Estimation of Rate and Mean Functions From Truncated Recurrent Event Data

X. Joan HU and Jerald F. LAWLESS

This article investigates data on recurrent events that arise from sources such as warranty claims, where the observation period for a unit is unknown until it experiences at least one event. This creates a type of truncation in the data. We consider nonparametric estimation of means and rates of the event occurrences with such "zero-truncated" data and examine the case where the population size and the distribution of observation times across units are at least approximately known. We study the behaviors of the proposed estimators by simulation. We examine some car warranty data by applying the methodology developed here and considering the underlying assumptions.

KEY WORDS: Nonparametric estimation; Population size; Reliability; Truncation distribution; Zero-truncated data.

1. INTRODUCTION

Field failure data provide important information about the reliability of manufactured products. Because follow-up of selected products is expensive, there is much interest in the utilization of information from sources such as failure reports or warranty claim records (see, for example, Kalbfleisch, Lawless, and Robinson 1991 and Lawless and Kalbfleisch 1992). But a difficulty with such data is that the time origin for a unit is often unknown until it experiences at least one event (e.g., failure or warranty claim), thus creating a type of truncation in the data (see, for example, Kalbfleisch and Lawless 1988, Lawless and Kalbfleisch 1992, and Suzuki 1985a,b, 1987). This article deals with data on recurrent events that arise from such sources, although applications of the methodology are not restricted to warranty or field failure problems. For example, a sociological study that tracked repeated utilizations of a social service by a population of individuals might not know of an individual's existence until the first utilization occurred.

The statistical aspects of the topic are as follows. An individual or unit experiences recurrent events over time $t > 0$; we let $N_i(t)$ denote the number of events over $(0, t]$ for unit i. Unit i is potentially observed (i.e., its times of event occurrence are recorded) over the time interval $(0, \tau_i]$; we call τ_i the observation or truncation time. We consider both discrete and continuous time. When t is discrete, we will assume that it takes on values $t = 1, 2, 3, \ldots$, and that the τ_i's likewise take on integer values. A set of M units, $i = 1, 2, \ldots, M$, is assumed to have iid event occurrence processes $\{N_i(t), t > 0\}$ and observation times $\tau_i, i = 1, \ldots, M$, that are determined independently of the event processes. The statistical objective is to estimate the mean function $\Lambda(t) = E\{N_i(t)\}$ and the corresponding rate, defined as $\lambda(t) = \Lambda'(t)$ when time is continuous and as $\lambda(t) = E\{N_i(t) - N_i(t-1)\}$ when time is discrete. The rate or mean function does not fully specify the probability structure of the recurrent events (unless the events happen to follow a Poisson process), but in many applications the rate and mean functions are of much interest (see, for example, Sun and Kalbfleisch 1993, Thall and Lachin 1988, and the examples herein).

A novelty in the situations we consider is that unit i and τ_i are observed only if $N_i(\tau_i) > 0$; that is, provided that at least one event has occurred. Otherwise the τ_i value is unknown, and we may even be unaware of the unit's existence. The motivation for studying these situations came from attempts to utilize warranty data. Manufacturers collect information about failures or repairs that result in warranty claims, and it is of considerable interest to estimate the mean number of repairs $\Lambda(t)$ and the rate $\lambda(t)$ per unit from warranty records. "Time" may be elapsed calendar time since the unit was sold (i.e., "age" of the unit) or some measure of usage, such as accumulated mileage on automobiles. The problem is to estimate $\Lambda(t), t > 0$ from warranty data collected up to some point in time. For a unit sold before then, the observation time τ_i is a function of when the unit was sold, the type of warranty plan, and possibly the usage history of the unit. In most situations the value of τ_i becomes known only when the unit has its first claim. Moreover, in some situations—for example when "time" is a measure of usage—the τ_i's may never be known exactly but only approximately. We give a pair of specific examples.

Example 1. Many products, such as tools or small household appliances, have a fixed time warranty, say 1 year from the date of sale. In this case, if unit i is sold at calendar time x_i, X is the current calendar time, and t represents the age of the unit, then $\tau_i = \min(X - x_i, 1)$, with time measured in years. With products of this type, however, the manufacturer usually receives date of sale information for at most a small fraction of units. Thus the date of sale and truncation time τ_i become known for most units only when a claim arises.

Example 2. For automobiles, the manufacturer is informed of the date of sale for each car. But because war-

X. Joan Hu is a Postdoctoral Fellow at Health Canada, B57 EHC, Tunneyis Pasture, Ottawa, Ontario, Canada K1A 0L2. Jerald F. Lawless is the GM and NSERC Chair Professor, Department of Statistics and Actuarial Science, University of Waterloo, Ontario, Canada N2L 3G1. This research was supported in part by grants to the second author from the Natural Sciences and Engineering Research Council of Canada and the Manufacturing Research Corporation of Ontario. The authors thank two referees and an associate editor for remarks that led to major improvements in the exposition.

Journal of the American Statistical Association
March 1996, Vol. 91, No. 433, Theory and Methods

ranties usually have both calendar time and mileage limits, the observation times τ_i are typically observed or estimated only when a claim is made. For example, if the warranty coverage is for 2 years or 24,000 miles and t represents age of the car in years, then $\tau_i = \min(X - x_i, 2, y_i)$, where x_i is the calendar time of sale, X is the current calendar time, and y_i is the age at which the car mileage reaches 24,000. Although x_i is known, y_i is not. But it can be estimated from the mileage at the time of the first claim; this is usually done under the generally reasonable assumption that in their first few years, cars accumulate mileage approximately linearly with time. If t represented mileage instead of age, then we would have $\tau_i = \min(w_i(2), w_i(X - x_i), 24{,}000)$, where $w_i(a)$ is the mileage on car i at age a. Once again, τ_i may be estimated from the mileage at the time of the first claim. When the τ_i's are estimated as here, an "errors in variables" aspect is introduced, but we will ignore this in the sequel.

Poisson processes frequently provide good models for recurrent events, and estimation of mean and rate of occurrence functions for them has been studied extensively (see, for example, Cox and Lewis 1966 and Crow 1974). Estimation of Poisson rates of occurrence with incomplete data on observation time has been discussed by Suzuki (1987). Kalbfleisch and Lawless (1988), Lawless and Kalbfleisch (1992), Suzuki (1985a,b, 1993), and Suzuki and Kasashima (1993) have considered similar problems for the estimation of failure time distributions. Much of this work deals with right-truncation of lifetimes or failure times. Wang (1989, 1992) considered similar problems, motivated by AIDS data analysis. Our emphasis in this article is on nonparametric methods, with a desire to avoid the assumption that the recurrent event processes are Poisson; here there has been much less work. Nelson (1988) and Lawless and Nadeau (1995) discussed completely observed processes. Thall and Lachin (1988) and Sun and Kalbfleisch (1993) studied situations where only interval event counts were available.

Our article is different from previous work in several respects. We consider nonparametric estimation of means and rates of occurrence from truncated data and also consider situations where the distribution of observation times across units is at least approximately known. In cases where truncation is heavy, information about observation times allows much more precise estimation of event rates. Such information is also crucial to the development of robust estimators of rate and mean functions. If one has only truncated data, then it is essential to assume a specific event process to proceed, because we need to model the probability that an individual is observed (not truncated). Poisson processes often provide satisfactory models, but methods based on them may be nonrobust. In Section 4 we construct rate and mean function estimators that are valid for quite arbitrary recurrent event processes. In addition, we study the robustness of Poisson process estimators under some alternative models.

The article is organized as follows. Section 2 discusses nonparametric estimation of mean functions when only "zero-truncated" data are available. A Poisson process of events is assumed. Section 3 presents Poisson process methods when the total number of units and the distribution of observation times in the population are known; the increased information this provides is addressed. Section 4 gives robust estimators for rate and mean functions without any strong assumptions about the recurrent event process. Simulation is used to study the estimators in Sections 2, 3, and 4, and it is demonstrated that the Poisson estimates based on truncated data alone are nonrobust. Section 5 considers the situation where data are aggregated across units. Section 6 illustrates our results on some car warranty data. Finally, Section 7 presents some remarks concerning extensions to the present work.

2. ESTIMATION FROM ZERO-TRUNCATED POISSON PROCESS DATA

We assume in this section that $\{N_i(t) : t > 0\}, i = 1, \ldots, M$, are independent Poisson processes with common rate of occurrence function $\lambda(t)$ and mean function $\Lambda(t)$. The process i has an observation window $(0, \tau_i]$, where the τ_i's are determined independently of the event processes. We consider estimation of $\Lambda(t)$ when we are aware of only those processes with at least one event over $(0, \tau_i]$. For those processes the times of events and the value of τ_i are assumed known. The value of M is unknown in such a case.

Suppose for notational convenience that processes $i = 1, \ldots, m$ have at least one event and that for them the times of events t_{ij} ($j = 1, \ldots, n_i$, where $n_i = N_i(\tau_i)$) and observation times τ_i are observed. We then have the "zero-truncated" likelihood function

$$L_T = \prod_{i=1}^{m} \Pr\{n_i, t_{ij}\text{'s} | n_i \geq 1, \tau_i\}$$
$$= \prod_{i=1}^{m} \left\{ \prod_{j=1}^{n_i} \lambda(t_{ij}) \frac{\exp[-\Lambda(\tau_i)]}{1 - \exp[-\Lambda(\tau_i)]} \right\}, \quad (1)$$

where we use "$\Pr\{\cdot\}$" to represent either a probability or probability density, depending on whether the model is in discrete or continuous time. Estimation from (1) for parametric models $\lambda(t; \theta)$ poses no particular difficulties. Our objective is to develop nonparametric estimates, which are valuable for assessing the shape of $\Lambda(t)$ and checking parametric assumptions.

It is simplest to obtain nonparametric estimates of $\Lambda(t)$ by assuming that time is discrete; the estimates also apply to the continuous-time case, as we discuss later. Thus we suppose without loss of generality that t takes on values 1, 2, ..., and let $n_i(t)$ be the number of events observed at time t for unit i. Letting $\tau_{\max} = \max(\tau_1, \ldots, \tau_m)$, we can write $\log(L_T)$ as

$$l_T = \sum_{t=1}^{\tau_{\max}} n.(t) \log \lambda(t) - \sum_{t=1}^{\tau_{\max}} \sum_{i=1}^{m} \delta_i(t) \lambda(t)$$
$$- \sum_{i=1}^{m} \log\{1 - \exp[-\Lambda(\tau_i)]\}, \quad (2)$$

where $n.(t) = \sum_{i=1}^{m} \delta_i(t) n_i(t)$ and $\delta_i(t) = I(\tau_i \geq t)$ indicate whether $\tau_i \geq t$ is true or not. This gives

$$\frac{\partial l_T}{\partial \lambda(t)} = \frac{n.(t)}{\lambda(t)} - \sum_{i=1}^{m} \frac{\delta_i(t)}{1 - \exp[-\Lambda(\tau_i)]}, \qquad t = 1, \ldots, \tau_{\max}, \tag{3}$$

and

$$-\frac{\partial^2 l_T}{\partial \lambda(t) \partial \lambda(s)} = \frac{n.(t) I(s = t)}{\lambda(t)^2} - \sum_{i=1}^{m} \frac{\delta_i(t)\delta_i(s) \exp[-\Lambda(\tau_i)]}{\{1 - \exp[-\Lambda(\tau_i)]\}^2}. \tag{4}$$

Estimates $\hat{\lambda}_T(t), t = 1, \ldots, \tau_{\max}$, are obtained by setting (3) equal to zero, and then $\hat{\Lambda}_T(t) = \sum_{s=1}^{t} \hat{\lambda}_T(s)$. This can be done using Newton's method, but if $\tau_{\max}$ is large, then it is simpler to use the iteration procedure

$$\tilde{\lambda}(t)^{(j+1)} = \frac{n.(t)}{\sum_{i=1}^{m} \frac{\delta_i(t)}{1 - \exp[-\tilde{\Lambda}^{(j)}(\tau_i)]}}, \tag{5}$$

where $\tilde{\lambda}(t)^{(j)}$ $(j = 1, 2, \ldots)$ is the jth iterate toward $\hat{\lambda}_T(t)$. The procedure (5) is self-consistent and may be shown to converge to a root of (3) by arguments of Turnbull (1976). We note that $\hat{\lambda}_T(t) = 0$ if $n.(t) = 0$, so (5) has to be carried out only for t values at which there is at least one event.

The procedure also applies to continuous-time processes, by letting $\tau_{\max}$ be large enough so that all distinct event times and τ_i values can be associated with one of $1, 2, \ldots, \tau_{\max}$. The values $\hat{\lambda}_T(t)$ are zero except when t is an event time and so do not give a particularly attractive estimate of the continuous rate function $\lambda(t)$, but the mean function estimate $\hat{\Lambda}_T(t)$ is attractive and is in fact a nonparametric maximum likelihood estimator (MLE) with typical properties of MLE's, as we discuss later. (More appealing nonparametric estimates of $\lambda(t)$ may be obtained by smoothing, but this is beyond the scope of this article.)

Assuming that $\Pr\{N(\tau) > 0\} > 0$, it follows from standard parametric maximum likelihood theory that for the discrete-time case, the $\hat{\lambda}_T(t)$'s and $\hat{\Lambda}_T(t)$'s are consistent and, conditional on m and τ_i's, have limiting normal distributions as $m \to \infty$. Asymptotic covariance matrices for the $\hat{\lambda}_T(t)$'s and $\hat{\Lambda}_T(t)$'s may be obtained in the usual way from the inverse of the observed information matrix with entries given by (4) evaluated at $\hat{\lambda}_T(t)$ $(t = 1, \ldots, \tau_{\max})$. These may be used to construct confidence intervals for $\Lambda(t)$, but if t is large, then the inversion of large $(t \times t)$ matrices is involved. In the continuous-time case it is possible to prove the consistency of $\hat{\Lambda}_T(t)$ and that $\sqrt{m}\{\hat{\Lambda}_T(t) - \Lambda(t)\}$ converges to a mean zero Gaussian process for $0 < t < \tau_{\max}$, conditional on m and τ_i's. (This takes us away from the main theme of this article and will be described elsewhere.) But there is no simple representation or estimate of the asymptotic variance for $\hat{\Lambda}_T(t)$. Such circumstances are not unusual and occur for various other situations involving truncated or interval-censored data.

If we use the nonparametric estimate $\hat{\Lambda}_T(t)$ for informal graphical checks on parametric assumptions, then confidence limits are not essential. If they are wanted, then our approach is to consider a weakly parametric model for which variance estimates are more easily obtained. The family of models with piecewise constant $\lambda(t)$ functions is convenient; in Appendix B we outline estimation for this family and in Section 6 we use it in an example. Thus our approach may be summarized as: use the nonparametric estimate $\hat{\Lambda}_T(t)$ for graphical model checking but use the weakly parametric estimate if variance estimates are also required. Provided that the piecewise constant rate model has a moderately large number of parameters, estimates based on it agree very closely with $\hat{\Lambda}_T(t)$.

We remark that it may be possible to develop variance estimates for $\hat{\Lambda}_T(t)$ based on some form of resampling or the jackknife. How best to do this is not obvious, but it is under investigation.

3. POISSON ESTIMATION WITH KNOWN POPULATION SIZE AND TRUNCATION DISTRIBUTION

In many situations the number of units and the observation time distribution $G(\cdot)$ is more or less known. For Example 1 in Section 1, the manufacturer might have a reasonably accurate estimate of sales over time. Similarly, for Example 2, car manufacturers usually have estimates of the distribution of mileage accumulation rates for the population of cars. This allows $G(\cdot)$ to be estimated, as we illustrate in Section 6. We now consider estimation of $\lambda(t)$ and $\Lambda(t)$ when $G(\cdot)$ and M are known.

3.1 Nonparametric Maximum Likelihood

Parametric models are readily fitted by maximizing the likelihood (6). We once again focus on nonparametric estimation of $\Lambda(t)$. The data consist of $n_i = N_i(\tau_i)$, the t_{ij}'s $(j = 1, \ldots, n_i)$, and τ_i if $n_i \geq 1$, plus the knowledge of the number of units with $n_i = 0$. As in Section 2, we consider the discrete-time case and assume for convenience that units $i = 1, \ldots, m$ are those with $n_i \geq 1$. The likelihood function is then

$$L_{TK} = \prod_{i=1}^{m} \left\{ \prod_{j=1}^{n_i} \lambda(t_{ij}) \exp[-\Lambda(\tau_i)] dG(\tau_i) \right\} \times \left\{ \int_0^\infty \exp[-\Lambda(\tau)]\, dG(\tau) \right\}^{M-m}. \tag{6}$$

We consider discrete-time models for τ_i as well, and define $g(s) = dG(s) = \Pr(\tau_i = s)$. Defining $\delta_i(t) = I(t \leq \tau_i)$ as in Section 2, we can write $\log(L_{TK})$ as

$$\begin{aligned} l_{TK} = & \sum_{t=1}^{\tau_{\max}} n.(t) \log(\lambda(t)) \\ & - \sum_{t=1}^{\tau_{\max}} \sum_{i=1}^{m} \delta_i(t) \lambda(t) + (M - m) \\ & \times \log \left(\sum_{\tau=1}^{\infty} \exp[-\Lambda(\tau)] g(\tau) \right) \\ & + \sum_{i=1}^{m} \log(g(\tau_i)). \end{aligned}$$

We then have

$$\frac{\partial l_{TK}}{\partial \lambda(t)} = \frac{n.(t)}{\lambda(t)} - \sum_{i=1}^{m} \delta_i(t) - (M-m)\,\frac{A(t)}{A(1)} \qquad (7)$$

and

$$-\frac{\partial^2 l_{TK}}{\partial \lambda(t)\partial \lambda(s)} = \frac{n.(t)}{\lambda(t)^2}\, I(s=t) - (M-m) \times \left\{ \frac{A(s\wedge t)A(1) - A(s)A(t)}{A(1)^2} \right\}, \qquad (8)$$

when $t, s = 1, \ldots, \tau^*$, where $A(t) = \sum_{s=t}^{\tau^*} \exp[-\Lambda(s)]g(s)$, $s \wedge t = \min(s,t)$, and $\tau^* = \sup\{s\colon g(s) > 0\} \geq \tau_{\max}$. Note that $A(1) = \Pr(n_i = 0)$ here. We obtain the nonparametric MLE $\hat{\lambda}_{TK}(t)$'s by solving equations $\partial l_{TK}/\partial\lambda(t) = 0$, $t = 1, \ldots, \tau^*$ and then $\hat{\Lambda}_{TK}(t) = \sum_{s=1}^{t} \hat{\lambda}_{TK}(s)$.

A convenient algorithm for obtaining the $\hat{\lambda}_{TK}(t)$'s is

$$\tilde{\lambda}(t)^{(j+1)} = \frac{n.(t)}{\sum_{i=1}^{m} \delta_i(t) + (M-m)\,\frac{\tilde{A}^{(j)}(t)}{\tilde{A}^{(j)}(1)}}, \qquad (9)$$

where $\tilde{A}^{(j)}(t) = \sum_{s=t}^{\tau^*} \exp[-\tilde{\Lambda}^{(j)}(s)]g(s)$. It may be shown that this is an EM algorithm for (6), based on complete data for which all τ_i's $(i = 1, \ldots, M)$ are known, and converges to a root of (7) (see, for example, Turnbull 1976). Note that for $t > \tau_{\max}, \tilde{\lambda}(t)^{(j)} = 0$, for $j = 1, \ldots$, and thus $\hat{\lambda}_{TK}(t) = 0$. Note also that for $t \leq \tau_{\max}, \tilde{A}^{(j)}(t)$ in (9) is

$$\sum_{s=t}^{\tau_{\max}-1} \exp[-\tilde{\Lambda}^{(j)}(s)]g(s) + \exp[-\tilde{\Lambda}^{(j)}(\tau_{\max})]\bar{G}(\tau_{\max}),$$

where $\bar{G}(t) = \sum_{s=t}^{\tau^*} g(s)$.

The same comments that were made in Section 2 about the $\hat{\lambda}_T(t)$ and $\hat{\Lambda}_T(t)$ apply to the estimators $\hat{\lambda}_{TK}(t)$ and $\hat{\Lambda}_{TK}(t)$. In particular, $\hat{\Lambda}_{TK}(t)$ is consistent and asymptotically normal in either the discrete-time or continuous-time framework, but asymptotic variance estimation is complicated. Our approach is to use $\hat{\Lambda}_{TK}(t)$ for graphical model checks but to use a weakly parametric model with a piecewise constant rate function when variance estimates are wanted. Estimation for such models is described in Appendix B.

In Eample 2 of Section 1, if t represents mileage, then it would be better to allow the τ_i's to come from different distributions. In particular, we could use $\tau_i \sim G_{a_i}(\cdot)$ with $a_i = X - x_i$, because τ_i depends on the time since the car was sold. In this context, Suzuki and Kasashima (1993) gave a nonparametric estimate for the distribution of mileage to the first failure or claim. If we assume that $\tau_i \sim G_{a_i}(\cdot)$ with $a_i \in \{1, \ldots, K\}$ and $G_{a_i}(\cdot)$'s known, the likelihood corresponding to (6) is

$$L_{TK} = \prod_{i=1}^{m} \left\{ \prod_{j=1}^{n_i} \lambda(t_{ij}) \exp[-\Lambda(\tau_i)] dG_{a_i}(\tau_i) \right\} \times \prod_{k=1}^{K} \left\{ \int_0^\infty \exp[-\Lambda(\tau)] dG_k(\tau) \right\}^{m_k}, \qquad (10)$$

where m_k is the size of $\{i\colon a_i = k$, and $N_i(\tau_i) = 0\}$. We can obtain the MLE of $\Lambda(t)$ by a straightforward extension of the methods in this section.

3.2 Information Added by Knowing M and $G(\cdot)$

If information about M and the distribution $G(\tau)$ is available, then much more precise inferences are possible under heavy truncation. We demonstrate this by comparing the estimators based on the truncated data likelihood (1) and the more informative likelihood (6). We use simulation to compare the estimators in a realistic finite-sample context.

We did two simulations, with events occurring according to a homogeneous Poisson process. Our objective is to compare the behaviors of the estimators $\hat{\Lambda}_T(t)$ and $\hat{\Lambda}_{TK}(t)$. We also compare them to the estimator $\hat{\Lambda}_C(t)$, which would be obtained if we knew every $\tau_i, (i = 1, \ldots, M)$, and not just those for which $n_i \geq 1$. This is the Nelson–Aalen estimator (see Andersen, Borgan, Gill, and Keiding 1993) $\hat{\Lambda}_C(t) = \sum_{s=1}^{t} \hat{\lambda}_C(s) = \sum_{s=1}^{t} [n.(s)/\sum_{i=1}^{M} \delta_i(s)]$. Our program was written in $C+$, but we used the functions runif and rexp for the uniform distribution and the exponential distribution in $S+$ for generating the random variables needed. We used the algorithms (5) and (9) to get $\lambda_T(t)$ and $\lambda_{TK}(t)$ and terminated the iterations when $\sum_{t=1}^{150} |\tilde{\lambda}(t)^{(j+1)} - \tilde{\lambda}(t)^{(j)}| \leq .0001$. The simulations are as follows:

1. We used $M = 150$ and generated $\tau_i, i = 1, \ldots, 150$, from the uniform distribution on (0, 300], and 150 corresponding time-homogeneous Poisson processes $N_i(t), i = 1, \ldots, 150$, with rate $\lambda = .01$. Here truncation is not too heavy; the expected value of m is between 102 and 103. Based on the generated data, we obtained the estimates $\hat{\Lambda}_T(t), \hat{\Lambda}_{TK}(t)$, and $\hat{\Lambda}_C(t)$. We repeated the simulation $n = 100$ times and found from the sample means of the estimates that all the estimators are essentially unbiased. Figure 1(a) presents the corresponding sample mean squared errors.

2. We took $\lambda = .001$ and kept the other aspects of the foregoing simulation unchanged. Now truncation is heavy, and the expected value of m is between 20 and 21. Once again, the estimators are essentially unbiased, except for some slight positive bias in $\hat{\Lambda}_T(t)$, as judged by the sample means of the estimates. We give the sample mean squared errors of the three estimators of $\Lambda(t)$ in Figure 1(b).

The simulations show that the estimators based on L_T and L_{TK} have similar mean squared errors when truncation is fairly light and are about as efficient as $\hat{\Lambda}_C(t)$. When truncation is heavier, the information added by knowing M and $G(\cdot)$ plays an important role; L_T is relatively much less informative, but $\hat{\Lambda}_{TK}(t)$ is still about as efficient as $\hat{\Lambda}_C(t)$. Obviously, the heavier the truncation, the more important the knowledge of M and $\bar{G}(\cdot)$.

We remark that mean squared errors for a fourth estimator (denoted by $\hat{\Lambda}_{SM}(t)$ in the figures), which is introduced in Section 4, are also shown in Figure 1.

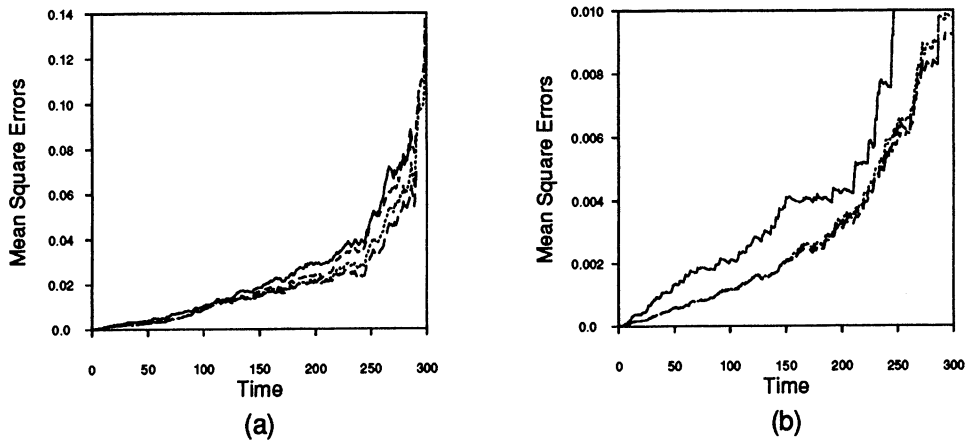

Figure 1. Simulation With $M = 150$, $G(\cdot) = U(0, 300)$, and $(N(t), t > 0)$ a Time-Homogeneous Poisson Process With $E\{N(t)\} = \lambda t$: (a) plot of the sample mean squared errors of the estimates for 100 simulated processes with $\lambda = .01$; (b) plot of the sample mean squared errors of the estimates for 100 simulated processes with $\lambda = .001$, where ——, $\hat{\Lambda}_T(t)$; $\cdots$, $\hat{\Lambda}_{TK}(t)$; - - -, $\hat{\Lambda}_{SM}(t)$; and – – –, $\hat{\Lambda}_C(t)$.

4. ROBUST ESTIMATION

4.1 A Robust Estimator

We consider another estimator that can be used if M and $G(\cdot)$ are known. If observation times are independent of the event processes, then $E\{\delta_i(t)n_i(t)\} = \bar{G}(t)\lambda(t)$, and

$$n.(t) - M\bar{G}(t)\lambda(t) = 0 \tag{11}$$

is a unbiased estimating equation (i.e., the left side has expectation zero) for each $t = 1, 2, \ldots$. Solving (11), we get

$$\hat{\lambda}_{SM}(t) = \frac{n.(t)}{M\bar{G}(t)} \tag{12}$$

and $\hat{\Lambda}_{SM}(t) = \sum_{s=1}^{t} \hat{\lambda}_{SM}(s)$, for $t = 1, \ldots, \tau^*$. Note that (11) is valid even if the event process is not Poisson but $E\{n_i(t)\} = \lambda(t)$, and that $\hat{\lambda}_{SM}(t)$ is in fact unbiased.

Under mild conditions, we can prove (see Appendix A) that $\sqrt{M}(\hat{\Lambda}_{SM}(t) - \Lambda(t)) \xrightarrow{d} N(0, \sigma_t^{(SM)^2}), t = 1, \ldots, \tau^*$, with $\sigma_t^{(SM)^2}$ estimated consistently by

$$\hat{V}_{SM}(t) = \frac{1}{M}\sum_{i=1}^{M}\left\{\sum_{s=1}^{t}\left[\frac{\delta_i(s)n_i(s)}{\bar{G}(s)} - \hat{\lambda}_{SM}(s)\right]\right\}^2. \tag{13}$$

The estimators $\hat{\Lambda}_{SM}(t)$ and $\hat{V}_{SM}(t)$ can be written in integral forms:

$$\hat{\Lambda}_{SM}(t) = \int_0^t \frac{dN.(s)}{M\bar{G}(s)},$$

$$\hat{V}_{SM}(t) = \frac{1}{M}\sum_{i=1}^{M}\left\{\int_0^t \frac{1}{\bar{G}(s)}\left[\delta_i(s)\,dN_i(s) - \frac{dN.(s)}{M}\right]\right\}^2,$$

with $dN.(s) = n.(s)ds$. These expressions define valid nonparametric estimators in the case of continuous-time processes.

In Figure 1 we show the mean squared errors of $\hat{\Lambda}_{SM}(t)$ based on the data generated in the simulation study of Section 3. We observe that $\hat{\Lambda}_{SM}(t)$ is about as efficient as $\hat{\Lambda}_{TK}(t)$ for the situations considered.

4.2 Robustness Comparison of Estimators

As noted in Sections 2 and 3, the estimators there are based on a Poisson process and may not be robust to departures from that model. To check the robustness of the various estimators, we consider a mixed Poisson model (see Lawless 1987), where each unit i has an associated random variable α_i such that events for it occur according to a Poisson process with rate function $\alpha_i\lambda(t)$. Such models are plausible when units function in different environments. The α_i's were taken to have a gamma distribution with mean 1, so that $E\{N_i(t)\} = \Lambda(t)$ still holds. The variance of $N_i(t)$ under a mixed Poisson model is $\Lambda(t)[1 + \text{var}(\alpha_i)\Lambda(t)]$. The larger either $\text{var}(\alpha_i)$ or $\Lambda(t)$, the more overdispersed the $N_i(t)$'s relative to a Poisson distribution. We did two simulations, as described next, using the same numerical procedures as in the simulations of Section 3:

1. $\tau_i, i = 1, \ldots, 150$, are generated as earlier from $U(0, 300)$; $\alpha_i, i = 1, \ldots, 150$, are from a gamma distribution with mean 1 and variance 1, generated by using the function rgamma in $S+$. The 150 counting processes, $N_i(t), i = 1, \ldots, 150$, are then independent homogeneous Poisson processes with rate function $\lambda_i(t) = \alpha_i\lambda(t) = .01\alpha_i$. Figure 2(a1)–(a2) presents the sample means and sample mean squared errors of $\hat{\Lambda}_T(t), \hat{\Lambda}_{TK}(t), \hat{\Lambda}_{SM}(t)$, and $\hat{\Lambda}_C(t)$ obtained by repeating the simulation $n = 100$ times.

2. We generated data in the same way as for simulation 1 except that $\alpha_i, i = 1, \ldots, 150$ were from a gamma distribution with mean 1 and variance .2. Figure 2(c1)–(c2) gives the sample means and sample mean squared errors of the estimates for $n = 100$ simulations.

These two simulations were each repeated with $\lambda(t) = .001$. The corresponding results are shown in Figure 2(b1)–(b2) and 2(d1)–(d2).

From Figure 2(a1)–(d1), we see corroboration of the fact that $\hat{\Lambda}_{SM}(t)$ and $\hat{\Lambda}_C(t)$ are unbiased but $\hat{\Lambda}_T(t)$ is biased. The bias depends on how overdispersed $N_i(t)$ is, and it can be substantial when overdispersion is large. Figure 2(a2)–(d2) shows a similar efficiency for $\hat{\Lambda}_{SM}(t)$ and $\hat{\Lambda}_C(t)$. The behavior of $\hat{\Lambda}_{TK}(t)$, which is based on the Poisson assump-

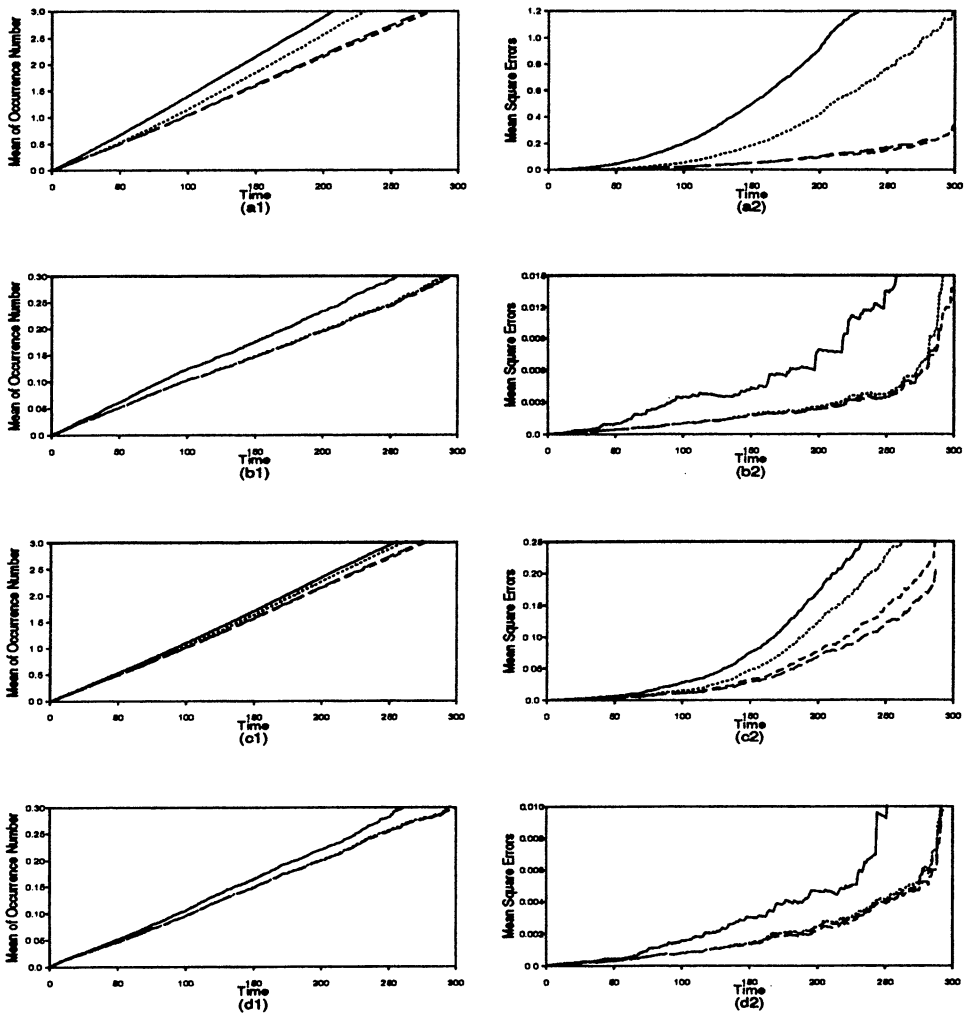

Figure 2. Simulation with $M = 150$, $G(\cdot) = U(0, 300)$, and $(N(t), t > 0)$ a Mixed Poisson Process, With $E\{N(t)|\alpha\} = \lambda t\alpha$, $\alpha \sim \Gamma(a, \sigma^2)$, and $E\{\alpha\} = a$, $V\{\alpha\} = \sigma^2$: (a1) plot of sample means of the estimates for 100 simulated processes with $\lambda = .01$, $a = \sigma = 1$; (a2) plot of the sample mean squared errors of the estimates corresponding to (a1); (b1) and (b2) plots of the results given by the simulation corresponding to (a1) and (a2) but with $\lambda = .001$; (c1) and (c2) plots of the results given by the simulation corresponding to (a1) and (a2) but with $\sigma^2 = .2$; (d1) and (d2) plots of the results given by the simulation corresponding to (c1) and (c2) but with $\lambda = .001$, where —— $\hat{\Lambda}_T(t)$; $\cdots$ $\hat{\Lambda}_{TK}(t)$; --- $\hat{\Lambda}_{SM}(t)$; - - - $\hat{\Lambda}_C(t)$.

tion, is apparent. When truncation is heavy (giving small m) it is more or less unbiased and is comparable to $\hat{\Lambda}_{SM}(t)$ and $\hat{\Lambda}_C(t)$ [see Fig. 2(b1)–(b2) and 2(d1)–(d2)], but when truncation is light (giving large m), it is biased and has considerably larger mean squared error than $\hat{\Lambda}_{SM}(t)$ and $\hat{\Lambda}_C(t)$ [see Fig. 2(a1)–(a2) and 2(c1)–(c2)]. The reason that $\hat{\Lambda}_{TK}(t)$ performs well under heavy truncation is that information from the truncated sample is dominated by the supplementary information about $G(\cdot)$ (Hu 1995). When $G(\cdot)$ is available, the results in Figures 1 and 2 suggest that $\hat{\Lambda}_{SM}(t)$ is generally to be preferred, especially given its simplicity.

4.3 Effect of Imprecise Information About $G(\cdot)$ and M

If we know $G(\cdot)$ and M, then much more efficient estimation of $\Lambda(t)$ is possible than if we have only zero-truncated data, especially when truncation is heavy. In addition, the estimator $\hat{\Lambda}_{SM}(t)$ of Section 4.1 is robust. We now consider the effect of misspecifying either $G(\cdot)$ or M, since they will often be known only approximately.

Overestimation or underestimation of M results in underestimation or overestimation of $\lambda(t)$ or $\Lambda(t)$ for the estimators in Sections 3.1 and 4.1. In particular, for $\hat{\lambda}_{SM}(t)$ given by (12), we estimate $(M/\tilde{M})\lambda(t)$ rather than $\lambda(t)$, where $\tilde{M}$ is the value (an estimate of the true value M) used in (12).

Misspecification of $G(\cdot)$ is a little more complicated to study. If we know M but assume the distribution of observation time is $\tilde{G}(\cdot)$ while the true one is $G(\cdot)$, then $\hat{\lambda}_T(t)$ and $\hat{\lambda}_C(t)$ are unaffected, because they do not depend on $G(\cdot)$. But we note that

$$E\left\{\frac{\partial l_{TK}}{\partial \lambda(t)}\right\} = M \Pr(N_i(\tau_i) = 0) \times \left[\frac{\partial \log \Pr(N_i(\tau_i) = 0)}{\partial \lambda(t)} - \frac{\partial \log \Pr(N_i(\tau_i) = 0|\tilde{G}(\cdot))}{\partial \lambda(t)}\right]$$

does generally equal zero if $\tilde{G}(\cdot) \neq G(\cdot)$, nor does the expectation of (11),

$$E\left\{\sum_{i=1}^{M} \delta_i(t) n_i(t) - \bar{\tilde{G}}(t)\lambda(t)\right\} = M\lambda(t)[\bar{G}(t) - \bar{\tilde{G}}(t)].$$

The estimates $\hat{\Lambda}_{TK}(t)$ and $\hat{\Lambda}_{SM}(t)$ would then be inconsistent. We can investigate the extent of the bias in the estimators; this will not be great if $\tilde{G}(\cdot)$ is a good approximation to $G(\cdot)$.

To examine the effect of misspecification of $G(\cdot)$, we performed some simulations. We generated τ_i from $N(150, 70^2) = G(\cdot)$ by using the function rnorm in $S+$, and $N_i(t)$ in the same way as the simulations in Section 3, for $i = 1, \ldots, 150$. We obtained estimates of $\Lambda(t)$ by taking $G(\cdot)$ as $\tilde{G}(\cdot) = U(0, 300]$. Each process was repeated 100 times. To conserve space we do not show figures of sample means and mean squared errors, but the simulations showed that $\hat{\Lambda}_{TK}(t)$ and $\hat{\Lambda}_{SM}(t)$ are not badly biased, and the efficiencies are similar to that of $\hat{\Lambda}_C(t)$, which does not use $G(\cdot)$, especially when truncation is heavy. The estimators in Sections 3.1 and 4.1 are thus not greatly affected by mild inaccuracies in M or $G(\cdot)$. But in situations where we are uncertain about their true values, we recommend checking the sensitivity of estimates $\hat{\Lambda}(t)$ and variance estimates $\hat{V}(t)$ to variations in $G(\cdot)$ or M.

5. DATA AGGREGATED ACROSS UNITS

Sometimes there is no linkage of events for individual units, in which case all we observe is the pair (t, τ) for each observed event. For example, warranty claims may be recorded according to the time of the claim and the time at which the product unit was sold, with no historical record kept for each unit. In this case, if M and $G(\cdot)$ are unknown, then we are able to obtain only the likelihood function based on $\Pr(t_l^* | t_l^* \le \tau_l^*, \tau_l^*)$ which, assuming that events occur according to a Poisson process, is

$$L_{T_2} = \prod_{l=1}^{n} \frac{\lambda(t_l^*)}{\Lambda(\tau_l^*)} = \prod_{i=1}^{m} \prod_{j=1}^{n_i} \frac{\lambda(t_{ij})}{\Lambda(\tau_i)}, \tag{14}$$

where $n = \sum_{i=1}^{m} n_i$ is the total number of events across all units, and (t_l^*, τ_l^*)'s are the observed (t, τ) values.

From (14) it is clear that we can estimate only $\lambda(t)/\Lambda(\tau_{\max})$, where $\tau_{\max} = \max(\tau_l^*)$, and not the absolute values of the $\lambda(t)$'s or $\Lambda(t)$'s. In fact, the estimation problem is similar to one for truncated failure time data (see, e.g., Kalbfleisch and Lawless 1991, Wang 1989, and Wang, Jewell, and Tsai 1986), and the nonparametric estimate from (14) can be given in closed form as follows:

Denote $\lambda(t)/\Lambda(t)$ by $\lambda^*(t)$. Then, for $t \le \tau_{\max}$,

$$\frac{\lambda(t)}{\Lambda(\tau_{\max})} = \lambda^*(t) \prod_{x=t+1}^{\tau_{\max}} [1 - \lambda^*(x)].$$

We can rewrite (14) as

$$L_{T_2} = \prod_{l=1}^{n} \frac{\lambda^*(t_l^*)}{\Lambda^*(\tau_l^*)} = \prod_{t=1}^{\tau_{\max}} \lambda^*(t)^{n(t)} [1 - \lambda^*(t)]^{n^*(t)}, \tag{15}$$

where $n(t) = \#\{l : t_l^* = t\} = n.(t)$ and $n^*(t) = \#\{l: t_l < t \le \tau_l\}$. The MLE of $\lambda^*(t)$ based on (15) is

$$\hat{\lambda}^*(t) = \frac{n.(t)}{n.(t) + n^*(t)}.$$

If M and $G(\cdot)$ are known, then, under the discrete Poisson process assumption, the data consisting of the pairs (t_l^*, τ_l^*) give the likelihood function

$$L_{T_{k2}} = \left\{ \prod_{l=1}^{n} \lambda(t_l^*) g(\tau_l^*) \right\} \exp\left\{ -M \sum_{s=1}^{\tau_{\max}} \sum_{t=1}^{s} \lambda(t) g(s) \right\},$$

which is proportional to

$$\prod_{t=1}^{\tau_{\max}} \lambda(t)^{n.(t)} \exp\left\{ -M \sum_{t=1}^{\tau_{\max}} \lambda(t) \bar{G}(t) \right\}. \tag{16}$$

Maximization of (16) gives the estimates

$$\hat{\lambda}(t) = \frac{n.(t)}{M\bar{G}(t)},$$

the same as $\hat{\lambda}_{SM}(t)$ in Section 4.1. As we show there, this estimate is robust to departures from the Poisson process.

6. AN EXAMPLE

In situations where "time" is calendar time and truncation arises because observation of individuals has to cease at some point, the truncation times τ_i are typically independent of the event processes. Example 1 of Section 1 is of this type. The methods of this article then apply, and in particular the mean function estimator $\hat{\Lambda}_{SM}(t)$ of Section 4.1 is robust and valid for essentially any practical situation. Occasionally, however, we encounter situations where the τ_i's may not be independent of the event processes; automobile warranty data as described in Example 2 of Section 1 sometimes falls into this category, because of simultaneous age and mileage limits placed on warranty coverage. Because it will allow us to illustrate more thoroughly both our methodology and the assumptions on which it is based, we consider here some real warranty data for a system (unknown to us for proprietary reasons) on a particular car model.

We consider a group of 8,394 cars manufactured over a 2-month period; this is a subset of a larger group discussed by Kalbfleisch et al. (1991). As of the final data base update, all $M = 8{,}394$ cars had been sold. They generated a total of 1,134 claims from 823 different cars. Let δ_i equal 1 if car i had a warranty claim and zero otherwise, and let $N_i(t)$ indicate the total number of claims from car i up to time t, where t can be either age (i.e., days since the car was sold) or mileage of the car. Denote by (a_{ij}, m_{ij}) the age and mileage of car i at the jth claim. If the observation time period for car i is denoted by $(0, \tau_i]$, we have $\{\tau_i, (a_{ij}, m_{ij} : j = 1, \ldots, N_i(\tau_i))\}$ as the data for car i, provided $\delta_i = 1$; that is, there is at least one claim. Otherwise all we know is the date of sale of the car and that there was no claim from it. The age and mileage limits of the warranty plan are 1 year and 12,000 miles. These and the rates at which cars accumulate miles determine the τ_i's. In the following, we obtain estimates of $\Lambda(t) = E\{N_i(t)\}$, the expected cumulative number of repairs per car as a function of age. Rates and means as a function of mileage can also be given.

We assume that the mileage accumulation rate for car i is u_i, so that $m_i(a) = u_i a$ is the mileage at age a. Although this ignores fluctuations in mileage accumulation, it is a reasonably practical assumption for many cars in their first 2 or 3 years. With a 1-year/12,000-mile warranty, the end of observation time for car i is $\tau_i = \min(X - x_i, 365, 120/u_i)$, where age is measured in days and mileage is measured in hundred miles, x_i is the day the car was sold, and X is the day of the final data update. For these cars, $X - x_i$ exceeded 365 days for 7,356 of 8,394 cars.

Customer surveys had been carried out for cars of the same type as in the warranty data base, in which the mileage at 1 year was obtained for each car sampled. We use this supplementary information to estimate $\bar{G}_2(t) = \Pr\{120/u_i \geq t\}$ and then estimate $\bar{G}(t) = \Pr\{\tau_i \geq t\}$ as $\hat{\bar{G}}_2(t)$ times $\hat{\bar{G}}_1(t) = \sum_{i=1}^{M} I[365 \wedge (X - x_i) \geq t]/M$, an estimate of $\bar{G}_1(t) = \Pr\{365 = \Pr\{365 \wedge (X - x_i) \geq t\}$, where $a \wedge b$ is used to denote $\min(a, b)$. We could stratify cars according to their dates of sale for the sake of precision, but it turns out that this gives almost exactly the same estimates we obtained before; hence we present the slightly simpler unstratified analysis. The estimated $\bar{G}(t)$ is shown in Figure 3. Note that $\hat{\bar{G}}(365) = .53$, but that $\hat{\bar{G}}(t) = 0$ for $t > 365$; we show $\hat{\bar{G}}(t)$ only for $t \leq 365$.

We began by obtaining the estimate $\hat{\Lambda}_T(t)$ of Section 2 and, using $\hat{\bar{G}}(t)$, the estimates $\hat{\Lambda}_{TK}(t)$ and $\hat{\Lambda}_{SM}(t)$ of Sections 3 and 4. These are shown in Figure 4, labeled "NT," "NTK," and "NSM," along with approximate .95 pointwise confidence limits based on approximating piecewise constant intensity models (see App. B) for $\hat{\Lambda}_{TK}(t)$ and $\hat{\Lambda}_{SM}(t)$. It is seen that $\hat{\Lambda}_{TK}(t)$ and $\hat{\Lambda}_{SM}(t)$ are almost identical but that $\hat{\Lambda}_T(t)$ is very different, and quite implausible based on the observed data. There are several potential sources for this discrepancy:

a. The Poisson process model on which $\hat{\Lambda}_T(t)$ and $\hat{\Lambda}_{TK}(t)$ are based may be inadequate. We have seen in Section 4 that under certain types of departures from a Poisson process, $\hat{\Lambda}_T(t)$ is a badly biased estimator of the true mean function, although if truncation is heavy like here ($m/M \doteq .10$), then $\hat{\Lambda}_{TK}(t)$ is satisfactory.
b. The truncation times τ_i may not be independent of the event processes. In particular, cars that have higher mileage accumulation rates u_i might also have higher event (claim) rates. In this case the τ_i's would not be independent of the event processes, because cars with high u_i's leave the warranty plan earlier and so have smaller τ_i's. This would affect the properties of all three estimators, because even the robust estimator $\hat{\Lambda}_{SM}(t)$ assumes independence.
c. The distribution of mileage accumulation rates in the survey sample that we used to estimate $G(\tau)$ may be different from that for the cars included in the warranty plan. In this case any differences are thought to be small, and we do not investigate this assumption here.

Because few claims per car are observed (an average of about .14), it is not possible to thoroughly examine the probabilistic structure of the event processes for individual cars. Indeed, this is one reason why we seek to estimate just the marginal mean function $\Lambda(t)$ for the processes. But it is possible to consider the situations (a) and (b) by fitting plausible extended models. Two that are generally useful and that may be adapted to the zero-truncated data presented here are as follows. For situation a, we consider mixed Poisson models in which a random effect or "frailty" α_i is associated with car i, as in Section 4.2. This is valuable because in many situations individual units have roughly Poisson event processes, but different environmental conditions or manufacturing variabilities generate extra-Poisson variation (Lawless 1987). For situation (b), it is useful to fit Poisson or mixed Poisson models with a covariate designed to deal with dependent censoring times; we illustrate this later.

Describing the fitting and diagnosis of such models in detail would extend the article unduly. For nontruncated data, Andersen et al. (1993) and Lawless (1987) have provided general methodology. For zero-truncated data, as in this ar-

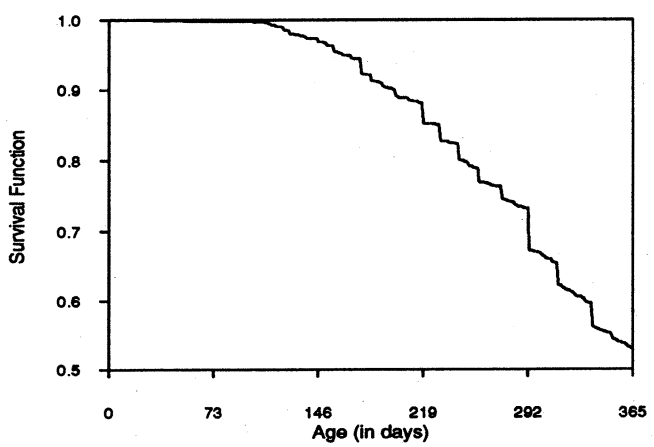

Figure 3. Plot of the Estimated $\bar{G}(\tau)$, $\tau \in (0, 365)$, Based on the Customer Sample Survey.

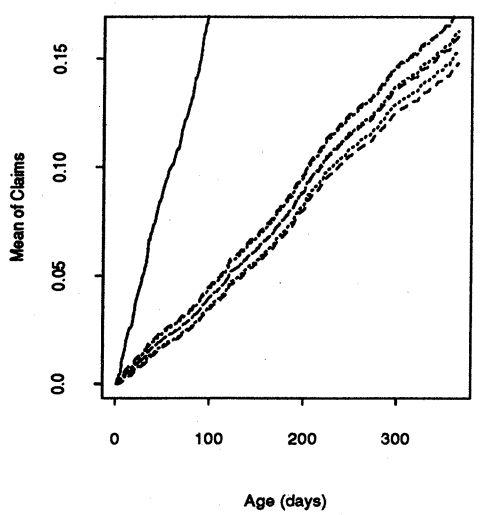

Figure 4. Estimates of $\Lambda(t)$ and Approximate 95% Confidence Limits Based on $\hat{\Lambda}_{TK}(t)$ and $\hat{\Lambda}_{SM}(t)$ —— NT; ··· NTK; --- NSM.

ticle, research is currently underway. Thus we simply outline how we have examined these models and then discuss the impact of our investigation on the present example.

Mixed Poisson models may be fitted by replacing the Poisson process with a mixed Poisson process in L_T (1) or L_{TK} (6), as done for nontruncated data by Lawless (1987). The simplest model is the negative binomial process, where the frailties α_i have a gamma distribution with mean 1 and variance σ^2. If σ^2 is close to zero, then there is little extra-Poisson variation, but preliminary fits with parametric rate functions indicate that $\hat{\sigma}^2$ here is in the range of 5.0. Because $\text{var}\{N_i(\tau)\} = \Lambda(\tau)\{1 + \sigma^2\Lambda(\tau)\}$ for this model, at $\tau = 1$ year we have a variance-to-mean ratio for $N_i(\tau)$ of approximately $1 + (5)(.16) = 1.80$, where we have used the estimate $\hat{\Lambda}_{SM}(t) \doteq .16$ at $t = 365$ days, shown in Figure 4. The simulation results in Section 4.2 suggest that with this degree of overdispersion, and with the heavy truncation here, $\hat{\Lambda}_T(t)$ may be expected to overestimate $\Lambda(t)$ considerably.

Nonindependence of the τ_i's and the claim processes may be investigated as follows. It is possible that the mileage accumulation rate u_i affects the rate function, so we consider models in which

$$\lambda_i(t) = \lambda_0(t)\exp(\beta u_i), \tag{17}$$

where $\lambda_0(t)$ is a baseline rate function and β is a parameter. Other forms of dependency of the rate $\lambda_i(t)$ on u_i can be considered, if necessary. Poisson or mixed Poisson models may once again be fitted via L_T or L_{TK}. Note that given u_i and the date of sale x_i, the observation time τ_i for automobile i is fixed, and thus τ_i and the claim process are conditionally independent, given u_i. Fits for parametric models (Hu 1995) indicate that the mileage accumulation rate indeed has a significant effect, with a $\hat{\beta}$ and standard error of about 3.17 and .16. In this case the accumulation rates u_i, being variable across the population of cars, act as a frailty effect and would lead $\hat{\Lambda}_T(t)$ to overestimate $\Lambda(t)$. But unlike the earlier mixed Poisson model, the model (17) implies that the τ_i's are nonindependent of the event processes, and so the robust estimator $\hat{\Lambda}_{SM}(t)$ may not necessarily estimate $E\{N_i(t)\}$ well. Note from (17) that

$$E\{N_i(t)\} = \Lambda_0(t)E\{\exp(\beta u_i)\}, \tag{18}$$

where the expectation on the right is over the distribution of u_i. Using the estimates from (17) and the distribution of u_i estimated from the survey data, we obtain the estimate labeled "FPM" in Figure 5. We also present a semiparametric fit of $E\{N_i(t)\}$ for the model (17) with nonparametric $\lambda_0(t)$ and $\hat{\beta} = 3.33$ and labeled "SPM," as well as $\hat{\Lambda}_{SM}(t)$ and $\hat{\Lambda}_{TK}(t)$ from Figure 4. We observe that the four estimates are close to each other. In particular, $\hat{\Lambda}_{SM}(t)$ (and $\hat{\Lambda}_{TK}(t)$) are very similar to "SPM" for the first 6 months, then diverge somewhat thereafter. Straightforward calculation shows that if (17) is correct and date of sale x_i is independent of u_i (as is supported by the survey data), then (in the discrete-time case)

$$E\{\hat{\Lambda}_{SM}(t)\} = \sum_{s=1}^{t} \lambda_0(s)E\left\{\exp(\beta u_i)|u_i \le \frac{120}{s}\right\}, \tag{19}$$

rather than (18). Investigation of (19) indicates that for s under 6 months, $E\{\exp(\beta u_i)|u_i \le 120/s\} \doteq E\{\exp(\beta u_i)\}$, because most mileage rates are in the range of 6,000–24,000 miles per year. Thus for this example, $\hat{\Lambda}_{SM}(t)$ and $\hat{\Lambda}_{TK}(t)$ provide reasonable estimates of $E\{N_i(t)\}$, even though the τ_i's are not strictly independent of the claims processes.

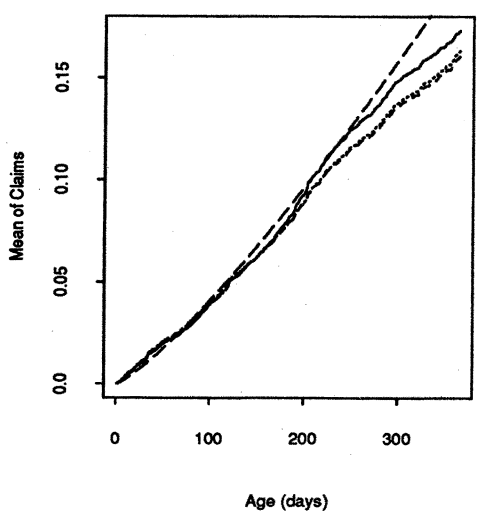

Figure 5. Diagnostic Check—Comparison of the Estimates Based on Different Models. —— SPM; ··· NTK; --- NSM; –– FPM.

7. CONCLUSION

This article deals with the estimation of rate functions $\lambda(t)$ or mean functions $\Lambda(t)$ for recurrent events when the data are zero-truncated. If the events follow a Poisson process, then the truncated data likelihood L_T of Section 2 is available, but simulations and large-sample calculations show that the MLE $\hat{\Lambda}_T(t)$ can be badly biased when events do not follow a Poisson process. But in many applications there is information about the distribution of observation times (τ_i's) as well as the total number of the units, and in this case the paper shows (i) that the MLE based on the Poisson likelihood L_{TK} (6) is reasonably robust under heavy truncation, and (ii) that the estimator $\hat{\Lambda}_{SM}(t)$ introduced in Section 4.1 is both robust against Poisson departures generally and very efficient under the Poisson model.

Several further investigations would be worthwhile. One of theoretical and practical interest is the determination of asymptotic variances and variance estimates for $\hat{\Lambda}_T(t)$ and $\hat{\Lambda}_{TK}(t)$ under a continuous-time Poisson model. Another is robust estimation of $\Lambda(t)$ when there is limited supplementary information about $G(\tau)$.

This article has not dealt in any detail with the estimation or assessment of full models for the event processes but has instead focused on rate and mean function estimation. Further study of model fitting and assessment with zero-truncated data would be valuable, especially in situations where the truncation times may not be independent of the event processes. Investigation of this topic is in progress. Finally, covariates may be introduced into the models used here. The use of a covariate to assess the independence of observation times was illustrated briefly in Section 6.

Estimation for fully parametric regression models is reasonably straightforward, but semiparametric models need further study.

APPENDIX A: ASYMPTOTICS FOR $\hat{\Lambda}_{SM}(t)$

We give the proof in a format that applies to discrete time and continuous time.

Theorem 1. Suppose that $N_i(t), i = 1, \ldots, M$ are independent counting processes with $E\{N_i(t)\} = \Lambda(t) = \int_0^t \lambda(s)\,ds$, where $\lambda(t)$ is right-continuous and has left-hand limits, and that observation times τ_i are iid random variables with distribution $G(\cdot)$, independent of $N_i(t)$. If $\mathrm{cov}(dN_i(t), dN_i(s)) = c(t,s)dtds$ exists and is integrable for $s, t \in (0, \tau^*]$, where $\tau^* = \sup\{s\colon \bar{G}(s) > 0\} < \infty$ and $\bar{G}(\tau^*) > 0$ with $\bar{G}(s) = \Pr\{\mathcal{T} \geq s\}$, then

$$\lim_{M\to\infty} \hat{\Lambda}_{SM,M}(t) = \Lambda(t), \text{a.s. } \forall\, t \in (0, \tau^*] \tag{A.1}$$

and

$$\sqrt{M}(\hat{\Lambda}_{SM,M}(t) - \Lambda(t)) \xrightarrow{d} \mathcal{G}_{SM}(t), \text{as } M \to \infty, \tag{A.2}$$

where $(\mathcal{G}_{SM}(t), t \in (0, \tau^*])$ a Gaussian process with mean zero and covariance process $\mathrm{cov}(\mathcal{G}_{SM}(t), \mathcal{G}_{SM}(s))$, which is given later.

Proof. Note that, for all $t \in (0, \tau^*]$,

$$\hat{\Lambda}_{SM,M}(t) = \frac{1}{M}\sum_{i=1}^{M} X_i^{(SM)}(t)$$

with $X_i^{(SM)}(t) = \int_0^t \delta_i(s)/\bar{G}(s)\,dN_i(s), i = 1, \ldots, M$, are iid random variables. Moreover,

$$E\{X_i^{(SM)}(t)\} = \Lambda(t),$$

and

$$\begin{aligned}
&\mathrm{cov}(X_i^{(SM)}(t), X_i^{(SM)}(s)) \\
&= \int_0^t \int_0^s \frac{\bar{G}(u \vee v)}{\bar{G}(u)\bar{G}(v)}\, c(u,v)\,du\,dv \\
&\quad + \int_0^t \int_0^s \left[\frac{\bar{G}(u \vee v)}{\bar{G}(u)\bar{G}(v)} - 1\right] \lambda(u)\lambda(v)\,du\,dv,
\end{aligned}$$

denoted by $\int_0^t \int_0^s c_{SM}(u,v)\,du\,dv$. By the strong law of large numbers, we have

$$\hat{\Lambda}_{SM,M}(t) \xrightarrow{\text{a.s.}} \Lambda(t)$$

for all $t \in (0, \tau^*]$. Also for all $\alpha, \beta > 0$, by the central limit theorem,

$$\begin{aligned}
&\alpha\sqrt{M}(\hat{\Lambda}_{SM,M}(t) - \Lambda(t)) + \beta\sqrt{M}(\hat{\Lambda}_{SM,M}(s) - \Lambda(s)) \\
&= \frac{1}{\sqrt{M}} \sum_{i=1}^{M} [\alpha(X_i^{(SM)}(t) - \Lambda(t)) + \beta(X_i^{(SM)}(t) - \Lambda(t))] \xrightarrow{d} \\
&\quad N(0, \sigma_{SM}^{\alpha,\beta}(t,s)),
\end{aligned}$$

where $\sigma_{SM}^{\alpha,\beta}(t,s) = \alpha^2 \int_0^t \int_0^t c_{SM}(u,v)\,du\,dv + 2\alpha\beta \int_0^t \int_0^s c_{SM}(u,v)\,du\,dv + \beta^2 \int_0^s \int_0^s c_{SM}(u,v)\,du\,dv$. The tightness of $\{\sqrt{M}(\hat{\Lambda}_{SM,M}(t) - \Lambda(t))\}$ can be shown by noting

$$\begin{aligned}
\Pr\{|Y_M(t) - Y_M(t_1)| &\geq \gamma, |Y_M(t_2) - Y_M(t)| \geq \gamma\} \\
&\leq \frac{1}{\gamma^2}\,[F(t_2) - F(t_1)]^2,
\end{aligned}$$

for $0 < t_1 \leq t \leq t_2 \leq \tau^*, \gamma > 0$ and $M \geq 1$, where $Y_M(t) = \sqrt{M}(\hat{\Lambda}_{SM,M}(t) - \Lambda(t))$ and $F(t) = \int_0^t \sqrt{c_{SM}(u,u)}\,du$, a nondecreasing, continuous function on $[0, \tau^*]$ (Billingsley 1968, chap. 3). Therefore, (A.2) holds with

$$\mathrm{cov}(\mathcal{G}_{SM}(t), \mathcal{G}_{SM}(s)) = \int_0^t \int_0^s c_{SM}(u,v)\,du\,dv.$$

Corollary 1. Under the conditions in the theorem, $\sqrt{M}(\hat{\Lambda}_{SM,M}(t) - \Lambda(t)) \xrightarrow{d} N(0, \sigma_t^{(SM)^2})$, for all $t \in (0, \tau^*])$, and $\sigma_t^{(SM)^2}$ is estimated consistently by

$$\hat{V}_{SM}(t) = \frac{1}{M}\sum_{i=1}^{M} \left\{ \int_0^t \frac{1}{\bar{G}(s)} \left[\delta_i(s)\,dN_i(s) - \frac{dN.(s)}{M}\right]\right\}^2.$$

The consistent variance estimate can be obtained by noting

$$\begin{aligned}
&\mathrm{var}(\hat{\Lambda}_{SM,M}(t)) \\
&\quad = \frac{1}{M}\int_0^t \int_0^t \frac{1}{\bar{G}(u)\bar{G}(v)}\, \mathrm{cov}\{\delta_i(u)\,dN_i(u), \delta_i(v)\,dN_i(v)\},
\end{aligned}$$

and that $\mathrm{cov}\{\delta_i(u)dN_i(u), \delta_i(v)dN_i(v)\}$ is estimated by

$$\frac{1}{M}\sum_{i=1}^{M} \left[\delta_i(u)dN_i(u) - \frac{dN.(u)}{M}\right]\left[\delta_i(v)dN_i(v) - \frac{dN.(v)}{M}\right]$$

for $u, v \in (0, \tau^*]$.

APPENDIX B: POISSON PROCESSES WITH PIECEWISE CONSTANT INTENSITIES

Consider the piecewise constant intensity function

$$\lambda(t) = \lambda_j \quad \text{for} \quad a_{j-1} < t \leq a_j,$$

where $a_0 = 0 < a_1 < \cdots < a_K \leq \infty$ are specified. Considering such models with K fairly small (say in the 3–10 range) provides enough flexibility to model most practical situations and allows the easy calculation of estimates and standard errors, based on either of the likelihood functions L_T or L_{TK}. Using L_T given by (1), we obtain the maximum likelihood equations

$$\frac{\partial l_T}{\partial \lambda_j} = \frac{n_{\cdot j}}{\lambda_j} - \sum_{i=1}^{m} \frac{w_{ij}}{1 - \exp[-\Lambda(\tau_i)]}, \qquad j = 1, \ldots, K,$$

where

$$w_{ij} = \frac{\partial \Lambda(\tau_i)}{\partial \lambda_j} = I(\tau_i > a_{j-1})[(\tau_i \wedge a_j) - a_{j-1}].$$

The observed information matrix has entries

$$-\frac{\partial^2 l_T}{\partial \lambda_j \partial \lambda_l} = I(j = l)\,\frac{n_{\cdot j}}{\lambda_j^2} - \sum_{i=1}^{m} \frac{w_{ij} w_{il} \exp[-\Lambda(\tau_i)]}{\{1 - \exp[-\Lambda(\tau_i)]\}^2}.$$

The equations $\partial l_T/\partial \lambda_j = 0$ $(j = 1, \ldots, K)$ are readily solved to yield the estimates $\hat{\lambda}_j$'s, and the inverse of the observed information matrix evaluated at $\hat{\lambda}_1, \ldots, \hat{\lambda}_K$ provides asymptotic variance and covariance estimates for the $\hat{\lambda}_j$'s.

Using L_{TK} given by (6), we get

$$\frac{\partial l_{TK}}{\partial \lambda_j} = \frac{n_{\cdot j}}{\lambda_j} - \sum_{i=1}^{m} w_{ij} - (M - m)\,\frac{A_j}{A_0},$$

where $A_0 = \int_0^\tau \exp[-\Lambda(u)]\,dG(u)$ and $A_j = -\partial A_0/\partial \lambda_j, j = 1, \ldots, K$. The observed information matrix has entries

$$-\frac{\partial^2 l_{TK}}{\partial \lambda_j \partial \lambda_l} = I\,(j = l)\,\frac{n_{\cdot j}}{\lambda_j^2} - (M - m)\,\frac{A_{jl}A_0 - A_j A_l}{A_0^2},$$

where $A_{jl} = -\partial A_j/\partial\lambda_l = \int_0^\tau I(u > a_{j-1})[(u \wedge a_j) - a_{j-1}]I(u > a_{l-1})[(u \wedge a_l) - a_{l-1}]\exp[-\Lambda(u)]\,dG(u)$. As with L_T, estimates $\hat{\lambda}_1, \ldots, \hat{\lambda}_K$ and variance estimates are readily obtained.

[Received April 1994. Revised May 1995.]

REFERENCES

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993), *Statistical Models Based on Counting Processes*, New York: Springer-Verlag.

Billingsley, P. (1968), *Convergence of Probability Measures*, New York: John Wiley.

Cox, D. R., and Lewis, P. A. W. (1966), *The Statistical Analysis of Series of Events*, London: Methuen.

Crow, L. H. (1979), "Reliability Analysis for Complex, Repairable Systems," in *Reliability and Biometry*, eds. F. Proschan and R. J. Serfling, Philadelphia: SIAM, pp. 379–410.

Hu, X. J. (1995), "Estimation From Truncated Data With Supplementary Information, With Application to Field Reliability," Ph.D. thesis, University of Waterloo, Dept. of Statistics and Actuarial Science.

Kalbfleisch, J. D., and Lawless, J. F. (1988), "Estimation of Reliability From Field Performance Studies" (with discussion), *Technometrics*, 30, 365–388.

——— (1991), "Regression Models for Right-Truncated Data With Applications to AIDS Incubation Times and Reporting Lags," *Statistica Sinica*, 1, 19–32.

Kalbfleisch, J. D., Lawless, J. F., and Robinson, J. A. (1991), "Methods for the Analysis and Prediction of Warranty Claims," *Technometrics*, 33, 273–285.

Lawless, J. F. (1987), "Regression Methods for Poisson Process Data," *Journal of the American Statistical Association*, 82, 808–815.

Lawless, J. F., and Kalbfleisch, J. D. (1992), "Some Issues in the Collection and Analysis of Field Reliability Data," in *Survival Analysis: State of the Art*, eds. J. P. Klein and P. K. Goel, Dordrecht: Kluwer, pp. 141–152.

Lawless, J. F., and Nadeau, J. C. (1995), "Some Simple Robust Methods for the Analysis of Recurrent Events," *Technometrics*, 37, 158–168.

Nelson, W. B. (1988), "Graphical Analysis of System Repair Data," *Journal of Quality Technology*, 20, 24–35.

Sun, J., and Kalbfleisch, J. D. (1993), "The Analysis of Current Status Data on Point Processes," *Journal of the American Statistical Association*, 88, 1449–1454.

Suzuki, K. (1985a), "Estimation Method of Lifetime Based on the Record of Failures During the Warranty Period," *Journal of the American Statistical Association*, 80, 68–72.

——— (1985b), "Nonparametric Estimation of Lifetime Distribution from Incomplete Field Data," *Technometrics*, 27, 263–271.

——— (1987), "Analysis of Field Failure Data from a Nonhomogeneous Poisson Process," *Reports on Statistical Applications in Research*, 34, 6–15.

——— (1993), "Estimation of Lifetime Distribution Using the Relationship of Calendar Time and Usage Time," *Reports on Statistical Applications in Research*, 40, 10–22.

Suzuki, K., and Kasashima, T. (1993), "Estimation of Lifetime Distribution from Incomplete Field Data with Different Observational Periods," Technical Report UEC-CAS-93-02. University of Electro-Communications, Dept. of Communications and Systems Engineering, Japan.

Thall, P. F., and Lachin, J. M. (1988), "Analysis of Recurrent Events: Nonparametric Methods for Random-Interval Count Data," *Journal of the American Statistical Association*, 83, 339–347.

Turnbull, B. W. (1976), "The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data," *Journal of the Royal Statistical Society*, Ser. B, 38, 290–295.

Wang, M. C. (1989), "A Semiparametric Model for Randomly Truncated Data," *Journal of the American Statistical Association*, 84, 742–748.

——— (1992), "The Analysis of Retrospectively Ascertained Data in the Presence of Reporting Delays," *Journal of the American Statistical Association*, 87, 396–406.

Wang, M. C., Jewell, N. P., and Tsai, W. Y. (1986), "Asymptotic Properties of the Product Limit Estimate Under Random Truncation," *The Annals of Statistics*, 14, 1597–1605.