Lifetime Data Analysis, 1, 227–240 (1995)

Methods for the Estimation of Failure Distributions and Rates from Automobile Warranty Data

JERRY LAWLESS
Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

JOAN HU
Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

JIN CAO
Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

Keywords: censored or truncated data; failure times; recurrent events; usage processes.

Received April 10, 1995; accepted May 30, 1995

Abstract. We consider the occurrence of warranty claims for automobiles when both age and mileage accumulation may affect failure. The presence of both age and mileage limits on warranties creates interesting problems for the analysis of failures. We propose a family of models that relates failure to time and mileage accumulation. Methods for fitting the models based on warranty data and supplementary information about mileage accumulation are presented and illustrated on some real data. The general problem of modelling failures in equipment when both time and usage are factors is discussed.

1. Introduction

In modelling the reliability of systems on automobiles and other types of equipment, it is often important to consider both the age of the equipment (i.e. the length of time since it was introduced into service) and its cumulative usage which, in the case of cars, is usually represented by mileage. For automobiles, warranty coverages have both age and mileage limits, and manufacturers want to model the occurrence of failures or other events as functions of age and mileage. Such models are needed to predict reliability or to assess changes to warranty plans, and when used in conjunction with explanatory variables can suggest opportunities for reliability improvement.

Field tracking studies that follow specific cars over time and record usage along with reliability events (hereafter termed "failures", for convenience) are expensive to conduct and as a result relatively little data is obtained in this way. It is therefore important to extract as much information as possible from warranty claims data which record the age and mileage of failures occurring while each car is under warranty. It is well known, however, that estimating failure distributions or rates from warranty data is problematic (e.g Lawless and Kalbfleisch 1992): even assuming that failures are correctly diagnosed, the fact that warranties have both age and mileage limits biases the recording of failures. For example, if there are two year and 24,000 mile limits then cars which accumulate mileage rapidly will not have all of their failures up to age two years reported. Conversely,

cars which accumulate mileage slowly will not have all of their failures up to 24,000 miles reported. To address this problem we need to have information both about the way that failures are related to age and mileage, and the variation in mileage accumulation across the population of cars in service.

The objectives of this paper are to model the dependence of failures on age and mileage, and to estimate failure distributions and rates from warranty claims data supplemented by information about mileage accumulation. Section 2 describes notation, the general types of models considered, and the kind of warranty data that we seek to utilize. Section 3 presents a specific family of models and associated inference procedures that may be used with warranty data. Section 4 illustrates the proposed methodology on data which motivated this research, and Section 5 concludes with some comments.

Although this paper deals with automobile warranty data, the concepts and models introduced apply more generally to equipment for which both age and some measure of cumulative usage are related to reliability. There are also points of contact with recent research on multiple time scales (e.g. Oakes 1995) and on time-dependent marker processes (e.g. De Gruttola and Tu, 1992; Jewell and Kalbfleisch, 1992; Self and Pawitan, 1992) in survival analysis, where information about cumulative exposures or other factors related to survival are considered. There are distinctive features about car warranty data, however, which make the problems described in this paper rather different from the usual survival-marker process applications.

2. Main Concepts and Notation

2.1. *Mileage Accumulation (Usage) and Failure*

We let $t \geq 0$ denote age (time since sale) and $u_i(t)$ denote the mileage at age t for the i'th car in some population. The mileage history $U_i = \{u_i(t), t \geq 0\}$ gives the (non-decreasing) mileage curve $u_i(t)$ over the lifetime of the car. We will consider both recurrent events and times to specific single events (failures). For the case of single failures, let T_i denote the age of car i at failure and $T_i^{(u)}$ the mileage; the two time variables are related by

$$T_i^{(u)} = u_i(T_i). \tag{2.1}$$

The effect of the mileage accumulation process on failure will be modelled through the distribution of T_i given (i.e. conditional on) U_i. This automatically specifies the joint distribution of $\left(T_i, T_i^{(u)}\right)$ given U_i. Unconditional distributions of $(T_i, T_i^{(u)})$, T_i or $T_i^{(u)}$ require the additional specification of a model for the U_i's in the population.

A model for T_i given U_i may be specified in terms of the hazard function

$$h(t|U_i) = \lim_{\Delta t \downarrow 0} Pr\{T_i < t + \Delta t | T_i \geq t, U_i\}/\Delta t. \tag{2.2}$$

Recurrent events or multiple types of failures may be handled similarly, by considering

event intensity functions conditional on U_i. A conditional Poisson process for recurrent events would, for example, be specified by

$$\lambda(t \mid U_i) = \lim_{\Delta t \downarrow 0} Pr\{\text{event in } [t, t + \Delta t) \mid H_t, U_i\}/\Delta t, \tag{2.3}$$

where H_t represents the history of events on the automobile up to age t. In fact, the models (2.2) and (2.3) will be assumed to depend on U_i only through $\{u_i(s), s \leq t\}$, but the present notation is convenient.

The mileage curve has the status of an "external" time-dependent covariate in (2.2) or (2.3) (e.g. Kalbfleisch and Prentice 1980, Section 5.3). In taking this approach we ignore the possibility that the usage of equipment may depend on its prior history of failures and treat mileage accumulation as something that is determined independently of the failure processes. This is a reasonable assumption for cars during the early part of their lives and, in particular, during warranty periods. Models for the U_i's are introduced in Section 3.

We remark that for some car systems and, more generally, for certain systems in other types of equipment, failures may depend primarily on only one of usage or age. In this case either $T_i^{(u)}$ or T_i, respectively, would be independent of U_i for the case of single failures, with an analogous condition for multiple or recurrent events. Much previous work on the estimation of failure distributions as functions of usage have implicitly assumed that $T_i^{(u)}$ is independent of U_i (e.g. Suzuki 1993, Suzuki and Kasashima 1993). To avoid systematic bias it is important that we be able to check such assumptions; the methods of Section 3 allow us to do this.

2.2. *Warranty and Mileage Accumulation Data*

For cars there is typically a record of when each vehicle entered service (was sold) and then subsequent records of the age and mileage at each failure occurring while the car is under warranty. Three typical mileage accumulation curves are shown in Figure 1, along with the location of a failure for each. Age and mileage limits T^o and u^o are also shown: if the warranty plan has these limits then a failure is observed (i.e. recorded in the warranty data) only if $t_i \leq T^o$ and $u_i(t_i) \leq u^o$. Thus in Figure 1 the failure on the middle curve would be recorded, but not those on the lower or higher curves. Since we know the number of cars entering service we would only know in these cases that no failure occurred inside the region $\{t \leq T^o,\ u_i(t) \leq u^o\}$. This unusual type of censoring and the fact that usage is recorded only at failure times leads to interesting estimation problems.

Let us, in particular, consider the case of single failure times. The probability density for a failure at age $T_i = t$ and mileage $T_i^{(u)} = t^{(u)}$ inside the warranty observation region is

$$\int_{\substack{U_i: \\ u_i(t)=t^{(u)}}} f(t \mid U_i)\, dP(U_i), \quad t \leq T^0, t^{(u)} \leq u^0 \tag{2.4}$$

where $f(t \mid U_i) = h(t \mid U_i) exp\{-\int_0^t h(s \mid U_i) ds\}$ is the density function of T_i given U_i corresponding to (2.2), and we use $dP(U_i)$ to represent the distribution of mileage accumulation curves. The probability that car i does not experience the failure in the

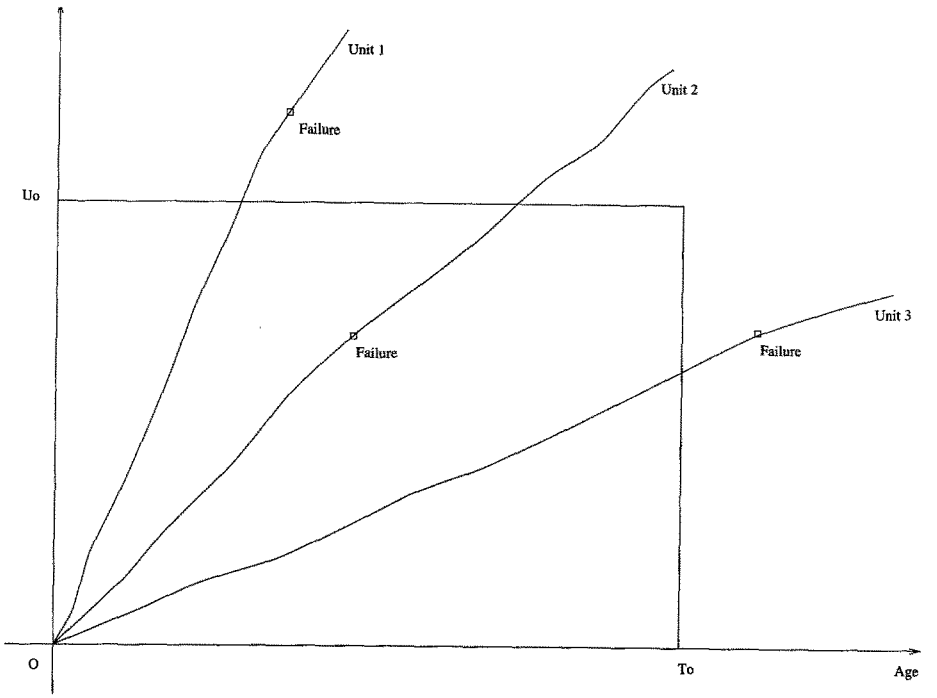

Figure 1. Mileage accumulation and failure

warranty region is, conversely,

$$Pr\left\{T_i > \min(T^o, u_i^{-1}(u^o))\right\}$$
$$= \int_{\substack{U_i: \\ u_i(T^o) \le u^o}} S(T^o \mid U_i) dP(U_i) + \int_{\substack{U_i: \\ u_i(T^o) > u^o}} S(u_i^{-1}(u^o) \mid U_i) dP(U_i), \tag{2.5}$$

where $S(t \mid U_i) = exp\{-\int_0^t h(s|U_i)ds\}$ is the survivor function corresponding to (2.2).

To evaluate (2.4) or (2.5) we require both a model for T_i given U_i and a model for U_i. The information about the distribution of mileage accumulation curves in the warranty data is limited, and confounded with failure information. It is important that $dP(U_i)$ be estimable from other sources and fortunately data to do this are typically available from customer surveys and field tracking studies. We discuss this further in Section 3, where specific models are introduced. The example in Section 4 describes some actual mileage accumulation information.

3. A Family of Models nd Estimation Methods

3.1. Models

There are various ways one might model usage processes and their relationship to failure. We require models that will be tractable and estimable from the type of data described in Section 2.2 and so introduce here a somewhat over-simplified family of models which, however, capture the essential features of age-mileage failures for automobiles.

We assume that the mileage accumulation curve U_i for a car can be represented as a function of age t and a vector of parameters α_i,

$$u_i(t) = m(t; \alpha_i) \quad t \geq 0. \tag{3.1}$$

The α_i's are assumed to vary from car to car, and we suppose that the parameters $\alpha_1, \ldots, \alpha_M$ for a population of M cars are generated as independent random variables with common distribution function $G(\alpha) = Pr(\alpha_i \leq \alpha)$. To connect the mileage curve and failure, we note that for single failures (2.2) may be expressed as $h(t \mid U_i) = h(t \mid \alpha_i)$ and that for recurrent events following a Poisson process, (2.3) becomes $\lambda(t|\alpha_i)$.

Mileage accumulation tends to be roughly linear over the first few years of a car's life, so we will work with the special case of (3.1),

$$u_i(t) = \alpha_i t \quad t \geq 0 \tag{3.2}$$

with the α_i's having distribution function $G(\alpha)$ and density $g(\alpha)$. This model has been used by others such as Suzuki (1993) and Suzuki and Kasashima (1993) and although it ignores seasonal effects or other short-term fluctuations in mileage accumulation rates, it provides reasonable results in most situations. The effect of departures from (3.2) will be briefly considered later in this section.

We choose to employ parametric models relating failure and α_i. This allows us to extrapolate failure probability calculations to age and mileage values beyond the warranty limits T^o and u^o and thus to estimate longer term reliability and assess the effect of increasing the warranty's age or mileage limits. Non- or semi-parametric models are more difficult to handle with the type of data considered in this paper, but we do present simple nonparametric estimates in Section 4 that may be used in special situations. The types of models described here are similar to ones used in biostatistics to relate marker processes and failures for subjects in longitudinal studies (e.g. De Gruttola and Tu, 1992). The type of data and the objectives in those situations are, hovever, different from ours.

There are two rather obvious approaches to modelling the dependence of failure on α_i in (3.2): proportional hazards and accelerated failure time models (e.g. Lawless 1982, Chapter 6). For single failure times, the former would assume $h(t \mid \alpha_i)$ to be of the form $h_0(t)\phi(\alpha_i)$ and the latter would assume it to be of the form $h_0[t\phi(\alpha_i)]\phi(\alpha_i)$, where in each case $\phi(\alpha_i)$ is a positive-valued function and $h_0(.)$ is a baseline hazard function. We will employ an accelerated failure time approach here; as we show below it has the advantage of providing simple special cases in which either of T_i or $T_i^{(u)}$ is independent of U_i.

We will describe the single failure time model first. The survivor function of T_i given α_i is assumed to be of the specific form

$$S(t; \alpha_i) = Pr(T_i > t \mid \alpha_i) = S_0(t\alpha_i^{\beta}), \tag{3.3}$$

where β is an unknown parameter and $S_0(t) = S_0(t; \theta)$ is a baseline survivor function specified up to a vector of parameters θ. Note that the survivor function of $T_i^{(u)}$ given α_i is, from (3.2)

$$Pr(T_i^{(u)} > t^{(u)} \mid \alpha_i) = S_0(t^{(u)}\alpha_i^{\beta-1}). \tag{3.4}$$

Thus, when $\beta = 0$ we have T_i independent of α_i, and when $\beta = 1$, $T_i^{(u)}$ is independent of α_i.

Accelerated failure time models for recurrent events are defined in a similar way. We consider a Poisson process model only; in this case, given α_i, recurrent events are assumed to follow a Poisson process with intensity function of the form

$$\lambda(t \mid \alpha_i) = \alpha_i^{\beta}\lambda_0(t\alpha_i^{\beta}), \tag{3.5}$$

where $\lambda_0(t)$ is a baseline intensity function specified up to a parameter vector θ. As with single failure times, the cases $\beta = 0$ and $\beta = 1$ imply that the recurrent event process in terms of age and mileage, respectively, is independent of α_i.

We comment on two features of the assumed models. As noted, (3.2) is an oversimplification. We could if desired replace (3.2) with a stochastic process for $u_i(t)$ which had mean function $\alpha_i t$, conditional on α_i. A convenient approach would then be to assume that given α_i, failure times are independent of U_i. In this case, however, the likelihood contributions (3.6) and (3.7) are replaced by much more complicated calculations. Since car-to-car variation in α_i values tends to dominate within-car variation around the trend curves $\alpha_i t$, (3.2) should provide reasonably adequate inferences in the current situation. A second point is that we have assumed $G(\alpha)$ in (3.8) and (3.9) to be known when, in practice, it is estimated from some data source. It is possible to allow for the fact that G is estimated in the calculation of standard errors for estimates $\hat{\theta}$, $\hat{\beta}$, obtained by maximizing (3.8) or (3.9) below. We discuss this in the example of Section .

We will consider and illustrate specific models in Section 4, but first we briefly discuss parameter estimation.

3.2. Estimation

We assume that the distribution $G(\alpha)$ is either known or estimated from information external to the warranty failures, and use the warranty data to estimate the parameters θ and β in (3.3) or (3.5) via maximum likelihood. There are two types of observations, described in Section 2.2, which give two types of likelihood contributions.

For single failure times the likelihood contributions are based on (2.4) or (2.5), depending on whether car i had an observed failure under warranty or not. For the model (3.2) these

become, respectively,

$$f(t_i \mid \alpha_i)g(\alpha_i) \tag{3.6}$$

$$\int Pr\left\{T_i > \min\left(T^0,\ u^o/\alpha_i\right) \mid \alpha_i\right\} dG(\alpha_i), \tag{3.7}$$

where in (3.6) $\alpha_i = t_i^{(u)}/t_i$. We will make a small adjustment to (3.7) that is useful when some of the cars in the warranty data set have been sold recently. If, when the data are assembled, car i has reached age a_i, then T^o in (3.7) should be replaced with min $(T^o,\ a_i)$. Since the dates of sale for all M cars in the data set are known, the a_i's can be computed for every car.

Using the family of models (3.3), we obtain the following likelihood function for θ and β:

$$L(\theta, \beta) = \prod_{i=1}^{m} \left\{\alpha_i^{\beta} f_0(t_i\alpha_i^{\beta}; \theta)\right\} \prod_{i=m+1}^{M} \int_0^{\infty} S_0\left\{\alpha^{\beta} \min\left(T^o, a_i, u^o/\alpha\right); \theta\right\} dG(\alpha), \tag{3.8}$$

where $f_0(t) = -S_0'(t)$ is the baseline failure density function, the cars experiencing failures are labelled $i = 1, \ldots, m$, and $\alpha_i = t_i^{(u)}/t_i (i = 1, \ldots, m)$. The likelihood (3.8) is relatively easy to maximize with respect to θ and β. The lognormal and Weibull distributions frequently fit the failure time data well, and we illustrate the implementation of (3.8) with a Weibull model in Section 4.

For recurrent events we assume that car $i (i = 1, \ldots, m)$ has $n_i > 0$ claims at times $t_{ij} (j = 1, \ldots, n_i)$. Based on the model (3.5), this produces the likelihood function

$$L(\theta, \beta) = \prod_{i=1}^{m} \prod_{j=1}^{n_i} \left\{\alpha_i^{\beta} \lambda_0(t_{ij}\alpha_i^{\beta}; \theta)\right\} e^{-\Lambda_0(\tau_i\alpha_i^{\beta};\theta)} \prod_{i=m+1}^{M} \int_0^{\infty} e^{-\Lambda_0(\tau_i\alpha_i^{\beta};\theta)} dG(\alpha_i), \tag{3.9}$$

where $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ and $\tau_i = \min(T^o, a_i, u^o/\alpha_i)$.

4. Example

We consider for illustration some real warranty data for a specific system on a car. The warranty in question was for one year or 12,000 miles and the data we consider here included warranty claims for $M = 8394$ cars manufactured in one plant during a two month period. Warranty claims were recorded up to 18 months after the first car was sold, but there were nevertheless, some cars that had been in service less than one year when the final data update was made. Among the M cars, $m = 823$ had at least one warranty claim; the car's age and mileage at the time of each claim are available. The dates of sale for all M cars are also known.

Information about the distribution of mileage accumulation rates α_i in (3.2) in the population of cars under warranty is available from a customer survey. A survey of 607 cars of the same type and approximate usage location as those in the warranty data base was taken,

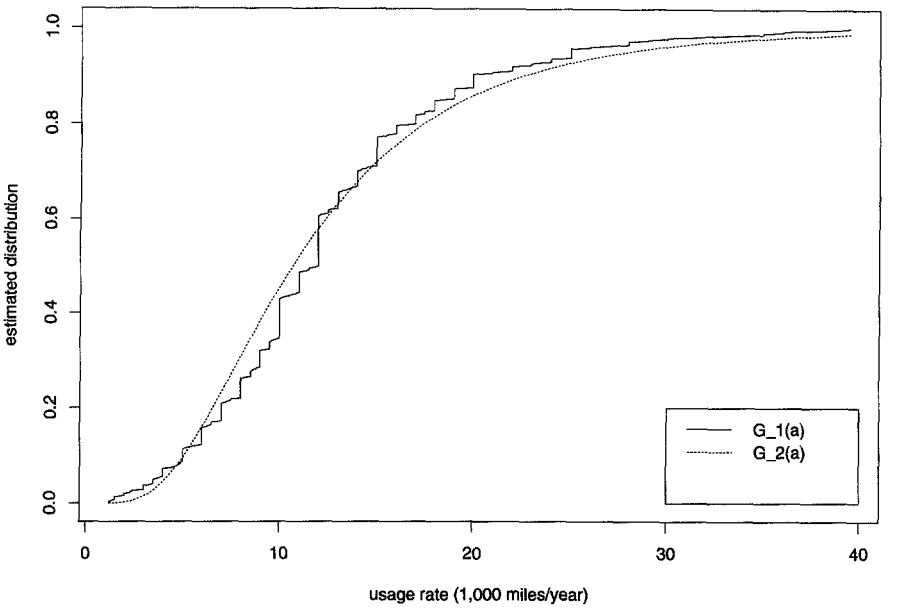

Figure 2. Estimated usage rate distributions

and the data included the mileage at age one year for each car. This allows us to estimate the distribution $G(\alpha)$ in Section 3. We show two estimates in Figure 2: the empirical c.d.f. based on the sample of 607, denoted as $G_1(\alpha)$, and a lognormal distribution fitted to the data, denoted as $G_2(\alpha)$. The units used for α_i are thousands of miles per year. In the latter case the mean and standard deviation of $\log \alpha_i$ are 2.37 and .58. In the calculations below we used $G_1(\alpha)$.

For failure time data it is convenient to re-express (3.8) in terms of the distribution of log failure times and to consider models for which the baseline distributions of log failure time are of location-scale form (Lawless 1982, Chapter 1). Define

$$Y_i = \log T_i,\ X_i = \log \alpha_i,\ y_i^o = \min(T^o, \log a_i),\ x_i^* = \log u^o - y_i^o\ ,$$

and assume that the distribution of Y_i given $X_i = x$ has survivor function

$$Pr(Y_i \geq y \mid X_i = x) = S_1\left(\frac{y + \beta x - \mu}{\sigma}\right), \tag{4.1}$$

where $-\infty < \mu < \infty$ and $\sigma > 0$ are location and scale parameters, β has precisely the same meaning as in Section 3, and $S_1(\cdot)$ is a survivor function defined on $(-\infty, \infty)$. The explicit relationship between $S_0(\cdot)$ in (3.3) or (3.8) and $S_1(\cdot)$ is given by $S_0(t) = S_1[(\log t - \mu)/\sigma]$.

Under (4.1), the log likelihood arising from (3.8) is

$$\ell(\mu, \beta, \sigma) = \sum_{i=1}^{m} \left\{ -\log\sigma + \log f_1\left(\frac{y_i + \beta x_i - \mu}{\sigma}\right)\right\} + \sum_{i=m+1}^{M} \log P_i(\mu, \beta, \sigma), \tag{4.2}$$

where $f_1(z) = -S_1'(z)$ and

$$P_i(\mu, \beta, \sigma) = \int_{-\infty}^{x_i^*} S_1\left(\frac{y_i^o + \beta x - \mu}{\sigma}\right) dG_X(x) + \int_{x_i^*}^{\infty} S_1\left[\frac{\log u^o + (\beta - 1)x - \mu}{\sigma}\right] dG_X(x)\,, \tag{4.3}$$

where $G_X(x)$ is the c.d.f. of $X_i = \log\alpha_i$.

We may maximize (4.2) and obtain variance estimates using Newton's method and the observed information matrix. The required derivatives of $\ell(\mu, \beta, \sigma)$ are straightforward but algebraically tedious to write down. Alternatively, (4.2) may be maximized with a general purpose optimizer that returns an estimate of the Hessian matrix at the maximum $(\hat{\mu}, \hat{\beta}, \hat{\sigma})$.

We will describe the results of fitting a Weibull model to the times of first failure in the warranty data described above. In this case $S_0(t)$ of (3.3) and $S_1(z)$ of (4.1) are of the forms

$$S_0(t) = \exp\left\{-(t/\theta_1)^{\theta_2}\right\} \quad t > 0$$

$$S_1(z) = \exp\{-\exp(z)\} \quad -\infty < z < \infty$$

where in (4.1), (4.2) and (4.3), $\mu = \log\theta$, and $\sigma = \theta_2^{-1}$. The maximum likelihood estimates and their standard errors, as estimated from the inverse of the observed information matrix (Lawless 1982, p. 523) are

$$\hat{\theta}_1 = 60.45(s.e.\ 12.52) \quad \hat{\theta}_2 = 1.128(.0382) \quad \hat{\beta} = .928(.0715),$$

with age t in years and mileage rate α_i in thousands of miles per year. The estimated survivor function for the distribution of time to first failure is then (see (3.3))

$$\hat{Pr}\,(T_i > t \mid \alpha_i) = \exp\left\{-\left(t\alpha_i^{\hat{\beta}}/\hat{\theta}_1\right)^{\hat{\theta}_2}\right\} \tag{4.4}$$

It is possible to compute standard errors taking into account that $G(\alpha)$ is not known exactly, but is estimated from a survey of 607 cars. When working with a parametric model for $G(\alpha)$, such as the lognormal distribution fitted above and shown in Figure 2, results of Gong and Samaniego (1981) and Parke (1986) show how to make the necessary adjustments. When a nonparametric estimate of $G(\alpha)$ is used, as here, results of Hu and Lawless (1995b) on pseudo likelihood estimation with supplementary information may be applied. When this is done the standard errors for $\hat{\theta}_1$, $\hat{\theta}_2$ and $\hat{\beta}$ increase to 16.73, .0389, and .1027, respectively.

We carried out diagnostic checks on the fitted model in two ways. One was to examine plots of truncated residuals, defined as follows. Let $\tau_i = \min\left(T_0, a_i, u^0/\alpha_i\right)$ represent the effective censoring time for car i; then $T_i = t_i$ is observed if and only if $t_i \leq \tau_i$. Conditional on α_i, a_i, and the event $T_i \leq \tau_i$, the quantity

$$e_i = \frac{F\left(T_i \mid \alpha_i\right)}{F\left(\tau_i \mid \alpha_i\right)}, \tag{4.5}$$

where $F\left(t \mid \alpha_i\right) = Pr\left(T_i \leq t \mid \alpha_i\right)$, is uniformly distributed on (0,1). We thus define residuals $\hat{e}_i$ for cars with observed failures by replacing T_i with t_i and $F(t \mid \alpha)$ with its estimate in (4.5). Figures 3a and 3b show an index plot $\left(\hat{e}_i \text{ vs. } i\right)$ and a uniform probability plot $\left(\hat{e}_{(i)} \text{ vs. } i/824\right)$ of the $\hat{e}_i$'s for the 823 cars with failures. No lack of fit is suggested.

A second check was to estimate the probability of failures under warranty and the probability of failure before $t = 1$ year, for various usage (mileage) rates α. The former is estimated via (3.7); we obtain the value .10, which is very close to the fraction 823/8394 of the cars observed to have failures. The latter probabilities increase with the mileage rate. The mean rate for these cars is about 14 (thousand miles per year), and gives a probability of failure by 1 year of about .14, consistent with the observed data.

Let us further examine the fitted model and also consider nonparametric estimation. It is noted that there is not strong evidence against the value $\beta = 1$ which, by (3.4), implies that the mileage $T_i^{(u)}$ at failure is more or less independent of the mileage accumulation rate. If $T_i^{(u)}$ is independent of α_i then we may obtain a simple nonparametric estimate of the distribution of $T_i^{(u)}$, as follows. For each car, define $\delta_i(s) = 1$ if the car is observed at mileage s. That is,

$$\delta_i(s) = 1 \text{ iff } s \leq \min\left[u_i\left(a_i\right), u_i\left(T^0\right), u^0\right],$$

where $T^0 = 1$ year and $u^0 = 12$ thousand miles. We do not know the $\delta_i(s)$'s for each of the $M = 8394$ cars, but we can estimate

$$p_i(s) = Pr\left\{\delta_i(s) = 1\right\}$$

by using the known distribution $G(\alpha)$ of mileage rates and the dates of sale. It follows that if the $\delta_i(s)$'s are independent of the $T_i^{(u)}$'s, then the c.d.f. of $T_i^{(u)}$ is estimated by

$$\hat{F}_u\left(t^{(u)}\right) = \hat{Pr}\left(T_i^{(u)} \leq t^{(u)}\right) = \int_0^{t^{(u)}} \frac{\sum_{i=1}^M dN_i(s)}{\sum_{i=1}^M p_i(s)},$$

where $dN_i(s)$ is the observed number of failures on car i at mileage s. This estimate may be rewritten as

$$\hat{F}_u\left(t^{(u)}\right) = \sum_{s_j^* \leq t^{(u)}} \frac{dN_{\cdot}\left(s_j^*\right)}{p_{\cdot}\left(s_j^*\right)}, \tag{4.6}$$

where the s_j^*'s are the distinct mileages at which failures were observed across all cars, $dN_{\cdot}\left(s_j^*\right)$ is the total number of failures reported at s_j^*, and $p_{\cdot}\left(s_j^*\right) = \sum_{i=1}^M p_i\left(s_j^*\right)$.

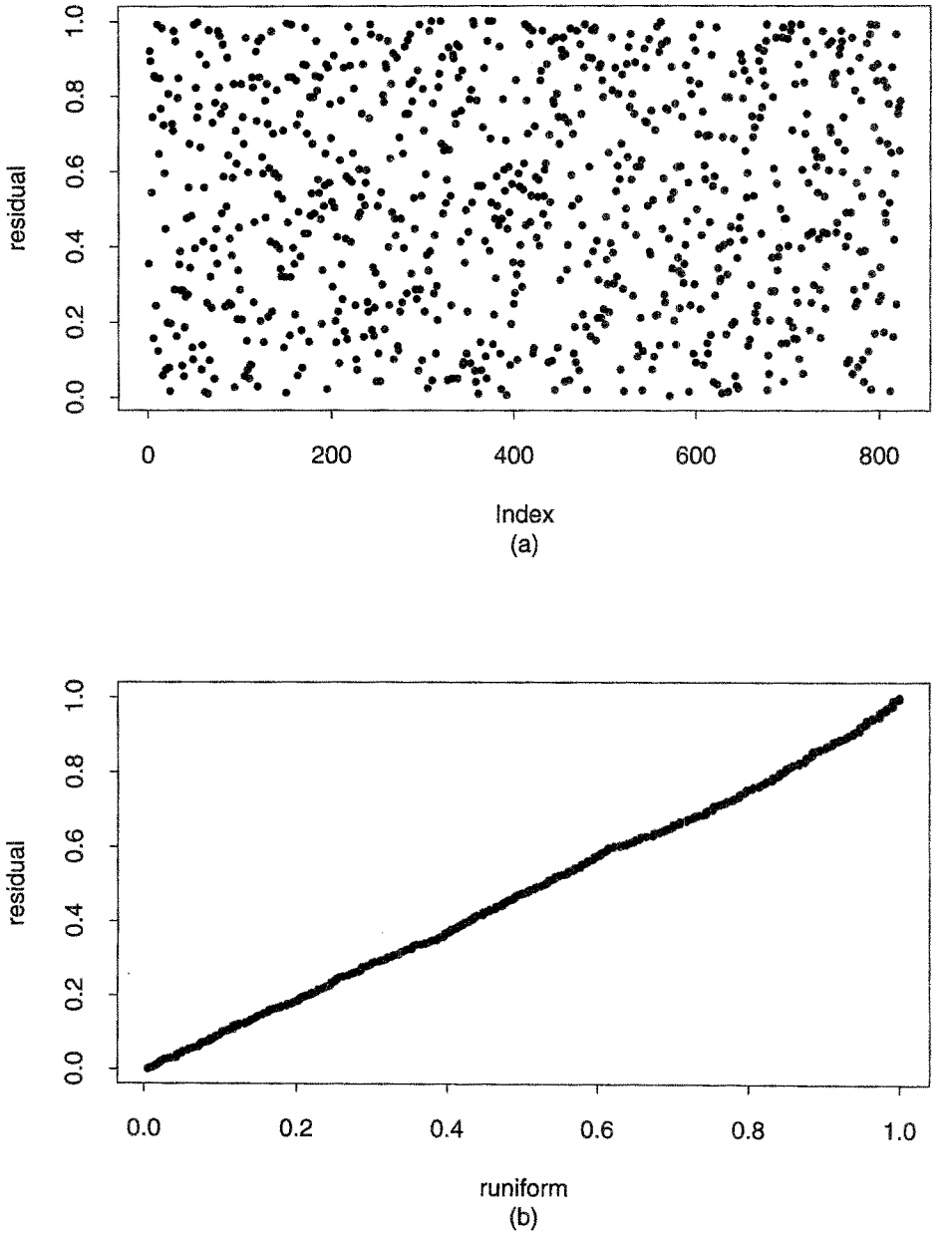

Figure 3. Index and probability plots of residuals

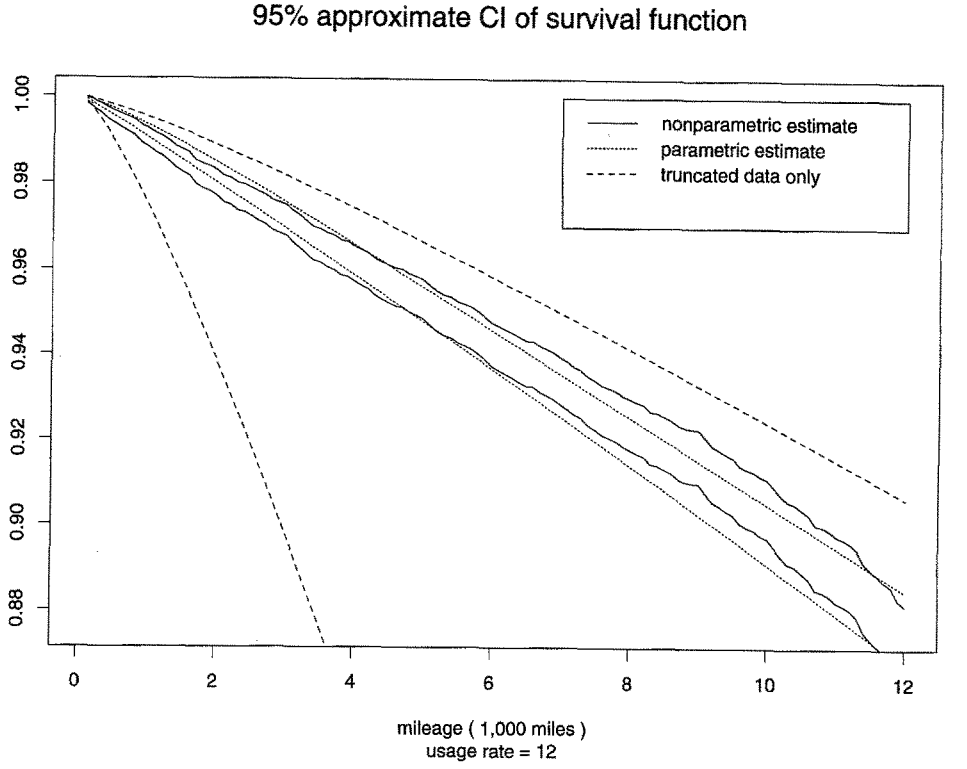

Figure 4. Confidence limits for $Pr(T_i^{(u)} > s)$

It is easily seen that (4.6) is unbiased, and its variance may be estimated by

$$\hat{V}(t^{(u)}) = \sum_{i=1}^{M} \left\{ \int_0^{t^{(u)}} \frac{dN_i(s) - p_i(s)d\hat{F}_u(s)}{p_\cdot(s)} \right\}^2 . \tag{4.7}$$

In Figure 4 we show pointwise approximate .95 confidence limits for $Pr(T_i^{(u)} > s)$ computed two ways: (i) using the Weibull model derived from (4.4) with the estimates $\hat{\theta}_1, \hat{\theta}_2, \hat{\beta}$ and usage rate $\alpha = 12$ (since $\hat{\beta}$ is close to 1, the usage rate has relatively little effect) and (ii) using the nonparametric estimate (4.6). In each case confidence limits were obtained as $\hat{Pr}(T_i^{(u)} > s) \pm 1.96$ standard errors. The estimates are seen to be in good agreement, thus lending further credence to the Weibull model. Note that the estimates are shown only up to $s = 12$ thousand miles, since no failures are observed beyond that point. Figure 4 also includes confidence limits based solely on the observed failures (i.e. the truncated data only, involving 823 cars), as described in Section 5.

We remark that the nonparametric procedure outlined here may be applied to estimate failure time distributions or recurrent event mean functions whenever the censoring times

τ_i for units are independent of the failure times for the units. In the current example this condition is not met when we consider age T_i at failure, because the mileage accumulation rates α_i affect both T_i and τ_i. However, for mileage $T_i^{(u)}$ at failure, the censoring mileage $\tau_i^{(u)} = \min[u_i(a_i), u_i(T^0), u^0]$ is more or less independent of $T_i^{(u)}$.

We conclude this example with some remarks on the use of the fitted model (4.4) to assess the effect of changes to warranty coverage. If, for example, we wish to estimate the probability or expected number of claims if the plan had 2 year, 24 thousand mile limits then (4.4) and the assumed distribution $G(\alpha)$ of mileage rates allow us to do this. In particular, with a T^0 year, u^0 thousand mile warranty the probability of no claim for a car is obtained from (3.7). With $T^0 = 2, u^0 = 24$ we estimate the probability of a claim to be .20, using the failure model (4.4). An obvious warning is of course that in making this estimate we are extrapolating the Weibull model well beyond the range of the current data. Similarly, extrapolations to very low or very high mileage accumulation rates should be treated with caution.

5. Concluding Remarks

Our objective has been to consider the rather interesting problems that arise with failure data obtained under a warranty scheme for automobiles. The fact that failure may depend on both age and mileage accumulation and the presence of both age and mileage limits in warranty coverage creates difficulties for modelling and analysis. A secondary objective has been to discuss models for failure when both time and usage of a product may be factors. The latter topic is closely related to work on time-dependent marker processes and multiple time scales (e.g. De Gruttola and Tu, 1992; Jewell and Kalbfleisch, 1992; Self and Pawitan, 1992; Oakes, 1995), where the difficulty of formulating tractable joint models for failure and marker processes has been noted. In this paper we have adopted a simple model based on linear mileage accumulation: this model seems adequate for the current application and, in any event, the type of censoring created by the warranty plan makes it difficult to fit or assess more complex models.

Murthy and Wilson (1991) consider models similar to those in Section 3.1, and also discuss models where $(T_i, T_i^{(u)})$ are assumed to have some specific family of bivariate distributions. Their objective is to study costs associated with different types of warranties, and they do not consider any inference procedures. The second type of model does not make assumptions about variation in usage and is less flexible in utilizing usage information obtained from sources external to the warranty data. It is also not easily extended to deal with recurrent events. However, it would be interesting to compare the distributions for $(T_i, T_i^{(u)})$ generated by models like those in Section 3.1 with some of the common bivariate failure time models.

We have noted earlier that it is important to have external information about the mileage accumulation processes. Because the warranty plan severely censors failure times, it is not possible to estimate $G(\alpha)$ from the warranty data alone. We can fit the model (3.3) using only the warranty data by considering the distribution of observed failure times t_i, conditional on mileage rates α_i, censoring times τ_i, and the fact that $t_i \leq \tau_i$. Hu and Lawless (1995a,b) consider such types of truncated data and demonstrate that they are

relatively uninformative about the parameters θ and β in models like (3.3). For precise estimation it is important to utilize additional information about mileage accumulation and the number of cars experiencing no failure under warranty. By way of illustration, we show in Figure 4 confidence limits for $Pr(T_i^{(u)} > s)$ based only on the truncated data for the 823 cars with failures; they are extremely wide relative to those based on the methods of Section 4.

Acknowledgment

This research was supported in part by grants to the first author from General Motors Canada, The Manufacturing Research Corporation of Ontario, and the Natural Sciences and Engineering Research Council of Canada. We thank Diane Gibbons and Jeff Robinson for assistance with the data.

References

1. V. De Gruttola and X. M. Tu, "Modeling the relationship between progression of CD-4 lymphocyte count and survival time," *AIDS Epidemiology: Methodological Issues* (N. P. Jewell, K. Dietz, and V. T. Farewell, eds.), pp. 275–296, Birkhäuser: Boston, 1992.
2. G. Gong and F. G. Samaniego, "Pseudo maximum likelihood estimation: Theory and applications," *Ann. Statist.* vol. 9 pp. 861–869, 1981.
3. X. J. Hu and J. F. Lawless, "Estimation of rate and mean functions from truncated recurrent event data," to appear in *J. Amer. Statist. Assoc.*, 1995a.
4. X. J. Hu and J. F. Lawless, "Estimation from truncated lifetime data with supplementary information," unpublished manuscript, 1995b.
5. N. P. Jewell and J. D. Kalbfleisch, "Marker models in survival analysis and applications to issues associated with AIDS," *AIDS Epidemiology: Methodological Issues* (N. P. Jewell, K. Dietz, and V. T. Farewell, eds.), pp. 211–230, Birkhäuser: Boston, 1992.
6. J. D. Kalbfleisch and R. L. Prentice, *The Statistical Analysis of Failure Time Data*, John Wiley and Sons: New York, 1980.
7. J. F. Lawless, *Statistical Models and Methods for Lifetime Data*, John Wiley and Sons: New York, 1982.
8. J. F. Lawless and J. D. Kalbfleisch, "Some issues in the collection and analysis of field reliability data," In *Survival Analysis: State of the Art*, pp. 141–152, Kluwer: Amsterdam, 1992.
9. D. N. P. Murthy and R. J. Wilson, "Modelling two-dimensional failure free warranties," in *Proc. 5th Int. Symp. of Applied Stoch. Models and Data Analysis*, Granada, Spain, 1991.
10. D. Oakes, "Multiple time scales in survival analysis," *Lifetime Data Analysis* vol. 1, pp. 7–18, 1995.
11. W. R. Parke, "Pseudo maximum likelihood estimation: the asymptotic distribution," *Ann. Statist.* vol. 14 pp. 355–357, 1986.
12. S. Self and Y. Pawitan, "Modeling a marker of disease progression and onset of disease," *AIDS Epidemiology: Methodological Issues* (N. P. Jewell, K. Dietz, and V. T. Farewell, eds.), pp. 231–255, Birkhäuser: Boston, 1992.
13. K. Suzuki, "Estimation of lifetime distribution using the relationship of calendar time and usage time," *Rep. Statist. Appl. Res.* vol. 40 pp. 10–22, 1993.
14. K. Suzuki and T. Kasashima, "Estimation of lifetime distribution from incomplete field data with different observational periods," Tech. Report, UEC-CAS-93-02, University of Electro-Communications, Department of Communications and System Engineering, Tokyo, Japan, 1993.