, it is put into field use. During the field operation, preventive maintenance is adopted to further improve the reliability and reduce operational cost. By and large, two well-known approaches for PM are the age and the block replacements. The age replacement strategy involves replacement upon failure or upon reaching a predetermined age, whichever occurs first. The block replacement policy replaces the system at a sequence of scheduled time points and upon unexpected failures. A main advantage of the block replacement is its simplicity, as it is unnecessary to keep detailed records about the failure times or ages of the system. For both policies, denote τ as the scheduled replacement interval. If a burnt-in unit survives τ, it is replaced with a new burnt-in unit with replacement cost c_r. Otherwise, if it fails before τ, it is replaced with a new burnt-in unit with cost $c_e + c_r$, where c_e is the extra cost due to unexpected failure. For these two kinds of replacements, the replacement times are assumed negligible.

The decision variables in this problem include the burn-in duration b, the cut-off level ξ_b and the preventive maintenance interval τ. To decide on the optimal settings of these decision variables, two burn-in cost models are developed based on the age and the block replacement policies, respectively.

5.3 Two Burn-In Models Based on Life Cycle Costs

5.3.1 Burn-In Cost

The expected burn-in cost is computed based on the "per-item-output" point of view, meaning that it calculates the amount of money a manufacturer need to pay in order to obtain an accepted burnt-in unit (Liu and Mazzuchi 2008). To obtain this cost, imagine that there is a burn-in lot where units are sequentially subjected to burn-in. For each unit, the cost incurred is $c_0 b + c_s + c_m$. Let $M - 1$ be the number of units scrapped until the first unit with $L(b) < \xi_b$ is obtained. Then the expected cost until obtaining this survival unit can be calculated as follows.

$$C_b(b,\xi_b) = (c_0 b + c_s + c_m)\cdot E[M]. \tag{5.3}$$

Here M can be simply regarded as a geometric random variable with success probability $\Pr(L(b) < \xi_b)$. Therefore, $E[M]$ can be computed as

$$E[M] = 1/\Pr(L(b) < \xi_b).$$

The probability that the degradation level of a burn-in unit does not exceed ξ_b can be computed as follows. After burn-in, the wear level of a normal unit follows a normal distribution $N(\beta_1 b, \sigma^2 b)$ and the wear level of a weak unit follows another normal distribution $N(\beta_2 b, \sigma^2 b)$. Therefore, the probability that a unit is accepted right after the burn-in test is

$$\Pr(L(b) < \xi_b) = p_1 \Phi\left(\frac{\xi_b - \beta_1 b}{\sigma\sqrt{b}}\right) + p_2 \Phi\left(\frac{\xi_b - \beta_2 b}{\sigma\sqrt{b}}\right). \tag{5.4}$$

On the other hand, the main purpose of burn-in is to identify the weak units. Therefore, an index of interest is the screening strength of the burn-in procedure, which is defined as the proportion of defectives detected during the burn-in test:

$$SS(b,\xi_b) = (p_2 - p_{2,b})/\,p_2,$$

where $p_{2,b}$ is the proportion of remaining weak units after burn-in. Consider a unit that is accepted after burn-in. The probability that it belongs to the weak class can be calculated via the Bayes formula as

$$p_{2,b} = \Pr\{\text{weak item} \mid L(b) < \xi_b\} = p_2 \Phi\left(\frac{\xi_b - \beta_2 b}{\sigma\sqrt{b}}\right) / \Pr(L(b) < \xi_b). \tag{5.5}$$

Similarly, the probability that it belongs to the normal class is

$$p_{1,b} = \Pr\left\{\text{normal item} \mid L(b) < \xi_b\right\} = p_1 \Phi\left(\frac{\xi_b - \beta_1 b}{\sigma\sqrt{b}}\right) / \Pr\left(L(b) < \xi_b\right). \quad (5.6)$$

Therefore, the screening strength for the burn-in test with duration b and cut-off level ξ_b can be specified as

$$SS(b, \xi_b) = 1 - \Phi\left(\frac{\xi_b - \beta_2 b}{\sigma\sqrt{b}}\right) / \Pr\left(L(b) < \xi_b\right). \quad (5.7)$$

As can be seen from Equation (5.7), a longer burn-in duration or/and a lower cut-off level leads to a higher screening strength, indicating a more stringent burn-in criterion under which more units will be rejected.

When a unit survives the burn-in procedure with duration b and cut-off level ξ_b, it is put into field operation. In next following, burn-in maintenance models will be formulated under the age replacement and the block replacement policies will be built, respectively.

5.3.2 Age Replacement

The age replacement strategy involves replacement upon failure or upon reaching a predetermined age τ, whichever occurs first. After replacement, the replacement process renews. Therefore, the replacement process is indeed a renewal process with each replacement, either scheduled or unexpected, as a renewal. The undiscounted long run average cost per unit time under this replacement policy can thus be computed based on the theory of renewal reward processes (Ross 2007, Chapter 7.4).

We shall first compute the reliability of a burnt-in unit at time τ. To compute this reliability, the distribution of the remaining lifetime of a burnt-in unit is required. Denote by T_i the remaining lifetimes for a normal (i = 1) and a weak (i = 2) unit, respectively. Failure is defined to be the event that the wear crosses a fixed threshold of value Y_f. Conditional on the wear level at the outset of field operation, i.e. $L_b(0)=u$, $u < Y_f$, the survival probability for the normal and the weak units at time τ can be computed based on (5.2) as

$$\bar{F}_{T_i}(\tau \mid u, \beta_i) = \bar{F}_{IG}\left(\tau; \frac{Y_f - u}{\beta_i}, \frac{(Y_f - u)^2}{\sigma^2}\right), \tag{5.8}$$

where $\bar{F}_{IG}(t;\lambda,\mu)$ is the survival function of an inverse Gaussian (IG) distribution given by $\bar{F}_{IG}(t;\lambda,\mu) = 1 - F_{IG}(t,\lambda,\mu)$.

Denote by T the remaining lifetime of a burnt-in unit. The survival function of T can be calculated by first conditional on the class of this unit and then conditional on the degradation level right after burn-in, which is given by

$$\bar{F}_T(\tau \mid b, \xi_b) = \sum_{i=1}^{2} p_{i,b} / \Phi\left(\frac{\xi_b - \beta_i b}{\sigma\sqrt{b}}\right) \cdot \int_{-\infty}^{\xi_b} \bar{F}_{IG}\left(\tau; \frac{Y_f - x}{\beta_i}, \frac{(Y_f - x)^2}{\sigma^2}\right) d\Phi\left(\frac{x - \beta_i b}{\sigma\sqrt{b}}\right). \tag{5.9}$$

Substituting (5.5) and (5.6) into (5.9) yields

$$\bar{F}_T(\tau \mid b, \xi_b) = \frac{1}{\Pr\left(L(b) < \xi_b\right)} \cdot \sum_{i=1}^{2} p_i \cdot \int_{-\infty}^{\xi_b} \bar{F}_{IG}\left(\tau; \frac{Y_f - x}{\beta_i}, \frac{(Y_f - x)^2}{\sigma^2}\right) d\Phi\left(\frac{x - \beta_i b}{\sigma\sqrt{b}}\right). \tag{5.10}$$

Therefore, the expected field operation cost between two replacements is

$$c_e\left[1 - \bar{F}_T(\tau \mid b, \xi_b)\right] + c_r .$$

Denote by T_r the time interval between two sequential replacements. If the expectation of T_r is known, the undiscounted long run average cost per unit time under the age replacement policy can be computed based on the renewal reward theory as

$$C_{AR}(\tau, b, \xi_b) = \frac{C_b(b, \xi_b) + c_e\left[1 - \bar{F}_T(\tau \mid b, \xi_b)\right] + c_r}{E\left[T_r \mid b, \xi_b\right]} . \tag{5.11}$$

In the following, we shall proceed to compute $E\left[T_r \mid b, \xi_b\right]$. Before proceeding to the main results, the following theorem is useful.

Theorem 5.1. Consider a unit whose wear follows the Wiener process with linear drift $\{Y(t) = \beta t + \sigma B(t); t \geq 0\}$. The unit is replaced when its degradation first exceeds Y_f or when a specified mission time τ is reached, whichever comes first. Letting Z be the time to replacement of this unit, we have

$$E\left[Z \mid Y_f\right] = \tau \bar{F}_{IG}\left(\tau; \frac{Y_f}{\beta}, \frac{Y_f^2}{\sigma^2}\right) + \frac{Y_f}{\beta}\bar{F}_{IG}\left(\frac{Y_f^2}{\tau\beta^2}; \frac{Y_f}{\beta}, \frac{Y_f^2}{\sigma^2}\right). \tag{5.12}$$

Proof: Given a fixed threshold level Y_f, the first passage time for $Y(t)$ follows an inverse Gaussian distribution as $Z \sim IG\left(Y_f / \beta, Y_f^2 / \sigma^2\right)$. Therefore, the expected length of a cycle is

$$E\left[Z \mid Y_f\right] = \tau \bar{F}_{IG}\left(\tau; \frac{Y_f}{\beta}, \frac{Y_f^2}{\sigma^2}\right) + \int_0^\tau x dF_{IG}\left(x; \frac{Y_f}{\beta}, \frac{Y_f^2}{\sigma^2}\right). \tag{5.13}$$

Because $F_{IG}\left(t; Y_f / \beta, Y_f^2 / \sigma^2\right) = F_{IG}\left(t\beta / Y_f; 1, Y_f\beta / \sigma^2\right)$, the second term on the right-hand side can be rewritten as

$$\begin{aligned}\int_0^\tau x dF_{IG}\left(x; \frac{Y_f}{\beta}, \frac{Y_f^2}{\sigma^2}\right) &= \int_0^\tau x dF_{IG}\left(\frac{x\beta}{Y_f}; 1, \frac{Y_f\beta}{\sigma^2}\right) = \frac{Y_f}{\beta}\int_0^{\tau\beta/Y_f} z dF_{IG}\left(z; 1, \frac{Y_f\beta}{\sigma^2}\right) \\ &= \frac{Y_f}{\beta}\int_0^{\tau\beta/Y_f}\left(\frac{Y_f\beta}{2\pi\sigma^2 z}\right)^{1/2}\exp\left(-\frac{(z-1)^2 Y_f\beta}{2\sigma^2 z}\right)dz.\end{aligned}$$

Letting $y = z^{1/2}$, we have $dy = 0.5z^{-1/2}dz$. Therefore, the above integral can be re-organized as

$$\int_0^\tau x F_{IG}\left(x; \frac{Y_f}{\beta}, \frac{Y_f^2}{\sigma^2}\right) = \frac{Y_f}{\beta}\int_0^{(\tau\beta/Y_f)^{1/2}}\left(\frac{2Y_f\beta}{\pi\sigma^2}\right)^{1/2}\exp\left(-\frac{\left(y^2-1\right)^2 Y_f\beta}{2\sigma^2 y^2}\right)dy. \tag{5.14}$$

On the other hand, Shuster (1968) found that

$$F_{IG}\left(t;1,\frac{Y_f\beta}{\sigma^2}\right)=\int_{t^{-1/2}}^{\infty}\left(\frac{2Y_f\beta}{\sigma^2\pi}\right)^{1/2}\exp\left[-\frac{\left(y^2-1\right)^2Y_f\beta}{2\sigma^2y^2}\right]dy\,. \tag{5.15}$$

Letting $t=\infty$, we can find that

$$F_{IG}\left(\infty;1,\frac{Y_f\beta}{\sigma^2}\right)=\int_{0}^{\infty}\left(\frac{2Y_f\beta}{\sigma^2\pi}\right)^{1/2}\exp\left[-\frac{\left(y^2-1\right)^2Y_f\beta}{2\sigma^2y^2}\right]dy=1\,. \tag{5.16}$$

Substituting (5.15) and (5.16) into Equation (5.14) yields

$$\int_0^{\tau}xdF_{IG}\left(x;\frac{Y_f}{\beta},\frac{Y_f^2}{\sigma^2}\right)=\frac{Y_f}{\beta}\bar{F}_{IG}\left(\frac{Y_f}{\tau\beta};1,\frac{Y_f\beta}{\sigma^2}\right)=\frac{Y_f}{\beta}\bar{F}_{IG}\left(\frac{Y_f^2}{\tau\beta^2};\frac{Y_f}{\beta},\frac{Y_f^2}{\sigma^2}\right). \tag{5.17}$$

The theorem follows if we substitute (5.17) to (5.13).

■

Evidently, there are two possibilities to trigger a replacement. The first term on the right hand side of (5.12) depicts replacement due to reaching the mission time, while the second item computes the expected time to replacement due to hitting the threshold before the mission time.

Denote by $T_{r,i}$ the time to replace a normal (i = 1) and a weak (i = 2) unit, respectively. Based on Theorem 5.1, the mean time to replace a normal (i = 1) and a weak (i = 2) unit with initial wear level $L_b\left(0\right)=u$, $u<Y_f$, is

$$\begin{aligned}E\left[T_{r,i}\,|\,Y_f-u\right]=&\,\tau\bar{F}_{IG}\left(\tau;\frac{Y_f-u}{\beta_i},\frac{\left(Y_f-u\right)^2}{\sigma^2}\right)\\&+\frac{Y_f-u}{\beta}\bar{F}_{IG}\left(\frac{\left(Y_f-u\right)^2}{\tau\beta_i^2};\frac{Y_f-u}{\beta_i},\frac{\left(Y_f-u\right)^2}{\sigma^2}\right)\end{aligned}. \tag{5.18}$$

Therefore, the mean value of T_r can be calculated by first conditional on the class of the unit and then conditional on the degradation level right after burn-in, which is given by

$$E\left[T_r \mid b, \xi_b\right] = \frac{1}{\Pr\left(L(b) < \xi_b\right)} \cdot \sum_{i=1}^{2} p_i \int_0^{\xi_b} E\left[T_{r,i} \mid Y_f - x\right] d\Phi\left(\frac{x - \beta_i b}{\sigma\sqrt{b}}\right). \quad (5.19)$$

By substituting (5.19) back to Equation (5.11), the undiscounted long run average cost per unit time can be readily computed. The integral can be computed by some numerical methods. A numerical procedure is provided here.

To compute $Q = \int_0^{\xi_b} E\left[T_{r,i} \mid Y_f - x\right] d\Phi\left(\frac{x - \beta_i b}{\sigma\sqrt{b}}\right)$, the following procedure can be implemented.

- Uniformly divide the interval $\left[0, \Phi\left(\frac{\xi_b - \beta_i b}{\sigma\sqrt{b}}\right)\right]$ into n subintervals. Denote by u_j, j = 1, 2, …, n the midpoint of the j^{th} interval.
- Compute $x_j = \Phi^{-1}\left(u_j; \beta_i b, \sigma^2 b\right)$ where $\Phi^{-1}\left(u_j; \beta_i b, \sigma^2 b\right)$ is the u_j quantile of the normal distribution with mean $\beta_i b$ and variance $\sigma^2 b$.
- Approximate Q as

$$Q = \frac{1}{n} \Phi\left(\frac{\xi_b - \beta_i b}{\sigma\sqrt{b}}\right) \sum_{j=1}^{n} E\left[T_{r,i} \mid Y_f - x_j\right]$$

This procedure partitions unevenly the interval $\left[-\infty, \xi_b\right]$ to give greater resolution in high probability density regions. It is very accurate when a reasonable large n is chosen. However, a large n renders a long computation time. According to our numerical trials, n = 100 is accurate enough while the run time is moderate.

5.3.3 Block Replacement

Under the block replacement policy, a system undergoes scheduled replacements at a sequence of equally spaced time points independent of the failure history. After each scheduled replacement, the replacement process starts anew. Therefore, the replacement process is again a renewal process with each scheduled replacement as a renewal, which is somewhat different from the age replacement case. The length of a cycle is a constant τ. To compute the undiscounted long run average cost per unit time, the expected cost of each cycle is needed, which is derived as follows.

Between two scheduled replacements, there may be some unexpected failures. By assuming negligible repair time, the mean number of unexpected failures within τ is given by

$$N_{BR}\left(\tau \mid b,\xi_b\right)=\sum_{n=1}^{\infty}F_T^{(n)}\left(\tau \mid b,\xi_b\right)=F_T\left(\tau\right)+\int_0^{\tau}F_T\left(t-x\right)dN_{BR}\left(x \mid b,\xi_b\right),$$

where $F_T^{(n)}\left(t \mid b,\xi_b\right)$ is the n-fold convolution of $F_T\left(t \mid b,\xi_b\right)$ given by (5.10). $N_{BR}\left(t \mid b,\xi_b\right)$ is the renewal function for the field replacement process with no preventive maintenance. This function can be efficiently computed by the Riemann-Stietjes (RS) sums method proposed in Xie (1989).

Costs incurred from this policy consist of the preventive replacement cost and the replacement costs from unexpected failures. The expected total field operation cost within two schedule replacements is

$$\left(c_e+c_r\right)N_{BR}\left(\tau \mid b,\xi_b\right)+c_r\,.$$

Therefore, the undiscounted long run average cost per unit time under the block replacement policy can be specified as

$$C_{BR}\left(\tau,b,\xi_b\right)=\frac{\left(C_b\left(b,\xi_b\right)+c_e+c_r\right)N_{BR}\left(\tau \mid b,\xi_b\right)+c_r+C_b\left(b,\xi_b\right)}{\tau}\,. \qquad (5.20)$$

5.3.4 Model Optimization

Model optimization is not a major concern in our study. Although it is very difficult to analytically optimize Equations (5.11) and (5.20), there are many efficient optimization packages that are able to numerically solve these two models. More specifically, there are only three decision variables in both (5.11) and (5.20). Most optimization techniques are able to efficiently tackle this type of low dimensional problems. Alternatively, the genetic algorithm can be used Ye *et al.* (2010).

5.4 Illustrative Example

For illustration, the example provided by Peng *et al.* (2009) is revisited with some modifications. Consider a MEMS device equipped with a micro-engine, whose major failure mechanism attributes to the wear degradation threshold failure. Here, we assume that the population of the MEMS devices consists of a majority of normal units as well as a small proportion of weak class. The proportion of the normal and the weak items are $p_1 = 0.95$ and $p_2 = 0.05$, respectively. Degradation of a normal MEMS follows a Wiener process with drift coefficient $\beta_1 = 8.4832\times10^{-9}$ μm^3/revolution, and variance coefficient $\sigma = 6.002\times10^{-8}$. On the other hand, a weak item degrades with the same variance coefficient but with a larger drift parameter $\beta_2 = 2.6875\times10^{-8}$ μm^3/revolution. A unit fails when its degradation exceeds $Y_f = 1.546\times10^{-5}$ μm^3. Some degradation paths of this imaginary population are illustrated in Figure 5.1.

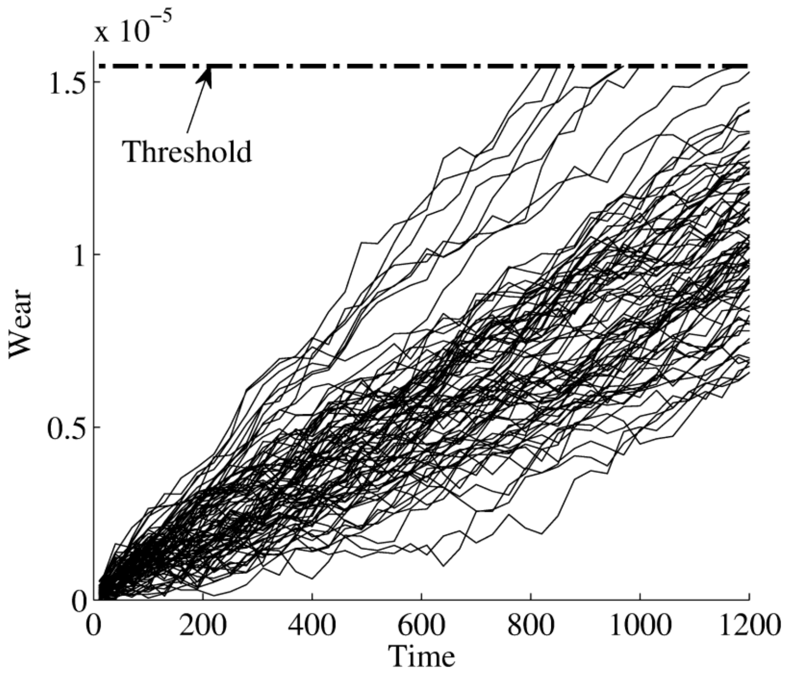


Figure 5.1. Simulated degradation paths

As visualized in Figure 5.1, there are some non-conforming units whose degradations are faster than other units in the population. If these units are put into field use, substantial costs due to unexpected early failures would be generated. Burn-in is thus necessary to scrap these units. To demonstrate our models, the cost structure is assumed as follows.

$$c_e = 1000,\ c_r = 50,\ c_0 = 0.08,\ c_s = 19, \text{ and } c_m = 1.$$

The burn-in maintenance models developed above are used to simultaneously decide the optimal burn-in and preventive maintenance decisions.

5.4.1 Age Replacement

By implementing the Nelder-Mead Simplex method (Lagarias *et al.* 1999), Equation (5.11) can be efficiently minimized. The optimal burn-in time is $b^* = 58.2$, the optimal cut-off point is $\xi_b^* = 1.021 \times 10^{-6}$, and the optimal replacement interval is 1190. With these burn-in and maintenance settings, the optimal average cost rate is 0.0728 and the screening strength (5.7) is 86.0%. The remaining proportion of defective units reduces to 0.70%, which is significantly lower than the original 5%. If burn-in is not performed, the optimal preventive maintenance interval is 1246 while the optimal average cost per unit time is 0.0826. Notably, a rational burn-in test is able to reduce the cost by 11.84%. To visualize the impact of the burn-in settings on the total costs, we fix the replacement interval at 1136 and plot the contour. The contour plot is given in Figure 5.2. The sensitivity of the total cost to the burn-in profiles is readily visualized in this figure.

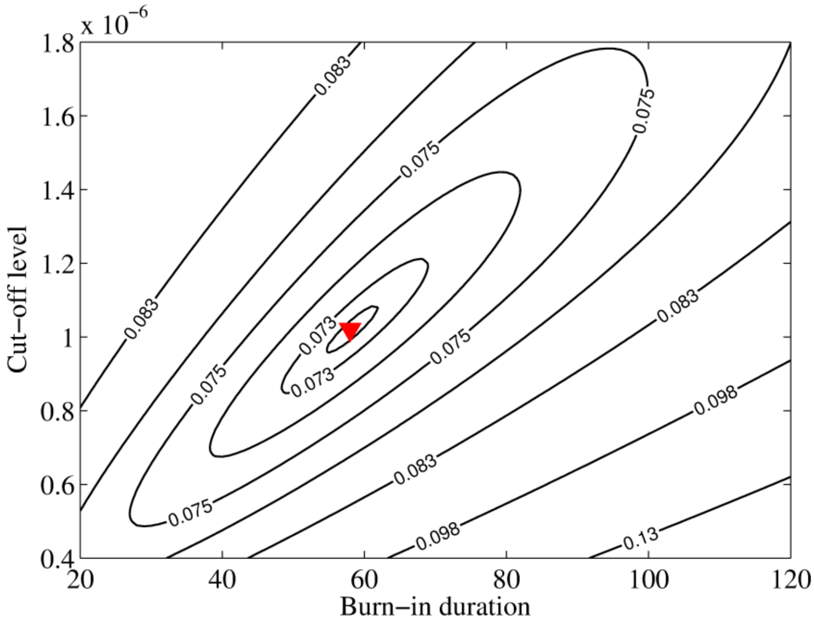


Figure 5.2. Impact of burn-in settings on the total costs with age replacement interval $\tau^* = 1190$

5.4.2 Block Replacement

Similar to age replacement scenario, the optimal burn-in settings and the optimal block replacement period can be obtained by minimizing (5.20). The optimal settings are $b^* = 59.7$, $\xi_b^* = 1.037 \times 10^{-6}$ and $\tau^* = 1178$, respectively. The corresponding optimal average cost rate is 0.0736 per unit time. Under this burn-in scheme, the screening strength (5.7) is 86.8%. Without burn-in, the optimal replacement interval should be 1231, while the optimal average cost per unit time is 0.0855, which is much higher than the cost with moderate burn-in. This highlights the necessities of burn-in testing. The contour plot is given in Figure 5.3 to illustrate the sensitivity of the total cost to the burn-in profiles.

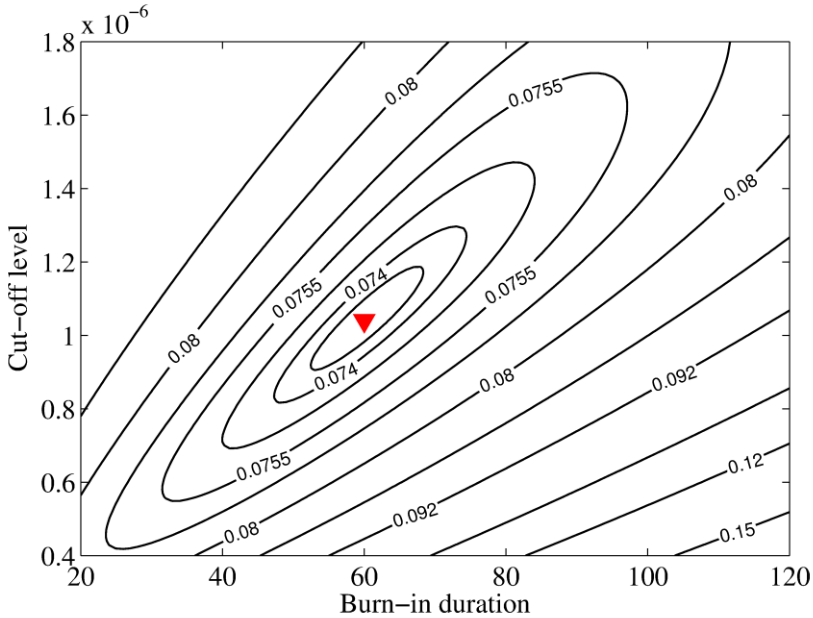


Figure 5.3. Impact of burn-in settings on the total costs with block replacement interval $\tau^* = 1178$

5.4.3 Comparison

Obviously, under both policies, burn-in is able to significantly cut down the cost. In addition, it is worth noting that the block replacement policy yields a slightly higher long-run average cost compared with that of the age replacement policy. This is because block replacement may replace a relatively new item if a failure occurs right before the scheduled maintenance time points, as this policy did not keep track of the replacement history. To reduce this probability and thus cutting down the replacement cost, the optimal replacement interval for this policy should be shorter, which is indeed so in this example. However, the increment in cost is not significant in our example, which is only 1.09% on average. Meanwhile, the optimal burn-in settings have little differences under these two maintenance policies. This can be explained using the concept of screening strength. The purpose of burn-in is to achieve certain screening strength and reduce the defective subpopulations. However, a high screening strength would increase the burn-in cost. The optimal screening strength is a trade-off between burn-in cost and field operation cost. Because the block replacement policy has a higher operational cost, it should have a slightly higher screening strength to further enhance the field reliability compared with that of the age replacement policy.

In addition to the cost reduction, burn-in is also able to enhance performance of a field unit. To demonstrate this fact, the survival functions of a unit with and without burn-in are depicted in Figure 5.4.

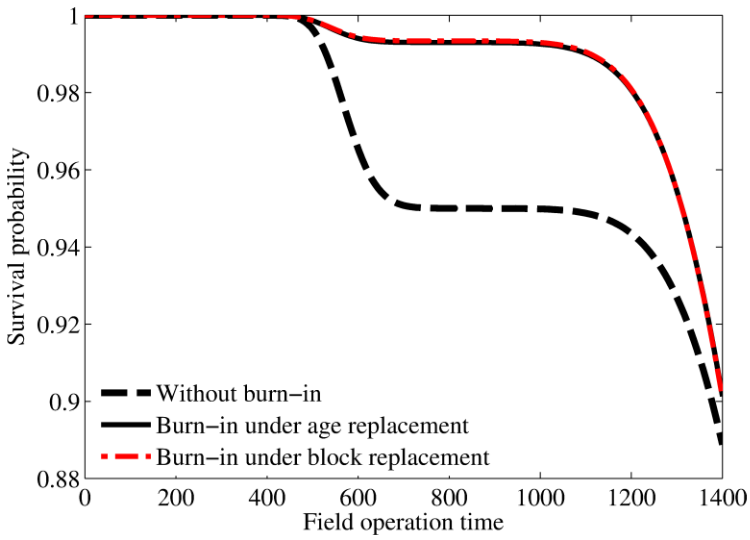


Figure 5.4. Survival functions of a unit with and without burn-in

From this figure, we can see that reliability of a unit has been greatly improved with proper burn-in, while the effects of burn-in based on these two replacement policies are quite similar. Without burn-in, the reliability function decreases very fast during around t = 400 and 600. This is due to failures from the weak class. With moderate burn-in, most defective units will be eliminated from the main population, and thus the reliability only slightly decreases during

t = 400 and 600 due to the remaining defective units. On the other hand, Figure 5.4 indicates that it needs at least 400 units of time to fail a defective unit. To identify these defectives, a failure-based burn-in has to suggest a duration greater than 400, while the degradation-based burn-in only requires a duration of around 60. This further highlights the importance of degradation-based burn-in testing.

5.4.4 Impact of the Defective Proportion

The motivation of conducting burn-in is to cope with the weak subpopulation. Therefore, optimal burn-in decisions depends heavily on percentage of the weak units. To get some insight into the impact of the defective proportion, a sensitivity analysis is conducted. In this sensitivity analysis, all parameter settings are the same as those in Section 5.1, expect the proportion of the weak class. We vary p_2 from 0.01 to 0.09 with step size 0.01, and obtain the optimal burn-in and maintenance settings as well as the optimal costs. The results are depicted in Table 5.1.

When the proportion of the weak class is small, say 1%, burn-in is not necessary as it is costly compared with possible field failures. But it is interesting to observe from Table 5.1 that as the percentage of the defective units increases, both the optimal burn-in duration and the associated screening strength increase. This reflects the efficacy of burn-in in dealing with the weak class. Moreover, the optimal preventive replacement intervals decrease with p_2. This is because the proportion of remaining defective units is increasing with p_2, although the screening strength increases. A shorter preventive replacement interval should be adopted to avoid possible failures due to the remaining weak class.

Table 5.1. Optimal burn-in and maintenance settings and the associated costs

p_2	Age replacement					Block replacement				
	b^*	$\xi_b^*\left(\times 10^{-6}\right)$	τ^*	C^*	SS(%)	b^*	$\xi_b^*\left(\times 10^{-6}\right)$	τ^*	C^*	SS(%)
0.01	0	∞	1223	0.0647	0	0	∞	1211	0.0655	0
0.02	31.9	0.7281	1205	0.0692	60.9	33.8	0.7507	1193	0.0700	63.7
0.03	44.4	0.8717	1198	0.0707	75.8	46.0	0.8898	1196	0.0715	77.3
0.04	52.3	0.9581	1193	0.0719	82.3	53.8	0.9749	1181	0.0727	83.3
0.05	58.2	1.021	1190	0.0728	85.9	59.7	1.037	1178	0.0736	86.8
0.06	63.0	1.071	1187	0.0737	88.4	64.4	1.087	1175	0.0745	89.0
0.07	66.9	1.112	1185	0.0744	90.1	68.4	1.128	1173	0.0753	90.6
0.08	70.3	1.147	1184	0.0751	91.2	71.8	1.163	1171	0.0760	91.8
0.09	73.3	1.179	1182	0.0758	92.2	74.8	1.194	1170	0.0767	92.6

5.5 Conclusions

In this study, two degradation-based burn-in maintenance models have been proposed. Optimal burn-in and maintenance settings are determined based on minimizing the average cost per unit time. The illustrative example shows the effectiveness of the degradation-based method compared with the traditional failure-based burn-in approach. These two models are motivated by the infant mortality in some MEMS devices, but they also have potential applications in laser devices (Tsai et al. 2011, Ye et al. 2012e), LED lamps (Tseng and Peng 2004), etc. They also have important links to the active area of system health management.

CHAPTER 6 DEGRADATION-BASED BURN-IN PLANNING UNDER COMPETING RISKS

6.1 Introduction

Most products may fail in many different ways, known as competing risks. According to the failure mechanism, a failure mode can either be a degradation-threshold (DT) failure or a catastrophic failure. A DT failure, also called soft failure, occurs when a measurable physical degradation reaches a critical threshold level which is often specified by industrial standards; while a catastrophic failure causes instant product failure. Both kinds of failure modes may be subject to infant mortality. For example, failure rate of a catastrophic failure mode might be decreasing, indicating some units will fail very early. Thus we consider two different classifications of failure modes: DT failure / catastrophic failure and infant mortality failure / normal failure.

To identify and eliminate units with infant mortality, engineers often resort to burn-in by activating all infant mortality failure modes during the test for a certain duration. Although products with competing risks are common in practice, current research on burn-in modeling for such products is far from satisfactory. All existing burn-in models pooled all failure modes together and resorted to the overall failure rate. However, it would be beneficial to differentiate different failure modes, as it improves the estimation accuracy, and allows a burn-in practitioner to understand the failure mechanism and to justify the necessity of burn-in. Moreover, these models implicitly assumed that all failure modes are activated during burn-in. If a normal failure mode can be kept dormant during burn-in, unnecessary product aging due to burn-in would be mitigated. For example, we are often able to partially operate a complex system, say, scanning electron microscope. If only parts of the system are prone to bad joints during assembly, it would be desirable to burn-in the system by activating these parts only. In addition, most burn-in models dealt with systems with binary states, i.e., failed or working, and did not make use of any degradation information. Nowadays, many modern products are so well designed and manufactured that they are highly reliable. It may take a very long time for a defective unit to fail even under highly accelerated stresses. Therefore, if a DT failure mode has infant mortality, degradation-based burn-in that bases the screening decision on the product's degradation level after burn-in will be more effective.

An motivating example of this study is from Huang and Askin (2003). An electronic device is subject to two *independent* failure modes, i.e., solder/Cu pad interface fracture which is regarded as a catastrophic failure, and light intensity degradation which is a DT failure mode. The light intensity degradation is measured by the percentage drop of the original light intensity. The device fails if the drop of light intensity exceeds 40 percent of its original value or if an interface fracture occurs. These two failure modes can be activated separately during tests. To assess these two modes, two *different* life tests are conducted under *normal* environmental stresses, each with 10 samples. The first test activates the fracture failure mode only, and all the 10 units are followed to failure; the second test activates the DT mode only, and each unit is inspected every 500h until 4000h. Data from these two tests are tabulated in Tables 6.1 and 6.2.

Table 6.1. Solder/Cu pad interface fracture lifetime data

Sample ID	1	2	3	4	5	6	7	8	9	10
Lifetime (h)	13320	17424	18600	20256	23496	24000	25176	27408	28776	29952

Table 6.2. Light intensity degradation data (in percentage relative to the original measurement)

Sample ID	Inspection time (hours)								
	0	500	1000	1500	2000	2500	3000	3500	4000
11	0	2.5	3.3	4.1	5	5.7	6.5	7.3	8.1
12	0	2.1	2.9	3.7	4.4	5.2	6	6.7	7.5
13	0	2	2.7	3.5	4.3	5	5.8	6.5	7.2
14	0	1.7	2.4	3.2	3.9	4.6	5.4	6.1	6.8

15	0	0.4	1	1.7	2.3	2.9	3.5	4.1	4.7
16	0	0	0.6	1.1	1.7	2.3	2.9	3.4	4
17	0	0	0.5	1.1	1.7	2.2	2.8	3.3	3.9
18	0	0	0.3	0.9	1.5	2	2.6	3.1	3.6
19	0	0	0	0.5	1	1.5	2.1	2.6	3.1
20	0	0	0	0.2	0.7	1.2	1.7	2.2	2.7

A simple Weibull plot shows a good fit to the data in Table 6.1. The estimated shape parameter is greater than 1, indicating that the interface fracture is a normal failure mode. However, a plot of the degradation paths shows that some units exhibit unacceptably high degradation rates, as can be observed from Table 6.2 and Figure 6.1 (Section 6.5). Burn-in should be used to identify these units so as to improve field reliability. Because these two modes can be induced *individually*, we are able to activate the light intensity degradation without inducing the catastrophic failures during burn-in. This is desirable as inducing a normal failure mode incurs unnecessary damages to the product.

However, normal failure modes may have to be activated in some other scenarios. Meeker and Escobar (1998) presented a GaAs laser example of this kind. Most laser devices undergo degradation-based burn-in test before delivered to customers (Johnson 2006). The degradation of a laser device manifests in an increasing operating current. The device fails when the degradation exceeds the threshold, or when a sudden failure occurs (Meeker and Escobar 1998 example 13.5). Possible reasons for the sudden failures include inadvertent shocks and unobserved sudden changes in its physical states. These catastrophic failures have to be activated at the outset of burn-in.

The above two examples suggest that competing risks are not uncommon for products with infant mortality. The purpose of this chapter is to develop a burn-in planning framework for products with independent multiple failure modes. Based on this framework, legitimate burn-in strategies for products in these two examples can be scheduled. Because the trauma failure data are not provided in Meeker and Escobar (1998), we will focus on the electronic device example and build three degradation-based burn-in models. We also propose several methods to cope with the effect of statistical uncertainty on the optimal burn-in decisions.

This chapter is organized as follows. Section 6.2 develops a general burn-in framework for products with competing risks. Based on this framework, three degradation-based burn-in models are built in Section 6.3. The cost functions are established and the optimal cut-off levels are derived. Section 6.4 discusses three methods to deal with the statistical uncertainty issue. In section 6.5, validity of our models is verified by the electronic device example. Section 6.6 concludes the chapter.

6.2 A Burn-In Planning Framework under Competing Risks

Many products are prone to multiple failure modes. We confine to the case where all the failure modes are independent. Behaviors of these modes can be accurately assessed through a carefully designed life/degradation tests (Ye et al. 2012a). Test information is collected and analyzed to identify sources of infant mortalities. If some infant mortality modes belong to the DT class, degradation-based burn-in should be considered. Otherwise, we have to consider traditional failure-based method. During burn-in, all infant mortality modes should be activated to identify weak units. On the other hand, we shall try to avoid activating normal failure modes, if possible, to prevent unnecessary system deterioration. Based on these analyses, mathematical models that quantify effects of all these failure modes can be built to help decide on the optimal burn-in settings. The burn-in planning framework for products with independent multiple failure modes is summarized as follows.

- Specify all failure modes and classify them into the DT class and the catastrophic failure class. Use degradation tests to assess degradation behavior of the DT failure modes. Use ALT to estimate lifetime distributions of the catastrophic failure modes.

- Based on results of the degradation/life tests, classify these failure modes into the infant mortality failure class and the normal failure class. In the normal failure class, specify all failure modes that can be avoided during burn-in and keep them dormant during this test.
- If there is any DT failure mode in the infant mortality class, consider degradation-based burn-in approach. Otherwise, consider failure-based method. Specify the objective of burn-in, e.g., minimize certain cost or maximize certain performance index, and build the corresponding model. This model should take all normal failure modes into account.
- Parameters in this model may be directly obtained from previous study or expert opinions, or it may have to be estimated from results of ALT and degradation test. In the latter scenario, if parameter uncertainty is large, it should be taken into account during model optimization.

Remark 1. Although burn-in is to cope with infant mortality modes, burn-in models should always embrace the normal failure modes, even if they are dormant during burn-in. Ignorance of the normal failure modes would render inferior burn-in decisions with higher costs.

Remark 2. When the infant mortality class includes more than one DT modes, each mode should be assigned a cut-off level. If a DT mode is normal, it can be treated as a catastrophic mode when building burn-in models.

Remark 3. If there are more than one infant mortality modes, it is operationally more convenient to simultaneously activate them and assign to them a common burn-in time, even if they can be activated individually.

6.3 Degradation-Based Models under Competing Risks

Throughout the chapter, we discuss burn-in under nominal use condition. If it is conducted under accelerated stresses, the time scale can be easily transformed to the nominal condition based on the physics of this product (Escobar and Meeker 2006). In the electronic device example, the DT failure, i.e., light intensity degradation, is an infant mortality mode while the interface fracture is normal and can be avoided during burn-in. The framework developed in Section 6.2 is used to help decide on the optimal burn-in settings. In addition, we also develop two other models to enrich the family of degradation-based burn-in models. This section focuses more on degradation modeling. We implicitly assume all parameters are known. This is true when information about these failure modes is available from previous studies or expert knowledge. The Gamma process with random effect introduced by Lawless and Crowder (2004) is found to be well-fit to the light degradation data, and thus will be introduced first.

6.3.1 Preliminaries: Gamma Process with Random Effect

Consider a Gamma process $\{Y(t), t \geq 0\}$ with random effect Θ. Given Θ, the process has independent and Gamma distributed increments, i.e., for $0 \leq u < t$, $Y(t) - Y(u)$ follows $\mathrm{Gamma}(\eta_t - \eta_u, \Theta)$ with PDF

$$f(y;\Theta,\eta) = \frac{\Theta(\Theta y)^{\eta_t - \eta_u - 1}}{\Gamma(\eta_t - \eta_u)} \exp(-\Theta y), \tag{6.1}$$

where $\eta_t = \eta(t)$ is a given, monotone increasing and differentiable function of t with $\eta_0 = 0$. A mathematically tractable model results when Θ conforms to $\mathrm{Gamma}(k, \lambda)$. Unconditional PDF of $Y(t) - Y(u)$ can then be obtained by integrating Θ out of (6.1), which yields

$$\frac{k\left[Y(t)-Y(u)\right]}{\lambda\left(\eta_t-\eta_u\right)} \sim F_{2(\eta_t-\eta_u),2k}, \tag{6.2}$$

where $F_{m,n}$ is the F-distribution with degree of freedom (m, n).

The random effect Θ is unknown but fixed for a unit. Given the degradation level $Y(b)=y_b$ at time b, it can be shown that conditional distribution of the random effect Θ follows $\text{Gamma}\left(\eta(b)+k,\lambda+y_b\right)$. This relation implies that given $Y(b)=y_b$,

$$\left(\frac{\eta_b+k}{\eta_t-\eta_b}\right)\left(\frac{Y(t)-Y(b)}{Y(b)+\lambda}\right) \sim F_{2(\eta_t-\eta_b),2\eta_b+2k}. \tag{6.3}$$

For more details about this process, see Lawless and Crowder (2004).

6.3.2 Problem Formulation

Consider a non-repairable product sold with a preset mission time, e.g., a warranty period, of duration τ. Degradation of its key quality characteristic $\{Y(t),t\geq 0\}$ follows a Gamma process with a Gamma distributed random effect Θ, $\Theta \sim \text{Gamma}(k,\lambda)$. Y_f is a fixed degradation threshold for this process, e.g., in the electronic device example, Y_f = 40. We assume that this DT mode is subject to infant mortality. In addition to this mode, the product is also prone to a catastrophic failure with cumulative distribution function (CDF) $G(\cdot)$ and survival function (SF) $\bar{G}(\cdot)$, which is deemed to be a normal failure mode.

The per unit burn-in cost includes set-up cost c_s, the burn-in operational cost which is proportional to the burn-in duration with proportionality of c_0, and disposal cost. Functionality of a burn-in unit is not monitored during burn-in. After the test with duration b, if a unit has failed (e.g., due to the catastrophic failure), it is scraped with cost c_p. Otherwise, its degradation level is measured non-destructively with measurement cost c_{mea}. If the degradation level exceeds the cut-off level ξ_b, this unit is rejected with disposal cost of c_d (e.g., reworked or sold at a lower price). An accepted unit will be put into field operation. If it fails within the mission time τ, some handling and administrative cost c_f is incurred. Otherwise, a gain of K is generated.

6.3.3 Degradation-Based Burn-in Model with Single Failure Mode

To start, we shall build a model for products with single failure mode, i.e., the DT failure. The purpose of burn-in is simply to identify units with high degradation rates. Denote $\xi_{1,b}$ as the cut-off degradation level with burn-in duration b. To determine the optimal cut-off level $\xi_{1,b}^*$, a cost function should be established first. Because there is no catastrophic failure, all units will not fail during burn-in and thus should be measured after the test. Therefore, the expected burn-in cost can be expressed as

$$c_0 b + \left(c_s + c_{mea}\right) + c_d \cdot \Pr\left(Y\left(b\right) \geq \xi_{1,b}\right),$$

where the screening probability can be obtained based on Equation (6.2) as

$$\Pr\left(Y\left(b\right) \geq \xi_{1,b}\right) = 1 - F_{2\eta_b, 2k}\left(\frac{k\xi_{1,b}}{\lambda\eta_b}\right). \tag{6.4}$$

With probability $\Pr\left(Y\left(b\right) < \xi_{1,b}\right)$, a burnt-in unit is accepted and put into field operation. The field operation cost of this unit can be expressed as

$$c_f - \left(c_f + K\right) \cdot \Pr\left(Y\left(b+\tau\right) \leq Y_f \mid Y\left(b\right) \leq \xi_{1,b}\right).$$

The conditional probability is the probability that this unit survives the mission time τ, which is given by

$$\Pr\left(Y\left(b+\tau\right) \leq Y_f \mid Y\left(b\right) \leq \xi_{1,b}\right) = \frac{1}{\Pr\left(Y\left(b\right) < \xi_{1,b}\right)} \int_0^{\xi_{1,b}} \Pr\left(\Delta Y_b \leq Y_f - u \mid Y\left(b\right) = u\right) f_{Y(b)}\left(u\right) du, \tag{6.5}$$

where $\Delta Y_b = Y(b+\tau) - Y(b)$. Based on (6.3), (6.5) can be expressed as

$$\Pr\left(Y(b+\tau) \le Y_f \mid Y(b) < \xi_{1,b}\right) = \frac{1}{\Pr\left(Y(b) < \xi_{1,b}\right)} \int_0^{\xi_{1,b}} F_{2\Delta\eta_b, 2\eta_b+2k}\left(\left(\frac{\eta_b + k}{\Delta\eta_b}\right)\left(\frac{Y_f - u}{u+\lambda}\right)\right) f_{Y(b)}(u)\,du \,, \tag{6.6}$$

where $\Delta\eta_b = \eta_{b+\tau} - \eta_b$. Summing up the mean burn-in cost and field operation cost, the expected total cost $E\left[C\left(b,\xi_{1,b}\right)\right]$ for a unit is given by

$$\begin{aligned} E\left[C\left(b,\xi_{1,b}\right)\right] &= c_0 b + \left(c_s + c_{mea}\right) + c_d \cdot \Pr\left(Y(b) \ge \xi_{1,b}\right) + \\ &\Pr\left(Y(b) < \xi_{1,b}\right)\left[c_f - \left(c_f + K\right) \cdot \Pr\left(Y(b+\tau) \le Y_f \mid Y(b) \le \xi_{1,b}\right)\right]. \end{aligned} \tag{6.7}$$

The optimal cut-off level $\xi_{1,b}^*$ can be obtained by minimizing Equation (6.7) over $\xi_{1,b}$ with b fixed. It can be shown that when b is fixed, $E\left[C\left(b,\xi_{1,b}\right)\right]$ is convex in $\xi_{1,b}$. Therefore, the minimum of $\xi_{1,b}$ is achieved when $\frac{\partial}{\partial \xi_{1,b}} E\left[C\left(b,\xi_{1,b}\right)\right] = 0$. The result is given in Theorem 6.1.

To simplify the notation, define

$$\Lambda = \left(c_f - c_d\right) / \left(c_f + K\right). \tag{6.8}$$

THEOREM 6.1. Suppose that degradation path of a product follows the Gamma process with random effect and the total cost function is given by (6.7). For fixed b we have the following.

(a) If $0 \le \Lambda \le 1$, the optimal cut-off level $\xi_{1,b}^*$ is

$$\xi_{1,b}^* = \frac{\left(\eta_b + k\right) Y_f - \Delta\eta_b \lambda F^{-1}_{2\Delta\eta_b, 2\eta_b+2k}\left(\Lambda\right)}{\left(\eta_b + k\right) + \Delta\eta_b F^{-1}_{2\Delta\eta_b, 2\eta_b+2k}\left(\Lambda\right)}, \tag{6.9}$$

where $F^{-1}_{2\Delta\eta_b, 2\eta_b+2k}(\cdot)$ is the percentile function of the F-distribution with degree of freedom

$\left(2\Delta\eta_b, 2\eta_b + 2k\right)$. In addition, if $\eta(\cdot)$ is a concave function, $\xi_{1,b}^*$ is increasing in *b*.

(b) If $\Lambda < 0$, the optimal cut-off degradation level is $\xi_{1,b}^* = \infty$.

(c) If $\Lambda > 1$, the optimal cut-off degradation level is $\xi_{1,b}^* = 0$.

Proof: For fixed *b*, differentiating (6.7) with respect to $\xi_{1,b}$ yields

$$\frac{\partial}{\partial \xi_{1,b}} E\left[C\left(b, \xi_{1,b}\right)\right] = f_{Y(b)}\left(\xi_{1,b}\right)\left[-c_d + c_f - \left(c_f + K\right) \cdot \Pr\left(Y\left(b+\tau\right) \le Y_f \mid Y\left(b\right) = \xi_{1,b}\right)\right]$$

The second term on the right hand side is increasing in $\xi_{1,b}$. Therefore, $E\left[C\left(b, \xi_{1,b}\right)\right]$ is a convex function of $\xi_{1,b}$.

The above derivative implies that for every $\xi_{1,b} > 0$, if $\Lambda = \left(c_f - c_d\right)/\left(c_f + K\right) < 0$, we have $\frac{\partial}{\partial \xi_{1,b}} E\left[C\left(b, \xi_{1,b}\right)\right] < 0$. Therefore, $\xi_{1,b}^* = \infty$.

Similarly, if $\Lambda > 1$, then $\frac{\partial}{\partial \xi_{1,b}} E\left[C\left(b, \xi_{1,b}\right)\right] > 0$ for every $\xi_{1,b} > 0$. Therefore, $\xi_{1,b}^* = 0$.

When $0 \le \Lambda \le 1$, the optimal cut-off point $\xi_{1,b}^*$ satisfies

$$c_f - \left(c_f + K\right) \cdot \Pr\left(Y\left(b+\tau\right) \le Y_f \mid Y\left(b\right) = \xi_{1,b}^*\right) = c_d .$$

Re-organizing this equality yields

$$\Pr\left(Y\left(b+\tau\right) \le Y_f \mid Y\left(b\right) = \xi_{1,b}^*\right) = \Lambda . \tag{6.10}$$

The left-hand side of (6.10) can be computed with the help of (6.3), which is given by

$$\Pr\left(\Delta Y_b \le Y_f - \xi_{1,b}^* \mid Y\left(b\right) = \xi_{1,b}^*\right) = F_{2\Delta\eta_b, 2\eta_b + 2k}\left(\left(\frac{\eta_b + k}{\Delta\eta_b}\right)\left(\frac{Y_f - \xi_{1,b}^*}{\xi_{1,b}^* + \lambda}\right)\right). \tag{6.11}$$

Substituting this result to (6.10) yields

$$\frac{Y_f - \xi_{1,b}^*}{\xi_{1,b}^* + \lambda} = \frac{\Delta\eta_b}{\eta_b + k} F_{2\Delta\eta_b, 2\eta_b + 2k}^{-1}(\Lambda).$$

The solution of this equation is no other than (6.9).

Let $d_1 = \Delta\eta_b$, $d_2 = \eta_b + k$ and $a = \frac{Y_f - \xi_{1,b}^*}{\xi_{1,b}^* + \lambda}$. d_1 and d_2 are differentiable functions with respect to b. Then the right-hand side of (6.11) can be rewritten as

$$F_{2\Delta\eta_b, 2\eta_b + 2k}\left(\left(\frac{\eta_b + k}{\Delta\eta_b}\right)\left(\frac{Y_f - \xi_{1,b}^*}{\xi_{1,b}^* + \lambda}\right)\right) = F_{2d_1, 2d_2}\left(ad_2 / d_1\right) = \frac{\int_0^{\frac{a}{a+1}} t^{d_1 - 1}(1-t)^{d_1 - 1}\, dt}{\int_0^1 t^{d_1 - 1}(1-t)^{d_1 - 1}\, dt}.$$

Next we need to verify that $F_{2d_1, 2d_2}\left(ad_2 / d_1\right)$ is increasing in b, i.e.

$$\frac{d}{db} F_{2d_1, 2d_2}\left(ad_2 / d_1\right) > 0.$$

Based on the chain rule for composite functions, this derivative can be expressed as

$$\frac{d}{db} F_{2d_1, 2d_2}\left(ad_2 / d_1\right) = \frac{\partial}{\partial d_1} F_{2d_1, 2d_2}\left(ad_2 / d_1\right) \cdot \frac{dd_1}{db} + \frac{\partial}{\partial d_2} F_{2d_1, 2d_2}\left(ad_2 / d_1\right) \cdot \frac{dd_2}{db}. \quad (6.12)$$

Because $\eta(t)$ is strictly increasing, we have

$$\frac{dd_2}{db} = \frac{d\eta(b)}{db} > 0. \quad (6.13)$$

When $\eta(t)$ is a concave in t,

$$\frac{dd_1}{db} = \frac{d\eta(\tau + b)}{db} - \frac{d\eta(b)}{db} \le 0. \quad (6.14)$$

The two partial derivatives in (6.12) can be obtained through direct calculation.

$$\frac{\partial}{\partial d_2} F_{2d_1, 2d_2}\left(ad_2 / d_1\right) = \frac{\int_0^{\frac{a}{a+1}} \ln(1-x) x^{d_1 - 1}(1-x)^{d_2 - 1} dx \int_0^1 z^{d_1 - 1}(1-z)^{d_2 - 1} dz - \int_0^{\frac{a}{a+1}} x^{d_1 - 1}(1-x)^{d_2 - 1} dx \int_0^1 \ln(1-z) \cdot z^{d_1 - 1}(1-z)^{d_2 - 1} dz}{\left(\int_0^1 x^{d_1 - 1}(1-x)^{d_2 - 1} dx\right)^2}$$

$$= \frac{\int_0^{\frac{a}{a+1}} \ln(1-x)\cdot x^{d_1-1}(1-x)^{d_2-1}dx \int_{\frac{a}{a+1}}^1 z^{d_1-1}(1-z)^{d_2-1}dz - \int_0^{\frac{a}{a+1}} x^{d_1-1}(1-x)^{d_2-1}dx \int_{\frac{a}{a+1}}^1 \ln z\cdot z^{d_1-1}(1-z)^{d_2-1}dz}{\left(\int_0^1 x^{d_1-1}(1-x)^{d_2-1}dx\right)^2}$$

The numerator of the above expression is greater than zero because

$$\int_0^{\frac{a}{a+1}} \ln(1-x)\cdot x^{d_1-1}(1-x)^{d_2-1}dx \int_{\frac{a}{a+1}}^1 z^{d_1-1}(1-z)^{d_2-1}dz = \int_0^{\frac{a}{a+1}} \int_{\frac{a}{a+1}}^1 \ln(1-x)\cdot x^{d_1-1}(1-x)^{d_2-1} z^{d_1-1}(1-z)^{d_2-1}dzdx$$

$$> \int_0^{\frac{a}{a+1}} \int_{\frac{a}{a+1}}^1 \ln(1-z)\cdot x^{d_1-1}(1-x)^{d_2-1} z^{d_1-1}(1-z)^{d_2-1}dzdx = \int_0^{\frac{a}{a+1}} x^{d_1-1}(1-x)^{d_2-1}dx \int_{\frac{a}{a+1}}^1 \ln z\cdot z^{d_1-1}(1-z)^{d_2-1}dz .$$

Therefore,

$$\frac{\partial}{\partial d_2} F_{2d_1,2d_2}\left(ad_2/d_1\right) > 0 . \tag{6.15}$$

Similarly, we can prove that

$$\frac{\partial}{\partial d_1} F_{2d_1,2d_2}\left(ad_2/d_1\right) < 0 . \tag{6.16}$$

Substituting (6.13)-(6.16) back to (6.12), we can conclude that

$$\frac{d}{db} F_{2d_1,2d_2}\left(ad_2/d_1\right) > 0 .$$

When *b* increases, to ensure that $F_{2d_1,2d_2}\left(ad_2/d_1\right) = \Lambda$, *a* should decrease. That is, we should increase $\xi_{1,b}^*$. In sum, when $\eta(t)$ is concave, $\xi_{1,b}^*$ is increasing in *b*.

■

The condition that $\eta(\cdot)$ is concave is necessary for $\xi_{1,b}^*$ to be increasing in *b*. For example if $\eta(b) = \exp(b)$, we find that $\xi_{1,b}^*$ may not be monotonically increasing in *b*. After $\xi_{1,b}^*$ is determined, the optimal burn-in duration b^* can be obtained by minimizing (6.7) with $\xi_{1,b}$

fixed at $\xi_{1,b}^{*}$.

It can be seen from Theorem 6.1 that optimal cut-off levels do not depend on the cost parameters of burn-in operation, i.e., c_0, c_s and c_{mea}. This is because at the time of making the screening decision, the burn-in operational cost can be regarded as sunk cost. It is also interesting to see that Λ serves like a normalize risk measure. When it is large (e.g., a large c_f and a small K), $\xi_{1,b}^{*}$ would be small, indicating a stringent criterion under which more units will be scrapped. Conversely, small Λ leads to a looser criterion.

6.3.4 Two Failure Modes with Normal Failures Inactive during Burn-In

In this model, we consider the scenario where there is a normal failure mode but only the DT mode is activated during burn-in. This scenario fits into the electronic device example, as the interface fracture is normal, and can be avoided during burn-in. Denote $\xi_{2,b}$ as the cut-off degradation level with burn-in duration b. Because the normal mode is inactive during burn-in, all units will not fail during burn-in and thus should be measured after the test. The expected burn-in cost is

$$c_0 b + c_s + c_{mea} + c_d \cdot \Pr\left(Y\left(b\right) \geq \xi_{2,b}\right).$$

If $Y\left(b\right) < \xi_{2,b}$, a burnt-in unit is put into field use. Denote $P_2\left(b,\tau\right)$ as the probability that this unit survives the mission time. It should be noted that the normal failure mode is active during field use. Therefore, this probability is given by

$$P_2\left(b,\tau\right) = \Pr\left(Y\left(b+\tau\right) \leq Y_f \mid Y\left(b\right) < \xi_{2,b}\right) \cdot \bar{G}\left(\tau\right).$$

The expected field operation cost is

$$c_f \cdot \left(1 - P_2\left(b,\tau\right)\right) - K \cdot P_2\left(b,\tau\right).$$

The mean total cost $E\left[C\left(b,\xi_{2,b}\right)\right]$ per unit can thus be expressed as

$$\begin{aligned} E\left[C\left(b,\xi_{2,b}\right)\right] &= c_0 b + c_s + c_{mea} + c_d \cdot \Pr\left(Y\left(b\right) \ge \xi_{2,b}\right) \\ &+ \Pr\left(Y\left(b\right) < \xi_{2,b}\right)\left[c_f \cdot \left(1 - P_2\left(b,\tau\right)\right) - K \cdot P_2\left(b,\tau\right)\right]. \end{aligned} \tag{6.17}$$

The optimal cut-off level $\xi_{2,b}^*$ for each burn-in time *b* can be obtained by minimizing (6.17) over $\xi_{2,b}$ with *b* fixed. The result is encapsulated in Theorem 6.2.

THEOREM 6.2. Suppose that in addition to the DT failure, there is a normal failure mode that can be avoided during burn-in. When the expected cost function is given by (6.7), we have the following.

(a) If $0 \le \Lambda \le \bar{G}\left(\tau\right)$, the optimal cut-off level with *b* fixed is

$$\xi_{2,b}^* = \frac{\left(\eta_b + k\right)Y_f - \Delta\eta_b \lambda F^{-1}_{2\Delta\eta_b, 2\eta_b + 2k}\left(\Lambda / \bar{G}\left(\tau\right)\right)}{\left(\eta_b + k\right) + \Delta\eta_b F^{-1}_{2\Delta\eta_b, 2\eta_b + 2k}\left(\Lambda / \bar{G}\left(\tau\right)\right)}. \tag{6.18}$$

In addition, if $\eta\left(t\right)$ is concave in *t*, $\xi_{2,b}^*$ is increasing in *b*.

(b) If $\Lambda < 0$, the optimal cut-off degradation level is $\xi_{2,b}^* = \infty$.

(c) If $\Lambda > \bar{G}\left(\tau\right)$, the optimal cut-off degradation level is $\xi_{2,b}^* = 0$.

Proof: For fixed *b*, differentiating (6.17) with respect to $\xi_{2,b}$ yields

$$\frac{\partial}{\partial \xi_{2,b}} E\left[C\left(b,\xi_{2,b}\right)\right] = f_{Y(b)}\left(\xi_{2,b}\right)\left[c_f - c_d - \left(c_f + K\right) \cdot \Pr\left(Y\left(b+\tau\right) \le Y_f \mid Y\left(b\right) = \xi_{2,b}\right)\bar{G}\left(\tau\right)\right]$$

Similar to the proof of Theorem 6.1, we can show that for fixed *b*, (6.17) is convex in $\xi_{2,b}$. For every $\xi_{2,b} > 0$,

$$\Lambda < 0 \text{ implies } \frac{\partial}{\partial \xi_{2,b}} E\left[C\left(b, \xi_{2,b}\right)\right] < 0 \text{ and}$$

$$\Lambda > \bar{G}(\tau) \text{ implies } \frac{\partial}{\partial \xi_{2,b}} E\left[C\left(b, \xi_{2,b}\right)\right] > 0,$$

which corresponds to the cases of $\xi_{2,b}^* = \infty$ and $\xi_{2,b}^* = 0$, respectively. When $0 \le \Lambda \le \bar{G}(\tau)$, the optimal cut-off point $\xi_{2,b}^*$ satisfies

$$c_f - c_d = \left(c_f + K\right) \cdot \bar{G}(\tau) \Pr\left(Y(b+\tau) \le Y_f \mid Y(b) = \xi_{2,b}^*\right).$$

Solving this equation yields (6.18).

Following the line of the proof of Theorem 6.1, we can proof that if $\eta(t)$ is concave in *t*, $\xi_{2,b}^*$ is increasing in *b*. Therefore, the theorem follows.

■

Similarly, the optimal cut-off level $\xi_{2,b}^*$ does not depend on the burn-in operational cost (excluding the disposal cost). When there is a normal failure mode, $\xi_{2,b}^*$ is smaller compared with $\xi_{1,b}^*$. This means that when the product deteriorates due to some other mechanisms, e.g., some normal failure modes, the only way we can enhance the reliability is to adopt a more stringent criterion for the infant mortality modes.

6.3.5 Two Failure Modes with Normal Failures Active during Burn-In

To provide more insight into our burn-in planning framework, we further consider the case where there is a normal failure mode which is catastrophic and has to be activated during burn-in. After burn-in with duration *b*, some units would fail due to catastrophic failures. The proportion is *G*(*b*) and thus the expected disposal cost is $c_p G(b)$. Degradation level of a functioning unit is measured. With probability $\Pr\left(Y\left(b\right)\geq\xi_{3,b}\right)$, its degradation would exceed the cut-off degradation level $\xi_{3,b}$, and is rejected. Otherwise, this unit is accepted and put into field use with expected field operation cost

$$c_f\cdot\left(1-P_3\left(b,\tau\right)\right)-K\cdot P_3\left(b,\tau\right),$$

where $P_3\left(b,\tau\right)$ is the probability of fulfilling the mission

$$P_3\left(b,\tau\right)=\Pr\left(Y\left(b+\tau\right)\leq Y_f\mid Y\left(b\right)<\xi_{3,b}\right)\cdot\bar{G}\left(b+\tau\right)/\bar{G}\left(b\right).$$

Therefore, the expected cost function $E\left[C\left(b,\xi_{3,b}\right)\right]$ is given by

$$\begin{gathered}E\left[C\left(b,\xi_{3,b}\right)\right]=c_s+c_pG\left(b\right)+c_{mea}\bar{G}\left(b\right)+c_d\cdot\bar{G}\left(b\right)\Pr\left(Y\left(b\right)\geq\xi_{3,b}\right)\\ \bar{G}\left(b\right)\Pr\left(Y\left(b\right)<\xi_{3,b}\right)\left[c_f-\left(c_f+K\right)P_3\left(b,\tau\right)\right]\end{gathered}. \quad (6.19)$$

Similarly, we can determine $\xi^*_{3,b}$ through minimizing (6.19) over $\xi_{3,b}$ with *b* fixed. The result is summarized in Theorem 6.3.

THEOREM 6.3. Suppose that in addition to the DT failure, there is a normal failure mode that has to be activated during burn-in. When the mean cost function is given by (6.19), we have the following.

(a) If $0\leq\Lambda\leq\bar{G}\left(b+\tau\right)/\bar{G}\left(b\right)$, the optimal cut-off degradation level is

$$\xi^*_{3,b}=\frac{\left(\eta_b+k\right)Y_f-\Delta\eta_b\lambda F^{-1}_{2\Delta\eta_b,2\eta_b+2k}\left(\Lambda\cdot\bar{G}\left(b\right)/\bar{G}\left(b+\tau\right)\right)}{\left(\eta_b+k\right)+\Delta\eta_bF^{-1}_{2\Delta\eta_b,2\eta_b+2k}\left(\Lambda\cdot\bar{G}\left(b\right)/\bar{G}\left(b+\tau\right)\right)}. \quad (6.20)$$

(b) If $\Lambda<0$, the optimal cut-off degradation level is $\xi^*_{3,b}=\infty$.

(c) If $\Lambda>\bar{G}\left(b+\tau\right)/\bar{G}\left(b\right)$, the optimal cut-off degradation level is $\xi^*_{3,b}=0$.

■

The proof is similar to those of Theorems 6.1 and 6.2, and thus is skipped here. Both the models in this section and in Section 6.3.4 can be readily generalized to the cases of multiple normal failure modes.

6.4 Optimization under Parameter Uncertainty

Sometimes, the process/distribution parameters have to be estimated from testing data, e.g., the electronic device example, and thus are subject to estimation uncertainties. Denote $\boldsymbol{\Upsilon}$ as the vector of parameters to estimate. A challenge faced us is how to take this risk into consideration.

6.4.1 Naïve Approach: The Plug-In Method

A traditional approach to cope with this issue is to simply take the maximum likelihood (ML) estimate $\hat{\Upsilon}$ and substitute it into the models in (6.7), (6.17) and (6.19). Optimal burn-in settings can then be determined through optimizing the cost functions by utilizing Theorems 6.1-6.3. This approach is appropriate when sufficient data are available to ensure small estimation error. When the uncertainty issue is severe, however**,** $\hat{\Upsilon}$ may take values significantly different from $\boldsymbol{\Upsilon}$, and the optimal solution found using this approach may be far removed from optimal. Admittedly, in the electronic device example, parameters are estimated from 20 samples and thus the uncertainty is significant. This method needs calibration to take into account parameter uncertainties, especially when the data size is small.

6.4.2 Standard Approach: Resorting to Expectation

In fact, the three models built in Section 3 rely on the mean costs, as both the degradation process and the catastrophic failures are stochastic. The estimated parameters $\hat{\Upsilon}$ are subject to uncertainties and can also be treated as random variables, conditional on which the cost functions take the forms of (6.7), (6.17) and (6.19), respectively. In order to obtain the unconditional mean cost, we need to take the expectation of the cost functions over these distribution parameters. Denote $E_{\hat{\Upsilon}}\left[C\left(b,\xi_{i,b}\right)\middle|\hat{\Upsilon}\right]$, i = 1, 2, 3, as the conditional mean cost per unit. It is noted that Theorems 6.1-6.3 are no longer applicable here. It is extremely difficult, if not impossible, to derive close form expression for the unconditional mean cost, as the distribution of $\hat{\Upsilon}$ is complicated. We recommend using bootstrapping method (Efron and Tibshirani 1993) to generate N sample estimates, computing the conditional mean cost for each estimate, and then averaging over the costs to approximate the unconditional cost. N = 1,000 is use in this study. The stochastic approximation algorithm (SAA) can be used to locate the optimal burn-in duration b^* and the corresponding optimal cut-off level. The theory and effectiveness of the simultaneous perturbation stochastic approximation (SPSA) algorithm have been well established (Spall 2003). Some Matlab codes are available in the SPSA web site (www.jhuapl.edu/spsa). A two-dimensional contour plot is also helpful in visualizing the optimal settings.

6.4.3 A Robust Perspective: Chance Constraint

The expectation-based method does not possess some build-in robustness in the sense that the realized cost is often higher than the expected value. This is not good news for manufacturers because they may underestimate the total cost that should be put into reserve to meet future expense. A justifiable means is to use chance constraint such that upper bound of the resulting cost would be controllable with high probability. This method also avoids the over-conservatism issue faced with the worst-case analysis. In this study, the chance constraint model essentially minimizes the upper α quantile of the costs as follows.

$$\begin{aligned} &\underset{b,\xi_{i,b}\geq 0}{\text{minimize}} \;\; y \\ &\text{subject to } \; \Pr_{\hat{\Upsilon}}\left\{E_{\hat{\Upsilon}}\left[C\left(b,\xi_{i,b}\right)\middle|\hat{\Upsilon}\right]\leq y\right\}\geq 1-\alpha \end{aligned} . \tag{6.21}$$

The optimal burn-in settings suggested by (6.21) give a 1 – α guarantee that the total cost will less than y^*, the optimal value of (6.21). Since the cost function adopts a complex form and the distribution for $\hat{\Upsilon}$ is unknown, this problem cannot be solved analytically. However, (local) optimum can be obtained by using simulation in conjunction with optimization algorithms that are derivative-free or that use numerical gradients, e.g., the mesh adaptive direct search (MADS) algorithm (Audet and Dennis 2006). Similarly, we use N = 1,000 for the bootstrapping. A detail procedure to solve (6.21) is as follows.

Procedure to Solve the Chance Constraint Problem

- Select an appropriate N such that αN is an integer. Generate N sample estimates of $\hat{\Upsilon}$ using parametric bootstrapping method.
- For each b and $\xi_{i,b}$, compute the approximated total cost by substituting these N sample estimates into the conditional cost function. Sort these costs from smallest to largest and use the $(N-\alpha N)^{\text{th}}$ sorted sample cost as an approximation to the smallest y that satisfies the constraint in (14). Denote this y value as $\hat{y}_\alpha\left(b,\xi_{i,b}\right)$
- Use certain optimization algorithm that does not require analytic derivative to minimize $\hat{y}_\alpha\left(b,\xi_{i,b}\right)$ as a function of b and $\xi_{i,b}$.

6.4.4 Additional Remarks

Calibration of the naïve approach can also be done by asymptotic expansions instead of simulation (Barndorff-Nielsen and Cox 1996). But due to the complexity, this method is not discussed here.

When parameter uncertainly is large, the expectation approach is recommended when the manufacturer is risk neutral, and the chance constraint approach when risk averse. On the other hand, the plug-in method is applicable when the parameters are known from other sources (e.g., a previous study), or when enough data from burn-in and in-operation failure/degradation are collected after the burn-in procedure is set-up. The latter case is appropriate for the electronic device example. These in-operation data update the ML estimates, whose consistency ensures minor uncertainty with large data size.

6.5 An Illustrative Example

The cost model developed in Section 6.3.4 is applied to the electronic device example. The optimization approaches presented in Section 6.4 are applied to determine the optimal burn-in settings. In this example, the degradation threshold is $Y_f = 40$. The following cost profile is adopted for illustrative purpose.

$$\tau = 2 \text{ years}, c_0 = 0.01, c_s = 0.1, c_{mea} = 0.1, c_d = -40, c_f = 1{,}000, K = 500.$$

Here a negative disposal cost means that the manufacturer is able to sell a unit at a lower price without any penalty cost if he deems that its quality is not high enough. As we have discussed before, the parameters need to be estimated from test data.

6.5.1 Data Analysis

Consider the degradation data in Table 6.2. Not all units start degradation from time 0, meaning that there are some slow starters. This may be due to limitations in measurement precision. In this analysis, when the degradation values are zero, we treat them as missing data. The modified degradation paths for the 10 units are depicted in Figure 6.1.

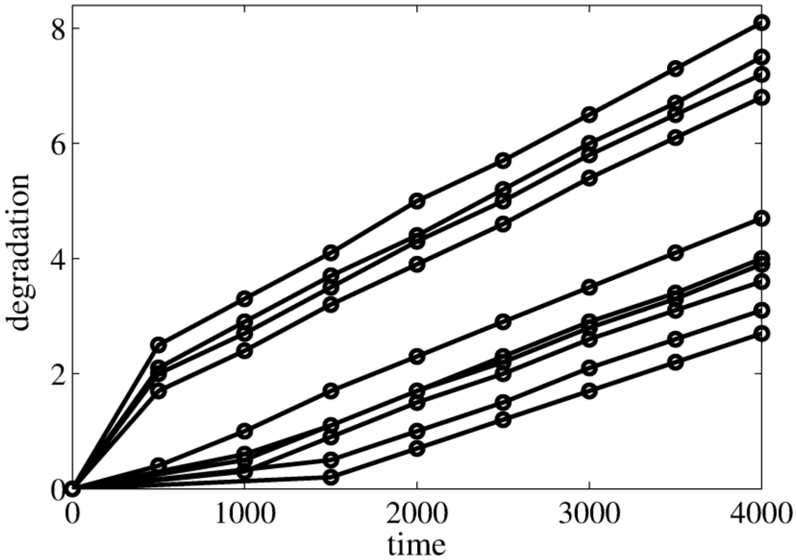


Figure 6.1. The degradation path of the test units.

Figure 6.1 shows that the degradation paths are approximately linear when t > 500. A homogeneous Gamma process with random effect may be appropriate. We assume that $\eta(t)=\beta t$, where β is a parameter to estimate. The likelihood function have been derived in Ye *et al.* (2012c), and is briefly described here.

Assume n units are under test. For the ith unit, $\mathbf{t}_i=\left\{t_{i:0}=0,t_{i:1},...,t_{i:m_i}\right\}$ is the set of check points. Define $\Delta Y_{i:j}=Y_i\left(t_{i:j}\right)-Y_i\left(t_{i:j-1}\right)$ for j = 1, 2, …, m_i. The joint density for $\Delta Y_{i:1},...,\Delta Y_{i:m_i}$ is

$$f_{\Delta Y_{i:1},...,\Delta Y_{i:m_i}}\left(\Delta y_{i:1},...,\Delta y_{i:m_i}\right)=\frac{\lambda^k\Gamma\left(\eta\left(t_{i:m_i}\right)+k\right)}{\left(\lambda+y_i\left(t_{i:m_i}\right)\right)^{\eta\left(t_{i:m_i}\right)+k}\Gamma(k)}\prod_{j=1}^{m_i}\frac{\left(\Delta y_{i:j}\right)^{\Delta\eta_{i:j}-1}}{\Gamma\left(\Delta\eta_{i:j}\right)},$$

where $\Delta\eta_{i:j}=\eta\left(t_{i:j}\right)-\eta\left(t_{i:j-1}\right)$. The log-likelihood function can thus be expressed as

$$l(\theta,k,\lambda)=\sum_{i=1}^{n}\log f_{\Delta Y_{i:1},\ldots,\Delta Y_{i:m_i}}\left(\Delta y_{i:1},\ldots,\Delta y_{i:m_i}\right).$$

ML estimates of this process are listed in Table 6.3.

Table 6.3. MLE for the degradation parameters.

parameters	β	λ	k
MLE	0.00812	1.3052	9.1778
Standard errors	(0.0015)	(0.8362)	(5.1833)

For comparison purpose, the estimated CDF for the time to threshold-defined failures from the method of Huang and Askin (2003) (H-A method), the Gamma process with random effect (Gamma method), and the Kaplan-Meier (KM) estimates are depicted in Figure 6.2. The KM estimation uses pseudo failure times obtained by fitting each degradation path and extrapolating to the threshold (Meeker and Escobar 1998, p. 339). Compared with the H-A method, the estimated CDF based on the Gamma method lies within the 95% pointwise confidence bound of the KM estimates. Therefore, the Gamma process presents an attractive alternative to describe the light degradation.

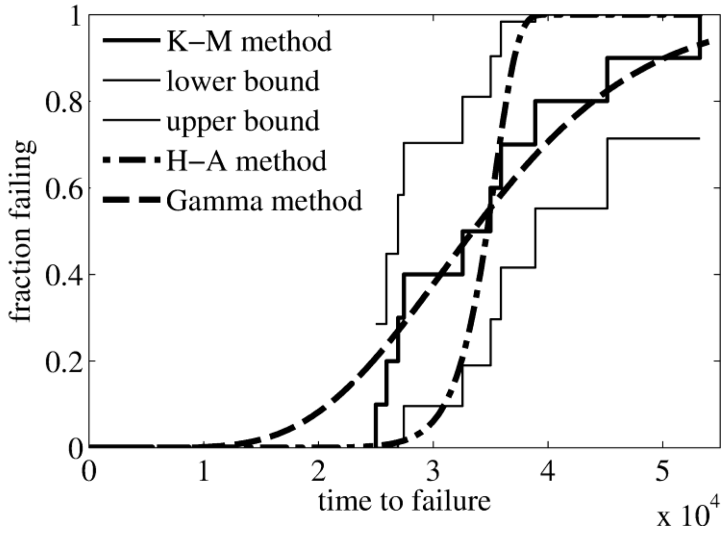


Figure 6.2. Comparison of the Gamma method and the H-A method in estimating CDF of the time to threshold-defined failure.

The Weibull distribution is able to fit the traumatic failure data well. The Weibull plot is given in Figure 6.3. The MLE for the shape parameter is 4.4012 (1.5176), the scale parameter is 25023 (1898.0). Therefore, the SF for the catastrophic failure time is

$$\bar{G}(t) = \exp\left[-\left(\frac{t}{25023}\right)^{4.4012}\right]. \tag{6.22}$$

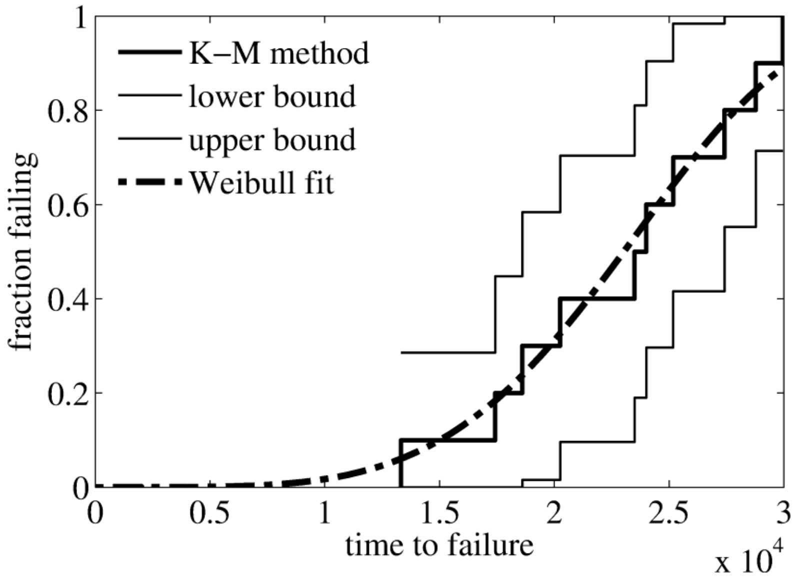


Figure 6.3. Using Weibull distribution to fit the catastrophic failure data

6.5.2 The Plug-In Approach

This device is subject to two failure modes while the catastrophic mode can be avoided during burn-in. Therefore, the cost model (6.17) can be applied to identify weak units. We first ignore the parameter uncertainties and apply the plug-in approach in Section 6.4.1. The optimal cut-off points for each burn-in time can be determined by (6.18), after which optimal burn-in duration can be determined by simple search method. The optimal burn-in scheme is to burn-in a unit for b^* = 497h with cut-off level ξ^*_{2,b^*} = 1.25, leading to the optimal cost of –166.6. The total cost without burn-in is –162.8. Burn-in reduces the cost by 2.32%. In addition, burn-in improves the field reliability. Originally, 19% of the product would ultimately fail due to DT mode. This proportion reduces to 15% with burn-in. To verify the correctness of Theorem 6.2, Figure 6.4 gives the contour of the expected total cost with respect to b and $\xi_{2,b}$. The dashed line is the optimal cut-off levels determined by (6.18).

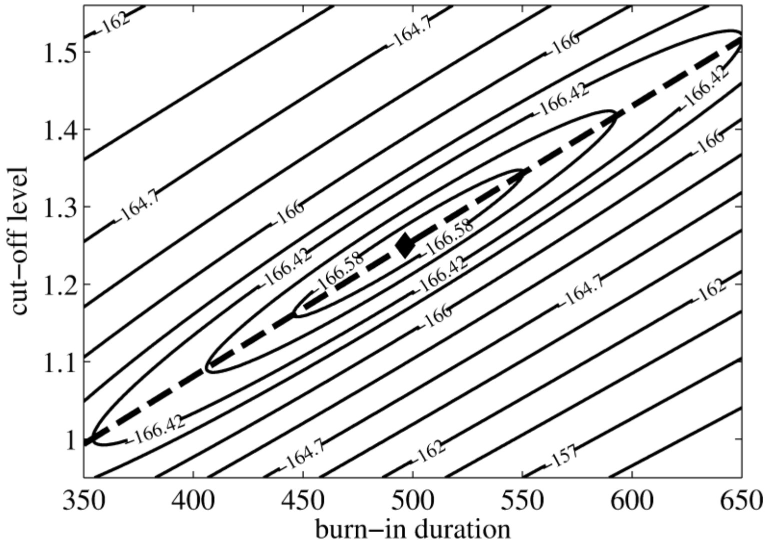


Figure 6.4. The expected total cost by treating the ML estimates as the true values: The dashed line is the optimal cut-off level determined by Theorem 6.2 and the diamond point is the optimal burn-in scheme.

This figure shows that for each fixed b, the cut-off level on the dashed line yields the lowest total cost. In addition, Figure 6.4 also indicates that $\xi_{2,b}^{*}$ is increasing in b, which is concordant with Theorem 6.2 because $\eta(t)$ is linear and thus is concave.

6.5.3 Resorting to Expectation

In view of the fact that the ML estimates themselves are random variables, we can average over them to obtain the unconditional expected cost. To use the SPSA algorithm, we follow the implementation guidance provided by Spall (2003, Chap. 7.5). The optimal burn-in duration is 288h with optimal cut-off level 1.01. The associated optimal cost is –186.4. Without burn-in, the unconditional expected cost is –179.9. Again, products undergone burn-in enjoy a relative cost reduction of 3.61%. A contour plot is provided in Figure 6.5. This figure tends to suggest that this local optimal solution is indeed global optimal.

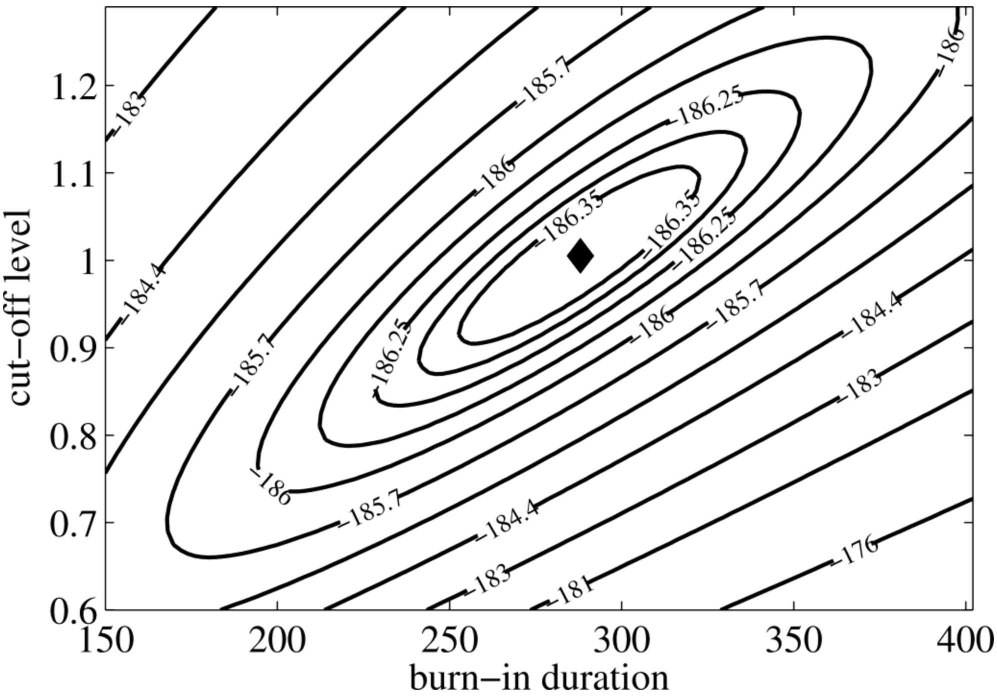


Figure 6.5. The unconditional expected total cost by treating the ML estimates as random variables: The diamond point is the optimal burn-in scheme.

6.5.4 Using Chance Constraint

To make the optimal burn-in scheme more robust, the chance constraint method can be employed. We set $\alpha = 0.05$ and apply bootstrapping in conjunction with the MADS algorithm, as described in Section 6.4.3, to solve (6.21). In this procedure, different starting points are tried. However, each time the algorithm converges to some point with a near-zero cut-off level and a (local) optimal cost greater than the disposal cost c_d. This is due to the problem of limited data and the large uncertainty associated with the ML estimates. Therefore, according to this approach, a risk-averse manufacture may either not launch this device, or reduce the uncertainty and re-evaluate this product by testing more units. A simulation study is carried out here to examine effects of sample size on the optimal solutions.

In the simulation, *n* units are used for the light degradation test and additional *n* units for the interface fracture test. Assume that values of the ML estimates based on these 2*n* units are the same as those in Section 6.5.1. Different values of *n* lead to different optimal solutions, as depicted in Table 6.4.

Table 6.4. Optimal solutions of the chance constraint method with different sample sizes.

n	10	30	50	100	200
b*	n/a	760.9	668.6	636.2	574.5
ξ^*_{2,b^*}	n/a	1.15	1.16	1.28	1.27
EC*	-40	-54.5	-83.3	-100.5	-119.3

Table 6.4 shows that when the sample size gets larger, the chance constraint approach would suggest a meaningful burn-in setting. In addition, when the sample size increases, the optimal solution approaches the settings suggested by the plug-in method. It is also interesting to note that when the uncertainty gets smaller, this approach tends to suggest a slacker burn-in policy (i.e., a smaller burn-in time and a higher cut-off level). This indicates some degree of reservation and robustness of the chance constraint approach in dealing with uncertainty.

6.6 Conclusions

This chapter develops a general burn-in planning framework for products with independent competing risks. This framework suggests identifying all failure modes, classifying them into the right classes, activating all infant mortality modes during burn-in and trying to keep the normal modes dormant. In addition, degradation-based burn-in approach is recommended when some DT modes have infant mortality. In view of the prevalence of multiple failure modes, this framework would furnish a good guide for burn-in practitioners. Based on this framework, three degradation-based burn-in models are developed, one of which is applied to the electronic device example. In addition, three approaches are proposed to deal with the issue of parameter uncertainties.

CHAPTER 7 BI-OBJECTIVE BURN-IN OPTIMIZATION

7.1 Introduction

Burn-in test is a cost-intensive procedure. The total book costs for a burnt-in product include both burn-in costs and the tangible field failure costs, e.g. handling and administrative costs. Many burn-in models have been developed to help decide on the optimal burn-in duration by minimizing the total book costs. Nevertheless, field failures also affect customer satisfactions and lead to intangible cost such as losses of reputation and customer loyalty. These intangible losses often have more significant effects on product benefits, and even on the overall company benefits, as the customer royalty has great impacts on customer retentions and the first purchase decisions of new customers (cf. Reichheld and Teal 2001). The intangible costs are much more difficult to quantify compared with the book costs. These costs associate directly with the performance, especially reliability, of a burnt-in unit. Accordingly, many researchers have proposed a variety of performance-based burn-in models. Although the cost and the performance measures are often correlated, optimal burn-in decisions based on these two measures are different. Optimizing the cost measure often yields a poorer system performance and vice versa. Therefore, we need a unifying burn-in framework for integrating these two measures and then striving for the optimal trade-off. This framework allows the manufacturer to obtain a best-compromise burn-in strategy by specifying the relative weights between these two measures. Naturally, this framework is able to include most existing burn-in models as its special cases.

We then apply the framework to model a system-level burn-in problem. For a complex system, we may use a non-homogeneous Poisson process (NHPP), often with bathtub arrival rate, to model the overall failure process, e.g., see Mi (1996) and Sheu and Chien (2004) to name a few. Alternatively, we can decompose the system into component level so as to trace sources of infant mortalities, and then model lifetime distribution of the system as a function of the reliability of each component position. There are a number of reasons in favor of this approach. One major justification is that many organizations adopt a modular approach to reuse proven component blocks and approaches, if possible, in subsequent products (Turner 2010). Infant mortality rates of these blocks are available from past experiences. Some efforts has been found in this topic; see Reddy and Dietrich (1994), Pohl and Dietrich (1999) and Kim and Kuo (2005, 2009) among others. All these studies stressed the problem from the perspective of either cost or performance. Our bi-objective model jointly considers these two objectives, and is flexible and easy to quantify the penalty of unmet performance. There are some unsolved computational difficulties associated with this modeling approach. In this chapter, we propose two numerical techniques to overcome the computational difficulties.

The remainder of the chapter is organized as follows. We start with a brief discussion of the background of burn-in and burn-in models in Section 7.2. Section 7.3 presents the proposed unifying framework that takes the book costs and field performance into consideration. Based on this framework, we build a system level burn-in model in Section 7.4. Section 7.5 investigates properties of this model and proposes numerical methods to optimize the objective function. An algorithm combining the grid search and an approximation of Riemann–Stieltjes integral (RS Sum) is proposed. In addition, bounds for the reliability function and the delayed renewal function are derived, based on which an approximation method is proposed. In Section 7.6, accuracy of the RS sums method and tightness of the bounds are evaluated through simulation. An illustrative example is provided in Section 7.7. Section 7.8 deals with some concluding comments.

7.2 Background and Problem Formulation

7.2.1 Quality Variations and Burn-In

There are four potential sources of quality issues in manufactured products. The first source of quality variation is from non-conforming components that sneak through supplier's screening. The remaining sources of quality problems are associated with the assembly process described as follows.

First of all, some damages may be inflicted to components when they are installed to the designated positions. Moreover, connection points that link the components to the substrates may also be faulty; these are commonly referred as component connection defects in the system. Last but not least, the assembly process may also damage the substrate. Whenever this type of defect occurs, the system will breakdown. Therefore, this is referred to as series connection defect.

These four sources of quality variations are illustrated in Figure 7.1.

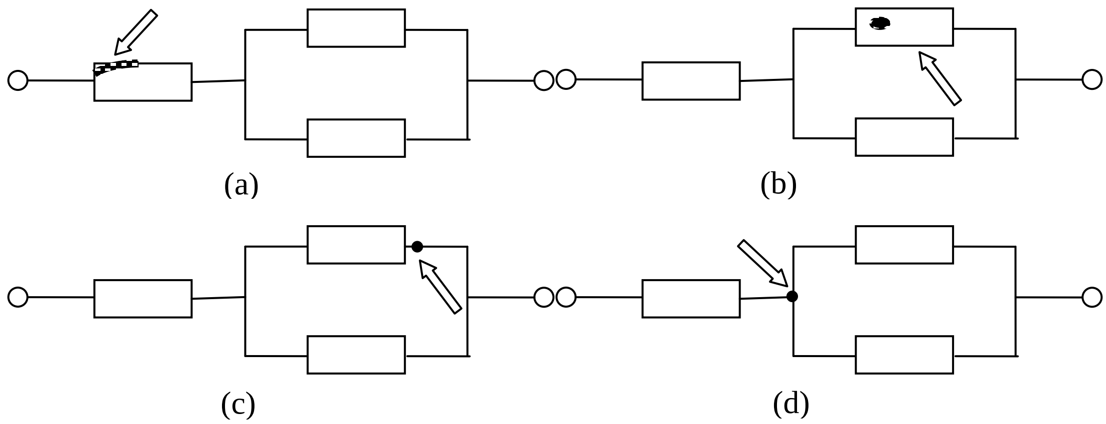


Figure 7.1. Sources of quality variations: (a) Non-conforming components; (b) Component defect; (c) Component connection defect; and (d) Series connection defect.

These four sources of quality variations account for the initial high failure rate, well-known as the infant mortalities, of a product. Burn-in has long been proven to be effective in dealing with the infant mortality, and has been widely used in the semi-conductor industries. They are often conducted under harsh environments such as higher temperature, humility, vibrations and elevated voltage. MIL-STD-883G (2006) specified 6 basic burn-in conditions for microelectronic devices. Latent flaws such as the material defects, software faults and manufacturing defects are precipitated during burn-in.

Models are needed to quantify the effects of burn-in and help decide on the optimal burn-in strategy. Generally speaking, burn-in models can be classified into two groups, i.e. those of cost-based and performance-based. These two classes of models are briefly encapsulated below.

7.2.2 Burn-In to Minimize Total Book Costs

Typical costs associated with burn-in include

- Fixed set-up cost.
- Burn-in operational cost which is approximately linear in the burn-in duration *b*.
- Disposal cost of a weak unit when the product is non-repairable, or repair cost when the product is repairable.

Denote the burn-in costs to obtain a survivor by $C_B(b)$. After a burnt-in unit is put into field use, unexpected failure of the unit would engender some book costs, e.g. field repair cost and warranty losses. Most extant models consider either maintenance cost or warranty costs. We summarize them as follows.

- Minimization of joint burn-in and maintenance costs, e.g. Mi (1996), Sheu and Chien (2004), Cha and Finkelstein (2010a). After burn-in, a legitimate preventive maintenance schedule is able to further reduce the total costs. Models in this class seek to simultaneously determine the optimal burn-in duration and the maintenance intervals.

- Minimization of joint burn-in and warranty costs, e.g. Mi (1999), Yun et al. (2002), Sheu and Chien (2005), Yuan and Kuo (2010) and Ye *et al.* (2011a). Most products are sold with warranty. Breakdown of a burnt-in unit within the warranty period causes warranty claims and transfer to warranty costs in return. We can use burn-in to achieve a balance between burn-in costs and warranty costs. Different warranty policies lead to different cost structures.

The field operation costs are denoted as $C_O(b)$. Objective function of a cost-based model can be expressed as

$$J_1(b) = C_B(b) + C_O(b). \tag{7.1}$$

The cost-based burn-in models attempt to determine the optimal burn-in scheme through minimizing $J_1(b)$.

7.2.3 Burn-In to Optimize Field Performance

Sometimes, the manufacturer may be more concerned with the intangible losses, such as reputation of the brand and the customer royalty. Unmet performance is associated with the user's perceived risks and satisfaction with this product, and thus can be used to represent these losses. The performance objective can be either survival probability over a pre-specified mission time, mean residual life (MRL), or percentile of the residual life. These criteria are summarized and discussed as follows.

- Minimization of the expected number of field failures over a mission time. This measure is directly related to repair costs and warranty costs. Minimizing this measure leads to low book costs. The major concern of this approach is how to model subsequent failures after the first failure.

- Maximization of the survival probability over a mission time, e.g. Mi (1994b), Kim and Kuo (2005, 2009). Even when a product is repairable, maximizing the survival probability presents an attractive alternative because it directly relates to customer satisfaction. In fact, this is often the target of product design (Murthy *et al.* 2009).
- Maximization of the percentile functions of the residual life, e.g., Ye *et al.* (2011a). This criterion leads to the longest mission time that the product can offer given a pre-specified proportion of failures.
- Maximization of the mean residual life (MRL), e.g. Bebbington *et al.* (2007). The MRL is a very important performance index in the context of reliability engineering. If there is no scheduled preventive maintenance, this criterion results in lowest long term average field operation costs.

Denote $J_2(b)$ as such objective functions. Performance-based burn-in seeks to find an optimal burn-in duration so as to optimize $J_2(b)$.

7.3 Burn-In Decision-Making: A Bi-Objective Framework

7.3.1 Bi-Objective Framework: A Meta Model

In this study, we propose to treat the cost and the performance objectives as two independent goals. A bi-objective framework which clearly includes both cost and performance as its special cases can thus be built. The most straightforward approach to solving a multiple-objective problem is to combine these objectives by using weighted sum of the multiple objective functions. This requires manufacturer to normalize the objective function prior to the construction of a specific burn-in model.

For the cost objective, we propose using $J_1(0) = C_O(0)$ as the normalized constant. Then

$1-J_1(b)/J_1(0)$ is the relative cost reduction from burn-in. On the other hand, we propose using $J_2(0)$ as the normalized constant with regard to the performance objective. Then $J_2(b)/J_2(0)-1$ is the relative improvement of the performance through burn-in when the performance objective is the-larger-the-better, e.g. the MRL. When this objective is the-smaller-the-better, the relative improvement of the performance is $1-J_2(b)/J_2(0)$.

After normalization, weights are assigned to these two normalized objectives to represent their relative importance. Denote w_1 and w_2 as the weights assigned to the cost and the performance objectives, respectively. Here, we consider only convex combinations as it is always possible to normalize w_1 and w_2 so that $w_1 + w_2 = 1$.

With these set-ups in place, a bi-objective framework can be constructed. The cost objective is the-smaller-the-better. If the performance objective is the-larger-the-better, the bi-objective framework is

$$\begin{aligned} &\min \quad J = w_1 \cdot J_1(b)/J_1(0) - (1-w_1)\cdot J_2(b)/J_2(0) \\ &s.t. \qquad g_i(b) \le 0, \quad i = 1,2,\ldots,n. \end{aligned} \tag{7.2}$$

where $g_i(b)$ are possible constraints imposed on the burn-in procedure. These constraints may include time limit on the burn-in duration and limited burn-in facilities.

If the performance objective is the-smaller-the-better, the bi-objective framework is

$$\begin{aligned} &\min \quad J = w_1 \cdot J_1(b)/J_1(0) + (1-w_1)\cdot J_2(b)/J_2(0) \\ &s.t. \qquad g_i(b) \le 0, \quad i = 1,2,\ldots,n. \end{aligned} \tag{7.3}$$

The procedure to use this framework is visualized in Figure 7.2.

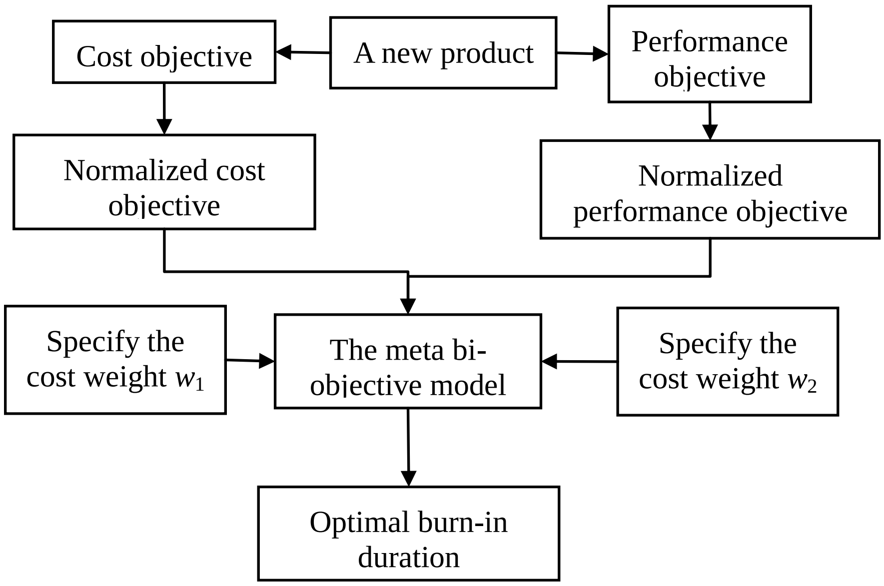


Figure 7.2. The procedure to construct a specific bi-objective model

By varying w_1, a series of optimal burn-in durations can be obtained. When $w_1 = 0$, the model reduces to a purely performance-based one. On the other hand when $w_1 = 1$, this model is purely cost-based, and the optimal burn-in duration achieves a balance between burn-in costs and costs due to field failures. When $0 < w_1 < 1$, there is a trade-off between performance and costs under the bi-objective framework, though there are some correlations between these two objectives.

An alternative way to understand the framework is to treat the performance objective as a penalty, under which we are able to convert the bi-objective model into a purely cost-based one. Some other studies also take the unmet performance as a penalty. For example, Toyota uses a multiple of six times of the repair cost for a field failure to measure the reputation cost, while the Westinghouse uses a multiple of four times (Balachandran and Radhakrishnan 2005). In effect, it is difficult to quantify the costs for such penalty. Our framework effectively fixes the problem by simply requiring the manufacturers to specify the weights that they would like to assign to these two objectives.

7.3.2 Determination of the Weights

The weights w_1 and w_2 are determined according to decision maker's preference. A weight can be interpreted as the relative worth of that objective when compared to the other objectives. The optimal solution is relative to a manufacturer's particular preference structure. Therefore, the solution to a specific bi-objective model is indeed the best-compromise solution.

If the manufacturer believe that these two objectives are equally important, the weights can be set as $w_1 = w_2 = 0.5$. However if the manufacturers are not quite sure about the relative importance, they can perform pair-wise comparison judgment on a set of criteria, e.g. the current flow, the reputation and the customer royalty, with respect to these two objectives. The analytic hierarchy process (AHP) can be effectively applied to determine the weights (Saaty 1994). The procedure is quite standard and goes beyond the scope of this work, thus is not discussed here.

7.4 A Bi-Objective System Level Burn-In Model

In this section, we shall use the framework to build a concrete bi-objective system-level burn-in model. Survival probability of a burnt-in system is adopted as the performance objective as it directly relates to customer satisfaction. We analyze the system in the component level so as to trace down the origin of quality variations. The burnt-in system reliability and burn-in costs can be determined in terms of analyzing the component reliability.

7.4.1 Decompose the system to component level

Imagine a complex piece of equipment as a large number of component positions into each of which there is inserted a component. A component position includes the component installed in it as well as the connections that connect the component to the system. As illustrated in Figure 7.1(a)-(c), possible defects in the component position include component defect and component connection defect.

Denote by K the number of component positions in this system. A component designated to component position I, $I=1,2,...,K$ is called a Type i component. A normal Type i component, i.e. before assembly, has CDF $F_i(t)$ and SF $\bar{F}_i(t)$. After assembly, it becomes defective with probability p_i. A defective Type i component has CDF $H_i(t)$. Therefore, the SF of the Type i component after assembly, denoted as $\bar{F}_i^*(t)$, can be expressed as

$$\bar{F}_i^*(t)=(1-p_i)\bar{F}_i(t)+p_i\bar{H}_i(t). \tag{7.4}$$

The number of series connection defects is denoted as N_0. The number of connection defects in position I is N_i, i = 1, 2, …, k. N_i, i = 0, 1, …, k, are non-negative integer-valued *r.v.*s with Pmf $\pi_i(n)$. Assume that the times to detect the series connection defects and the component connection defects are *i.i.d. r.v.*s with CDF $G(t)$ and SF $\bar{G}(t)$.

Appropriate distributions for N_i, i = 0, 1, …, k, and for time to detect of a connection defect can be found from past experiences if some proven component blocks are reused. For example, a Bernoulli distribution is appropriate for N_i, i = 1, 2, …, k since we may be concerned with whether connection defect exists in position I or not. On the other hand, there can be a number of series connection defects in a system, especially when the system is large and complex. Therefore, the Poisson and the negative binomial distributions are plausible models for N_0.

With the help of Test during Burn-in (cf. Kececioglu and Sun 1997), functionalities of all components in the system can be monitored through standard burn-in testing facilities so that any component failures and failures resulting from connection defects can be immediately detected, regardless of the system structure. To simplify the formulation, some more assumptions are made below.

(a) Lifetimes of all components and connection defect failure times are all independent.

(b) If a Type *i* component fails during burn-in, it is replaced with a normal one with SF $\bar{F}_i(t)$, at the cost of $C_{r,i}$. When a connection defect is detected, it is perfectly removed with repair cost C_d. The process of failure detection, trouble location, and replacement is assumed to consume no appreciable time.

(c) The unit time burn-in operational cost is a constant C_0. After burn-in with duration b, a mission time τ is set for a burnt-in system.

(d) Once the system fails during field operation, we assume that repair is not allowed and a breakdown cost of C_f is incurred.

Remark 1: We do not further divide the normal components into the conforming and the non-conforming ones due to the fact that lifetime of a non-conforming component is usually much longer than time to defect failures (Jiang and Murthy 2009). The principal purpose of system level burn-in is to screen the assembly defects illustrated in Figure 7.1(b)-(d).

Remark 2: Assumption (d) is a legitimate simplification for many commercial products with a warranty contract, e.g. the renewing replacement free warranty. For products sold under other warranty policy, the probability that it fails more than two times during the warranty period is pretty small, e.g. see Meeker *et al.* (2009) for some justifications.

7.4.2 Burnt-In System Reliability

Component failures at position I constitute a delayed renewal process $\{Y, X_1, X_2, ..., X_n, ...\}$ where Y has SF $\bar{F}_i^*(t)$ and X_n, $n \geq 1$ have SF $\bar{F}_i(t)$ (cf. Ross 1996, Chap. 3.5). When the system undergoes a burn-in procedure with duration b, remaining lifetime of the component in position I is the excess life of the delayed renewal process. Its SF $\bar{F}_{i,b}^*(t)$ is given by the renewal equation as

$$\bar{F}_{i,b}^*(t) = \bar{F}_i^*(t+b) + \int_0^b \bar{F}_i(t+b-x)\,dm_i^*(x), \tag{7.5}$$

where $m_i^*(t)$ is the renewal function for the delayed renewal process.

For a specific connection defect, if it is detected during the burn-in test, it is perfectly removed. For example, if a cavity in a weld spot results in position failure during burn-in, we may expect that this cavity will be perfectly rectified. The probability that this defect is not detected during burn-in but detected within the mission time τ is $G(\tau+b)-G(b)$. Therefore, the probability that this defect is not detected within the mission time is $G(b)+\bar{G}(\tau+b)$. The SF of component position I can thus be expressed as

$$R_{I,\tau}(b) = \bar{F}_{i,b}^*(\tau)\cdot\sum_{n=0}^{\infty}\left[\bar{G}(\tau+b)+G(b)\right]^n \pi_i(n). \tag{7.6}$$

If a Bernoulli distribution is assumed for N_i with

$$\Pr\{N_i = 1\} = 1 - \Pr\{N_i = 0\} = q_i,$$

Equation (7.6) boils down to

$$R_{I,\tau}(b) = \bar{F}_{i,b}^*(\tau)\cdot\left\{1 - q_i\left[\bar{G}(b) - \bar{G}(\tau+b)\right]\right\}. \tag{7.7}$$

The series connection defects can be handled in a similar vein. If we assume that N_0 follows a Poisson distribution with rate μ_0, the probability that no series connection defect is detected within the mission time is

$$R_{0,\tau}(b)=\exp\left(-\mu_0\left(\bar{G}(b)-\bar{G}(b+\tau)\right)\right). \tag{7.8}$$

Once the system structure is known, the system reliability can be readily derived based on reliabilities of all component positions and reliability of the series connections. Let ϕ denote the system structure. The system reliability, denoted as $R_{\phi,\tau}(b)$, is a function of the burn-in duration *b*, the system structure ϕ, the reliability of each component position $R_{I,\tau}(b)$, *I*=1, 2, …, *K*, the number of series connection defects N_0 and the mission time τ. To reflect these relations, we express $R_{\phi,\tau}(b)$ as

$$R_{\phi,\tau}(b)=\Phi\left(R_{0,\tau}(b),R_{1,\tau}(b),\ldots,R_{K,\tau}(b)\right). \tag{7.9}$$

7.4.3 Cost Formulation

The burn-in cost includes burn-in operational cost and rectification of connection defect failures and component replacements. During burn-in, if a specific connection defect is detected, it is permanently removed with cost C_d. The probability of this occurrence is $G(b)$. There are N_i connection defects in position *I*. For each connection defect, we define a corresponding indicator variable Z_n. Then we have a sequence of independent indicator variables $\{Z_1,Z_2,\cdots,Z_{N_i}\}$ such that

$$P\{Z_n=1\}=1-P\{Z_n=0\}=G(b).$$

All these random variables are independent of N_i. Obviously, N_i is a stopping time for the sequence $\{Z_n;n\geq 1\}$. Therefore, the number of connection defects detected during burn-in in component position *I* is $\sum_{n=1}^{N_i}Z_n$ where by convention $\sum_{n=1}^{N_i}Z_n=0$ when $N_i=0$. Its mean value is given by

$$M_i(b) = E\left\{\sum_{n=1}^{N_i} Z_n\right\} = EN_i \cdot EZ_n = G(b) \cdot \sum_{n=0}^{\infty} n\pi_i(n). \tag{7.10}$$

The total number of connection defects detected is the sum of connection failures in all component positions as well as in the series connections. Its expected value is given by

$$M(b) = G(b) \cdot \sum_{i=0}^{k} \sum_{n=0}^{\infty} n\pi_i(n). \tag{7.11}$$

If a Type *i* component fails during burn-in, it is replaced with a functionally equivalent component with no component defect, i.e. a normal one. Let $\eta_i(b)$ be the number of component replacements in position *I* during burn-in. The replacement cost in position *I* is $C_{r,i} \cdot \eta_i(b)$. Its expected value is $C_{r,i} \cdot m_i^*(b)$, where $m_i^*(b) = E\{\eta_i(b)\}$ is the renewal function of the delayed renewal process. The expected number of renewals until *b* can be expressed as

$$m_i^*(b) = \sum_{n \geq 1} F_i^* * F_i^{(n-1)}(b), \tag{7.12}$$

where $F_i^{(n-1)}(b)$ is the (*n*-1)-fold convolution of $F_i(b)$. Therefore, the expected cost due to component replacements during burn-in is given by $\sum_{i=1}^{k} C_{r,i} m_i^*(b)$. The expected burn-in cost can thus be expressed as

$$C_B(b) = C_d \cdot M(b) + \sum_{i=1}^{k} C_{r,i} m_i^*(b) + C_0 b. \tag{7.13}$$

Once the system fails during field operation, a breakdown cost of C_f is incurred. This cost includes the warranty cost and the administrative cost. The probability that the burnt-in system fails in field operation is $1 - R_{\phi,\tau}(b)$, Therefore, the total field failure cost is

$$C_O = C_f \cdot \left(1 - R_{\phi,\tau}(b)\right).$$

It follows that the cost objective is given by

$$J_1(b) = C_B(b) + C_f \cdot \left(1 - R_{\phi,\tau}(b)\right). \quad (7.14)$$

Note by passing that our cost function adopts a simple form by only capturing the principal cost components, but yet is able to quantify the costs incurred by both burn-in and field operation. With the burnt-in reliability function (7.9) and the total cost function (7.14) in place, we are in a position to develop the bi-objective system level burn-in model.

7.4.4 The Bi-Objective Model

The reliability objective can be normalized as $R_{\phi,\tau}(b)/\tilde{R}_{\phi,\tau}$, where

$$\tilde{R}_{\phi,\tau} = \Phi\left(\bar{F}_1(\tau), \bar{F}_2(\tau), \ldots, \bar{F}_K(\tau)\right)$$

is the system reliability without any assembly defects. Under this context, the ratio $R_{\phi,\tau}(b)/\tilde{R}_{\phi,\tau}$ can be interpreted as the screening strength. On the other hand, we normalize the cost objective as $J_1(b)/J_1(0)$, where $J_1(0)$ is the total cost without burn-in. Using (7.2), the objective function is a convex combination of the screening strength and the normalized cost as

$$\begin{aligned} \min_b \quad & w_1 \frac{J_1(b)}{J_1(0)} - (1 - w_1) \frac{R_{\phi,\tau}(b)}{\tilde{R}_{\phi,\tau}} \\ s.t. \quad & 0 \le b \le b_U, \end{aligned}$$

where b_U is the maximum allowable burn-in duration that might be imposed by the manufacturer. When the manufacturer does not have such constraint, we can simply set $b_U = \infty$. When $w_1 = 0$, this problem reduces to the performance-based model proposed by Kim and Kuo (2009). When $w_1 = 1$, this problem is purely cost-based.

7.5 Model Analysis and Optimization

In this section, some properties of the bi-objective system-level burn-in model are investigated, after which a numerical algorithm is introduced to search for the optimal burn-in duration. For systems consisting of a few components, and systems comprising many identical components, modeling and analysis using our algorithm are quite efficient. However, many systems comprise a large variety of different components, which may render an insufferably long computational time if our methodology is applied. For these complex systems, we derive lower and upper bounds for the burnt-in reliability functions and the delayed renewal functions. These bounds are used to approximate the objective function. Simulation studies reveal that these bounds are very tight.

7.5.1 Bounds for the Optimal Burn-In Duration

As w_1 varies from 0 to 1, the optimal burn-in duration changes. Denote this optimal duration as $b_{w_1}^*$. Then the optimal burn-in time that maximize the system reliability is b_0^* while the optimal burn-in time that minimize the total cost function is b_1^*. Lower and upper bounds for the optimal burn-in time are desirable as they define the range of all possible choices. Some useful results are given in Theorem 7.1.

Theorem 7.1. When $0 \le w_1 \le 1$, the optimal burn-in time $b_{w_1}^*$ of the bi-objective model is decreasing in w_1. Moreover, $b_{w_1}^*$ is not greater than b_0^* and is not less than b_1^*, i.e. $b_1^* \le b_{w_1}^* \le b_0^*$ for any w_1.

Proof: When $w_1 \neq 0$, by substituting (7.14) into the objective function of the bi-objective model, we can rewrite the objective function as

$$J\left(b \mid w_1\right)=\frac{C_B\left(b\right)+C_f}{J_1\left(0\right)}-\left[\frac{1-w_1}{w_1\tilde{R}_{\phi,\tau}}+\frac{C_f}{J_1\left(0\right)}\right]R_{\phi,\tau}\left(b\right). \tag{7.15}$$

For a specific choice of w_1, its optimal burn-in time is $b^*_{w_1}$. For any $b<b^*_{w_1}$, we have

$$J\left(b \mid w_1\right)\geq J\left(b^*_{w_1} \mid w_1\right). \tag{7.16}$$

A useful result shall be proved first. For any $b<b^*_{w_1}$, because the burn-in cost is an increasing function of burn-in duration, we can know

$$C_B\left(b\right)<C_B\left(b^*_{w_1}\right).$$

If we suppose

$$R_{\phi,\tau}\left(b\right)\geq R_{\phi,\tau}\left(b^*_{w_1}\right),$$

we have $J\left(b \mid w_1\right)<J\left(b^*_{w_1} \mid w_1\right)$ from (7.15), which contradicts with (7.16). Therefore for any $b<b^*_{w_1}$, the following relation should always hold

$$R_{\phi,\tau}\left(b\right)<R_{\phi,\tau}\left(b^*_{w_1}\right). \tag{7.17}$$

Consider another choice w_1' with $w_1'<w_1$, $J\left(b \mid w_1'\right)$ can be expressed as

$$J\left(b \mid w_1'\right)=J\left(b \mid w_1\right)+\left(\frac{1-w_1}{w_1}-\frac{1-w_1'}{w_1'}\right)\frac{R_{\phi,\tau}\left(b\right)}{\tilde{R}_{\phi,\tau}}. \tag{7.18}$$

Note that the second term on the right hand size is negative. By substituting (7.16) and (7.17) into (7.18), we have that for any $b<b^*_{w_1}$,

$$J\left(b \mid w_1'\right)>J\left(b^*_{w_1} \mid w_1'\right).$$

Therefore,

$$b^*_{w_1'}\geq b^*_{w_1}. \tag{7.19}$$

Inequality (7.19) implies that $b^*_{w_1}$ is decreasing in w_1. When $w_1\to 1$, the model reduces to the case of minimizing the cost function. When $w_1 \to 0$, the model reduces to the case of minimizing the expected number of field failures. Therefore, $b_{w_1}^*$ is not greater than b_0^* and $b_{w_1}^*$ is not less than b_1^*. Therefore, the theorem holds.

■

This theorem provides meaningful insights into the burn-in problems. When $w_1 = 1$, the model is purely cost-based. The optimal burn-in duration achieves a balance between the burn-in cost and costs due to field failures. When $w_1 < 1$, there is a trade-off between reliability and costs in the bi-objective model, though there are some correlations between these two elements. An alternative way to understand the model is to treat the reliability objective as a penalty, under which we are able to convert the bi-objective model into a purely cost-based one. If more weight is assigned to the total costs, a relatively short burn-in duration can be employed since burn-in is costly. This is at the expense of a lower system reliability. If the reliability is important, which is embodied in a small w_1, a longer burn-in duration is desirable to attain a higher screening strength, at the price of greater costs.

Remark: Given the burnt-in reliability (7.9) and the total cost (7.14) as above, the optimal burn-in duration under the performance-based burn-in model, i.e. $w_1 = 0$ is always greater than that under the cost-based burn-in model, i.e. $w_1 = 1$.

7.5.2 An Optimization Algorithm

Kim and Kuo (2009) derived some close-form expressions of $\bar{F}_{i,b}^*(t)$ for some simple distributions. However, these expressions are very complicated. For general distributions, however, $\bar{F}_{i,b}^*(t)$ in (7.5) and $m_i^*(b)$ in (7.12) have no closed-form expressions. Therefore, it is almost impossible to evaluate the objective function in the bi-objective model by exact methods, let along the optimal solution. An algorithm using the concept of RS sums is proposed in this section to numerically compute (7.5) and (7.12). Based on this method, we propose a new algorithm to optimize the bi-objective burn-in model.

RS sums method

Consider a delayed renewal process $\{Y, X_1, X_2, ..., X_n, ...\}$ where Y has SF $\bar{F}^*(t)$ and X_n, $n \geq 1$ have SF $\bar{F}(t)$. Denote $W(b)$ as the excess life of this process at time b and $\bar{F}_b^*(t)$ as the SF of $W(b)$. By conditioning on the last renewal before b, we have

$$\bar{F}_b^*(t) = \bar{F}^*(t+b) + \int_0^b \bar{F}(t+b-x)\,dm^*(x), \tag{7.20}$$

where $m^*(t)$ is the renewal function for the delayed renewal process. Conditional on the first arrival, $m^*(t)$ can be expressed as

$$m^*(t) = F^*(t) + \int_0^t F^*(t-x)\,dm(x), \tag{7.21}$$

where $m(t)$ is the renewal function for the renewal process $\{X_1, X_2, ..., X_n, ...\}$ with renewal equation

$$m(t) = F(t) + \int_0^t F(t-x)\,dm(x). \tag{7.22}$$

Xie (1989) proposed a simple numerical method to compute $m(t)$ based on the RS sums. To use this method, the interval $[0, b]$ is uniformly divided with $0 = t_0 < t_1 < \ldots < t_n = b$. Then $m(t_j)$, $1 \leq j \leq n$ can be calculated iteratively. Its expression was derived by Xie (1989) and is given by

$$m(t_j) \approx \frac{F(t_j) + \sum_{k=1}^{j-1} F(t_j - t_{k-1/2})[m(t_k) - m(t_{k-1})] - F(t_j - t_{j-1/2})m(t_{j-1})}{1 - F(t_j - t_{j-1/2})}, \tag{7.23}$$

where $m(t_0) = 0$ and $t_{j-1/2} = \frac{1}{2}\left(t_{j-1} + t_j\right)$. Once $m(t_1)$, $m(t_2)$, …, $m(t_j)$ are calculated, we can approximate $m^*(t_j)$, $1 \le j \le n$ based on the RS sums method, which is given by

$$m^*\left(t_j\right) \approx \bar{F}^*\left(t_j\right) + \sum_{k=1}^{j} \bar{F}^*\left(t_j - t_{k-1/2}\right)\left[m\left(t_k\right) - m\left(t_{k-1}\right)\right]. \tag{7.24}$$

Similarly, we can use the RS sums method to approximate $\bar{F}_{t_j}^*\left(t\right)$, $1 \le j \le n$, which is given by

$$\bar{F}_{t_j}^*\left(t\right) = \bar{F}^*\left(t + t_j\right) + \sum_{k=1}^{j} \bar{F}\left(t + t_j - t_{k-1/2}\right)\left[m^*\left(t_k\right) - m^*\left(t_{k-1}\right)\right]. \tag{7.25}$$

The proposed numerical algorithm

As will be verified in Section 7.6, the RS sums method is very accurate. Technically, we can compute the objective function by the RS sums method and then use some optimization algorithms that do not require analytical gradients, e.g. the simplex search method (cf. the FMINSEARCH function in Matlab®), to find the optimum iteratively. However, the RS method itself is also an iterative algorithm. An iterative algorithm invoking another iteration algorithm is often time-consuming. We have noticed that there is only one decision variable in the bi-objective model. Grid search technique has been shown to be powerful and computationally effective in solving problems with a small number of variables (Lerman 1980). By using the grid search, we only need to invoke the RS algorithm once to compute $\bar{F}_{i,\bar{b}}^*\left(\tau\right)$, where $\bar{b}$ is an upper bound for the burn-in duration. Then $\bar{F}_{i,t_j}^*\left(\tau\right)$, $1 \le j < n$ are by-products in the course of obtaining $\bar{F}_{i,\bar{b}}^*\left(\tau\right)$. The general procedure to combine the grid search and the RS method is as follows.

Procedure 1.

(i) Determine an upper bound $\bar{b}$ for the burn-in duration. Uniformly divide the interval $[0, \bar{b}]$ with points $0 = t_0 < t_1 < \ldots < t_n = \bar{b}$.

(ii) Consider component position I. For $1 \le j \le n$, compute $m_i(t_j)$ by using (7.23). Then for

$1 \le j \le n$, compute $m_i^*\left(t_j\right)$ in (7.12) by using (7.24). After that, for $1 \le j \le n$, compute $F_{i,t_j}^*\left(\tau\right)$ of (7.5) by using (7.25).

(iii) For $1 \le j \le n$, compute the system reliability $R_{\phi,\tau}(t_j)$ and the total cost $J_1(t_j)$ by using the results from Step (ii).

(iv) For a given w_1, substitute $R_{\phi,\tau}(t_j)$ and $J_1(t_j)$ into the bi-objective model and determine a grid point t_j that maximizes the objective function. A plot of the objective function versus the burn-in time is also helpful in the determination of the optimal burn-in time.

■

When $b_U < \infty$, we can set $\bar{b} = b_U$. when $b_U = \infty$, we will show how to determine $\bar{b}$ in the next subsection. For computational purpose, n should not be too large. According to our simulations in Section 7.6, $n = 1000$ is accurate enough while the run time is moderate.

When there are not too many distinct component positions in a system, the above algorithm requires moderate computational time, and is very accurate. If the system comprises multifarious components, the algorithm may not be efficient enough. In practice, duration of the burn-in procedure is short because (a) burn-in is costly and (b) latent defects lead to failures easily. When b is small, some accurate approximations for $R_{\phi,\tau}(b)$ and $m_i^*\left(t\right)$ can be derived. The remainder of this section focuses on deriving such bounds.

7.5.3 Bounds for Burnt-In Reliability and Delayed Renewal Function

In the following, we shall first derive tight bounds for $R_{\phi,\tau}(b)$. Both the upper and the lower bounds, or the average of them, can be used to approximate $R_{\phi,\tau}(b)$. The upper limit $\bar{b}$ in the new algorithm can also be determined through these bounds.

Theorem 7.2 provides upper and lower bounds for the reliability function $R_{\phi,\tau}(b)$. The second part of this theorem requires that the CDF F_i, i = 1, 2, …, k, for the normal components are new better than used (NBU). It should be noted that this is a legitimate assumption because without assembly defects, it is commonly believed that a conforming component exhibits a gradually increasing failure rate. For more discussions about properties of NBU, interested readers are referred to Lai and Xie (2006). In this theorem, it is implicitly assumed that the system has a monotonic structure, i.e. a coherent system.

Theorem 7.2.

(a) When a system undergoes burn-in with duration b, an upper bound for the system reliability after burn-in is given by

$$R_{\phi,\tau}^{U}(b)=\Phi\left(R_{0,\tau}(b),R_{1,\tau}^{U}(b),R_{2,\tau}^{U}(b),\ldots,R_{K,\tau}^{U}(b)\right), \tag{7.26}$$

where

$$R_{I,\tau}^{U}(b)=\left[\bar{F}_i^{*}(\tau+b)+\bar{F}_i(\tau)F_i^{*}(b)/\bar{F}_i(b)\right]\cdot\sum_{n=0}^{\infty}\left[\bar{G}(\tau+b)+G(b)\right]^{n}\pi_i(n). \tag{7.27}$$

An lower bound for the system reliability after burn-in is given by

$$R_{\phi,\tau}^{L}(b)=\phi\left(R_{0,\tau}(b),R_{1,\tau}^{L}(b),R_{2,\tau}^{L}(b),R_{K,\tau}^{L}(b)\right) \tag{7.28}$$

where

$$R_{I,\tau}^{L}(b)=\left[\bar{F}_i^{*}(\tau+b)+\bar{F}_i(\tau+b)F_i^{*}(b)\right]\cdot\sum_{n=0}^{\infty}\left[\bar{G}(\tau+b)+G(b)\right]^{n}\pi_i(n) \tag{7.29}$$

(b) If we further know F_i is new better than used (NBU), i.e. $\bar{F}_i(t+b)\le\bar{F}_i(t)\bar{F}_i(b)$ for any t, b > 0, (7.27) can be further tightened by using

$$R_{I,\tau}^{U}(b)=\left[\bar{F}_i^{*}(\tau+b)+\bar{F}_i(\tau)F_i^{*}(b)\right]\cdot\sum_{n=0}^{\infty}\left[\bar{G}(\tau+b)+G(b)\right]^{n}\pi_i(n). \tag{7.30}$$

Proof: Consider component position I. Let $S_{N_i(b)}$ denote the time of the last component replacement before *b*. The SF of the excess life $W_i(b)$ can be obtained by conditioning on $S_{N_i(b)}$ as

$$\bar{F}_{i,b}^{*}(\tau) = \bar{F}_i^{*}(\tau + b) + \int_0^b \frac{\bar{F}_i(\tau + b - u)}{\bar{F}_i(b - u)} d\Pr\left\{S_{N_i(b)} \leq u\right\}.$$

It is easy to see that

$$\max\left\{\frac{\bar{F}_i(\tau + b - u)}{\bar{F}_i(b - u)} : 0 \leq u \leq b\right\} \leq \frac{\bar{F}_i(\tau)}{\bar{F}_i(b)}.$$

According to properties of the Riemann-Stieltjes integral (cf. Thm. 10.10 in Protter 1998),

$$\bar{F}_{i,b}^{*}(\tau) \leq \bar{F}_i^{*}(\tau + b) + \frac{\bar{F}_i(\tau)}{\bar{F}_i(b)}\left\{\Pr\left\{S_{N_i(b)} \leq b\right\} - \Pr\left\{S_{N_i(b)} \leq 0\right\}\right\} = \bar{F}_i^{*}(\tau + b) + \frac{\bar{F}_i(\tau)}{\bar{F}_i(b)}\left[F_i^{*}(b) - 0\right].$$

Therefore,

$$\bar{F}_{i,b}^{*}(\tau) \leq \bar{F}_i^{*}(\tau + b) + \frac{\bar{F}_i(\tau)}{\bar{F}_i(b)} F_i^{*}(b), \tag{7.31}$$

and thus (7.27) follows. Most of the time, a system is assumed to have a monotone structure. From the definition of coherent system (Barlow and Proschan 1996), the following inequality follows.

$$R_{\phi,\tau}(b) \leq R_{\phi,\tau}^{U}(b).$$

If we further know that F_i is NBU, We have $\bar{F}_i(t + b) \leq \bar{F}_i(t)\bar{F}_i(b)$. Then

$$\max\left\{\frac{\bar{F}_i(\tau + b - u)}{\bar{F}_i(b - u)} : 0 \leq u \leq b\right\} = \bar{F}_i(\tau).$$

Following the same procedure as that used in deriving (7.31), we arrive at (7.30).

Consider Equation (7.5),

$$\bar{F}_{i,b}^{*}(t) = \bar{F}_i^{*}(t + b) + \int_0^b \bar{F}_i(t + b - x)\, dm_i^{*}(x).$$

It is easy to see that

$$\min\left\{\bar{F}_i\left(\tau+b-u\right):0\le u\le b\right\}=\bar{F}_i\left(\tau+b\right).$$

According to properties of the Riemann-Stieltjes integral,

$$\bar{F}_{i,b}^{*}\left(\tau\right)\ge\bar{F}_i^{*}\left(\tau+b\right)+\bar{F}_i\left(\tau+b\right)m_i^{*}\left(b\right). \tag{7.32}$$

We have known that

$$m_i^{*}\left(b\right)=\sum_{n\ge1}F_i^{*}*F_i^{(n-1)}\left(b\right)\ge F_i^{*}\left(b\right).$$

Substituting this result into (7.32) yields

$$\bar{F}_{i,b}^{*}\left(\tau\right)\ge\bar{F}_i^{*}\left(\tau+b\right)+\bar{F}_i\left(\tau+b\right)F_i^{*}\left(b\right).$$

Therefore,

$$R_{I,\tau}\left(b\right)\le R_{I,\tau}^{L}\left(b\right).$$

For a system with monotonic structure, the following relationship holds.

$$R_{\phi,\tau}\left(b\right)\ge R_{\phi,\tau}^{L}\left(b\right)$$

■

Both the lower and the upper bounds do not involve the renewal functions and the SF of the excess lives for the delayed renewal processes, and thus can be easily computed.

Moreover, these bounds can be used to determine $\bar{b}$. By plotting the upper bound (7.26) and lower bound (7.28) for the burnt-in system reliability in the same figure, the upper limit $\bar{b}$ can be determined as follows.

Procedure 2.

- First, locate the maximum of the lower bound curve, which is denoted by $\left(b_1,R_{\phi,\tau}^{L*}\right)$.
- Draw a horizontal line traversing $\left(b_1,R_{\phi,\tau}^{L*}\right)$. This line penetrates the upper bound curve

and separates it to two parts. We shall focus on the upper part of this curve and denote $[b_2, b_3]$ as range of the *b* values of this upper part.

- Then b_0^* should be somewhere at $[b_2, b_3]$. Therefore, b_3 can serve as the upper limit, as can be known from Theorem 7.1.

In our bi-objective burn-in model, the delayed renewal function in (7.12) is also difficult to compute. We can use the lower bound below for this function.

Lower Bound for the Delayed Renewed Function. When a system undergoes burn-in with duration *b*, a lower bound for the expected replacement in component position *I* is given by

$$m_i^*(b) \geq F_i^*(b). \tag{7.33}$$

■

This result can be easily obtained by noting that, from (7.12), we have

$$m_i^*(b) = \sum_{n\geq 1} F_i^* * F_i^{(n-1)}(b) \geq F_i^*(b).$$

Tightness of this bound is examined in Section 7.6.

7.5.4 Approximation Method for the Bi-Objective Model

With the bounds provided in Theorems 7.2, an approximation method is proposed to approximate the objective function. We propose approximating the objective function by using

$$R_{\phi,\tau}(b) \approx \left[R_{\phi,\tau}^U(b) + R_{\phi,\tau}^L(b)\right]/2,$$

where $R_{\phi,\tau}^U(b)$ and $R_{\phi,\tau}^L(b)$ are given by (7.26) and (7.30), respectively. The delayed renewal functions are approximated by using (7.33). After these approximations, the approximated objective function in the bi-objective model has close form expression, and can be computed easily and efficiently. The optimum can be easily located by some simple but yet efficient one-dimensional search algorithm, e.g., the bi-sectional search algorithm.

7.6 Simulation Studies

In this section, simulation studies are carried out to verify the accuracy of the RS sums method, and the tightness of bounds provided in Theorem 7.2 and Equation (7.33).

7.6.1 Accuracy of the RS Sums Method

We examine the accuracy of the RS sums method by comparing (7.25) with the exact values of $\bar{F}_b^*(t)$. In fact, closed form expressions for $\bar{F}_b^*(t)$ rarely exist. Fortunately, if we assume that lifetime of a normal component follows $EXP(\lambda_1)$ and that lifetime of a defective component follows $EXP(\lambda_0)$ with defective probability *p*, the SF of the excess life can be derived as

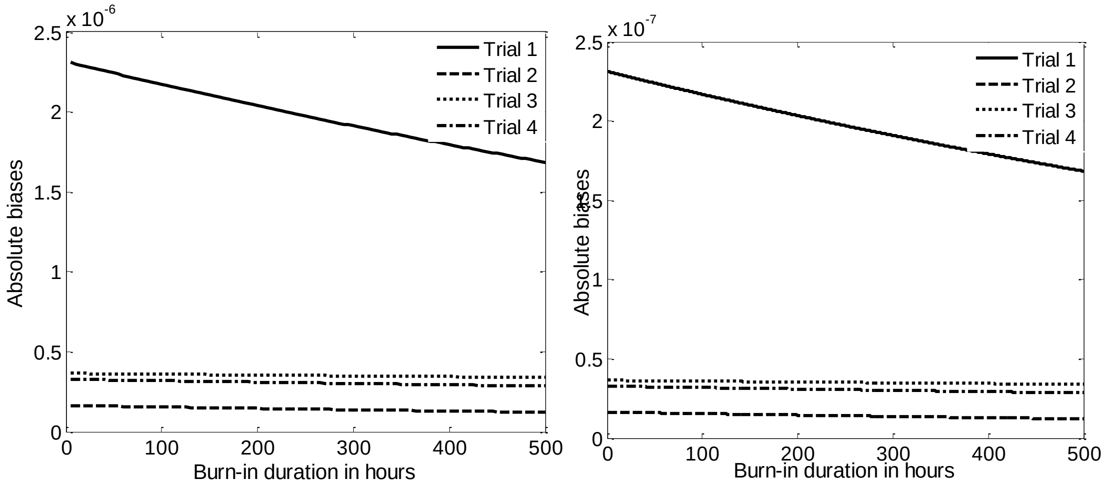
$$\bar{F}_b^*(t) = p\bar{H}(t+b) + \left(1 - p\bar{H}(b)\right)\bar{F}(t). \tag{7.34}$$

where $\bar{H}(t) = \exp(-\lambda_0 t)$ and $\bar{F}(t) = \exp(-\lambda_1 t)$. There are three parameters, i.e. λ_0, λ_1 and *p*. To evaluate the accuracy of the RS sums method, we examine the parameter settings in Table 7.1, which are analogous to a 2^{3-1} design. Values of the other parameters are τ = 1 year and $\bar{b}$ = 500 hours.

Table 7.1. Parameter settings

Trial	$1/\lambda_0$	$1/\lambda_1$	p
1	7E-4	3E-6	0.05
2	7E-4	1E-6	0.01
3	3E-4	3E-6	0.01
4	3E-4	1E-6	0.05

We try to plot the exact values of (7.34) and the approximations obtained from (7.25) on the same figure. Nevertheless, the approximation is so good that it is unable to distinguish these two curves. Therefore, we depict the absolute values of the difference in Figure 7.3.



(a) (b)

Figure 7.3. Absolute values of the difference between the exact values and those from the RS sums method: (a) n = 100 and (b) n = 1000

As visualized in Figure 7.3, the absolute biases of the RS sums method are quite small, which indicates the accuracy of this method. Another interesting observation from this simulation is that the absolute biases are of order $O\left(n^{-1}\right)$, as can be seen from the two plots in Figure 7.3. Actually, n = 100 is accurate enough, as the order of magnitude of the absolute biases is about 10^{-6}.

7.6.2 Accuracy of the Burnt-In Reliability Bounds

Consider a component position where there are no connection defects. Suppose a normal component has SF $\bar{F}$ while a defective one has SF $\bar{H}$. The defective probability is *p*. To examine the accuracy of the bounds, we focus on the difference between the upper and the lower bounds. According to (7.27) and (7.29), we have

$$\bar{F}^*(\tau+b)+\bar{F}(\tau+b)F^*(b)\le \bar{F}_b^*(\tau)\le \bar{F}^*(\tau+b)+\bar{F}(\tau)F^*(b)/\bar{F}(b).$$

That is,

$$\Delta_1(b,\tau)=\left[\bar{F}(\tau)/\bar{F}(b)-\bar{F}(\tau+b)\right]F^*(b). \qquad (7.35)$$

If we further know that *F* is NBU, the gap between the upper and the lower bounds can be narrowed down to

$$\Delta_2(b,\tau)=\left[\bar{F}(\tau)-\bar{F}(\tau+b)\right]F^*(b). \qquad (7.36)$$

We let τ = 1 year and $\bar{b}$ = 500 hours. Several combinations of *p*, $\bar{F}$ and $\bar{F}^*$ are examined, as listed in Table 7.2. To make the results comparable, these combinations are carefully chosen such that the expected lifetimes of a normal and a defective component are approximately 2×10^5 and 10^3, respectively.

Table 7.2. Some combinations of the normal and defective components

Comb.	Normal	p_i	Defective
1	$EXP\left(2\times10^5\right)$	0.03	$EXP\left(10^3\right)$
2	$\text{Weibull}\left(10^6/4.4,\ 3\right)^*$	0.06	$\text{Weibull}\left(10^3/3,\ 1/3\right)$
3	$\text{gamma}\left(10^5,2\right)^{**}$	0.06	$\text{gamma}(500,2)$
4	$\ln N(10.7,\ 1)$	0.03	$\ln N(6.6,\ 0.8)$
5	$IG\left(2\times10^5,10^5\right)$	0.06	$IG\left(10^3,4\times10^2\right)$

* The SF for $\text{Weibull}(\alpha,\beta)$ is $\exp\left(-(t/\alpha)^\beta\right)$

** The PDF for $\text{gamma}(\alpha,\beta)$ is $\dfrac{t^{\beta-1}}{\alpha^\beta\Gamma(\beta)}\exp(-t/\alpha)$

The *Fs* in the first three combinations are NBU, (7.36) can thus be used. The gaps are depicted in Figure 7.4(a). For the 4th and the 5th combinations, it is well known that their *F*s are not NBU, (7.35) need to be used. The gaps are depicted in Figure 7.4(b).

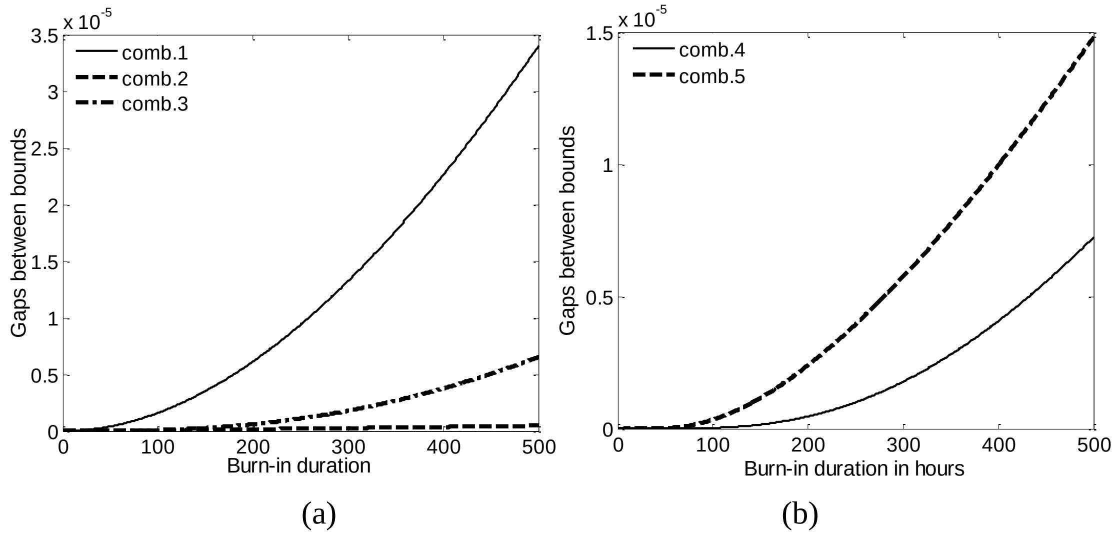


Figure 7.4. Gaps between the lower and the upper bounds

From Figure 7.4, we can see that the gaps between the lower and the upper bounds are very narrow. Therefore, they are tight enough to approximate the burnt-in reliability.

7.6.3 Accuracy of the Lower Bound for the Delayed Renewal Function

Next, we will examine the lower bound (7.33) for the delayed renewal functions. For illustrative purpose, we only consider Comb. 2 and 4 in Table 7.2. Because it is almost impossible to derive explicit expressions for the delayed renewal function, we use simulation to determine this function. To ensure the accuracy, we simulate each delayed renewal process for $N_{sim} = 10^7$ times. During each simulation, arrival time of each renewal is recorded. Then the simulated $m_i^*(t)$ can be calculated as follows.

$$m_i^*(t) = \frac{\text{total no. of renewals before } t}{N_{sim}}$$

The simulated $m_i^*(t)$ and the lower bound (7.33) are depicted on the same figure to compare the tightness of the bound, as shown in Figure 7.5.

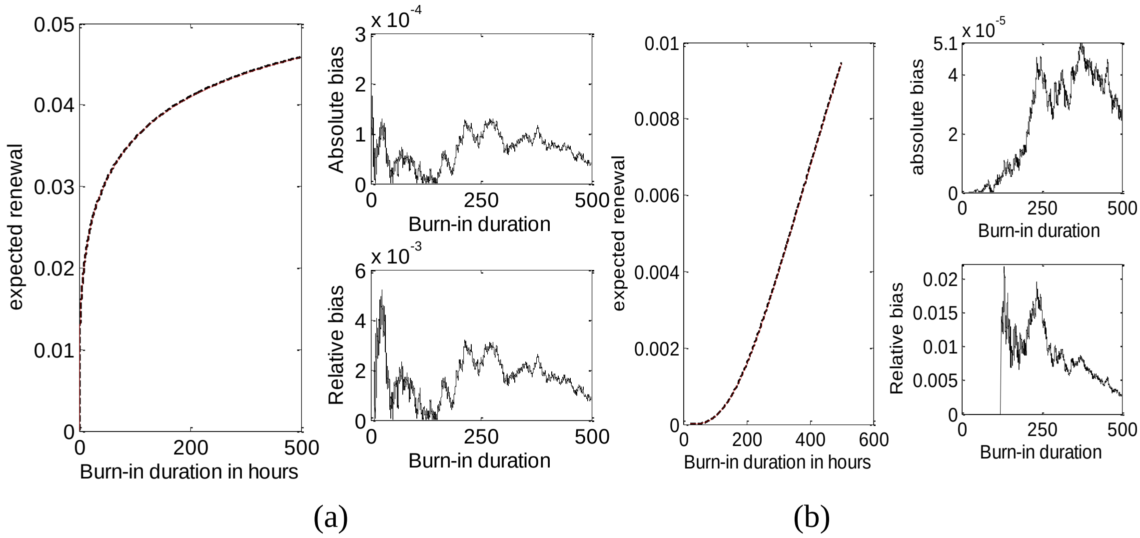


Figure 7.5. Lower bound of $m_i^*(t)$ and the biases compared with the simulated values: (a) for Comb. 2; (b) for Comb. 4

In Figure 7.5, $m_i^*(t)$ approximated by (7.33) and computed from simulation almost overlap, which manifests the tightness of (7.33). This is because burn-in duration is often not too long. More than 1 failure in a component position is extremely rare. It should be noted that the biases in Figure 7.5 are not very stable, this is because we use simulated $m_i^*(t)$ in lieu of the exact ones.

7.7 An Illustrative Example

We illustrate our methods by analyzing the parallel-series system in detail. Its structure is depicted in Figure 7.6.

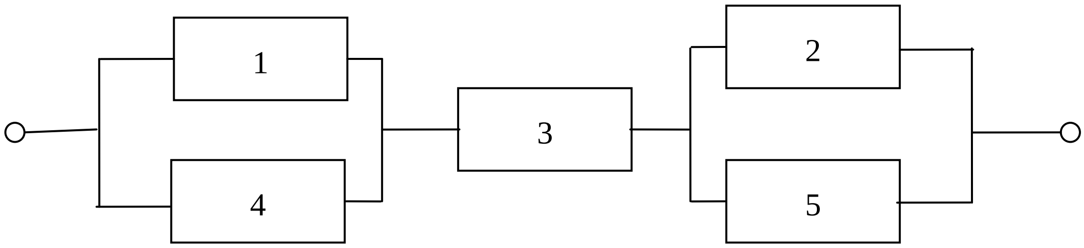


Figure 7.6.Structure of the parallel-series system

Suppose that the distribution for G is $\text{Weibull}\left(2\times10^{3},0.85\right)$, $N_0 \sim \text{Poisson}\left(\mu_0=0.018\right)$ and $N_i \sim \text{Bernoulli}\left(q_i\right)$ for i=1, 2, …, k. The system reliability function can be derived as

$$\begin{aligned} R_{\phi,\tau}(b) &= \exp\left(-\mu_0\left(\bar{G}(b)-\bar{G}(b+\tau)\right)\right)\times \\ &\quad \left\{1-\left[1-R_{1,\tau}(b)\right]\left[1-R_{2,\tau}(b)\right]\right\}\cdot R_{3,\tau}(b)\cdot\left\{1-\left[1-R_{4,\tau}(b)\right]\left[1-R_{5,\tau}(b)\right]\right\}, \end{aligned}$$

where $R_{I,\tau}(b)$, I = 1, 2, …, K, are given by (7.7). SF for each type of component and the values of q_i are given in Table 7.3. For illustrative purpose, a mixture of exponential distributions is used for Type 1 component to represent the effect of imperfect inspection. The cost profiles are given in Table 7.4.

Table 7.3. SF and q_i for Type 1-5 components

i	$\bar{F}_i(t)$	p_i	$\bar{H}_i(t)$	q_i
1	$0.99EXP\left(8\times10^{5}\right)+0.01EXP\left(3\times10^{5}\right)$	0.03	$EXP\left(4\times10^{2}\right)$	0.01
2	$\text{Weibull}\left(5\times10^{5},\ 1.2\right)$	0	———	0.006
3	$\text{gamma}\left(3\times10^{5},3\right)$	0.05	$\text{gamma}\left(2\times10^{2},0.8\right)$	0.02
4	$\ln N(11.6,\ 1.2)$	0.04	$\ln N(6.1,\ 0.9)$	0.015
5	$IG\left(4\times10^{4},1.5\times10^{5}\right)$	0.02	$IG\left(5\times10^{2},3\times10^{2}\right)$	0.04

Table 7.4. The cost configuration

Parameters	$C_{r,1}$	$C_{r,2}$	$C_{r,3}$	$C_{r,4}$	$C_{r,5}$	C_d	C_f	C_0
Values	\$55	\$48	\$75	\$64	\$21	\$5	\$249	\$0.03/h

Assume that τ = 10 years, the upper limit can be obtained from Procedure 2, which is $\bar{b} = 2283$. Within this upper limit, the optima of the bi-objective model for different choices of w_1 can be located by using Procedure 1. We also use the approximation method described in Section 7.4.4 to find the optima. Results from these two methods are depicted in Figure 7.7.

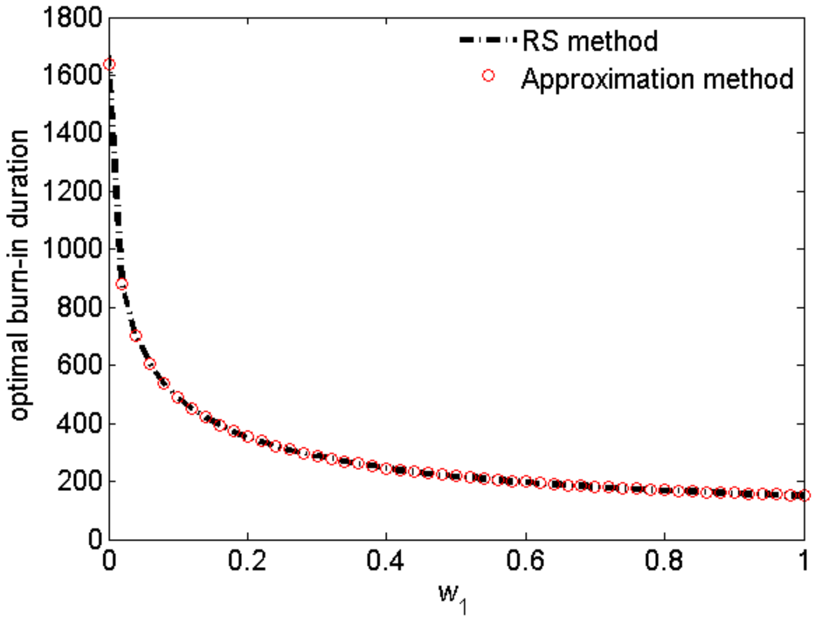


Figure 7.7. Optimal burn-in durations determined by the new algorithm and the approximation method.

As can be seen from Figure 7.7, optimal burn-in durations determined through the new algorithm and the approximation method tally very well. These two algorithms are also quite efficient in this example. For each choice of w_1, both algorithm successfully terminate within 2 seconds by using Matlab® in our desktop with Intel Core 2 Duo CPU E6750 @ 2.66G Hz. Figure 7.7 also suggests that the optimal burn-in duration is decreasing in w_1, which is concordant with Theorem 7.1.

Another interesting observation from Figure 7.7 is that the curve is rather flat when $w_1 > 0.4$.

This means that within this interval, the optimal burn-in decision is not sensitive to the weight. However when w_1 is relatively small, a small variation of w_1 would lead to significant change of the decision. These results imply that if the manufacturers deem that the cost is relatively important, then it is not necessary to evaluate the weight accurately. But if they believe that the performance is more important, the relative weight should be estimated with more care.

7.8 Conclusions

Many cost-based and performance-based burn-in models have been proposed in the literature. Performance- and cost-based burn-in lay stress on product's intangible losses and book costs, respectively. A legitimate burn-in strategy should allow the manufacturer to make a trade-off between these two objectives. This study bridges the gap between the cost-based and the performance-based burn-in models by developing a unifying framework for the determination of optimal burn-in durations. An intuitive way to understand this framework is to regard it as a cost model by treating the performance objective as a cost penalty. As such, this framework facilitates quantification of the intangible losses by simply requiring the relative weight of the performance objectives. From another point of view, this model allows for incorporation of manufacturers' preferences in burn-in decision makings.

We successfully use this framework to build a bi-objective system-level burn-in model by decomposing the system into component level. Many companies adopt a modular approach and reuse proven component blocks in new product design. Analyzing the system in component level enables us to reuse infant mortality data of these blocks when making burn-in decisions for a new product. However, difficulties with regard to this approach are that evaluation of the bi-objective model requires dealing with the delayed renewal processes. We develop two numerical techniques to solve this problem. The RS approach is an iterative algorithm which is often time-consuming. But due to the special solution method of this approach, it can be perfectly combined with the grid search technique. When the system is comprised of a large variety of components, we propose using approximation method. Lower and upper bounds for the burnt-in system reliability are derived. They are used to approximate the objective function. Both of these two numerical techniques are very accurate, as validated by our simulation and numerical example.

CHAPTER 8 CONCLUSIONS AND FUTURE RESEARCH

Motivated by a couple of practical problems, this thesis has developed several burn-in models for different types of products under complex failure processes from some new perspectives. These models are able to furnish as theoretical as well as practical guidance for burn-in practitioners in dealing with optimal burn-in decisions. The contributions of this thesis are summarized as follows.

Chapter 3 investigates a burn-in criterion based on change point of the PRL-*p* function. When the probability of warranty failure is pre-specified, this change point naturally gives rise to an optimal burn-in duration. Moreover, the maximal PRL-*p* represents the maximum allowable warranty period when the expected field return is set at *p*. We present some properties of this change point, and derive the asymptotic distribution of its parametric maximum likelihood estimator and that of the corresponding PRL-*p*. The procedure is applied to estimate a set of desirable burn-in duration and the corresponding warranty period. An example using the modified Weibull extension model is given to illustrate the procedure. The methodologies are then applied to the car engine problem faced by Volvo.

Chapter 4 proposes and studies a new burn-in modeling approach for repairable products sold with a two-dimensional warranty. More specifically, we characterize two types of failures, i.e., normal and defect failures, and develop both performance and cost-based burn-in models under the non-renewing free repair warranty policy. Our models subsume the special cases of one-dimensional warranty, allow different failure modes to have distinct accelerated relationships, and take the consumer usage heterogeneity into consideration. Under some mild assumptions, it is shown that the optimal burn-in usage rate should be as high as possible, provided that no extraneous failure modes are introduced. Furthermore, we show that the optimal burn-in duration determined from the performance-based model is not shorter than that from the cost-based model. Numerical examples are used to demonstrate the benefits of burn-in. In addition, the sensitivity analysis reveals the importance of designed reliability in terms of defect detection.

Motivated by the infant mortalities in many Micro-Electro-Mechanical Systems (MEMS), Chapter 5 develops degradation-based burn-in maintenance models under the age and the block based maintenances, respectively. Both models assume that the product population comprises weak and normal subpopulations. Degradation of the product follows the Wiener processes while the weak and the normal subpopulations possess distinct drift parameters. The objective of joint burn-in and maintenance decisions is to minimize the long run average cost per unit time during field use by properly choosing the burn-in settings and the preventive replacement intervals. An example of the MEMS devices is used to demonstrate effectiveness of these two models.

Chapter 6 develops a burn-in planning framework for products with competing risks. Existing burn-in approaches are confined to single failure mode based on the assumption that this failure mode is subject to infant mortality. Considering the prevalence of multiple modes of failures and the high reliability of modern products, our framework differentiates between normal and infant mortality failure modes and recommends degradation-based burn-in approaches. This framework is employed to guide the burn-in planning for an electronic device subject to both degradation-threshold failure, which is an infant mortality mode and can be modeled by a Gamma process with random effect, and a catastrophic mode, which is normal. Three cost-based burn-in models are built and the optimal cut-off degradation levels are derived. Their validity is demonstrated by the electronic device example. We also propose three approaches to deal uncertainties due to parameter estimation.

Chapter 7 develops a bi-objective framework for burn-in decision makings to achieve an optimal trade-off between the cost and the performance objectives. Decisions derived from performance-based burn-in models often yield high total book costs, but cost-based burn-in models often lead to relatively poorer product performance compared with the former ones. Under the proposed framework, a convex combination of the cost and performance objectives allows manufacturers to specify the relative weights and achieve a best-compromise solution. Based on this framework, we build a system-level burn-in model by analyzing system failures at the component level. We prove that the optimal burn-in duration is decreasing in the weight assigned to the normalized cost. To obtain the optimal burn-in duration, we develop an efficient numerical algorithm that combines an approximation to Riemann-Stieltjes Integral and the grid search technique. We also derive tight bounds for the burnt-in reliability function and the delayed renewal functions. These bounds are then used to approximate the objective function of the burn-in model.

On the whole, a number of contributions have been achieved in this thesis. Nevertheless, some further research is necessary to extend our research. Some possible topics for future research are as follows.

Chapter 3 focuses on the parametric inference. Further research on nonparametric estimation and the corresponding asymptotic behaviors of the PRL-*p* function may be explored in the future. Chapter 4 is a first step towards modeling burn-in for products sold with a two-dimensional warranty. We have assumed that all defect failures are *iid*. It can be modified to the case where different defects have different failure distributions. It is also possible that some defects occur only if usage or usage rate is above some threshold. Other warranty policies other than the FRW, e.g. non-renewing replacement free policy and some renewing policies reviewed by Murthy and Blischke (2006) should be considered. Other servicing strategies (e.g. Murthy and Jack 2007) and shapes for the warranty region (e.g. Murthy *et al.* 1995) also need further study. In addition, other objective functions including customer satisfactions can be considered. It would be important to examine the effect of usage heterogeneity on optimal burn-in decisions under the three existing approaches summarized in Chapter 4.1. Chapter 5 implicitly assumes that degradation of a burnt-in unit is not monitored during field use. This is reasonable for small-size items, or products that are not very expensive. But when the product is expensive, we may also monitor its field degradation and make the dynamic maintenance decision. This deserves further investigation. In addition, sometimes, the measurement error of the degradation is not negligible. It would be interesting to see how burn-in decisions changes under this scenario. Chapter 6 studies optimal burn-in planning under independent competing risks. The case of dependent competing risks deserves further investigation. In this chapter, we consider the case where degradation level is measured only after burn-in, i.e., single inspection point. When the measurement cost is low and the unit time burn-in operational cost is high, it might be more cost effective to consider multiple inspection points, each associated with a cut-off level. In addition, products are often produced and shipped in batch (Huang and Ye 2010). Burn-in under such a setting will be different because of limited burn-in population and batch and batch variations. Burn-in models under this circumstance also need investigation.

REFERENCES

Arney, S. (2001) Designing for MEMS reliability. *MRS Bulletin*, 26 (4), 296-299.

Attardi, L., Guida, M. and Pulcini, G. (2005) A mixed-Weibull regression model for the analysis of automotive warranty data. *Reliability Engineering & System Safety*, 87 (2), 265-273.

Audet, C. and Dennis, J.E. (2006) Mesh adaptive direct search algorithms for constrained optimization. *SIAM Journal on Optimization*, 1 (17), 188–217.

Bagdonavicius, V. and Nikulin, M.S. (2002) *Accelerated Life Models: Modeling and Statistical Analysis*, Chapman and Hall/CRC, London.

Balachandran, K.R. and Radhakrishnan, S. (2005) Quality implications of warranties in a supply chain. *Management Science*, 51 (8), 1266-1277.

Barlow, R.E. and Proschan, F. (1996) *Mathematical Theory of Reliability*, SIAM, Philadelphia.

Barndorff-Nielsen, O.E. and Cox, D.R. (1996) Prediction and asymptotics. *Bernoulli*, 2 (4), 319-340.

Bebbington, M., Lai, C.D. and Zitikis, R. (2007) Optimum burn-in time for a bathtub shaped failure distribution. *Methodology and Computing in Applied Probability*, 9 (1), 1-20.

Bebbington, M., Lai, C.D. and Zitikis, R. (2008) Estimating the turning point of a bathtub-shaped failure distribution. *Journal of Statistical Planning and Inference*, 138 (4), 1157-1166.

Block, H.W., Jong, Y.K. and Savits, T.H. (1999) Bathtub functions and burn-in. *Probability in the Engineering and Informational Sciences*, 13 (4), 497-507.

Cha, J.H. (2000) On a better burn-in procedure. *Journal of Applied Probability*, 37 (4), 1099-1103.

Cha, J.H. (2001) Burn-in procedures for a generalized model. *Journal of Applied Probability*, 38 (2), 542-553.

Cha, J.H. (2003) A further extension of the generalized burn-in model. *Journal of Applied Probability*, 40 (1), 264-270.

Cha, J.H. (2005) On optimal burn-in procedures - A generalized model. *IEEE Transactions on Reliability*, 54 (2), 198-206.

Cha, J.H. (2006) An extended model for optimal burn-in procedures. *IEEE Transactions on Reliability*, 55 (2), 189-198.

Cha, J.H. and Mi, J. (2007) Optimal burn-in procedure for periodically inspected systems. *Naval Research Logistics*, 54 (7), 720-731.

Cha, J.H., Yamamoto, H. and Yun, W.Y. (2008) Optimal burn-in for minimizing total warranty cost. *IEICE Transactions on Fundamentals of Electronics Communications and Computer Sciences*, E91A (2), 633-641.

Cha, J.H. and Finkelstein, M. (2010a) Burn-in by environmental shocks for two ordered subpopulations. *European Journal of Operational Research*, 206 (1), 111-117.

Cha, J.H. and Finkelstein, M.S. (2010b) Stochastically ordered subpopulations and optimal burn-in procedure. *IEEE Transactions on Reliability*, 59 (4), 635-643.

Cha, J.H. (2011) A survey of burn-in and maintenance models for repairable systems, in *Replacement Models with Minimal Repair*, Springer, London, pp. 179-203.

Cha, J.H. and Finkelstein, M. (2011) Burn-in and the performance quality measures in heterogeneous populations. *European Journal of Operational Research*, 210 (2), 273-280.

Chang, D.S. (2000) Optimal burn-in decision for products with an unimodal failure rate function. *European Journal of Operational Research*, 126 (3), 534-540.

Chen, N., Chen, Y., Li, Z., Zhou, S. and Sievenpiper, C. (2011) Optimal variability sensitive condition-based maintenance with a Cox PH model. *International Journal of Production Research*, 49 (7), 2083-2100.

Chen, Z. (2000) A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function. *Statistics & Probability Letters*, 49 (2), 155-161.

Chou, K. and Tang, K. (1992) Burn-in time and estimation of change point with Weibull exponential mixture distribution. *Decision Sciences*, 23 (4), 973-990.

Chukova, S. and Johnston, M.R. (2006) Two-dimensional warranty repair strategy based on minimal and complete repairs. *Mathematical and Computer Modelling*, 44 (11-12), 1133-1143.

Coit, D.W. and Smith, A.E. (2002) Genetic algorithm to maximize a lower-bound for system time-to-failure with uncertain component Weibull parameters. *Computers & Industrial Engineering*, 41 (4), 423-440.

Crowder, M.J. (2001) *Classical Competing Risks*, Chapman & Hall/CRC Press, Boca Raton.

Csörgő, M. and Csörgő, S. (1987) Estimation of percentile residual life. *Operations Research*, 35 (4), 598-606.

Drapella, A. and Kosznik, S. (2002) Combining preventive replacement and burn-in procedures. *Quality and Reliability Engineering International*, 18 (5), 423-427.

Efron, B. and Tibshirani, R. (1993) *An Introduction to the Bootstrap*, Chapman & Hall/CRC Press, New York.

Escobar, L.A. and Meeker, W.Q. (2006) A review of accelerated test models. *Statistical Science*, 21 (4), 552-577.

Folks, J.L. and Chhikara, R.S. (1978) The inverse Gaussian distribution and its statistical application--A review. *Journal of the Royal Statistical Society-Series B*, 40 (3), 263-289.

Franco-Pereira, A.M., Lillo, R.E., Romo, J. and Shaked, M. (2011) Percentile residual life orders. *Applied Stochastic Models in Business and Industry*, 27 (3), 235-252.

Gupta, R.D. and Kundu, D. (2001) Generalized exponential distribution: Different method of estimations. *Journal of Statistical Computation and Simulation*, 69 (4), 315-337.

Hogan, T.J., Barker, P. and Douglass, M.R. (2003) Burn-in test reduction for the Digital Micromirror Device. *Proceedings of SPIE*, 4980 317-323.

Hong, Y., Meeker, W.Q. and McCalley, J.D. (2009) Prediction of remaining life of power transformers based on left truncated and right censored lifetime data. *The Annals of Applied Statistics*, 3 (2), 857-879.

Hong, Y.L. and Meeker, W.Q. (2010) Field-failure and warranty prediction based on auxiliary use-rate information. *Technometrics*, 52 (2), 148-159.

Huang, B. and Ye, Z.S. (2010) The effects of lumpy demand and shipment size constraint: A response to "Revisit the note on supply chain integration in vendor-managed inventory". *Decision Support Systems*, 48 (2), 421-425.

Huang, W. and Askin, R.G. (2003) Reliability analysis of electronic devices with multiple competing failure modes involving performance aging degradation. *Quality and Reliability Engineering International*, 19 (3), 241-254.

Huang, Y.S. and Yen, C. (2009) A study of two-dimensional warranty policies with preventive maintenance. *IIE Transactions*, 41 (4), 299-308.

Jack, N., Iskandar, B.P. and Murthy, D.N.P. (2009) A repair-replace strategy based on usage rate for items sold with a two-dimensional warranty. *Reliability Engineering & System Safety*, 94 (2), 611-617.

Jeong, J.H., Jung, S.H. and Costantino, J.P. (2008) Nonparametric inference on median residual life function. *Biometrics*, 64 (1), 157-163.

Jiang, R. and Jardine, A.K.S. (2007) An optimal burn-in preventive-replacement model associated with a mixture distribution. *Quality and Reliability Engineering International*, 23 (1), 83-93.

Jiang, R. and Murthy, D.N.P. (2009) Impact of quality variations on product reliability. *Reliability Engineering & System Safety*, 94 (2), 490-496.

Joe, H. and Proschan, F. (1984) Percentile residual life functions. *Operations Research*, 32 (3), 668-678.

Johnson, L.A. (2006) Laser diode burn-in and reliability testing. *IEEE Communications Magazine*, 44 (2), 7-10.

Jung, S.H., Jeong, J.H. and Bandos, H. (2009) Regression on quantile residual life. *Biometrics*, 65 (4), 1203-1212.

Kar, T.R. and Nachlas, J.A. (1997) Coordinated warranty & burn-in strategies. *IEEE Transactions on Reliability*, 46 (4), 512-518.

Kececioglu, D. and Sun, F.B. (1994) Mixed-Weibull parameter estimation for burn-in data using the Bayesian approach. *Microelectronics and Reliability*, 34 (10), 1657-1679.

Kececioglu, D. and Sun, F.B. (1997) *Burn-In Testing: Its Quantification and Optimization*, Prentice Hall, New Jersey.

Kim, K.O. and Kuo, W. (2003) A general model of heterogeneous system lifetimes and conditions for system burn-in. *Naval Research Logistics*, 50 (4), 364-380.

Kim, K.O. and Kuo, W. (2005) Some considerations on system burn-in. *IEEE Transactions on Reliability*, 54 (2), 207-214.

Kim, K.O. and Kuo, W. (2009) Optimal burn-in for maximizing reliability of repairable non-series systems. *European Journal of Operational Research*, 193 (1), 140-151.

Kim, K.O. and Kuo, W. (2011) Component and System Burn-in for Repairable Systems. *IIE Transactions*, 43 (11), 773-782.

Lagarias, J.C., Reeds, J.A., Wright, M.H. and Wright, P.E. (1999) Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM Journal on Optimization*, 9 (1), 112-147.

Lai, C.D., Xie, M. and Murthy, D.N.P. (2003) A modified Weibull distribution. *IEEE Transactions on Reliability*, 52 (1), 33-37.

Lai, C.D. and Xie, M. (2006) *Stochastic Ageing and Dependence for Reliability*, Springer, New York.

Launer, R. (1993) Graphical techniques for analyzing failure data with the percentile residual-life function. *IEEE Transactions on Reliability*, 42 (1), 71-75.

Lawless, J.F. and Crowder, M.J. (2004) Covariates and random effects in a gamma process model with application to degradation and failure. *Lifetime Data Analysis*, 10 (3), 213-227.

Lawless, J.F., Crowder, M.J. and Lee, K.A. (2009) Analysis of reliability and warranty claims in products with age and usage scales. *Technometrics*, 51 (1), 14-24.

Lawrence, M.J. (1966) An investigation of the burn-in problem. *Technometrics*, 8 (1), 61-71.

Lee, Y.C., Parviz, A., Chiou, J.A. and Chen, S. (2003) Packaging for microelectromechanical and nanoelectromechanical systems. *IEEE Transactions on Advanced Packaging*, 26 (3), 217-226.

Lerman, P.M. (1980) Fitting segmented regression models by grid search. *Journal of the Royal Statistical Society-Series C*, 29 (1), 77-84.

Lindsay, B.G. and Li, B. (1997) On second-order optimality of the observed Fisher information. *The Annals of Statistics*, 25 (5), 2172-2199.

Liu, X. and Mazzuchi, T.A. (2008) The optimal burn-in: State of the art and new advances for cost function formulation, in *Recent Advances in Reliability and Quality in Design*, Springer, London, pp. 137-182.

Liu, X. and Tang, L.C. (2010) Accelerated life test plans for repairable systems with multiple independent risks. *IEEE Transactions on Reliability*, 59 (1), 115-127.

Ma, Y. and Yin, G. (2010) Semiparametric median residual life model and inference. *Canadian Journal of Statistics*, 38 (4), 665-679.

Majeske, K.D. (2003) A mixture model for automobile warranty data. *Reliability Engineering & System Safety*, 81 (1), 71-77.

Majeske, K.D. (2007) A non-homogeneous Poisson process predictive model for automobile warranty claims. *Reliability Engineering & System Safety*, 92 (2), 243-251.

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

Meeker, W.Q., Escobar, L.A. and Hong, Y.L. (2009) Using accelerated life tests results to predict product field reliability. *Technometrics*, 51 (2), 146-161.

Mi, J. (1993) Discrete bathtub failure rate and upside - down bathtub mean residual life.

Naval Research Logistics, 40 (3), 361-371.

Mi, J. (1994a) Burn-in and maintenance policies. *Advances in Applied Probability*, 26 (1), 207-221.

Mi, J. (1994b) Maximization of a survival probability and its application. *Journal of Applied Probability*, 31 (4), 1026-1033.

Mi, J. (1995) Bathtub failure rate and upside-down bathtub mean residual life. *IEEE Transactions on Reliability*, 44 (3), 388-391.

Mi, J. (1996) Minimizing some cost functions related to both burn-in and field use. *Operations Research*, 44 (3), 497-500.

Mi, J. (1997) Warranty policies and burn-in. *Naval Research Logistics*, 44 (2), 199-209.

Mi, J. (1999) Comparisons of renewable warranties. *Naval Research Logistics*, 46 (1), 91-106.

Mi, J. (2003) Optimal burn-in time and eventually IFR. *Journal of the Chinese Institute of Industrial Engineers*, 20 (5), 533-543.

MIL-STD-883G (2006) *Test Methods and Procedures for Microelectronics*, Department of Defense, Washington.

Mudholkar, G.S., Srivastava, D.K. and Freimer, M. (1995) The exponentiated Weibull family: A reanalysis of the bus-motor-failure data. *Technometrics*, 37 (4), 436-445.

Murthy, D.N.P., Iskandar, B.P. and Wilson, R.J. (1995) Two-dimensional failure-free warranty policies: Two-dimensional point process models. *Operations Research*, 43 (2), 356-366.

Murthy, D.N.P., Xie, M. and Jiang, R. (2004) *Weibull Models*, Wiley, Hoboken.

Murthy, D.N.P. and Blischke, W.R. (2006) *Warranty Management and Product Manufacture*, Springer London.

Murthy, D.N.P. and Jack, N. (2007) Warranty servicing. *Encyclopedia of Statistics in Quality and Reliability*, 2091-2097.

Murthy, D.N.P., Rausand, M. and Virtanen, S. (2009) Investment in new product reliability.

Reliability Engineering & System Safety, 94 (10), 1593-1600.

Nelson, W. (1990) *Accelerated Testing: Statistical Models, Test Plans and Data Analyses*, Wiley, New Jersey.

Nguyen, D.G. and Murthy, D.N.P. (1982) Optimal burn-in time to minimize cost for products sold under warranty. *IIE Transactions*, 14 (3), 167-174.

Nikulin, M.S., Limnios, N. and Balakrishnan, N. (2009) *Advances in Degradation Modeling: Applications to Reliability, Survival Analysis, and Finance*, Springer Verlag, Boston.

Peng, H., Feng, Q. and Coit, D.W. (2009) Simultaneous quality and reliability optimization for microengines subject to degradation. *IEEE Transactions on Reliability*, 58 (1), 98-105.

Pham, H. and Lai, C. (2007) On recent generalizations of the Weibull distribution. *IEEE Transactions on Reliability*, 56 (3), 454-458.

Pohl, E.A. and Dietrich, D.L. (1999) Optimal stress screening strategies formulti component systems sold under warranty: The case of phase type lifetimes. *Annals of Operations Research*, 91 (0), 137-161.

Protter, M.H. (1998) *Basic Elements of Real Analysis*, Springer, New York.

Rai, B. and Singh, N. (2006) Customer-rush near warranty expiration limit, and nonparametric hazard rate estimation from known mileage accumulation rates. *IEEE Transactions on Reliability*, 55 (3), 480-489.

Reddy, R.K. and Dietrich, D.L. (1994) A 2-level environmental-stress-screening (ESS) model: a mixed-distribution approach. *IEEE Transactions on Reliability*, 43 (1), 85-90.

Reichheld, F.F. and Teal, T. (2001) *The Loyalty Effect: The Hidden Force behind Growth, Profits, and Lasting Value*, Harvard Business Press.

Ross, S.M. (1996) *Stochastic Processes*, Wiley, New York.

Ross, S.M. (2007) *Introduction to Probability Models*, Academic Press.

Saaty, T.L. (1994) *Fundamentals of Decision Making and Priority Theory with The Analytic Hierarchy Process*, RWS Publications, Pittsburgh.

Sheu, S.H. and Chien, Y.H. (2004) Minimizing cost-functions related to both burn-in and field-operation under a generalized model. *IEEE Transactions on Reliability*, 53 (3), 435-439.

Sheu, S.H. and Chien, Y.H. (2005) Optimal burn-in time to minimize the cost for general repairable products sold under warranty. *European Journal of Operational Research*, 163 (2), 445-461.

Shuster, J. (1968) On the inverse Gaussian distribution function. *Journal of the American Statistical Association*, 63 (324), 1514-1516.

Smith, R. and Bain, L. (1975) An exponential power life-testing distribution. *Communications in Statistics-Theory and Methods*, 4 (5), 469-481.

Spall, J.C. (2003) *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*, Wiley, Hoboken.

Suzuki, K., Nakamoto, T. and Matsuo, Y. (2010) Optimum specimen sizes and sample allocation for estimating Weibull shape parameters for two competing failure modes. *Technometrics*, 52 (2), 209-220.

Tang, Y., Xie, M. and Goh, T.N. (2003) Statistical analysis of a Weibull extension model. *Communications in Statistics-Theory and Methods*, 32 (5), 913-928.

Tanner, D.M. (2009) MEMS reliability: Where are we now? *Microelectronics Reliability*, 49 (9-11), 937-940.

Tsai, C.C., Tseng, S.T. and Balakrishnan, N. (2011) Optimal burn-in policy for highly reliable products using Gamma degradation process. *IEEE Transactions on Reliability*, 60 (1), 234-245.

Tseng, S.T. and Tang, J. (2001) Optimal burn-in time for highly reliable products. *International Journal of Industrial Engineering*, 8 (4), 329-338.

Tseng, S.T., Tang, J. and Ku, I.H. (2003) Determination of burn-in parameters and residual life for highly reliable products. *Naval Research Logistics*, 50 (1), 1-14.

Tseng, S.T. and Peng, C.Y. (2004) Optimal burn-in policy by using an integrated Wiener process. *IIE Transactions*, 36 (12), 1161-1170.

Turner, M.D. (2010) A practical application of quantitative accelerated life testing in power systems engineering. *IEEE Transactions on Reliability*, 59 (1), 91-101.

Ulusoy, S.K., Mazzuchi, T.A. and Perlstein, D. (2011) Bayesian calculation of optimal burn in time using the joint criteria of cost and delivered reliability. *Quality and Reliability Engineering International*, 27 (4), 569-580.

van Noortwijk, J.M. (2009) A survey of the application of gamma processes in maintenance. *Reliability Engineering & System Safety*, 94 (1), 2-21.

Wang, H.Z. (2002) A survey of maintenance policies of deteriorating systems. *European Journal of Operational Research*, 139 (3), 469-489.

Watson, G.S. and Wells, W.T. (1961) On the possibility of improving the mean useful life of items by eliminating those with short lives. *Technometrics*, 3 (2), 281-298.

Weiss, G.H. and Dishon, M. (1971) Some economic problems related to burn-in programs. *Reliability, IEEE Transactions on*, 20 (3), 190-195.

Wu, C.C., Chou, C.Y. and Huang, C. (2007) Optimal burn-in time and warranty length under fully renewing combination free replacement and pro-rata warranty. *Reliability Engineering & System Safety*, 92 (7), 914-920.

Wu, S. and Xie, M. (2007) Classifying weak, and strong components using ROC analysis with application to burn-in. *IEEE Transactions on Reliability*, 56 (3), 552-561.

Xie, M. (1989) On the solution of renewal-type integral equations. *Communications in Statistics-Simulation and Computation*, 18 (1), 281-293.

Xie, M. (1991) *Software reliability modelling*, World Scientific Pub Co Inc.

Xie, M. and Lai, C.D. (1996) Reliability analysis using an additive Weibull model with bathtub-shaped failure rate function. *Reliability Engineering & System Safety*, 52 (1), 87-93.

Xie, M., Tang, Y. and Goh, T.N. (2002) A modified Weibull extension with bathtub-shaped failure rate function. *Reliability Engineering & System Safety*, 76 (3), 279-285.

Xie, M., Goh, T.N. and Tang, Y. (2004) On changing points of mean residual life and failure rate function for some generalized Weibull distributions. *Reliability Engineering & System*

Safety, 84 (3), 293-299.

Yan, L. and English, J.R. (1997) Economic cost modeling of environmental-stress-screening and burn-in. *IEEE Transactions on Reliability*, 46 (2), 275-282.

Ye, Z.S., Li, Z. and Xie, M. (2010) Some improvements on adaptive genetic algorithms for reliability-related applications. *Reliability Engineering & System Safety*, 95 (2), 120-126.

Ye, Z.S., Tang, L.C. and Xie, M. (2011a) A burn-in scheme based on percentiles of the residual life. *Journal of Quality Technology*, 43 (4), 334-345.

Ye, Z.S., Tang, L.C. and Xu, H.Y. (2011b) A distribution-based systems reliability model under extreme shocks and natural degradation. *IEEE Transactions on Reliability*, 60 (1), 246-256.

Ye, Z.S., Chan, P.S., Xie, M. and Ng, H.K.T. (2012a) Statistical inference for the extreme value distribution under adaptive Type-II progressive censoring schemes. *Journal of Statistical Computation and Simulation*, to appear.

Ye, Z.S., Murthy, D.N.P., Xie, M. and Tang, L.C. (2012b) Optimal burn-in for repairable products sold with a two-dimensional warranty. *IIE Transactions*, to appear.

Ye, Z.S., Shen, Y. and Xie, M. (2012c) Degradation-based burn-in with preventive maintenance. *European Journal of Operational Research*, to appear.

Ye, Z.S., Xie, M., Tang, L.C. and Chen, N. (2012d) Efficient semiparametric estimation of gamma processes for deteriorating products. *Technometrics*, to appear.

Ye, Z.S., Xie, M., Tang, L.C. and Shen, Y. (2012e) Degradation-based burn-in planning under competing risks. *Technometrics*, 54 (2).

Yuan, T. and Kuo, Y. (2010) Bayesian analysis of hazard rate, change point, and cost-optimal burn-in time for electronic devices. *IEEE Transactions on Reliability*, 59 (1), 132-138.

Yun, W.Y., Lee, Y.W. and Ferreira, L. (2002) Optimal burn-in time under cumulative free replacement warranty. *Reliability Engineering & System Safety*, 78 (2), 93-100.