Statistical Science
1997, Vol. 12, No. 1, 1–19

Burn-In

Henry W. Block and Thomas H. Savits

Abstract. A survey of recent research in burn-in is undertaken. The emphasis is on mixture models, criteria for optimal burn-in and burn-in at the component or system level.

Key words and phrases: Bathtub failure rate, mixture models, TTT plot, TTT transform, cost functions.

0. INTRODUCTION

Burn-in is a widely used engineering method to eliminate weak items from a standard population. The standard population usually consists of various engineering systems composed of items or parts, or components which are assembled together into systems. The components operate for a certain amount of time until they fail, as do the systems composed of these components. The systems might be electronic systems such as circuit boards and the components would be various types of chips and printed circuits. A typical mechanical system is an air conditioner and the components are condensor, fan, circuits and so forth. Usually within any population of components there are strong components with long lifetimes and weak components with much shorter lifetimes. To insure that only the strong components reach the customer, a manufacturer will subject all of the components to tests simulating typical or even severe use conditions. In theory, the weak components will fail, leaving only the strong components. This type of testing can also be carried out on systems after they are assembled in order to determine the weak or strong systems or to uncover defects incurred during assembly. These tests are known as burn-in. One important issue is to determine the optimal time for burn-in. Burn-in is more typically applied to electronic than to mechanical systems.

We give a survey of recent burn-in research with emphasis on mixture models (which are used to describe populations with weak and strong components), criteria for optimal burn-in and whether it is better to burn in at the system or component level. After some background, we give a brief description of the types of statistical distributions which model the lifetimes of components for which burn-in is relevant. The remainder of the paper is devoted to explicating recent promising developments in burn-in. Because of the authors's interests, most emphasis will be placed on probability modeling for burn-in, but some statistical topics will also be covered. We will not review the fairly extensive engineering literature on burn-in since this has been done in several review articles which we cite at the end of Section 1.

Section 1 contains several illustrative examples and an introduction to some references for additional background on burn-in. The distributions which are used to describe the lifetimes of components which can benefit from burn-in are given in Section 2. An important family of distributions is one in which the failure rate functions have a bathtub shape. In particular, distributions which arise as mixtures are singled out for emphasis since many bathtub-shaped failure rates arise in this way. In Section 3, various criteria are described which have been used to determine optimal burn-in times. Section 3.1 considers general criteria and Section 3.2 covers various cost structures. Section 4 discusses two recent mixture models. The first of these (Section 4.1) examines a typical heterogeneous population to which burn-in is often applied and how this translates into renewal intensity behavior. The second of these proposes a general mixture model. A related result involves the asymptotic failure rate of a mixture model in terms of the asymptotic failure rates of the components of the mixture. The question of whether it is better to burn in at the component or the system level is discussed in Section 5. In Section 6, we consider an important tool, the TTT transform, which is used for approximating burn-in times. Section 7 gives a brief introduction to some recent sequential burn-in procedures involving optimal control. Section 8 gives a discussion with an indication of some future research directions.

Henry W. Block and Thomas H. Savits are Professors, Department of Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260 (e-mail: hwb@stat.pitt.edu, ths@stat.pitt.edu).

1. BACKGROUND AND SIMPLE EXAMPLES

Many manufacturers and users of electronic components and systems, as a matter of course, subject these systems and/or components to initial testing for a fixed period of time under conditions which range from typical to those which approximate a worst-case scenario. A typical regimen is to introduce for a period of time some vibration and temperature elevation for a device. In a particular context this is sometimes known as "shake and bake." At the end of this period, those components and/or systems which do not survive this period of testing may be discarded (scrapped), analyzed for defects and/or repaired. Those which survive this period may be sold, placed into service or subjected to further testing. Although these procedures have a variety of names depending on the area of application, we use the term burn-in to describe them all. The time period will be called the burn-in period. We illustrate some of these ideas with the following three examples.

EXAMPLE 1. Rawicz (1986) considers 30-watt long-life lamps manufactured by the Pacific Lamp Corporation (Vancouver, Canada) which were designed "for 5,000 hours of constant work in severe environmental conditions at 120 V." These are installed on billboards where it is difficult and expensive to replace them. It turns out that a certain small percentage of these lamps tend not to last the requisite 5,000 hours but fail relatively early. Obviously it would be beneficial if this subpopulation of lamps could be identified and eliminated before being placed on a billboard. The procedure recommended involves stressing all of the lamps at a high voltage (240 V) for a short period of time, which causes the weak lamps to fail rather quickly while the stronger lamps do not fail during this period. The lamps which do not fail are the lamps potentially capable of surviving the 5,000 hours of constant work. Often the burn-in weakens the surviving devices. In this particular application, however, the surprising result is that the surviving lamps are actually improved. This was thought to occur since the high thermal treatment seemed to relax structural stresses caused by the fabrication process.

EXAMPLE 2. In the *AT&T Reliability Manual* (Klinger, Nakada and Menendez, 1990) an electronics switching system (the 5ESS Switch) is discussed. Immediately after manufacture this system is operated at room temperature (25°C) for 12 hours, during which "volume-call" testing is performed; that is, 1,000 calls are simulated and passed through each of the five to eight modules of the switch. The system is then subjected for up to 48 hours to the high temperature (50°C) which can occur within the switch if the air conditioning should fail. The first part of this procedure is to find and eliminate early system failures, and the second part simulates use in an extreme case which might occur. The objective of the second part is to accelerate aging, so that weak systems fail. It also provides data which can be used to see how this equipment compares to certain standards set for it.

EXAMPLE 3. Jensen and Petersen (1982) consider a piece of measuring equipment made up of approximately 4,000 components. They focus on several critical types of these components. One of these, called an IC-memory circuit, accounts for 35 of the 4,000 components. The bimodal Weibull distribution (i.e., a mixture of two Weibulls) is used to model this type of component and has the following survival function:

$$\bar{F}(t) = p\exp[-(t/n_1)^{\beta_1}] + (1-p)\exp[-(t/n_2)^{\beta_2}].$$

From the data, the values $p = 0.015$, $\beta_1 = 0.25$, $n_1 = 30$, $\beta_2 = 1$ and $n_2 = 10$ have been determined, but an explicit method is not given.

We illustrate the results of Block, Mi and Savits (1993) (which is discussed in Section 4.2) to obtain the optimal burn-in time for a reasonable cost function (we use CF1 of Section 3.2 in this example).

Assume that we would like to plan a burn-in for components of this type so that those surviving burn-in should function for a mission time of $\tau = 60$ units. If a circuit fails before the end of burn-in a cost $c_0 = q_0 C$, where $0 < q_0 < 1$, is incurred. If it fails after burn-in but before the mission time is over, a cost of C is incurred. If an item survives burn-in and the mission time, a gain of $K = kC$ is obtained. For illustrative purposes, we choose $q_0 = 0.5$ and $k = 0.05$.

We apply Theorem 2.1 of Block, Mi and Savits (1993). Let f be the density of the bimodal Weibull given above. It is not hard to show that $g(t) = f(t+\tau)/f(t)$ is increasing in t (either directly or by standard results) and goes from 0 (as $t \to 0$) to 1 (as $t \to \infty$). By the cited results an optimal burn-in time $0 < b^* < \infty$ exists and satisfies

$$g(b^*) = \frac{C - c_0}{C + K}.$$

For the values above we obtain the equation $g(b^*) = 0.476$, and solving graphically yields $b^* = 102.9$.

Even though we present Example 2 as an example of burn-in, in the *AT&T Reliability Manual* (Klinger, Nakada and Menendez, 1990), Example 2 is called a system reliability audit. Other terms which are often used are “screen” and “environmental stress screening” (ESS). *The AT&T Manual* (Klinger, Nakada and Menendez, 1990, page 52) defines a screen to be an application of some stress to 100% of the product to remove (or reduce the number of) defective or potentially defective units. Fuqua (1987, pages 11 and 44) concurs with the 100% but states that this may be an inspection and stress is not required. Fuqua (1987, page 11) describes ESS as a series of tests conducted under environmental stresses to disclose latent part and workmanship defects. Nelson (1990, page 39) is more specific and describes ESS as involving accelerated testing under a combination of random vibration and thermal cycling and shock.

Burn-in is described by the *AT&T Manual* (Klinger, Nakada and Menendez, 1990, page 52) as one effective method of screening (implying 100%) using two types of stress (temperature and electric field). Nelson (1990, page 43) describes burn-in as running units under design or accelerated conditions for a suitable length of time. Tobias and Trindade (1995, page 297) restrict burn-in to high stress only and require that it be done prior to shipment. Bergman (1985, page 15) defines burn-in in a more general way as a pre-usage operation of components performed in order to screen out the substandard components, often in a severe environment. Jensen and Petersen (1982) have more or less the same definition as Bergman.

For the purposes of this paper we use the term burn-in in a general way, similar to the usage of Jensen and Petersen (1982) and of Bergman (1985). We think of it as some pre-usage operation which involves usage under normal or stressed conditions. It can involve either 100% of the product or some smaller subgroup (especially in the case of complex systems as in Example 2) and it is not limited to eliminating weak components.

Many of the traditional engineering ideas concerning burn-in are discussed in the handbook of Jensen and Petersen (1982). This book is intended as a handbook for small or moderate-size electronics firms in order to develop a burn-in program. Consequently the book should be viewed in this spirit. Emphasis is on easy-to-apply methods and on graphical techniques. One important contribution of the book is to popularize the idea that components and systems to which burn-in is applied have lifetimes which can be modeled as mixtures of statistical distributions. Specifically components either come from “freak” or “main” populations and their lifetimes can be modeled as mixtures of Weibull distributions. Systems are assumed to inherit this dichotomous behavior, but the weaker population is called an “infant mortality” population. This population arises partly because of defects introduced by the manufacturing process.

Most reliabilty books familiar to the statistics community do not discuss burn-in. We mention three applied reliability books which discuss this topic. The first of these is the book by Tobias and Trindade (1995), which has a section on burn-in covering some basics. An engineering reliability book by Fuqua (1987) delineates the uses of burn-in (see Section 2.4 and Chapter 14) for electronic systems at the component, module (intermediate between component and system) and system level. Most useful is the *AT&T Reliability Manual* (Klinger, Nakada and Menendez, 1990), which discusses a particular burn-in distribution used at AT&T along with a variety of burn-in procedures and several examples of burn-in. Two papers which review the engineering literature on burn-in are Kuo and Kuo (1983) and Leemis and Beneke (1990).

2. BURN-IN DISTRIBUTIONS

For which components or systems is burn-in effective? Another way of posing this question is by asking, “For which distributions (which model the lifetimes of components or systems) is burn-in effective?” First, it seems reasonable to rule out classes of distributions which model wearout. The reason for this is that objects which become more prone to failure throughout their life will not benefit from burn-in since burn-in stochastically weakens the residual lifetime. Consequently, distributions which have increasing failure rate or other similar aging properties are generally not candidates for burn-in.

For burn-in to be effective, lifetimes should have high failure rates initially and then improve. Since those items which survive burn-in have the same failure rate as the original, but shifted to the left, burn-in, in effect, eliminates that part of the lifetime where there is a high initial chance of failure. The class of lifetimes having bathtub-shaped failure rates has this property. For this type of distribution the failure rate starts high (the infancy period), then decreases to approximately a constant (the middle life) and then increases as it wears out (old age). As suggested by the parenthetical remarks, this distribution is thought to describe human life and other biological lifetimes. Certain other mechanical and electronic lifetimes also can be approximated by these distributions. This type

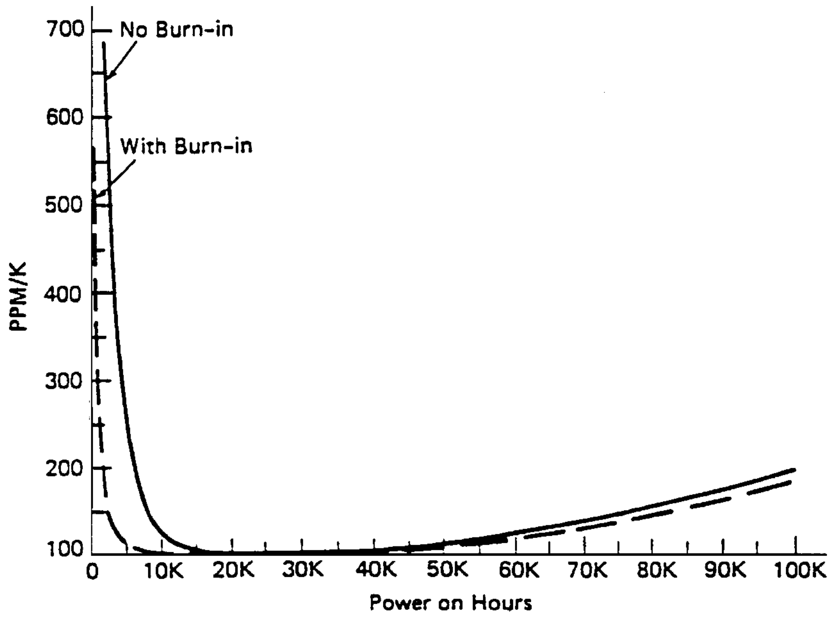

FIG. 1. *Burn-in improvement example* ($K = 1{,}000$ *hours*; $PPM/K =$ *parts per million per* 1,000 *hours*).

of distribution would seem to be appropriate for burn-in, since burn-in eliminates the high-failure infancy period, leaving a lifetime which begins near its former middle life (see Figure 1).

It turns out that there are reasons why many systems and components have bathtub-shaped failure rates. As described by Jensen and Petersen (1982), many industrial populations are heterogeneous and there are only a small number of different subpopulations. Although members of these subpopulations do not strictly speaking have bathtub-shaped failure rates, sampling from them produces a mixture of these subpopulations and these mixtures often have bathtub-shaped failure rates. For a simple example, assume that there are two subpopulations of components each of which is exponential, one with a small mean and one with a large mean. Sampling produces a distribution with decreasing failure rate which is a special case of the bathtub failure rate. An intuitive explanation of why this occurs is easy to give. Initially the higher failure rate of the weaker subpopulation dominates until this subpopulation dies out. After that, the lower failure rate of the stronger subpopulation takes over so that the failure rate decreases from the higher to the lower level. This type of idea, about the eventual domination of the strongest subpopulation, carries through for very general mixtures. See Block, Mi and Savits (1993, Section 4). A subjectivist explanation of the fact that mixing exponentials produces a decreasing failure rate distribution was given by Barlow (1985), who argued that even though a model may be exponential, information may change our opinion about the failure rate.

The mixture of two exponentials mentioned above produces a special case of the bathtub failure rate where no wearout is evident. Models of this type with no wearout are thought to be sufficient for modeling the lifetimes of certain electronic components, since these components tend to become obsolete before they wear out. Mixing two distributions which are more complex than exponentials yields distributions with more typical bathtub-shaped failure rates, as can be seen in the following example. A typical bathtub curve is given in Figure 8.2 of Tobias and Trindade (1995, page 238) which we reproduce in Figure 1.

This distribution is realized as a mixture of a lognormal and a Weibull distribution (both of which are used to model defectives) and another distribution (which models the population of normal devices),

$$\begin{aligned} F(t) &= 0.002\Phi\left(\frac{\ln(t/2{,}700)}{0.8}\right) \\ &\quad + 0.001\left(1-\exp\left[-\left(\frac{t}{400}\right)^{0.5}\right]\right) \\ &\quad + 0.997\Bigg(1-\exp\left(-10^{-7}t\right) \\ &\qquad \cdot\left[1-\Phi\left(\frac{\ln(t/975{,}000)}{0.8}\right)\right]\Bigg), \end{aligned}$$

where Φ is the standard normal cdf. Notice that the left tail of the distribution is very steep. This tail represents the period where many failures occur. Burn-in is utilized in order to remove this part of the tail. The dotted line represents the resulting distribution after a burn-in of several hours at an accelerated temperature. The point at which the curve flattens out and stops decreasing is at about 20K. This is called the first change point.

Many papers have appeared in the statistical literature providing models and formulas for bathtub-shaped failure rates. See Rajarshi and Rajarshi (1988) for a review of this topic and many references. One easy way of obtaining some of these is by mixing standard life distributions such as the exponential, gamma and Weibull. See Vaupel and Yashin (1985) for some illustrations of various distributions or Mi (1991) for an example of a simple mixture of gammas which has a bathtub-shaped failure rate. *The AT&T Reliability Manual* (Klinger, Nakada and Menendez, 1990) gives another model (called the AT&T model) for the failure rate of an electronics component. The early part of the failure rate is modeled by a Weibull with decreasing failure rate, and the latter part is modeled by an exponential (i.e., constant). It does not have a part describing wearout since the manual claims that the AT&T electronic equipment tends not to wear out before it is replaced. The AT&T model has been used extensively by Kuo and various co-authors (e.g., see Chien and Kuo, 1992) to study optimal burn-in for integrated circuit systems. This model is also called the Weibull–exponential model in the statistical literature (e.g., see Boukai, 1987).

Since mixtures are emphasized in this review we point out one apparent anomoly mentioned by Gurland and Sethuraman (1994). In that paper it is observed that when even strongly increasing failure rate distributions are mixed with certain other distributions, their failure rate tends to decrease after a certain point. This is not surprising in the light of the previously mentioned result of Block, Mi and Savits (1993), which gives that asymptotically the failure rate of a mixture tends to the asymptotic failure rate of the strongest component of the mixture. Since the failure rate of the strongest component is the smallest, the failure rate of the mixture is often eventually decreasing to this smallest value.

Most definitions of bathtub-shaped failure rates assume the failure rate decreases to some change point (t_1), then remains constant to a second change point (t_2), then increases. The case $t_1 = t_2$ (i.e., no constant portion) is often adequate as an assumption in some theoretical results. We give the definition below.

DEFINITION 1. A random lifetime X with distribution function $F(t)$, survival function $\bar{F}(t) = 1 - F(t)$, density $f(t)$ and failure rate $r(t) = f(t)/\bar{F}(t)$ is said to have a *bathtub-shaped failure rate* if there exist points $0 \le t_1 \le t_2 \le \infty$, called *change points*, such that

$$r(t) \text{ is } \begin{cases} \text{decreasing for } 0 \le t < t_1, \\ \text{constant for } t_1 \le t < t_2, \\ \text{increasing for } t_2 \le t < \infty. \end{cases}$$

We have restricted the above definition to continuous lifetimes, but discrete lifetimes can be handled similarly (see Mi, 1993, 1994c). Further we often shorten the phrase bathtub-shaped failure rate to *bathtub failure rate* or even *bathtub distribution*. A bathtub curve is called *degenerate* if either the decreasing or increasing part is not present (i.e., it is either always increasing or always decreasing).

3. OPTIMAL BURN-IN

In this section we consider some basic criteria for determining the optimal burn-in time for a lifetime. In general, we consider lifetimes with a bathtub-shaped failure rate having change points t_1 and t_2 (see Definition 1). As exemplified in Figure 1, burn-in often takes place at or before the first change point t_1. In fact, in the following, various optimality criteria lead to such a burn-in time. In Section 3.1 we focus on performance based criteria. The more realistic situation involving cost structures is considered in Section 3.2 and these are based in part on the criteria of Section 3.1.

3.1 Performance-Based Criteria

In this section we consider the problem of performance-based criteria in which the more general assumption of a cost structure is not made. Many of these criteria are basic concepts which can and should be incorporated into a general cost structure. Cost structures are considered in Section 3.2.

The paper of Watson and Wells (1961) was one of the first statistical papers to study the question of burn-in. These authors were interested in conditions under which the mean residual life (after burn-in) was larger than the original mean lifetime. Maximizing the mean residual life is one of the criteria we examine in this section. We now list several criteria for determining burn-in. Criteria C1, C2 and C4 deal with only one component. Criterion C3 deals with components which are replaced at failure with other identical components.

C1. Let τ be a fixed mission time and let $\bar{F}$ be the survival function of a lifetime. Find b which maximizes $\bar{F}(b + \tau)/\bar{F}(b)$, that is, find b such that, given survival to time b, the probability of completing the mission is as large as possible.

C2. Let X be a lifetime. Find the burn-in time b which maximizes $E[X - b|X > b]$, that is, find the burn-in time which gives the largest mean residual life.

C3. Let $\{N_b(t),\ t \geq 0\}$ be a renewal process of lifetimes which are burned in for b units of time (i.e., where F is the original lifetime distribution and the interarrival distribution has survival function $\bar{F}_b(t) = \bar{F}(b + t)/\bar{F}(b)$). For fixed mission time τ, find b which minimizes $E[N_b(\tau)]$, which is the mean number of burned-in components which fail during the mission time τ.

The next criterion involves the *α-percentile residual life function*. The α-percentile residual life is defined by

$$q_\alpha(b) = F_b^{-1}(\alpha) = \inf\{x \geq 0\colon \bar{F}_b(x) \leq 1 - \alpha\}$$

(see Joe and Proschan, 1984, for further details).

C4. For a fixed α, $0 < \alpha < 1$, find the burn-in time b which maximizes $\tau = q_\alpha(b)$, that is, find the burn-in time which gives the maximal warranty period τ for which at most α% of items will fail.

Criterion C2 has been studied by several authors. The first of these, Watson and Wells (1961), examined various parametric distributions. Lawrence (1966) obtained bounds on the mean residual life. Park (1985) gave some results on the mean residual life for a bathtub distribution. One result was that the optimal burn-in time b^* occurs before the first change point t_1. Mi (1994b) obtained the same result for criteria C1 and C3, that is, $b^* \leq t_1$. Launer (1993) introduced criterion C4 and also showed that the optimal b^* occurs before t_1. This type of result is important since it provides an upper bound for burn-in.

The fact that optimal burn-in for a bathtub distribution takes place before the first change point is not unusual. In fact, it is intuitive that burn-in should occur before this change point since this is where the failure rate of such a lifetime stops improving. We shall see in Section 3.2 that the result also holds true for many cost structures.

In another direction, Mi (1994b) compared optimal burn-in times for two mission times $\tau_1 \leq \tau_2$ for criterion C1. He showed the intuitive result that $b_2^* \leq b_1^*$. An extension to random mission times was also considered.

In criterion C3, a burned-in unit that failed during field use was replaced with another burned-in unit. If instead of replacing this unit, a minimal repair is performed (see Barlow and Proschan, 1965), then the total number of minimal repairs is a nonhomogeneous Poisson process with mean function $-\ln[\bar{F}(b + \tau)/\bar{F}(b)]$. Thus if we want to minimize the expected number of minimal repairs in the interval $[0, \tau]$, it suffices to maximize the quantity $\bar{F}(b + \tau)/\bar{F}(b)$. But this is just criterion C1.

3.2 Cost Functions and Burn-in

Several cost functions have been proposed to deal with burn-in. A discussion of many of these is given in the review papers of Kuo and Kuo (1983) and Leemis and Beneke (1990). Also see Nguyen and Murthy (1982). In this section we discuss a few of the recent models involving cost functions for burn-in. In all cases we are interested in finding the burn-in time which minimizes the cost. Cost functions CF1 and CF4 are used in subsequent sections. In general these cost functions build upon and elaborate the criteria of Section 3.1. Cost function CF1 is basic, while CF2 and CF4 incorporate C2; CF3 uses C1.

CF1. A component or system with lifetime X is burned-in for time b. If it fails to survive b units of time a cost c_0 is incurred. If it survives b units of time, then it incurs a second cost C, $C > c_0$, if it does not survive past an additional mission time τ or it incurs a gain of K if it does survive τ. Consequently, if F is the distribution function of the component or system the expected cost as a function of b is

$$c_1(b) = c_0 F(b) + C[F(b+\tau) - F(b)] - K\bar{F}(b+\tau).$$

CF2. If instead of a mission time after the burn-in we consider a gain proportional to the mean residual life (with proportionality constant K), the expected cost becomes

$$c_2(b) = c_0 F(b) - K\,\frac{\int_b^\infty \bar{F}(t)\,dt}{\bar{F}(b)}.$$

The next criteria involve costs for in-shop repair. If a device fails burn-in, it is scrapped at a cost $c_s > 0$ and another unit is burned-in. This process is continued until a unit survives burn-in time b. A device which survives burn-in is then put into field use. The cost for burn-in is assumed to be proportional to the time it takes to obtain a unit which survives burn-in with proportionality constant c_0. Mi (1994a) derives the expression for the expected cost as

$$k(b) = c_0\,\frac{\int_0^b \bar{F}(t)\,dt}{\bar{F}(b)} + \frac{c_s F(b)}{\bar{F}(b)}.$$

The complete cost also includes additive field costs, and this is reflected in the following cost functions.

CF3. In this case, after a burned-in item is obtained, a cost of C is incurred if the burned-in device does not survive the mission time τ and a gain of K if it survives the mission. Thus the total cost function is given by

$$c_3(b) = k(b) + C\,\frac{F(b+\tau) - F(b)}{\bar{F}(b)} - K\,\frac{\bar{F}(b+\tau)}{\bar{F}(b)},$$

where $k(b)$ is as above.

CF4. If instead of a mission time, a gain is taken proportional to the mean residual time, the cost function in CF3 is modified to

$$c_4(b) = k(b)\ - K\,\frac{\int_b^\infty \bar{F}(t)\,dt}{\bar{F}(b)}.$$

The cost function CF1 was introduced by Clarotti and Spizzichino (1990). These authors obtained conditions for an optimal burn-in time b^* and applied their results to a mixed exponential model. See also Section 4.2, where an extension of the mixed exponential model to a general mixture model is discussed. The cost function CF2 is a variant of CF1. The cost functions CF3 and CF4 are discussed in Mi (1991, 1995). As in Section 3.1, the respective authors show that the optimal burn-in time b^* satisfies $b^* \le t_1$ for cost functions CF2–CF4, where t_1 is the first change point for the assumed bathtub distribution.

4. MIXTURE MODELS

In this section we consider recent mixture models. This is the typical model described in Section 2 to which burn-in is applicable. In both Arjas, Hansen and Thyregod (1991) and Block, Mi and Savits (1993) an underlying mixture distribution is used to model the life of components. The latter paper discusses burn-in applications although the former paper does not.

The paper of Arjas, Hansen and Thyregod (1991) discussed in Section 4.1 is an interesting mix of modeling and estimation and uses ideas and techniques from the reliability theory, life testing (engineering reliability) and survival analysis literature. The methods developed are applied to an example involving printed circuit boards. In Section 4.2 we discuss results of Block, Mi and Savits (1993). A more general mixture model than in Arjas, Hansen and Tyregod (1991) is examined. A recent paper of Spizzichino (1995) discusses another model for mixtures in heterogeneous populations.

4.1 A Reliability Growth Model

Arjas, Hansen and Thyregod (1991) consider a renewal process approach to reliability growth where heterogeneity of the underlying part structure is shown to translate into renewal intensity behavior. Although burn-in per se is not discussed in this paper, the lifetimes discussed are of the type to which burn-in is typically applied. This section also provides a background for Section 4.2, which considers mixed lifetimes.

The basic process involves the lifetimes of parts placed in two or more sockets where, upon failure, a failed part is replaced by a new part of the same type. The first and subsequent lifetimes for one socket are designated by $X_1, X_2, \ldots$. These lifetimes are assumed independent. The lifetimes are also assumed to come from a heterogeneous population. It is natural to model these lifetimes using a random hazard rate so that the distribution of the lifetime can be written as a mixed exponential, that is,

$$P\{X_k > x) = \int_0^\infty e^{-\lambda x}\,d\phi(\lambda),$$

where ϕ is the distribution of the random hazards. The aim of the paper is to study the renewal process of one socket or the superimposed renewal process of several. As is well known, this mixed distribution has a decreasing failure rate.

If $N(t)$ is the renewal process for one socket, it can be shown that $V(t) = EN(t)$ is concave and the rate of occurrence of failures for the renewal process, $v(t) = (d/dt)\ V(t)$, is decreasing. Various results for this type of renewal process can be obtained, and comparisons can be made with processes where sockets are minimally repaired rather than replaced. For minimal repair, the associated process is the nonhomogeneous Poisson process. (See Block and Savits, 1995, for many comparisons of this type.)

Parametric estimation is considered by these authors for the bimodal (i.e., mixture of two) exponential case. The bimodal Weibulls (and exponentials) are the principal examples of the Jensen and Petersen (1982) monograph on burn-in. The distribution for the life length of the part is the three-parameter mixture of two exponential distributions with distribution function

$$F(x) = \pi\,(1 - \exp(-\lambda_0 x)) + (1-\pi)\,(1 - \exp(-\lambda_1 x)), \quad x > 0.$$

It is assumed that inferior parts cannot be distinguished from a standard part. Two cases are considered: (a) the case where sockets are observed individually and (b) the case where sockets are only

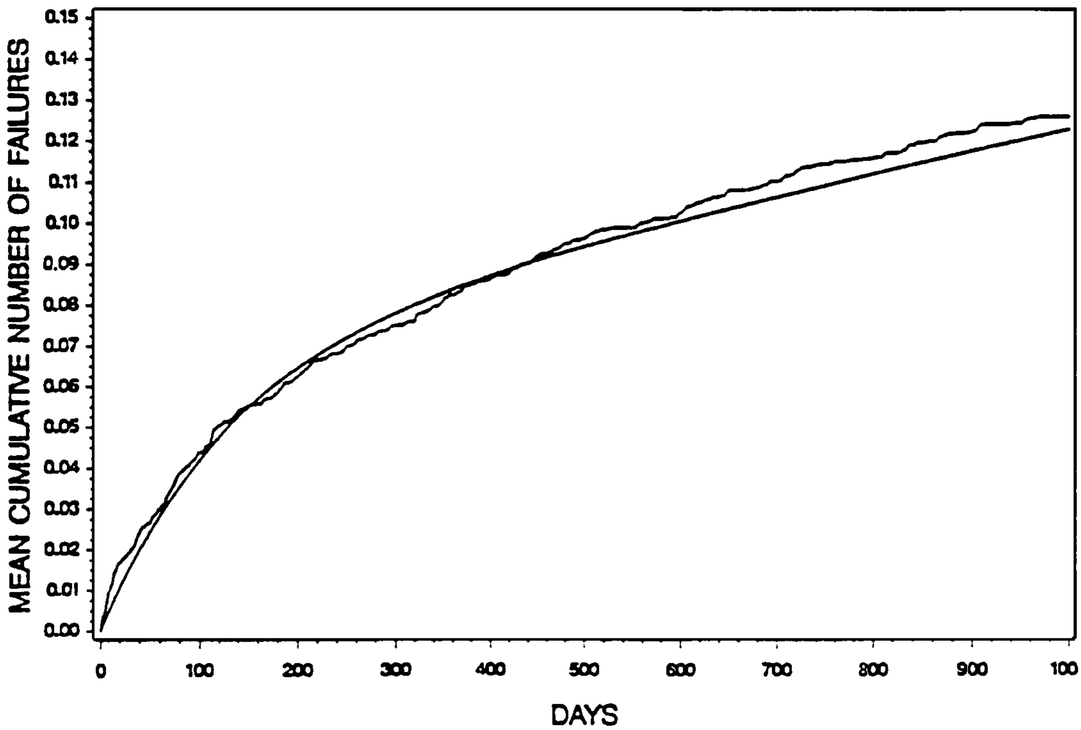

FIG. 2. *Comparing two estimates; the step curve comes from* $\widehat{V}_{N-A}(t)$ *and the smooth curve comes from* $\bar{V}(t)$.

observed as aggregated data. In case (a), the maximum likelihood estimation is straightforward. Right censoring is permitted and the likelihood or log-likelihood function is standard. In case (b), times between failures are not independent and so either (1) an approximation by a corresponding nonhomogeneous Poisson process is used or (2) it is assumed, in the case when the number of failures is less than the number of sockets, that each socket has experienced at most one failure and so the techniques of (a) apply.

An example is given where the system is a printed circuit board consisting of 560 parts (sockets) and there are 3,481 systems from which data was collected for five years. Maximum likelihood estimates were obtained computationally for the three parameters and were used to estimate the cumulative number of occurrence of failures $\bar{V}(t) = E\bar{N}(t)$, where $\bar{N}(t)$ is the superimposed renewal process. The model can be assessed graphically by calculating $\bar{N}^0(t)$, the counting process obtained as the sum of the individual system processes, and then using the Nelson–Aalen estimate

$$\widehat{V}_{N-A}(t) = \sum_{s \le t} \frac{\Delta \bar{N}^0(s)}{R(s)},$$

where $0 < T_1^0 < T_2^0 < \cdots < T^0_{\bar{N}^0(t)} < t$ are all of the failure times, $\Delta \bar{N}^0(s)$ is 1 for each T_i^0 and $R(s)$ denotes the number of active systems older than s. This yields Figure 2, which compares these two estimates. The step curve comes from $\widehat{V}_{N-A}(t)$ and the smooth curve comes from $\bar{V}(t)$. Confidence bounds are also obtained in this paper using several methods.

4.2 A General Mixture Model

As mentioned in Section 1, one explanation for a bathtub-shaped failure rate that is often given by engineers is that it is due to mixtures of populations, some weak and some strong. In Block, Mi and Savits (1993), a general mixture model was investigated. A goal of that paper was to determine optimal burn-in for the cost function CF1 of Clarotti and Spizzichino (1990). Some results of independent interest, however, were also obtained. They are summarized below.

For the general mixture model, it is assumed that each member of the subpopulation, indexed by $\lambda \in S$, has a positive density $f(t, \lambda)$ on $(0, \infty)$. The density of the resulting mixed population is then given by

$$f(t) = \int_S f(t, \lambda) P(d\lambda), \qquad (4.1)$$

where P is the mixing distribution.

The first results concern the monotonicity of the ratio $g(t) = f(t + \tau)/f(t)$ for a fixed mission time

$\tau > 0$. This is a new type of aging property that seems appropriate for burn-in since it is related to a notion of beneficial aging. More specifically, if we require the ratio $f(t+\tau)/f(t)$ to be increasing in $t > 0$ for *each* $\tau > 0$, then f must be log-convex and hence belongs to the class of distributions which have a decreasing failure rate. Furthermore, certain bathtub failure rates which can be realized as mixtures have this monotonicity property.

Before we can state this result, we recall the definition of *reverse regular of order* 2 (RR_2). A nonnegative function $k(x, y)$ on $A \times B$ is said to be RR_2 if

$$k(x_1, y_1)k(x_2, y_2) \leq k(x_1, y_2)k(x_2, y_1)$$

whenever $x_1 < x_2$ in A and $y_1 < y_2$ in B. Alternatively, we require that the ratio

$$\frac{k(x+\Delta, y)}{k(x, y)}$$

be decreasing in $y \in B$ for each $x \in A$ and $\Delta > 0$.

The following is a preservation result for a monotonicity property with a fixed mission time τ. Let the family of positive densities $\{f(t, \lambda)\colon \lambda \in S\}$ be RR_2 on $(0, \infty) \times S$ and let $\tau > 0$ be a fixed mission time. Suppose the ratio

$$g(t, \lambda) = \frac{f(t+\tau, \lambda)}{f(t, \lambda)}$$

is increasing in $t > 0$ for each $\lambda \in S$. Then, for the mixture density f given in (4.1), the ratio

$$g(t) = \frac{f(t+\tau)}{f(t)} = \frac{\int_S f(t+\tau, \lambda)\ P(d\lambda)}{\int_S f(t, \lambda)\ P(d\lambda)}$$

is increasing in $t > 0$. A more general result that does not require the RR_2 condition is given in Block, Mi and Savits (1993, Theorem 3.1).

A second result of interest in the paper of Block, Mi and Savits (1993) pertains to the limiting behavior of the failure rate for the mixed population. Heuristically, it states that the failure rate of the mixture tends to the strongest subpopulation. Under certain technical conditions it is shown that the failure rate of the mixed population converges to a constant α as $t \to \infty$. Here $\alpha = \inf\{a(\lambda)\colon \lambda \in S\}$ and $a(\lambda) = \lim_{t\to\infty} r(t, \lambda)$ with $r(t, \lambda)$ the failure rate of the λ-subpopulation. (The discrete version is considered in Mi, 1994c.)

Clarotti and Spizzichino (1990) also show for the mixture of exponentials model that if one mixing distribution P_1 is less than P_2 in the sense of likelihood ratio ordering, then the optimal burn-in times b_i^* for the cost function CF1 are ordered as $b_1^* \leq b_2^*$. The same result also holds for the general mixture model. See Block, Mi and Savits (1993) for details.

5. COMPONENT VERSUS SYSTEM BURN-IN

In this section we deal with the important issue of at which stage burn-in is most effective. Consider a system composed of individual components. Is it better to burn in all the components or is it better to assemble the components and burn in the system? If there are modules and subassembly systems similar questions can be asked. The component level is usually the least expensive stage at which to consider burn-in. Assembly of even burned-in components usually introduces defects, so burn-in at higher levels would seem to have some value. In this section we consider some preliminary work in which this question is considered, but under the simplifying assumption that no defects are introduced upon assembly. By a system here we mean a coherent system in the sense of Barlow and Proschan (1981, page 6).

There are three possible actions we want to consider which constitute different methods for burning-in the system:

(i) Burn in component i for a time β_i, $i = 1, \ldots, n$, and then assemble the system with the burned-in components.
(ii) Burn in component i for a time β_i, $i = 1, \ldots, n$, assemble the system with the burned-in components and then perform an additional burn-in of the system for a time b.
(iii) Assemble the system with new components and then burn in the system for a time period b.

Since (i) and (iii) are special subcases of (ii), we can do no better than (ii). However, is it possible that we can do just as well with one of the other two actions?

In Block, Mi and Savits (1994, 1995), this question was considered for three different criteria: (1) maximizing the probability that the system will survive a fixed mission time (or warranty period) τ; (2) maximizing the system mean residual life; and (3) maximizing the α-percentile (system) residual life $\tau = q_\alpha(b)$ for a fixed α, $0 < \alpha < 1$. In each case it was shown that one can do as well with burn-in at the component level only.

This result can be extended to criteria which have a type of monotonicity property. More specifically, the result can be shown to hold for any criterion determined by a functional ϕ defined on the class of life distributions which is monotone in stochastic order, that is, in the case of maximizing (minimizing) the objective function ϕ, we require that $\phi(F) \leq (\geq)\ \phi(G)$ whenever $F \leq^{st} G$ [i.e., $\bar{F}(t) \leq \bar{G}(t)$ for all $t \geq 0$]. Thus, for such criterion, burn-in at the system level is precluded by effective burn-in at the component level.

It should be noted that this result does not apply to the cost function criteria considered in Section 3.2 since they are not monotone in stochastic order. Also, if the act of assembling the components degrades the system, an additional burn-in at the system level might be required. Whitbeck and Leemis (1989) have considered a model for dealing with this problem.

6. TOTAL TIME ON TEST (TTT)

In this section we describe a primarily graphical technique which has been useful in burn-in. One consequence of this technique is in obtaining approximate burn-in times.

In a life test, failure times are observed until all or some portion of the items fail. A way to summarize the behavior is through the *total time on test* (TTT) *statistics*. Let $0 = x_{(0)} < x_{(1)} < x_{(2)} < \cdots < x_{(n)}$ be an ordered sample from a continuous lifetime distribution with finite mean. In this section, to avoid technical problems, we assume the distribution function F is strictly increasing on $(0, \infty)$. The TTT statistics are defined by

$$T_i = \sum_{j=1}^{i} (n - j + 1)(x_{(j)} - x_{(j-1)})$$

and

$$u_i = \frac{T_i}{n} \quad \text{for } i = 1, \dots, n.$$

Notice that $u_n = \bar{x}_n$. Moreover, if F_n is the empirical distribution function and $F_n^{-1}(x) = \inf\{t \mid F_n(t) \geq x\}$, then $F_n^{-1}(i/n) = x_{(i)}$ for $i = 1, \dots, n$. Consequently,

$$\int_0^{F_n^{-1}(i/n)} \bar{F}_n(x)\, dx = u_i, \quad i = 1, \dots, n.$$

This suggests a distributional analog called the *TTT transform*, traditionally denoted by H_F^{-1}. It is defined by

$$H_F^{-1}(t) = \int_0^{F^{-1}(t)} \bar{F}(u)\, du.$$

The *scaled TTT transform* is given by

$$\phi_F(t) = \frac{H_F^{-1}(t)}{H_F^{-1}(1)} = \frac{H_F^{-1}(t)}{\mu},$$

where μ is the mean of F. Although these concepts were discussed earlier, one of the first systematic expositions was given in Barlow and Campo (1975).

One of the principle uses of the TTT concept has been in obtaining approximate optimal solutions for age replacement and also in obtaining approximate optimal burn-in times. We briefly describe the procedure for burn-in and note that the procedure for age replacement is similar.

We consider the cost function CF4 of Section 3 as an example and describe how an optimal burn-in time b^* can be obtained using the TTT transform. This example is taken from Mi (1991). The cost function can be written as

$$c_4(b) = -c_s + \frac{c_0 \int_0^b \bar{F}(t)\, dt + c_s - K \int_b^\infty \bar{F}(t)\, dt}{\bar{F}(b)}.$$

The optimal burn-in is obtained by minimizing this function. Letting $u = F(b)$, minimizing the above is equivalent to maximizing

$$M_F(u) = \frac{\alpha - \phi_F(u)}{1 - u},$$

where $\alpha = (-c_s + K\mu)/(c_0 + K)\mu$ and μ is the mean of F. The function $M_F(u)$ is the slope of the line segment connecting the points $(1, \alpha)$ and $(u, \phi_F(u))$. Consequently we need only find the point on the graph of ϕ_F, the scaled TTT transform, for which the above slope is largest.

If n items with lifetime X are put on test, a TTT plot (i.e., the graph of ϕ_{F_n}) can be obtained. Since the TTT transform is the asymptotic version of the TTT plot, an estimate of the optimal burn-in can be obtained. If the point $(i/n, T_i/T_n)$ maximizes $M_{F_n}(u)$, then $x_{(i)} = F_n^{-1}(i/n)$ is the ordered value giving an estimate of the optimal burn-in. We illustrate this in Figure 3. Other similar burn-in applications can be found in Bergman and Klefsjo (1985). See also the review article by Bergman (1985), which gives other applications of the TTT transform.

7. SEQUENTIAL BURN-IN AND OPTIMAL CONTROL

A theory of sequential burn-in has been proposed in Spizzichino (1991), and some extensions of this have been initiated by the same author and some of his colleagues. This extends the previous material which deals with mainly one component or system, or components which are independent and identically distributed. The more general situation where the components are not assumed independent is treated by Spizzichino and colleagues, who assume components are exchangeable. A mixture model for strong and weak exchangeable components has been proposed by Spizzichino (1995). We give a brief introduction to this work. Several representative papers are contained in Barlow, Clarotti and Spizzichino (1993).

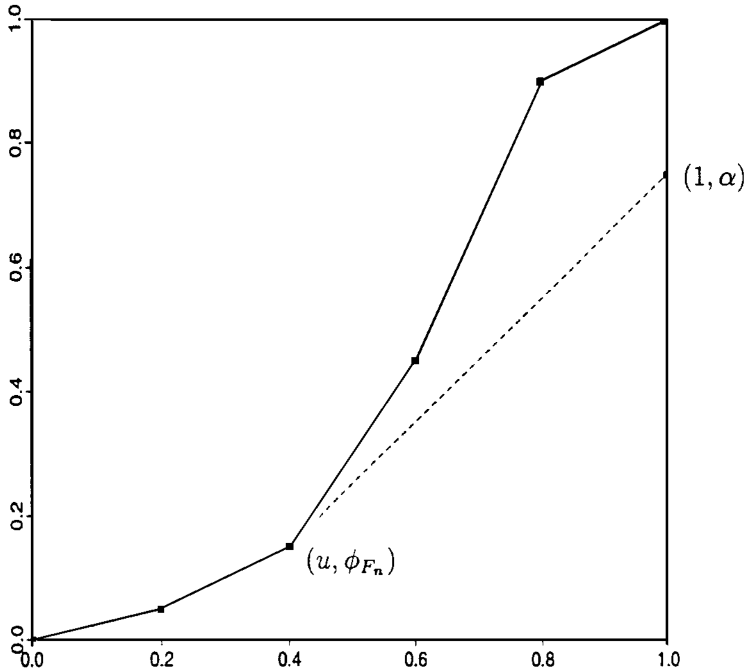

FIG. 3. *Estimate of the optimal burn-in.*

As background we mention a paper of Marcus and Blumenthal (1974), who considered a sequential burn-in procedure. The stopping rule they suggested is as follows: observe failure times and stop when the time between failures exceeds a fixed value. This is reasonable for a lifetime which has a high initial failure rate that becomes smaller. Properties of this rule are studied and tables for its use are given.

In Spizzichino (1991), failure times of n components which are assumed to be exchangeable are observed. One burn-in time is chosen initially and if all components survive it, they are put into field operation. If there is a failure before this time, a new additional burn-in time is chosen (which may depend on the first failure) and the procedure repeats. A cost structure based on the one in Clarotti and Spizzichino (1990), that is, CF1 from Section 3, is given. A sequential burn-in strategy is defined and this is shown to be optimal. A particular case is mentioned where the exchangeable distribution is a mixture of exponentials. This case is further explored in Costantini and Spizzichino (1991), where a strategy is proposed for reducing this to an optimal stopping problem for a two-dimensional Markov process. Further details are given in Costantini and Spizzichino (1990) and in Caramellino and Spizzichino (1996).

A related approach for optimal screening (a type of burn-in) is given in Iovino and Spizzichino (1993). A general unifying model is proposed by Spizzichino (1993). Some very recent research on optimal burn-in of software is given in Barlow, Clarotti and Spizzichino (1994).

8. DISCUSSION AND AREAS FOR DEVELOPMENT

In this review of recent developments in burn-in we have discussed a variety of problems. We recapitulate some of these ideas in this section along with some future research directions.

A basic assumption on a lifetime for which burn-in is appropriate is that it has a bathtub-shaped failure rate. This type of lifetime often arises because a population consists of a mixture of weak and strong subpopulations. One question for which a satisfactory answer has not been determined is for which mixtures does the failure rate have a bathtub shape.

As described in Section 3, the intuitive result that burn-in should occur before the first change point of a bathtub failure rate has been demonstrated for a wide variety of criteria and cost functions, but in an ad-hoc way. The authors are currently working on a unified result for an even broader class of objective functions.

The handbook of Jensen and Petersen (1982) presents a wide array of graphical and heuristic statistical techniques for burn-in. Many of these are applied to mixtures which model weak and strong components. At the time the book was written, statistical techniques and procedures for mixtures were less well understood than they are at the present time. It would be useful if many of the intuitively plausible and useful techniques given in this handbook were updated and put on a firmer statistical foundation. One example of this is the paper of Arjas, Hansen and Tyregod (1991) (see Section 4.1), who develop estimation techniques for renewal processes where the underlying distribution is a mixture of exponentials.

The material of Section 7 on sequential burn-in and optimal control appears to be a fruitful area of research. It seems evident that this direction should be expanded and further investigated.

The development of new ideas on burn-in goes hand-in-hand with developments in accelerated life testing. In fact, burn-in is most often accomplished in an accelerated environment. A related topic is degradation, in which instead of the lifetime of a component the emphasis is on a measure of the quality of the component as it wears out. If the environment is accelerated, the question of burn-in in conjunction with this accelerated degradation becomes of interest. For recent developments on accelerated degradation, see Nelson (1990) and Meeker and Escobar (1993).

An area of reliability where burn-in techniques might be applicable and vice-versa is the topic of software reliability modeling. In this area one problem involves removing errors (bugs) from the software. An assumption which is made is that when bugs are detected and removed no new bugs are introduced. In this case the software is improved since the number of bugs remaining is decreased. Consequently, the rate at which bugs are discovered is decreased. This rate is analogous to the left tail of a failure rate with infant mortality present. Since the time at which the testing should stop is of interest, and this is analogous to the burn-in period for a lifetime of the type discussed in this paper, there should be some transfer between the ideas of both of these fields. To date there have been a few applications of burn-in ideas to finding the time at which to stop testing the software. The paper of Barlow, Clarotti and Spizzichino has been mentioned. See also Section 6 of Singpurwalla and Wilson (1994), who review the optimal testing of software.

ACKNOWLEDGMENTS

We would like to thank Leon Gleser, Rob Kass and Jie Mi for helpful comments as well as a referee for a careful reading and perceptive comments. David Coit supplied us with the Fuqua reference and some recent papers and we thank him. Discussions with Allan Sampson, Leon Gleser and David Coit were also helpful.

The work of the authors was partially supported by NSA Grant MDA-904-90-H-4036 and by NSF Grant DMS-92-03444.

REFERENCES

Arjas, E., Hansen, C. K. and Thyregod, P. (1991). Estimation of mean cumulative number of failures in a repairable system with mixed exponential component lifetimes. *Technometrics* **33** 1–12.

Barlow, R. E. (1985). A Bayes explanation of an apparent failure rate paradox. *IEEE Transactions on Reliability* **R-34** 107–108.

Barlow, R. E. and Campo, R. (1975). Total time on test processes and applications to failure data analysis. In *Reliability and Fault Tree Analysis* (R. E. Barlow, J. B. Fussell and N. D. Singpurwalla, eds.) 451–481. SIAM, Philadelphia.

Barlow, R. E., Clarotti, C. A. and Spizzichino, F., eds. (1993). *Reliability and Decision Making*. Chapman and Hall, New York.

Barlow, R. E., Clarotti, C. A. and Spizzichino, F. (1994). Optimal burn-in of software. Unpublished report.

Barlow, R. E. and Proschan, F. (1965). *Mathematical Theory of Reliability*. Wiley, New York.

Barlow, R. E. and Proschan, F. (1981). *Statistical Theory of Reliability and Life Testing*. To Begin With, Silver Spring, MD.

Bergman, B. (1985). On reliability theory and its applications. *Scand. J. Statist.* **12** 1–41.

Bergman, B. and Klefsjo, B. (1985). Burn-in models and TTT transforms. *Quality and Reliability International* **1** 125–130.

Block, H. W., Mi, J. and Savits, T. H. (1993). Burn-in and mixed populations. *J. Appl. Probab.* **30** 692–702.

Block, H. W., Mi, J. and Savits, T. H. (1994). Some results on burn-in. *Statist. Sinica* **4** 525–534.

Block, H. W., Mi, J. and Savits, T. H. (1995). Burn-in at the component and system level. In *Lifetime Data: Models in Reliability and Survival Analysis* (N. P. Jewell, A. C. Kimber, M.-L. T. Lee and G. A. Whitmore, eds.) 53–58. Kluwer, Dordrecht.

Block, H. W. and Savits, T. H. (1995). Comparisons of maintenance policies. In *Stochastic Orders and Their Applications* (M. Shaked and J. G. Shanthikumar, eds.) Chap. 15. Academic Press, New York.

Boukai, B. (1987). Bayes sequential procedure for estimation and for determination of burn-in time in a hazard rate model with an unknown change-point parameter. *Sequential Anal.* **6** 37–53.

Caramellino, L. and Spizzichino, F. (1996). WBF properties and statistical monotonicity of the Markov process associated to Schur-constant survival functions. *J. Multivariate Anal.* **56** 153–163.

Chien, W.-T. K. and Kuo, W. (1992). Optimal burn-in simulation on highly integrated circuit systems. *IIE Transactions* **24**(5) 33–43.

Clarotti, C. A. and Spizzichino, F. (1990). Bayes burn-in decision procedures. *Probability in the Engineering and Informational Sciences* **4** 437–445.

Costantini, C. and Spizzichino, F. (1990). Optimal stopping of the burn-in of conditionally exponential components. Unpublished report.

Costantini, C. and Spizzichino, F. (1991). Optimal stopping of life testing: use of stochastic orderings in the case of conditionally exponential lifetimes. In *Stochastic Orders and Decision Making Under Risk* (K. Mosler and M. Scarsini, eds.) 95–103. IMS, Hayward, CA.

FUQUA, N. B. (1987). *Reliability Engineering for Electronic Design*. Dekker, New York.

GURLAND, J. and SETHURAMAN, J. (1994). Reversal of increasing failure rates when pooling failure data. *Technometrics* **36** 416–418.

IOVINO, M. G. and SPIZZICHINO, F. (1993). A probabilistic approach for an optimal screening problem. *J. Ital. Statist. Soc.* **2** 309–335.

JENSEN, F. and PETERSEN, N. E. (1982). *Burn-in*. Wiley, New York.

JOE, H. and PROSCHAN, F. (1984). Percentile residual life functions. *Oper. Res.* **32** 668–678.

KLINGER, D. J., NAKADA, Y. and MENENDEZ, M. A., eds. (1990). *AT&T Reliability Manual*. Van Nostrand Rheinhold, New York.

KUO, W. and KUO, Y. (1983). Facing the headaches of early failures: a state-of-the-art review of burn-in decisions. *Proceedings of the IEEE* **71** 1257–1266.

LAUNER, R. L. (1993). Graphical techniques for analyzing failure data with percentile residual-life functions. *IEEE Transactions on Reliability* **R-42** 71–75.

LAWRENCE, M. J. (1966). An investigation of the burn-in problem. *Technometrics* **8** 61–71.

LEEMIS, L. M. and BENEKE, M. (1990). Burn-in models and methods: a review. *IIE Transactions* **22** 172–180.

MARCUS, R. and BLUMENTHAL, S. (1974). A sequential screening procedure. *Technometrics* **16** 229–234.

MEEKER, W. Q. and ESCOBAR, L. A. (1993). A review of recent research and current issues in accelerated testing. *Internat. Statist. Rev.* **61** 147–168.

MI, J. (1991). Burn-in. Ph.D. dissertation, Dept. Mathematics and Statistics, Univ. Pittsburgh.

MI, J. (1993). Discrete bathtub failure rate and upside down failure rate. *Naval Research Logistics* **40** 361–371.

MI, J. (1994a). Burn-in and maintenance policies. *Adv. in Appl. Probab.* **26** 207–221.

MI, J. (1994b). Maximization of a survival probability and its applications. *J. Appl. Probab.* **31** 1026–1033.

MI, J. (1994c). Limiting behavior of mixtures of discrete lifetime distributions. Technical report, Dept. Statistics, Florida International Univ.

MI, J. (1995). Minimizing some cost functions related to both burn-in and field use. *Oper. Res.* To appear.

NELSON, W. (1990). *Accelerated Testing*. Wiley, New York.

NGUYEN, D. G. and MURTHY, D. N. P. (1982). Optimal burn-in time to minimize cost for products sold under warranty. *IIE Transactions* **14** 167–174.

PARK, K. S. (1985). Effect of burn-in on mean residual life. *IEEE Transactions on Reliability* **R-34** 522–523.

RAJARSHI, S. and RAJARSHI, M. B. (1988). Bathtub distributions: a review. *Comm. Statist.* **17** 2597–2621.

RAWICZ, A. H. (1986). Burn-in of incandescent sign lamps. *IEEE Transactions on Reliability* **R-35** 375–376.

SINGPURWALLA, N. and WILSON, S. (1994). Software reliability modeling. *Internat. Statist. Rev.* **62** 289–317.

SPIZZICHINO, F. (1991). Sequential burn-in procedures. *J. Statist. Plann. Inference* **29** 187–197.

SPIZZICHINO, F. (1993). A unifying model for optimal design of life testing and burn-in. In *Reliability and Decision Making* (R. E. Barlow, C. A. Clarotti and F. Spizzichino, eds.) 189–210. Chapman and Hall, New York.

SPIZZICHINO, F. (1995). A probabilistic model for heterogeneous populations and related burn-in design problems. Scientific note, Dept. Mathematics, Univ. Rome "Guido Castelnuovo."

TOBIAS, P. A. and TRINDADE, D. C. (1995). *Applied Reliability*. Van Nostrand, New York.

VAUPEL, J. W. and YASHIN, A. I. (1985). Heterogeneity's ruse: some surprising effects of selection on population dynamics. *Amer. Statist.* **39** 176–185.

WATSON, G. S. and WELLS, W. T. (1961). On the possibility of improving mean useful life of items by eliminating those with short lives. *Technometrics* **3** 281–298.

WHITBECK, C. W. and LEEMIS, L. M. (1989). Component vs. system burn-in techniques for electronic equipment. *IEEE Transactions on Reliability* **38** 206–209.

Comment: "Burn-In" Makes Us Feel Good

Nicholas J. Lynn and Nozer D. Singpurwalla

1. PREAMBLE

Block and Savits, henceforth BS, have made many contributions to the mathematics of burn-in and are eminently qualified to put together a review article on this topic. Indeed, what they provide here is an authoritative survey of the technical aspects of the subject. All those who work in reliability should thank them for this and their other writings in this arena. Our intent here is not to challenge BS on the mathematics of burn-in, which undoubtedly is their territory. Rather, we take exception to their interpretation and their view of burn-in. Our main concern is that BS view burn-in as a mathematical rather than as an engineering problem. The authors are not to be faulted for this because their perspective of burn-in is, regrettably, guided by engineers who do reliability rather than by engineers who do engineering! Consequently, this survey does a good job of reporting that which is known and

Nozer D. Singpurwalla is Professor, Department of Operations Research and Department of Statistics, George Washington University, Washington, DC 20052-0001 (e-mail: nozer@gwuvm.gwu.edu). Nicholas Lynn works with Professor Singpurwalla.

written on the topic; unfortunately, that which is known is subject to debate. The result is that BS have adopted a limited view of burn-in and have refrained from a discussion of its foundational issues. Our commentary—actually an article—is written with the hope of filling these gaps and providing an alternative perspective on burn-in; in the sequel we provide some new results on mixtures of distributions that are germane to burn-in.

"Burn-in" is commonly used in engineering reliability, statistical simulation and medical sensitivity testing. In this article we discuss the philosophical underpinnings of burn-in, and make three claims. Our first claim is that the main purpose served by burn-in is psychological, that is, relating to belief. Our second claim is that burn-in is dictated by the interaction between predictive failure rates and utilities. Consequently, burn-in may be performed even if the predictive failure rate is increasing and the utility of the time on test decreasing. An example is the burn-in phase of statistical simulation, which mirrors burn-in testing of engineering components. Our third claim is that the famous "bathtub" curve of reliability and biometry rarely has a physical reality. Rather, as shown in Theorem 2, it is the manifestation of one's uncertainty.

2. INTRODUCTION AND OVERVIEW

2.1 Background

What is "burn-in"? The answer depends on whom you ask: an engineer, a simulator or a survivor (biostatistician). Each explains burn-in differently. Our goal is to argue, using a minimum of mathematics, that there is a unifying theme underlying burn-in and, therefore, there must be a single answer to the question that is posed.

First, let us see how engineers view burn-in. To an engineer, burn-in is a procedure for eliminating "weak" items from a population (cf. Block, Mi and Savits, 1993). The population is assumed to consist of two homogenous subpopulations: "weak" and "strong." Burn-in is achieved by testing each item in the population for the *burn-in period*, and commissioning to service those items that survive the test. The items that fail the test are judged weak.

To a simulator, burn-in is the time phase during which an algorithm, such as a "Gibbs sampler," attains its theoretical convergence (usually the weak convergence of distributions); see, for example, Besag, Green, Higdon and Mengersen (1995). Biostatisticians do not use the term burn-in, but the notion of "sensitivity testing" a new drug for a short period of time parallels the thinking of engineers.

2.2 Misconceptions about Burn-in

There appear to be at least two misconceptions about the engineer's view of burn-in. The first is that items that are judged to have exponential life distributions (or distributions that have an increasing failure rate) should *not* be subjected to a burn-in (cf. Clarotti and Spizzichino, 1990). The second misconception is that the *sole* purpose of burn-in is the elimination of weak items from a population.

The causes of the first misconception are a failure to appreciate the role of utility in burn-in and a failure to distinguish between what Barlow (1985) refers to as the "model failure rate" and the "predictive failure rate." Burn-in decisions should be based on the predictive failure rate, not the model failure rate. In fact, if the predictive life distribution is a mixture of exponential distributions, then burn-in must be contemplated; it should be performed if the costs of testing compensate for the avoidance of risk of in-service failures.

The cause of the second misconception is a failure to appreciate the fact that, fundamentally, there are two reasons for performing a burn-in test: *psychological* (i.e., those pertaining to belief) and *physical* (i.e., those pertaining to a change in the physical or the chemical composition of an item).

2.3 Objectives

The aim of this article is to argue that the two reasons given above cover the entire spectrum of burn-in, be it in engineering reliability, in simulation or in sensitivity testing. Also, fundamentally, since the concepts of physics, chemistry and biology influence belief (or psychology), there is only one reason for burn-in, namely, psychological. In what follows (see Section 4), we will attempt to justify our point of view. We will also point out that in one's day-to-day life, the psychology of burn-in is routinely practiced. In Section 5 we give examples of circumstances which provide a physical motivation for burn-in. Section 6 explores the role of utility in burn-in, and Section 7, entitled "An anatomy of failure rates with decreasing segments," leads us to a discussion of optimal burn-in times. A consequence of the material of Section 7 is our claim (Theorem 2) that the famous "bathtub curve" of reliability is rarely a physical reality; rather, it is often the manifestation of one's subjective belief. This may come as a surprise to many.

3. NOTATION AND TERMINOLOGY

Suppose that T, the time to failure of an item, has a distribution function $F(t) = P(T \leq t)$ and a survival function $\overline{F}(t) = 1 - F(t)$. Assume that

$f(\cdot)$, the probability density function of $F(\cdot)$, exists. If $F(\cdot)$ is indexed by a parameter $\boldsymbol{\theta}$ so that $P(T \le t|\boldsymbol{\theta}) = F(t|\boldsymbol{\theta})$, then $h(t|\boldsymbol{\theta})$, the *model failure rate function* of $F(\cdot)$, is defined as

$$h(t|\boldsymbol{\theta}) = \frac{f(t|\boldsymbol{\theta})}{\overline{F}(t|\boldsymbol{\theta})}, \quad t \ge 0. \tag{1}$$

The function F is said to have an *increasing* (*decreasing*) *model failure rate* if $h(t|\boldsymbol{\theta})$ is monotonically increasing (decreasing) in t, where we use increasing (decreasing) in place of nondecreasing (nonincreasing) throughout. The function F is said to have a constant model failure rate if $h(t|\boldsymbol{\theta})$ is constant in t. It is well known that $h(t|\boldsymbol{\theta})$ is constant in t if and only if $\overline{F}(t|\boldsymbol{\theta}) = e^{-\theta t}$, an exponential distribution.

In keeping with our claim to use a minimum of mathematics, we will call $h(t|\theta)$ a bathtub curve if it satisfies the following definition.

A function $g(t)$ is said to be a *bathtub curve* if there exists a point $u > 0$ such that $g(t)$ is strictly decreasing for $t < u$ and strictly increasing for $t > u$.

In the context of burn-in, which is de facto a limited life test, F having an increasing (decreasing) model failure rate implies that burn-in results in a depletion (enhancement) of useful life. When F has a constant model failure rate, burn-in results in neither a depletion nor an enhancement of useful life.

Since the parameter $\boldsymbol{\theta}$ is always unknown, we need to specify a distribution for it; let the density of this distribution be denoted $\pi(\boldsymbol{\theta})$. Then $h(t)$, the *predictive failure rate function* of F, is given as

$$\begin{aligned} h(t) &= \frac{\int f(t|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int \overline{F}(t|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}} \\ &= \int h(t|\boldsymbol{\theta})\left(\frac{\pi(\boldsymbol{\theta})\overline{F}(t|\boldsymbol{\theta})}{\int \overline{F}(t|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}}\right)d\boldsymbol{\theta}. \end{aligned} \tag{2}$$

Thus

$$h(t) = \int h(t|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|t)\,d\boldsymbol{\theta}, \tag{3}$$

where $\pi(\boldsymbol{\theta}|t)$ denotes the density of the distribution of $\boldsymbol{\theta}$ given that $T \ge t$.

Note that, contrary to what many believe, $h(t) \ne \int h(t|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}$; also, if $\pi(\boldsymbol{\theta})$ is degenerate, the model and predictive failure rates agree.

We conclude this section with a statement of the following important *closure* (*under mixtures*) *theorem*.

THEOREM 1 [Barlow and Proschan (1975) page 103]. *If the model failure rate $h(t|\boldsymbol{\theta})$ is decreasing in t, then the predictive failure rate $h(t)$ is decreasing in t, for any $\pi(\boldsymbol{\theta})$.*

A consequence of this theorem is the result that if $\int F(t|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}$, the *predictive life distribution*, is a mixture of exponential distributions, then the predictive failure rate will be strictly decreasing.

4. THE PSYCHOLOGICAL ASPECT OF BURN-IN

In this section, we argue that burn-in is a process of learning, where by learning we mean a reduction of uncertainty. The *optimal burn-in time* is the time at which the amount of information that is gleaned from the test balances the costs of the test, where costs include the depletion of useful life. We are prompted to make this claim as a consequence of observing that engineers subject *every* item that they use to a short life test prior to commissioning. This is true even of items that have an increasing model failure rate. For such items, burn-in would deplete useful life. When asked why every item is subjected to a burn-in, the answer has been that burn-in gives a "warm feeling" or "confidence" about an item's survivability. Thus engineering practice is contrary to the statistical literature, which seems to imply that only items having a strictly decreasing model failure rate should be subjected to a burn-in.

How can one explain engineers's actions which are contrary to the literature? Our explanation is that, with burn-in, we are learning by observing, so burn-in must be contemplated whenever we have uncertainty about the model failure rate, be it increasing, constant or decreasing. The depletion of useful life which occurs is the price that we pay for additional knowledge about the failure rate. The optimal burn-in time represents the optimal trade-off between knowledge and cost, and it may be greater than zero if the predictive failure rate is decreasing or has a decreasing segment.

5. THE PHYSICAL ASPECT OF BURN-IN

Is uncertainty about the model failure rate the only circumstance under which a burn-in should be contemplated? The answer is no, because burn-in may also be done in those situations wherein the act of using the item physically enhances its survivability. Examples include the work hardening of ductile materials and the self-sharpening of drill bits. Under such circumstances the model failure rate is decreasing, and one would contemplate a burn-in, even if the model failure rate were known with certainty. The predictive failure rate is of course decreasing by Theorem 1.

To summarize, burn-in should be contemplated for all items whose predictive failure rate is either monotonically decreasing or has a decreasing segment. Burn-in should be performed if the costs of testing compensate for the avoidance of risk, either because of our added knowledge or the physical enhancement of survivability, both of which make us feel good; hence the title of this paper.

6. THE ROLE OF UTILITY IN BURN-IN

Implicit in everything we have said above is the assumption that the event of interest is failure and that there is a positive utility associated with survival. Thus, neglecting costs associated with testing, a reduction in the predictive failure rate corresponds to an increase in the expected life and, therefore, the expected utility. Thus burn-in is only considered when the predictive failure rate is decreasing or has a decreasing segment.

However, the term "failure rate" is misleading, since we never stipulate that T is the time to failure. Indeed, T may represent the time to any event of interest, and the utility associated with the time before that event's occurrence may be negative. One example arises in statistical simulation, where an algorithm, such as the Gibbs sampler, is subjected to a burn-in to ensure its (weak) convergence. The idea here is that the algorithm experiences a phenomenon that is akin to work hardening, in the sense that each run is a stepping stone toward convergence. However, in this example, we define T to be the time until a specified convergence criterion is met; the utility associated with this time is negative. Furthermore, the model failure rate is increasing, since convergence becomes increasingly likely with each step of the algorithm.

Should we perform burn-in in this case? The answer to this question comes from a consideration of the costs. We conclude that burn-in should be contemplated whenever the predictive failure rate has an increasing segment; when the predictive failure rate is increasing, we will burn in indefinitely (i.e., until convergence is achieved). Burn-in will not be performed when the predictive failure rate is decreasing. These conclusions are opposite to those of Sections 4 and 5. Indeed, the simulation problem may be thought of as the dual (or "mirror image") of the usual engineering problem.

This scenario raises two important issues: (i) that, in essence, burn-in is a decision problem and cannot be answered without consideration of utility; and (ii) that the material of this paper is not restricted to the analysis of failure—rather, it applies to any situation where we have uncertainty surrounding the time of an event's occurrence.

7. AN ANATOMY OF FAILURE RATES WITH DECREASING SEGMENTS

7.1 Decreasing Failure Rates

We start off by asking the question, "What causes F to have a monotonically decreasing predictive failure rate?" Three reasons come to mind. These are (i) the *physics of failure* of an item, (ii) the *physical mixing* of *several* items, each having a decreasing but *known* model failure rate and (iii) the *subjective* (or *psychological*) *mixing* of a decreasing but *unknown* model failure rate. One may claim that (ii) above is a special case of (iii); however, it is helpful to distinguish between the two. We first elaborate on each of these and then address the question of a bathtub failure rate.

7.2 The Physics of Failure

The best examples of items whose time-to-failure distribution F has a monotonically decreasing model failure rate are those which experience work hardening. Examples include the curing of concrete slabs and the self-sharpening of drill bits. In all the above cases, the chemical bonds which hold together the atoms of a material strengthen over time or with use, making their failure increasingly unlikely over time.

In Figure 1 we illustrate several forms of decreasing model failure rates for an item and the resulting predictive failure rate, which by the closure under mixture theorem (Theorem 1), must also be decreasing. If $\pi(\boldsymbol{\theta}^*)$ was degenerate at $\boldsymbol{\theta}^*$, then there would be only one model failure rate (corresponding to $\boldsymbol{\theta}^*$) and the corresponding predictive failure rate would be $\boldsymbol{\theta}^*$.

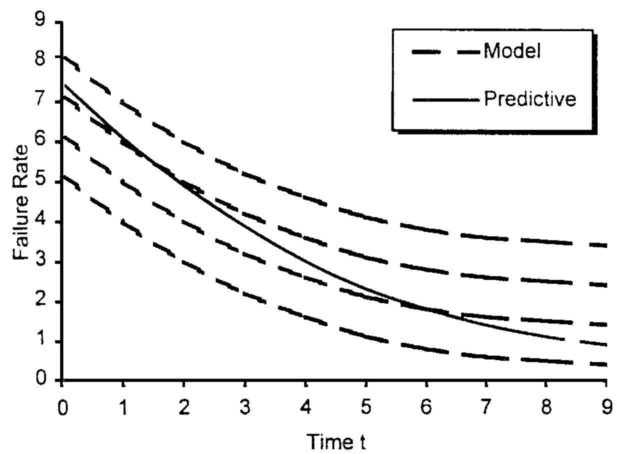

FIG. 1. *Monotonically decreasing model and predictive failure rates.*

7.2.1 *Physical mixing.* By "physical mixing" we mean the act of physically putting together several probabilistically heterogenous items that are otherwise indistinguishable, and inquiring about the stochastic behavior of an item picked (at random) from the mixture. For example, suppose that a bin contains n items, with the ith item having model failure rate $h(t|\boldsymbol{\theta}_i)$, $i = 1, 2, \ldots, n$, where $\boldsymbol{\theta}_i$ is assumed known. Suppose that the n items are otherwise indistinguishable, so that the model failure rate of an item picked at random from the bin is unknown. However, if the $\boldsymbol{\theta}_i$'s have a probability mass function $P(\boldsymbol{\theta}_i)$, then the predictive failure rate of the item picked at random is given by the discrete mixture

$$h(t) = \sum_{i=1}^{n} h(t|\boldsymbol{\theta}_i)P(\boldsymbol{\theta}_i|t), \tag{4}$$

where $P(\boldsymbol{\theta}_i|t)$ is the probability mass of $\boldsymbol{\theta}_i$ given that $T \geq t$.

Suppose now that each $h(t|\boldsymbol{\theta}_i)$ is decreasing in t. Then, by the closure under mixtures theorem (Theorem 1), $h(t)$ is also decreasing in t. In Figure 2 we illustrate this phenomenon via model failure rates that are constant. In fact, the closure under mixtures theorem was motivated by the physical mixing of constant model failure rates.

7.2.2. *Subjective* (*or psychological*) *mixing*. The notion of what we refer to as "subjective mixing" parallels that of physical mixing, except that now one does not conceptualize a mixing process that is prompted by physically putting together several heterogenous items. Rather, one acts as a subjectivist (in the sense of de Finetti and Savage), and

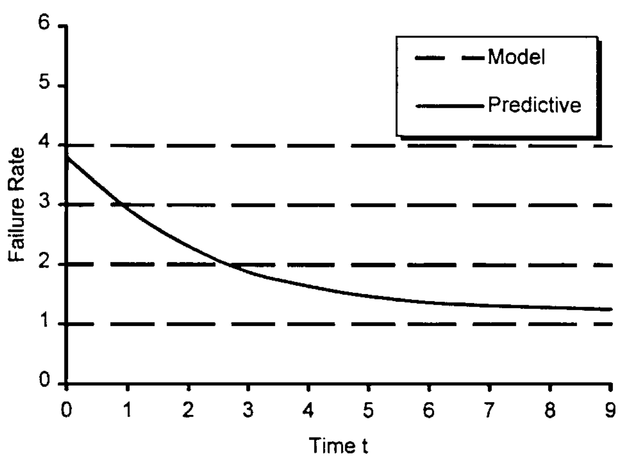

FIG. 2. *Monotonically decreasing predictive failure rate under physical mixing.*

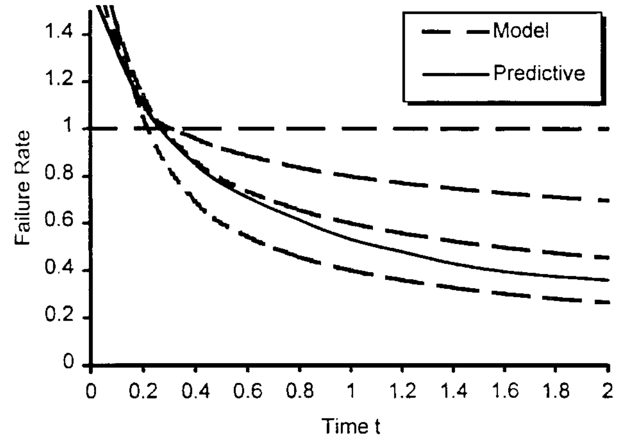

FIG. 3. *Monotonically decreasing predictive failure rate under subjective mixing over the Weibull shape parameter θ.*

mixes over the different model failure rates that are suggested by an unknown $\boldsymbol{\theta}$, via a prior $\pi(\boldsymbol{\theta})$ over $\boldsymbol{\theta}$.

Specifically, suppose that an item has a model failure rate $h(t|\boldsymbol{\theta})$ with $\boldsymbol{\theta}$ unknown. Let $\pi(\boldsymbol{\theta})$ reflect one's subjective opinion about the different values of $\boldsymbol{\theta}$; that is, $\pi(\boldsymbol{\theta})$ is the prior on $\boldsymbol{\theta}$. Suppose that $h(t|\boldsymbol{\theta})$ is decreasing in t for all values of $\boldsymbol{\theta}$. Then, by the closure under mixtures theorem, the predictive failure rate of the item $h(t)$ will also decrease in t. For example, suppose that $\boldsymbol{\theta} = \theta$ and that $F(t|\theta) = \exp(-t^{\theta})$, $\theta \geq 0$, $t > 0$, a Weibull distribution with shape parameter θ. If $\theta = 1$, $h(t|\theta) = 1$, a constant, whereas if $\theta < 1$, $h(t|\theta)$ decreases in t. Thus if $\pi(\theta)$ has support $(0, 1]$, then $h(t)$ decreases in t; see Figure 3.

Thus in the two scenarios of physical and subjective mixing, the predictive failure rate is decreasing in t, suggesting that burn-in should be contemplated.

7.3 Bathtub Failure Rates

We now ask the question, "What causes F to have a decreasing and then increasing failure rate?" It is difficult to think of an example from the physical sciences for which one could come up with a convincing argument about the changing behavior of chemical bonds. That is, the bonds must initially strengthen with use and then weaken. In the biological context, it has been conjectured that the immune system initially improves with age but then gets worse, and so a use of the bathtub curve in human mortality tables has a biological justification. However, the most convincing argument—at least to us—is that of mixing, either due to physical or, more likely, subjective

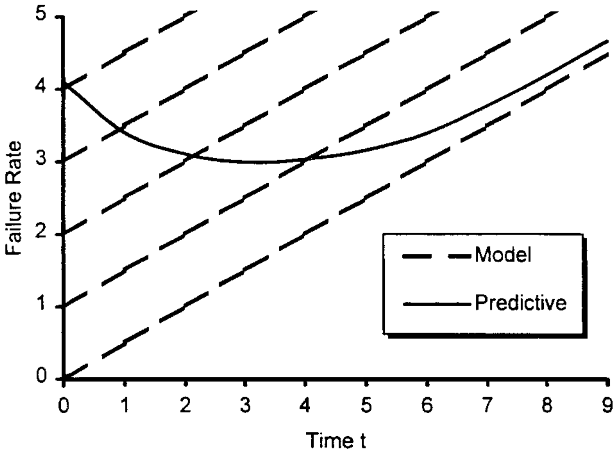

FIG. 4. *Subjective mixing of increasing model failure rates resulting in bathtub-shaped predictive failure rate.*

causes. As an example of the above suppose that F is a Rayleigh distribution, truncated at the left at zero, so that $h(t|\theta) = 2t + \theta$, with θ unknown. Let $\pi(\theta)$ have support $[0, \infty)$. Then it can be shown (see Theorem 2 below) that the predictive failure rate of F initially decreases and then increases, like a bathtub curve (see Figure 4). Gurland and Sethuraman (1994, 1995), discuss other cases wherein the mixture of increasing model failure rates could result in decreasing predictive failure rates. Their results suggest that in the presence of uncertainty, it is unusual for the predictive failure rate to be increasing.

The example depicted in Figure 4 suggests the following theorem, which is a generalization of the situation discussed.

THEOREM 2. *Suppose that* $h(t|\theta) = \alpha(t) + \theta$, *where* $\theta \geq 0$ *is unknown and* $\alpha(\cdot)$ *is convex. Let* $\pi(\theta)$ *describe our uncertainty about* θ, *and let* $V(\theta|t)$ *denote the variance of* θ *given* $T > t$. *Then* $h(t)$ *has a bathtub shape if*

$$\mathrm{Var}(\theta|0) > \frac{d}{dt}\alpha(0),$$

in which case the minimum occurs when $\mathrm{Var}(\theta|t) = (d/dt)\alpha(t)$.

This result follows from the fact that $h(t) = \alpha(t) + E[\theta|t]$, where $E[\theta|t]$ is a decreasing, convex function of t. The above theorem, as well as the example, show that the popular bathtub curve of reliability is not necessarily physically realistic. Rather, it is a consequence of belief produced by the process of subjectively mixing increasing model failure rates having certain properties. Note how the shape of the predictive failure rate is directly linked to our uncertainty, via the prior variance.

8. THE OPTIMAL BURN-IN TIME

The question, "When should we burn-in?", leads us naturally to the issue of an optimal burn-in time. To address this issue, let us first put into perspective the circumstances under which the predictive failure rate has a decreasing segment. These are (i) mixing due to uncertainty about constant, increasing or decreasing model failure rates and (ii) model failure rates which are strictly decreasing because of physical circumstances, but about which we are certain.

Under (i) above, burn-in can be viewed as a process of learning, that is, a reduction of uncertainty about T. To see this, suppose that the optimal burn-in time is $\tau \geq 0$. When the burn-in test shows that $T > \tau$, our predictive ability about T sharpens (via added knowledge about $\boldsymbol{\theta}$). If the burn-in test shows that $T \leq \tau$, then $\overline{F}$ is degenerate at some $t \in (0, \tau]$, and the item tested is declared a weak one. Thus for predictive failure rates given by mixtures, be they decreasing or bathtub, burn-in gives us added knowledge. The price we pay for this knowledge is the cost of testing and the depletion of useful life if the model failure rate is increasing in t. The optimal τ is a trade-off between the costs and the utility of reduced uncertainty (see Theorem 3, below). Clearly, burn-in should not be done if (i) the predictive failure rate $h(t)$ is increasing in t or (ii) our trade-off calculations show that $\tau = 0$; see Theorem 2.

8.1 The Scenario of Indefinite Burn-in: Eternal Happiness

A situation of interest is that of $h(t|\boldsymbol{\theta})$ strictly decreasing in t, with $\pi(\boldsymbol{\theta})$ being degenerate. If the costs of burn-in are zero, then $\tau \to \infty$, because burn-in enhances useful life. This implies that indefinite burn-in leads to eternal happiness! However, since $\boldsymbol{\theta}$ is an abstraction (just a Greek symbol to de Finetti), a degenerate $\pi(\boldsymbol{\theta})$ is not realistic, and thus eternal happiness is a myth. If the costs of burn-in are greater than zero, then τ is the time at which the costs of burn-in and the utility of enhanced life due to burn-in balance out.

The above matters are summarized and quantified via the following theorem due to Clarotti and Spizzichino (1990)—extended further by Block, Mi and Savits (1993).

Suppose that $\overline{F}$ has a density f, and suppose that $g(t) \equiv f(t+s)/f(t)$ increases in t for all $s > 0$. Let c_1 denote the cost if $T < \tau$, let C be the cost if

$\tau \leq T \leq \tau + s$ (where s can be viewed as the mission time) and let $-K$ be the reward if $T > \tau + s$. Then:

THEOREM 3 [Clarotti and Spizzichino (1990)].

(i) *Burn in indefinitely, iff*

$$\lim_{t \to \infty} g(t) < \frac{C - c_1}{C + K} = v.$$

(ii) *Do not burn in, iff* $g(0) \geq v$.
(iii) *Burn in for time* $\tau > 0$, *iff* $g(\tau) \equiv v$.

Note that the indefinite burn-in of (i) above is different from the indefinite burn-in of eternal happiness discussed above. The former is based on costs of testing and in-service failure; the latter assumes that the costs of burn-in are zero.

9. CONCLUDING COMMENTS

Let us return to the original question: "What is burn-in"? We argue that it is primarily a mechanism for learning.

The model failure rate describes the physical process of aging. The predictive failure rate describes our changing beliefs about an item as we observe it surviving. Since burn-in is performed for a psychological purpose, it is only natural to base burn-in calculations upon the predictive failure rate. The modeland predictive failure rates may have very different forms. Indeed, while the famous bathtub curve rarely has a physical motivation, it arises quite naturally in our minds.

ACKNOWLEDGMENTS

The authors would like to thank Dennis Lindley and Al Marshall for their helpful comments on an earlier draft of this paper. Supported by Army Research Office Grant DAAH04-93-G-0020 and Air Force Office of Scientific Research Grant AFOSR-F-49620-95-1-0107.

REFERENCES

BARLOW, R. E. and PROSCHAN, F. (1975). *Statistical Theory of Reliability: Probability Models*. Holt, Rinehart & Winston, New York.

BESAG, J., GREEN, P., HIGDON, D. and MENGERSEN, K. (1995). Bayesian computation and stochastic systems (with discussion). *Statist. Sci.* **10** 3–66.

GURLAND, J. and SETHURAMAN, J. (1995). How pooling data may reverse increasing failure rates. *J. Amer. Statist. Assoc.* **90** 1416–1423.