Brazilian Journal of Probability and Statistics
Volume 22, Number 1, Pages 23–33
ISSN 0103-0752

INTENSITY MODELS FOR PARAMETRIC ANALYSIS OF RECURRENT EVENTS DATA

FRANCISCO LOUZADA-NETO

ABSTRACT. In this paper we discuss a general parametric class of models for analyzing multiple failure data where the study subjects can experience repeated events. The idea is to model the global time, combining the two time scales, total time and interval time (intervals between successive events), and the event counts in a hybrid model, and to decide about their appropriateness in the light of the data. The general framework allows a broad class of hazard or intensity models including the Poisson and renewal process models as very special cases. Simulation results suggest that the test procedures perform well even for small and moderate sample sizes. A Real dataset illustrate the methodology.

1. Introduction

Lifetime data where more than one event is observed on each subject arise in areas such as biomedical studies, criminology, demography, manufacturing and industrial reliability. An offender may be convicted several times. Several tumors may be observed for an individual. Recurrent pneumonia episodes arise in patients with human immunodeficiency syndrome. A piece of equipment may experience repeated failures or warranty claims.

For this kind of data, for each individual i we observe the total number, m_i, of events (lifetimes) occurred over the time period $(0, \tau_i]$, the ordered epochs of the m_i lifetimes at times $0 \leq t_{i1} < t_{i2} < \cdots < t_{im_i} \leq \tau_i$, and, additionally, we may have covariate information on each subject defined by a vector z and a vector of censoring indicators. In such studies, interest lies on understanding and characterizing the event occurrence process for individual subjects and on treatment comparisons based on the time to each distinct event, the number of events, the type of events and the interdependencies between events, explaining the nature of variation between subjects in terms of treatments, fixed covariates or other factors (maybe unobservable ones).

Approaches often used to model recurrent event data, which allow us to learn about an individual process, are those based on Poisson and renewal processes. A Poisson representation for the individual process is employed for modelling the total time on study, while a renewal process is employed for modelling the interval time, that is, the time from

Key words and phrases. Intensity models, maximum likelihood estimation, multivariate lifetime data, Poisson process, renewal process, simulation.
Received October/2004, May/2007.

previous event (Cox and Isham, 1980, Cox and Lewis, 1966, Andersen, Borgan, Gill and Keiding, 1993, Lawless, 1987, Follmann and Goldberg, 1988, Prentice and Williams and Peterson, 1981).

Lipschutz and Snapinn (1997) discussed that the total time modelling seems natural if the events can be viewed as unrelated, and therefore independent, so it is reasonable to assume that they can be developing simultaneously, with the risk for the occurrence of each event beginning at the same time. In this context, the ordered events simply reflect the order of their occurrence in time. For example, patients with human immunodeficiency syndrome can experience recurrence of opportunistic infections. They may be at risk for a type of infection, such as superficial candidiasis, at several sites from the beginning of the study, with the first, second and subsequent events reflecting the order of their occurrence as time progress. Interval time modelling, however, is appropriate for situations where the risk for the next event does not begin until after the previous event has occurred, such as myocardial infarction, superficial candidiasis at a single site and warranty claims on a particular system. The parameters of the model reflect the relative risk of the next event from the time of the previous event. Among others, Cox (1972b) and Lawless and Thiagarajah (1996) considered modulated renewal and Poisson processes, which accommodate the two time scales.

In this paper we discuss a general parametric hybrid scale intensity model for analyzing multiple failure data for use in studies where individuals can experience recurrent events. The general framework accommodate a broad class of intensity models including the Poisson and renewal process models as very special cases. The idea is to combine the two time scales, total time and interval time (intervals between successive events), and the event counts in a hybrid model, and to decide about their appropriateness in the light of the data. In the model with covariates, we assume a proportional intensity baseline function. We envisage applications in which a moderate or large number of individuals is observed and the number of events per individual may be quite small. A brief review of intensity models is given in Section 2. In Section 3 we define the hybrid scale intensity model. A very special case of the model is given in Section 4 where we discuss the estimation procedure and computational issues. In Section 5 we present the results of a simulation study on the error rates for hypothesis tests and on quantification of the gain in the precision of parameter estimates obtained by simplifying the model when a possibly small or moderate number of recurrent events per individual is observed. A real numerical example is provided in Section 6. Some final remarks in Section 7 conclude the paper.

2. Model formulation

Recurrent event data can be conveniently modelled by considering intensity-based models. Suppose that n individuals may experience a single type of recurrent event. Let $m_i(t)$ denote the number of events occurring for the ith individual over $(0, t]$. Assume that the ith individual is observed over the interval $(0, \tau_i]$, where τ_i is determined independently of $m_i(t)$. Let $t_{i1} \leq t_{i2} \leq \cdots$ denote the continuous failure times for the ith individual, and let $x_{ij} = t_{ij} - t_{i,j-1}$ be the time intervals between successive events, with $t_{i0} = 0$. For simplicity, we will consider an arbitrary individual and drop the subscript i.

The intensity function at time t is defined as (Cox and Isham, 1980, p. 9; Lawless, 1995)

$$h\{t \mid \mathcal{M}(t)\} = \lim_{\Delta t \to 0} \frac{\text{pr}\{m(t, t + \Delta t) = 1 \mid \mathcal{M}(t)\}}{\Delta t}, \tag{2.1}$$

where $m(t, t + \Delta t)$ denotes the number of events over the small interval $[t, t + \Delta t)$, and $\mathcal{M}(t) = \{m(s) : s < t\}$ denotes the history of the process up to time t.

Equation (2.1) represents the instantaneous rate of failure at time t given the history of the process $\mathcal{M}(t)$ up to time t. The failure process will be assumed orderly, so that the limiting probability of two or more failures in the interval $[t, t + \Delta t)$, given that at least one failure occurs in it, tends to zero, as $\Delta t \to 0$.

Several intensity-based models with varying degrees of memory can be established as particular cases of (2.1). We shall divide them in two types; those which depend primarily on the total time t, and those which depend primarily on the interval time x. Important total-time type models are the nonhomogeneous Poisson and pure birth process, while important interval-time type models are the renewal process and its natural generalization to a semi-Markov process.

The intensity function for a nonhomogeneous Poisson process is specified by considering, from (2.1),

$$h\{t \mid \mathcal{M}(t)\} = h_0(t), \tag{2.2}$$

where $h_0(t) \geq 0$ (Lawless, 1982, p. 495). That is, the dependence is only on elapsed time t. For

$$h\{t \mid \mathcal{M}(t)\} = m(t)h_0(t), \tag{2.3}$$

we obtain a nonhomogeneous pure birth process. A renewal process is obtained for $h\{t \mid \mathcal{M}(t)\} = h_0(t - t_{m(t)})$, which can be rewritten as

$$h\{t \mid \mathcal{M}(t)\} = h_0(v_t), \tag{2.4}$$

where $v_t = t - t_{m(t)}$ is the backward recurrence time from t to the previous event (or the time origin). A particular semi-Markov process is obtained if, in (2.4), $h_0(\cdot)$ is replaced by $h_{0j_t}(\cdot)$, with $j_t = m(t) = 1, 2, \ldots$, where j_t is the jth point (event), so that an individual moves to stratum j_t following his $(j_t - 1)$th failure and remains there until the j_tth failure or censoring takes place (Prentice, Williams and Peterson, 1981).

It is natural to specify regression models for (2.1). Following Cox (1972a, 1972b) we can consider (2.1) to be the product of an nonnegative arbitrary function of time and a nonnegative function of covariates. From (2.2) the nonnhomogeneous Poisson process model with covariates is defined as

$$h\{t \mid z, \mathcal{M}(t)\} = h_0(t)g(\beta^T z), \tag{2.5}$$

where $g(\cdot)$ is a known positive function that equals one when its argument is zero, β is a vector of unknown regression parameters, and $h_0(t) \geq 0$ denotes the baseline intensity function for an individual with $z = 0$.

In the same way, from (2.4), the renewal process model with covariates is defined as

$$h\{t \mid z, \mathcal{M}(t)\} = h_0(v_t)g(\beta^T z), \tag{2.6}$$

where $h_0(v_t) \geq 0$ denotes the baseline intensity function.

Models (2.5) and (2.6) are semiparametric if $h_0(\cdot)$ is left arbitrary, and fully parametric if $h_0(\cdot)$ is specified up to a vector of parameters, θ. A variety of functional forms can be often employed for $g(\cdot)$, but the simplest and most natural is $\exp(\cdot)$. More general expressions for model (2.6) are obtained by setting different baseline intensity functions to each event j_t, so that $h_0(\cdot)$ is replaced by $h_{0j_t}(\cdot)$.

3. A general class of intensity model

A general class of intensity model which allows for multiple time scales in the baseline intensity function, $h_0(\cdot)$, and permit dependency on the event counts can be defined as

$$h\{t \mid z, \mathcal{M}(t)\} = q_1(v_t; \theta_1) q_2(t; \theta_2) q_3(j_t; \theta_3) g(\beta^T z), \tag{3.1}$$

where $g(\cdot)$ and β are defined as before, and $q_1(\cdot)$, $q_2(\cdot)$, and $q_3(\cdot)$ are positive functions denoting the parametric baseline intensity functions on the interval time, v_t, total time, t, and event counts, j_t, respectively, with unknown parameter vectors θ_1, θ_2 and θ_3. The covariates are assumed to be fixed and therefore not affected by the event process. For simplicity, hereafter v and j will refer to v_t and j_t, respectively.

Model (3.1) covers a wide spectrum of intensity-based models. For instance, (3.1) reduces to the nonhomogeneous Poisson process model (2.5) in its parametric version if $q_1(\cdot) = q_3(\cdot) = 1$. For $q_2(\cdot) = q_3(\cdot) = 1$ we obtain the renewal process model (2.6). A hybrid Poisson/renewal intensity model is obtained if $q_3(\cdot) = 1$. More general models can be obtained with a more complex degree of memory structure, particularly if different baseline intensity functions are set for the j-th event and for time-dependent covariates.

From the practical point of view, an advantage of modelling the total time jointly with the intervals, as in (3.1), is that we can accommodate situations in which the process is regarded as a superposition of two no observable processes, one depending primarily on the total time t and one depending primarily on the interval time x. Also, we can accommodate situations where the basic process is, for instance, a Poisson process which is perturbed (multiplicatively) by a renewal one and vice-versa.

Noticing that we can rewrite $t = x_1 + \cdots + x_{j-1} + x = t_{j-1} + v$, the corresponding cumulative intensity function is given by

$$H\{t; \theta_1, \theta_2, \theta_3, \beta \mid z, \mathcal{M}(t)\} = H_0\{t; \theta_1, \theta_2, \theta_3 \mid \mathcal{M}(t)\} g(\beta^T z), \tag{3.2}$$

where $H_0\{t; \theta_1, \theta_2, \theta_3 \mid \mathcal{M}(t)\} = \int_0^t h_o\{a; \theta_1, \theta_2, \theta_3 \mid \mathcal{M}(a)\}\, da$ is the cumulative baseline function, with $h_0\{t; \theta_1, \theta_2, \theta_3 \mid \mathcal{M}(t)\} = q_1(v; \theta_1) q_2(t; \theta_2) q_3(j; \theta_3)$ denoting the baseline intensity function. In general, the integral in $H_0(\cdot)$ needs to be evaluated numerically, since only in special cases analytical solutions are feasible.

Assuming that each failure time t_j has an associated indicator variable defined by $\delta_j = 1$ if t_j is an observed failure time and $\delta_j = 0$ if t_j is a right-censored observation, and that individual i has an associated covariate vector z_i, the contribution to the likelihood from an individual's interval time, x_j, which starts at t_{j-1}, is

$$L_{j|j-1,\ldots,1} = \{h_j(t_j \mid z, \mathcal{M}(t_j))\}^{\delta_j} e^{-H_j(t_j|z,\mathcal{M}(t_j))}, \tag{3.3}$$

where $h_j(t_j \mid z, \mathcal{M}(t_j))$ and $H_j(t_j \mid z, \mathcal{M}(t_j))$ are the intensity and cumulative intensity functions over the period of time from t_{j-1}.

Suppose that n individuals may experience a total of $m_1, \ldots, m_n$ recurrent events. From (3.3), given a specified parametric model for the intensity function $h_j(t_j \mid z, \mathcal{M}(t_j))$, the maximum likelihood estimates (MLEs) can be obtained by direct maximization of the overall likelihood

$$L(\theta_1, \theta_2, \theta_3, \beta) = \prod_{i=1}^{n} \prod_{j=1}^{m_i} \{h_{ij}(t_{ij} \mid z_i, \mathcal{M}(t_{ij})\}^{\delta_{ij}} e^{-H_{ij}(t_{ij} \mid z_i, \mathcal{M}(t_{ij}))}. \tag{3.4}$$

We discuss the estimation procedure further in the next section.

Although it will not be used in the paper latter, it is interesting to note, following Lawless (1987), that (3.3) can be decomposed as

$$L_{j|j-1,\ldots,1} = L_j^{(1)}(\theta_1, \theta_2) L_j^{(2)}(\theta_1, \theta_2, \theta_3, \beta) \tag{3.5}$$

where, from (3.1) and (3.2),

$$L_j^{(1)}(\theta_1, \theta_2) = \{\frac{q_1(a; \theta_1) q_2(a; \theta_2)}{\int_0^{x_j} q_1(a; \theta_1) q_2(a; \theta_2)\, da}\}^{\delta_j} \tag{3.6}$$

and

$$\begin{aligned} L_j^{(2)}(\theta_1, \theta_2, \theta_3, \beta) =& \{g(\beta^T z) q_3(j; \theta_3) \int_0^{x_j} q_1(a; \theta_1) q_2(a; \theta_2)\, da\}^{\delta_j} \\ & \exp\{-g(\beta^T z) q_3(j; \theta_3) \int_0^{x_j} q_1(a; \theta_1) q_2(a; \theta_2)\, da\}. \end{aligned} \tag{3.7}$$

Thus, there is the possibility of obtaining the MLEs $\hat{\theta}_1$ and $\hat{\theta}_2$ by maximizing $L^{(1)} = \prod_i \prod_j L_j^{(1)}(\cdot)$ without having to consider θ_3 and β, and then estimating $\hat{\theta}_3$ and $\hat{\beta}$ by maximizing $L^{(2)} = \prod_i \prod_j L_j^{(2)}(\cdot)$ with θ_1 and θ_2 fixed at $\hat{\theta}_1$ and $\hat{\theta}_2$. Although this procedure may yield poor estimates, the estimates produced can often provide a good starting point for any iteration procedure designed to maximize (3.4).

Further, focusing on (3.7), if all individuals are observed over the same time period, say $(0, \tau]$, and there are no censored lifetimes, we can decompose $L^{(2)} = \prod_i \prod_j L_j^{(2)}(\cdot)$ as

$$L^{(2)} = L^{(3)}(\beta) L^{(4)}(\theta_1, \theta_2, \theta_3, \beta), \tag{3.8}$$

where

$$L^{(3)}(\beta) = \prod_{i=1}^{n} \{\frac{g(\beta^T z_i)}{\sum_{i=1}^{n} g(\beta^T z_i)}\}^{m_i}, \tag{3.9}$$

which is the Cox likelihood for this situation, which depends only the counts m_i, and

$$\begin{aligned} L^{(4)}(\theta_1, \theta_2, \theta_3, \beta) =& \{q_3(j; \theta_3) \int_0^{\tau} q_1(a; \theta_1) q_2(a; \theta_2)\, da \sum_{i=1}^{n} g(\beta^T z_i)\}^{m} \\ & \exp\{-q_3(j; \theta_3) \int_0^{\tau} q_1(a; \theta_1) q_2(a; \theta_2)\, da \sum_{i=1}^{n} g(\beta^T z_i)\}. \end{aligned} \tag{3.10}$$

4. A special parametric intensity model

To examine the fully parametric approach we shall consider a particular flexible parametrization of (3.1), where $q_1(v;\theta_1) = q_1(v;\alpha,\gamma) = \alpha\gamma(\alpha v)^{\gamma-1}$, $q_2(t;\theta_2) = q_2(t;\alpha,\phi) = (1+\alpha\phi t)$, $q_3(j;\theta_3) = q_3(j;\psi) = \psi^{j-1}$ and $g(\beta^T z) = \exp(\beta^T z)$. That is, we shall consider the intensity function

$$h\{t;\alpha,\beta,\gamma,\phi,\psi \mid z, \mathcal{M}(t)\} = \alpha\gamma(\alpha v)^{\gamma-1}(1+\alpha\phi t)\psi^{j-1}e^{\beta^T z}, \tag{4.1}$$

where $\alpha,\ \gamma,\ \phi,\ \psi > 0$. We can have $\phi < 0$ provided that $1+\alpha\phi t > 0$ for any potentially observed value of t. We will further assume that $\beta^T z = \beta_1 z_1 + \beta_2 z_2 + \cdots$ has no intercept term, which is absorbed by α.

An advantage of this parametrization is its relatively easy interpretation. While γ, ϕ and ψ are dimensionless numbers, α denotes the exchange rate between the time scales (total and interval times). Moreover, the renewal component, $q_1(\cdot)$, is driven by a Weibull-type model, while the Poisson component, $q_2(\cdot)$, works as a time dependent Poisson process part. The event count function, $q_3(\cdot)$, penalizes large numbers of events for $j > 1$. An exponentially proportional covariate effect $g(\cdot)$ completes the formulation. Apart from β and α, the three parameters γ, ϕ and ψ represent departures from the Poisson intensity model, but often there would be some indication on general grounds that some of the parameters should be omitted.

Several important intensity models can be obtained as particular cases of (4.1). For instance, for $\phi = 0$ and $\psi = 1$, it is an ordinary Weibull renewal model for the interval times. It is a special nonhomogeneous Poisson process model if $\gamma = \psi = 1$. With $\gamma = \psi = 1$ and $\phi = 0$ it reduces to an ordinary homogeneous Poisson process model.

An important measure in survival studies is the relative risk between pacients on two different covariate levels, say, z_1 and z_2. Although two time scales are involved in the intensity function (4.1), in the light of its proportional structure, the relative risk betwen pacients on two different covariate levels z_1 and z_2 is given by

$$\frac{h\{t;\alpha,\beta,\gamma,\phi,\psi \mid z_1, \mathcal{M}(t)\}}{h\{t;\alpha,\beta,\gamma,\phi,\psi \mid z_2, \mathcal{M}(t)\}} = e^{\beta^T(z_1-z_2)}. \tag{4.2}$$

Following (3.4) and (4.1), the contribution of an interval time x_j starting at t_{j-1} from an arbitrary individual to the log-likelihood is

$$\begin{aligned} l_{j|j-1,\ldots,1} =& \delta_j\{\beta^T z + \log\gamma + \gamma\log\alpha + (\gamma-1)\log x_j + \log(1+\alpha\phi t_j) + (j-1)\log\psi\} \\ & - \{1+\alpha\phi t_{j-1} + \alpha\phi\frac{\gamma}{\gamma+1}x_j\}(\alpha x_j)^{\gamma}\psi^{j-1}e^{\beta^T z}. \end{aligned} \tag{4.3}$$

In our experience, the log-likelihood

$$l = \sum_{i=1}^{n}\sum_{j=1}^{m_i} l_{i,j|j-1,\ldots,1} \tag{4.4}$$

is straightforward to be maximized using a standard routine, such as `nlmin` in the package `S-Plus`, which finds a local maximum of a function using a general quasi-Newton method (Seber and Wild, 1989, ch. 13). It may be appropriate to consider reparametrization. We avert numerical problems from parameters with unbounded ranges.

Large-sample inference for the parameters can be based on the MLEs and their estimated standard errors, or, preferably, on the profile likelihood, the latter being invariant under reparametrization. Further, asymptotic approximations to the likelihood ratio statistics (LRS) distribution are likely to be more accurate in small or moderate samples than are asymptotic approximations to the distribution of the MLEs. Formal goodness-of-fit tests are feasible in principle. Model (4.1) is a nested-type model, and it provides an easy way to test whether or not a particular case fits a dataset by fitting special models and comparing the fit with the saturated model. For instance, considering (4.1), we can use LRS for testing goodness-of-fit of hypotheses such as $H_0 : \psi = 1$, $H_0 : \phi = 0$, $H_0 : \gamma = 1$, $H_0 : \psi = 1, \phi = 0$, $H_0 : \psi = 1, \gamma = 1$, $H_0 : \phi = 0, \gamma = 1$ and $H_0 : \psi = 1, \phi = 0, \gamma = 1$, which postulate the special cases of (4.1). Although it is well known that, under general grounds, the LRS test will work well if there is no boundary problems on the hypothesis test, which is our case, the adequacy of these procedures for small and moderate sample sizes is studied via simulation in the next section.

5. A simulation study

It is common in practice to find studies with a possibly small or moderate number of recurrent events per individual on a moderate number of individuals. In order to assess the applicability of the asymptotic results in such situations a simulation study was performed to estimate the error rates of the hypothesis tests. The study was based on generated samples of unit exponential random variables, assuming that each one of n individuals experienced the same number of recurrent events $m = 2, 4, 7, 15, 30$ with $n = 20, 50, 100$. In addition to setting the parameters at the null point $\psi = 1, \phi = 0, \gamma = 1$ and at $\alpha = 1$, we also set $\beta = 0.7$ and considered a dichotomous covariate z equals to -1 and 1, indexing a control group and a treatment group, respectively. The same number of individuals, $n/2$, were considered at each covariate level. A case study is defined by n and m, and by one of the three different hypothesis tests, which will be set up below. Thus, 45 different cases were simulated, each with $1,000$ samples.

For survival data, it is definitively important to address the impact of censoring. Then, the overall study described above was repeated with righ-censored samples with $m = 4, 7, 15, 30$. For $m = 4$ we consider one censored observation, for $m = 7$ we consider two censored observations, for $m = 15$ we considered four censored observation and for $m = 30$ we considered nine censored observations, which correspond approximately to 25, 30, 30 and 30 percent of censoring.

Considering (4.1), we use the LRS to test the following hypothesis: **(a)** $H_0 : \psi = 1, \phi = 0, \gamma = 1$ against $H_1 : \psi \neq 1, \phi = 0, \gamma = 1$, **(b)** $H_0 : \psi \neq 1, \phi = 0, \gamma = 1$ against $H_1 : \psi \neq 1, \phi \neq 0, \gamma = 1$, and **(c)** $H_0 : \psi \neq 1, \phi \neq 0, \gamma = 1$ against $H_1 : \psi \neq 1, \phi \neq 0, \gamma \neq 1$. The LRS was treated as a chi-squared distribution with 1 degree of freedom. Table 1 presents the empirical rejection rates for different numbers m of recurrent events per individual and different number n of individuals. The empirical rejection rates are close to the nominal rejection rate 0.05. For small values of m and n however the empirical rejection rates are within about 0.05 of the nominal. The censoring effect is to present estimate empirical rejection rates within about 0.10 of the nominal.

TABLE 1. Confidence intervals for β_1: coverages (%) and lengths; unbalanced design and skewed errors.

For $m = 4, 7, 15$ and 30, in each cell the left result correspond to the complete samples while the righ result correspond to the censored samples.

Test	n	m				
		2	4	7	15	30
	20	0.082	0.079/0.122	0.062/0.103	0.055/0.098	0.056/0.078
(a)	50	0.076	0.066/0.097	0.054/0.092	0.054/0.084	0.053/0.071
	100	0.058	0.053/0.079	0.051/0.075	0.050/0.069	0.052/0.065
	20	0.083	0.081/0.112	0.063/0.101	0.052/0.098	0.056/0.077
(b)	50	0.075	0.069/0.093	0.053/0.093	0.054/0.089	0.054/0.073
	100	0.054	0.055/0.072	0.053/0.071	0.051/0.069	0.051/0.069
	20	0.085	0.077/0.114	0.060/0.099	0.052/0.097	0.055/0.078
(c)	50	0.069	0.065/0.098	0.052/0.097	0.053/0.088	0.056/0.071
	100	0.053	0.050/0.078	0.052/0.073	0.050/0.071	0.052/0.068

It is of practical interest to quantify the gain in the precision of parameter estimates obtained by simplifying the model when a possibly small or moderate number of recurrent events per individual is observed. In many applications it will be sensible to examine simple submodels of (4.1), in which only one or two of the three paramenters γ, ϕ and ψ are not omitted.

For instance, three simplified models are considered assuming that there is no covariate information on each individual, i.e., considering model (4.1) without the term $\exp(\beta^T z)$, hereafter model (1). Model (2) will refer to the full model (1) without the event count effect, that is,

$$h\{t; \alpha, \beta, \gamma, \phi \mid z, \mathcal{M}(t)\} = \alpha\gamma(\alpha v)^{\gamma-1}(1 + \alpha\phi t), \tag{5.1}$$

model (3) is model (1) without the Poisson component,

$$h\{t; \alpha, \beta, \gamma, \psi \mid z, \mathcal{M}(t)\} = \alpha\gamma(\alpha v)^{\gamma-1}\psi^{j-1}, \tag{5.2}$$

while model (4) is the renewal-type intensity model,

$$h\{t; \alpha, \beta, \gamma \mid z, \mathcal{M}(t)\} = \alpha\gamma(\alpha v)^{\gamma-1}, \tag{5.3}$$

which is driven by a Weibull-type model.

Assuming that there is no censoring, a numerical study, based on the parameters observed information matrix (calculated numerically since no maijor simplification is possible), was carried out with the same specification above. We considered a $1,000$ generated randon samples of exponential random varibles. For $m = 2, 4, 7, 15, 30$ and $n = 20, 50, 100$, we calculated the relative efficiency (RE) of the MLE obtained by considering the reduced models (2), (3) and (4) relative to the MLE obtained by considering the full model (1). The RE for $\widehat{\alpha_l}$ is defined to be $\text{RE}(\widehat{\alpha_r}, \widehat{\alpha_l}) = \text{var}(\widehat{\alpha_l})/\text{var}(\widehat{\alpha_r})$, where $r = 2, 3, 4$ corresponds to the estimates obtained by considering models (2), (3) and (4), respectively, and $\widehat{\alpha_l}$ is the MLE of α obtained by considering model (1). We define the RE for the $\widehat{\gamma}$, $\widehat{\phi}$ and $\widehat{\psi}$ analogously.

TABLE 2. RE of the estimates obtained by considering the reduced models.

(2), (3) and (4) relative to the estimates obtained under the model (1) for same *m* values and $n = 50$.

RE	*m*				
	2	4	7	15	30
$(\widehat{\alpha}_2, \widehat{\alpha}_1)$	1.10	1.19	1.18	1.11	1.06
$(\widehat{\alpha}_3, \widehat{\alpha}_1)$	1.31	1.09	1.06	1.02	1.01
$(\widehat{\alpha}_4, \widehat{\alpha}_1)$	2.44	3.02	3.25	3.46	3.58
$(\widehat{\gamma}_2, \widehat{\gamma}_1)$	1.11	1.11	1.10	1.06	1.03
$(\widehat{\gamma}_3, \widehat{\gamma}_1)$	1.67	1.33	1.20	1.10	1.04
$(\widehat{\gamma}_4, \widehat{\gamma}_1)$	1.68	1.33	1.20	1.10	1.04
$(\widehat{\phi}_2, \widehat{\phi}_1)$	1.28	1.67	2.18	3.58	5.89
$(\widehat{\psi}_3, \widehat{\psi}_1)$	1.27	1.67	2.18	3.58	5.89

A case study is defined by n and m, and by one of the three different hypothesis tests, which will be set up below. Thus, 45 different cases were simulated.

Table 2 shows the RE of the estimates obtained by considering the reduced models (2), (3) and (4) relative to the estimates under the model (1) for $m = 2, 4, 7, 15, 30$. The RE are very similar under n variantion and only the results related to $n = 50$ are shown. Overall, the parameter estimates are obtained with more precision under simpler models than under the full model (1).

6. REANALYZING THE MAMMARY TUMORS DATA

The hybrid scale model (4.1) was fitted to a dataset extracted from Table 1 of Gail, Santner and Brown (1980) on the times to development of mammary tumors for 48 rats in a carcinogenicity experiment. The rats were assigned randomly to a treatment group (23 rats) and the remaining 25 to a control group. The animals experienced different number of tumors and no censored lifetimes were observed. The covariate was set equal to 0 for the control group and 1 for the treatment group.

The LRS for testing the full model (4.1) against the submodels it contains in a hierarchical manner were obtained (the models are all nested). The LRS for testing the full model against the model with $\psi = 1$ is 3.32, while that for testing model with $\psi = 1$ against the model with $\psi = 1$ and $\phi = 0$ is 0.64. For testing the model with $\psi = 1$ and $\phi = 0$ against the model with $\psi = 1$, $\phi = 0$ and $\gamma = 1$ the LRS is 0.42. Using the asymptotic χ^2 distribution of the LRS, the significance levels are 0.07, 0.42, and 0.52, respectively. This gives evidence in favor of a homogeneous Poisson intensity model, but with event count effect. The MLE of the covariate effect β indexing the homogeneous Poisson intensity model with event count effect is equal to -0.29 with 90% profile confidence interval given by $(-0.44, -0.15)$, implying in a relative risk between the rats on the treatment group to the rats on the control group of 0.75 with 90% profile confidence interval given by $(0.64, 0.86)$. This is clear evidence of the treatment benefit, which is in broad agreement with Lawless (1995).

7. Concluding remarks

The hybrid scale intensity models developed in this paper can be used effectively for analyzing recurrent event data. The special model (4.1) is easily interpretable, and contains several common intensity models as particular cases. These models provide a convenient approach for analyzing data where the covariates affect the intensity functions proportionally. Besides, although there are gains in precision on the parameter estimates obtained under simpler models than under the full model our simulation results suggest that the test procedures performs well even for small and moderate sample sizes.

Although the specific parametric forms for $q_1(\cdot)$, $q_2(\cdot)$ and $q_3(\cdot)$ in (4.1) have analytical convenience and are flexible, they are not critical, in principle, for the overall approach to hold and alternative forms could be considered. However, studies of model mispecification should be considered further by assuming alternative forms for $q_1(\cdot)$, $q_2(\cdot)$ and $q_3(\cdot)$, since, as pointed out by a referee, the LRS tests for model reduction would be seriously compromised if the full model is incorrectly specified. For instance, in the light of the mannary tumors data, there is no formal justification for using the specific parametric forms of $q_1(\cdot)$, $q_2(\cdot)$ and $q_3(\cdot)$ in (4.1) and this numerical example should be seem as an illustration for the overall procedure.

Specifying hybrid intensity models with a non-proportional regression structure may have physical appeal and can be considered. For instance we could consider an accelerated failure time model (Cox and Oakes, 1984, p. 64). This would however introduce extra difficulties in the analysis and needs further work. Time-varying covariates are natural in several situations, but are not considered under our approach. The problem is however not straightforward in our case, and should be investigated further.

As pointed out by a referee, in our formulation, the covariates have the same impact on the time intervals between failures as well as on the total time. However, one could have a given treatment with different effects on the time scales. This situations is out of the scope of the paper and needs further work.

Acknowledgements

I am very grateful to D.R. Cox, A.C. Davison, D. Firth, V. Isham, J. Carpenter and J.G. Leite for their comments and suggestions on this work. I am also very grateful to the Editorial board and to one of the referees for his suggestions on the paper, addressing several important issues. F. Louzada-Neto is supported by CNPq grant number 300045/2003-4.

References

Abramowitz, M. and Stegun, I. (1972), *Handbook of Mathematical Functions*, New York: Dover Publications.

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

Cox D.R. (1972a), ‘Regression models and life tables (with discussion)’, *Journal of the Royal Statistical Society B* **34**, 187–220.

Cox D.R. (1972b), *The statistical analysis of dependencies in point processes. In Stochastic Point Processes*, Ed. P.A.W. Lewis, pp. 55–66, New York: Wiley.

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

Cox, D.R. and Isham, V. (1980), *Point Processes*, London: Chapman and Hall.

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

Follmann, D.A. and Goldberg, M.S. (1988), 'Distinguishing heterogeneity from decreasing hazard rates', *Technometrics* **30**, 389–96.

Gail, M.H., Santner, T.J. and Brown, C.C. (1980), 'An analysis of comparative carcinogenesis experiments based on multiple times to tumour', *Biometrics* **36**, 255–66.

Johnson, N.L., Kotz, S. and Balakrishnan. N. (1994), *Continuous Univariate Distributions 1*, New York: Wiley.

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

Lawless, J.F. (1987), 'Regression methods for Poisson process model', *Journal of the American Statistical Association* **82**, 808–15.

Lawless, J.F. (1995), 'The analysis of recurrent events for multiple subjects', *Applied Statistics* **44**, 487–98.

Lawless, J.F. and Nadeau, C. (1995), 'Some simple robust methods for the analysis of recurrent events', *Technometrics* **37**, 158–68.

Lawless, J.F. and Thiagarajah, K. (1996), 'A point-process model incorporating renewals and time trends, with application to repairable systems', *Technometrics* **38**, 131–38.

Lipschutz, K.H. and Snapinn, S.M. (1997), 'Discussion of the paper by Wei, L.J. and Glidden, D.V.', *Statistics in Medicine* **16**, 846–48.

McGilchrist, C.A. and Aisbett, C.W. (1991), 'Regression with frailty in survival analysis', *Biometrics* **47**, 461–66.

Nelson, W. (1988), 'Graphical analysis of system repair data', *Journal of Quality Technology* **20**, 24–35.

Nelson, W. (1995), 'Confidence limits for recurrence data — Applied to cost or number of product repairs', *Technometrics* **37**, 147–57.

Oakes, D. (1997), Models-based and/or marginal analysis for multiple event-time data, 'In Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis', Ed. D.Y. Lin and T.R. Fleming, pp. 85–98. New York: Springer.

Pepe, M.S. and Cai, J. (1993), 'Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates', *Journal of the American Statistical Association* **88**, 811–20.

Prentice, R.L., Williams, B.J. and Peterson, A.V. (1981), 'On the regression analysis of multivariate failure time data', *Biometrika* **68**, 373–79.

Ross, G.J.S. (1990), *Nonlinear Estimation*, New York: Springer-Verlag.

Seber, G.A.F. and Wild, C.J. (1989), *Nonlinear Regression*, New York: Wiley.

Wei, L.J., Lin, D.Y. and Weissfeld, L. (1989), 'Regression analysis of multivariate incomplete failure time data by modeling marginal distributions', *Journal of the American Statistical Association* **84**, 1065–73.

Departamento de Estatística, Universidade Federal de São Carlos, CP 676, São Carlos/SP, 13565-906, Brazil

E-mail address, F. Louzada-Neto: dfln@power.ufscar.br