© Board of the Foundation of the Scandinavian Journal of Statistics 2003. Published by Blackwell Publishing Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA Vol 30: 651–665, 2003

Bayesian Semiparametric Regression for Median Residual Life

ALAN E. GELFAND

Duke University

ATHANASIOS KOTTAS

University of California, Santa Cruz

ABSTRACT. With survival data there is often interest not only in the survival time distribution but also in the residual survival time distribution. In fact, regression models to explain residual survival time might be desired. Building upon recent work of Kottas & Gelfand [*J. Amer. Statist. Assoc.* 96 (2001) 1458], we formulate a semiparametric median residual life regression model induced by a semiparametric accelerated failure time regression model. We utilize a Bayesian approach which allows full and exact inference. Classical work essentially ignores covariates and is either based upon parametric assumptions or is limited to asymptotic inference in non-parametric settings. No regression modelling of median residual life appears to exist. The Bayesian modelling is developed through Dirichlet process mixing. The models are fitted using Gibbs sampling. Residual life inference is implemented extending the approach of Gelfand & Kottas [*J. Comput. Graph. Statist.* 11 (2002) 289]. Finally, we present a fairly detailed analysis of a set of survival times with moderate censoring for patients with small cell lung cancer.

Key words: censoring, Dirichlet process mixing, residual survival curve, skewness, split densities

1. Introduction

In analysing survival data there is often interest not only in the survival time distribution but also in the residual survival time (or residual life) distribution, i.e. the distribution of survival time given survival past some time say t_0. Just as regression models are employed to explain survival time, such models might be sought to explain residual survival time.

In modelling survival data in the presence of covariates $\mathbf{x}$, typically, either a proportional hazard (PH) model is adopted where the cumulative hazard $H(t; \mathbf{x}) = H_0(t) \exp(\mathbf{x}^T\boldsymbol{\beta})$ or an accelerated failure time (AFT) model is adopted where the cumulative hazard $H(t; \mathbf{x}) = H_0(t \exp(\mathbf{x}^T\boldsymbol{\beta}))$ (see e.g. Cox & Oakes, 1984). In either case a regression is induced for the residual survival time distribution. The AFT setting is the focus of this paper but we do devote section 6 to an alternative version arising under the PH specification.

More precisely, in the AFT setting, for a fixed $\mathbf{x}$, the survival function $S(t; \mathbf{x}) = \exp\{-H_0(t \exp(\mathbf{x}^T\boldsymbol{\beta}))\}$ whence the random variable $U = H_0(T \exp(\mathbf{x}^T\boldsymbol{\beta}))$ is distributed as Exp(1). More importantly, $\log T = \mathbf{x}^T\tilde{\boldsymbol{\beta}} + \epsilon$, where $\epsilon = \log(H_0^{-1}(U))$ and $\tilde{\boldsymbol{\beta}} = -\boldsymbol{\beta}$. If H_0 is an arbitrary cumulative hazard function, ϵ has an arbitrary distribution in R^1. So, the AFT specification is a natural candidate for a semiparametric regression model. That is, we have a parametric component supplied by $\boldsymbol{\beta}$ in the linear regression and a non-parametric component supplied by a class of distributions for ϵ. For the PH specification, instead the baseline cumulative hazard H_0 (or perhaps the hazard function itself) belongs to a non-parametric class of functions.

In the sequel, we formulate a semiparametric median residual life regression model induced by a semiparametric AFT regression model. For a sufficiently rich class of distributions for ϵ, the mean need not exist for all members. But then a mean regression for log survival time is not well defined. As the median always exists a median regression for log

survival time is less restrictive. In the special case of no covariates we can supply a fully non-parametric model under which the class of distributions for log T is dense within all distributions on R^1.

In the classical literature, to fit a median regression model, the least squares criterion is replaced by the least absolute deviations criterion, resulting in what is referred to as L_1 regression. The computational difficulties of this method may partially explain its limited usage as do the inferential limitations with smaller sample sizes (see e.g. Rousseeuw & Leroy, 1987, for a further discussion of L_1 regression). Recently, in this literature, attention has been focused on semiparametric inference procedures for median regression models under censoring, providing models for survival data that are essentially alternatives to an AFT specification (see Ying *et al.*, 1995; Yang, 1999). The estimation techniques are extensions of those for the non-censored case and inference is asymptotic.

A Bayesian approach enabling exact inference given the data may be appealing. The Bayesian literature on non-parametric methods has grown rapidly since the theoretical background for the construction of priors on function spaces was developed, e.g. the work of Ferguson (1973, 1974) on the Dirichlet process. Markov chain Monte Carlo (MCMC) methods (Gelfand & Smith, 1990; Smith & Roberts, 1993; Tierney, 1994), made their practical use feasible. Walker *et al.* (1999) provide a summary of some of the methods in Bayesian non-parametrics. Semiparametric regression modelling is especially attractive in this context; see, for instance, Brunner (1995) and Kuo & Mallick (1997) as well as Gelfand (1999) who offers a review. For median regression modelling, there is only the work of Kottas & Gelfand (2001), summarized below, as well as that of Walker & Mallick (1999), using a Pólya tree prior, and Hanson & Johnson (2002), using a mixture of Pólya trees prior.

Kottas & Gelfand (2001) propose two classes of median regression models introducing two fully Bayesian modelling approaches for the error distribution, a semiparametric and a fully non-parametric one. Both models are based on mixtures with Dirichlet process priors placed on the mixing distributions. The resulting families of error distributions are rich enough to allow for extra variability, skewness and general tail behaviour. Posterior inference, for both models, is carried out through the use of Gibbs sampling. Furthermore, using ideas from Gelfand & Kottas (2002), they demonstrate how full inference for rather general functionals of the underlying error distribution can be achieved. The methodology can be extended to incorporate censoring by introducing latent variables. However, in the presence of moderate to substantial censoring, fitting the more general fully non-parametric error model becomes unstable relative to the semiparametric error model as illustrated in Kottas & Gelfand (2001). As censoring is arguably the most distinguishing feature of survival data, we employ here the semiparametric specification.

The primary contribution of this paper is to show how the approach of Kottas & Gelfand (2001), applied to survival data, can accommodate the induced median residual life regression. As a result, in a flexible semiparametric modelling framework, exact inference about essentially all features of the residual survival distribution can be obtained.

Estimation approaches for the mean residual life and median residual life functions do exist in the classical literature (see e.g. Alam & Kulasekera, 1993; Kochar *et al.*, 2000, where further references can be found). However, these are either based on parametric assumptions or are limited to asymptotic inference in non-parametric settings. Moreover, essentially all the existing work ignores covariate information with the exception of Maguluri & Zhang (1994) where a mean residual life regression model is studied. In the Bayesian literature, we have only found Johnson (1999), which provides an expression for the posterior expectation of mean residual life for a specific type of interval data under a Dirichlet process prior for the survival distribution.

© Board of the Foundation of the Scandinavian Journal of Statistics 2003.

Hence the paper is organized as follows. In section 2, we review the basics of residual life distributions. In section 3, we review Dirichlet process mixing, which we use to model random distributions. Section 4 considers the special case of residual life with no covariates. Section 5 brings in covariate information. Section 6 presents an alternative modelling formulation, based on the PH specification, modelling the baseline cumulative hazard using a Gamma process. Section 7 explicitly details the modelling, the fitting approach and computational issues. Section 8 considers a set of survival times with moderate censoring for patients with small cell lung cancer. The paper ends with a summary and brief discussion in section 9.

2. The residual life distribution

Let $S(t)$ denote the survival function for the continuous random variable T with support on R^+, i.e. $S(t) = P(T > t)$, $t > 0$. We define the residual survival function at t_0 as

$$S_{t_0}(t) = P(T > t \mid T > t_0) = \frac{S(t)}{S(t_0)}, \quad t > t_0.$$

Equivalently, the conditional distribution of residual life $T - t_0 \mid T > t_0$ is defined. If $S(t)$ is differentiable the associated residual survival density function at t_0 is $f_{t_0}(t) = (S(t_0))^{-1}(-dS(t)/dt)$, $t > t_0$.

The mean residual life at t_0 is

$$E(T \mid T > t_0) = \int_{t_0}^{\infty} t f_{t_0}(t)dt = t_0 + (S(t_0))^{-1} \int_{t_0}^{\infty} S(t)dt, \tag{1}$$

provided $tS(t) \to 0$ as $t \to \infty$. The right-most expression in (1) shows immediately that the mean residual life need not exist. On the other hand, the median residual life η_{t_0} satisfies

$$\int_{\eta_{t_0}}^{\infty} f_{t_0}(t)dt = 0.5$$

equivalently $0.5 = S_{t_0}(\eta_{t_0}) = S(\eta_{t_0})/S(t_0)$. So formally $\eta_{t_0} = S^{-1}(0.5S(t_0))$. Strictly speaking, the mean and median residual life should have t_0 subtracted from the given expressions. However, in the sequel, we suppress this known translation.

We note that the set of functionals $\{\eta_{t_0} : t_0 \geq 0\}$ does not uniquely determine $S(t)$ (Gupta & Langford, 1984). This is not a problem for us as we are modelling $S(t)$ to begin with and inferring about the induced S_{t_0} and η_{t_0}.

For the general PH model given in section 1, $H(t; \mathbf{x}) = H_0(t) \exp(\mathbf{x}^T\boldsymbol{\beta})$, it is immediate that

$$\eta_{t_0}(\mathbf{x}) = H_0^{-1}(H_0(t_0) + (\log 2)\exp(-\mathbf{x}^T\boldsymbol{\beta})) \tag{2}$$

while for the general AFT model in section 1, $H(t; \mathbf{x}) = H_0(t \exp(\mathbf{x}^T\boldsymbol{\beta}))$, we have

$$\eta_{t_0}(\mathbf{x}) = \exp(-\mathbf{x}^T\boldsymbol{\beta})H_0^{-1}((\log 2) + H_0(t_0 \exp(\mathbf{x}^T\boldsymbol{\beta}))). \tag{3}$$

It is evident that in either case we still have a linear regression for the median residual life but on a transformed scale.

If H_0 is known in (2) or in (3) the link function for the regression is known. For instance, in the special case of a Weibull hazard, the only example of both a PH and an AFT specification,

$$\eta_{t_0}(\mathbf{x}) = (t_0^{\gamma} + (\log 2)\exp(-\mathbf{x}^T\boldsymbol{\beta}))^{1/\gamma}, \tag{4}$$

where γ is the shape parameter of the Weibull distribution with survival function parameterized as $S(t; \mathbf{x}) = \exp(-t^{\gamma} e^{\mathbf{x}^T\boldsymbol{\beta}})$.

© Board of the Foundation of the Scandinavian Journal of Statistics 2003.

Of course in the semiparametric context, either the cumulative hazard or the error distribution is modelled non-parametrically so that the link function is unknown. As we propose to model the distribution directly, rather than the hazard, we now turn to our approach for non-parametric specifications for the distribution of T. From the Bayesian perspective if this distribution is unknown it is assumed to be random from a class of distributions and a prior is specified over this class. The priors we propose arise through Dirichlet process mixing. In the case of no covariates we can place this prior on a dense class within the class of all distributions on R^1 by modelling the distribution of $\log T$. When the regression is introduced we are modelling the error distribution. In order to define a median regression for $\log T$ a median zero distribution is required for the errors. As noted in section 1, we consider a flexible family of median zero distributions proposed in Kottas & Gelfand (2001).

3. Dirichlet process mixture models

Following Ferguson (1973), a distribution G on Θ follows a Dirichlet process $\mathrm{DP}(vG_0)$ if, given an arbitrary finite measurable partition, $B_1,\ldots,B_r$ of Θ, the joint distribution of $(G(B_1),\ldots,G(B_r))$ is Dirichlet $(vG_0(B_1),\ldots,vG_0(B_r))$, where $G(B_i)$ and $G_0(B_i)$ denote the probability of set B_i under G and G_0, respectively. Here, G_0 is a specified distribution on Θ and $v > 0$ is a precision parameter. Let $K(\cdot;\boldsymbol{\theta})$ be a parametric family of distribution functions (c.d.f.'s), indexed by $\boldsymbol{\theta} \in \Theta$, with associated densities, $k(\cdot;\boldsymbol{\theta})$. If G is proper we define the mixture distribution

$$F(\cdot;G) = \int K(\cdot;\boldsymbol{\theta})G(d\boldsymbol{\theta}). \tag{5}$$

In (5) it is useful to think of $G(d\boldsymbol{\theta})$ as the conditional distribution of $\boldsymbol{\theta}$ given G. Differentiating both sides of (5) with respect to $(\cdot)$ defines $f(\cdot;G) = \int k(\cdot;\boldsymbol{\theta})G(d\boldsymbol{\theta})$.

If G is random say $G \sim \mathrm{DP}(vG_0)$, then $F(\cdot;G)$ is random. Letting $D = \{Y_i,\ i = 1,\ldots,n\}$ denote a sample from $F(\cdot;G)$ and using the bracket notation of Gelfand & Smith (1990), we write its posterior as $[F(\cdot;G) \mid D]$. Functionals of $F(\cdot;G)$, for which we use the generic notation $Q(F(\cdot;G))$, are of interest with posteriors denoted by $[Q(F(\cdot;G)) \mid D]$.

In the context of (5), suppose for each Y_i, $i = 1,\ldots,n$ we introduce a latent $\boldsymbol{\theta}_i$ and assume that, given the $\boldsymbol{\theta}_i$'s, the Y_i's are conditionally independent, distributed as $k(\cdot;\boldsymbol{\theta}_i)$. Assume further that the $\boldsymbol{\theta}_i$'s are conditionally independent and identically distributed given G. As a result, marginalizing over the $\boldsymbol{\theta}_i$'s, the Y_i's are still independent, now conditionally on G, with joint density $\prod_{i=1}^n f(y_i;G) = \prod_{i=1}^n \int k(y_i;\boldsymbol{\theta}_i)G(d\boldsymbol{\theta}_i)$. Adding $G \sim \mathrm{DP}(vG_0)$, possibly along with hyperpriors on v (see Escobar & West, 1995) and/or the parameters of G_0, completes the Bayesian model specification. Such Dirichlet process mixture models were originally studied by Antoniak (1974) and Lo (1984). In particular, Antoniak (1974) noted that this Bayesian model can be marginalized over G; a result that forms the basis of several MCMC algorithms (Escobar, 1994; West *et al.*, 1994; Escobar & West, 1995; Bush & MacEachern, 1996; MacEachern & Müller, 1998; Neal, 2000) which can be implemented to obtain samples from the posterior $[\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_n \mid D]$ resulting after the marginalization over G.

Gelfand & Mukhopadhyay (1995) describe how to use these samples to infer about linear functionals associated with $F(\cdot;G)$. They show how posterior expectations of linear functionals and products of linear functionals can be computed. Restriction to posterior moments of linear functionals severely limits inference. Gelfand & Kottas (2002) provide a computational approach to obtain the entire posterior distribution for more general functionals. Hence, full inference is available for many population features and for comparing a feature across

© Board of the Foundation of the Scandinavian Journal of Statistics 2003.

populations. Briefly, note that for a linear functional Q, $Q(F(\cdot; G)) = \int Q(K(\cdot; \boldsymbol{\theta}_0))G(d\boldsymbol{\theta}_0)$. Now, instead of marginalizing over G in $[\boldsymbol{\theta}_0, \boldsymbol{\theta}, G \mid D] \propto [D \mid \boldsymbol{\theta}][\boldsymbol{\theta}_0, \boldsymbol{\theta} \mid G][G]$, observe that

$$[\boldsymbol{\theta}_0, \boldsymbol{\theta}, G \mid D] \propto [\boldsymbol{\theta}_0 \mid G][G \mid \boldsymbol{\theta}][\boldsymbol{\theta} \mid D].$$

Hence given the posterior sample $\boldsymbol{\theta}_b^* = (\boldsymbol{\theta}_{b1}^*, \ldots, \boldsymbol{\theta}_{bn}^*)$, $b = 1, \ldots, B$, for each $\boldsymbol{\theta}_b^*$ draw $G_b^* \sim [G \mid \boldsymbol{\theta}_b^*]$ which is an updated Dirichlet process $\mathrm{DP}(\nu' G_{0b}')$, where $\nu' = \nu + n$ and $G_{0b}' = (\nu + n)^{-1}(\nu G_0 + \sum_{i=1}^n \delta_{\boldsymbol{\theta}_{bi}^*})$, with δ_a a degenerate distribution at a (Ferguson, 1973). Then draw independent $\boldsymbol{\theta}_{0lb}^* \sim G_b^*$, for $l = 1, \ldots, L$. Finally, $Q_b^* \equiv Q(F(\cdot; G_b^*)) = L^{-1} \sum_{l=1}^L Q(K(\cdot; \boldsymbol{\theta}_{0lb}^*))$ is a Monte Carlo integration for a realization from $[Q(F(\cdot; G)) \mid D]$.

To obtain an approximate realization from $[G \mid \boldsymbol{\theta}_b^*]$ we use the constructive definition of Sethuraman (1994) with a partial sum approximation that we justify through certain convergence results. A practical rule to choose the number of terms in the partial sum approximation is also provided. Sampling from the posterior of the 'c.d.f.-at-a-point' functional $F(y_0; G)$, for a grid of points, we can invert to obtain samples from the posterior of any quantile functional. Other functionals of interest can also be handled.

If we write $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)})$, we might place a Dirichlet process prior on $\boldsymbol{\theta}^{(1)}$, i.e. $\boldsymbol{\theta}^{(1)} \sim G$ where $G \sim \mathrm{DP}(\nu G_0)$ with a parametric prior on $\boldsymbol{\theta}^{(2)}$, yielding $F(\cdot; G, \boldsymbol{\theta}^{(2)}) = \int K(\cdot; \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)})G(d\boldsymbol{\theta}^{(1)})$, a semiparametric specification.

4. The case of no covariates

Following the development of the previous section, we seek to create a random residual life distribution for T given t_0, which we denote by $F_{t_0}^T(\cdot; G)$. In fact, we model $Y = \log T$ given $y_0 = \log t_0$, i.e. $F_{y_0}^Y(\cdot; G)$. The closure of the family of densities corresponding to distributions

$$F^Y(\cdot; G) = \int \Phi\Big(\frac{\cdot - \mu}{\sigma}\Big) G(d\mu, d\sigma), \tag{6}$$

where Φ is the standard normal distribution function, contains all densities on R^1 (Ferguson, 1983; Lo, 1984). Hence if $G \sim \mathrm{DP}(\nu G_0)$, (6) provides a random realization from a class of distributions that is dense in the entire class of distributions on R^1.

Then, extending the notation of section 2, $S_{y_0}^Y(\cdot; G) = S^Y(\cdot; G)/S^Y(y_0; G)$, on (y_0, ∞), where $S^Y(\cdot; G) = 1 - F^Y(\cdot; G)$ and $\eta_{y_0}(G) = (S_{y_0}^Y)^{-1}(0.5; G)$. Straightforwardly, $\eta_{t_0}(G) = \exp(\eta_{y_0}(G))$ and $S_{t_0}^T(\cdot; G) = S^T(\cdot; G)/S^T(t_0; G)$ where $S^T(t; G) = S^Y(\log t; G)$.

How can the methodology of Gelfand & Kottas (2002), described briefly at the end of the previous section, be used to obtain full inference regarding $S_{t_0}^T(t; G)$, for fixed t, and $\eta_{t_0}(G)$, i.e. the posteriors given data D, $[S_{t_0}^T(t; G) \mid D]$ and $[\eta_{t_0}(G) \mid D]$? Following section 3, for a grid of t values say $t_{(1)} < t_{(2)} < \cdots < t_{(K)}$ and a posterior sample G_b^*, $b = 1,2, \ldots, B$, we can create a $K \times B$ matrix say V where $V_{kb} = F^Y(y_{(k)}; G_b^*)$ is a realization from $[F^Y(y_{(k)}; G) \mid D]$ with $y_{(k)} = \log t_{(k)}$. But then $W = J - V$, where J is the $K \times B$ matrix with all its elements equal to 1, is such that W_{kb} is a realization from $[S^Y(y_{(k)}; G) \mid D]$ and hence a realization from $[S^T(t_{(k)}; G) \mid D]$. In fact, the kth row of W provides a posterior sample from $[S^T(t_{(k)}; G) \mid D]$. But also, the bth column of W provides K points on a random posterior realization of the curve $S^T(\cdot; G)$. With interpolation, we obtain essentially a posterior realization of this curve.

Next, suppose we divide all rows of W by the first row resulting in a matrix $W^{(1)}$. Now the entries in the kth row of $W^{(1)}$, $k > 1$, are posterior samples from $[S_{t_{(1)}}^T(t_{(k)}; G) \mid D]$ and the entries in the bth column of $W^{(1)}$ provide $K - 1$ points on a random posterior realization of $S_{t_{(1)}}^T(\cdot; G)$. Again, interpolation enables essentially a posterior realization of this curve. But then, appropriate inversion of this curve supplies essentially a realization from $[\eta_{t_{(1)}}(G) \mid D]$. The B columns of $W^{(1)}$ provide B posterior realizations of $[S_{t_{(1)}}^T(\cdot; G) \mid D]$ and B samples from

© Board of the Foundation of the Scandinavian Journal of Statistics 2003.

$[\eta_{t_{(1)}}(G) \mid D]$. Hence, given $t_{(1)}$, posterior inference for the residual life curve and for the median residual life (in fact, any quantile of the residual life distribution) is immediate.

Evidently, if we divide all rows of W by the second row we can obtain posterior inference for the residual life curve and for the median residual life given $t_{(2)}$. So now, an overall computational strategy is clear. Choose the set of t's to be sufficiently dense over the portion of R^+ of interest, to include all t's for which we seek the residual life distribution and such that, beyond the largest t of interest, there are enough t's to provide an adequate domain for the residual life distribution associated with this t.

Lastly, the technical discussion in Gelfand & Kottas (2002, section 3.2) clarifies that, as the residual life functional is defined through a ratio of bounded linear functionals, the partial sum approximation approach described at the end of section 3 and above yields convergence in probability to the exact functional. For the median residual life functional convergence in probability emerges using theorem 4 from Gelfand & Kottas (2002).

5. The regression case

As explained in sections 1 and 2, with covariates, we first formulate a semiparametric median regression model for log survival time which in turn, induces a regression for the residual life distribution and the median residual life. Explicitly, we have

$$Y = \log T = \mathbf{x}^T \boldsymbol{\beta} + \epsilon, \tag{7}$$

where $\mathbf{x} = (1, x_1, \ldots, x_p)^T$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ and ϵ has a median zero distribution.

Next, we summarize briefly the semiparametric median zero family of distributions on R^1 proposed in Kottas & Gelfand (2001).

Let $f(\cdot; \theta)$ be a symmetric (about 0) unimodal density on R^1 where $\theta > 0$ is a scale parameter. Define

$$p(\cdot; \theta, \gamma) = \gamma^{-1} f(\cdot \gamma^{-1}; \theta) 1_{(-\infty,0)}(\cdot) + \gamma f(\cdot \gamma; \theta) 1_{[0,\infty)}(\cdot), \tag{8}$$

where $\gamma > 0$. Any member of this family, with $\gamma \neq 1$, is a skewed distribution with the type and amount of skewness depending on the value of γ. The case of symmetry corresponds to $\gamma = 1$, yielding $p(\cdot; \theta, 1) \equiv f(\cdot; \theta)$, while for $\gamma < 1 (> 1)$ the resulting distribution is right (left) skewed. γ controls the rate at which the density drops off on the positive and negative axes. Regardless, the mass remains 0.5 on each so that the median is 0 but a discontinuity of the density occurs at the origin. The unique mode is still at 0. The c.d.f. of (8) is $P(\cdot; \theta, \gamma) = F(\cdot \gamma^{-1}; \theta) 1_{(-\infty,0)}(\cdot) + F(\cdot \gamma; \theta) 1_{[0,\infty)}(\cdot)$, where $F(\cdot; \theta)$ is the c.d.f. associated with $f(\cdot; \theta)$. The densities in (8) are closely related to the split densities introduced by Geweke (1989) as importance sampling densities.

Properties of (8) are developed in Kottas & Gelfand (2001). Here we only note that, to clarify how γ affects the skewness of (8), we might reparametrize to a skewness functional. As, for a general f, moments associated with (8) need not exist, the Bowley coefficient (Groeneveld & Meeden, 1984) being free of moments, is useful. This coefficient,

$$\delta(\theta, \gamma) \equiv \frac{q_{0.75}(\theta, \gamma) + q_{0.25}(\theta, \gamma) - 2q_{0.5}(\theta, \gamma)}{q_{0.75}(\theta, \gamma) - q_{0.25}(\theta, \gamma)} = \frac{1 - \gamma^2}{1 + \gamma^2} \tag{9}$$

under (8) by straightforward calculation. Evidently, $\delta \in (-1, 1)$ with $\delta = 0$ indicating symmetry and $\delta = 1(-1)$ indicating extreme right (left) skewness.

To introduce Dirichlet process mixing to (8) we consider general scale mixtures of $p(\cdot; \theta, \gamma)$. For a proper G, consider

© Board of the Foundation of the Scandinavian Journal of Statistics 2003.

$$f(\cdot; G, \gamma) = \int p(\cdot; \theta, \gamma) G(d\theta). \tag{10}$$

Mixing on θ to create the semiparametric family in (10) preserves median 0 while enriching the class of models in terms of their dispersion. Note that, even if the mean of $p(\cdot; \theta, \gamma)$ exists, the mean of (10) need not exist for arbitrary G. Also, the class (10) is not made richer by one choice of $f(\cdot; \theta)$ in (8) versus another. Attractively, δ is again given by (9) for the mixture (10). If G is assumed to arise from a Dirichlet process prior, the full inference approach described in section 3 will be applicable.

Returning to (7), for a sample of survival times T_i, $i = 1, \ldots, n$ we now assume that the ϵ_i are distributed according to (10). That is, $Y_i \sim f(\cdot - \mathbf{x}_i^T \boldsymbol{\beta}; G, \gamma)$, which, as in section 2, induces a linear regression, on a transformed scale, for the residual life curve and for the median residual life.

6. An alternative version using the PH model

An alternate approach to introduce a regression into the residual life function or the median residual life is in the PH setting. For instance, Kalbfleisch (1978) models the unknown baseline cumulative hazard function using a Gamma process. Alternatively, the extended Gamma process (Dykstra & Laud, 1981) can be used to model the hazard function itself.

To illustrate with the former, first in the absence of covariates $S_{t_0}(t) = \exp\{-(H_0(t) - H_0(t_0))\}$, $t > t_0$. Define $r_{t_0}(t) = H_0(t) - H_0(t_0)$. If H_0 comes from a Gamma process GP(cR), where R is a specified cumulative hazard and $c > 0$, i.e. for any t, $H_0(t) \sim \text{Gamma}(cR(t), c)$, a Gamma distribution with mean $R(t)$ and variance $R(t)/c$ then $r_{t_0}(t) \sim \text{Gamma}(c(R(t) - R(t_0)), c)$. With n ordered observations $t_{(1)} < t_{(2)} < \cdots < t_{(n)}$ the vector $\mathbf{r} = (r_0(t_{(1)}), r_{t_{(1)}}(t_{(2)}), \ldots, r_{t_{(n-1)}}(t_{(n)}))$, where $r_{t_{(i-1)}}(t_{(i)}) = H_0(t_{(i)}) - H_0(t_{(i-1)})$, has components that are, a priori, independent Gamma variables and, clearly $\sum_{j=1}^{i} r_{t_{(j-1)}}(t_{(j)}) = H_0(t_{(i)})$, $i = 1, \ldots, n$.

Hence with posterior samples $\mathbf{r}_b^*$, $b = 1, \ldots, B$ from $[\mathbf{r} \mid D]$ and interpolation we can obtain a posterior realization from $[H_0(\cdot) \mid D]$ and thus from $r_{t_0}(\cdot) = H_0(\cdot) - H_0(t_0)$ given D for any t_0. But $r_{t_0}(\eta_{t_0}(H_0)) = \log 2$ determines $\eta_{t_0}(H_0)$ so each posterior realization $\mathbf{r}_b^*$ yields an $\eta_{t_0,b}^*$.

In practice, the grid of t values arising from the data will neither be fine enough or regular enough to adequately handle the required interpolation. Additional sampling from the Gamma process will be needed. To introduce covariates, now $S_{t_0}(t; \mathbf{x}) = \exp\{-(H_0(t) - H_0(t_0)) \exp(\mathbf{x}^T \boldsymbol{\beta})\}$, $t > t_0$, so we can define $r_{t_0}(t; \mathbf{x}) = r_{t_0}(t) \exp(\mathbf{x}^T \boldsymbol{\beta})$. Now posterior realizations from $\mathbf{r}$ and $\boldsymbol{\beta}$ enable posterior realizations from $\eta_{t_0}(\mathbf{x})$.

7. Modelling details, model fitting and computational issues

Here we present explicit details for the regression case based on the class of error distributions presented in section 5. We discuss prior specification, simulation-based model fitting and inference for the residual life distribution and median residual life function.

To specify (10), we choose a split normal for the kernel of the mixture. Hence (8) becomes

$$p(\cdot; \theta, \phi) = f_N(\cdot \mid 0, \phi\theta) 1_{(-\infty, 0)}(\cdot) + f_N(\cdot \mid 0, \theta/\phi) 1_{[0, \infty)}(\cdot), \tag{11}$$

where $f_N(\cdot \mid \mu, \sigma^2)$ denotes the $N(\mu, \sigma^2)$ density and we have reparameterized from γ to $\phi = \gamma^2$. We adopt a Dirichlet process prior for G whence the semiparametric model is completed by specifying parametric priors for $\boldsymbol{\beta}$ and ϕ. We take a multivariate normal prior for the former

© Board of the Foundation of the Scandinavian Journal of Statistics 2003.

and a Gamma(a, b) prior (with mean a/b) for the latter. Letting $Y_i = \log T_i$, $i = 1, \ldots, n$, the resulting full Bayesian model has the hierarchical structure:

$$\begin{aligned} Y_i \mid \boldsymbol{\beta}, \phi, \theta_i &\overset{\text{ind.}}{\sim} p(y_i - \mathbf{x}_i^T\boldsymbol{\beta}; \theta_i, \phi), \quad i = 1, \ldots, n, \\ \theta_i \mid G &\overset{\text{i.i.d.}}{\sim} G, \quad i = 1, \ldots, n, \\ G &\sim \mathrm{DP}(vG_0), \\ \boldsymbol{\beta} &\sim N_{p+1}(\boldsymbol{\mu}, \Sigma), \\ \phi &\sim \mathrm{Gamma}(a, b), \end{aligned} \tag{12}$$

where the base distribution G_0 for the Dirichlet process is taken to be an IGamma(s_1, s_2) (with mean $s_2/(s_1 - 1)$, provided $s_1 > 1$). In fact, the components of the vector $\boldsymbol{\beta}$ are assumed a priori independent, hence $\Sigma = \mathrm{diag}(\sigma_0^2, \sigma_1^2, \ldots, \sigma_p^2)$ with $\boldsymbol{\mu} = (\mu_0, \mu_1, \ldots, \mu_p)^T$. The hyperparameters are fixed. Prior specification can be accomplished in a rather non-informative fashion using only a rough range, r, for survival time on the logarithmic scale provided either from previous studies or from the data in hand. To specify G_0, we work with the parametric version of the model, emerging when $v \to 0^+$, that replaces the first three stages in (12) with $Y_i \mid \boldsymbol{\beta}, \phi, \theta \overset{\text{ind.}}{\sim} p(y_i - \mathbf{x}_i^T\boldsymbol{\beta}; \theta, \phi)$, $i = 1, \ldots, n$, and $\theta \sim \mathrm{IGamma}(s_1, s_2)$. θ being a scale parameter, we set the mean, $s_2/(s_1 - 1)$, of G_0 equal to $(r/6)^2$ or perhaps $(r/4)^2$. G_0 is fully specified by taking $s_1 = 2$ implying infinite variance. Adding a prior for v does not complicate the fitting details (Escobar & West, 1995), but is not done here because we have found very little sensitivity to the choice of its value. For the data set of section 8, we experimented with values of v between 0.5 and 10 obtaining essentially identical posterior inference (taking eventually $v = 1$). For the regression coefficients, we follow the standard approach, assuming $\mu_j = 0$ and large variances σ_j^2, $j = 0, 1, \ldots, p$. Finally, we need to supply the hyperparameters a and b, corresponding to the prior of ϕ. Seeking a specification that, a priori, does not favour skewness we centre this prior around 1, yielding $a = b$, assuming large variance. We note that a choice of $a < 1$ is not reasonable since then the prior has an asymptote at 0 which may strongly affect the behaviour of the posterior. In the example of section 8, we take $a = b = 2.5$, implying a priori a range for ϕ roughly from 0 to 3.5 and $P(\phi < 1) = 0.584$.

To obtain inferences for the vector of regression coefficients, for the skewness in the error distribution (through the parameter ϕ), and for functionals of the residual life distribution we need the joint posterior $[\boldsymbol{\beta}, \phi, \boldsymbol{\theta} \mid D]$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ and $D = \{y_i, \mathbf{x}_i, i = 1, \ldots, n\}$, obtained upon marginalization over G. In fact, we resort to simulation-based fitting of the model, employing a Gibbs sampler (Gelfand & Smith, 1990) whose full conditionals are briefly described next; see Kottas & Gelfand (2001, appendix A) for the complete implementation details.

The full conditional for ϕ is a generalized inverse Gaussian distribution that can be sampled efficiently using a ratio of uniforms generation method given in Dagpunar (1988, p. 133). Following Escobar & West (1995), the full conditional for each θ_i, $i = 1, \ldots, n$, is a mixed distribution with point masses at $\theta_i = \theta_j$, $j = 1, \ldots, n$, $j \neq i$ and continuous mass on an inverse Gamma distribution. The required weights are easily computed resulting in straightforward draws from these full conditionals. Finally, the full conditionals for the regression coefficients β_j, $j = 0, 1, \ldots, p$, can be expressed as piecewise densities with components that are truncated normals, which we sample using the suggestion of Devroye (1986, p. 38).

The algorithm can be readily modified to incorporate censoring. In particular, a combination of Gibbs sampling and data augmentation can be employed to handle left, right or interval censored survival times. Briefly, for any T_i which are, say, right censored at V_i, i.e. we only know $V_i < T_i$, we can retain $Y_i = \log T_i$ in (12) but employ an additional updating step in the Gibbs sampler using the structure induced by the Dirichlet process prior. In particular,

© Board of the Foundation of the Scandinavian Journal of Statistics 2003.

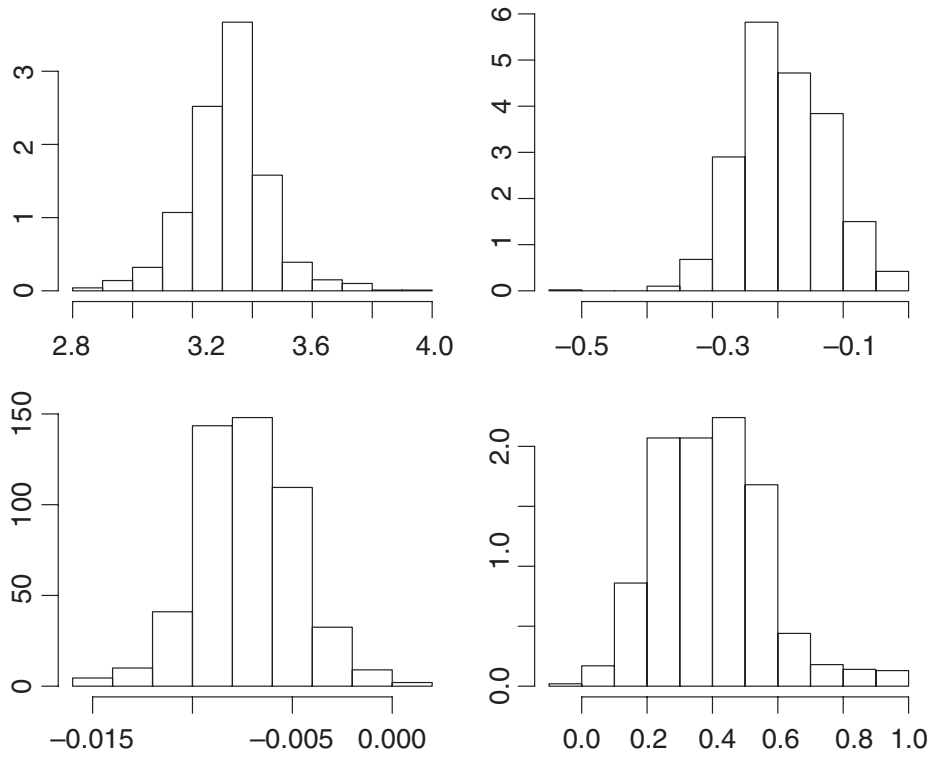

Fig. 1. Posteriors of regression coefficients (β_0 upper left, β_1 upper right, β_2 lower left) and skewness functional δ (lower right) under the semiparametric regression model.

given $\boldsymbol{\beta}$, ϕ and $\boldsymbol{\theta}$, we update Y_i from $\int p(y_i - \mathbf{x}_i^T\boldsymbol{\beta}; \theta_0, \phi)[\theta_0 \mid \boldsymbol{\theta}]d\theta_0$ subject to $Y_i > \log V_i$. Here $[\theta_0 \mid \boldsymbol{\theta}]$ is a mixed distribution with point masses at θ_i, $i = 1, \ldots, n$ and continuous mass on G_0. But then, given the observed Y_i's and the updated censored Y_i's, we update $\boldsymbol{\beta}$, ϕ and the θ_i's as above. Again, we refer to Kottas & Gelfand (2001, section 5) for more details.

The posterior sample $(\boldsymbol{\beta}_b^*, \phi_b^*, \boldsymbol{\theta}_b^*)$, $b = 1, \ldots, B$ obtained from the Gibbs sampler yields inference for the residual life distribution using a simple modification of the approach described in section 4 for the no covariates case. All that is needed here is to apply the method for the specific combination of covariate levels say $\mathbf{x}_0$. Hence, instead of (6), we work with

$$F(\cdot - \mathbf{x}_0^T\boldsymbol{\beta}; G, \phi) = \int P(\cdot - \mathbf{x}_0^T\boldsymbol{\beta}; \theta, \phi)G(d\theta),$$

where $P(\cdot; \theta, \phi)$ is the c.d.f. of (11). Now the $K \times B$ matrix consists of entries $F(y_{(k)} - \mathbf{x}_0^T\boldsymbol{\beta}_b^*; G_b^*, \phi_b^*)$, $k = 1, \ldots, K$, $b = 1, \ldots, B$. All the other details are the same as with the no covariates case. We note that drawing G_b^* is only done once at iteration b for all the $\mathbf{x}_0$'s of interest. Finally, in the regression context, the range of inferences that can be reported is broader. Fixing $\mathbf{x}_0$, we can compare the posteriors of median residual life for several conditional t_0's of interest. But also fixing t_0, we can observe how $\eta_{t_0}(\mathbf{x})$ evolves with $\mathbf{x}$. In particular, for a continuous covariate, working with a grid of its possible values, we obtain the posterior estimate with ranges of uncertainty for the median residual life regression curve.

© Board of the Foundation of the Scandinavian Journal of Statistics 2003.

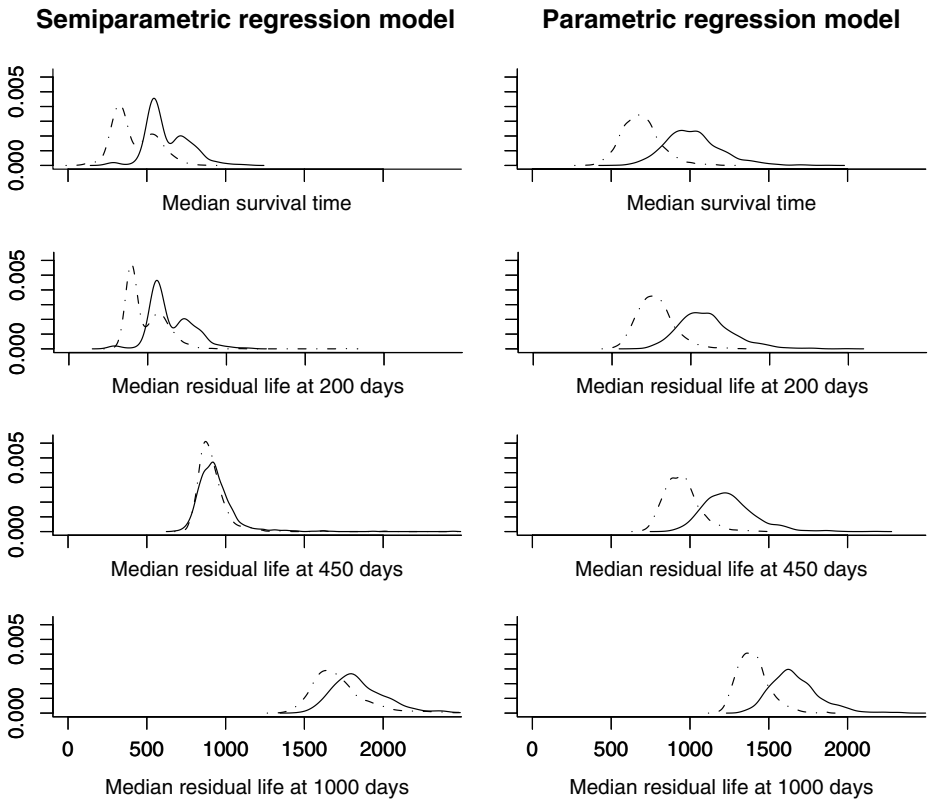

Fig. 2. Posteriors of median survival time and median residual life at $t_0 = 200$, 450 or 1000 days, for a 47-year-old patient receiving treatment B (dashed-dotted lines) or treatment A (solid lines). The left column corresponds to the semiparametric model and the right column to the parametric model.

8. Data analysis

We illustrate the methodology with a data set involving censoring analysed using median regression models initially by Ying *et al.* (1995) and later by Yang (1999), Walker & Mallick (1999) and Kottas & Gelfand (2001). It consists of survival times in days for 121 patients with small cell lung cancer. Among them, 98 are observed with the remaining 23 right censored. Each patient was randomly assigned to one of two treatments A and B, achieving 62 and 59 patients, respectively. Also available is the patient's age at entry. We fit model (12) to this data set with $Y_i = \log_{10} T_i$, $x_{i1} = 1$ if the ith patient is receiving treatment A and 2 otherwise and $x_{i2} = i$th patient's entry age. Following the suggestions of section 7 regarding the prior hyperparameters, we take $v = 1$, $s_1 = 2$, $s_2 = 0.203$, $\mu_j = 0$, $\sigma_j^2 = 50$, $j = 0$, 1, 2 and $a = b = 2.5$. The value of s_2 corresponds to a range $r = 1.8$ with $s_2 = (r/4)^2$, a rather vague specification given that the smallest (observed) survival time on the log scale is equal to 1.919 with the largest (censored) being 3.297. Kottas & Gelfand (2001) offer a comparison of posterior results under $s_2 = 0.203$ and $s_2 = 1.5$, a dramatically larger value, revealing robustness of model (12).

The posteriors of the regression coefficients (see Fig. 1) provide evidence that survival time decreases with increasing age and that treatment A is better. Moreover a right-skewed error distribution is clearly favoured as the posterior of the skewness functional $\delta = (1 - \phi)/(1 + \phi)$,

© Board of the Foundation of the Scandinavian Journal of Statistics 2003.

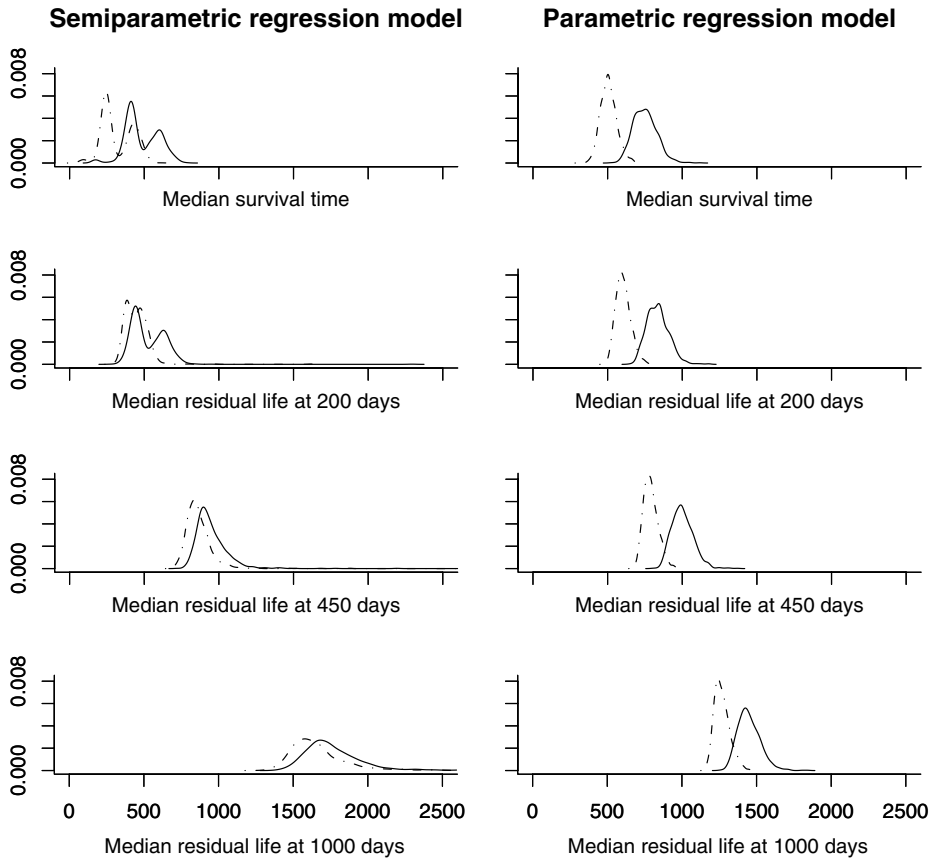

Fig. 3. Posteriors of median survival time and median residual life at $t_0 = 200$, 450 or 1000 days, for a 63-year-old patient receiving treatment A (solid lines) or treatment B (dashed-dotted lines). The left column has the posteriors under the semiparametric model and the right under the parametric model.

also given in Fig. 1, indicates; see Kottas & Gelfand (2001) for further illustrations with functionals of the survival distribution at certain combinations of covariate values. In particular, the posteriors of median survival time for both treatments at certain ages are bimodal, an interesting feature that the model captures. See Figs. 2–4 where we include the posteriors of median survival time for three values of age.

In the interest of comparing the results from the semiparametric model (12) with a parametric analysis, we consider a PH regression model with a Weibull baseline cumulative hazard, $H(t_i; x_{i1}, x_{i2}) = t_i^{\gamma} \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})$, assuming a priori that $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T \sim N_3((0,0,0)^T, \mathrm{diag}(\tau_0^2, \tau_1^2, \tau_2^2))$ and $\gamma \sim \mathrm{Gamma}(c, d)$. Under this model, the median residual life has the convenient form given in (4) and therefore its posterior immediately emerges if we sample from $[\boldsymbol{\beta}, \gamma \mid D]$. To this end, we employ Gibbs sampling with auxiliary variables in the spirit of Damien *et al.* (1999). We performed prior sensitivity analysis with values for τ_j^2, $j = 0, 1, 2$ ranging from 50 to 400 and values for c and d corresponding to Gamma distributions with spread ranging from (0, 10) to (0, 20) and medians from 1.5 to 2.5. The differences in the resulting posteriors were inconsequential in all cases yielding both age and treatment as significant covariates and favouring an increasing baseline hazard function. The final results are based on the choice $c = 1.4$, $d = 0.7$ and $\tau_j^2 = 50$, $j = 0, 1, 2$ under which 95

© Board of the Foundation of the Scandinavian Journal of Statistics 2003.

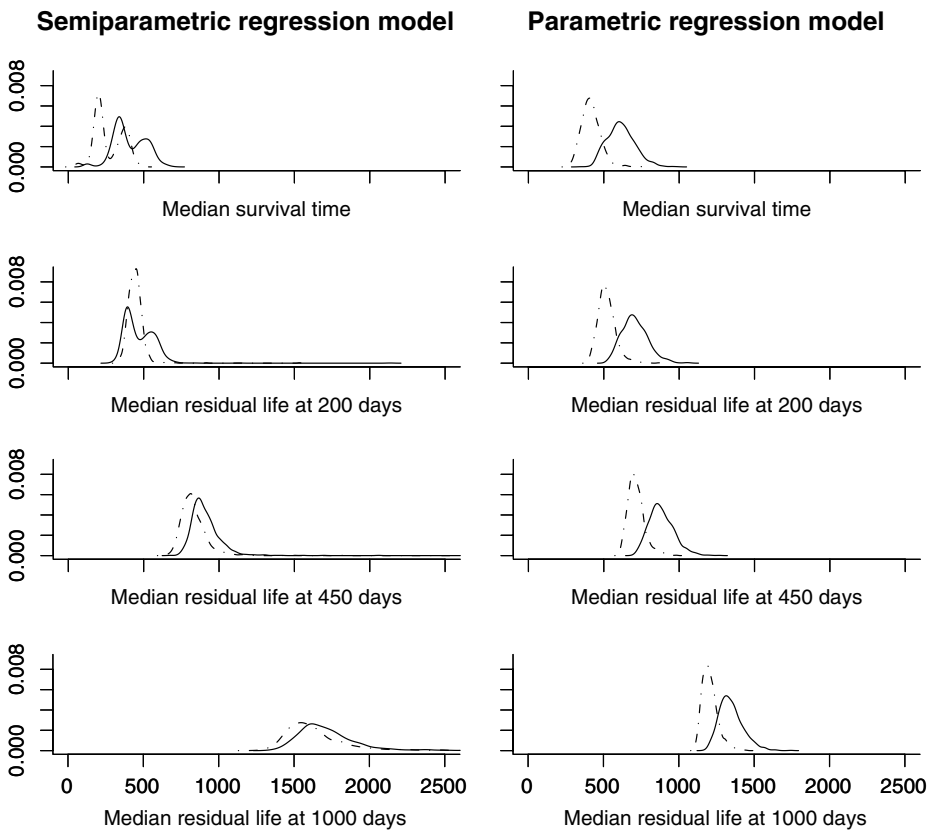

Fig. 4. Posteriors of median survival time and median residual life at $t_0 = 200$, 450 or 1000 days, for a 74-year-old patient receiving treatment A (solid lines) or treatment B (dashed-dotted lines). The left column has the posteriors under the semiparametric model and the right under the parametric model.

per cent posterior interval estimates for β_1, β_2 and γ are (0.1639, 0.9443), (0.0013, 0.0510) and (1.1708, 1.6121), respectively.

Turning to inference for the residual life distribution, we first study the median residual life for six combinations of covariate values and three values of t_0. Specifically, we consider both treatments for three values of age, 47, 63 and 74 years, corresponding to the 0.05, 0.5 and 0.95 quantiles, respectively, of the observed values of entry age. Moreover we take $t_0 = 200$, 450 and 1000 days, values which correspond roughly to the 0.2, 0.5 and 0.9 quantiles, respectively, of the 98 observed survival times. Figures 2–4 provide all the resulting posteriors of median residual life under both models. The posteriors of median survival time (i.e. median residual life at $t_0 = 0$) are also included. The greater flexibility of the semiparametric regression model is evident, for instance, suggesting bimodality in the posteriors of median survival time and median residual life at 200 days. (We do not think this is an artefact of the prior. It arises only in some of the posteriors and, in fact, occurs at the smallest t_0 we considered, the choice where the data provides the most information.) The posteriors from the semiparametric regression model suggest a decrease of median residual life with age, particularly evident at $t_0 = 200$ days, and superiority of treatment A. Finally, based on the semiparametric regression model, in Fig. 5, we plot the posterior predictive residual survival functions at $t_0 = 200$ and 450 days for the six combinations of covariate levels considered above.

© Board of the Foundation of the Scandinavian Journal of Statistics 2003.

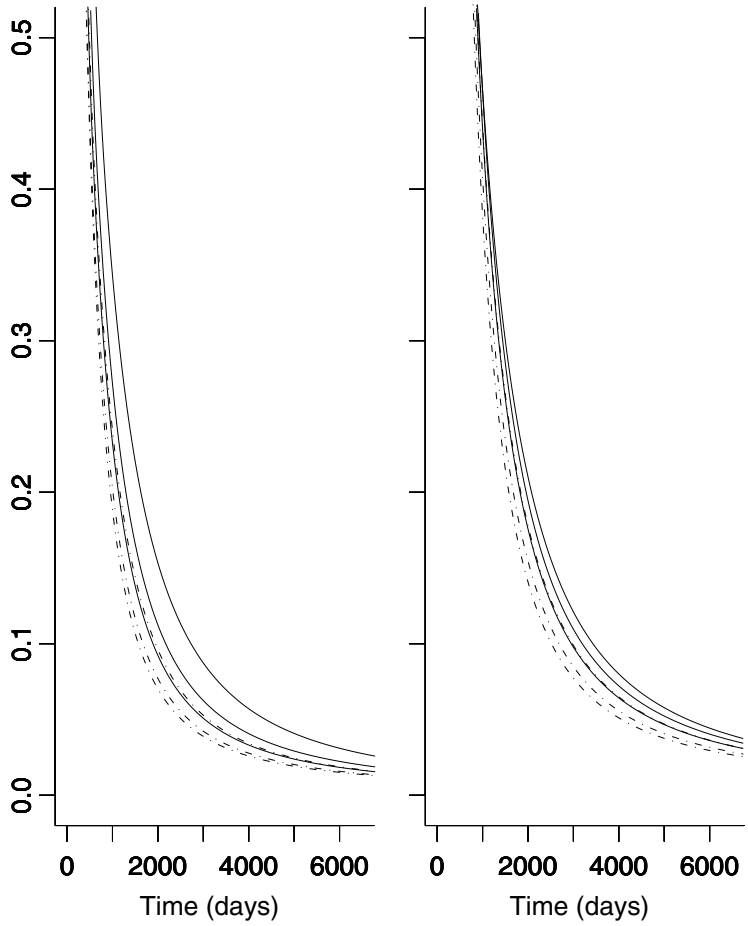

Fig. 5. Under the semiparametric regression model, posterior predictive residual survival functions at $t_0 = 200$ days (left) and 450 days (right) for treatment A (solid lines) and treatment B (dashed-dotted lines) at ages 47, 63 and 74 years (upper, middle and lower curve, respectively, in each case).

9. Summary and concluding remarks

In this paper, building on earlier work for median regression modelling, we have developed a semiparametric regression model for median residual life. Moreover, we have shown how full and exact inference for other features of the residual life distribution can be obtained. The model is induced by semiparametric AFT median regression modelling for log survival time, based on a Dirichlet process mixture for the error distribution. We have provided approaches for prior specification and simulation-based model fitting and inference. Finally, we have applied the methodology to the analysis of a data set and included comparison with results from a standard parametric model.

Any particular specification for the initial regression model, either an AFT or a PH specification, necessarily induces certain (potentially restrictive) structure on the corresponding regression for the residual life distribution. The Bayesian non-parametric models we have proposed allow data-driven deviations from commonly used AFT or PH specifications resulting in more flexible regression models to explain residual survival time. An alternative

© Board of the Foundation of the Scandinavian Journal of Statistics 2003.

approach would attempt to provide Bayesian non-parametric models directly for classes of percentile residual life functions or mean residual life functions in order to achieve inference for median or mean residual life regression.

Acknowledgements

The work of the first author was supported in part by NSF grant DMS 99-71206. The authors thank Deb Sinha for suggesting this problem to them and two referees for kind words and useful suggestions.

References

Alam, K. & Kulasekera, K. B. (1993). Estimation of the quantile function of residual life time distribution. *J. Statist. Plann. Inference* **37**, 327–337.

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to nonparametric problems. *Ann. Statist.* **2**, 1152–1174.

Brunner, L. J. (1995). Bayesian linear regression with error terms that have symmetric unimodal densities. *J. Nonparametr. Statist.* **4**, 335–348.

Bush, C. A. & MacEachern, S. N. (1996). A semiparametric Bayesian model for randomised block designs. *Biometrika* **83**, 275–285.

Cox, D. R. & Oakes, D. (1984). *Analysis of survival data*. Chapman & Hall/CRC, New York.

Dagpunar, J. (1988). *Principles of random variate generation*. Clarendon Press, Oxford.

Damien, P., Wakefield, J. & Walker, S. (1999). Gibbs sampling for Bayesian non-conjugate and hierarchical models using auxiliary variables. *J. Roy. Statist. Soc. Ser. B* **61**, 331–344.

Devroye, L. (1986). *Non-uniform random variate generation*. Springer-Verlag, New York.

Dykstra, R. L. & Laud, P. (1981). A Bayesian nonparametric approach to reliability. *Ann. Statist.* **9**, 356–367.

Escobar, M. D. (1994). Estimating normal means with a Dirichlet process prior. *J. Amer. Statist. Assoc.* **89**, 268–277.

Escobar, M. D. & West, M. (1995). Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* **90**, 577–588.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209–230.

Ferguson, T. S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2**, 615–629.

Ferguson, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. In *Recent advances in statistics* (eds M. H. Rizvi, J. S. Rustagi & D. Siegmund). pp. 287–302. Academic Press, New York.

Gelfand, A. E. (1999). Approaches for semiparametric Bayesian regression. In *Asymptotics, non-parametrics and time series* (ed. S. Ghosh). pp. 615–638. Marcel Dekker, New York.

Gelfand, A. E. & Kottas, A. (2002). A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models. *J. Comput. Graph. Statist.* **11**, 289–305.

Gelfand, A. E. & Mukhopadhyay, S. (1995). On nonparametric Bayesian inference for the distribution of a random sample. *Canad. J. Statist.* **23**, 411–420.

Gelfand, A. E. & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85**, 398–409.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* **57**, 1317–1339.

Groeneveld, R. A. & Meeden, G. (1984). Measuring skewness and kurtosis. *The Statistician* **33**, 391–399.

Gupta, R. C. & Langford, E. S. (1984). On the determination of a distribution by its median residual life function: a functional equation. *J. Appl. Probab.* **21**, 120–128.

Hanson, T. & Johnson, W. O. (2002). Modelling regression error with a mixture of Pólya trees. *J. Amer. Statist. Assoc.* **97**, 1020–1033.

Johnson, W. O. (1999). Survival analysis for interval data. In *The IMA volumes in mathematics and its applications, Volume 114: Diagnosis and prediction* (ed. S. Geisser). pp. 75–90. Springer, New York.

Kalbfleisch, J. D. (1978). Non-parametric Bayesian analysis of survival time data. *J. Roy. Statist. Soc. Ser. B* **40**, 214–221.

© Board of the Foundation of the Scandinavian Journal of Statistics 2003.

Kochar, S. C., Mukerjee, H. & Samaniego, F. J. (2000). Estimation of a monotone mean residual life. *Ann. Statist.* **28**, 905–921.

Kottas, A. & Gelfand, A. E. (2001). Bayesian semiparametric median regression modelling. *J. Amer. Statist. Assoc.* **96**, 1458–1468.

Kuo, L. & Mallick, B. (1997). Bayesian semiparametric inference for the accelerated failure-time model. *Canad. J. Statist.* **25**, 457–472.

Lo, A. Y. (1984). On a class of Bayesian non-parametric estimates: I. Density estimates. *Ann. Statist.* **12**, 351–357.

MacEachern, S. N. & Müller, P. (1998). Estimating mixture of Dirichlet process models. *J. Comput. Graph. Statist.* **7**, 223–238.

Maguluri, G. & Zhang, C.-H. (1994). Estimation in the mean residual life regression model. *J. Roy. Statist. Soc. Ser. B* **56**, 477–489.

Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *J. Comput. Graph. Statist.* **9**, 249–265.

Rousseeuw, P. J. & Leroy, A. M. (1987). *Robust regression and outlier detection*. John Wiley, New York.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica* **4**, 639–650.

Smith, A. F. M. & Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *J. Roy. Statist. Soc. Ser. B* **55**, 3–23.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Ann. Statist.* **22**, 1701–1762.

Walker, S. G., Damien, P., Laud, P. W. & Smith, A. F. M. (1999). Bayesian nonparametric inference for random distributions and related functions (with discussion). *J. Roy. Statist. Soc. Ser. B* **61**, 485–527.

Walker, S. G. & Mallick, B. K. (1999). A Bayesian semiparametric accelerated failure time model. *Biometrics* **55**, 477–483.

West, M., Müller, P. & Escobar, M. D. (1994). Hierarchical priors and mixture models, with application in regression and density estimation. In *Aspects of uncertainty: a tribute to D. V. Lindley* (eds A. F. M. Smith & P. Freeman). pp. 363–386. Wiley, New York.

Yang, S. (1999). Censored median regression using weighted empirical survival and hazard functions. *J. Amer. Statist. Assoc.* **94**, 137–145.

Ying, Z., Jung, S. H. & Wei, L. J. (1995). Survival analysis with median regression models. *J. Amer. Statist. Assoc.* **90**, 178–184.

Received April 2002, in final form February 2003

Athanasios Kottas, Department of Applied Mathematics and Statistics, Baskin School of Engineering, University of California at Santa Cruz, Santa Cruz, CA 95064, USA.
E-mail: thanos@ams.ucsc.edu

© Board of the Foundation of the Scandinavian Journal of Statistics 2003.