Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Biometrics. Author manuscript; available in PMC 2015 May 14.

Published in final edited form as:

Biometrics. 2014 March ; 70(1): 192–201. doi:10.1111/biom.12104.

Accelerated Hazards Model based on Parametric Families Generalized with Bernstein Polynomials

Yuhui Chen[1,*], **Timothy Hanson**[1,**], and **Jiajia Zhang**[2,***]

[1]Department of Statistics, University of South Carolina, Columbia, SC, U.S.A

[2]Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC, U.S.A

Summary

A transformed Bernstein polynomial that is centered at standard parametric families, such as Weibull or log-logistic, is proposed for use in the accelerated hazards model. This class provides a convenient way towards creating a Bayesian non-parametric prior for smooth densities, blending the merits of parametric and non-parametric methods, that is amenable to standard estimation approaches. For example optimization methods in SAS or R can yield the posterior mode and asymptotic covariance matrix. This novel nonparametric prior is employed in the accelerated hazards model, which is further generalized to time-dependent covariates. The proposed approach fares considerably better than previous approaches in simulations; data on the effectiveness of biodegradable carmustine polymers on recurrent brain malignant gliomas is investigated.

Keywords

Accelerated Hazard Model; Bayesian Nonparametric Prior; Survival Analysis; Time Dependent Covariate

1. Introduction

As an alternative to proportional hazards (Cox, 1972) and accelerated failure time (e.g. Cox and Oakes, 1984) models, the accelerated hazards (AH) model (Chen and Wang, 2000) was proposed for its ability to capture the gradual effects of a treatment; this lag period often exists before a treatment is fully effective (e.g., Zucker and Lakatos, 1990). Denote $h_{\mathbf{x}}(\cdot)$ as the hazard for an individual with covariate vector $\mathbf{x}$ and $h_0(\cdot)$ the baseline hazard for an individual with $\mathbf{x} = \mathbf{0}$. The proportional hazards (PH) model can be written as $h_{\mathbf{x}}(t) = h_0(t)e^{\boldsymbol{\beta}\mathbf{x}}$, the accelerated failure time (AFT) model is expressed as $h_{\mathbf{x}}(t) = h_0(te^{\boldsymbol{\beta}\mathbf{x}})e^{\boldsymbol{\beta}\mathbf{x}}$, and the AH model is

*chen33@email.sc.edu
**hansont@email.sc.edu
***jzhang@mailbox.sc.edu

8. Supplementary Materials

Web Appendix results referenced in Section 2, a supplementary Table and Figures from Section 2, and sample R Code referenced in Section 3 are available with this paper at the Biometrics website on Wiley Online Library.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$h_{\mathbf{x}}(t)=h_0(te^{\mathbf{x}'\beta}), \quad (1)$$

implying density and survival functions

$$f_{\mathbf{x}}(t)=S_0(e^{\mathbf{x}'\beta}t)^{e^{-\mathbf{x}'\beta}-1}f_0(e^{\mathbf{x}'\beta}t), \quad S_{\mathbf{x}}(t)=S_0(e^{\mathbf{x}'\beta}t)^{e^{-\mathbf{x}'\beta}}, \quad (2)$$

where $\boldsymbol{\beta}$ is a vector of unknown regression parameters. The exponentiated *j*th regression effect $e^{-\beta_j}$ is interpreted as a factor of how much more (or less) time is required to reach the same failure risk when the *j*th predictor x_j is increased by one.

One of the main differences among these three models is the risk effect at the initial time $t = 0$. It is easily seen that the PH and AFT models assume that, in general, $h_{\mathbf{x}}(0) \neq h_{\mathbf{z}}(0)$ for $\mathbf{x} \neq \mathbf{z}$, implying an immediate treatment effect, whereas $h_{\mathbf{x}}(0) = h_{\mathbf{z}}(0)$ under AH, allowing the treatment to take effect gradually. Note that this 'gradual effect' may not be appropriate for other non-treatment covariates such as cancer stage, e.g. stage IV versus II at diagnosis could substantially impact the hazard immediately; here the PH model may be more appropriate.

The limited application of the AH model is due to the lack of efficient and reliable estimation methods. Chen and Wang (2000) estimate the regression effects $\boldsymbol{\beta}$ via non-smooth rank-type estimating equations, and Chen (2001) improved the rank-based variance estimation procedure. Zhang et al. (2011) proposed an efficient semiparametric estimation method for the AH model based on a kernel-smoothed approximation of the profile likelihood function. However, profile likelihood methods may have convergence issues and often underestimate the variance when the sample size is small, or even moderate; we find this to be true in simulations in Section 5.

Historically, much of Bayesian survival analysis has considered variants of the PH model built from independent increments priors on the baseline (e.g. Kalbfleisch, 1978; Ibrahim, Chen, and Sinha, 2001). This paper develops a Bayesian semiparametric AH model and novel generalization to allow time-dependent covariates, built on a suitably transformed Bernstein polynomial. Inference is straightforward to obtain using standard maximization routines; we make several recommendations for obtaining inference in R, and make code available to interested users in an online appendix.

Bernstein (1912) gave a constructive proof of the Weierstrauss theorem using what are now termed 'Bernstein polynomials'. The Bernstein polynomial is a type of Bézier curve, and more generally a special case of a B-spline with certain restrictions on the B-spline knots. A statistician recognises a Bernstein polynomial basis function as a beta density with integer parameters. We use Bernstein polynomials as a means to easily add flexibility to existing parametric families, in this paper the Weibull and log-logistic distributions. Random probability measures G that have a Polya tree prior (Lavine, 1992) are centered at a distribution G_0 in the sense that $E\{G(A)\} = G_0(A)$ for any measurable A. However, the Polya tree posterior joint density is not continuous, often making inference challenging.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Furthermore, Polya tree densities, like histograms, have discontinuities; some find this troubling. We seek to build a flexible nonparametric prior that is centered at a given parametric family, but with a smooth likelihood, allowing for the use of standard maximization techniques and asymptotic inference via normal approximations.

Starting with a parametric family as an initial 'washed canvas' then adding detail though the Bernstein polynomial has two advantages. First, there may be sound theoretical reasons why a particular family is approximately appropriate. For example, the Weibull distribution is derived as the asymptotic distribution of the first failure in a series of independent components; when the first component fails, the whole system fails. Second, there is existing, well-tested software for fitting parametric models, and this software can serve as a source of initial values, and customized sequential fitting procedures for the nonparametric extension – this is our approach.

The purpose of this paper is three-fold. First, we introduce a generalization of existing parametric families that does not immediately "take the leap" into infinite-dimensional in Section 2; second, we apply this novel approach to a neglected survival model: the AH model in Section 3. Finally, we suggest an approach to extend the AH model to time-dependent covariates that are step-functions in Section 4. Sections 5 and 6 provide evidence that the procedure works in simulated and real data situations, including data on the effectiveness of biodegradable carmustine polymers on recurrent brain malignant gliomas. Section 7 concludes the paper.

2. Transformed Bernstein Polynomial Priors

2.1 Bernstein Polynomial Prior

Petrone (1999a,b) is the first to give a comprehensive treatment to using Bernstein polynomials for Bayesian density estimation. Define the beta density with parameters a and b as

$$\beta(x;a,b)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}I_{[0,1]}(x),$$

with $I_A(\cdot)$ the usual indicator function for the set A and $\Gamma(\cdot)$ the usual gamma function. A Bernstein polynomial with J components is given by

$$f_{\mathbf{w}_J}(x)=\sum_{j=1}^{J}w_{J_j}\beta(x;j;J-j+1)\overset{def}{=}\sum_{j=1}^{J}w_{J_j}b_{J_j}(x) \quad (3)$$

is used, where $\mathbf{w}_J = (w_{J1}, \ldots, w_{JJ})'$ is a vector of weights summing to unity such that $0 < w_j < 1$ for $j = 1, \ldots, J$.

Integrating (3) gives

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$F_{\mathbf{w}_J}(x)=\sum_{j=1}^{J} w_{Jj} B_{Jj}(x) \stackrel{def}{=} \sum_{j=1}^{J} w_{Jj} \int_0^x b_{Jj}(s)ds. \quad (4)$$

A natural way to obtain a random cumulative distribution function (cdf), $F(x)$, on [0, 1] is to assign a joint distribution to $(J, \mathbf{w}_J)$. The (marginal) prior probability for J is denoted $p(J)$. For a given J, the vector of weights $\mathbf{w}_J$ conditionally follows a Dirichlet distribution, written as

$$\mathbf{w}_J | J \sim \text{Dirichlet}(\alpha_{J1}, \ldots, \alpha_{JJ}), \quad J \sim p(J). \quad (5)$$

Most papers utilizing a finite mixture of continuous components assume the weights $\mathbf{w}_J$ are Dirichlet. An exception is Petrone (1999a) who instead chooses $w_{Jj} = F(j/J) - F((j-1)/J)$ for $j = 1, \ldots, J$ and takes $F \sim DP(MF_0)$, a Dirichlet process with precision M centered at F_0, typically $F_0 = U(0, 1)$.

Bernstein polynomials have been used for density estimation on bounded domains, typically by transforming data to lie in the interval [0, 1]. We now consider an approach to estimating a suitable transformation from the data automatically, including transformations for data that lie on unbounded domains. A useful property of Bernstein polynomials (and more generally B-splines) is that if the weights are identical, $w_1 = \cdots = w_J = J^{-1}$, then $f_{\mathbf{w}J}(x) = 1$ and $F_{\mathbf{w}J}(x) = x$ for $x \in [0, 1]$; so by linearity if $E(w_j) = J^{-1}$, for all $j \in (1, 2, \ldots, J)$, then $E\{f_{\mathbf{w}J}(x)\} = 1$ and $E\{F_{\mathbf{w}J}(x)\} = x$. Consider a standard parametric family of survival densities $\{S_\theta(\cdot) : \theta \in \Theta\}$; we use the Weibull $S_\theta(t) = \exp(-\theta_2 t^{\theta_1})$ and log-logistic $S_\theta(t) = \left[1 + \left(\frac{t}{\theta_1}\right)^{\theta_2}\right]^{-1}$ families in applications. Define a random survival function $S_0(\cdot)$ for a baseline group with all covariates set equal to zero as

$$S_0(t) = F_{\mathbf{w}_J}\{S_\theta(t)\} = \sum_{j=1}^{J} w_{Jj} B_{Jj}\{S_\theta(t)\}. \quad (6)$$

Clearly, under (5), if $\boldsymbol{\alpha}_J = M\mathbf{1}_J$, where $\mathbf{1}_J$ is a vector of J one's and $M > 0$, $E\{S_0(t)\} = S_\theta(t)$ for all $t > 0$. For notational simplicity, we suppress the dependence of $S_0(t)$ on $(J, \mathbf{w}_J)$ and θ, e.g. $S_0(t) = S_0(t|J, \mathbf{w}_J, \theta)$. The prior acts somewhat like a B-spline, but where knot locations are more dense in areas of higher mass under the parametric family G_θ, also the basis functions do not have finite support unless G_θ does. So the prior automatically picks knot locations guided by an overall parametric G_θ but then allows substantial deviations from G_θ through the additional parameters $\mathbf{w}_J$.

We say that the random $S_0(\cdot)$ in (6) has a transformed Bernstein polynomial prior (TBPP). Unlike Polya tree priors (Lavine, 1992), the random densities

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$f_0(t)=\sum_{j=1}^{J} w_{J_j} b_{J_j}\{S_\theta(t)\} f_\theta(t) \quad (7)$$

are quite smooth, and therefore amenable to standard estimation approaches involving asymptotic normal approximations; also note that $E\{f_0(t)\} = f_\theta(t)$ when $\alpha_J = M\mathbf{1}_J$. Inference is easy and relatively quick to obtain in contrast to Markov chain Monte Carlo (MCMC) algorithms, which require iteratively sampling the posterior. However, like Polya trees, S_0 is centered at a given parametric family, yielding more efficient estimation when the parametric model approximately holds, but conveying substantial robustness when data deviate markedly from the parametric family. Furthermore, a fit of the parametric model provides excellent starting values for the estimation routine; this is especially so for the AH model with a Weibull centering distribution.

2.2 Prior specification

The number of basis functions J provides scale or resolution for possible departures of S_0 from S_θ, and the parameter M governs how stochastically "pliable" S_0 is relative to S_θ. We fix M at a reasonable value and leave polynomial order J to be estimated from the data. As mentioned in the previous section, several authors have used small, fixed values, e.g. $J = 3$ or $J = 5$. Others have considered a prior on J, including Mallick and Walker (2003); Chang et al. (2005); and Petrone (1999a,b), who implements very clever Gibbs sampling stratgies by introducing latent variables. These first two use reversible jump (Green, 1995) to sample J in a trans-dimensional Gibbs sampler. Walker and Mallick (2003) use a Poisson prior for J with mean 4 truncated to $J \leq 16$; Chang et al. (2005) consider $J \leq 10$ and $J \leq 20$ in simulations and examples. We also consider a prior $J \sim p(J)$ with an upper bound $J \leq K$, and choose $K = 15$ based on simulations and the discussion in the online appendix. Table 1 in the online appendix gives the mean L_1 distance of (7) from the centering uniform distribution for some values of M and J. When $J = 15$, $M = 1$ and $M = 2$ give average L_1 distances of 0.29 and 0.21, respectively, *a priori* routinely allowing 20% or 30% of the mass to be moved away from S_θ.

For the sample sizes used in this paper, $J \leq 15$ has worked very well. In general, however, simulations indicate that K should be mildly increased with the sample size n to accommodate greater resolution. In such cases, a joint prior on J and M would be ideal, perhaps one that fixes the mean and variance of the L_1 distance $\|f_{\mathbf{w}_J} - 1\|_1$. Under mild conditions, Theorem 4 in Petrone and Wasserman (2002) shows that the posterior density converges (as $n \to \infty$) to the $\tilde{f_K}$ that minimizes the Kullback-Leibler divergence to the true density for a fixed J.

Note that $J = 15$ introduces 14 additional free parameters to the overall shape given by θ. For the TBPP, capping J off at $J \leq K \approx n/10$ implies that roughly 10 observations inform each weight in $\mathbf{w}_K$ if S_0 approximately follows S_θ. We have found simply setting $K = 15$ to give good results for the sample sizes considered in this paper. For much higher sample sizes, Petrone (1999a,b) considered higher values of K, e.g. $K = 100$.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Although we consider a at prior $p(\boldsymbol{\theta}, \boldsymbol{\beta}) \propto 1$ in the rest of the paper, it is possible to incorporate some prior information for $(\boldsymbol{\theta}, \boldsymbol{\beta})$ through the centering Weibull model. When $\mathbf{w}_J = \mathbf{1}_J/J$, i.e. $\mathbf{w}_J$ is fixed at its prior mean, the Weibull regression model is obtained. Thus any prior specification approach for Weibull regression would reasonably work, for example the conditional median approach of Bedrick, Christensen, and Johnson (2000).

3. Accelerated Hazards Model

Denote the lifetime of patient i as t_i^* with right-censoring time $\mathbf{c} = (c_1, \ldots, c_n)'$. The observed event time is $t_i = \min\{t_i^*, c_i\}$; take the collection of all n event times to be $\mathbf{t} = (t_1, t_2, \ldots, t_n)'$ The censoring indicator for the ith subject is $\delta_i = 0$ if $t_i^* > c_i$, otherwise $\delta_i = 1$. The covariates for the ith subject is defined as $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})'$ of dimension p. The whole of the observed data are included in the set $\mathcal{D} = \{(t_i, \delta_i, \mathbf{x}_i)\}_{i=1}^n$.

Fix J. Momentarily dropping the subscript on $(\mathbf{w}_J, \boldsymbol{\theta}_J, \boldsymbol{\beta}_J)$, the conditional likelihood function for the AH model is written:

$$\mathcal{L}(\mathbf{w}, \boldsymbol{\theta}, \boldsymbol{\beta}) = \prod_{i=1}^{n} S_0(e^{\mathbf{x}_i'\boldsymbol{\beta}} t_i) e^{\frac{1}{\mathbf{x}_i'\boldsymbol{\beta}} - \delta_i} f_0(e^{\mathbf{x}_i'\boldsymbol{\beta}} t_i)^{\delta_i} \quad (8)$$

with $S_0(\cdot)$ and $f_0(\cdot)$ defined in (6) and (7) respectively. Assuming independent priors $p(\boldsymbol{\theta})$, $p(\boldsymbol{\beta})$, and $p(\mathbf{w})$, the posterior distribution of $(\mathbf{w}, \boldsymbol{\theta}, \boldsymbol{\beta})$ given J is

$$p(\mathbf{w}, \boldsymbol{\theta}, \boldsymbol{\beta} | J, \mathcal{D}) \propto \mathcal{L}(\mathbf{w}, \boldsymbol{\theta}, \boldsymbol{\beta}) p(\boldsymbol{\theta}) p(\boldsymbol{\beta}) p(\mathbf{w}). \quad (9)$$

We assume the improper prior $p(\boldsymbol{\theta}, \boldsymbol{\beta}) \propto$; 1 independent of $\mathbf{w} \sim$ Dirichlet($M\mathbf{1}_J$); here we take $M = 1$ or $M = 2$ following the discussion in Section 2. Let $\boldsymbol{\psi}_J = (\mathbf{w}_J', \boldsymbol{\theta}_J', \boldsymbol{\beta}_J')'$ be all of the model parameters. The posterior mean $\hat{\boldsymbol{\psi}}_J$ of parameters $\boldsymbol{\psi}_J$ is directly obtained from maximizing (9). Since the $S_{\boldsymbol{\theta}}(t)$ are assumed to be sufficiently smooth in $\boldsymbol{\theta}$ and t, standard asymptotic theory requires that $(\boldsymbol{\psi}_J | J, \mathcal{D})$ approximately follows a multivariate normal distribution with an approximate covariance matrix which is the inverse of the Hessian matrix evaluated at $\hat{\boldsymbol{\psi}}_J$, i.e.

$$(\boldsymbol{\psi}_J | J, \mathcal{D}) \sim N_d \left(\hat{\boldsymbol{\psi}}_J, \sum_{\hat{\boldsymbol{\psi}}_J} \right) \quad (10)$$

where N_d is a multivariate normal distribution with dimension $d = p + J - 1 + q$; q is the number of parameters in $\boldsymbol{\theta}$, in this paper $q = 2$.

The posterior of $(J | \mathcal{D})$ can be evaluated using Bayes' rule. For any value of $\boldsymbol{\psi}_J$,

$$p(J | \mathcal{D}) \propto \frac{\mathcal{L}(\boldsymbol{\psi}_J) p(\boldsymbol{\beta}) p(\boldsymbol{\theta}) p(\mathbf{w}_J) p(J)}{p(\boldsymbol{\psi}_J | J, \mathcal{D})} \quad (11)$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Evaluating the approximation (10) at the posterior mode $\hat{\boldsymbol{\psi}}_J$ simplifies (11) to

$$\hat{p}(J|\mathscr{D}) \propto \mathscr{L}(\hat{\boldsymbol{\psi}}_J)p(\hat{\boldsymbol{w}}_J)p(J)|2\pi\sum_{\hat{\psi}_J}|^{1/2}. \quad (12)$$

We now describe how to obtain inference for the regression effects $\boldsymbol{\beta}$ and common functionals of survival. We assume an upper bound on the degree of the polynomial $J \leq K$, so that $p(J) = 0$ for $J > K$. To simplify notation, denote $p(J|\mathcal{D}) = p_J$ in (12), where $J \in \{1, 2, \ldots, K\}$, and $E(\boldsymbol{\beta}|J, \mathcal{D}) = \hat{\boldsymbol{\beta}}_J$ and $cov(\boldsymbol{\beta}|J, \mathcal{D}) = \boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}_J}$, both well-approximated using standard asymptotic theory applied to (9); $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}_J}$ is a $p \times p$ submatrix of $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\psi}}_J}$. The posterior mean of $\boldsymbol{\beta}$ is obtained via iterated expectation:

$$E(\boldsymbol{\beta}|\mathscr{D}) = E_J(E(\boldsymbol{\beta}|J, \mathscr{D})) = \sum_{J=1}^{K} E(\boldsymbol{\beta}|J, \mathscr{D})p(J|\mathscr{D}) \approx \sum_{J=1}^{K} \hat{\boldsymbol{\beta}}_J \hat{p}_J \quad (13)$$

The posterior covariance is given by iterated covariance:

$$cov(\hat{\boldsymbol{\beta}}|\mathscr{D}) = E_J(cov(\boldsymbol{\beta}|J, \mathscr{D})) + cov_J(E(\boldsymbol{\beta}|J, \mathscr{D})) \quad (14)$$

where

$$E_J(cov(\boldsymbol{\beta}|J, \mathscr{D})) = \sum_{J=1}^{K} \hat{p}_J \sum_{\hat{\boldsymbol{\beta}}_J},$$
$$cov_J(E(\boldsymbol{\beta}|J, \mathscr{D})) = \sum_{J=1}^{K} \hat{p}_J \hat{\boldsymbol{\beta}}_J (\hat{\boldsymbol{\beta}}_J)' - E(\boldsymbol{\beta}|\mathscr{D})E(\boldsymbol{\beta}|\mathscr{D})'.$$

Credible intervals for regression coefficients are found numerically using a grid search, noting that the jth coefficient β_j follows a mixture of K normal distributions. Using obvious notation, $\beta_j|\mathscr{D} \sim \sum_{J=1}^{K} \hat{p}_J N(\hat{\beta}_{Jj}, \hat{\sigma}^2_{Jj})$.

Survival function estimates are obtained similarly, by first computing survival conditionally on J; then the unconditional survival function estimate will be obtained by averaging with respect to the posterior J. The estimated survival function given J is obtained through (2):

$$\hat{S}_{\mathbf{x}}(t|J, \mathscr{D}) = \left[\sum_{j=1}^{J} \hat{w}_{Jj} B_{Jj}\{S_{\hat{\boldsymbol{\theta}}_J}(e^{\mathbf{x}'\hat{\boldsymbol{\beta}}_J} t)\}\right]^{1/e^{\mathbf{x}'\hat{\boldsymbol{\beta}}_J}},$$

and so $\hat{S}_{\mathbf{x}}(t|\mathscr{D}) = \sum_{J=1}^{K} \hat{p}_J \hat{S}(t|J, \mathscr{D})$. Similarly, $\hat{f}_{\mathbf{x}}(t|\mathscr{D}) = \sum_{J=1}^{K} \hat{p}_J \hat{f}(t|J, \mathscr{D})$ and $\hat{h}_{\mathbf{x}}(t|\mathcal{D}) = \hat{f}_{\mathbf{x}}(t|\mathcal{D})/\hat{S}_{\mathbf{x}}(t|\mathcal{D})$.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

3.1 Computational Issues

Petrone (1999a,b) proposed MCMC procedures for bounded density estimation using Bernstein-Dirichlet priors. These algorithms perform well, but are inefficient for large sample sizes, as they require the updating a latent component indicator for each datum. Furthermore, Petrone's data-augmentation schemes will not immediately work in this regression context as the observations from S_0 are not directly observed. Petrone and Wasserman (2002) proposed a more efficient estimation method called sieve maximum likelihood estimation (SMLE) which can also handle the parameter space dimension changing with J. This method applies maximum likelihood to each model with fixed J, then averages conditional estimates with respect to model-based weights computed either from AIC or BIC. Our estimation method is similar to the SMLE in Petrone and Wasserman (2002), but within the context of a full regression model on an unbounded domain, rather than just density estimation, and we apply a true Bayesian approach to estimating the weights $P(J|\mathcal{D})$. Specifically, we consider the prior $p(J) \propto \nu^{J-1}$ where $\nu \in (0, 1)$, penalizing for more complexity in the form of larger J. We have found this prior to work very well in simulations and real data analyses. Alternatively, one might rather consider just picking one J (e.g. $J = 15$ or $J = 20$) enriching the baseline parametric model indexed by $\boldsymbol{\theta}$ with (e.g. 14 or 19) additional parameters to add detail to the overall shape provided by $S_{\boldsymbol{\theta}}$. Picking one largish J could be justified because Bernstein polynomials have a very important and fascinating property: all Bernstein polynomials of degree 1, 2, 3, …, $J-1$ are included in the Bernstein polynomial of degree J; i.e. lower degrees are formally nested within larger. In simulations and the data analysis in Sections 5 and 6.1 we use model averaging with $J \leq 15$; in the data analysis of Section 6.2 we simply pick $J = 15$.

To obtain the posterior mode $\hat{\boldsymbol{\psi}}_J$ and accompanying Hessian matrix, there are canned functions in most statistical software. We tried several packages in R and found the ucminf function in the ucminf library (written by Hans B. Nielsen, Stig B. Mortensen, and Douglas Bates) to be the most stable.

To facilitate model fitting we transformed all parameters to lie in $\mathbb{R} = (-\infty, \infty)$. In both Weibull and log-logistic families, the elements of $\boldsymbol{\theta} = (\theta_1, \theta_2)'$ are positive; we work with the natural log of these parameters instead. Relaxing the restriction on $\mathbf{w}_J$ is a bit more complicated. Borrowing from the multivariate logistic normal distribution (Aitchison and Shen, 1980), we instead work with $\mathbf{v}_J = (v_1, \ldots, v_{J-1})'$ through

$$w_j = \frac{e^{v_j}}{1+\sum_{k=1}^{J-1} e^{v_k}} \text{ for } j = 1, \ldots, J, \quad (15)$$

defining $v_J = 0$. Under $\mathbf{w}_J \sim \text{Dirichlet}(M\mathbf{1}_J)$, the induced prior on $\mathbf{v}_J$ is:

$$p(\mathbf{v}_J) = \frac{\Gamma(MJ)}{\Gamma(M)^J} \prod_{j=1}^{J} \left[\frac{e^{v_j}}{1+\sum_{k=1}^{J-1} e^{v_k}} \right]^M. \quad (16)$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Note that this is the only part of the posterior density inolving M; placing a prior on M would yield a full conditional distribution of (16) times the prior $p(M)$. Working with $\hat{\boldsymbol{\psi}}_J = (\boldsymbol{\beta}, \log \theta_1, \log \theta_2, \nu_1, \ldots, \nu_{J-1})'$ enhances the asymptotic normality approximation for the posterior of $\tilde{\boldsymbol{\psi}}_J$ used in (12).

Optimization routines require good starting values to be efficient. In our simulations and data analysis in Sections 5 and 6.1, we center the TBPP at the Weibull distribution $S_{\boldsymbol{\theta}}(t) = e^{-\lambda t^{\alpha}}$; for $J = 1$, the baseline survival function S_0 is then reduced to this centering distribution. Under the accelerated hazard model with Weibull baseline $S_{\boldsymbol{\theta}}$, $S_{\mathbf{x}}(t) = S_0(e^{\mathbf{x}'\beta} t)^{\frac{1}{e^{\mathbf{x}'\beta}}} = e^{-\lambda (e^{\mathbf{x}'\beta})^{\alpha-1} t^{\alpha}}$. Note under ordinary Weibull regression, we have $S_{\mathbf{x}}(t) = e^{-e^{\mathbf{x}'\boldsymbol{\beta}^*}\lambda^* t^{\alpha^*}}$, therefore we can obtain the initial values of (β^*, α^*, λ^*) from a maximum likelihood fit of the Weibull model; e.g. proc lifereg in SAS or survreg in R. Specifically, let (β^*, λ^*, α^*) be estimated under the Weibull regression, then solving $\alpha^* = \alpha$, $\lambda^* = \lambda$, and $\beta^* = \beta(\alpha - 1)$ gives the initial values for the AH model parameters ($\boldsymbol{\beta}$, α, λ) at $J = 1$, the simple parametric Weibull distribution.

For $J > 1$, an iterative procedure is used to obtain starting values for maximization of the posterior conditional on a value of J. We can recursively use the values from the fitted model at J 1 to be the initial values for the model at J using the "nesting" property of Bernstein polynomials, namely

$$b_{J-1,j}(x) = \left[\frac{J-j}{J}\right] b_{Jj}(x) + \left[\frac{j}{J}\right] b_{J,j+1}(x),$$

see Sauer (1999, Proposition 2.3). Some algebra reveals that the initial values for the next J is recursively obtained by

$$\begin{gathered} w_{Jj} = \tfrac{j-1}{J} w_{J-1,j-1} + \tfrac{J-j}{J} w_{J-1,j} \text{ for } j \in \{2, 3, \ldots, J-1\}, \\ w_{J1} = \tfrac{J-1}{J} w_{J-1,1} \text{ and } w_{JJ} = \tfrac{J-1}{J} w_{J-1,J-1}. \end{gathered} \quad (17)$$

The previous posterior modes $\hat{\boldsymbol{\theta}}_{J-1}$ and $\hat{\boldsymbol{\beta}}_{J-1}$ are used as starting values for $\boldsymbol{\theta}_J$ and $\boldsymbol{\beta}_J$. In this manner, the centering distribution $S_{\boldsymbol{\theta}}$, regression parameters $\boldsymbol{\beta}$, and "adjustments" to S_0 encapsulated in $\mathbf{w}_J$ are iteratively refined as J increases. Depending on the level of spatial inhomogeneity in the true S_0, the (conditional on J) posterior means S_0 and $\boldsymbol{\beta}$ "converge" in the sense that they change very little after a certain J. In our experience over several data sets, this is usually a small number, less than $K = 15$. Instead of model averaging, as we pursue in this paper, one could look at the L_1 distance in successive S_0 and stop after $L_1 < \varepsilon$ for some small ε. Sample R code for fitting the model and obtaining common loci of inference such as regression coefficients, surivival and hazard curves are available in the online appendix.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

3.2 Broad assessment of model fit

Cox-Snell residual plots (Nelson, 1972) provide an overall visual assessment of model fit. The *i*th Cox-Snell residual is $r_i = -\log \hat{S}_{\mathbf{x}_i}(t_i)$, the log-estimated survival function for the *i*th subject evaluated at their event time. This residual is a by-product of evaluating the likelihood for all models considered here, and thus provides a means for the relative comparison of model fit. If the model is correct the pairs $\{(r_i, \delta_i)\}_{i=1}^n$ are approximately a censored random sample from an exp(1) distribution, and the estimated integrated hazard plot should be approximately straight with slope one.

A Bayesian approach to fitting via MCMC allows the computation of the log pseudo marginal likelihood (LPML, Geisser and Eddy, 1979), a leave-one-out cross validated measure of a model's ability to predict the observed data. The *i*th conditional predictive ordinate CPO_i is the predictive density ($\delta_i = 1$) or survival function ($\delta_i = 0$) evaluated at t_i, but based on the observed data leaving out $\mathcal{D}_i = (t_i, \delta_i, \mathbf{x}_i)$, denoted $\mathcal{D}_{-i} = \{\mathcal{D}_j : j \neq i\}$: $\text{CPO}_i = f_{\mathbf{x}_i}(t_i|\mathcal{D}_{-i})^{\delta_i} S_{\mathbf{x}_i}(t_i|\mathcal{D}_{-i})^{1-\delta_i}$. The LPML is the log of the product of these, $\text{LPML} = \sum_{i=1}^n \log \text{CPO}_i$. Exponentiated differences in LPML lead to a so called "pseudo Bayes factor" giving evidence in favor of one model over another much like traditional Bayes factors. The LPML is straightforward to compute for censored data using an approach described in Section 10.1 of Chen, Shao, and Ibrahim (2000).

4. Accelerated Hazards with Time-Dependent Covariates

One strength of PH is that the model is easily extended to handle covariates that change in time, such as blood pressure, cholesterol, age, et cetera; the AH model can be similarly generalized. For each individual *i*, assume $\mathbf{x}_i(t) = (x_{i1}(t), \ldots, x_{ip}(t))'$ is a step function that changes at the m_i ordered times $\mathbf{r}_i = (r_{i1}, \ldots r_{im_i})'$

$$\mathbf{x}_i(t) = \sum_{j=1}^{m_i} \mathbf{z}_{ij} I_{[r_j, r_{j+1})}(t),$$

where $r_{i1} = 0$ and $r_{m_i+1} = \infty$ (Hanson et al., 2009). The survival time $t_i \geq r_m$. As in the fixed covariates case, we assume a baseline survival function S_0. This distribution corresponds to the situation where $\mathbf{x}(t) = \mathbf{0}$ for all $t \geq 0$. Under the AH model, the hazard function with time-dependent covariates (TDC) is expressed as

$$h_{\mathbf{x}}(t) = h_0(te^{\mathbf{x}(t)'\beta}) \quad (18)$$

The cumulative hazard function for the *i*th individual under the accelerated hazards model can be expressed as

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$H_i(t_i)=\int_0^{t_i} h_0(se^{\mathbf{x}_i(s)'\beta})ds=\sum_{j=1}^{m_i-1}\int_{r_{ij}}^{r_{i,j+1}} h_0(se^{\mathbf{z}_{ij}'\beta})ds+\int_{r_{m_i}}^{t_i} h_0(se^{\mathbf{z}_{im_i}'\beta})ds.$$

This leads to the survival function

$$S_i(t_i)=\left[\frac{S_0(e^{\mathbf{z}_{im_i}'\beta}t_i)}{S_0(e^{\mathbf{z}_{im_i}'\beta}r_{m_i})}\right]^{e^{-\mathbf{z}_{im_i}'\beta}}\prod_{j=1}^{m_i-1}\left[\frac{S_0(e^{\mathbf{z}_{ij}'\beta}r_{i,j+1})}{S_0(e^{\mathbf{z}_{ij}'\beta}r_{ij})}\right]^{e^{-\mathbf{z}_{ij}'\beta}} \quad (19)$$

and density defined through (19)

$$f_i(t_i)=S_i(t_i)\frac{f_0(e^{\mathbf{z}_{im_i}'\beta}t_i)}{S_0(e^{\mathbf{z}_{im_i}'\beta}t_i)}. \quad (20)$$

Baseline survival $S_0(\cdot)$ corresponds to an individual with zero covariates for all time; $\mathbf{x}(t) \equiv 0$; $S_0(\cdot)$ is assigned a TBPP as in the fixed-covariates case. The likelihood is given by

$$\mathcal{L}(\psi_J)=\prod_{i=1}^{n} S_i(t_i)^{1-\delta_i} f_i(t_i)^{\delta_i} \quad (21)$$

where $S_i(t_i)$ and $f_i(t_i)$ defined in (19) and (20) respectively, and as usual S_θ is a specified parametric family.

5. Simulation Study

The performance of the TBPP for survival analysis under the AH model is studied in this section. The Weibull family, $S_\theta(t) = e^{-\theta_2 t^{\theta_1}}$, centers S_0; the priors are $p(\theta, \beta) \propto 1$ independent of $\mathbf{w}_J|J \sim \text{Dirichlet}(\mathbf{2}_J)$. Two sample sizes are considered, $n = 300$ and $n = 500$. The degree J is capped at $K = 15$ with the prior $p(J) \propto 0.8^{J-1}$, which dies down to $p(15) \approx 0.01$. The true baseline survival functions S_0 encompass several situations: (i) the baseline survival function defined in (6), where $S_\theta(t) = e^{-e^{-2}t^3}$ and $J = 6$ components with the weights $\mathbf{w}_6 = (0.3, 0.15, 0.05, 0.05, 0.15, 0.3)$ (this baseline density is bimodal); (ii) a Weibull baseline survival function $S_0 = Weibull(0.5, 0.5)$; (iii) the log-logistic baseline survival function $LLogis(2, 1)$, where the first parameter is the shape parameter and the second is the scale parameter; and (iv) the log-normal baseline survival function $LN(0, 1)$. Two covariates $\mathbf{x} = (x_1, x_2)$ are considered with x_1 generated from standard normal distribution and x_2 from a Bernoulli distribution with probability 0.5. The true coefficients are assigned as $\beta = (1, -1)$. Censoring rates were fixed at 15% and 30% by simulating $c_i \sim U(0, a)$ for different values of a. For each setting of baseline S_0 and sample size n, 200 simulated data sets were generated and the posterior mean $E(\beta|\mathcal{D})$ for each data set obtained as described in Section 3. Bias, mean square error (MSE), standard deviation (StDev), and coverage probability are given in Table 1, where coverage probability is the proportion of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

95% credible intervals that contain the true parameter value out of the 200 simulated data sets.

For comparison, Table 1 also lists results from the profile likelihood method of Zhang et al. (2011) and the Gehan rank-type method of Chen and Wang (2000). For $n = 300$ the profile likelihood method failed to converge for several simulated data sets using either a log-logistic or a log-normal baseline distribution, hence the "NaN" values in the table. With either increasing sample size or smaller censoring rates, all estimation methods are more precise as more information is included. In terms of the coverage probability, our proposed method is considerably more reliable compared to the other two existing methods for the simulation cases we looked at. Across all simulation cases, our method gives mean squared errors essentially as good as or better than the two other approaches simply fixing the upper bound $J \leq K = 15$. Although we fixed $K = 15$, in general we recommend increasing K with the sample size to achieve better resolution.

We also compared the normal approximation to MCMC sampling from this simulation scenario for $n = 50$, 100, 200, and 500. Accurate posterior modes are difficult to obtain from MCMC output, so posterior means and medians tend to be used as they are stably estimated. Thus we compared posterior modes from the normal approximation to posterior means from the MCMC, as well as posterior standard deviations. In terms of estimating the regression coefficients $\boldsymbol{\beta}$, the two approaches provide regression estimates within 5% of each other at $n = 50$; this drops to 1% at $n = 500$. However, the normal approximation provides larger standard deviation estimates at all four sample sizes, as much as 50% greater, indicating that the true posterior density is perhaps lighter tailed than multivariate normal; this is plausible as the Weibull distribution has lighter tails than the normal, and the TBPP inherits the tail behavior of the centering family.

The estimated centering parameters $\boldsymbol{\theta}$ and weights $\mathbf{w}_{10}$ were quite different at smaller sample sizes, but do converge, although more slowly than for $\boldsymbol{\beta}$, as the sample size is increased. Posterior densities for the elements of $\mathbf{w}_{10}$ all show some degree of left skew, even at higher sample sizes. Although $\boldsymbol{\beta}$ and $\mathbf{w}_{10}$ can differ somewhat, the posterior mean baseline survival densities $f_0(\cdot)$ are quite similar across the sample sizes. This would seem to indicate, much like mixtures of Polya trees, there is "weak identifiability" in the sense that similar baseline survival shapes can be obtained from different settings of $\boldsymbol{\theta}$ and $\mathbf{w}_J$.

6. Real Data Illustrations

6.1 Brain Tumor Study

Brem et al. (1995) conducted a randomized, placebo-controlled clinical trial to evaluate the effectiveness of biodegradable carmustine (BCNU) polymers on recurrent brain malignant gliomas. In total, 110 of 222 participants were randomized to the BCNU polymer treatment group, and 112 were randomized to the placebo polymer control group. After the tumor was removed, a BCNU or placebo polymer was placed into the cavity. The BCNU polymer gradually releases BCNU for three weeks following placement; due to the gradual effect of treatment, the AH model is a plausible fit to these data.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Chen and Wang (2000) considered only the treatment effect in the first 52 weeks when BCNU polymer was applied; survival times larger than 52 weeks are right-censored. We fit the TBPP AH model to these data with $M = 1$ and $K = 15$ as well as the TBPP PH model; integrated Cox-Snell residual plots show marked lack of fit for both models (not shown). Following Chen (2001) and Zhang et al. (2011), we thus include covariates age and resect75, which indicates ≥ 75% tumor resection. The estimated $\boldsymbol{\beta}$ and corresponding 95% credible intervals (CI) are in Table 2 as profile likelihood results (Zhang et al., 2011), the Gehan rank-type method (Chen and Wang, 2000), and the PH model fit through partial likelihood. Estimated survival and hazard functions for the treatment and control groups are plotted in Figure 1 for age = 48 and resect75 = 1.

Overall, the treatment group exhibits lower hazard risk than the control group. However, since our estimated 95% CI for β_1 includes 0, we do not have a significant difference between two groups, in contrast to the approaches of Chen and Wang (2000) and Zhang et al. (2011). We also fit a Bayesian TBPP AH model fixing $M = 1$ and $J = 15$ via MCMC, as well as a Bayesian TBPP PH model, and the AH model of Zhang et al. (2011), and obtained the integrated Cox-Snell residual plots for all three in Figure 2. All three models show no gross lack of fit, although the PH model shows some deviations from the line $y = x$ in the tail. The LPML is −783 for the TBPP AH model and −790 for the TBPP PH model, giving a pseudo Bayes factor of about $e^7 \approx 1000$ in favor of the AH model. Although none of the models show any *gross* lack of fit, the AH model provides significantly better out of sample prediction than the PH model for these data. The estimated value of $e^{\hat{\beta}_1} \approx 0.7$ for the AH treatment coefficient implies that the risk of individuals in the treatment group at time t is equal to the risk of those in the control group at time $0.7t$; note that the baseline hazard is increasing. Put another way, since $e^{-\hat{\beta}_1} \approx 1.4$, the treatment group will take 40% more time to achieve the same (arbitrary) elevated risk as the control group.

6.2 Stanford Heart Transplant Study

Crowley and Hu (1977) first analyzed the well-known Stanford heart transplant data using the Cox model with TDC. Recently, Hanson et al. (2009) considered Bayesian semiparametric models for analyzing the data using the Cox model with TDC and two generalizations of the AFT model; they called those three models CTD (Cox, 1972), PKTD (Prentice and Kalbfleisch, 1979), and COTD (Cox and Oakes, 1984) respectively. Each of the three models was fit with a mixture of Polya trees (MPT) prior assigned to the baseline survival function S_0. Here we will fit this data to the AH model with TDC using the TBPP assigned to the baseline survival function. To compare with their fitting results, we consider the same covariates: transplant indicator, patient baseline age, and mismatch scores. Those patients who entered in the transplant program but did not receive a heart transplant before they died serve as a control group; they have $\mathbf{x}_i(t) = (x_{i1}(t), x_{i2}(t), x_{i3}(t)) = (0, 0, 0)$ for all $t > 0$. For patients receiving a heart transplant, let s_i when the transplant occurred. These individuals have

$$\mathbf{x}_i(t) = \begin{Bmatrix} (0,0,0) & t < s_i \\ (z_{i1}, z_{i2}, z_{i3}) & t \geq s_i \end{Bmatrix},$$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

where $z_{i1} = 1$, z_{i2} = age − 35, and z_{i3} = mismatch − 0.5. Since Hanson et al. (2009) centered S_0 at the log-logistic family, we follow suit, specifically, $S_{\theta}(t)=(1+(\frac{t}{\theta_1})^{\theta_2})^{-1}$.

Instead of normal approximations, MCMC is instead used here; MCMC allows us to easily compute the LPML to compare to other approaches. The random quantities based on (21) are θ, $\mathbf{w}_J$, and β. Again we assume (θ, β) has an improper flat prior and $\mathbf{w}_J \sim$ Dirichlet($\mathbf{1}_J$). To relax boundary restrictions on $\mathbf{w}_J$, we take the transformation (15) and fit to $\mathbf{v}_{J-1}$. A block-updated version of the adaptive random walk Metropolis-Hastings algorithm of Haario et al. (2005) is applied to (θ, $\beta|\mathbf{v}_{J-1}$, $\mathcal{D}$) and ($\mathbf{v}_{J-1}|\theta, \beta, \mathcal{D}$) in turn.

We fit two models with $J = 1$ (parametric log-logistic model) and $J = 15$; we call them AHTD and AHTD-TBPP respectively. The results are displayed in Table 3.

Based on the LPML, the two models predict similarly; in fact, the parametric model with a log-logistic baseline fits slightly better than the semiparametric model. Both models imply a significant difference in survival between patients with and without heart transplants. For comparison purposes, consider two individuals who have the same covariates (age 35 and mis-match score 0.5, so $\mathbf{z} = (1, 0, 0)$), except for one does not receive a heart transplant and the other has one at $s = 6$ months. Under this setup, e^{β_1} reflects how the hazard of death is re-scaled when the patient receives a heart transplant. Based on the AHTD, a 95% credible interval for e^{β_1} is (1.99, 25.53) with posterior median of 6.96. Although the hazard for death is decreasing for both groups (see Figure 6 in the online appendix), a given risk of dying takes about 7 times longer to achieve for those not receiving the heart transplant. Hanson et al. (2009) found LPML values of −468.0, −467.0, and −464.1 for the PKTD, COTD, and Cox models respectively; the AHTD model fares better than all of these.

7. Discussion

The approach in this paper enriches a given class of densities by adding detail to an overall parametric shape; we apply the method to the underused AH model. Specifically, we advocate estimating a transformation of a Bernstein polynomial by centering S_0 at the parametric distribution G_{θ}, posterior inference averages over θ. Our approach uses standard maximization routines an adaptive MCMC algorithm to obtain inference. The AH model is further generalized to accommodate time-dependent covariates.

Although the AH model is tailored for use with covariates that take effect gradually, it may not be appropriate for factors such as cancer stage that have an immediate impact on the hazard; here the PH model could provide superior fit. Etezadi-Amoli and Ciampi (1987), and later Chen and Jewell (2001) consider a natural extension of AH, PH, and AFT models $h_{\mathbf{x}}(t) = e^{\mathbf{x}'\beta} h_0\{e^{\mathbf{x}'\gamma}t\}$. Under this model, setting elements of β or γ to zero allows a covariate to have a conditionally AH or PH effect on the hazard, respectively; setting the same element of β and γ to be equal implies an AFT effect. We have begun exploring approaches to fitting this general model and implementing Bayes factors for testing the types of effect each covariate might have on the hazard, i.e. AH, PH, AFT, or extended. The more general model could, for example, accommodate an AH effect for treatment but a PH effect for cancer stage.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work is supported by NCI grant R03CA165110 and the University of South Carolina. The authors thank the editor, associate editor, and referee for suggestions that greatly improved the readability of the paper.

References

Aitchison J, Shen S. Logistic-normal distributions: Some properties and uses. Biometrika. 1980; 67:261–272.

Bedrick EJ, Christensen R, Johnson WO. Bayesian accelerated failure time analysis with application to veterinary epidemiology. Statistics in Medicine. 2000; 19:221–237. [PubMed: 10641026]

Bernstein S. Démonstration du Théoreme de Weierstrass fondée sur le calcul des Probabilités. Communications of the Kharkov Mathematical Society. 1912; 13:1–2.

Brem H, Piantadosi S, Burger P, Walker M, Selker R, Vick N, Black K, Sisti M, Brem S, Mohr G, Muller P, Morawetz R, Schold S. Placebo-controlled trials of safety and efficacy of intraoperative controlled delivery by biodegradable polymers of chemotherapy of recurrent gliomas. The polymerbrain tumor treatment group. The Lancet. 1995; 345:1008–1012.

Chang I, Hsiung C, Wu Y, Yang C. Bayesian survival analysis using Bernstein polynomials. Scandinavian Journal of Statistics. 2005; 32:447–466.

Chen, M-H.; Shao, Q-M.; Ibrahim, JG. Monte Carlo Methods in Bayesian Computation. Springer-Verlag; New York: 2000.

Chen Y. Accelerated hazards regression model and its adequacy for censored survival data. Biometrics. 2001; 57:853–860. [PubMed: 11550937]

Chen Y, Jewell N. On a general class of semiparametric hazards regression models. Biometrika. 2001; 88:687–702.

Chen Y, Wang M. Analysis of accelerated hazards models. Journal of the American Statistical Association. 2000; 95:608–618.

Cox D. Regression models and life tables (with discussion). Journal of the Royal Statistical Society, Series B. 1972; 34:187–208.

Cox, D.; Oakes, D. Analysis of survival data. London: Chapman and Hall; 1984.

Crowley J, Hu M. Covariance analysis of heart transplant data. Journal of the American Statistical Society. 1977; 77:27–36.

Etezadi-Amoli J, Ciampi A. Extended hazard regression for censored surival data with covariates: A spline approximation for the baseline hazard function. Biometrics. 1987; 43:181–192.

Geisser S, Eddy W. A predictive approach to model selection. Journal of the American Statistical Association. 1979; 74:153–160.

Green P. Reversible jump MCMC computation and Bayesian model determination. Biometrika. 1995; 82:711–732.

Haario H, Saksman E, Tamminen J. Componentwise adaptation for high dimensional MCMC. Computational Statistics. 2005; 20:265–273.

Hanson T, Johnson W, Laud P. Semiparametric inference for survival models with step process covariates. The Canadian Journal of Statistics. 2009; 37:60–79.

Ibrahim, JG.; Chen, M-H.; Sinha, D. Bayesian Survival Analysis. New York: Springer; 2001.

Kalbfleisch JD. Nonparametric Bayesian analysis of survival time data. Journal of the Royal Statistical Society, Series B. 1978; 40:214–221.

Lavine M. Some aspects of Polya tree distributions for statistical modeling. The Annals of Statistics. 1992; 20:1222–1235.

Mallick B, Walker S. A Bayesian semiparametric transformation model incorporating frailties. Journal of Statistical Planning and Inference. 2003; 112:159–174.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Nelson W. Theory and applications of hazard plotting for censored failure data. Technometrics. 1972; 14:945–965.

Petrone S. Random Bernstein polynomials. Scandinavian Journal of Statistics. 1999; 26:373–393.

Petrone S. Bayesian density estimation using Bernstein polynomials. The Canadian Journal of Statistics. 1999; 27:105–126.

Petrone S, Wasserman L. Consistency of Bernstein polynomial posteriors. Journal of the Royal Statistical Society, Series B. 2002; 64:79–100.

Prentice RL, Kalbfleisch JD. Hazard Rate Models with Covariates. Biometrics. 1979; 35:25–39. [PubMed: 497336]

Sauer, T. Multivariate Bernstein polynomials, convexity and related shape properties. In: Pena, JM., editor. Shape Preserving Representations in Computer Aided Design. Nova Science Publishers; 1999.

Zhang J, Peng Y, Zhao O. A new semiparametric estimation method for accelerated hazards model. Biometrics. 2011; 67:1352–1360. [PubMed: 21457194]

Zucker D, Lakatos E. Weighted log rank type statistics for comparing survival curves when there is a time lag in the effectiveness of treatment. Biometrika. 1990; 77:853–864.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

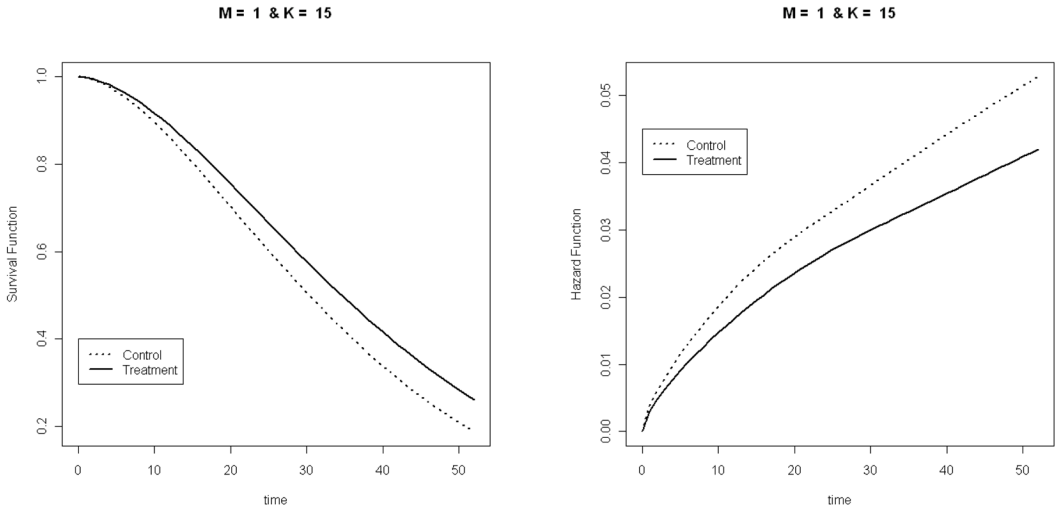

Figure 1.
Estimated survival and hazard functions for $M = 1$ and $K = 15$; age is set to 48 years and Resect75 is one.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

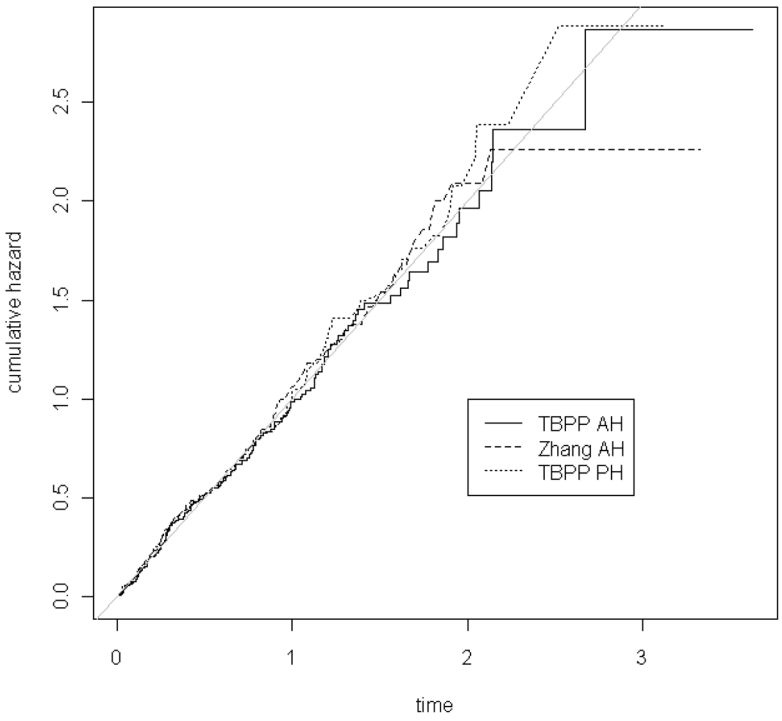

Figure 2.
Integrated Cox-Snell residual plots for TBPP AH, TBPP PH, and the AH model of Zhang et al. (2011).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Biometrics. Author manuscript; available in PMC 2015 May 14.

Table 1

Estimated bias, mean square error, standard deviation, and coverage probability (CP) from the proposed method with different censoring rates (CR) and sample sizes.

***n* = 300**													
		TBPP				**Profile**				**Gehan Rank-type**			
CR	$\hat{\boldsymbol{\beta}}$	**Bias**	**MSE**	**StDev**	**CP**	**Bias**	**MSE**	**StDev**	**CP**	**Bias**	**MSE**	**StDev**	**CP**
(i) Mixed Weibull baseline survival function													
15%	$\hat{\beta}_1$	0.001	0.002	0.041	0.81	0.166	0.036	0.061	0.13	0.002	0.004	0.045	0.74
	$\hat{\beta}_2$	0.005	0.004	0.065	0.84	−0.150	0.034	0.100	0.67	0.003	0.010	0.083	0.78
30%	$\hat{\beta}_1$	0.010	0.003	0.056	0.95	0.217	0.054	0.085	0.26	0.009	0.008	0.058	0.78
	$\hat{\beta}_2$	−0.013	0.008	0.086	0.96	−0.218	0.063	0.131	0.67	−0.0100	0.019	0.102	0.85
(ii) Weibull baseline survival function													
15%	$\hat{\beta}_1$	0.010	0.027	0.161	0.95	−0.084	0.030	0.147	0.89	0.003	0.045	0.213	0.94
	$\hat{\beta}_2$	0.011	0.081	0.271	0.95	0.100	0.080	0.252	0.91	0.004	0.133	0.305	0.90
30%	$\hat{\beta}_1$	0.049	0.038	0.180	0.96	−0.012	0.034	0.169	0.94	0.074	0.077	0.243	0.94
	$\hat{\beta}_2$	−0.051	0.094	0.305	0.97	−0.001	0.091	0.286	0.95	−0.052	0.159	0.325	0.87
(iii) Log-logistic baseline survival function													
15%	$\hat{\beta}_1$	0.047	0.020	0.145	0.98	1.034	188.0	NaN	NaN	−0.011	0.018	0.096	0.81
	$\hat{\beta}_2$	−0.096	0.065	0.247	0.98	−0.435	22.5	0.240	0.86	−0.027	0.066	0.208	0.88
30%	$\hat{\beta}_1$	0.033	0.041	0.175	0.96	0.120	0.069	NaN	NaN	−0.011	0.030	0.106	0.78
	$\hat{\beta}_2$	−0.033	0.080	0.280	0.97	−0.121	0.146	NaN	NaN	0.035	0.083	0.221	0.85
(iv) Log-normal baseline survival function													
15%	$\hat{\beta}_1$	0.037	0.033	0.171	0.96	0.064	0.046	0.132	0.86	−0.038	0.068	0.122	0.87
	$\hat{\beta}_2$	−0.044	0.107	0.296	0.95	−0.057	0.1477	0.248	0.85	−0.016	0.180	0.267	0.84
30%	$\hat{\beta}_1$	0.036	0.032	0.190	0.98	0.369	15.2	NaN	NaN	−0.017	0.023	0.115	0.86

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

n = 300

		TBPP				Profile				Gehan Rank-type			
CR	$\hat{\beta}$	**Bias**	**MSE**	**StDev**	**CP**	**Bias**	**MSE**	**StDev**	**CP**	**Bias**	**MSE**	**StDev**	**CP**
	$\hat{\beta}_2$	−0.027	0.072	0.310	0.98	0.270	21.6	NaN	NaN	−0.004	0.077	0.255	0.92

n = 500

		TBPP				Profile				Gehan Rank-type			
CR	$\hat{\beta}$	**Bias**	**MSE**	**StDev**	**CP**	**Bias**	**MSE**	**StDev**	**CP**	**Bias**	**MSE**	**StDev**	**CP**
(i) Mixed Weibull baseline survival function													
15%	$\hat{\beta}_1$	0.003	0.001	0.036	0.96	0.150	0.025	0.050	0.11	0.005	0.003	0.040	0.83
	$\hat{\beta}_2$	−0.003	0.004	0.057	0.92	−0.143	0.026	0.082	0.59	−0.002	0.008	0.074	0.89
30%	$\hat{\beta}_1$	0.010	0.002	0.042	0.94	0.164	0.031	0.058	0.18	0.003	0.004	0.043	0.78
	$\hat{\beta}_2$	−0.010	0.004	0.064	0.95	−0.162	0.034	0.091	0.59	−0.015	0.012	0.078	0.87
(ii) Weibull baseline survival function													
15%	$\hat{\beta}_1$	0.017	0.014	0.125	0.98	−0.047	0.015	0.118	0.92	0.020	0.025	0.157	0.95
	$\hat{\beta}_2$	0.004	0.041	0.210	0.97	0.067	0.040	0.199	0.96	−0.008	0.074	0.237	0.90
30%	$\hat{\beta}_1$	0.026	0.019	0.135	0.97	−0.005	0.018	0.130	0.96	0.025	0.029	0.166	0.94
	$\hat{\beta}_2$	−0.007	0.055	0.231	0.94	0.032	0.053	0.220	0.95	−0.019	0.092	0.250	0.90
(iii) Log-logistic baseline survival function													
15%	$\hat{\beta}_1$	0.004	0.037	0.111	0.93	0.040	0.040	0.097	0.86	−0.024	0.077	0.080	0.84
	$\hat{\beta}_2$	−0.066	0.035	0.190	0.96	−0.060	0.052	0.171	0.90	0.014	0.034	0.163	0.91
30%	$\hat{\beta}_1$	0.015	0.015	0.122	0.95	0.070	0.026	0.110	0.84	0.001	0.015	0.087	0.83
	$\hat{\beta}_2$	−0.037	0.037	0.200	0.95	−0.064	0.050	0.187	0.86	−0.002	0.040	0.170	0.89
(iv) Log-normal baseline survival function													
15%	$\hat{\beta}_1$	0.036	0.017	0.125	0.97	0.047	0.024	0.102	0.86	−0.000	0.012	0.090	0.89
	$\hat{\beta}_2$	−0.047	0.044	0.217	0.96	−0.043	0.061	0.186	0.88	−0.021	0.044	0.208	0.95

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

***n* = 500**

		TBPP				Profile				Gehan Rank-type			
CR	$\hat{\boldsymbol{\beta}}$	**Bias**	**MSE**	**StDev**	**CP**	**Bias**	**MSE**	**StDev**	**CP**	**Bias**	**MSE**	**StDev**	**CP**
30%	$\hat{\beta_1}$	0.027	0.022	0.135	0.95	0.064	0.030	0.113	0.82	−0.002	0.018	0.089	0.82
	$\hat{\beta_2}$	−0.046	0.055	0.233	0.95	−0.064	0.077	0.204	0.85	−0.026	0.055	0.204	0.90

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Estimated β and 95% CI for brain tumor data; * significance at the 5% level.

	TBPP	Zhang, Peng, and Zhao	Chen and Wang	PH
Treatment	−0.341 (0.238)	−0.648* (0.134)	−0.672* (0.272)	−0.232 (0.151)
Age	0.040* (0.010)	0.053* (0.008)	0.011* (0.004)	0.025* (0.006)
Resect75	−1.078* (0.281)	−0.967* (0.223)	−0.368 (0.357)	−0.6502* (0.166)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Inference for Stanford heart transplant data

Models	AHTD	AHTD-TBPP
LPML	−463.0	−463.9
Status	1.94 (0.69,3.24)	1.74 (0.39,3.18)
Age-35	−0.100 (−0.173, −0.032)	−0.090(−0.166, −0.029)
Mismatch-0.5	−0.91 (−1.93,0.04)	−0.86(−1.88,0.06)