

# Bayesian Semiparametric Regression for Median Residual Life

Alan E. Gelfand and Athanasios Kottas\*

## 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 and Gelfand (2001) we formulate a semiparametric median residual life regression model induced by a semiparametric accelerated failure time regression model. We employ the median since for the rich nonparametric class of distributions over which we model, the mean need not always exist.

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 nonparametric settings. No regression modeling of median residual life appears to exist. The Bayesian modeling is developed through Dirichlet process mixing. The models are fitted using Gibbs sampling. Residual life inference is implemented extending the approach of Gelfand and Kottas (to appear).

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.

---

\*Alan E. Gelfand is Professor in the Department of Statistics at the University of Connecticut, Storrs, CT. Athanasios Kottas is Visiting Assistant Professor in the Institute of Statistics and Decision Sciences at Duke University, Durham, NC. 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.

# 1 Introduction

In analyzing survival data there is often interest not only in the survival time distribution but also in the residual survival time (or residual life) distribution. That is, we are concerned with the distribution of survival time given survival past some time say  $t_0$ . In fact, just as regression models are employed to explain survival time, such models might be sought to explain residual survival time.

In modeling survival data in the presence of covariates  $\mathbf{x}$  typically either a proportional hazard (PH) model is adopted where the cumulative hazard  $H(t; \mathbf{x}) = H_0(t) \exp(\mathbf{x}^T \beta)$  or an accelerated failure time (AFT) model is adopted where the cumulative hazard  $H(t; \mathbf{x}) = H_0(t \exp(\mathbf{x}^T \beta))$  (see, e.g., Cox and Oakes, 1984). In either case a regression is induced for the residual survival time distribution. In this regard, the AFT setting is the focus of this paper with details supplied in the ensuing sections. However, 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  $S(t; \mathbf{x}) = \exp\{-H_0(t \exp(\mathbf{x}^T \beta))\}$  whence the random variable  $U = H_0(T \exp(\mathbf{x}^T \beta))$  is distributed as  $Exp(1)$ . More importantly,  $\log T = -\mathbf{x}^T \beta + \epsilon$  where  $\epsilon = \log(H_0^{-1}(U))$ . If  $H_0$  is an arbitrary cumulative hazard function,  $\epsilon$  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  $\beta$  in the linear regression and a nonparametric component supplied by a class of distributions for  $\epsilon$ . For the PH specification we also have a natural setting for a semiparametric regression. The only difference is that now the baseline cumulative hazard  $H_0$  (or perhaps the hazard function itself) belongs to a nonparametric class of functions.

In the sequel we formulate a semiparametric median residual life regression model induced by a semiparametric AFT regression model. In particular, to allow for a sufficiently rich class of distributions for  $\epsilon$ , the mean need not exist for all members of the class. But then a mean regression for log survival time is not well defined. Since the median always exists and, if desired, can be constrained to be 0, a median regression for log survival time is less restrictive. In the special case of no covariates we can supply a fully nonparametric 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 (for example the possibility of a non-unique solution) compared to the simplicity of least squares may partially explain its limited usage as do the inferential limitations with smaller sample sizes. (See, e.g., Rousseeuw and Leroy, 1987, for a fuller 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 accelerated failure time specification (see, Ying, et al., 1995 and Yang, 1999). The estimation techniques are extensions of the noncensored case and inference is again asymptotic.

A Bayesian approach enabling exact inference given the data may be appealing. The Bayesian literature on nonparametric 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 and Smith, 1990, Smith and Roberts, 1993, Tierney, 1994), made their practical use feasible. Walker, et al. (1999) provide a summary of some of the methods in Bayesian nonparametrics. Semiparametric regression modeling is especially attractive in this context. See, for instance, Brunner (1995) and Kuo and Mallick (1997) as well as Gelfand (1999) who offers a review. For median regression modeling there is only the work of Kottas and Gelfand (2001), summarized below, as well as that of Walker and Mallick (1999), described also in Walker, et al. (1999), using a Pólya tree prior, and the very recent work of Hanson and Johnson (2001) using a mixture of Pólya tree priors.

Kottas and Gelfand (2001) propose two classes of median regression models introducing two fully Bayesian modeling approaches for the error distribution, a semiparametric and a fully nonparametric 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 behavior. Posterior inference, for both models, is carried out through the use of Gibbs sampling, with the

implementation details being relatively simple. Furthermore, using ideas from Gelfand and Kottas (to appear), 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. However, in the presence of moderate to substantial censoring, fitting the more general fully nonparametric error model becomes unstable relative to the semiparametric error model as illustrated in Kottas and Gelfand (2001). Since 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 and Gelfand (2001) can be extended to the case of median residual life regression. As a result, in a flexible semiparametric modeling framework, exact inference about essentially all features of the residual survival distribution can be obtained. We are unaware of any other work in the literature which provides this.

Estimation approaches for the mean residual life and median residual life functions do exist in the classical literature (see, e.g., Alam and Kulasekera, 1993 and 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 nonparametric settings. Moreover, essentially all the existing work ignores covariate information with the exception of Maguluri and 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.

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 modeling formulation, based on the PH specification, modeling the baseline cumulative hazard using a Gamma process. Section 7 explicitly details the modeling, the fitting approach and computational issues. Finally, section 8 considers a set of survival times with moderate censoring for patients with small cell lung cancer.

## 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

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

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

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

provided  $tS(t) \rightarrow 0$  as  $t \rightarrow \infty$ . The right-most expression in (1) shows immediately that the mean residual life need not exist. On the other hand, the median residual life  $\eta_{t_0}$  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  $\eta_{t_0} = S^{-1}(0.5S(t_0))$ .

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

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

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

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

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

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

Note that under the PH assumption, the difference  $H_0(\eta_{t_0}(\mathbf{x}_1)) - H_0(\eta_{t_0}(\mathbf{x}_2))$ , for any two combinations of covariate values  $\mathbf{x}_1$  and  $\mathbf{x}_2$ , does not depend on  $t_0$ . In fact, if  $H_0^{-1}$  can be approximated by a first order Taylor's series expansion, the same holds (approximately) for  $\eta_{t_0}(\mathbf{x}_1) - \eta_{t_0}(\mathbf{x}_2)$ . This suggests a potential drawback to an underlying PH regression model for studying median residual life. Imposing such approximate constancy regardless of  $t_0$  is intuitively very restrictive. The data set we consider in section 8 clearly illustrates this point. The AFT regression model does not force any such structure as is evident from the term  $H_0(t_0 \exp(\mathbf{x}^T \beta))$  in (3).

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,

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

where  $\gamma$  is the shape parameter of the Weibull distribution.

Of course in the semiparametric context, either the cumulative hazard or the error distribution is modeled nonparametrically so that the link function is unknown. Since we propose to model the distribution directly, rather than the hazard, we now turn to our approach for nonparametric 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 modeling the distribution of  $\log T$ . When the regression is introduced we are modeling 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 the introduction, we consider a flexible family of median zero distributions proposed in Kottas and Gelfand (2001).

### 3 Dirichlet Process Mixture Models

Following Ferguson (1973), a distribution  $G$  on  $\Theta$  follows a Dirichlet process  $DP(\nu G_0)$  if, given an arbitrary finite measurable partition,  $B_1, \dots, B_r$  of  $\Theta$ , the joint distribution

of  $(G(B_1), \dots, G(B_r))$  is *Dirichlet* $(\nu G_0(B_1), \dots, \nu G_0(B_r))$  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  $\Theta$  and  $\nu > 0$  is a precision parameter. Let  $K(\cdot; \theta)$  be a parametric family of distribution functions (c.d.f.'s), indexed by  $\theta \in \Theta$ , with associated densities,  $k(\cdot; \theta)$ . If  $G$  is proper we define the mixture distribution

$$F(\cdot; G) = \int K(\cdot; \theta) G(d\theta). \quad (5)$$

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

If  $G$  is random say  $G \sim DP(\nu G_0)$ , then  $F(\cdot; G)$  is random. Letting  $D = \{Y_i, i = 1, \dots, n\}$  denote a sample from  $F(\cdot; G)$  and using the bracket notation of Gelfand and 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  $Y_i, i = 1, \dots, n$  we introduce a latent  $\theta_i$  and assume that the  $Y_i$ 's are conditionally independent given the  $\theta_i$ 's. Assume further that the  $\theta_i$ 's are conditionally independent and identically distributed given  $G$ . As a result, marginalizing over the  $\theta_i$ 's, the  $Y_i$ 's are still independent, now conditionally on  $G$ , with joint density  $\prod_{i=1}^n f(y_i; G) = \prod_{i=1}^n \int k(y_i; \theta_i) G(d\theta_i)$ . Adding  $G \sim DP(\nu G_0)$ , possibly along with hyperpriors on  $\nu$  (see Escobar and 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, Müller and Escobar, 1994, Escobar and West, 1995, Bush and MacEachern, 1996, MacEachern and Müller, 1998 or Neal, 2000) which can be implemented to obtain samples from the posterior  $[\theta_1, \dots, \theta_n \mid D]$  resulting after the marginalization over  $G$ .

Gelfand and 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 and Kottas (to appear) provide a computational approach to obtain the entire posterior distribution for more

general functionals. Hence, exact inference is available for many population features and for comparing a feature across populations. Briefly, note that for a linear functional  $Q$ ,  $Q(F(\cdot; G)) = \int Q(K(\cdot; \theta_0))G(d\theta_0)$ . Now, instead of marginalizing over  $G$  in  $[\theta_0, \theta, G \mid D] \propto [D \mid \theta][\theta_0, \theta \mid G][G]$ , observe that this joint posterior is proportional to  $[\theta_0 \mid G][G \mid \theta][\theta \mid D]$ . Hence given the posterior sample  $\theta_b^*$ ,  $b = 1, \dots, B$ , for each  $\theta_b^*$  draw  $G_b^* \sim [G \mid \theta_b^*]$  and then  $\theta_{0lb}^* \sim G_b^*$ , for  $l = 1, \dots, L$ . Finally,  $Q_b^* \equiv Q(F(\cdot; G_b^*)) = L^{-1} \sum_{l=1}^L Q(K(\cdot; \theta_{0lb}^*))$  is a Monte Carlo integration for a realization from  $[Q(F(\cdot; G)) \mid D]$ . To obtain an approximate realization from  $[G \mid \theta_b^*]$ , which is an updated Dirichlet process (Ferguson, 1973), we use the constructive definition of Sethuraman (1994) with a partial sum approximation which 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  $\theta = (\theta^{(1)}, \theta^{(2)})$ , we might place a Dirichlet process prior on  $\theta^{(1)}$ , i.e.,  $\theta^{(1)} \sim G$  where  $G \sim DP(\nu G_0)$  with a parametric prior on  $\theta^{(2)}$ , yielding  $F(\cdot; G, \theta^{(2)}) = \int K(\cdot; \theta^{(1)}, \theta^{(2)})G(d\theta^{(1)})$ , 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  $F_T(\cdot; G, t_0)$ . In fact, we model  $Y = \log T$  given  $y_0 = \log t_0$ , i.e.,  $F_Y(\cdot; G, y_0)$ . The closure of the family of densities corresponding to distributions

$$F_Y(\cdot; G) = \int \Phi\left(\frac{\cdot - \mu}{\sigma}\right)G(d\mu, d\sigma), \quad (6)$$

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

Then, modifying slightly the notation of section 2,  $S_Y(\cdot; G, y_0) = S_Y(\cdot; G)/S_Y(y_0; G)$ ,  $\cdot > y_0$ , where  $S_Y(\cdot; G) = 1 - F_Y(\cdot; G)$  and  $\eta_{y_0}(G) = S_Y^{-1}(0.5; G, y_0)$ .

Straightforwardly,  $\eta_{t_0}(G) = \exp(\eta_{y_0}(G))$  and  $S_T(\cdot; G, t_0) = S_T(\cdot; G)/S_T(t_0; G)$  where  $S_T(t; G) = S_Y(\log t; G)$ .

How can the methodology of Gelfand and Kottas (to appear), described briefly at the end of the previous section, be used to obtain full inference regarding  $S_T(t; G, t_0)$ , for fixed  $t$ , and  $\eta_{t_0}(G)$ , i.e., the posteriors given data  $D$ ,  $[S_T(t; G, t_0) \mid D]$  and  $[\eta_{t_0}(G) \mid D]$ ? Following section 3, for a grid of  $t$  values say  $t_{(1)} < t_{(2)} < \dots < t_{(K)}$  and a posterior sample  $G_b^*$ ,  $b = 1, 2, \dots, B$ , 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  $k^{\text{th}}$  row of  $W$  provides a posterior sample from  $[S_T(t_{(k)}; G) \mid D]$ . But also, the  $b^{\text{th}}$  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 to a matrix  $W^{(1)}$ . Now the entries in the  $k^{\text{th}}$  row of  $W^{(1)}$ ,  $k > 1$ , are posterior samples from  $[S_T(t_{(k)}; G, t_{(1)}) \mid D]$  and the entries in the  $b^{\text{th}}$  column of  $W^{(1)}$  provide  $K - 1$  points on a random posterior realization of  $S_T(\cdot; G, t_{(1)})$ . 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  $[S_T(\cdot; G, t_{(1)}) \mid D]$  and  $B$  samples from  $[\eta_{t_{(1)}}(G) \mid D]$ . 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 and Kottas (to appear, section 3.2) clarifies that, since the residual life functional is defined through a ratio of bounded linear function-

als, 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 and Kottas (to appear).

## 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

$$Y = \log T = \mathbf{x}^T \beta + \epsilon \quad (7)$$

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

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

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

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

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  $\gamma$ . 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.  $\gamma$  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 and Gelfand (2001). Here we only note that to clarify how  $\gamma$  affects the skewness of (8) we might reparametrize to a skewness functional.

Since, for a general  $f$ , moments associated with (8) need not exist, the Bowley coefficient (Groeneveld and Meeden, 1984) being free of moments, is useful. This coefficient,

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

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

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

Mixing on  $\theta$  to create the semiparametric family in (10) preserves median 0 while enriching the class of models in terms of their dispersion. Attractively,  $\delta$  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  $T_i$ ,  $i = 1, \dots, n$  we now assume that the  $\epsilon_i$  are distributed according to (10). That is,  $Y_i \sim f(\cdot - \mathbf{x}_i^T \beta; G, \gamma)$  which, as in section 2, induces a linear regression on a transformed scale on the residual life curve and on 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 and 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

$t_{(1)} < t_{(2)} < \dots < t_{(n)}$  the vector  $\mathbf{r} = (r_0(t_{(1)}), r_{t_{(1)}}(t_{(2)}), \dots, r_{t_{(n-1)}}(t_{(n)}))$ , where  $r_{t_{(i-1)}}(t_{(i)}) = H_0(t_{(i)}) - H_0(t_{(i-1)})$ , has components which are, a priori, independent Gamma variables and, clearly  $\sum_{j=1}^i r_{t_{(j-1)}}(t_{(j)}) = H_0(t_{(i)})$ ,  $i = 1, \dots, n$ .

Hence with posterior samples  $\mathbf{r}_b^*$ ,  $b = 1, \dots, B$  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 be comfortable with the required interpolation. Additional sampling from the Gamma process will be needed. To introduce covariates, now  $S_{t_0}(t; \mathbf{x}) = \exp\{-(H_0(t) - H_0(t_0)) \exp(\mathbf{x}^T \beta)\}$ ,  $t > t_0$ , so we can define  $r_{t_0}(t; \mathbf{x}) = r_{t_0}(t) \exp(\mathbf{x}^T \beta)$ . Now posterior realizations from  $\mathbf{r}$  and  $\beta$  enable posterior realizations from  $\eta_{t_0}(\mathbf{x})$ .

The version one employs to model covariates should depend on the particular application and data set. In view of the structure on the median residual life, discussed in section 2, that a PH regression model imposes, the version described above might be restrictive even under the nonparametric specification for the baseline cumulative hazard function.

## 7 Modeling 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

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

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

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

$$\begin{aligned}
Y_i | \beta, \phi, \theta_i &\stackrel{ind.}{\sim} p(y_i - \mathbf{x}_i^T \beta; \theta_i, \phi), \quad i = 1, \dots, n \\
\theta_i | G &\stackrel{i.i.d.}{\sim} G, \quad i = 1, \dots, n \\
G &\sim DP(\nu G_0) \\
\beta &\sim N_{p+1}(\mu, \Sigma) \\
\phi &\sim Gamma(a, b),
\end{aligned} \tag{12}$$

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  $\beta$  are assumed a priori independent, hence  $\Sigma = diag(\sigma_0^2, \sigma_1^2, \dots, \sigma_p^2)$  with  $\mu = (\mu_0, \mu_1, \dots, \mu_p)^T$ . All the hyperparameters are assumed fixed. Full prior specification can be accomplished in a rather noninformative fashion. Essentially all that we use is a rough range  $r$  for survival time on the logarithmic scale provided either from previous studies or from the data in hand. Based on that we specify the mean  $s_2/(s_1 - 1)$  of the base measure  $G_0$  setting it 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  $\nu$  does not complicate the fitting details (Escobar and 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, values of  $\nu$  as large as 100 produced essentially identical posterior inference with smaller values. For the regression coefficients we follow the standard approach assuming  $\mu_j = 0$  and large variances  $\sigma_j^2$ ,  $j = 0, 1, \dots, p$ . Finally, we need to supply the hyperparameters  $a$  and  $b$ , corresponding to the prior of  $\phi$ . Seeking a specification that, a priori, does not favour skewness we center 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 behavior of the posterior. In the example of section 8 we take  $a = b = 2.5$ , implying a priori a range for  $\phi$  roughly from 0 to 3.5 and  $P(\phi < 1) = 0.584$ .

To obtain inference for the vector of regression coefficients, for the skewness in the error distribution, through the parameter  $\phi$ , and for functionals of the residual life distribution we need the joint posterior  $[\beta, \phi, \theta | D]$ , where  $\theta = (\theta_1, \dots, \theta_n)$  and  $D = \{y_i, \mathbf{x}_i, i = 1, \dots, n\}$ , obtained upon marginalization over  $G$ . In fact, we resort to simulation-based fitting of

the model, employing a Gibbs sampler (Gelfand and Smith, 1990) whose full conditionals are briefly described next. See Kottas and Gelfand (2001, Appendix A) for the complete implementation details.

The full conditional for  $\phi$  is a generalized inverse Gaussian distribution which can be sampled efficiently using a ratio of uniforms generation method given in Dagpunar (1988, p. 133). Following Escobar and West (1995), the full conditional for each  $\theta_i$ ,  $i = 1, \dots, n$ , is a mixed distribution with point masses at  $\theta_i = \theta_j$ ,  $j = 1, \dots, n$ ,  $j \neq i$  and continuous mass on an inverse Gamma distribution. The required weights are easily computed rendering drawing from these full conditionals straightforward. Finally, the full conditionals for the regression coefficients  $\beta_j$ ,  $j = 0, 1, \dots, p$ , 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. Again, we refer to Kottas and Gelfand (2001, section 5) for more details.

The posterior sample  $(\beta_b^*, \phi_b^*, \theta_b^*)$ ,  $b = 1, \dots, B$  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

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

where  $P(\cdot; \theta, \phi)$  is the c.d.f. of (11). Now the  $K \times B$  matrix consists of entries  $F(y_{(k)} - \mathbf{x}_0^T \beta_b^*; G_b^*, \phi_b^*)$ ,  $k = 1, \dots, K$ ,  $b = 1, \dots, B$ . 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.

## 8 Data Analysis

We illustrate the methodology with a data set involving censoring analyzed using median regression models initially by Ying, et al. (1995) and later by Yang (1999), Walker and Mallick (1999) and Kottas and 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  $i^{\text{th}}$  patient is receiving treatment A and 2 otherwise and  $x_{i2} = i^{\text{th}}$  patient's entry age. Following the suggestions of section 7 regarding the prior hyperparameters, we take  $\nu = 1$ ,  $s_1 = 2$ ,  $s_2 = 0.203$ ,  $\mu_j = 0$ ,  $\sigma_j^2 = 50$ ,  $j = 0,1,2$  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 and 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). We have also experimented with other values for  $s_2$ ,  $a$ ,  $b$ ,  $\sigma_j^2$  and  $\nu$  again obtaining stable posterior inference.

The posteriors of the regression coefficients (see Figure 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  $\delta = (1 - \phi)/(1 + \phi)$ , also given in Figure 1, indicates. See Kottas and 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 Figures 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  $\beta = (\beta_0, \beta_1, \beta_2)^T \sim N_3((0, 0, 0)^T, \text{diag}(\tau_0^2, \tau_1^2, \tau_2^2))$  and  $\gamma \sim \text{Gamma}(c, d)$ . Under this model, the median residual life has the convenient form given in (4) and therefore its posterior immediately emerges if we sample from  $[\beta, \gamma | D]$ . 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  $\tau_j^2$ ,  $j = 0,1,2$  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% posterior interval estimates for  $\beta_1$ ,  $\beta_2$  and  $\gamma$  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 6 combinations of covariate values and 3 values of  $t_0$ . Specifically, we consider both treatments for 3 values of age, 47, 63 and 74 years, corresponding to the 0.05, 0.5 and 0.95 percentiles, 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 percentiles, 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, being able to capture the bimodality in the posteriors of median survival time and median residual life at 200 days. We also note that, under the parametric model, for each fixed value of age  $x_2 = 47, 63$  or 74 years, the separation between the posteriors of  $\eta_{t_0}(x_1 = 1, x_2)$  and  $\eta_{t_0}(x_1 = 2, x_2)$  is roughly the same for the 3 values of  $t_0$  illustrating the remark of section 2 regarding the effect of the PH assumption on median residual life. Such structure is clearly not supported by the semiparametric model as Figures 2-4 indicate, the most notable being Figure 2. Hence the semiparametric analysis for median residual life provides an informal way to check the goodness of fit of a PH regression model. 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 Figure 5 we plot the posterior predictive residual survival functions at  $t_0 = 200$  and 450 days for the 6 combinations of covariate levels considered above.

## References

- Alam, K, and Kulasekera, K.B. (1993), "Estimation of the Quantile Function of Residual Life Time Distribution," *Journal of Statistical Planning and Inference*, 37, 327-337.
- Antoniak, C.E. (1974), "Mixtures of Dirichlet Processes With Applications to Nonparametric Problems," *The Annals of Statistics*, 2, 1152-1174.
- Brunner, L.J. (1995), "Bayesian Linear Regression With Error Terms That Have Symmetric Unimodal Densities," *Journal of Nonparametric Statistics*, 4, 335-348.
- Bush, C.A., and MacEachern, S.N. (1996), "A Semiparametric Bayesian Model for Randomised Block Designs," *Biometrika*, 83, 275-285.
- Cox, D.R., and Oakes, D. (1984), *Analysis of Survival Data*, Chapman & Hall/CRC.
- Dagpunar, J. (1988), *Principles of Random Variate Generation*, Oxford: Clarendon Press.
- Damien, P., Wakefield, J., and Walker, S. (1999), "Gibbs Sampling for Bayesian Nonconjugate and Hierarchical Models Using Auxiliary Variables," *Journal of the Royal Statistical Society, Ser. B*, 61, 331-344.
- Devroye, L. (1986), *Non-Uniform Random Variate Generation*, New York: Springer-Verlag.
- Dykstra, R.L., and Laud, P. (1981), "A Bayesian Nonparametric Approach to Reliability," *The Annals of Statistics*, 9, 356-367.
- Escobar, M.D. (1994), "Estimating Normal Means With a Dirichlet Process Prior," *Journal of the American Statistical Association*, 89, 268-277.
- Escobar, M.D., and West, M. (1995), "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association*, 90, 577-588.
- Ferguson, T.S. (1973), "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, 1, 209-230.

- Ferguson, T.S. (1974), "Prior Distributions on Spaces of Probability Measures," *The Annals of Statistics*, 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 and D. Siegmund, New York: Academic Press, pp. 287-302.
- Gelfand, A.E. (1999), "Approaches for Semiparametric Bayesian Regression," in *Asymptotics, Nonparametrics and Time Series*, ed. Subir Ghosh, New York: Marcel Dekker, Inc., pp. 615-638.
- Gelfand, A.E., and Kottas, A. (to appear), "A Computational Approach for Full Nonparametric Bayesian Inference under Dirichlet Process Mixture Models," *Journal of Computational and Graphical Statistics*.
- Gelfand, A.E., and Mukhopadhyay, S. (1995), "On Nonparametric Bayesian Inference for the Distribution of a Random Sample," *The Canadian Journal of Statistics*, 23, 411-420.
- Gelfand, A.E., and Smith, A.F.M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398-409.
- Geweke, J. (1989), "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, 57, 1317-1339.
- Groeneveld, R.A., and Meeden, G. (1984), "Measuring Skewness and Kurtosis," *The Statistician*, 33, 391-399.
- Gupta, R.C., and Langford, E.S. (1984), "On the Determination of a Distribution by its Median Residual Life Function: A Functional Equation," *Journal of Applied Probability*, 21, 120-128.
- Hanson, T., and Johnson, W.O. (2001), "Modeling Regression Error with a Mixture of Pólya Trees," Technical Report, University of New Mexico.

- 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, New York: Springer, pp. 75-90.
- Kalbfleisch, J.D. (1978), "Non-parametric Bayesian Analysis of Survival Time Data," *Journal of the Royal Statistical Society, Ser. B*, 40, 214-221.
- Kocher, S.C., Mukerjee, H., and Samaniego, F.J. (2000), "Estimation of a Monotone Mean Residual Life," *The Annals of Statistics*, 28, 905-921.
- Kottas, A., and Gelfand, A.E. (2001), "Bayesian Semiparametric Median Regression Modeling," *Journal of the American Statistical Association*, 96.
- Kuo, L., and Mallick, B. (1997), "Bayesian Semiparametric Inference for the Accelerated Failure-Time Model," *The Canadian Journal of Statistics*, 25, 457-472.
- Lo, A.Y. (1984), "On a Class of Bayesian Nonparametric Estimates: I. Density Estimates," *The Annals of Statistics*, 12, 351-357.
- MacEachern, S.N., and Müller, P. (1998), "Estimating Mixture of Dirichlet Process Models," *Journal of Computational and Graphical Statistics*, 7, 223-238.
- Maguluri, G., and Zhang, C-H. (1994), "Estimation in the Mean Residual Life Regression Model," *Journal of the Royal Statistical Society, Ser. B*, 56, 477-489.
- Neal, R. M. (2000), "Markov Chain Sampling Methods for Dirichlet Process Mixture Models," *Journal of Computational and Graphical Statistics*, 9, 249-265.
- Rousseeuw, P.J., and Leroy, A.M. (1987), *Robust Regression and Outlier Detection*, New York: John Wiley.
- Sethuraman, J. (1994), "A constructive definition of Dirichlet priors," *Statistica Sinica*, 4, 639-650.
- Smith, A.F.M., and Roberts, G.O. (1993), "Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society, Ser. B*, 55, 3-23.

- Tierney, L. (1994), “Markov Chains for Exploring Posterior Distributions” (with discussion), *The Annals of Statistics*, 22, 1701-1762.
- Walker, S.G., Damien, P., Laud, P.W., and Smith, A.F.M. (1999), “Bayesian Nonparametric Inference for Random Distributions and Related Functions” (with discussion), *Journal of the Royal Statistical Society, Ser. B*, 61, 485-527.
- Walker, S.G., and Mallick, B.K. (1999), “A Bayesian Semiparametric Accelerated Failure Time Model,” *Biometrics*, 55, 477-483.
- West, M., Müller, P., and 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 and P. Freeman, New York: Wiley, pp. 363-386.
- Yang, S. (1999), “Censored Median Regression Using Weighted Empirical Survival and Hazard Functions,” *Journal of the American Statistical Association*, 94, 137-145.
- Ying, Z., Jung, S.H., and Wei, L.J. (1995), “Survival Analysis With Median Regression Models,” *Journal of the American Statistical Association*, 90, 178-184.

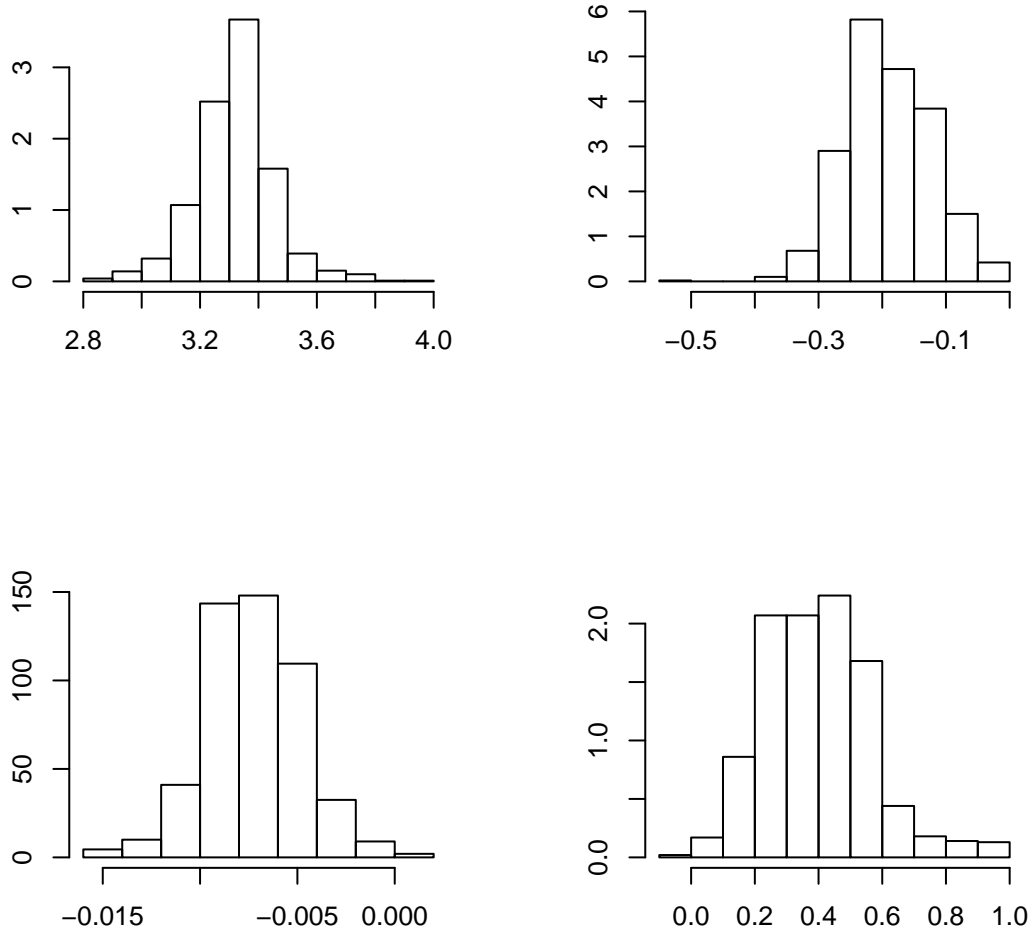


Figure 1: Posteriors of regression coefficients ( $\beta_0$  upper left,  $\beta_1$  upper right,  $\beta_2$  lower left) and skewness functional  $\delta$  (lower right) under the semiparametric regression model.

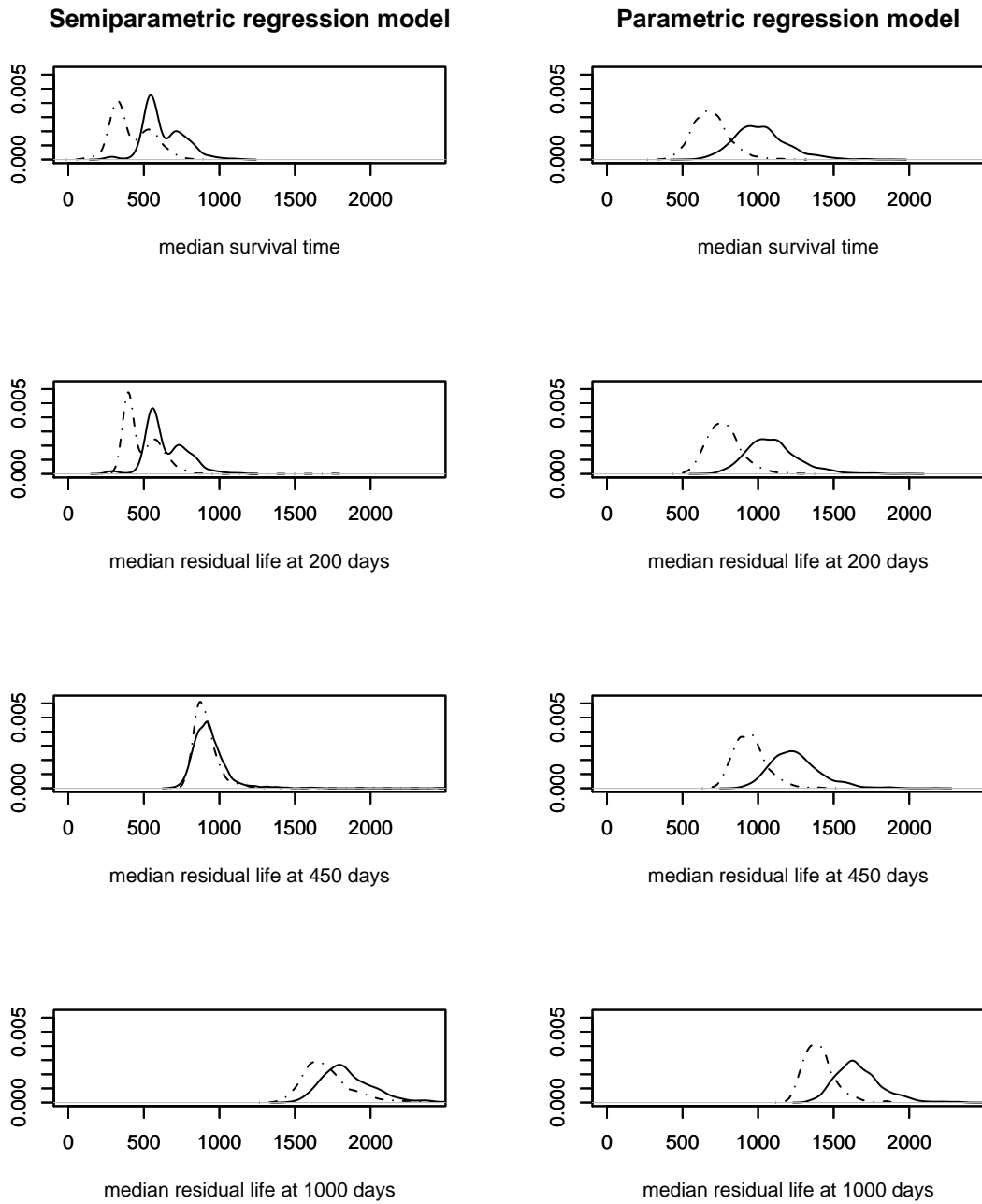


Figure 2: Posteriors of median survival time and median residual life at  $t_0 = 200, 450$  or 1000 days, for a 47 year old patient receiving treatment B (dashed-dotted lines) or treatment A (solid lines). The left column corresponds to the semiparametric model and the right column to the parametric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

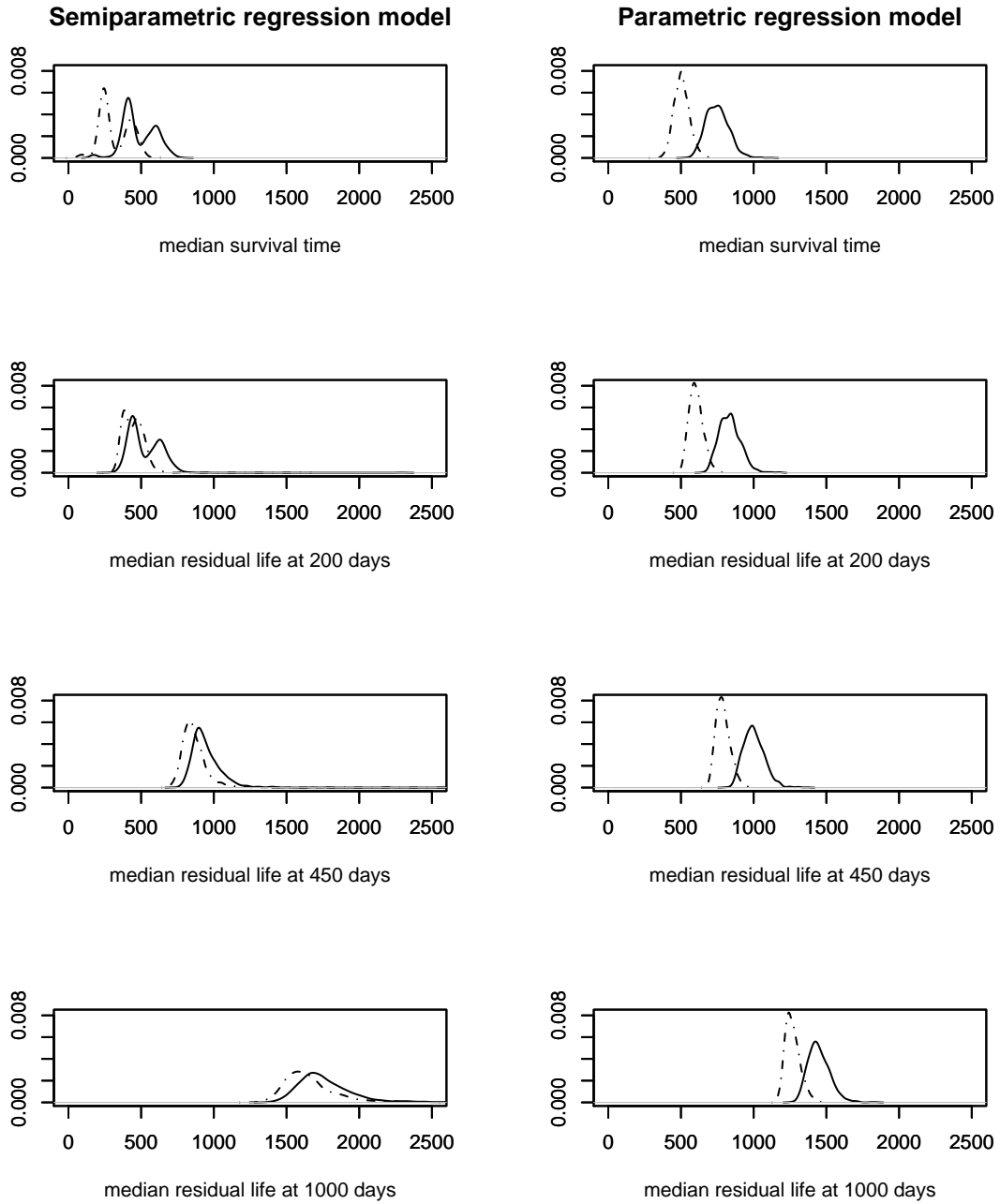


Figure 3: Posteriors of median survival time and median residual life at  $t_0 = 200, 450$  or 1000 days, for a 63 year old patient receiving treatment A (solid lines) or treatment B (dashed-dotted lines). The left column has the posteriors under the semiparametric model and the right under the parametric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

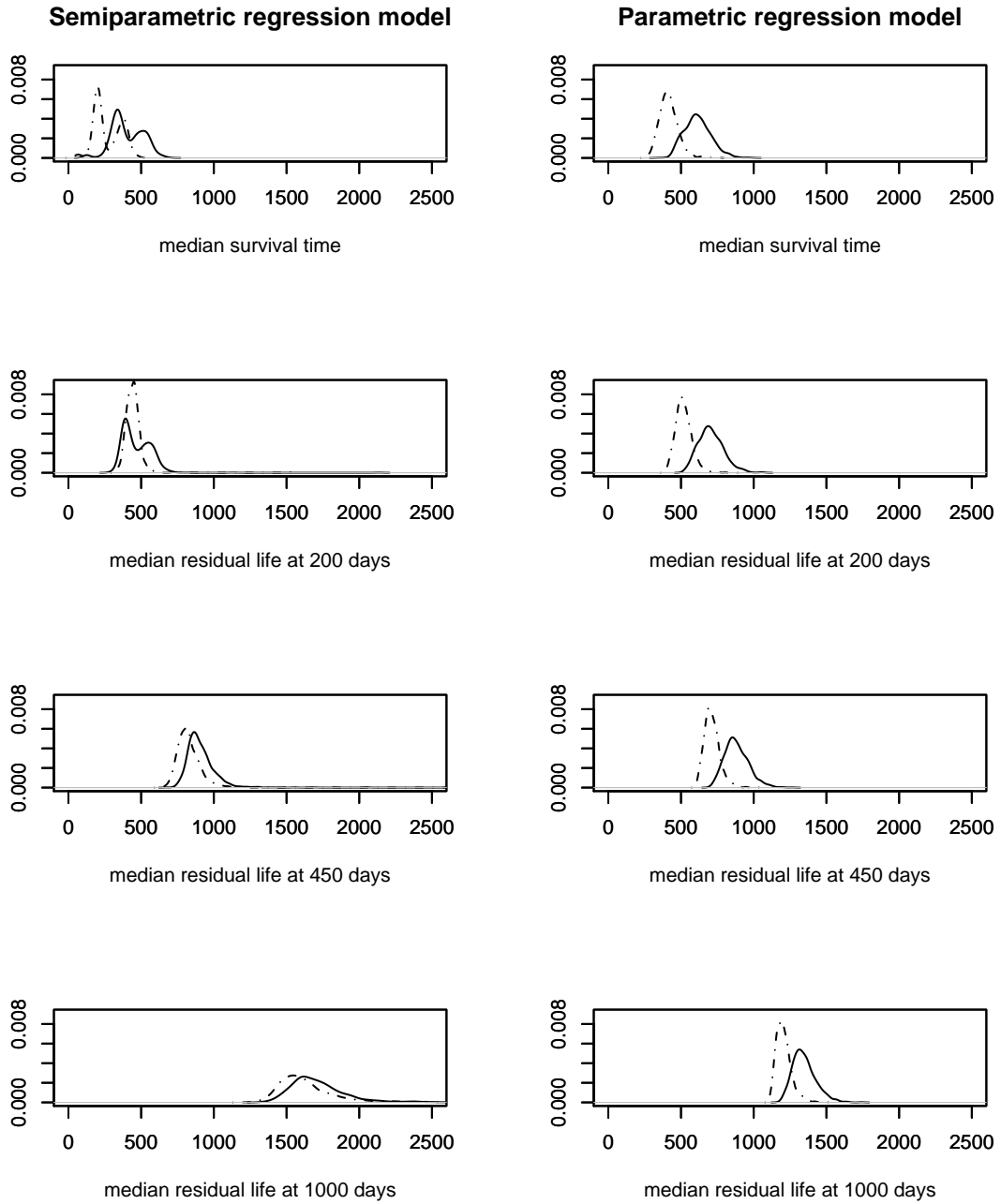


Figure 4: Posteriors of median survival time and median residual life at  $t_0 = 200, 450$  or 1000 days, for a 74 year old patient receiving treatment A (solid lines) or treatment B (dashed-dotted lines). The left column has the posteriors under the semiparametric model and the right under the parametric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

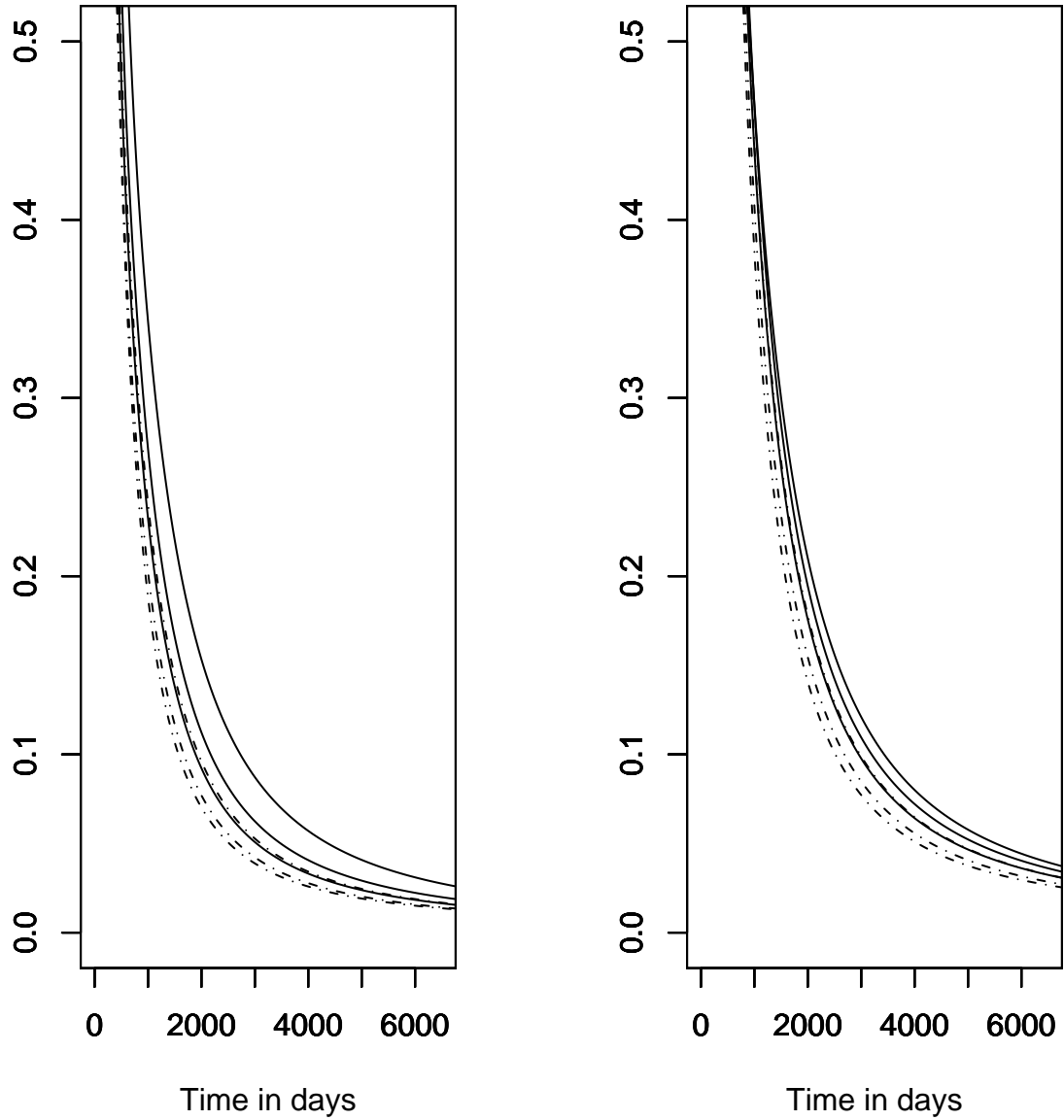


Figure 5: Under the semiparametric regression model, posterior predictive residual survival functions at  $t_0 = 200$  days (left) and 450 days (right) for treatment A (solid lines) and treatment B (dashed-dotted lines) at ages 47, 63 and 74 (upper, middle and lower curve, respectively, in each case).