Lifetime Data Analysis, 8, 263–276, 2002

Semiparametric Inference Methods for General Time Scale Models

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

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

Received May 23, 2000; Revised February 5, 2001; Accepted February 22, 2001

Abstract. In this paper we consider semiparametric inference methods for the time scale parameters in general time scale models (Oakes, 1995; Duchesne and Lawless, 2000). We use the results of Robins and Tsiatis (1992) and Lin and Ying (1995) to derive a rank-based estimator that is more efficient and robust than the traditional minimum coefficient of variation (min CV) estimator of Kordonsky and Gerstbakh (1993) for many underlying models. Moreover, our estimator can readily handle censored samples, which is not the case with the min CV method.

Keywords: accelerated failure time model, collapsible model, generalized residuals, ideal time scale, minimum coefficient of variation, linear rank estimator, separable scale model, unbiased estimating function

1. Introduction

In many survival analysis applications, modeling failure using a scale other than real (chronological) time may be more appropriate than using real time at failure as the time variable. For example, cumulative mileage may be a better measure of the age of a car than real time since purchase. More generally, composite scales that are based on real time and one or more measures of usage, stress or exposure may be considered. Kordonsky and Gertsbakh (1993), Oakes (1995) and Duchesne and Lawless (2000) consider this type of problem and they investigate the concept of *load invariant* or *ideal* time scale. Let x represent a fixed value of real (chronological) time and let $\mathbf{z}(x)$ represent the value of usage or exposure[1] measures at time x. Define $\mathcal{P}(x) = \{\mathbf{z}(u), 0 \leq u < x\}$, the usage path or history up to time x. Then an ideal time scale (ITS) is a function $\Phi[\cdot, \cdot]$ of x and $\mathcal{P}(x)$ such that the conditional survivor function of the real time at failure, X, given the whole usage history, $\mathcal{P} = \lim_{x\to\infty}\mathcal{P}(x)$, can be written as

$$\Pr[X > x|\mathcal{P}] = G(\Phi[x, \mathcal{P}(x)]), \tag{1}$$

where $G(\cdot)$ is a positive, 1−1, decreasing function. As stated, (1) merely says that an ITS is a one-to-one function of the survivor function of X given $\mathcal{P}$, and is not unique. The usefulness of (1) comes from the modeling choices that it suggests. Both $\Phi[\cdot, \cdot]$ and $G(\cdot)$ can be fully specified, depend on a vector of parameters, or one of $\Phi[\cdot, \cdot]$ or $G(\cdot)$ can be left arbitrary.

Modeling through (1) allows us to emphasize scales $\Phi[x, \mathcal{P}(x)]$ that are simple or easily interpreted. A class of models that emphasizes simple time scales has been suggested by Oakes (1995). These are termed *collapsible models* and are such that

$$\Pr[X > x|\mathcal{P}] = H[x, \mathbf{z}(x)], \tag{2}$$

where $\mathbf{z}(x)$ is the value of the usage or exposure measures in $\mathcal{P}$ at time x, and $H[\cdot, \cdot]$ is some decreasing function of x and the elements of $\mathbf{z}(x)$. It is easily seen that (2) is a special case of (1), where $\Phi[x, \mathcal{P}(x)] \equiv \Phi[x, \mathbf{z}(x)]$, i.e., the conditional survivor function only depends on the endpoint of the usage path $\mathcal{P}(x)$:

$$\Pr[X > x|\mathcal{P}] = G(\Phi[x, \mathbf{z}(x)]). \tag{3}$$

Examples of collapsible models seen in the literature (with only one usage factor) include

$$\Pr[X > x|\mathcal{P}] = G((1-\eta)x + \eta z(x))$$
$$\text{and} \quad \Pr[X > x|\mathcal{P}] = G(x^{1-\eta} z(x)^{\eta}),$$

where $z(x)$ are monotone functions of x.

Lin and Ying (1995) and others consider a time-scale model where

$$\Phi[x, \mathcal{P}(x); \beta] = \int_0^x \exp\{\beta^t \mathrm{z}(u)\}\, du, \tag{4}$$

where $\mathbf{z}(u)$ is the value of all the covariates in $\mathcal{P}$ at time u, and β is a vector of unknown parameters. This model is known as an *accelerated failure time model* (Cox and Oakes, 1984; Robins and Tsiatis, 1992; Bagdonavičius and Nikulin, 1997a). It is a natural model in some applications, but it is not generally collapsible. Note that the covariates, $\mathbf{z}(x)$, need not be monotone with this model.

Models where $G(\cdot)$ and $\Phi[\cdot, \cdot]$ in (1) are both specified parametrically can be handled by straightforward maximum likelihood (e.g., Oakes, 1995; Duchesne and Lawless, 2000). Our objective in this paper is to develop semiparametric inference for the parameters of a time scale, $\Phi[\cdot, \cdot]$, when $G(\cdot)$ is left unspecified. This case has yet to be thoroughly investigated in the literature. Kordonsky and Gertsbakh (1993, 1995a,b, 1997) proposed the *minimum coefficient of variation* (min CV) method, which consists of choosing the value of the time scale parameter that minimizes the squared sample coefficient of variation of the failure times in scale $\Phi[\cdot, \cdot]$. More precisely, consider collapsible models and a sample of n uncensored observations of the form $(x_i, \mathcal{P}_i), i = 1, 2, \ldots, n$, and assume that

$$\Pr[X > x|\mathcal{P}] = G(\Phi[x, \mathbf{z}(x); \eta]), \tag{5}$$

where η is a time scale parameter (vector) to be estimated from data. Let $t_i(\eta) = \Phi[x_i, \mathbf{z}_i(x_i); \eta]$ and $\overline{t(\eta)} = \sum t_i(\eta)/n$, where x_i is the observed real failure time. Then the min

CV estimator of η is given by

$$\hat{\eta}_{CV} = \arg\min_{\eta} \frac{\left(\sum_{i=1}^{n} t_i^2(\eta) - n\overline{t(\eta)}^2\right)/(n-1)}{\overline{t(\eta)}^2}.$$

Duchesne and Lawless (2000) show that this estimator is consistent for a special type of collapsible model, but more generally, the properties of the method are not known. Another practical limitation of the min CV estimator is that it cannot readily handle censoring.

In this paper, we propose a semiparametric method for inference about the time scale parameter η in (5), and study it and the min CV method. In Section 2, we use the approach of Robins and Tsiatis (1992) and Lin and Ying (1995) to provide an estimation procedure. We study its performance in finite samples and compare it to that of the min CV method through Monte Carlo simulations in Sections 3 and 4. An illustration and a discussion of model assessment are given in Section 5. Concluding remarks are given in Section 6.

2. Rank-Based Estimation Method

We wish to estimate η without specifying $G(\cdot)$ in

$$\Pr[X > x|\mathcal{P}] = G[t_{\mathcal{P}}(x;\eta)], \tag{6}$$

where $t_{\mathcal{P}}(x;\eta) = \Phi[x, \mathbf{z}(x);\eta]$ and $G(\cdot)$ is a survivor function. We assume that for a given x, (6) depends on $\mathbf{z}(s)$ only for $s \leq x$. It is convenient to introduce some further notation. Let X_i^*, $i = 1, \ldots, n$ denote a sequence of n independent failure times. Under right-censorship, we observe $X_i = \min(X_i^*, C_i)$ and $\delta_i = I[X_i^* \leq C_i]$, $i = 1, \ldots, n$, where C_i is the censoring time for the ith item and $I[\cdot]$ is the indicator function. For each item, we also observe $\mathcal{P}_i = \{\mathbf{z}_i(x), x \geq 0\}$; for convenience, we assume that $\mathbf{z}(0) = 0$. Finally, following Lin and Ying (1995), we make the usual assumption that given $\mathcal{P}_i$, X_i^* and C_i are independent, $i = 1, \ldots, n$, i.e., that censoring is uninformative.

2.1. Semiparametric Estimating Function for η

Under model (6), the log-likelihood for η is given by

$$l(\eta) = \sum_{i=1}^{n} \int_0^\infty Y_i(x)\Big(\ln\, \lambda[t_{\mathcal{P}_i}(x;\eta)] + \ln\, t'_{\mathcal{P}_i}(x;\eta)\Big) \Big(dN_i(x) - \lambda[t_{\mathcal{P}_i}(x;\eta)]t'_{\mathcal{P}_i}(x;\eta)dx\Big), \tag{7}$$

where $\lambda[u]$ is the hazard function corresponding to $G(u)$, $N_i(u)$ is a counting process taking value 0 if $u < X_i^*$ and 1 if $u \geq X_i{}^*$, $Y_i(u)$ is 1 if individual i is at risk of failing at real time u, 0 otherwise, and $t'_{\mathcal{P}_i}(x;\eta) = dt_{\mathcal{P}_i}(x;\eta)/dx$. From here on, we assume that $t_{\mathcal{P}}(x;\eta)$ is smooth in x and η and that it is strictly increasing in x. Moreover, we make the assumption that for each $\mathcal{P}$ and each η, the mapping

$$t_{\mathcal{P}}(\cdot;\eta) : \mathbb{R}^+ \mapsto \mathbb{R}^+$$

$$x \to t_{\mathcal{P}}(x;\eta)$$

is one-to-one, with inverse $t_{\mathcal{P}}^{-1}(\cdot;\eta)$, i.e

$$t_{\mathcal{P}}(x;\eta) = t \Leftrightarrow t_{\mathcal{P}}^{-1}(t;\eta) = x.$$

When $\lambda[\cdot]$ is specified parametrically, the likelihood score $\partial l(\eta)/\partial\eta$ is an unbiased estimating function.

In order to estimate η semiparametrically, we use a linear rank estimating function, generalized so as to allow for time-varying covariates. This method is discussed in detail in Robins and Tsiatis (1992), who use it to estimate the parameters of an accelerated failure time model.

First, define $\tilde{N}_i(t;\eta) = I[t_{\mathcal{P}_i}(X_i;\eta) \le t, \delta = 1]$ and $\tilde{Y}_i(t;\eta) = I[t_{\mathcal{P}_i}(X_i;\eta) \ge t]$ and let

$$d\hat{\Lambda}(t;\eta) = \frac{\sum_{j=1}^n \tilde{Y}_j(t;\eta) d\tilde{N}_j(t;\eta)}{\sum_{j=1}^n \tilde{Y}_j(t;\eta)}.$$

To obtain an estimator, say $\hat{\eta}$, of η we use the estimating function

$$\begin{aligned} \tilde{U}(\eta) &= \sum_{i=1}^n \int_0^\infty \tilde{Y}_i(t;\eta) Q(\mathcal{P}_i, t, \eta) \left(d\tilde{N}_i(t;\eta) - d\hat{\Lambda}(t;\eta) \right) \\ &= \sum_{i=1}^n \int_0^\infty \tilde{Y}_i(t;\eta) \left(Q[\mathcal{P}_i, t, \eta] - \frac{\sum_{j=1}^n \tilde{Y}_j(t;\eta) Q[\mathcal{P}_j, t, \eta]}{\sum_{j=1}^n \tilde{Y}_j(t;\eta)} \right) d\tilde{N}_i(t;\eta) \\ &= \sum_{i=1}^n \delta_i \left(Q[\mathcal{P}_i, t_{\mathcal{P}_i}, (x_i;\eta), \eta] - \overline{Q_i} \right), \end{aligned} \tag{8}$$

where the weight $Q[\mathcal{P}_i, t, \eta]$ is a function of the usage path for item i, of the time in scale $t_{\mathcal{P}}(\cdot;\eta)$ and of η, and $\overline{Q_i}$ is the average of the $Q[\mathcal{P}_j, t_{\mathcal{P}_i}(x_j;\eta), \eta]$'s of the individuals still at risk when (in scale $t_{\mathcal{P}}(\cdot;\eta)$) item i fails. In the Appendix we follow the arguments of Robins and Tsiatis (1992) to show that the score (8) is an unbiased estimating function, i.e., $E[\tilde{U}(\eta_0)] = 0$, where η_0 is the true value of η.

We would like to choose a weighting function Q in (8) that will yield an estimator of η that is as efficient as possible. Because (8) is an unbiased estimating function, it will yield consistent estimators for general Q. However, no choice of Q is uniformly optimal for every survivor function G in (6). Robins and Tsiatis (1992) derive the optimal weighting function under an exponential survivor function. In the Appendix, we use their results to derive the following weighting function, which is optimal under $G(t) = \exp(-\lambda t)$:

$$Q[\mathcal{P}, t, \eta^*] = \frac{\partial}{\partial\eta} \ln\, t'_{\mathcal{P}}(t_{\mathcal{P}}^{-1}(t;\eta^*);\eta)\Big|_{\eta=\eta^*} \tag{9}$$

We choose the exponential distribution here as it is the only continuous distribution with a constant hazard function, which we need in order to obtain a weighting function that does not involve any value of the hazard function.

Piecing together equations (8) and (9), we obtain the following "optimal" linear rank-type estimating function, i.e., an estimating function that may be used to obtain estimates of η when G is arbitrary, but is optimal when G is the exponential survivor function:

$$\begin{aligned}\tilde{U}_{opt}(\eta^*) &= \sum_{i=1}^{n} \int_0^\infty \tilde{Y}_i(t;\eta^*) \left[\frac{\partial}{\partial\eta} \ln\, t'_{\mathcal{P}_i}(x;\eta)\right]\Bigg|_{\substack{\eta=\eta^* \\ x=t_{\mathcal{P}_i}^{-1}(t;\eta^*)}} \\ &\quad \times\left\{d\widetilde{N}_i(t;\eta^*) - d\hat{\Lambda}(t;\eta^*)\right\} \\ &= \sum_{i=1}^{n} \delta_i \left(\frac{\partial}{\partial\eta} \ln t'_{\mathcal{P}_i}(x_i;\eta)\Big|_{\eta=\eta^*} - \overline{Q_i}\right),\end{aligned} \tag{10}$$

where $\overline{Q_i}$ is the average of the $\partial \ln t'_{\mathcal{P}}/\partial\eta$'s of the individuals still at risk when (in scale $t_{\mathcal{P}}(\cdot;\eta^*)$) individual i fails. Note that under the accelerated failure time model for time-varying covariates, i.e., with $t_p(\cdot;\eta)$ as in (4), we obtain the same score as Robins and Tsiatis (1992) and Lin and Ying (1995).

Because estimating functions of the form (8) or (10) are not continuous nor monotone in η in general, we find the estimator not by solving $\tilde{U}(\eta) = 0$, but by minimizing the length of $\tilde{U}$, i.e., the estimator of η is defined by

$$\hat{\eta} = \arg\min_{\eta\in\mathcal{N}} \tilde{U}(\eta)^t \tilde{U}(\eta), \tag{11}$$

where $\mathcal{N}$ is a compact subset of $\mathbb{R}^q, q = \dim(\eta)$. When η is unidimensional, this can be done with the golden section search algorithm (Press et al., 1992). When η is of higher dimension, Lin and Ying (1995) propose to use the simulated annealing algorithm described in Lin and Geyer (1992).

Robins and Tsiatis (1992) argue that $\sqrt{n}(\eta-\eta_0)$ is asymptotically normal with mean vector **0** in general settings while Lin and Ying (1995) show that $\hat{\eta}$ is strongly consistent under the accelerated failure time model.

To illustrate this method, let us consider the linear scale model $t_{\mathcal{P}}(\mathrm{x};\eta) = (1\eta)x + \eta z(x)$ (Kordonsky and Gertsbakh, 1993, 1995a,b and 1997; and Oakes, 1995). For smooth $\mathbf{z}(x)$

this is also an accelerated failure time model; just replace $\exp\{\beta^t \mathbf{z}(u)\}$ with $1 - \eta + \eta z'(u)$ in equation (4), where $z'(x) = dz(x)/dx$. Under this model, (10) becomes

$$\tilde{U}_{opt}(\eta) = \sum_{i=1}^{n} \delta_i \left(\frac{z'(x_i) - 1}{1 - \eta + \eta z'(x_i)} - \overline{Q_i} \right), \tag{12}$$

where $\overline{Q_i}$ is the average of the $(z'(x) - 1)/(1 - \eta + \eta z'(x))$'s still at risk when (in scale $(1 - \eta)x + \eta z(x)$) individual i fails. In order to compute the score (12) for a fixed value of η, we need the value of everyone's covariate $z(\cdot)$ and its derivative $z'(\cdot)$ at every failure time $(1 - \eta)x + \eta z(x)$. Thus, to minimize the square of (12) with respect to η, we need to observe $z(x)$ in continuous time for every individual. In many applications the usage measures $z(x)$ are parameterized (for example $z(x) = \theta x$ as in Lawless et al., 1995; Oakes, 1995; or Kordonsky and Gertsbakh, 1997), or $z'(x)$ is piecewise constant, as in Lin and Ying (1995), so the information needed is available.

2.2. Variance of $\tilde{U}(\eta)$ and Inference About η

Robins and Tsiatis (1992) argue that estimating functions of the form (8) are consistent and asymptotically multivariate normal with mean vector **0** and variance matrix that can be consistently estimated by $n^{-1}V(\eta)$, with

$$V(\eta) = \sum_{i=0}^{n} \int_0^{\infty} \tilde{Y}_i(t;\eta) \left[\frac{\sum_{j=1}^{n} \tilde{Y}_j(t;\eta) Q^{\otimes 2}(\mathcal{P}_j, t, \eta)}{\sum_{j=1}^{n} \tilde{Y}_j(t;\eta)} - \left(\frac{\sum_{j=1}^{n} \tilde{Y}_j(t;\eta) Q(\mathcal{P}_j, t, \eta)}{\sum_{j=1}^{n} \tilde{Y}_j(t;\eta)} \right)^{\otimes 2} \right] d\tilde{N}_i(t;\eta), \tag{13}$$

where the Q's are as defined earlier and, for any vector $\boldsymbol{u}$, $\boldsymbol{u}^{\otimes 2}$ denotes the outer product $\boldsymbol{u}\boldsymbol{u}^t$.

From this result, we use the following approximate $(1 - \alpha)100\%$ confidence regions for η, proposed by Lin and Ying (and indirectly by Robins and Tsiatis (1992)):

$$\{\eta : \tilde{U}(\eta)^t V^{-1}(\eta) \tilde{U}(\eta) \le \chi^2_{q;1-\alpha}\}, \tag{14}$$

where q is the number of parameters in η and $\chi^2_{q;1-\alpha}$ is the $(1 - \alpha)$th quantile of a chi-square distribution on q degrees of freedom.

Suppose we wish to test if one of the parameters of η, say η_1, is equal to some fixed value η_0. Let η_{-1} be the parameter vector without η_1. Following Lin and Ying (1995), we use the following quadratic form:

$$G(\eta_0) = \inf_{\substack{\eta_1 = \eta_0 \\ \eta_{-1}}} \tilde{U}(\eta)^t V^{-1}(\eta) \tilde{U}(\eta), \tag{15}$$

which is asymptotically chi-square with one degree of freedom under the hypothesis $\eta_1 = \eta_0$. We can get confidence intervals for η_1 of a form similar to that of (14) simply by inverting the quadratic form (15).

Lin and Ying (1995) suggest the use of $V^{-1}(\hat{\eta})$ rather than $V^{-1}(\eta)$ in (14). These two options are asymptotically equivalent; however, through Monte Carlo simulations (discussed in Duchesne, 1999), we found that the coverage of intervals of the form (14) was more stable, especially in cases where $\hat{\eta}$ was more variable.

Estimating function (10) apparently yields an estimator of η that is consistent and asymptotically normal for any distribution $G(\cdot)$ and time scale function $t_{\mathcal{P}}(\cdot;\eta)$ that are not too ill behaved, and this estimator has minimum asymptotic variance within the class (8) when $G(\cdot)$ is the exponential distribution. In the next section we assess normality and confidence interval coverage through a small simulation study.

3. Simulation Study of the Proposed Estimator

According to Robins and Tsiatis (1992), estimators obtained from (8) will be consistent and asymptotically normal. However, they do not give precise conditions on the form of the usage histories, $\mathcal{P}_i$, the form of the time scale function, $t_{\mathcal{P}_i}(x;\eta)$, or the distribution function, $G(\cdot)$, for these asymptotic results to be valid. Lin and Ying (1995) do give precise conditions for the validity of their asymptotic results, but their conclusions are restricted to the case of the accelerated failure time model (4) with $Q[\mathcal{P}_i, t, \eta] = z_i(t_{\mathcal{P}_i}^{-1}[t;\eta])$. Asymptotic analysis of semiparametric estimators for accelerated failure time models under other particular conditions can also be found (see for example Bagdonavičius and Nikulin (1997b) and Bordes (1999)), and a rigorous treatment of (10) can presumably be given. We consider instead some empirical investigation of the methodology in finite samples.

In this section, we examine the behavior of η as defined in the previous section and we look at the coverage of the confidence intervals defined by (14) through a simulation study based on some simple collapsible models. The first two models studied are part of the family of *separable scale models* (Duchesne and Lawless, 2000). Let us suppose that the usage paths are completely described by some vector of parameters, say θ. For example, items used at constant rates have usage paths of the form $\mathcal{P} = \{\theta x, x \geq 0\}$, and the value of θ entirely specifies $\mathcal{P}$. A separable scale model is of the form

$$\Pr[X > x|\mathcal{P}] = \Pr[X > x|\theta] = G[u(x)v(\theta;\eta)], \tag{16}$$

where $u(\cdot)$ is a non-negative, increasing function and $v(\cdot,\cdot)$ is positive-valued.

The first model that we study is the linear scale model where $y(x) = \theta x$ and $t_{\mathcal{P}}(x;\eta) = (1-\eta)x + \eta y(x)$, which is of the separable scale form with $u(x) = x$ and $v(\theta;\eta) = (1-\eta) + \eta\theta$. The second model is the multiplicative scale model with $y(x) = \theta x$ and $t_{\mathcal{P}}(x;\eta) = x^{1-\eta}y(x)^{\eta}$; this is a separable scale model with $u(x) = x$ and $v(\theta;\eta) = \theta^{\eta}$. Finally, we consider a model where $y(x) = x^{\theta}$ and $t_{\mathcal{P}}(x;\eta) = (1-\eta)x + \eta y(x)$; this is not a separable scale model. The validity of the estimator and of the confidence intervals has been studied for various choices of distribution by Duchesne (1999). Here we report on the

Table 1. Simulation results for the rank-type estimator (11) under different models. For each model, we give a sample average and standard deviation of the estimator, along with a proportion of coverage of the true value of the time scale parameter by nominal 95% confidence intervals. True value of η is 0.5.

	Linear scale			Mult. scale			Non-separable scale		
Cens. %	$\bar{\hat{\eta}}$	$S_{\hat{\eta}}$	cover.	$\bar{\hat{\eta}}$	$S_{\hat{\eta}}$	cover.	$\bar{\hat{\eta}}$	$S_{\hat{\eta}}$	cover.
0%	0.500	0.032	94.6%	0.501	0.022	94.9%	0.499	0.022	94.9%
20%	0.503	0.043	94.8%	0.502	0.032	94.9%	0.500	0.029	95.0%
60%	0.515	0.10	95.0%	0.505	0.054	94.2%	0.587	0.19	95.6%

results obtained with a Weibull survivor function with shape parameter value 3 and scale parameter value 1000; other choices of distributions affect the efficiency of the estimator, but not the coverage proportion of the intervals very much. For the separable scale models, the path parameters, θ, were simulated from atan(Θ) $\sim$ uniform(0, π/2), and in the non-separable scale case, we simulated from $\Pr[\Theta = \theta] = 1/3$, $\theta = 1/2$, 1 and 2, for computational reasons[2]. Censoring times (in the x scale) were generated independently from a normal distribution chosen so as to obtain specified censoring proportions. We used the golden section search algorithm (Press et al., 1992) to solve our estimating equations, as described in Section 2.1. We report on the simulations done with $\eta = 0.5$, the results not varying much for different values of η (see discussion in Duchesne (1999) and Duchesne and Lawless (2000)). The results of these simulations are summarized in Table 1. Averages and standard deviations of estimators are based on 2000 simulations of samples of size 100 and coverage proportions of 95% confidence intervals derived from result (14) are based on 10,000 samples of size 100. In all the simulations reported, the parameter η and its estimators are restricted to the range [0, 1].

Except in one case, the estimated bias of the estimators was small, and the coverage of the intervals based on (14) was very good for all the models at every censoring level. The larger bias and large standard deviation of $\hat{\eta}$ in the non-separable scale model with 60% censoring are due to an asymmetric distribution for $\hat{\eta}$ in that setting and, in particular, a discrete probability mass at $\hat{\eta} = 1$.

4. Comparison of Rank and min CV Estimators

We now compare the min CV and rank-based estimators of the time scale parameter through a simulation study based on the three models of Section 3. For each model, we simulated data using different survivor functions $G(\cdot)$ and different distributions for the usage paths. The distributions were chosen so as to reflect the practical settings of Kordonsky and Gertsbakh (1995a) and Lawless, Hu and Cao (1995), and also to see the impact of variations in the form and the variability of the distributions of Θ and $X|\Theta$ on the distribution of the estimators.

For each model we generated failure times in the ideal time scale from three different survivor functions: G_1 is Weibull with shape 3 and scale 1000, G_2 is Weibull with shape 1

Table 2. Efficiency (relative to maximum likelihood) for the min CV and rank-based estimators based on 2000 samples of size 100. True value of η is 0.5.

	Failure time distribution					
Path distribution	G_1		G_2		G_3	
	Linear scale model $t_{\mathcal{P}}(x;\eta) = x(1-\eta+\eta\theta)$					
	Rank	min CV	Rank	min CV	Rank	min CV
F_1	97.6%	99.8%	97.5%	63.3%	80.4%	60.6%
F_2	97.4%	99.7%	121%	92.1%	80.1%	59.6%
	Multiplicative scale model $t_{\mathcal{P}}(x;\eta) = x\theta^{\eta}$					
	Rank	min CV	Rank	min CV	Rank	min CV
F_1	98.9%	99.7%	98.9%	54.7%	80.4%	61.5%
F_2	97.5%	99.8%	81.7%	73.6%	79.5%	59.4%
	Model $t_{\mathcal{P}}(x;\eta) = (1-\eta)x + \eta x^{\theta}$					
	Rank	min CV	Rank	min CV	Rank	min CV
F_3	96.0%	98.5%	96.1%	62.3%	78.9%	59.2%

and scale 1000, and G_3 is log normal with location 6.7 and scale 0.35. Note that the mean and variance of G_1 and G_3 are equal. For the first two models, the path parameter, Θ, is generated using two distributions: F_1 represents atan(Θ) $\sim$ uniform(0, $\pi/2$) and F_2 represents atan(Θ) $\sim$ uniform($\pi/5$, $2\pi/5$). For the numerical considerations mentioned in Section 3, the path parameter for the third model is only generated using Pr $[\Theta = \theta] = 1/3$, $\theta = 1/2$, 1 and 2, which we denote as distribution F_3. For each model, we generated 2000 samples of size 100. The results are summarized in Table 2, where the min CV and rank estimators are compared with parametric maximum likelihood based on the true model.

From Table 2, the min CV estimator is virtually as good as the maximum likelihood estimator when the survivor function is Weibull with shape parameter 3. However, its efficiency is poorer in the other cases, and is less than that of the rank-based estimator.

For the two models generated from the distributions (G_2, F_2), none of the three estimators (MLE, rank, min CV) could accurately estimate the time scale parameter. The distributions of the estimators had probability masses at $\eta = 1.0$ and hence the efficiency comparisons under (G_2, F_2) reported in Table 2 are of limited interest. For these probability models, the variability in the usage paths is small and is coupled with a large variability in the distribution of $X|\Theta$, making it very difficult to distinguish between the "survivor function effect" and the "time scale effect" with samples of size 100.

5. An Example

Duchesne and Lawless (2000) consider data on the fatigue life of 30 steel specimens, provided by Kordonsky and Gertsbakh (1995a). The specimens were subjected to

Table 3. Time scale parameter estimates(95% confidence intervals), Kordonsky and Gertsbakh (1995) data.

Method	Linear scale	Mult. scale 1	Mult. scale 2
MLE, Weibull[1]	0.868(0.842, 0.899)	0.800(0.715, 0.885)	0.539(0.460, 0.632)
MLE, lognormal[1]	0.875(0.839, 0.906)	0.789(0.724, 0.854)	0.555(0.466, 0.643)
Rank	0.868(0.844, 0.910)	0.800(0.662, 0.930)	0.538(0.450, 0.693)
Minimum CV[2]	0.871(0.833, 0.898)	0.804(0.686, 0.896)	0.547(0.469, 0.643)

[1] Confidence intervals: likelihood ratio.
[2] Confidence intervals: bootstrap percentile.

alternating periods of low and high stress cycles, until a defined failure occured. Duchesne and Lawless followed Kordonsky and Gertsbakh (1995a) in letting x represent the cumulative number of low stress cycles and defining $y(x)$ as the cumulative number of high stress cycles corresponding to x. To a close approximation, $y_i(x) = \theta_i x$ for the ith specimen ($i = 1, \ldots, 30$).

Duchesne and Lawless (2000) fit linear collapsible models with $\Pr[X > x|\mathcal{P}] = G(x + \eta y)$. We consider the same model under the alternative parameterization with $t = (1 - \eta)x + \eta y$, where $0 \leq \eta \leq 1$, and in addition we consider two multiplicative scales of the form $t = x^{1-\eta}y(x)^{\eta}$. We let x be the cumulative number of low stress cycles in the first case and we let x be the total number of cycles in the second case. In both cases, we let $y(x)$ be the cumulative number of high stress cycles at x.

Estimates of η and 95% confidence intervals are given for all three time scale models using the four different inference methods in Table 3. For the three models considered, the inference methods agree well and seem to suggest that both the number of low and high stress cycles are important, but with a greater weight given to the number of high stress cycles. The agreement between the rank-based and other estimates is poorer for the multiplicative scales which, as we discuss below, do not appear ideal.

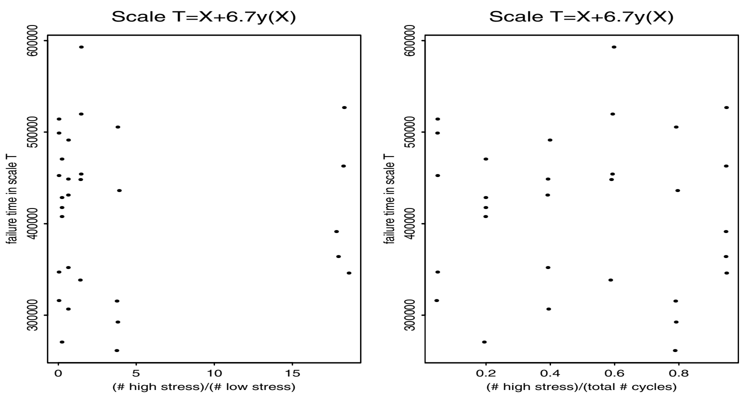

Figure 1. Failure times versus path features, scale $t = (1 - \eta)x + \eta y(x)$.

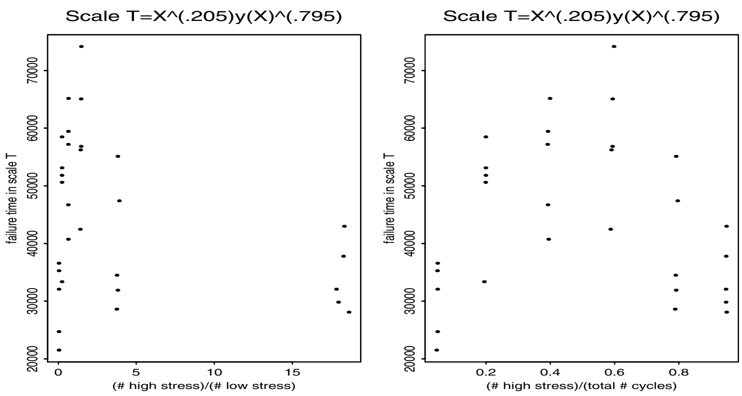

Figure 2. Failure times versus path features, scale $t = x^{1-\eta}y(x)^{\eta}$, x cumulative number of low stress cycles.

To find which of the three scales is more appropriate, we look at generalized residual plots as described by Duchesne and Lawless (2000). The estimated failure times $t_i(\hat{\eta})$ in a time scale are plotted against path features; if the time scale is ideal, no systematic departures from a horizontal band should be seen.

Figures 1 and 2 show plots for the linear TS and the first multiplicative TS. There does not seem to be any trend in the plots of Figure 1. However, the plots in Figure 2 (especially the right hand panel) show a quadratic trend, where the items with extreme usage rates tend to fail earlier than items with average usage rates. A similar pattern is seen for the second multiplicative TS. These plots thus indicate no problem with the ITS assumption for the linear time scale, but problems with the assumption that the multiplicative scales are ideal.

6. Conclusion

We have proposed a semiparametric estimator of the time scale parameter in models of the form (5), based on an approach suggested by Robins and Tsiatis (1992). This estimator performs well in simulations under models considered in Sections 3 and 4. We are able to derive confidence intervals for the time scale parameter, handle censoring, and the estimator seems more efficient than the min CV estimator in most of the models considered.

A rigorous examination of asymptotic properties for the rank-type and the min CV estimators has not been undertaken. Conditions on the form of the survivor function, $G(\cdot)$, the ideal time scale, $t_{\mathcal{P}}(x;\eta)$, and the usage histories, $\mathcal{P}$, for consistency and asymptotic normality of $\hat{\eta}$ would be of interest.

Finally, calculation of the Q_i's in the estimating function (8) requires a lot of information. In general, the value of the covariates, $\mathbf{z}_i(x)$, and their derivative, $\mathbf{z}_i'(x)$, are needed for every item at every failure time in the scale $t_{\mathcal{P}}(x;\eta)$ in order to compute the value of $\tilde{U}(\eta)$. This virtually means that the values of covariates along with their derivatives need to be observed in continuous time, unless the covariate paths are defined parametrically.

These conditions may be difficult to obtain in many settings. Methods of approximating $\tilde{U}(\eta)$ are needed in such cases.

Appendix A

A.1 Unbiasedness of (8)

Without loss of generality, we can rewrite (8) as

$$\tilde{U}(\eta) = \sum_{i=1}^{r} \Big(Q[\mathcal{P}_{(i)}, t_{\mathcal{P}_{(i)}}(X_{(i)};\eta), \eta] - \overline{Q_{(i)}} \Big),$$

where $r = \sum_{i=1}^{n} \delta_i$ is the number of observed failures, (i) is the label of the ith individual to fail in scale $t_{\mathcal{P}}(\cdot;\eta)$ and

$$\overline{Q_{(i)}} = \frac{\sum_{j=1}^{n} \tilde{Y}_j[t_{\mathcal{P}_{(i)}}(X_{(i)};\eta);\eta] Q[\mathcal{P}_j, t_{\mathcal{P}_{(i)}}(X_{(i)};\eta), \eta]}{\sum_{j=1}^{n} \tilde{Y}_j[t_{\mathcal{P}_{(i)}}(X_{(i)};\eta);\eta]}$$

is the average of the Q's of individuals still at risk when (in scale $t_{\mathcal{P}}(\cdot;\eta)$) individual (i) fails.

Now let us write $\tilde{U}(\eta) = \sum_{i=1}^{r} \tilde{U}_{(i)}(\eta)$ and prove that $E[\tilde{U}_{(i)}(\eta_0)]$ is zero. Let $T_{(i)}$ represent the time of the ith failure in scale $t_{\mathcal{P}}(\cdot;\eta_0)$ and $\mathcal{F}_t$ denote the risk set just prior to time t in scale $t_{\mathcal{P}}(\cdot;\eta_0)$ and all the usage histories. Then

$$\begin{aligned}
\mathrm{E}[\tilde{U}_{(i)}(\eta_0)] &= \mathrm{E}\big\{\mathrm{E}\big[\tilde{U}_{(i)}(\eta_0)|T_{(i)} = t, \mathcal{F}_t\big]\big\} \\
&= \mathrm{E}\big\{\mathrm{E}[Q[\mathcal{P}_{(i)}, T_{(i)}, \eta_0] - \overline{Q_{(i)}}|T_{(i)} = t, \mathcal{F}_t]\big\} \\
&= \mathrm{E}\big\{\mathrm{E}[Q[\mathcal{P}_{(i)}, t, \eta_0]|T_{(i)} = t, \mathcal{F}_t] - \overline{Q_{(i)}}\big\},
\end{aligned}$$

the last equality holding because we can calculate the value of the Q's of everybody still at risk at time t (and hence, their average) given $\mathcal{F}_t$. This implies that

$$\mathrm{E}\left[Q[\mathcal{P}_{(i)}, t, \eta_0]|T_{(i)} = t, \mathcal{F}_t\right] = \sum_{j=1}^{n} \tilde{Y}_j[t;\eta_0] Q[\mathcal{P}_j, t, \eta_0] \mathrm{Pr}[(i) = j].$$

Since $t_{\mathcal{P}}(\cdot;\eta_0)$ is an ITS, every individual still at risk at time t is equally likely to be the one that fails at that time. Hence,

$$\begin{aligned}
\mathrm{E}[Q[\mathcal{P}_{(i)}, t, \eta_0]|T_{(i)} = t, \mathcal{F}_t] &= \sum_{j=1}^{n} \frac{\tilde{Y}_j[t;\eta_0] Q[\mathcal{P}_j, t, \eta_0]}{\sum_{j=1}^{n} \tilde{Y}_j[t;\eta_0]} \\
&= \overline{Q_{(i)}},
\end{aligned}$$

which implies that $\mathrm{E}\big[\tilde{U}(\eta_0)\big] = 0$.

A.2 Derivation of (9)

Substituting our notation in Proposition A.1 of Robins and Tsiatis (1992), we get that the optimal choice for Q in (8) is

$$Q_{opt}[\mathcal{P}_i, t, \eta^*] \;\propto\; - \left[\frac{\partial \ln\ \lambda_0[t]}{\partial t} \frac{\partial t_{\mathcal{P}_i}(w_i; \eta)}{\partial \eta} \Bigg|_{\eta = \eta^*} + \frac{\partial}{\partial \eta}\ \ln\ \frac{\partial}{\partial w_i} t_{\mathcal{P}_i}(w_i; \eta) \Bigg|_{\eta = \eta^*} \right]$$

where $w_i = t_{\mathcal{P}_i}^{-1}(t; \eta^*)$ and "$\propto$" means "proportional to". Thus, when $G(\cdot)$ is the exponential survivor function, $\lambda_0[t] = \lambda$ and we get that

$$Q_{opt}[\mathcal{P}_i, t, \eta^*] \propto \frac{\partial}{\partial \eta}\ \ln\ t'_{\mathcal{P}_i}(t_{\mathcal{P}_i}^{-1}(t; \eta^*); \eta) \Bigg|_{\eta = \eta^*},$$

which yields (9).

Acknowledgments

The authors wish to thank the Natural Sciences and Engineering Research Council of Canada for their Research Grant support, and General Motors of Canada for their research support (JL). The authors are also thankful to Richard Cook, Jack Kalbfleisch, Jock MacKay, an Associate Editor and two anonymous referees for their helpful comments.

Notes

1. Usage, stress or exposure measures are assumed to be left-continuous, external time-varying covariates.
2. With these values of θ, we can get a closed form expression for $x = t_{\mathcal{P}}^{-1}(t; \eta)$ and thus do not have to resort to numerical methods to calculate x for given t, θ and η.

References

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

V. B. Bagdonavičius and M. S. Nikulin, "Asymptotical analysis of semiparametric models in survival analysis and accelerated life testing," *Statistics* vol. 29 pp. 261–283, 1997b.

L. Bordes, "Semiparametric additive accelerated life models," *Scandinavian Journal of Statistics* vol. 26 pp. 345–361, 1999.

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

T. Duchesne, "Multiple Time Scales in Survival Analysis," Doctoral dissertation: University of Waterloo, 1999.

T. Duchesne and J. Lawless, "Alternative time scales and failure time models," *Lifetime Data Analysis* vol. 6 pp. 157–179, 2000.

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, 1997.

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. 227–239, 1995.

D. Y. Lin and C. J. Geyer, "Computational methods for semiparametric linear regression with censored data," *Journal of Computational and Graphical Statistics* vol. 1 pp. 77–89, 1992.

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.

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

W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press: Cambridge, 1992.

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.