Lifetime Data Analysis, 6, 157–179 (2000)

Alternative Time Scales and Failure Time Models

THIERRY DUCHESNE duchesne@utstat.toronto.edu
Department of Statistics, University of Toronto, Toronto, ON, Canada, M5S 3G3

JERRY LAWLESS jlawless@setosa.uwaterloo.ca
Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada, N2L 3G1

Received January 6, 1999; Revised July 8, 1999; Accepted September 9, 1999

Abstract. In many reliability applications, there may not be a unique plausible scale in which to measure time to failure or assess performance. This is especially the case when several measures of usage are available on each unit. For example, the age, the total number of flight hours, and the number of landings are usage measures that are often considered important in aircraft reliability. Similarly, in medical or biological applications of survival analysis there are often alternative scales (e.g., Oakes, 1995). This paper considers the definition of a "good" time scale, along with methods of determining a time scale.

Keywords: accelerated lifetime models, exposure, reliability, survival analysis, time transformation, time-varying covariate, usage measure

1. Introduction

In many failure time applications there may not be a unique plausible scale in which to model or analyze failures. For example, with automobiles we might use age of the vehicle or miles driven as time scales in assessing reliability; with aircraft engines we might use total hours of operation since the last major overhaul, the total number of engine cycles, or perhaps some composite measure of duration and stress that takes account of engine speed and startup periods. In biological or medical applications, examples include problems involving exposure to hazardous substances (e.g., Oakes, 1995) and problems concerning the choice of a time origin in disease history or treatment studies (e.g., Farewell and Cox, 1979).

Methods of survival and failure time analysis assume that a time scale has been selected. The question of what constitutes a "good" time scale is often mentioned but has received rather little attention in the literature, with a few notable exceptions (e.g., Farewell and Cox, 1979; Oakes, 1995; Kordonsky and Gertsbakh, 1993, 1995a, 1995b, 1997a). Similarly, the question of how to select a time scale has not received much study. The purpose of this paper is to address these issues.

Because real world processes and their consequences occur in "real" or chronological time, it obviously plays an important role. Nevertheless, there are reasons for adopting an alternative time scale t in certain contexts:

1. The failure process may depend primarily on t, so that it is scientifically "sufficient". Care is needed in formalizing this concept, but, for example, the probability a unit survives beyond a certain chronological time x may depend only on its t-value at x.

2. A time scale t as described in 1. has advantages for experimental design and decision making. For example, if the failure process for an automobile component depends primarily on the mileage accumulated then reliability studies in which test vehicles are driven many miles over short time periods may be extrapolated to ordinary customer usage conditions. Environmental stresses such as temperature, humidity or road conditions may also be incorporated into time scales.

3. A time scale t for which the failure distribution is concentrated over a short interval is useful for monitoring systems and scheduling maintenance and replacement interventions. The same is true for individuals being monitored for the occurrence of some type of event.

4. The effect of treatments or covariates may be most directly or simply expressed on some scale other than chronological time; a similar comment applies to the choice of alternative time origins for chronological time (e.g., Breslow et al., 1983).

Farewell and Cox (1979), Kordonsky and Gertsbakh (1993, 1995a, 1995b, 1997a) and Oakes (1995) have discussed the selection of a time scale from linear combinations of specified variables. However, there has been little comprehensive discussion of time scales and that is our objective here; we provide a general framework and new results concerning time scale specification. In Section 2 we present a general conceptual framework in which the analysis of time scales is equivalent to the analysis of models of failure in the presence of time-varying covariates. In Section 3 we examine known classes of failure time models and show how they fit within our framework. Approaches to model fitting and time scale selection methods are discussed in Section 4, and new semiparametric approaches are proposed. In Section 5 we examine the concept of small variation in failure times and give new results for an approach due to Kordonsky and Gertsbakh (1993, 1995a, 1995b, 1997a) which involves choosing a scale for which the coefficient of variation (CV) for failure time is minimized. Section 6 presents examples involving automobile reliability and the fatigue life of steel, respectively. Some additional discussion is given in Section 7.

2. Ideal Time Scales

In the following, we let $x \geq 0$ represent chronological time measured from an origin that typically corresponds to the "birth" or introduction of a unit into service. The random variable X denotes the chronological time or age at failure for an item or unit. Our investigation of time scales is based on the idea of time-varying covariates (see Kalbfleisch and Prentice, 1980) associated with each unit. These are defined so as to be left-continuous functions of x, and represent factors that are thought to be related to failure. For each unit, we have covariates $z_1(x), \ldots, z_q(x)$, $x \geq 0$. A special class of covariates will be of particular interest later on: *usage* or *exposure* measures. We define usage measures to be external time-varying covariates that are non-decreasing in x. We sometimes derive usage measures $y_i(x)$ as fully specified increasing functionals of $\{z_1(u), \ldots, z_q(u); 0 \leq u \leq x\}$. For example, $z_1(x)$ might be the engine temperature for a car at age x, and $y_1(x)$ could represent the cumulative number of degree-hours on this car's engine at x.

It is convenient to denote $y_0(x) = x$ and let $\mathbf{y}(x) = (y_0(x), y_1(x), \ldots, y_p(x))^t$ and $\mathbf{z}(x) = (z_1(x), \ldots, z_q(x))^t$. We term $\mathcal{P}(x) = \{\mathbf{y}(u), 0 \leq u < x\}$ and $\mathcal{Z}(x) = \{\mathbf{z}(u), 0 \leq u < x\}$ as the *usage* and *covariate path* (or *history*) up to chronological time x, respectively. The set of all possible usage [covariate] histories $\mathcal{P}(x)$ $[\mathcal{Z}(x)]$ is denoted by $\mathbb{P}(x)$ $[\mathbf{Z}(x)]$, and for convenience we write $\mathbb{P}$ $[\mathbf{Z}]$ for $\mathbb{P}(\infty)$ $[\mathbf{Z}(\infty)]$. A *time scale* (TS) is then a function of chronological time and covariates. This is consistent with other authors such as Kordonsky and Gertsbakh (1993, 1995a, 1995b, 1997a) and Oakes (1995). Formally, we consider a TS to be a non-negative-valued functional $\Phi : \mathbb{R}^+ \otimes \mathbf{Z} \to \mathbb{R}^+$ that maps $(x, \mathcal{Z})$ to $\Phi[x, \mathcal{Z}(x)]$ and such that $\Phi[x, \mathcal{Z}(x)]$ is non-decreasing in x for all $\mathcal{Z} \in \mathbf{Z}$. The value $t_{\mathcal{Z}}(x) = \Phi[x, \mathcal{Z}(x)]$ is the operational time ("time" for short) at x for the covariate path $\mathcal{Z}$.

In this paper, we focus on covariates that are *external*. An external covariate (Kalbfleisch and Prentice 1980, Section 5.3) is one for which the history $\mathcal{Z}$ for a unit is determined independently of the failure process for that unit, such as an environmental or usage factor. In some applications one might want to consider time scales that incorporate "internal" measures such as the amount of physical deterioration in a unit. These are more difficult to address and we defer discussion to Section 7.

Covariate paths $\mathcal{Z}$ are assumed to vary from unit to unit in the population. Although they are often randomly determined, the fact that they are external means that the distribution of failure time conditional on the realized path $\mathcal{Z}$ for a unit is easily interpreted. For the chronological time of failure X, we make the reasonable assumption that

$$\Pr[X > x|\mathcal{Z}] = \Pr[X > x|\mathcal{Z}(x)], \; x \geq 0 \tag{1}$$

for all $\mathcal{Z} \in \mathbf{Z}$. By conditioning on $\mathcal{Z}$ we make $t_{\mathcal{Z}}(x) = \Phi[x, \mathcal{Z}(x)]$ a specified function of x, and we can therefore also determine the distribution of the generalized failure time $T = \Phi[X, \mathcal{Z}(X)]$ given $\mathcal{Z}$ from (1).

Some authors (e.g., Singpurwalla and Wilson, 1993 or Murthy et al., 1995) approach multiple time scales by modeling directly the joint distribution of an operational failure time T and chronological failure time X. In our framework a joint model for (T, X) can be obtained by specifying a distribution for the covariate paths $\mathcal{Z}$, and we can examine the effect of different paths or distributions for $\mathcal{Z}$ on the distribution of (T, X). The approach we take is more comprehensive; note for example that models for (T, X) alone do not use information about $\mathcal{Z}$ except for the single value $T = \Phi[X, \mathcal{Z}(X)]$. In addition, our approach allows alternative time scales to be considered in situations where the covariate paths $\mathcal{Z}$ are not random but are fixed by study design, as in accelerated life test experiments (c.f., Nelson, 1990, Chapter 10).

We would like to develop concepts that lead to "good" time scales. If $t_{\mathcal{Z}}(x) = \Phi[x, \mathcal{Z}(x)]$ is to be "sufficient" for the calculation of failure probabilities, then we want for t such that $t = \Phi[x, \mathcal{Z}(x)]$ for some x that

$$\begin{aligned} \Pr[T > t|\mathcal{Z}] &= \Pr[T > t] \\ &= G(t), \end{aligned} \tag{2}$$

where $T = \Phi[X, \mathcal{Z}(X)]$, and $G(\cdot)$ does not depend on $\mathcal{Z}$. In addition, we want t to change

whenever

$$\Pr[X > x|\mathcal{Z}] = S_0[x, \mathcal{Z}(x)]$$

does. This leads us to define an ideal time scale as follows:

Definition 2.1. $t_{\mathcal{Z}}(x) = \Phi[x, \mathcal{Z}(x)]$ is an *ideal time scale* if it is a one-to-one function of $S_0[x, \mathcal{Z}(x)]$. Furthermore, in that case

$$\begin{aligned}\Pr[X > x|\mathcal{Z}] &= G[t_{\mathcal{Z}}(x)] \\ &= \Pr[T > t_{\mathcal{Z}}(x)],\end{aligned}$$

where $G(\cdot)$ is a survivor function that does not depend on $\mathcal{Z}$.

This definition has been considered by various authors (e.g., Cinlar and Ozekici (1987) and Kordonsky and Gertsbakh (1993), who use the terms "intrinsic" and "load-invariant" time scales, respectively). Definition 2.1 implies that T is statistically independent of $\mathcal{Z}$ in situations where the covariate paths are random, but it is of course functionally dependent on $\mathcal{Z}$. Under this definition t is "sufficient" for describing the failure process as far as the covariates that make up $\mathcal{Z}$ are concerned. An investigation of ITS's is equivalent to an investigation of models for X conditional on $\mathcal{Z}$.

Any set of external covariates will generate an ITS, and an ITS is not necessarily "good" or useful. A search for a "good" ITS is equivalent to a search for "good" covariates and "good" models for $\Pr[X > x|\mathcal{Z}(x)]$; that is, we want models for which $\mathcal{Z}$ is highly predictive for X. We therefore now consider approaches to regression modeling; then we reconsider the concept of a "good" time scale and modeling strategies.

3. Models of Failure and Time Scales

If $S(x|\mathcal{Z}) = \Pr[X > x|\mathcal{Z}]$ is continuous at all x, let $h(x|\mathcal{Z})$ be the corresponding hazard function, so that

$$S(x|\mathcal{Z}) = \exp\left\{-\int_0^x h(u|\mathcal{Z})\,du\right\}. \tag{3}$$

We note that if $t_{\mathcal{Z}}(x) = \Phi[x, \mathcal{Z}(x)]$ is any ITS relative to $\mathcal{Z}$ then from Definition 2.1 we have

$$\Pr[X > x|\mathcal{Z}] = G[t_{\mathcal{Z}}(x)] \tag{4}$$

and

$$h(x|\mathcal{Z}) = h_G[t_{\mathcal{Z}}(x)]t'_{\mathcal{Z}}(x), \tag{5}$$

where $h_G(t) = -G'(t)/G(t)$ is the hazard function corresponding to the survivor function $G(t)$ and $t'_{\mathcal{Z}}(x) = dt_{\mathcal{Z}}(x)/dx$.

Sometimes the distribution of X given $\mathcal{Z}$ may have a non-zero probability mass at certain x values; for example, if equipment is turned on and off then there may be a non-zero probability of failure at the instant a unit is turned on (e.g., Follmann, 1990). Then we allow $S(x|\mathcal{Z}) = \Pr[X > x|\mathcal{Z}]$ to have jump discontinuities. The survivor function can be written as (e.g., Kalbfleisch and Prentice, 1980, Section 1.2.3)

$$S(x|\mathcal{Z}) = \exp\left\{-\int_0^x h_C(u|\mathcal{Z})\,du\right\} \prod_{u_j \le x} \left\{1 - h_D(u_j|\mathcal{Z})\right\}, \tag{6}$$

where $h_C(u|\mathcal{Z})$ is an integrable hazard function corresponding to the continuous part of $S(x|\mathcal{Z})$ and $u_1, u_2, \ldots$ are the jump points for $S(x|\mathcal{Z})$. The values $h_D(u_j|\mathcal{Z})$ are the discrete hazard function components

$$h_D(u_j|\mathcal{Z}) = \Pr[X = u_j | X \ge u_j, \mathcal{Z}],$$

and

$$S(x|\mathcal{Z}) = \exp\left\{-\int_0^x h_C(u|\mathcal{Z})\,du + \int_0^x \log[1 - h_D(u|\mathcal{Z})]\,dN(u|\mathcal{Z})\right\}, \tag{7}$$

where $dN(u|\mathcal{Z})$ equals 1 if a jump in $S(x|\mathcal{Z})$ occurs at u, and 0 otherwise. Note that

$$t_{\mathcal{Z}}(x) = \Phi[x, \mathcal{Z}(x)] = \int_0^x h_C(u|\mathcal{Z})\,du + \int_0^x \log[1 - h_D(u|\mathcal{Z})]\,dN(u|\mathcal{Z}) \tag{8}$$

is an ITS, with $G(t) = \exp(-t)$ in the format (2). Any one-to-one function of $t_{\mathcal{Z}}(x)$ also defines an ITS.

3.1. Collapsible Models

Oakes (1995) introduced the notion of collapsibility.

Definition 3.1. Let $y_1(x), \ldots, y_p(x)$ be a specified set of usage factors. Then the distribution of $X|\mathcal{P}$ is collapsible in $y_1(x), \ldots, y_p(x)$ if

$$\Pr[X > x|\mathcal{P}] = S_0[\mathbf{y}(x)]. \tag{9}$$

That is, the survival probability at chronological time x depends on the path $\mathcal{P}(x)$ up to x only through the endpoint $\mathbf{y}(x)$. In this case, ITS's are of the form

$$\begin{aligned} t_{\mathcal{P}}(x) &= \Phi[x, y_1(x), \ldots, y_p(x)] \\ &= \Phi[\mathbf{y}(x)]. \end{aligned} \tag{10}$$

It may seem that collapsibility is a very restrictive assumption. However, the possibility of defining measures $y_j(x)$ based on stress, usage or environmental factors allows considerable

flexibility. Note that collapsibility is defined in terms of usage measures, rather than general external covariates. It is not possible to define a function of $x, z_1(x), \ldots, z_q(x)$ that would be non-decreasing in x for all $\mathcal{Z} \in \mathbf{Z}$, unless the functions z_j's themselves are non-decreasing.

Collapsible models have not been studied or used much, but have some nice properties. One is that if we wish to compute the marginal distribution of X for some specified probability distribution on the usage paths $\mathcal{P}$, then (9) implies that

$$\Pr[X > x] = \mathrm{E}\{S_0[\mathbf{y}(x)]\}.$$

Thus the expectation involves only the distribution of $y_1(x), \ldots, y_p(x)$ for the given x and not the entire path $\mathcal{P}(x)$ up to x. A more general property that implies the first one concerns prediction through $\Pr[X > x + s | X \geq x, \mathcal{P}(x)]$. It is easily seen that

$$\begin{aligned}
\Pr[X > x + s | X \geq x, \mathcal{P}(x)] &= \mathrm{E}\left\{ \exp\left[-\int_x^{x+s} h(u|\mathcal{P}(u))\, du \right] \middle| \mathcal{P}(x) \right\} \\
&= \mathrm{E}\left\{ \frac{S_0[\mathbf{y}(x+s)]}{S_0[\mathbf{y}(x)]} \middle| \mathcal{P}(x) \right\} \\
&= \frac{\mathrm{E}\{S_0[\mathbf{y}(x+s)] | \mathcal{P}(x)\}}{S_0[\mathbf{y}(x)]}.
\end{aligned}$$

The required expectation concerns just the usage vector at time $x + s$ and not the entire path from x to $x + s$.

In some settings distinct usage paths never cross, e.g., when $y_j(x)$'s are linear in x. In that case all models for X given $\mathcal{P}$ are collapsible in the trivial sense that each $\mathcal{P}$ is identified by its $y(x)$ value at any given time x. However, the ITS (10) is not in general of simple functional form. The attraction of using collapsible models is to consider fairly simple parametric specifications $\Phi[y(x); \boldsymbol{\eta}]$ in (10), thus yielding easily interpreted operational times. Oakes (1995) and Kordonsky and Gertsbakh (1993, 1995a, 1995a, 1997a) consider models in which

$$t_{\mathcal{P}}(x) = \boldsymbol{\eta}^t \mathbf{y}(x)$$

is assumed linear in x and $y_1(x), \ldots, y_p(x)$, and G in (4) has a specified parametric form. Semiparametric approaches can also be adopted. For example, Kordonsky and Gertsbakh consider a method of estimating $\boldsymbol{\eta}$ without a parametric model for G; this is examined in Section 5. We outline approaches to model fitting and selection in Section 4, and a more detailed treatment of collapsible models will be given elsewhere.

3.2. Generalized Time Transform Models

A time transform (TT) or accelerated failure time (AFT) model is one in which (4) holds with the ITS

$$t_{\mathcal{Z}}(x) = \int_0^x \psi[\mathbf{z}(u)]\, du, \tag{11}$$

where ψ is a positive-valued function, and G is a survivor function. Such models have been studied in survival analysis (e.g., Cox and Oakes, 1984; Robins and Tsiatis, 1992; Lin and Ying, 1995) and also have an extensive history in reliability (e.g., Nelson, 1990, Chapter 10, Doksum and Hóyland, 1992), where the $z_j(u)$'s typically represent time-varying stress factors. The model with (11) written in terms of $r[\mathbf{z}(u)] = \psi^{-1}[\mathbf{z}(u)]$ is a probabilistic analog of Miner's (1945) rule.

Bagdonavičius and Nikulin (1997) have proposed models that extend the AFT model. Indeed, their formulations are sufficiently broad to include multiplicative hazards, proportional odds, and other models. We consider here only Model 2 of Bagdonavičius and Nikulin (1997): in this case Pr $[X > x|\mathcal{Z}]$ is of the form (4) but $t_{\mathcal{Z}}(x)$ is given by

$$t_{\mathcal{Z}}(x) = \int_0^x \psi[\mathbf{z}(u)]\, dG^{-1}S_0(u), \tag{12}$$

where S_0 is also a survivor function. If $G = S_0$ then (12) reduces to (11) and gives the AFT model. If we select $G(u) = \exp(-u)$, on the other hand, and let $S_0(u)$ be arbitrary, then we get

$$\Pr\,[X > x|\mathcal{Z}] = \exp\left\{-\int_0^x \psi[\mathbf{z}(u)]h_0(u)\, du\right\}, \tag{13}$$

where $h_0(u) = -S_0'(u)/S_0(u)$ is the hazard function corresponding to $S_0(u)$. This is the well known multiplicative hazards model (Cox, 1972).

The assumptions and approaches taken with AFT and collapsible models are rather different. With AFT models it is customary to specify $\psi[\mathbf{z}(u)]$ in (11) parametrically: for example, Lin and Ying (1995) and others consider $\psi[\mathbf{z}(u)] = \exp\{\boldsymbol{\beta}^t\mathbf{z}(u)\}$, and many reliability models (e.g., Nelson, 1990, Chapter 10) can be written in this form. However, it is clear that in this case the ITS

$$t_{\mathcal{Z}}(x) = \int_0^x \exp\{\boldsymbol{\beta}^t\mathbf{z}(u)\}\, du \tag{14}$$

cannot in general be expressed in the collapsible model form (10) for any usage measures $y_1(x), \ldots, y_p(x)$ that can be represented as fully specified functionals of $\{\mathbf{z}(u), 0 \le u \le x\}$.

3.3. Hazard-Based Specifications

When time-varying covariates are present it is convenient to specify failure time models via the hazard function. For arbitrary covariates $\mathbf{z}(x)$ we make the reasonable assumption that the hazard function of X given $\mathcal{Z}$ satisfies

$$h[x|\mathcal{Z}] = h[x|\mathcal{Z}(x)], \tag{15}$$

so that in the continuous case (3) becomes

$$\Pr\,[X > x|\mathcal{Z}] = \exp\left\{-\int_0^x h[u|\mathcal{Z}(u)]\, du\right\}. \tag{16}$$

As noted in (8), this can be considered as an ITS model with

$$t_{\mathcal{Z}}(x) = \int_0^x h[u|\mathcal{Z}(u)]\,du. \tag{17}$$

Discrete components can be incorporated into the models as in (8).

AFT models introduced in Section 3.2 assume that

$$h[x|\mathcal{Z}(x)] = \psi[\mathbf{z}(x)]h_G\left(\int_0^x \psi[\mathbf{z}(u)]\,du\right),$$

whereas the collapsible models of Section 3.1 assume that (replacing $\mathcal{Z}$ with $\mathcal{P}$)

$$h[x|\mathcal{P}(x)] = \frac{d\Phi[\mathbf{y}(x)]}{dx}h_G(\Phi[\mathbf{y}(x)]).$$

It was noted previously that AFT and collapsible models are generally quite distinct. A case where they intersect is for *additive hazard models* with positive covariates, for which

$$h[x|\mathcal{Z}(x)] = h_0(x) + \boldsymbol{\beta}^t\mathbf{z}(x), \tag{18}$$

with $h_0(x)$ a baseline hazard function. Then

$$t_{\mathcal{Z}}(x) = \int_0^x h[u|\mathcal{Z}(u)]\,du = H_0(x) + \boldsymbol{\beta}^t\mathbf{y}^*(x),$$

where $H_0(x) = \int_0^x h_0(u)\,du$ and $\mathbf{y}^*(x) = (y_1(x), \ldots, y_q(x))^t$, with

$$y_j(x) = \int_0^x z_j(u)\,du, \qquad j = 1, 2, \ldots, q.$$

By (10), this model is collapsible. Jewell and Kalbfleisch (1996) and Singpurwalla (1995) consider models of the form (18) along with stochastic models for the covariate processes, and develop some specialized prediction formulas of the type mentioned in Section 3.1.

Another widely-used family of models with time-varying covariates is the multiplicative hazards family (13), where $h[x|\mathcal{Z}(x)] = \psi[\mathbf{z}(x)]h_0(x)$. These can be viewed as generalized TT models arising from (12) but they are not in general collapsible.

3.4. "Good" Time Scales

It is not possible to give a universal definition of a good time scale. Rather, we identify qualities that a good scale t should possess: (i) scientific relevance, (ii) a parsimonious and accurate description of variation in failure times under varying usage, exposure or environmental conditions, (iii) a relatively "compact" distribution for operational time T, (iv) succinct and meaningful summarization of the effects of fixed or time-varying covariates that are of special interest. The selection of covariates to be considered for time scale construction and the relative weight given to qualities (i)-(iv) will vary according to the specific context.

In reliability and other areas the quality (iii) of "small" variation in T is often mentioned, but this is somewhat elusive. We restrict attention to ITS's since, given a set of time-varying factors, they are "sufficient" in the sense of (2). However, ITS's are defined only up to monotone transformations so that marginal variation in T cannot be the only consideration, even if it is adjusted for scale of measurement. In addition, the "goodness" of T is dependent on the existence and selection of "good" time-varying factors for $\mathcal{Z}$. The crucial issue is as for prediction using regression models: we want factors for which the relative variation in X is small, given $\mathcal{Z}$. This translates into small variation in T. Some aspects of the small variation concept are considered in Section 5.

Not all covariates in a given situation need to be used for a time scale. For simplicity we define ITS's in (2) as a function $\Phi[x, \mathcal{Z}(x)]$ of all covariates, but we could have additional covariates $\mathbf{w}(x)$ which are conditioned on but not used in T. In that case we replace (2) with the requirement that

$$\Pr\,[T > t|\mathcal{Z}, \mathcal{W}] = \Pr\,[T > t|\mathcal{W}],$$

where $\mathcal{W}$ is the covariate history $\{w(x); x \geq 0\}$. For example, in a clinical setting $\mathcal{W}$ might represent treatments or other discrete individual level factors such as sex. For simplicity we will continue to write T as in (2), however.

3.5. Modeling Strategies

There are two broad approaches with which we may investigate time scales. One is the traditional failure time analysis approach via models for X given the covariate history $\mathcal{Z}$. This leads to a specification of the conditional hazard function for X as $h(x|\mathcal{Z}(x))$, and

$$\Pr\,[X > x|\mathcal{Z}(x)] = \exp\left\{-\int_0^x h(u|\mathcal{Z}(u))\,du\right\} \tag{19}$$

in the continuous case. Standard modeling paradigms such as proportional hazards or accelerated failure time may be used. Within these frameworks we would search for the time scales by examining the cumulative hazard in (19), as described in Sections 3.2 and 3.3.

The other approach is to express (19) directly as a function of an ITS:

$$\Pr\,[X > x|\mathcal{Z}(x)] = G[t_{\mathcal{Z}}(x)]. \tag{20}$$

In this case we have the option of specifying G and $t_{\mathcal{Z}}(x)$ parametrically or semiparametrically. Collapsible models are a special case of (20) in which $t_{\mathcal{Z}}(x)$ is of the form $\Phi[\mathbf{y}^*(x)]$, where $\mathbf{y}^*(x)$ is a vector consisting of x and usage measures. With this approach the search for a time scale is emphasized, and models with meaningful or easily interpreted scales may readily be considered.

Example 3.1. (Linear TS models). Suppose there is a single usage factor $y(x)$ and no other time-varying covariates. Some authors (e.g., Kordonsky and Gertsbakh, 1993; Oakes, 1995)

have considered collapsible models for which

$$\Pr\,[X > x|\mathcal{P}] = G[\eta_0 x + \eta_1 y(x)], \tag{21}$$

so that $t_{\mathcal{P}}(x) = \eta_0 x + \eta_1 y(x)$ is an ITS. Here η_0 and η_1 are real parameters and G is a distribution function which may also involve unknown parameters.

It is interesting to compare an approach of Farewell and Cox (1979), which also emphasizes linear TS's. They search for a model for which

$$h(x|\mathcal{P}) = h_0[\eta_0 x + \eta_1 y(x)]. \tag{22}$$

It is readily seen that this model is not collapsible in general, since

$$\Pr\,[X > x|\mathcal{P}] = \exp\left\{-\int_0^x h_0[\eta_0 u + \eta_1 y(u)]\,du\right\}.$$

In fact, $\psi(x) = \eta_0 x + \eta_1 y(x)$ is not an ITS and so the Farewell-Cox use of linear TS's is quite different than that in (21).

4. Model Fitting and Time Scale Selection

In this section we outline some recent and some new approaches to estimation and time scale selection, in which we consider parametric specifications for $t_{\mathcal{Z}}(x)$ in (20). If a parametric specification for the distribution G is also adopted, then estimation via maximum likelihood is straightforward, as indicated by Oakes (1995) and reviewed in Subsection 4.1. In some cases we may prefer to leave G nonparametric; we discuss this in Subsection 4.2 and we propose some new approaches. A more detailed discussion will be given elsewhere.

4.1. Parametric Estimation Based on a model (20)

Parametric estimation based on a model (20) is easily implemented (e.g., Kordonsky and Gertsbakh, 1993 or Oakes, 1995). Suppose for example that $G(t) = G(t;\phi)$ and that $t_{\mathcal{Z}}(x) = t_{\mathcal{Z}}(x;\boldsymbol{\eta})$, where ϕ and $\boldsymbol{\eta}$ are parameter vectors. Consider a random sample of n units, of which some fail and some are censored. Let x_i denote the failure or censoring time (chronological time) for unit i, and let $\delta_i = 1$ if x_i is a failure time and $\delta_i = 0$ if it is a censoring time. All covariates $\mathcal{Z}$ are external and their marginal distribution is assumed to have no information about ϕ or $\boldsymbol{\eta}$. Thus we condition on $\mathcal{Z}$ and the likelihood function under a continuous model (20) is

$$\begin{aligned} L(\phi,\boldsymbol{\eta}) &= \prod_{i=1}^{n} f_G[t_{\mathcal{Z}_i}(x_i)]^{\delta_i} G[t_{\mathcal{Z}_i}(x_i)]^{1-\delta_i} \\ &= \prod_{i=1}^{n} \left\{-G'[t_{\mathcal{Z}_i}(x_i)]t'_{\mathcal{Z}_i}(x_i)\right\}^{\delta_i} G[t_{\mathcal{Z}_i}(x_i)]^{1-\delta_i}, \end{aligned} \tag{23}$$

where $f_G(t) = -G'(t)$ is the density function corresponding to $G(t)$ and $t'_{\mathcal{Z}}(x) = dt_{\mathcal{Z}}(x)/dx$. Note that (23) involves $t'_{\mathcal{Z}_i}(x_i)$ at failure times, even for collapsible models.

The above assumes implicitly that $t_{\mathcal{Z}}(x; \boldsymbol{\eta})$ is an ITS for some $\boldsymbol{\eta}$. Constructing a model that includes (20) as a sub-model can serve as a model assessment procedure. Suppose for example that we wish to consider a collapsible model (20) with $t_{\mathcal{Z}}(x) = t[x, y(x); \boldsymbol{\eta}]$ for some usage measure $y(x)$. We might consider the expanded family

$$\Pr\,[X > x|\mathcal{Z}(x)] = G\left\{t[x, y(x); \boldsymbol{\eta}]e^{\boldsymbol{\beta}'\mathbf{z}^*(x)}\right\}, \tag{24}$$

where $\mathbf{z}^*(x)$ is a vector of additional covariates which may include functions of x; if $\boldsymbol{\beta} = \mathbf{0}$ the collapsible model holds. This approach is also valuable when we wish to select a time scale based on certain factors, but other covariates are to be included in the model.

Farewell and Cox (1979) use this general approach to select a linear time scale $t = \eta_0 x + \eta_1 y(x)$ by defining $\mathbf{z}^*(x)$ as $-\eta_1 x + \eta_0 y(x)$ and finding $\boldsymbol{\eta}$ to give an estimate of $\beta = 0$. However, as discussed in Section 3.5, they employ a multiplicative hazards framework and their method does not give an ITS when $\mathbf{z}^*(x)$ is dropped.

4.2. Semiparametric Estimation

Various ad hoc approaches might be used to estimate the parameter $\boldsymbol{\eta}$ in $t_{\mathcal{Z}}(x; \boldsymbol{\eta})$ while leaving the distribution G in (20) unspecified. An ITS is one for which (2) holds, i.e., T is independent of $\mathcal{Z}$, so we could, for example, stratify the covariate paths $\mathcal{Z}_i$ $(i = 1, \ldots, n)$ into K groups that are homogeneous in some sense. If we let $\hat{S}_j(t; \boldsymbol{\eta})$ denote the empirical survivor function (the Kaplan-Meier estimate, assuming some units have censored failure times) then we might choose $\boldsymbol{\eta}$ to minimize some measure of disparity among the K distributions $\hat{S}_1(t; \boldsymbol{\eta}), \ldots, \hat{S}_K(t; \boldsymbol{\eta})$. The feasibility of this approach depends on our ability to form the K groups in an effective way and, in particular, so that the disparity among the $\hat{S}_j(t; \boldsymbol{\eta})$'s varies significantly as $\boldsymbol{\eta}$ varies.

Another possibility is to select $\boldsymbol{\eta}$ to make sample covariances between some function of values $t_{\mathcal{Z}_i}(x_i; \boldsymbol{\eta})$ and some function of $\mathcal{Z}_i$ equal to zero. A third semiparametric procedure proposed by Kordonsky and Gertsbakh (1993, 1995a, 1995b, 1997a) for linear time scales is to choose $\boldsymbol{\eta}$ so as to minimize the sample CV for the values $t_i = t_{\mathcal{Z}_i}(x_i; \boldsymbol{\eta})$, $i = 1, \ldots, n$; we consider this approach in Section 5 and find its usefulness to be quite limited.

The estimation of $\boldsymbol{\eta}$ can also be viewed as a semiparametric regression problem with X_i as response and $\mathcal{Z}_i$ as covariate. The model (20) is, however, rather nonstandard. If we instead consider (again in the absence of censoring) the response as $T_i = t_{\mathcal{Z}_i}(X_i; \boldsymbol{\eta})$ and note that we can without loss of generality force T_i to have mean 1, then we might estimate $\boldsymbol{\eta}$ by least squares. That is, we

$$\text{minimize} \quad \sum_{i=1}^{n} \left\{t_{\mathcal{Z}_i}(x_i; \boldsymbol{\eta}) - 1\right\}^2 \tag{25}$$

subject to

$$\frac{1}{n}\sum_{i=1}^{n} t_{\mathcal{Z}_i}(x_i;\boldsymbol{\eta}) = 1. \tag{26}$$

We could alternatively assume a specified mean for some monotonic function of T_i and minimize its variance; for example, force $\log t_{\mathcal{Z}_i}(X_i;\boldsymbol{\eta})$ to have mean 0 and minimize

$$\sum_{i=1}^{n}\left\{\log t_{\mathcal{Z}_i}(X_i;\boldsymbol{\eta})\right\}^2. \tag{27}$$

Least squares and minimum CV methods do not adapt easily to censored data. Another approach, better able to handle censored data, would be via rank regression methods (e.g., Robins and Tsiatis, 1992, Lin and Ying, 1995.) Duchesne and Lawless (1999) develop this approach.

The computation of estimates via such methods can be complex, and the efficiency of some methods is likely to be poor; considerable study is needed. We provide brief examples of estimation in Section 6.

We conclude this section with a remark on model assessment. Oakes (1995) and Section 4.1 mention model expansion as a means of model assessment. Informal checks based on the examination of generalized residuals $\hat{t}_i = t_{\mathcal{Z}_i}(x_i;\hat{\boldsymbol{\eta}})$ can also be employed. In particular, the $\hat{t}_i$'s should be roughly independent of functions of the covariate paths $\mathcal{Z}_i$; this can be used in both semiparametric and parametric settings, though as for all types of residual analysis, study of specific models is needed to assess effectiveness and the extent to which the $\hat{t}_i$'s mimic a random sample of $t_i's$. Distributional assumptions about G in (20) can be assessed through residuals, but this is likely to be effective only for models which are fitted parametrically.

5. Small Variation and Time Scale Selection

Some authors have suggested that one select a time scale that gives a compact distribution, or small variation, for failure time T. There is no obvious connection between this concept and that of ideal time scales. Indeed, any monotone increasing function of T is an ITS if T is, whereas measures of variation or relative variation are not invariant under general transformations.

Nevertheless, the concept of small variation is widely considered by practitioners in areas such as reliability. The main approach to time scale selection using this concept is due to Kordonsky and Gertsbakh (1993, 1995ab, 1997a); they suggest that if a parametric family of time scales $t_{\mathcal{P}}(x) = \Phi[\mathbf{y}_{\mathcal{P}}(x);\boldsymbol{\eta}]$ is being considered, then $\boldsymbol{\eta}$ be chosen so as to minimize the squared coefficient of variation of $T = t_{\mathcal{P}}(X)$, that is, $\mathrm{Var}(T)/\mathrm{E}^2(T)$. This is used in practice by considering complete (uncensored) samples in which chronological times x_i and corresponding usage measures $\mathbf{y}_i(x_i) = (y_{i0}(x_i) = x_i,\ y_{i1}(x_i),\ldots,y_{ip}(x_i))$ are observed for units $i = 1,\ldots,n$. Then the operational times $t_i(\boldsymbol{\eta}) = \Phi[\mathbf{y}_i(x_i);\boldsymbol{\eta}]$ are considered and

$\boldsymbol{\eta}$ is estimated by minimizing

$$\widehat{CV}^2[t_i(\boldsymbol{\eta})] = \frac{\widehat{\mathrm{Var}}[t_i(\boldsymbol{\eta})]}{\widehat{\mathrm{E}}^2[t_i(\boldsymbol{\eta})]}, \tag{28}$$

where $\widehat{\mathrm{Var}}$ and $\widehat{\mathrm{E}}$ denote sample variance and sample mean, respectively.

Kordonsky and Gertsbakh apply this method to linear time scale families $t_{\mathcal{P}}(x) = \boldsymbol{\eta}^t \mathbf{y}(x)$. In this case one easily obtains that the unrestricted minimum of (28) is obtained at

$$\widetilde{\boldsymbol{\eta}} \propto \widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\mu}}, \tag{29}$$

where $Y_{ij} = y_{ij}(X_i)$, $\widehat{\boldsymbol{\Sigma}}$ is the sample covariance matrix for $(X_i, Y_{i1}, \ldots, Y_{ip})^t$ and $\widehat{\boldsymbol{\mu}}$ is the sample mean vector for $(X_i, Y_{i1}, \ldots, Y_{ip})^t$. The CV is invariant under scale changes to T, and so any $\boldsymbol{\eta}$ proportional to the right side of (29) may be used. Kordonsky and Gertsbakh assume that $\boldsymbol{\eta} \geq \mathbf{0}$ and it may happen that the right side of (29) has negative elements. In the two-dimensional case Kordonsky and Gertsbakh select either $\boldsymbol{\eta}^t = (1,\ 0)$ or $\boldsymbol{\eta}^t = (0,\ 1)$ (corresponding to $T = X$ or $T = Y$, respectively), according to whichever gives the smaller $\widehat{CV}^2$ for $T(\boldsymbol{\eta})$.

It is not clear how this approach relates to those discussed in Sections 3 and 4. Note in addition that the minimum CV method may be used without any information on the path $\mathcal{P}(x)$ except for the endpoint $(X_i, Y_{i1}, \ldots, Y_{ip})$ and that it seemingly applies to situations where the paths are randomly determined, rather than fixed.

The theorem below is new and it identifies a setting in which the minimum CV method gives consistent estimates. We consider situations where the usage paths $\{\mathbf{y}(x),\ x \geq 0\}$ may be represented by a finite-dimensional parameter $\boldsymbol{\theta}$, in which case we write the ideal time scales as $t_{\mathcal{P}}(x) = \Phi[x, \boldsymbol{\theta}]$, where $\boldsymbol{\theta} = \boldsymbol{\theta}(\mathcal{P})$. This assumption is clearly too restrictive for many settings, but provides a reasonable approximation to reality in some. So far, proofs of validity of the minimum CV under more general conditions are elusive; we note that Kordonsky and Gertsbakh provide no proof of validity at all.

THEOREM 5.1 *Consider models for which* $Pr\,[X > x|\mathcal{P}] = G(\Phi[x, \boldsymbol{\theta}; \boldsymbol{\eta}_0])$, *where* $\boldsymbol{\theta}$ *is a random finite-dimensional parameter that completely identifies* $\mathcal{P}$ *and where the function* Φ *has the form*

$$\Phi[x, \boldsymbol{\theta}; \boldsymbol{\eta}] = u(x)v(\boldsymbol{\theta}; \boldsymbol{\eta}), \tag{30}$$

where $\boldsymbol{\eta}$ *is a finite-dimensional parameter. Also denote* $t(\boldsymbol{\eta}) = u(x)v(\boldsymbol{\eta}, \boldsymbol{\theta})$ *and* $T(\boldsymbol{\eta}) = u(X)v(\boldsymbol{\eta}, \boldsymbol{\Theta})$, *so that* $t(\boldsymbol{\eta}_0)$ *is an ITS. Then for any* $\boldsymbol{\eta}$ *we have*

$$CV^2[T(\boldsymbol{\eta})] \geq CV^2[T(\boldsymbol{\eta}_0)],$$

and if $CV^2[T(\boldsymbol{\eta})] = CV^2[T(\boldsymbol{\eta}_0)]$ *then* $t(\boldsymbol{\eta})$ *is also an ITS.*

Proof: Let $\mu_T = \mathrm{E}[T(\boldsymbol{\eta}_0)] = \mathrm{E}[T(\boldsymbol{\eta}_0)|\boldsymbol{\Theta} = \boldsymbol{\theta}]$ and $\sigma_T^2 = \mathrm{Var}[T(\boldsymbol{\eta}_0)] = \mathrm{Var}[T(\boldsymbol{\eta}_0)|\boldsymbol{\Theta} = \boldsymbol{\theta}]$; μ_t and σ_T^2 do not depend on $\boldsymbol{\theta}$ because of (2). Then

$$\begin{aligned}
CV^2[T(\boldsymbol{\eta})] &= \mathrm{Var}[T(\boldsymbol{\eta})]/\mathrm{E}^2[T(\boldsymbol{\eta})] \\
&= \frac{\mathrm{Var}_{\boldsymbol{\Theta}}[\mathrm{E}[T(\boldsymbol{\eta})|\boldsymbol{\Theta}]] + \mathrm{E}_{\boldsymbol{\Theta}}[\mathrm{Var}[T(\boldsymbol{\eta})|\boldsymbol{\Theta}]]}{\mathrm{E}^2_{\boldsymbol{\Theta}}[\mathrm{E}[T(\boldsymbol{\eta})|\boldsymbol{\Theta}]]} \\
&= \left\{\mathrm{Var}_{\boldsymbol{\Theta}}\left[\frac{v(\boldsymbol{\eta},\boldsymbol{\Theta})}{v(\boldsymbol{\eta}_0,\boldsymbol{\Theta})}\mathrm{E}[u(X)v(\boldsymbol{\eta}_0,\boldsymbol{\Theta})|\boldsymbol{\Theta}]\right]\right. \\
&\qquad \left. + \mathrm{E}_{\boldsymbol{\Theta}}\left[\left(\frac{v(\boldsymbol{\eta},\boldsymbol{\Theta})}{v(\boldsymbol{\eta}_0,\boldsymbol{\Theta})}\right)^2 \mathrm{Var}[u(X)v(\boldsymbol{\eta}_0,\boldsymbol{\Theta})|\boldsymbol{\Theta}]\right]\right\} \\
&\quad \times \mathrm{E}^{-2}_{\boldsymbol{\Theta}}\left[\frac{v(\boldsymbol{\eta},\boldsymbol{\Theta})}{v(\boldsymbol{\eta}_0,\boldsymbol{\Theta})}\mathrm{E}[u(X)v(\boldsymbol{\eta}_0,\boldsymbol{\Theta})|\boldsymbol{\Theta}]\right] \\
&= \frac{\mu_T^2\mathrm{Var}_{\boldsymbol{\Theta}}\left[\frac{v(\boldsymbol{\eta},\boldsymbol{\Theta})}{v(\boldsymbol{\eta}_0,\boldsymbol{\Theta})}\right] + \sigma_T^2\mathrm{E}_{\boldsymbol{\Theta}}\left[\left(\frac{v(\boldsymbol{\eta},\boldsymbol{\Theta})}{v(\boldsymbol{\eta}_0,\boldsymbol{\Theta})}\right)^2\right]}{\mu_T^2\mathrm{E}^2_{\boldsymbol{\Theta}}\left[\frac{v(\boldsymbol{\eta},\boldsymbol{\Theta})}{v(\boldsymbol{\eta}_0,\boldsymbol{\Theta})}\right]} \\
\Rightarrow CV^2[T(\boldsymbol{\eta})] &\geq \sigma_T^2/\mu_T^2 = CV^2[T(\boldsymbol{\eta}_0)].
\end{aligned}$$

In addition, equality can hold only if $v(\boldsymbol{\eta},\boldsymbol{\theta})/v(\boldsymbol{\eta}_0,\boldsymbol{\theta})$ is constant with respect to $\boldsymbol{\theta}$. (We suppose the distribution of $\boldsymbol{\Theta}$ is such that $\mathrm{Var}[v(\boldsymbol{\eta},\boldsymbol{\theta})/v(\boldsymbol{\eta}_0,\boldsymbol{\theta})] > 0$ implies this.) But in this case $t(\boldsymbol{\eta}) = ct(\boldsymbol{\eta}_0)$, where c does not depend on $\boldsymbol{\theta}$, which is also an ITS. ■

Theorem 5.1 thus shows that we can identify ITS's within the family (30) by finding $\boldsymbol{\eta}$'s that minimize $CV^2[T(\boldsymbol{\eta})]$. This coversseveral important situations. Suppose for instance that there is a single usage factor $y(x)$ and that usage paths $\mathcal{P}$ are of the form

$$y(x) = \theta x, \quad x \geq 0, \tag{31}$$

where θ has some distribution in the population of interest. Linear time scales as in model (21) can in this case be expressed as

$$t_{\mathcal{P}}(x) = \eta_1 x + \eta_2 y(x) = x(\eta_1 + \eta_2\theta),$$

which is of the form (30). Because CV and ITS's are both invariant to scale changes, we can consider the single parameter form $T(\eta) = X + \eta Y$. Minimization of $CV^2[T(\eta)]$ yields the unique value

$$\eta = \frac{\mu_Y\sigma_X^2 - \mu_X\sigma_{XY}}{\mu_X\sigma_Y^2 - \mu_Y\sigma_{XY}}, \tag{32}$$

which is consistent with (29).

To estimate the parameter $\boldsymbol{\eta}$ in an ITS model of the form (30) we can minimize the sample coefficient of variation (28) for a complete (uncensored) sample $t_1(\boldsymbol{\eta}), \ldots, t_n(\boldsymbol{\eta})$.

When $\boldsymbol{\eta}_0$ is unique, assuming finiteness and smoothness of $\mathrm{E}[T(\boldsymbol{\eta})]$ and $\mathrm{Var}[T(\boldsymbol{\eta})]$ in the neighborhood of $\boldsymbol{\eta}_0$ will ensure that $\widetilde{\boldsymbol{\eta}}$ minimizing (28) is a consistent estimator for $\boldsymbol{\eta}_0$.

There are some subtleties concerning minimum CV estimation. For example, if we assume $\eta_0 \geq 0$, under the parameterization $T(\eta) = X + \eta Y$ the right side of (32) is non-negative. However, the estimate $\tilde{\eta}$ based on minimization of (28) is obtained by replacing σ_X^2, σ_Y^2, σ_{XY}, μ_X and μ_Y in (32) with sample-based estimates, and it can be negative. In particular, it may be shown that $\tilde{\eta} < 0$ if either of the following two conditions holds:

$$\text{A: } \frac{\hat{\sigma}_X^2}{\hat{\mu}_X^2} < \frac{\hat{\sigma}_{XY}}{\hat{\mu}_X \hat{\mu}_Y} < \frac{\hat{\sigma}_Y^2}{\hat{\mu}_Y^2}$$

$$\text{B: } \frac{\hat{\sigma}_Y^2}{\hat{\mu}_Y^2} < \frac{\hat{\sigma}_{XY}}{\hat{\mu}_X \hat{\mu}_Y} < \frac{\hat{\sigma}_X^2}{\hat{\mu}_X^2}.$$

If case A holds then $\widehat{CV}^2[t_i(\eta)]$ is minimized subject to $\eta \geq 0$ at $\eta = 0$ and if B holds then it is minimized at $\eta = \infty$, thus leading to the choice of X or Y, respectively, as the ideal time scale. As Kordonsky and Gertsbakh note, cases A and B are automatically ruled out if $\hat{\sigma}_{XY} < 0$.

If $\widehat{CV}^2[T(\boldsymbol{\eta})]$ can be shown to be minimized at $\boldsymbol{\eta} = \boldsymbol{\eta}_0$ for other time scale models, then under mild smoothness assumptions we can estimate $\boldsymbol{\eta}_0$ consistently as here. However, as noted, we have at present a proof only for models of the form (30). A more detailed discussion of minimum CV and other semiparametric methods will be given elsewhere. The efficiency of minimum CV is considered in an example in Section 6.

Semiparametric estimation of ITS parameters $\boldsymbol{\eta}$ without specifying G in (20) was discussed in Section 4.2, where it was suggested that the independence of T and covariate path $\mathcal{Z}$ (or usage path $\mathcal{P}$) for an ITS be exploited. Theorem 5.1 holds because of this independence and it is this rather than the unconditional variability in $T(\boldsymbol{\eta})$ (i.e. variability in $T(\boldsymbol{\eta})$ when both X and $\mathcal{P}$ are considered random) that seems crucial.

6. Examples

We consider a pair of real examples which illustrate features of time scale selection.

6.1. Automobile Reliability

Lawless et al. (1995) considered the time to failure in automobile systems as a function of age (chronological time since sale) x and cumulative mileage $y(x)$, and fitted models based on warranty data. They considered a model where (20) is of the form

$$\Pr[X > x|\mathcal{P}] = G[x^{1-\eta} y(x)^{\eta}; \phi]. \tag{33}$$

They actually obtained this model by assuming that for a given automobile i we have $y_i(x) = \alpha_i x$, with α_i as a fixed covariate. They considered accelerated failure time models where $\Pr[X > x|\alpha_i]$ was of the form $G[x\alpha_i^{\eta}]$; this gives (33) since $\alpha_i = y_i(x)/x$. The

linearity assumption is widely used with automobiles, but (33) can be used more generally. It has the nice feature that if $\eta = 0$, then age x is the ITS and if $\eta = 1$ then mileage $y(x)$ is.

Several authors have considered linear time scales, as mentioned earlier. An alternative to (33) could be to consider

$$\Pr\,[X > x|\mathcal{P}] = G[x + \eta y(x); \phi]. \tag{34}$$

Either of models (33) or (34) can be fitted by maximum likelihood via (23), if we take G to have a specific parametric form; Lawless et al. (1995) assumed G to be Weibull. Oakes (1995) and Kordonsky and Gertsbakh (1997a) have also fitted Weibull models in conjunction with (34).

With $y(x) = \alpha x$, models (33) and (34) meet the requirements of Theorem 5.1; semi-parametric estimation via minimum CV should therefore yield consistent estimators of the parameter η. To assess minimum CV and compare it with parametric maximum likelihood and an ad hoc least squares approach, we performed some simulations based on the models fitted by Lawless et al. (1995). Although they had censored data, we consider only complete samples here because minimum CV handles only this case at present.

Model (33)

We simulated 1000 samples of size 100 from the model with (33) given by

$$\Pr\,[X > x|\alpha] = \exp\left\{-(x\alpha^{\eta}/\phi_1)^{\phi_2}\right\}, \tag{35}$$

where $\phi_1 = 60$ and $\phi_2 = 1$, and where $\log\alpha$ is normally distributed with mean $\mu = 2.37$ and standard deviation $\sigma = 0.58$. These parameters are selected to match the estimates in Lawless et al. (1995), and thus to be physically plausible. The units associated with these parameter values are years, x, and thousands of miles, $y(x)$. We repeated this procedure for three values of η_0, namely 0.1, 0.5 and 0.9. (For the data in Lawless et al. (1995) a value of η near 0.9 was indicated). For each sample, we computed the maximum likelihood estimates (MLE's) of the parameters under the Weibull model along with the minimum CV and least squares (see Lawless, 1982, page 331) estimates of η. We first computed the estimators of η with no constraint on their value. However, since an ITS must be non-decreasing in all the usage measures, η must be in the interval [0, 1]. We thus also obtained estimates of η constrained to [0, 1]. This was discussed for minimum CV in Section 5 and for least squares and maximum likelihood, we used the constrained optimization function `nlminb` in S-Plus.

Table 1 shows simulation means and variances of the three estimators of η for both the constrained and unconstrained versions. In both cases, the estimators have reasonably small bias, though it is interesting that the constrained estimates have slightly larger bias for the cases $\eta_0 = 0.1$ and $\eta_0 = 0.9$. However, the minimum CV estimator is less efficient than the least squares (LS) estimator, with an efficiency relative to maximum likelihood of

Table 1. Summary of simulations, model (33).

$\hat{\eta}$ unconstrained						
	$\eta_0 = 0.1$		$\eta_0 = 0.5$		$\eta_0 = 0.9$	
	$\widehat{\mathrm{E}}[\hat{\eta}]$	$\widehat{\mathrm{Var}}[\hat{\eta}]$	$\widehat{\mathrm{E}}[\hat{\eta}]$	$\widehat{\mathrm{Var}}[\hat{\eta}]$	$\widehat{\mathrm{E}}[\hat{\eta}]$	$\widehat{\mathrm{Var}}[\hat{\eta}]$
MLE	0.095	0.034	0.50	0.031	0.89	0.034
min CV	0.087	0.065	0.50	0.058	0.88	0.059
least squares	0.10	0.052	0.50	0.045	0.90	0.056
$\hat{\eta}$ constrained to [0,1]						
	$\eta_0 = 0.1$		$\eta_0 = 0.5$		$\eta_0 = 0.9$	
	$\widehat{\mathrm{E}}[\hat{\eta}]$	$\widehat{\mathrm{Var}}[\hat{\eta}]$	$\widehat{\mathrm{E}}[\hat{\eta}]$	$\widehat{\mathrm{Var}}[\hat{\eta}]$	$\widehat{\mathrm{E}}[\hat{\eta}]$	$\widehat{\mathrm{Var}}[\hat{\eta}]$
MLE	0.12	0.017	0.50	0.032	0.87	0.019
min CV	0.15	0.029	0.50	0.053	0.84	0.033
least squares	0.14	0.025	0.50	0.051	0.86	0.026

around 57% for the former compared to about 65% for the latter. If one considers a log-normal distribution instead of a Weibull in (35), the minimum CV estimator again would not outperform the LS estimator, the latter being the MLE in this setup. The LS procedure is thus as good or better than the minimum CV method for two popular lifetime distributions under model (33).

Model (34)

Several authors (e.g., Oakes, 1995; Kordonsky and Gertsbakh, 1997a) consider models of the linear form (34) when looking for an ITS. Advantages of such models include an easy interpretation of the parameter η and collapsibility. However, there is an intrinsic difficulty with the parameterization in (34). The parameter η is in many situations confounded with a scale parameter in G, and cannot be estimated precisely. This is easily seen when we write (34) in this form:

$$\Pr\,[X > x|\mathcal{P}] = G\left(\frac{x[1+\eta\alpha]}{\phi_1}; \phi_2\right), \tag{36}$$

where ϕ_1 is a scale parameter and ϕ_2 is a vector of other parameters. When $\eta\alpha$ is large compared to one, the model for $\Pr\,[X > x|\mathcal{P}]$ becomes approximately $G[x(\eta/\phi_1)\alpha; \phi_2]$, and we can only accurately estimate $\psi = \eta/\phi_1$ and ϕ_2. If there is not sufficient variability in the values of the slopes (α_i's) in the sample or if the variation in X given α is too large, inferences about the time scale parameter η may be very imprecise. Note also that in (36), η is not invariant to changes in the units of y, whereas in (35) it is. Restricting η to be in [0,1] by using the parameterization $t(\eta) = (1-\eta)x + \eta y$ does not eliminate these problems.

We simulated samples of size 100 from model (34), with

$$\Pr\,[X > x|\alpha] = \exp\left\{-[x(1+\eta\alpha)/\phi_1]^{\phi_2}\right\}, \tag{37}$$

where $\eta = 840$, $\phi_1 = 63800$ and $\phi_2 = 1$, so that the values of $\Pr\,[X > x|\alpha]$ obtained were close to the ones from the model (35) with $\eta = 0.9$, $\phi_1 = 60$ and $\phi_2 = 1$. As expected,

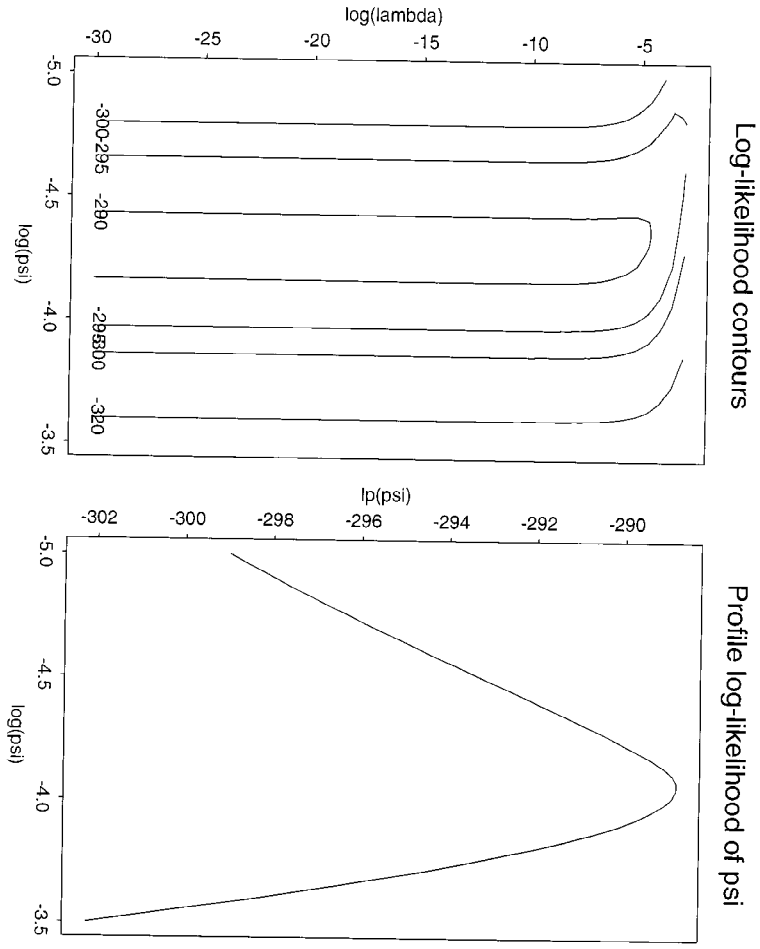

Figure 1. Contours of the log-likelihood evaluated at $\phi_2 = 1$ and profile log-likelihood of ψ for a typical sample from model (38).

it was nearly impossible to get precise estimates of η and ϕ_1 for samples generated from model (37) and the previous distribution of α, and both the MLE and minimum CV estimates varied greatly from sample to sample. Estimation in (37) is clarified if we use the alternate parameterization

$$\Pr\left[X > x|\alpha\right] = \exp\left\{-[x(\lambda + \psi\alpha)]^{\phi_2}\right\}, \tag{38}$$

where $\psi = \eta/\phi_1$ and $\lambda = 1/\phi_1$. In this case we can estimate ψ accurately. However, inferences about λ remain imprecise. Figure 1 shows the profile log-likelihood contours in ψ and λ as well as the log-likelihood profile in ψ for a single sample of size 100.

Table 2. Linear scale parameter estimates, Kordonsky and Gertsbakh (1995a) data.

Method	Estimate	95% confidence interval
MLE, Weibull	6.61	(5.33, 8.89)
MLE, lognormal	6.97	(5.20, 9.67)
Quasi-likelihood	6.91	(5.33, 9.75)
Rank	6.59	(5.40, 10.11)
Minimum CV	6.77	(4.99, 8.77)

Simulations with very large sample sizes support the consistency of the minimum CV^2 estimator $\tilde{\eta}$, but for samples of size 100 the enormous variability of both $\tilde{\eta}$ and the MLE make efficiency comparisons of little value. However, the difficulties discussed here do not always arise with linear time scale models. When the paths $\mathcal{P}$ in the sample are sufficiently variable and when the variation in X given $\mathcal{P}$ is sufficiently small, precise estimation of η is possible. The next example illustrates this, and considers interval estimation and model checking.

6.2. Fatigue Life of Steel Specimens

Kordonsky and Gertsbakh (1995a) gave data on the lifetime of steel specimens subjected to cyclic loading. Thirty specimens were divided into six groups of size 5, and the loading program for each group was a repeated sequence of high stress and low stress cycles. Kordonsky and Gertsbakh used the cumulative number of low stress cycles as their "real" time x, and the number of high stress cycles as their usage measure $y(x)$. Because of the large number of cycles applied and the loading programs, the relationship between x and $y(x)$ was well approximated by a straight line through the origin; slopes of the approximating lines varied from about 0.05 to 18. This large variability in slopes makes the estimation of a linear time scale relatively precise in spite of the small sample size. There were no censored observations.

Kordonsky and Gertsbakh (1995a) fit the linear TS model $t(\eta) = x + \eta y(x)$ by the minimum CV method, to get $\tilde{\eta}$=6.77. They did not consider interval estimation or the adequacy of the linear TS. We consider estimation of η in model (34) using four other methods, for comparison: (i) and (ii) maximum likelihood (ML) under Weibull and log-normal specifications for G in (34), (iii) a quasi-likelihood method which is similar to least squares, and (iv) a rank based method. Methods (iii) and (iv) are discussed in Duchesne and Lawless (1999).

Table 2 shows the estimates and approximate 95% confidence intervals. The confidence intervals for ML are based on the likelihood ratio statistic, those for the quasi-likelihood and rank based methods are based on variance estimates given by Duchesne and Lawless (1999), and that for minimum CV is based on the bootstrap percentile method (Efron and Tibshirani, 1993, Chapter 13). The estimates are quite precise and confidence intervals from the different methods are in good agreement. As one might expect, they indicate the importance of the number of high stress cycles.

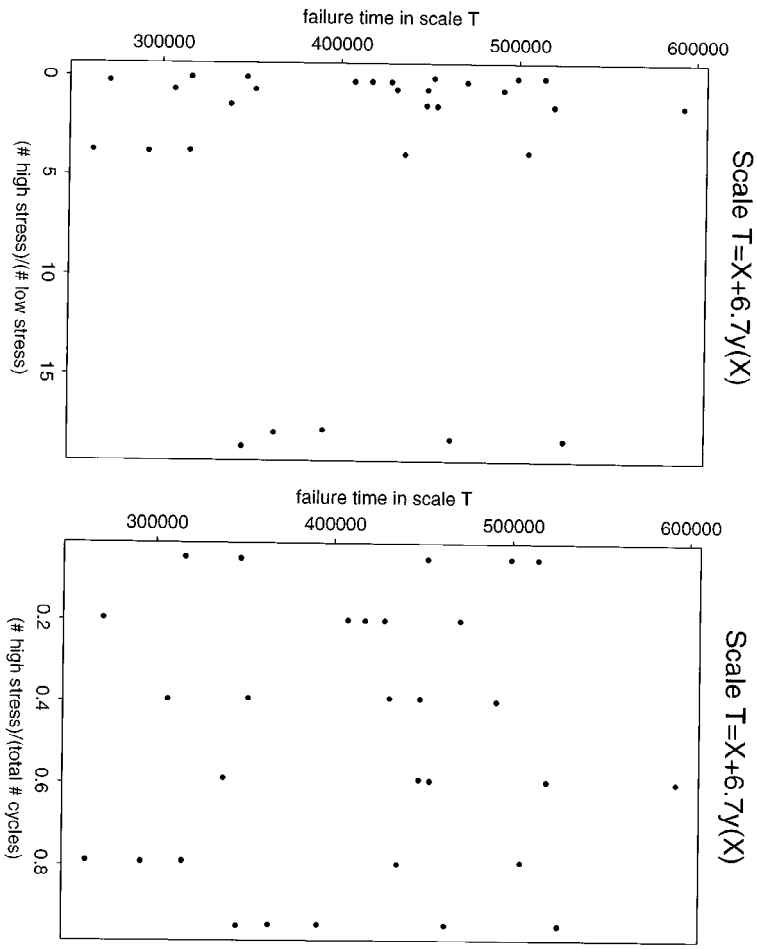

Figure 2. Failure times versus path features, scale $T = X + 6.7y(X)$.

For informal model checks, we use the fact that failure times $T_i(\eta_0)$ are independent of the usage path if the time scale is ideal. We therefore plot values $t_i(\hat{\eta})$ versus features of the usage paths. For example, Figure 2 shows a plot of the $t_i(\hat{\eta})$'s using the minimum CV estimate $\hat{\eta} = 6.77$ versus the ratio of high to low stress cycles and the proportion of high stress cycles in the specimen's loading program. This and other checks do not suggest any inadequacies in the model (34). We remark, on the other hand, that such plots do provide evidence against multiplicative scale models of the form (33) for this data set.

Probability plots of the $t_i(\hat{\eta})$'s can be used to check on specific functional forms for G in (34), such as the Weibull. However, they are not able to detect non-ideal time scales very

well; for that purpose plots as in Figure 2 are crucial. For the current data set, Weibull and log-normal probability plots suggest either model is consonant with the data, provided a linear TS is used.

7. Conclusion

The investigation of alternative time scales raises interesting problems. Key ideas are that we seek scales which "capture" most of the variation in failure times, given a set of time-varying covariates, and that the time scales be fairly easy to interpret. This led us to consider ideal time scales and models of the form (20), in which $t_{\mathcal{Z}}(x)$ is specified in terms of a parameter vector $\boldsymbol{\eta}$. As Section 6 indicates, care may be needed in the parameterization and fitting of such models.

The concept of small variation in operational failure times T is very useful but harder to pin down. The essential problem is one of identifying time-varying covariates which, when known and conditioned upon, make (conditional) variation in the chronological time of failure X small. There is no obvious general validity to approaches that seek to minimize unconditional relative variation in T, though there is a connection with conditional variation of X given $\mathcal{Z}$ in some cases, as shown in Section 5. Further study is needed.

Many methodological issues associated with time-scale models of the form (20) require further investigation, particularly ones associated with semiparametric estimation. Duchesne and Lawless (1999) investigate some rank based and quasi-likelihood methods.

This paper has excluded internal time-varying covariates, but there are clearly many settings where we might want to consider them. For example, in assessing the survival time or disease occurrence for persons infected with the human immunodeficiency virus (HIV) it is natural to investigate "biological" time scales which include information about variation in an individual's immune function across time. Similarly, in planning maintenance actions for a system or piece of equipment it is desirable to consider internal covariates measuring deterioration or degradation of the system; this is referred to as "condition-based" maintenance.

If the covariate history $\mathcal{Z}(x)$ has internal components we need to consider the joint distribution of X and $\mathcal{Z}$, and in particular $\Pr[X > x,\ \mathcal{Z}(x)]$. This can be written in product integral form in the absolutely continuous case as

$$\Pr[X > x,\ \mathcal{Z}(x)] = \prod_{(0,x]} \{1 - h(u|\mathcal{Z}(u))du\} \Pr[d\mathcal{Z}(u)|\mathcal{Z}(u),\ X > u],$$

where $h(u|\mathcal{Z}(u))$ is as previously the hazard function for X at u, conditional on the covariate history $\mathcal{Z}(u)$. This gives

$$\Pr[X > x,\ \mathcal{Z}(x)] = \exp\left\{-\int_0^x h(u|\mathcal{Z}(u))\,du\right\} \prod_{(0,x]} \Pr[d\mathcal{Z}(u)|\mathcal{Z}(u),\ X > u].$$

It is tempting to consider monotone functions of

$$t_{\mathcal{Z}}(u) = \int_0^x h(u|\mathcal{Z}(u))\,du \tag{39}$$

as operational time scales, but it should be noted that such scales are not ITS's in the sense of (2), and (1) does not hold. We hope to discuss these issues and time scales based on (39) in a future communication.

Finally, there are essentially no rigorous studies of time scale performance for reliability areas such as maintenance and replacement, in spite of the fact that composite time scales are often used (e.g., Kordonsky and Gertsbakh, 1997b). Work in this area would be welcome.

Acknowledgments

The authors wish to thank the Natural Sciences and Engineering Research Council of Canada for their Postgraduate Scholarship (TD) and Research Grant support (JL), and General Motors of Canada for their research support (JL). They also thank Richard Cook, Jack Kalbfleisch, Jock MacKay and three anonymous reviewers for useful comments.

References

V. B. Bagdonavičius and M. S. Nikulin, "Transfer functionals and semiparametric regression models," *Biometrika* vol. 84 pp. 365–378, 1997.

N. E. Breslow, J. H. Lubin, P. Marek, and B. Langholz, "Multiplicative models and cohort analysis," *Journal of the American Statistical Association* vol. 78 pp. 1–12, 1983.

E. Cinlar and S. Ozekici, "Reliability of complex devices in random environments," *Probability in the Engineering and Informational Sciences* vol. 1 pp. 97–115, 1987.

D. R. Cox, "Regression models and life-tables (with discussion)," *Journal of the Royal Statistical Society, series B* vol. 34 pp. 187–220, 1972.

D. R. Cox and D. Oakes, *Analysis of Survival Data*. Chapman and Hall: London, 1984.

K. A. Doksum and A. Hóyland, "Models for variable-stress accelerated life testing experiments based on Wiener processes and the inverse gaussian distribution," *Technometrics* vol. 34 pp. 83–91, 1992.

T. Duchesne and J. F. Lawless, "Semiparametric inference methods for alternative time scale models," unpublished manuscript, 1999.

B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. Chapman and Hall: New York, 1993.

V. T. Farewell and D. R. Cox, "A note on multiple time scales in life testing," *Applied Statistics* vol. 28 pp. 73–75, 1979.

D. A. Follmann, "Modelling failures of intermittently used machines," *Applied Statistics* vol. 39 pp. 115–124, 1990.

J. J. Gaynor, "The use of time dependent covariates in modelling data from an occupational cohort study," *Applied Statistics* vol. 36 pp. 340–351, 1987.

N. P. Jewell and J. D. Kalbfleisch, "Marker processes in survival analysis," *Lifetime Data Analysis* vol. 2 pp. 15–29, 1996.

J. D. Kalbfleisch and R. L. Prentice, *The Statistical Analysis of Failure Time Data*. Wiley: New York, 1980.

K. B. Kordonsky and I. Gertsbakh, "Choice of the best time scale for system reliability analysis," *European Journal of Operational Research* vol. 65 pp. 235–246, 1993.

K. B. Kordonsky and I. Gertsbakh, "System state monitoring and lifetime scales–I," *Reliability Engineering and System Safety* vol. 47 pp. 1–14, 1995a.

K. B. Kordonsky and I. Gertsbakh, "System state monitoring and lifetime scales–II," *Reliability Engineering and System Safety* vol. 49 pp. 145–154, 1995b.

K. B. Kordonsky and I. Gertsbakh, "Multiple time scales and the lifetime coefficient of variation: engineering applications," *Lifetime Data Analysis* vol. 2 pp. 139–156, 1997a.

K. B. Kordonsky and I. Gertsbakh, "Choice of the best time scale for preventive maintenance in heterogeneous environments," *European Journal of Operations Research* vol. 98 pp. 64-74, 1997b.

J. F. Lawless, *Statistical Models and Methods for Lifetime Data*. Wiley: New York, 1982.

J. F. Lawless, J. Hu, and J. Cao, "Methods for the estimation of failure distributions and rates from automobile warranty data," *Lifetime Data Analysis* vol. 1 pp. 227–240, 1995.

D. Y. Lin and Z. Ying, "Semiparametric inference for the accelerated life model with time-dependent covariates," *Journal of Statistical Planning and Inference* vol. 44 pp. 47–63, 1995.

N. A. Miner, "Cumulative damage in fatigue," *Journal of Applied Mechanics* vol. 12 pp. A159–164, 1945.

D. N. P. Murthy, B. P. Iskandar, and R. J. Wilson, "Two dimensional failure free warranty policies: two dimensional point process models," *Operations Research* vol. 43 pp. 356–366, 1995.

W. Nelson, *Accelerated Testing: Statistical Models, Test Plans and Data Analyses*. Wiley: New York, 1990.

D. Oakes, "Multiple time scales in survival analysis," *Lifetime Data Analysis* vol. 1 pp. 7–18, 1995.

J. Robins and A. A. Tsiatis, "Semiparametric estimation of an accelerated failure time model with time-dependent covariates," *Biometrika* vol. 79 pp. 311–319, 1992.

N. D. Singpurwalla, "Survival in dynamic environments," *Statistical Science* vol. 10 pp. 86–103, 1995.

N. D. Singpurwalla and S. Wilson, "The warranty problem: its statistical and game theoretic aspects," *SIAM Review* vol. 35 pp. 17–42, 1993.