2015

Bayesian Semi- and Non-parametric Analysis for Spatially Correlated Survival Data

Haiming Zhou
University of South Carolina

Bayesian Semi- and Non-parametric Analysis for Spatially Correlated Survival Data

by

Haiming Zhou

Bachelor of Science
University of Science and Technology of China 2009

Master of Science
Clemson University 2011

Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in

Statistics

College of Arts and Sciences

University of South Carolina

2015

Accepted by:

Timothy Hanson, Major Professor

Xianzheng Huang, Committee Member

Lianming Wang, Committee Member

Jiajia Zhang, Committee Member

Lacy Ford, Senior Vice Provost and Dean of Graduate Studies

Acknowledgments

First of all, I would like to express my heartfelt gratitude to my advisor, Dr. Timothy Hanson, for his constant guidance and enthusiastic encouragement during my PhD studies, without which this dissertation work would not have been possible. For his unwavering support, I am truly grateful. He is the best advisor in the world; I feel so warm and lucky having him as my advisor. I would also like to express my sincere appreciation to my committee, Dr. Xianzheng Huang, Dr. Lianming Wang and Dr. Jiajia Zhang, for their support towards the successful completion of my studies at USC. My grateful thanks are also extended to all professors in the Department of Statistics. I learned a lot from their classes. Finally, I would like to thank my friends for their accompany and encouragement which made my stay and studies at USC more enjoyable.

Abstract

Flexible incorporation of both geographical patterning and risk effects in cancer survival models is becoming increasingly important, due in part to the recent availability of large cancer registries. The analysis of spatial survival data is challenged by the presence of spatial dependence and censoring for survival times. Accurately modeling the risk factors and geographical pattern that explain the differences in survival is particularly of interest. Within this dissertation, the first chapter reviews commonly-used baseline priors, semiparametric and nonparametric Bayesian survival models and recent approaches for accommodating spatial dependence, both conditional and marginal. The last three chapters contribute three flexible survival models: (1) a proportional hazards model with areal-level covariate-adjusted frailties with application to county-level breast cancer survival data, (2) a marginal Bayesian nonparametric model for time to disease arrival of threatened amphibian populations, and (3) a generalized accelerated failure time model with spatial intrinsic conditionally autoregressive frailties with application to county-level prostate cancer data. An R package `spBayesSurv` is developed to examine all the proposed models along with some traditional spatial survival models.

Table of Contents

List of Tables

List of Figures

Introduction

Methodology for modeling spatially correlated survival data has attracted increasing attention across diverse studies due to the important role geographical information can play in predicting survival (i.e. failure time or time-to-event). There are two general types of spatial survival data: (a) point-referenced (or geostatistical), where survival times along with covariates are measured with exact locations that vary continuously over a study region; and (b) areal-referenced, where the study region is partitioned into a finite number of areal units with well-defined boundaries. For example, one of our motivating data sets aims at quantifying the effects of fungus spread on the time-to-disease distribution of threatened frog populations in Sequoia Kings Canyon National Park, where exact latitude and longitude are available for each frog population. But sometimes locational information may be only available at the areal level. Cancer data from the Surveillance Epidemiology and End Results (SEER) program of the National Cancer Institute involve the analysis of areal type survival data, since the location for each patient is only recorded at the county level.

This dissertation is organized as follows. Chapter 1 reviews commonly-used baseline priors, semiparametric and nonparametric Bayesian survival models and recent approaches for accommodating spatial dependence, both conditional and marginal. Chapter 2 presents a covariate-adjusted frailty proportional hazards model for the analysis of clustered right-censored data. Chapter 3 develops a marginal Bayesian nonparametric survival model for point-referenced right-censored data. Chapter 4 proposes a generalized accelerated failure time model with intrinsic conditionally autoregressive frailties for arbitrarily censored areal-referenced data.

Chapter 1

Background and Dissertation Overview[1]

Survival analysis has received a great deal of attention as a subfield of Bayesian nonparametrics over the last 50 years. In particular, the fitting of survival models that allow for sophisticated correlation structures has become common due to computational advances in the 1990s, in particular Markov chain Monte Carlo techniques. Very large, complex spatial datasets can now be analyzed accurately including the quantification of spatiotemporal trends and risk factors. This chapter reviews four nonparametric priors on baseline survival distributions in common use, followed by a catalogue of semiparametric and nonparametric models for survival data. Generalizations of these models allowing for spatial dependence are then discussed and broadly illustrated. Throughout, practical implementation through existing software is emphasized.

1.1 Introduction

This chapter reviews several semiparametric Bayesian survival models, and summarizes some recent proposals to allow for spatial and covariate-adjusted dependence among the survival times. Two generalizations of the accelerated failure time model that allow crossing cumulative hazards for different covariate combinations, and hence crossing survival curves, are also discussed.

[1]Zhou, H. and Hanson, T. (2015). Bayesian spatial survival models. In *Nonparametric Bayesian Methods in Biostatistics and Bioinformatics (Frontiers in Probability and the Statistical Sciences)*, to appear. P. Müller and R. Mitra, editors. Springer. Reprinted here with permission of Springer.

Four prior specifications in broad use are first reviewed in Section 1.2. A catalogue of Bayesian survival models is presented in Section 1.3. Section 1.4 discusses the incorporation of dependence among survival times across the models in Section 1.3, focusing mostly on spatial dependence followed by several real-data illustrations in Section 1.5. The chapter concludes with a short discussion in Section 1.6. Please note at the outset that, although a review is attempted, the cited papers and approaches are biased toward what the authors are aware of and have found useful.

1.2 A selection of nonparametric priors

A common starting point in the specification of a regression model for time-to-event data is the definition of a baseline survival function, S_0, that is modified (either directly or indirectly) by subject-specific covariates $\mathbf{x}$. Let T_0 be a random survival time from the baseline group (with all covariates equal to zero). The baseline survival function is defined by $S_0(t) = P(T_0 > t) = \exp\{-H_0(t)\}$ where $H_0(t)$ is the baseline cumulative hazard. For continuous outcomes, the baseline density and hazard functions are $f_0(t) = -\frac{d}{dt}S_0(t)$ and $h_0(t) = f_0(t)/S_0(t) = \frac{d}{dt}H_0(t)$, respectively. The cumulative distribution, survival, density and hazard functions for a member of the population with covariates $\mathbf{x}$ will be denoted by $F_{\mathbf{x}}(t)$, $S_{\mathbf{x}}(t)$, $f_{\mathbf{x}}(t)$, and $h_{\mathbf{x}}(t)$, respectively.

A wide variety of priors have been used in Bayesian survival analysis over the last 40 years. We focus on four of these: the gamma process, B-splines, Dirichlet process mixtures, and mixtures of Polya trees. Additional reviews can be found in Sinha and Dey (1997), Ibrahim et al. (2001), Müller and Quintana (2004), Hanson et al. (2005), Nieto-Barajas (2013), and Müller et al. (2015).

Gamma process

Kalbfleisch (1978) proposed the gamma process (GP) to model the cumulative hazard

function H_0 in the context of the proportional hazards (PH) model (Cox, 1972). Let $H_{\boldsymbol{\theta}}(t)$ be an increasing, left-continuous function on $[0, \infty)$ indexed by $\boldsymbol{\theta}$, where $H_{\boldsymbol{\theta}}(0) = 0$; typically $H_{\boldsymbol{\theta}}$ is parametric. Let $H_0(\cdot)$ be a stochastic process such that (i) $H_0(0) = 0$, (ii) $H_0(\cdot)$ has independent increments in disjoint intervals, and (iii) $H_0(t_2) - H_0(t_1) \sim \Gamma\{\alpha(H_{\boldsymbol{\theta}}(t_2) - H_{\boldsymbol{\theta}}(t_1)), \alpha\}$ for $t_2 > t_1$, where $\Gamma(\alpha, \beta)$ implies mean α/β. Then $\{H_0(t) : t \geq 0\}$ is said to be a GP with parameter $(\alpha, H_{\boldsymbol{\theta}})$ and denoted $H_0 \sim GP(\alpha, H_{\boldsymbol{\theta}})$.

Note that $\mathrm{E}\{H_0(t)\} = H_{\boldsymbol{\theta}}(t)$ so that H_0 is centered at $H_{\boldsymbol{\theta}}$. Also, $\mathrm{Var}\{H_0(t)\} = H_{\boldsymbol{\theta}}(t)/\alpha$ so that, similar to the Dirichlet process and Polya trees described below, the precision parameter α controls how "close" H_0 is to $H_{\boldsymbol{\theta}}$ and provides a prior measure of how certain one is that H_0 is near $H_{\boldsymbol{\theta}}$. Ferguson (1973) recast the Dirichlet process (DP) as a scaled GP.

The posterior of the GP is characterized by Kalbfleisch (1978); his results for the PH model simplify when no covariates are specified. With probability one, the GP is a monotone nondecreasing step function, implying that the corresponding survival function S_0 is a nonincreasing step function. Similar to the DP, matters are complicated by the presence of ties in the data with positive probability. When present in the observed data, such ties make the resulting computations awkward. Clayton (1991) described a Gibbs sampler for obtaining inferences in the PH model with a GP baseline.

Burridge (1981) and Ibrahim et al. (2001) suggest that the model as proposed by Kalbfleisch (1978) and extended by Clayton (1991) is best suited to grouped survival data. Walker and Mallick (1997) considered an approximation to the GP for continuous data. Define a partition of $(0, \infty)$ by $\{(a_{j-1}, a_j]\}_{j=1}^{J} \cup (a_J, \infty)$ where $0 = a_0 < a_1 < a_2 < \cdots < a_{J+1} = \infty$. Here, a_J is taken to be equal to be largest event time recorded. If $H_0 \sim GP(\alpha, H_{\boldsymbol{\theta}})$, then by definition $h_{0j} = H_0(a_j) - H_0(a_{j-1}) \overset{ind.}{\sim} \Gamma\{\alpha(H_{\boldsymbol{\theta}}(a_j) - H_{\boldsymbol{\theta}}(a_{j-1})), \alpha\}$. Walker and Mallick (1997) make this assumption for

the given partition and further assume that $h_0(t)$ is constant and equal to h_{0j} for $t \in (a_{j-1}, a_j]$, $j = 1, \ldots, J$, yielding a particular piecewise exponential model. So the piecewise exponential model, which has a long and fruitful history in both Bayesian and frequentist survival analysis, can be viewed as an approximation to the GP when gamma increments are used.

B-splines and Bernstein polynomials

A flexible and popular basis expansion approach to modeling functions over a finite interval $[a, b]$ is based on B-splines (de Boor, 2001). A B-spline is a piecewise-differentiable polynomial of a given degree d; $d = 2$ and $d = 3$ give quadratic and cubic B-splines, respectively. The B-spline is defined over the union of intervals with endpoints termed knots. The overall polynomial is continuous ($d \geq 1$) or differentiable ($d \geq 2$) over the range of the knots. Knots can be equispaced yielding a cardinal B-spline or else irregularly spaced. Computation is especially easy for equispaced knots and so we focus on that here; generalizations can be found in Kneib (2006). The B-spline includes polynomials of the same or lower degree as special cases; e.g. a quadratic B-spline includes all constant, linear, and parabolic functions over $[a, b]$.

For degree $d = 2$, the quadratic B-spline "mother" basis function is defined on $[0, 3]$

$$\varphi(x) = \left\{ \begin{array}{ll} 0.5x^2 & 0 \leq x \leq 1 \\ 0.75 - (x - 1.5)^2 & 1 \leq x \leq 2 \\ 0.5(3 - x)^2 & 2 \leq x \leq 3 \\ 0 & \text{otherwise} \end{array} \right\}.$$

Say the number of basis functions is J. The B-spline basis functions are shifted, rescaled versions of φ. Let $x_1, \ldots, x_n$ be event times of interest and $x_{(1)}, \ldots, x_{(n)}$ their order statistics. The j-th basis function is $B_j(x) = \varphi\left(\frac{x - x_{(1)}}{\Delta} + 3 - j\right)$, where $\Delta = \frac{x_{(n)} - x_{(1)}}{J-2}$. A B-spline is typically used with a rather large number of basis

functions J, e.g. 20–40. The B-spline model for an unknown function is

$$g(x) = \sum_{j=1}^{J} \theta_j B_j(x). \tag{1.1}$$

A global level of smoothness can be incorporated into a B-spline model by encouraging neighboring coefficients to be similar; the more regular the coefficients are, the less wiggly g is. The hazard can be modeled directly as $h_0(t) = g(t)$ with the constraint $\theta_j \geq 0$ (Wang and Dunson, 2011; Pan et al., 2014; Lin et al., 2015; Li et al., 2015b); typically $\theta_1, \ldots, \theta_J$ have exponential or gamma priors. Komárek and Lesaffre (2008) consider a limiting case of the B-spline order as a model for densities and model the $\theta_j \geq 0$ via a generalized logit transformation so that $\sum_{j=1}^{J} \theta_j = 1$.

Alternatively, to avoid the positivity constraints on θ_i, one can model $h_0(t) = \exp\{g(t)\}$ (Hennerfeind et al., 2006; Kneib and Fahrmeir, 2007) with $\theta_j \in \mathbb{R}$. Classical spline estimation on $\{(x_i, y_i)\}_{i=1}^{n}$ proceeds by minimizing $\sum_{i=1}^{n}(y_i - g(x_i))^2$ subject to the "wiggliness" penalty $\int_a^b |g''(x)|^2 dx \leq c$ for some $c > 0$. This is equivalent to maximizing a penalized log-likelihood. Borrowing from Eilers and Marx (1996), Lang and Brezger (2004) recast and developed this idea into a Bayesian framework. Let $\mathbf{D}_2 \in \mathbb{R}^{(J-2)\times J}$ and $\mathbf{D}_1 \in \mathbb{R}^{(J-1)\times J}$ be defined as

$$\mathbf{D}_2 = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & -2 & 1 \end{bmatrix} \text{ and } \mathbf{D}_1 = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & -1 \end{bmatrix}.$$

For equispaced cases, quadratic (and cubic) B-splines the penalty can be written as $\int_a^b |g''(x)|^2 dx = ||\mathbf{D}_2 \boldsymbol{\theta} \Delta||^2$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_J)$.

Optimization with the $\mathbf{D}_2$ penalty is equivalent to assuming a second order random-walk prior, that is, the improper prior $\mathbf{D}_2\boldsymbol{\theta} \sim N_{J-2}(\mathbf{0}, \lambda^{-1}\mathbf{I}_{J-2})$. As λ becomes large, $g''(x)$ is forced toward zero and $g(x)$ becomes linear. Alternatively, a first order random walk prior is given by $\mathbf{D}_1\boldsymbol{\theta} \sim N_{J-1}(\mathbf{0}, \lambda^{-1}\mathbf{I}_{J-1})$. When λ is large, adjacent basis functions are forced closer and $g'(x)$ is forced toward zero, yielding a constant $g(x)$.

The Bernstein polynomial is a special case of the B-spline with support $[0, 1]$ (Petrone, 1999a,b). A Bernstein polynomial prior for a function g on $[0, 1]$ is a discrete mixture of beta distributions with equispaced means and integer parameters; i.e. the functions $B_j(x)$ in (1.1) are

$$B_j(x) = \frac{\Gamma(J+1)}{\Gamma(j)\Gamma(J-j+1)} x^{j-1}(1-x)^{J-j}.$$

The resulting g is then transformed to $[0, b)$ ($b = \infty$ for some transformations) for use in baseline survival modeling (Gelfand and Mallick, 1995; Carlin and Hodges, 1999; Banerjee and Dey, 2005; Chang et al., 2005; Chen et al., 2014).

B-splines are now a standard tool for modeling hazard functions. Like the GP, the piecewise constant hazard is a special case, i.e. a first order B-spline with $d = 0$; piecewise exponental models have been used extensively in Bayesian survival analysis, e.g. Ibrahim et al. (2001). Existing approaches to modeling hazard functions using B-splines (Gray, 1992; Hennerfeind et al., 2006; Sharef et al., 2010) choose either equispaced knots over the spread of the observed data or knots at the empirical quantiles of the observed event times. Chen et al. (2014) and Li et al. (2015b) instead choose knot locations based on an approximation of underlying parametric family, e.g. $S_{\boldsymbol{\theta}}$ indexed by $\boldsymbol{\theta}$.

Dirichlet process mixture model

A random probability measure G follows a DP (Ferguson, 1973) with parameters (α, G_0), where $\alpha > 0$ and G_0 is an appropriate probability measure defined on $\mathbb{R}^d$, written as

$$G|\alpha, G_0 \sim DP(\alpha G_0), \tag{1.2}$$

if for any measurable nontrivial partition $\{B_l : 1 \leq l \leq k\}$ of $\mathbb{R}^d$, then the vector $(G(B_1), \ldots, G(B_k))'$ has a Dirichlet distribution with parameters $\{\alpha G_0(B_l) : l =$

$1, \ldots, k\}$. It follows that

$$G(B_l)|\alpha, G_0 \sim \text{Beta}(\alpha G_0(B_l), \alpha G_0(B_l^c)),$$

and therefore $E\{G(B_l)|\alpha, G_0\} = G_0(B_l)$ and $Var\{G(B_l)|\alpha, G_0\} = G_0(B_l)G_0(B_l^c)/(\alpha+1)$. Thus G is centered at G_0 with precision α. The DP was used by Susarla and Van Ryzin (1976) to model and estimate the survival function for right-censored data; Müller et al. (2015) provide R code to implement this approach.

If $G|\alpha, G_0 \sim DP(\alpha G_0)$, then the process can be represented by the stick-breaking representation (Sethuraman, 1994),

$$G(\cdot) = \sum_{i=1}^{\infty} w_i \delta_{\boldsymbol{\theta}}(\cdot), \tag{1.3}$$

where $\delta_{\boldsymbol{\theta}}(\cdot)$ is Dirac measure at $\boldsymbol{\theta}$, $w_i = V_i \prod_{j<i}(1 - V_j)$, with $V_i|\alpha \overset{iid}{\sim} \text{Beta}(1, \alpha)$, and $\boldsymbol{\theta}_i|G_0 \overset{iid}{\sim} G_0$. Note that $E(w_j) > E(w_{j+1})$ for all j, so the weights are stochastically ordered.

Convolving a DP with a parametric kernel, such as the normal, gives a DP mixture (DPM) model (Lo, 1984; Escobar and West, 1995). A simple DPM of Gaussian densities for continuous data $\epsilon_1, \ldots, \epsilon_n$ is given by

$$\epsilon_i|G \overset{iid}{\sim} \int N(\mu, \sigma^2) dG(\mu, \sigma^2), \tag{1.4}$$

where $N(\mu, \sigma^2)$ denotes the normal density with mean μ and σ^2, and the mixing distribution, G, is a random probability measure defined on $\mathbb{R} \times \mathbb{R}^+$, following a DP. The stick-breaking representation recasts (1.4) as a countably infinite mixture of normals given by

$$\epsilon_i|G \overset{iid}{\sim} \sum_{j=1}^{\infty} \left[V_j \prod_{k=1}^{j-1}(1 - V_k) \right] N(\mu_j, \sigma_j^2). \tag{1.5}$$

The prior distribution on ϵ_i is centered at the normal distribution; Griffin (2010) discusses prior specifications that control the "non-normalness" of this distribution.

Polya tree

A Polya tree (PT) successively partitions the reals $\mathbb{R}$ (or any other domain) into finer and finer partitions; each refinement of a partition doubles the number of partition sets by cutting the previous level's sets into two pieces; there are two sets at level 1, four sets at level 2, eight sets at 3, and so on. We focus on a PT centered at the standard normal density, that is, $N(0,1)$ is the *centering distribution* for the Polya tree. At level j, the Polya tree partitions the real line into 2^j intervals $B_{j,k} = (\Phi^{-1}((k-1)2^{-j}), \Phi^{-1}(k2^{-j}))$ of probability 2^{-j} under Φ, $k = 1, \ldots, 2^j$, where $\Phi(\cdot)$ is the cumulative distribution function of $N(0,1)$. Note that $B_{j,k} = B_{j+1,2k-1} \cap B_{j+1,2k}$. Given an observation ϵ is in set k at level j, i.e. $\epsilon \in B_{j,k}$, it could then be in either of the two offspring sets $B_{j+1,2k-1}$ or $B_{j+1,2k}$ at level $j+1$. The conditional probabilities associated with these sets will be denoted by $Y_{j+1,2k-1}$ and $Y_{j+1,2k}$. Clearly they must sum to one, and so a common prior for either of these probabilities is a beta distribution (Ferguson, 1974; Lavine, 1992, 1994; Walker and Mallick, 1997, 1999; Hanson and Johnson, 2002; Hanson, 2006a; Zhao et al., 2009), given by

$$Y_{j,2k-1}|c \overset{ind.}{\sim} \text{Beta}(cj^2, cj^2), \quad j = 1, \ldots, J; \; k = 1, \ldots, 2^{j-1},$$

where $c > 0$, which ensures that every realization of the process has a density, allowing the modeling of continuous data without the need of convolutions with continuous kernels.

The user-specified weight $c > 0$ controls how closely the posterior follows $N(0,1)$ in terms of L_1 distance (Hanson et al., 2008), with larger values forcing the PT process G closer to $N(0,1)$; often a prior is placed on c, e.g. $c \sim \Gamma(a,b)$. The PT is stopped at level J (typically $J = 5, 6, 7$); within the sets $\{B_{J,k} : k = 1, \ldots, 2^J\}$ at the level J, G follows $N(0,1)$ (Hanson, 2006a). The resulting model for data $\epsilon_1, \ldots, \epsilon_n$ is given by

$$\epsilon_i|G \overset{iid}{\sim} G, \tag{1.6}$$

where

$$G \sim PT_J(c, N(0,1)). \tag{1.7}$$

The corresponding density is given by

$$p(\epsilon|\{Y_{j,k}\}) = 2^J \phi(\epsilon) \prod_{j=1}^{J} Y_{j,\lceil 2^j \phi(\epsilon) \rceil}, \tag{1.8}$$

where $\lceil \cdot \rceil$ is the ceiling function, and so a likelihood can be formed. For the simple model, the PT is conjugate. Let $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)$. Then

$$Y_{j,2k-1}|\boldsymbol{\epsilon} \overset{ind.}{\sim} \text{Beta}\left(cj^2 + \sum_{i=1}^{n} I\{\lceil 2^j \phi(\epsilon_i) \rceil = 2k-1\}, cj^2 + \sum_{i=1}^{n} I\{\lceil 2^j \phi(\epsilon_i) \rceil = 2k\}\right),$$

and $Y_{j,2k} = 1 - Y_{j,2k-1}$.

Location μ and spread σ parameters are melded with expression (1.6) and the PT prior (1.7) to make a median-μ location-scale family for data $y_1, \ldots, y_n$, given by

$$y_i = \mu + \sigma\epsilon_i,$$

where the $\epsilon_i \mid G \overset{iid}{\sim} G$ and G follows a PT prior as in expression (1.7), with the restriction $Y_{1,1} = Y_{1,2} = 0.5$. Allowing μ and σ to be random induces a mixture of Polya trees (MPT) model for $y_1, \ldots, y_n$, smoothing out predictive inference (Lavine, 1992; Hanson and Johnson, 2002). Note that Jeffreys' prior under the normal model is a reasonable choice here (Berger and Guglielmi, 2001), and leads to a proper posterior (Hanson, 2006a).

1.3 Survival models

Proportional hazards

A proportional hazards (PH) model (Cox, 1972), for continuous data, is obtained by expressing the covariate-dependent survival function $S_{\mathbf{x}}(t)$ as

$$S_{\mathbf{x}}(t) = S_0(t)^{\exp(\mathbf{x}'\beta)}. \tag{1.9}$$

In terms of hazards, this model is

$$h_{\mathbf{x}}(t) = \exp(\mathbf{x}'\boldsymbol{\beta})h_0(t).$$

Note then that for two individuals with covariates $\mathbf{x}_1$ and $\mathbf{x}_2$, the ratio of hazard curves is constant and proportional to $\frac{h_{\mathbf{x}_1}(t)}{h_{\mathbf{x}_2}(t)} = \exp\{(\mathbf{x}_1 - \mathbf{x}_2)'\boldsymbol{\beta}\}$, hence the name "proportional hazards." Cox (1972) is the second most cited statistical paper of all time (Ryan and Woodall, 2005), and the PH model is easily the most popular semiparametric survival model in statistics, to the point where medical researchers tend to compare different populations' survival in terms of instantaneous risk (hazard) rather than mean or median survival as in common regression models. Part of the popularity of the model has to do with the incredible momentum the model has gained from how easy it is to fit the model through partial likelihood (Cox, 1975) and its implementation in SAS in the procedure `PHREG`. The use of partial likelihood and subsequent counting process formulation (Andersen and Gill, 1982) of the model has allowed ready extension to stratified analyses, proportional intensity models, frailty models, and so on (Therneau and Grambsch, 2000).

The first Bayesian semiparametric approach to PH models posits a gamma process as a prior on the baseline cumulative hazard $H_0(t) = \int_0^t h_0(s)ds$ (Kalbfleisch, 1978); partial likelihood emerges as a limiting case (of the marginal likelihood as the precision parameter approaches zero). The use of the gamma process prior in PH models, as well as the beta process prior (Hjort, 1990), piecewise exponential priors, and correlated increments priors are covered in Ibrahim et al. (2001) (pp. 47-94) and Sinha and Dey (1997). Other approaches include what are essentially Bernstein polynomials (Gelfand and Mallick, 1995; Carlin and Hodges, 1999) and penalized B-splines (Hennerfeind et al., 2006; Kneib and Fahrmeir, 2007). The last two models are available in the free software BayesX (Belitz et al., 2015) which can be called from R via the packages `R2BayesX` and `BayesX` (Umlauf et al., 2015). The BayesX functions allow for a general additive (including partially linear) PH model to be easily

fit, including time-dependent covariates; BayesX also accommodates spatial frailties, discussed in Section 1.4. PH models with Polya tree baselines were considered by Hanson (2006a), Hanson and Yang (2007), Zhao et al. (2009), and Hanson et al. (2009) and can be fit in the `SpBayesSurv` package for R.

Stratified PH model posits a separate hazard function across levels of strata $s = 1, \ldots, S$,

$$h_{\mathbf{x},s}(t) = \exp(\mathbf{x}'\boldsymbol{\beta})h_{0s}(t).$$

A version of this model based on Bernstein polynomials is given by Carlin and Hodges (1999); B-splines were considered by Cai and Meyer (2011). The stratified PH model can also be fit using SAS `PHREG` assuming piecewise exponential priors, i.e. piecewise constant baseline hazard functions. A version of the stratified model that SAS fits, but with a "Polya tree" type prior on the hazard was considered by Dukić and Dignam (2007). Note that BayesX can also fit stratified models based on B-splines by including a time-varying regression effect for the categorical strata variable.

Accelerated failure time

An accelerated failure time (AFT) model is obtained by expressing the covariate-dependent survival function $S_{\mathbf{x}}(t)$ as

$$S_{\mathbf{x}}(t) = S_0\{\exp(-\mathbf{x}'\boldsymbol{\beta})t\}. \tag{1.10}$$

This is equivalent to the linear model for the log transformation of the corresponding time-to-event response variable, T,

$$\log T = \mathbf{x}'\boldsymbol{\beta} + \epsilon, \tag{1.11}$$

where $\exp(\epsilon) \sim S_0$. The mean, median, and any quantile of survival for an individual with covariates $\mathbf{x}_1$ is changed by a factor of $\exp\{(\mathbf{x}_1 - \mathbf{x}_2)'\boldsymbol{\beta}\}$ relative to those with covariates $\mathbf{x}_2$.

An early frequentist least-squares teatment of the AFT model with right-censored data is due to Buckley and James (1979); the Buckley-James estimator is implemented in Frank Harrell's `Design` library for R (Alzola and Harrell, 2006). The R packages `emplik` and `bujar` have various extensions. More refined estimators followed in the 1990s (Ying et al., 1995; Yang, 1999) focusing on median-regression.

From a Bayesian nonparametric perspective, the first approach, based on a Dirichlet process prior, obtained approximate marginal inferences to the AFT model (Christensen and Johnson, 1988); a full Bayesian treatment using the Dirichlet process is not practically possible (Johnson and Christensen, 1989). Approaches based on Dirichlet process mixture models have been considered by Kuo and Mallick (1997), Kottas and Gelfand (2001) and Hanson (2006b). Dirichlet process mixtures "fix" the discrete nature of the Dirichlet process, as do other discrete mixtures of continuous kernels. We refer the reader to Komárek and Lesaffre (2007) for an alternative approach based on finite mixtures of normal distributions, and Komárek and Lesaffre (2008) based on an approximating B-spline, both available in the R package `bayesSurv`. Polya tree priors that have continuous densities can directly model the distribution of ϵ in expression (1.11) (Walker and Mallick, 1999; Hanson and Johnson, 2002; Hanson, 2006a; Hanson and Yang, 2007; Zhao et al., 2009). AFT models with Polya tree baseline densities can be fit in the `spBayesSurv` package for R.

Although PH is by far the most commonly-used semiparametric survival model, several studies have shown vastly superior fit and interpretation from AFT models (Hanson and Yang, 2007; Hanson, 2006a; Kay and Kinnersley, 2002; Orbe et al., 2002; Hutton and Monaghan, 2002). Cox pointed out himself (Reid, 1994) *"... the physical or substantive basis for ... proportional hazards models ... is one of its weaknesses ... accelerated failure time models are in many ways more appealing because of their quite direct physical interpretation ..."*. However, similar to the PH model, standard AFT models also impose constraints so that survival curves from different covariate

levels are not allowed to cross, which is unrealistic in many practical applications (e.g., De Iorio et al., 2009). For these data that do not follow AFT assumptions, we next discuss two generalizations of the AFT model that allow for crossing survival and hazard curves. The two approaches are the *linear dependent Dirichlet process mixture*, which can be interpreted as a mixture of parametric AFT models, and the *linear dependent tailfree process*, which is an AFT model with very general baseline functions that are covariate-dependent. Both augmentations are examples of "density regressions," allowing the entire survival density $f_{\mathbf{x}}(t)$ to change smoothly with covariates $\mathbf{x}$.

Linear dependent Dirichlet process

By considering a Dirichlet process mixture of normal distributions for the errors in (1.11) (Kuo and Mallick, 1997), the distribution for the log survival time is the distribution of ϵ_i, given by (1.5), shifted by the linear predictor $\eta_i = \mathbf{x}_i'\boldsymbol{\beta}$. Specifically,

$$y_i|\boldsymbol{\beta}, G \overset{ind.}{\sim} \sum_{j=1}^{\infty} w_j N(\mu_j + \mathbf{x}_i'\boldsymbol{\beta}, \sigma_j^2),$$

where $G(\cdot) = \sum_{j=1}^{\infty} w_j \delta_{(\mu_j, \sigma_j^2)}(\cdot)$ is a Dirichlet process. The interpretation of the components of $\boldsymbol{\beta}$ is as usual and the model can be fit using standard algorithms for Dirichlet process mixture models (Neal, 2000).

The linear dependent Dirichlet process mixture (LDDPM) (De Iorio et al., 2009; Jara et al., 2010, 2011; Zhou et al., 2015b) can be interpreted as a generalization of the previous model, which arises by additionally mixing over the regression coefficients, yielding a mixture of log-normal AFT models. Specifically, the LDDPM model is given by

$$y_i|G \overset{ind.}{\sim} \sum_{j=1}^{\infty} w_j N(\mathbf{x}_i'\boldsymbol{\beta}_j, \sigma_j^2), \tag{1.12}$$

where $\mathbf{x}_i$ now includes a '1' for the intercept, $w_i = V_i \prod_{j<i}(1 - V_j)$, with $V_i|\alpha \overset{iid}{\sim}$ Beta$(1, \alpha)$, and $\boldsymbol{\beta}_j \overset{iid}{\sim} N(\mathbf{m}_0, \mathbf{V}_0)$ and $\sigma_j^{-2} \overset{iid}{\sim} \Gamma(a_0, b_0)$.

The model trades easy interpretability offered by a single $\boldsymbol{\beta}$ for greatly increased flexibility. In particular, the LDDPM model does not stochastically order survival curves from different predictors $\mathbf{x}_{i_1}$ and $\mathbf{x}_{i_2}$, and both the survival and hazard curves can cross.

Linear dependent tailfree process

A Polya trees defines the conditional probabilities $Y_{j+1,2k-1}$ and $Y_{j+1,2k}$ as beta distributions. However, one can instead define a logistic regression for each of these probabilities, allowing the *entire* shape of the density to change with covariates; this is the approach considered by Jara and Hanson (2011). Given covariates $\mathbf{x}$, the linear dependent tailfree process (LDTFP) models $(Y_{j+1,2k-1}, Y_{j+1,2k})$ through logistic regressions

$$\log\{Y_{j+1,2k-1}(\mathbf{x})/Y_{j+1,2k}(\mathbf{x})\} = \mathbf{x}'\boldsymbol{\tau}_{j,k},$$

where $\mathbf{x}$ includes an intercept. There are $2^J - 1$ regression coefficient vectors $\boldsymbol{\tau} = \{\boldsymbol{\tau}_{j,k}\}$; e.g. for $J = 3$, $\{\boldsymbol{\tau}_{0,1}, \boldsymbol{\tau}_{1,1}, \boldsymbol{\tau}_{1,2}, \boldsymbol{\tau}_{2,1}, \boldsymbol{\tau}_{2,2}, \boldsymbol{\tau}_{2,3}, \boldsymbol{\tau}_{2,4}\}$. Let $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_n]'$ be the $n \times p$ design matrix. Following Jara and Hanson (2011), each is assigned an independent normal prior, $\tau_{j,k} \sim N_p\left(\mathbf{0}, \frac{2}{c(j+1)^2}\boldsymbol{\Psi}\right)$. Jara and Hanson (2011) discussed the case $\boldsymbol{\Psi} = n(\mathbf{X}'\mathbf{X})^{-1}$, generating a g-prior Zellner (1983) for the tailfree regression coefficients. By setting $\boldsymbol{\tau}_{0,1} \equiv \mathbf{0}$, the resulting LDTFP is almost surely a median-zero probability measure for every $\mathbf{x} \in \mathcal{X}$, important to avoid identifiability issues.

Augmenting (1.8), the random density is given by

$$g_{\mathbf{x}}(\epsilon) = \phi(\epsilon)2^J \prod_{j=1}^{J} Y_{j,\lceil 2^j \Phi(\epsilon)\rceil}(\mathbf{x}).$$

Since the $\{Y_{j,k}\}$ are modeled with logistic-normal distribution instead of beta, the resulting random density is a tailfree process. The final AFT model with LDTFP baseline is given by

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \sigma\epsilon_i, \quad \epsilon_i|\boldsymbol{\tau} \overset{ind.}{\sim} g_{\mathbf{x}_i}. \tag{1.13}$$

Unlike the LDDPM, the LDTFP separates survival into one distinct trend $\mathbf{x}'\boldsymbol{\beta}$ and an evolving log-baseline survival density $g_{\mathbf{x}}$. By forcing $g_{\mathbf{x}}$ to be median-zero, e^{β_j} gives a factor by how median survival changes when x_j is increased *just as in standard AFT models*. This heightened interpretability in terms of median-regression in the presence of heteroscedastic error allows a fit of the LDTFP model to easily relate covariates $\mathbf{x}$ to median survival.

The LDTFP models the probability of falling above or below quantiles of the $N(\mathbf{x}'\boldsymbol{\beta}, \sigma^2)$ distribution, but in terms of conditional probabilities. This model can be viewed as a particular kind of quantile regression model. Koenker and Hallock (2001) suggest that "*...instead of estimating linear conditional quantile models, we could instead estimate a family of binary response models for the probability that the response variable exceeded some prespecified cutoff values.*" However, Koenker and Hallock (2001) prefer the linear (in covariates) quantile specification because "*...it nests within it the iid error location shift model of classical linear regression.*" By augmenting a median-zero tailfree process with a general trend $\mathbf{x}'\boldsymbol{\beta}$ we accomplish the same objective, nesting the ubiquitous normal-errors linear model within a highly flexible median regression model, but with heteroscedastic error that changes shape with covariate levels $\mathbf{x} \in \mathcal{X}$.

Both the LDDPM and the LDTFP model the entire density at every covariate level $\mathbf{x} \in \mathcal{X}$, so full density and hazard estimates are available, accompanied by reliable interval estimates, unlike many median (and other quantile) regression models. Both models are implemented as user-friendly functions calling compiled FORTRAN in `DPpackage` or calling compiled C++ in `spBayesSurv` for `R`. These functions accommodate general interval-censored data (including current status data); the latter package also allows for spatial correlation. If only a trend function is desired one could instead use quantile regression models, such as the ones implemented in the excellent `quantreg` package in `R` (Koenker, 2008).

Proportional odds

The proportional odds (PO) model has recently gained attention as an alternative to the PH and AFT models. PO defines the survival function $S_{\mathbf{x}}(t)$ for an individual with covariate vector $\mathbf{x}$ through the relation

$$\frac{S_{\mathbf{x}}(t)}{1-S_{\mathbf{x}}(t)} = \exp\{-\mathbf{x}'\boldsymbol{\beta}\}\left(\frac{S_0(t)}{1-S_0(t)}\right). \tag{1.14}$$

The odds of dying before any time t are $\exp\{(\mathbf{x}_1-\mathbf{x}_2)'\boldsymbol{\beta}\}$ times greater for those with covariates $\mathbf{x}_1$ versus $\mathbf{x}_2$.

The first semiparametric approaches to PO models involving covariates are due to Cheng et al. (1995), Murphy et al. (1997), and Yang and Prentice (1999). A semiparametric frequentist implementation of the PO model is available in the package `timereg` (Martinussen and Scheike, 2006) for `R`. Bayesian nonparametric approaches for the PO model have been based on Bernstein polynomials (Banerjee and Dey, 2005), B-splines (Wang and Dunson, 2011; Lin and Wang, 2011), and Polya trees (Hanson, 2006a; Hanson and Yang, 2007; Zhao et al., 2009; Hanson et al., 2011).

The PH, AFT, and PO models all make overarching assumptions about the data generating mechanism for the sake of obtaining succinct data summaries. An important aspect associated with the Bayesian nonparametric formulation of these models is that, by assuming the *same, flexible model* for the baseline survival function, they are placed on a common ground (Hanson, 2006a; Hanson and Yang, 2007; Zhang and Davidian, 2008; Zhao et al., 2009; Hanson et al., 2011). Furthermore, parametric models are special cases of the nonparametric models. Differences in fit and/or predictive performance can therefore be attributed to the *survival* models only, rather than to additional possible differences in quite different nonparametric models or estimation methods.

Of the Bayesian approaches based on Polya trees considered by Hanson (2006a), Hanson and Yang (2007), Zhao et al. (2009) and Hanson et al. (2011), the PO model

was chosen over PH and AFT according to the log-pseudo marginal likelihood (LPML) criterion (Geisser and Eddy, 1979). In three of these works, the parametric log-logistic model, a special case of PO that also has the AFT property, was chosen. This may be due to the fact that the PO assumption implies that hazard ratios $\lim_{t\to\infty} \frac{h_{\mathbf{x}_1}(t)}{h_{\mathbf{x}_2}(t)} = 1$, that is, eventually everyone has the same risk of dying tomorrow. These authors also found that, everything else being equal, the actual semiparametric model chosen (PO, PH or AFT) affects prediction far more than whether the baseline is modeled nonparametrically. It is worth noting that none of these papers favored the semiparametric PH model in actual applications.

Other semiparametric models

PH, AFT, and PO are three of many semiparametric survival models used in practice. There are a few more hazard-based models including the additive hazards (AH) model (Aalen, 1980, 1989), given by

$$h_{\mathbf{x}}(t) = h_0(t) + \mathbf{x}'\boldsymbol{\beta},$$

which is implemented in the `timereg` package for `R`. An empirical Bayes approach to this model based on the gamma process was implemented by Sinha et al. (2009). Fully Bayesian approaches require an elaborate model specification to incorporate the rather awkward constraint $h_0(t) + \mathbf{x}'\boldsymbol{\beta} \geq 0$ for $t > 0$ (Yin and Ibrahim, 2005; Dunson and Herring, 2005). Recently, there has been some interest in the accelerated hazards model (Chen and Wang, 2000; Zhang et al., 2011; Chen et al., 2014), given by

$$h_{\mathbf{x}}(t) = h_0\{\exp(-\mathbf{x}'\boldsymbol{\beta})t\}.$$

This model allows hazard and survival curves to cross.

Finally, several interesting "super models" have been proposed in the literature, including non-proportional hazard regression models that include PH as a special

case (Devarajan and Ebrahimi, 2011), generalized odds-rate hazards models that include PH and PO as special cases (Dabrowska and Doksum, 1988; Scharfstein et al., 1998), Box-Cox transformation regression models that include PH and AH as special cases (Yin and Ibrahim, 2005; Martinussen and Scheike, 2006), and extended hazard regression models that include both PH and AFT as special cases (Chen and Jewell, 2001; Li et al., 2015b).

1.4 Spatial dependence

When survival data are spatially correlated, it is often of scientific interest to investigate possible spatial dependence in survival outcomes after adjusting for known subject-specific covariate effects. Such spatial dependence is often due to region-specific similarities in ecological and/or social environments that are typically not measurable. We next discuss two general approaches, *frailty* and *copula*, for incorporating spatial dependence into the semiparametric models presented in Section 1.3, followed by some other possibilities.

Spatial frailty modeling

Frailties have been frequently used to induce correlation among related survival times in models which have a linear predictor. The linear predictor is augmented $\eta_i = \mathbf{x}_i'\boldsymbol{\beta} + v_i$, where v_i is a random effect, termed "frailty," accounting for heterogeneity after adjusting for covariates. The so-called shared frailty models have one common random effect within each group, e.g. $v_i = z_{g_i}$ where $g_i \in \{1, \ldots, G\}$ is the group—e.g. county, hospital, family—to which observation i belongs. Early literature considered exchangeable frailties with $z_1, \ldots, z_G \overset{iid}{\sim} H$, where H was constrained to be mean or median zero to avoid confounding with the baseline function.

In the case of spatial survival data, one can extend the frailty model by including

a spatial effect, e.g.,

$$\eta_i = \mathbf{x}_i'\boldsymbol{\beta} + \gamma_i,\ \ \gamma_i = v_i + w_i,$$

where the frailty term γ_i incorporates the effects of both heterogeneity (via the non-spatial frailty v_i) and spatial dependence (through the spatial frailty w_i); note that, however, in applications often only spatial dependence is modeled ($\gamma_i = w_i$) or exchangeable dependence ($\gamma_i = v_i$). Spatial frailty models have been widely discussed in the literature and correspond to particular cases of hierarchical models. Such models are usually grouped into two general settings according to their underlying data structure: *point-referenced* (geostatistical) data, where the location $\mathbf{s}_i$ varies continuously throughout a fixed study region $\mathcal{D}$, and *areal* (lattice) data, where the study region is partitioned into a finite number of areal units with well-defined boundaries (Banerjee et al., 2014).

Point-referenced data modeling

In modeling point-referenced data, the non-spatial frailty term v_i is often specified $v_i \overset{iid}{\sim} N(0, \sigma^2)$, and the spatially correlated frailties $\mathbf{w} = (w_1, \ldots, w_n)$ can be specified to have a multivariate Gaussian distribution:

$$\mathbf{w} \sim N_n(\mathbf{0}, \theta^2\mathbf{R}), \tag{1.15}$$

where N_n denotes the n-dimensional Gaussian distribution, θ^2 measures the amount of spatial variation across locations, and the (i, j)th element of $\mathbf{R}$, denoted by $\mathbf{R}_{ij}$, is the correlation between w_i and w_j. An isotropic correlation function is commonly used to construct $\mathbf{R}$, where the correlation of any two subjects is a function solely of the distance d_{ij} between their locations $\mathbf{s}_i$ and $\mathbf{s}_j$, i.e., $\mathbf{R}_{ij} = \rho(d_{ij})$. A flexible, frequently used correlation function is the Matérn

$$\rho(d_{ij}) = \frac{(\phi d_{ij})^\nu K_\nu(\phi d_{ij})}{2^{\nu-1}\Gamma(\nu)}, \tag{1.16}$$

where K_ν is a modified Bessel function of the third kind, $\phi > 0$ measures the spatial decay over distance, and $\nu > 0$ is a parameter controlling the smoothness of the realized random field. Interested readers are referred to Banerjee et al. (2014) for further discussion of correlation functions. Note that the Matérn reduces to the exponential $\rho(d_{ij}) = \exp(-\phi d_{ij})$ for $\nu = 0.5$ and the Gaussian $\rho(d_{ij}) = \exp(-\phi^2 d_{ij}^2)$ when $\nu \to \infty$. Under the above prior specifications of exchangeable normal v_i and spatially correlated w_i, the resulting multivariate Gaussian distribution on frailties $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)$ is

$$\boldsymbol{\gamma} \sim N_n \left\{ \mathbf{0}, \theta^2 \mathbf{R} + \sigma^2 \mathbf{I} \right\}. \tag{1.17}$$

With this representation, the non-spatial effect variance σ^2 is often called the *nugget*, the spatial effect variance θ^2 is called the *partial sill*, and the total effect variance $\theta^2 + \sigma^2$ is called the *sill*. The rationale of including the nugget effect is that we don't expect all remaining individual heterogeneity to be accounted for by the spatial story, as other factors (e.g., measurement error, replication error, micro-scale error) may also potentially explain the heterogeneity. In Henderson et al. (2002), the term $\tau = \theta^2/(\theta^2 + \sigma^2)$ is called the *nugget effect* and interpreted as the proportion of the heterogeneity variance that is explained by spatial effects.

For posterior inference, MCMC requires computing the inverse and determinant of n-dimensional correlation matrix $\mathbf{R}$ in each iteration. With an increasing sample size n, such computation becomes very expensive and even unstable due to a large amount of numerical operations. This situation is often referred to as "the big n problem." Various approaches have been developed to approximate the correlation function such as predictive process models (Banerjee et al., 2008; Finley et al., 2009), sparse approximations (Furrer et al., 2006; Kaufman et al., 2008), and the full scale approximation (FSA) method (Sang and Huang, 2012). The last approximation is the summation of the former two approximations, which can capture both large- and small-scale spatial dependence. The FSA has been successfully applied to model

point-referenced survival data in Zhou et al. (2015b, Chapter 3) and implemented in the R package `spBayesSurv`.

Areal data modeling

In the case of areal data, the whole study region $\mathcal{D}$ is often partitioned into a finite number of areas, say $B_1, \ldots, B_G$, and a common frailty is assumed for the subjects within each area, i.e.

$$\eta_i = \mathbf{x}_i'\boldsymbol{\beta} + \gamma_{g_i},\ \gamma_j = v_j + w_j, j = 1, \ldots, G.$$

Here the non-spatial frailty v_j for each area is typically assigned a mean-zero normal distribution with variance σ^2. For the spatial frailty term w_j, there has been two general approaches. First, one can assume a fully specified mean-zero multivariate Gaussian distribution on $\mathbf{w} = (w_1, \ldots, w_G)$ with covariance matrix $\theta^2\mathbf{R}$, where $\mathbf{R}_{ij}$ is modeled using a traditional correlation function like the Matérn in (1.16) but with d_{ij} representing the distance between two areal centroids. Another way is to consider an intrinsic conditionally autoregressive (ICAR) model. Let $a_{ij} = 1$ if areas B_i and B_j share a nontrivial border (i.e. a connected curve in $\mathbb{R}^2$ that is more than one point) and $a_{ij} = 0$ otherwise; set $a_{ii} = 0$. Then the $G \times G$ matrix $\mathbf{A} = [a_{ij}]$ is called the adjacency matrix for the region $\mathcal{D}$. The ICAR prior is defined through the set of all conditional distributions

$$w_j | \{w_i : i \neq j\} \sim N\left(\bar{w}_j,\ \theta^2/a_{j+}\right),\ j = 1, \ldots, G, \tag{1.18}$$

denoted $\mathbf{w} \sim \text{ICAR}(1/\theta^2)$, where a_{j+} is the number of neighbors of area B_j, $\bar{w}_j = \frac{1}{a_{j+}} \sum_{i:a_{ij}=1} w_i$ is the sample mean of the a_{j+} values of the neighboring areal unit frailties, and θ^2/a_{j+} is the conditional variance. Note that the ICAR model induces an improper joint density, and the constraint $\sum_{j=1}^{G} w_j = 0$ is commonly used to avoid identifiability issues. Another common fix is to assume a proper CAR model

by multiplying the conditional mean $\bar{w}_j$ in (1.18) by a shrinkage scale parameter ρ, where $0 \leq \rho < 1$; it is generally difficult to estimate ρ and θ^2 simultaneously.

Related literature

Henderson et al. (2002) modeled the spatial structure of leukemia survival data using both district-level and point-referenced frailty effects in the context of the PH model. In their point-referenced analysis, a multivariate gamma distribution for $(e^{\gamma_1}, \ldots, e^{\gamma_n})'$ was constructed so that each marginal has a gamma distribution with mean 1 and variance $\sigma^2 + \theta^2$, and the correlation between e^{γ_i} and e^{γ_j} takes the form defined in (1.17). In their district-level analysis, they considered a linear predictor with individual frailties as $\eta_i = \mathbf{x}_i'\boldsymbol{\beta} + \gamma_i$, where $e^{\gamma_i}|\mu_{g_i} \sim \Gamma\left(1/\xi, 1/(\xi\mu_{g_i})\right)$. They then assumed a multivariate Gaussian distribution on the latent effects $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_G)$ with the correlation function between the ith and jth district modeled via the powered exponential and Matérn. They also considered the ICAR specification on $\boldsymbol{\mu}$ and found that the multivariate Gaussian via a Matérn correlation with $\nu = 2$ had the best fit based on the DIC goodness-of-fit criterion.

Pan et al. (2014) fitted the semiparametric PH model with ICAR frailties to interval censored data with the baseline hazard function modeled via B-splines. Lin et al. (2015) duplicated this model without the ICAR frailties. Using the same methodology, a special case of interval-censored data, current-status data, was presented in Cai et al. (2011). The aforementioned models can be fit in the `ICBayes` R package. Li and Ryan (2002) modeled the district-level frailty effect using a fully specified multivariate normal prior within the framework of PH, and applied the model to detect prognostic factors leading to childhood asthma. All of these approaches are essentially a special case of the general models previously presented in Kneib (2006) and Hennerfeind et al. (2006), which can be efficiently fit in the freely available program BayesX or the R package `R2BayesX`; the latter package uses compiled code and places

the B-spline prior on the log-hazard instead of the hazard. An advantage of the models fitted in BayesX is that both areal and point-referenced data are accommodated as well as nonparametric additive effects. In addition, the R package `spatsurv` can also fit the PH model with multivariate Gaussian frailties, where the baseline hazard is modeled either parametrically or nonparametrically via B-splines.

Banerjee and Carlin (2003) developed a semiparametric PH frailty model for capturing spatio-temporal heterogeneity in survival of women diagnosed with breast cancer in Iowa, using a mixture of beta densities baseline. Banerjee et al. (2003) applied the Weibull parametric PH frailty model to infant mortality data in Minnesota, where the county-level frailties were assumed to have either an uncorrelated zero-mean Gaussian prior, an ICAR prior, or a fully specified multivariate Gaussian prior as in (1.15). They showed that the fully specified prior provides the best model fitting in terms of DIC in the analysis of the infant data. Banerjee and Dey (2005) utilized the same frailty modeling technique for capturing spatial heterogeneity within the framework of semiparametric PO, found that the proper CAR prior yielded the best fit in the application to a subset of SEER breast cancer data. Zhao et al. (2009) considered either an AFT, PH, or PO model with ICAR frailties, where the baseline function was assumed to have a mixture of Polya trees prior. Zhang and Lawson (2011) and Wang et al. (2012) developed parametric and semiparametric AFT models with ICAR frailties, respectively. Chernoukhov (2013) extended the additive hazards model for allowing various spatial dependence structures in his dissertation. Zhou et al. (2015c, Chapter 4) extended the generalized model in (1.13) by allowing frailties accommodating spatial correlation via the ICAR prior distribution. The models proposed in Zhao et al. (2009) and Zhou et al. (2015c, Chapter 4) can be fit in the R package `spBayesSurv`. Other references focusing on spatial frailty modeling and its application include McKinley (2007), Diva et al. (2008), Darmofal (2009), Liu (2012), Ojiambo and Kang (2013), Dasgupta et al. (2014), Li et al. (2015a), and

among others.

Spatial copula modeling

Spatial copulas are just beginning to become popular in geostatistics. The use of copulas in the spatial context was first proposed by Bárdossy (2006), where the empirical variogram is replaced by empirical copulas to investigate the spatial dependence structure. The spatial copula approach offers an appealing way to separate modeling from the spatial dependence structure for multivariate distributions. Copulas completely describe association among random variables separately from their univariate distributions and thus capture joint dependence without the influence of the marginal distribution (Li, 2010). In the context of survival models, the idea of spatial copula approach is to first assume that the survival time T_i at location $\mathbf{s}_i$ marginally follows a model $S_{\mathbf{x}_i}(t)$ introduced in Section 1.3, then model the joint distribution of $(T_1, \ldots, T_n)'$ as

$$F(t_1, \ldots, t_n) = C(F_{\mathbf{x}_1}(t_1), \ldots, F_{\mathbf{x}_n}(t_n)), \tag{1.19}$$

where $F_{\mathbf{x}_i}(t) = 1 - S_{\mathbf{x}_i}(t)$ is the cumulative distribution function and the function C is an n-copula used to capture spatial dependence. If we let $U_i = F_{\mathbf{x}_i}(T_i)$, then the problem is reduced to constructing a copula for modeling the joint distribution of $\mathbf{U} = (U_1, \ldots, U_n)$. Hereafter we assume that U_i follows a uniform distribution on $[0, 1]$ for all locations $\mathbf{s}_i$; i.e., the survival model $S_{\mathbf{x}_i}(t)$ is assumed to be correctly specified. In fact, copulas are all the joint cumulative distribution functions on the unit hypercube with uniform marginal distributions. We refer interested readers to Nelsen (2006) for general introduction to copulas and to Smith (2013) for Bayesian approaches to copula modeling.

In the geostatistical framework, the multivariate spatial copula of $\mathbf{U}$ is often constructed so that for any selected two locations $\mathbf{s}_i$ and $\mathbf{s}_j$, the bivariate copula (i.e., joint distribution) of (U_i, U_j) does not depend on the locations $\mathbf{s}_i$ and $\mathbf{s}_j$ but on their

distance d_{ij} only. However, such construction is not a trivial task. Here we introduce a spatial version of the Gaussian copula and refer readers to Li (2010) for further discussion of other theoretical spatial copulas. Define $Z_i = \Phi^{-1}\{U_i\}$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function, then we have $Z_i \sim N(0,1)$ for all i. If we further assume that $\mathbf{Z} = (Z_1, \ldots, Z_n)'$ follows a multivariate normal distribution with mean zero and covariance $\mathbf{R}$, i.e., $\mathbf{Z} \sim N_n(\mathbf{0}, \mathbf{R})$, then the induced joint distribution of $\mathbf{U}$ is called the Gaussian copula, which is given by

$$C(u_1, \ldots, u_n) = \Phi_n\left(\Phi^{-1}\{u_1\}, \ldots, \Phi^{-1}\{u_n\}; \mathbf{R}\right), \tag{1.20}$$

where $\Phi_n(\ldots; \mathbf{R})$ denotes the distribution function of $N_n(\mathbf{0}, \mathbf{R})$. Note that all the diagonal elements of $\mathbf{R}$ are ones, so we refer to $\mathbf{R}$ as the correlation matrix thereafter. The Gaussian copula has a symmetrical density, which can be written as

$$c(u_1, \ldots, u_n) = |\mathbf{R}|^{-1/2} \exp\left\{\frac{1}{2}\mathbf{z}'(\mathbf{R}^{-1} - \mathbf{I})\mathbf{z}\right\}, \tag{1.21}$$

where $\mathbf{z} = (z_i, \ldots, z_n)'$ with $z_i = \Phi^{-1}\{u_i\}$ and $\mathbf{I}$ is the identity matrix. The spatial dependence structure of the Gaussian copula is induced by constructing the correlation matrix $\mathbf{R}$ using classical geostatistical models. For example, the (i,j)th element of $\mathbf{R}$ can be defined using the Matérn in (1.16) with a nugget effect τ, that is, $\mathbf{R}_{ij} = \tau\rho(d_{ij})$ for $i \neq j$, where $1 < \tau < 1$. Under the spatial Gaussian copula, the joint density of $(T_1, \ldots, T_n)$ takes the form

$$f(t_1, \ldots, t_n) = |\mathbf{R}|^{-1/2} \exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{R}^{-1} - \mathbf{I}_n)\mathbf{z}\right\} \prod_{i=1}^{n} f_{\mathbf{x}_i}(t_i), \tag{1.22}$$

where $z_i = \Phi^{-1}\{F_{\mathbf{x}_i}(t_i)\}$ and $f_{\mathbf{x}_i}(t_i)$ is the density function of T_i. The use of spatial copulas has not been widely applied for modeling survival data that are subject to spatial correlation. Li and Lin (2006) successfully applied the spatial Gaussian copula approach to a semiparametric PH model and proposed spatial semiparametric estimating equations that yield consistent and asymptotically normal estimators. Zhou et al. (2015b, Chapter 3) considered the LDDPM marginal model given in (1.12) using

the same Gaussian copula for capturing spatial dependence structure, where MCMC algorithms were used to obtain posterior inferences. Zhou et al. (2015b, Chapter 3) also provided a Bayesian version of the model considered in Li and Lin (2006) using piecewise exponential baseline specifications. The R package `spBayesSurv` can fit the aforementioned copula-based Bayesian survival models.

The spatial Gaussian copula approach can also be extended for fitting lattice data, for which constructing the correlation matrix $\mathbf{R}$ of $\mathbf{Z} = (Z_1, \ldots, Z_n)$ becomes a challenging task. One may consider a random effects model for $\mathbf{Z}$ based on the partition of the domain $\mathcal{D}$ into G districts, that is,

$$Z_i = \mu_{g_i} + \epsilon_i,\ \boldsymbol{\mu} \sim N_G(\mathbf{0}, \mathbf{B}\boldsymbol{\Omega}\mathbf{B}),\ \epsilon_i \overset{ind}{\sim} N\left(0, \frac{\sigma^2}{\omega_{g_i g_i} + \sigma^2}\right), g_i \in \{1, \ldots, G\}, \quad (1.23)$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_G)'$ are the random effects, $\boldsymbol{\Omega} = [\omega_{ij}]$ is a $G \times G$ matrix introducing spatial dependence to $\boldsymbol{\mu}$, $\mathbf{B} = \text{diag}\left(1/\sqrt{\omega_{11} + \sigma^2}, \ldots, 1/\sqrt{\omega_{GG} + \sigma^2}\right)$, and ϵ_i is the error term independent of the spatial random effects. Note that $Var(Z_i) = 1$. Popular models for Ω include multivariate Gaussian coupled with a spatial covariance function, ICAR, proper CAR and many others. Li et al. (2015b) derived the implied correlation matrix $\mathbf{R} = cov(\mathbf{Z})$ under the ICAR model, which only involves one unknown quantity ψ^*. A smaller value of ψ^* corresponds to stronger spatial dependence. With the specification of $\mathbf{R}$, one can model joint cumulative distribution function of $(T_1, \ldots, T_n)$ by

$$F(t_1, \ldots, t_n) = \Phi_n\left(\Phi^{-1}\{F_{\mathbf{x}_i}(t_i)\}, \ldots, \Phi^{-1}\{F_{\mathbf{x}_n}(t_n)\}; \mathbf{R}\right). \quad (1.24)$$

Other spatial dependence modelings

Zhao and Hanson (2011) considered a stratified PH model:

$$S_{\mathbf{x}_i} = S_{0g_i}(t)^{\exp(\mathbf{x}_i'\beta)},\ g_i \in \{1, \ldots, G\},$$

where each region-specific baseline $S_{0j}(\cdot)$ approximately follows a mixture of Polya trees prior centered at a parametric log-logistic family. The spatial dependence a-

mong the $\{S_{01}(\cdot), \ldots, S_{0G}(\cdot)\}$ is induced through proper CAR priors on the logit transformed Polya tree conditional probabilities $\{Y_{l,k}\}$. Hanson et al. (2012) extended this idea to fit a Bayesian semiparametric temporally stratified PH model with spatial frailties. Stratified AFT models with ICAR areal frailties are considered by Zhou et al. (2015c, Chapter 4).

In modeling areal data, spatial dependence is often due to unadjusted district-level risk factors that may potentially relate to survival outcomes. Zhao and Hanson (2011) note that spatial frailties serve as proxies to unmeasured region-level covariates, but are less-precise adjustments since region-level covariates (such as shortest distance to a clinic) are unlikely to sharply change at areal boundaries. Therefore it is natural to introduce spatial dependence by allowing frailties to depend on region-level covariates, especially when information is available on each region that may affect the survival outcome beyond the recorded covariates. For this reason, Zhou et al. (2015a, Chapter 2) proposed a region-level covariate adjusted frailty PH model. Specifically, with the linear predictor $\eta_i = \mathbf{x}_i'\beta + \gamma_{g_i}$, they assume an LDTFP prior on the frailties, i.e., $\gamma_j|\mathbf{z}_j \sim g_{\mathbf{z}_j}(\cdot)$, where $\mathbf{z}_j$ is a vector of region-level covariates. This model can be fit in the `DPpackage` for R.

1.5 Illustrations

Both of the frailty and copula modeling approaches are illustrated using real-life datasets. All the analyses are implemented using the R packages `spBayesSurv`. The fitted models are compared in terms of the log pseudo marginal likelihood (LPML) developed by Geisser and Eddy (1979). Note that the frailties used in frailty models are either exchangeable v_i or spatial w_i, but not both $v_i + w_i$.

SEER cancer data

The Surveillance Epidemiology and End Results (SEER) program of the National Cancer Institute (`seer.cancer.gov`) is an authoritative source of information on cancer incidence and survival in the US, providing county-level cancer data on an annual basis for particular states for public use. Areal-referenced SEER data have been analyzed by many authors in the context of spatial frailty models (e.g., Banerjee and Carlin, 2003; Banerjee and Dey, 2005; Zhao et al., 2009; Zhao and Hanson, 2011; Wang et al., 2012; Zhou et al., 2015a,c).

For illustration, we analyze a subset of the Iowa SEER breast cancer survival data, which consists of a cohort of 1073 Iowan women, who were diagnosed with malignant breast cancer starting in 1995, and enrollment and follow-up continued through the end of 1998. This data set has been analyzed in Zhao et al. (2009), and Zhou et al. (2015a, Chapter 2). The observed survival time, from 1 to 48, is defined as the number of months from diagnosis to either death or the last follow-up. Here we assume that only deaths due to metastasis of cancerous nodes in the breast are events, while the deaths from other causes are censored at the time of death. The right-censoring rate is 54.5%. For each patient, the observed survival time and county of residence at diagnosis are recorded. The considered individual-level covariates include age at diagnosis and the stage: local, regional, or distant, where two dummy variables are created for regional and distant, respectively, and the reference group is local. Zhou et al. (2015a, Chapter 2) point out that some county-level socioeconomic factors (e.g., median household income, poverty level, education, rurality) are also potentially associated with breast cancer and argue that rural counties present more heterogeneity in access to quality care and screening for breast cancer. Therefore, we also include a county-level covariate "Rural-Urban Continuum Codes" (RUCC) measuring degree of urbanization; see Zhou et al. (2015a, Chapter 2) for a detailed description.

Table 1.1 SEER breast cancer data: Posterior medians (95% credible intervals) of fixed effects from various models. Note the AFT model is parameterized as $S_{\mathbf{x}}(t) = S_0(e^{\mathbf{x}'\beta}t)$.

Model	Centered age	Regional stage	Distant stage	RUCC
PH/CAR	0.019 (0.013, 0.025)	0.26 (0.03, 0.48)	1.69 (1.45, 1.93)	-0.069 (-0.136, 0.002)
AFT/CAR	0.018 (0.011, 0.023)	0.22 (0.01, 0.43)	1.51 (1.26, 1.75)	-0.045 (-0.105, 0.013)
PO/CAR	0.030 (0.021, 0.038)	0.40 (0.12, 0.69)	2.59 (2.25, 2.95)	-0.087 (-0.174, -0.001)
PH/LDTFP	0.019 (0.013, 0.025)	0.27 (0.03, 0.49)	1.64 (1.43, 1.88)	-0.105 (-0.185, -0.041)

We fit each of the PH, AFT, and PO frailty models with a mixture of Polya trees prior on baseline survival $S_0(t)$ and the ICAR prior on the frailties $\boldsymbol{\gamma} \sim \text{ICAR}(\lambda)$, where the PH is centered at the Weibull $G_{\boldsymbol{\theta}}(t) = 1 - \exp\left\{-(e^{\theta_1}t)^{\exp(\theta_2)}\right\}$ and the AFT and PO are centered at the log-logistic $G_{\boldsymbol{\theta}}(t) = 1 - \{1 + (e^{\theta_1}t)^{\exp(\theta_2)}\}^{-1}$. We consider the following prior settings: $J = 4$, $c \sim \Gamma(5, 1)$, $\boldsymbol{\theta} \sim N_2(\hat{\boldsymbol{\theta}}, \hat{\mathbf{V}})$, $\boldsymbol{\beta} \sim N_p(\hat{\beta}, 30\hat{\boldsymbol{\Sigma}})$ and $\lambda \sim \Gamma(1, 1)$, where $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\beta}}$, $\hat{\mathbf{V}}$, and $\hat{\boldsymbol{\Sigma}}$ are maximum likelihood estimates from the underlying parametric model. Using the same priors, we also fit the above models with Gaussian exchangeable frailties and without frailties. For all models considered, a burn-in of $100,000$ iterations is followed by a run of $100,000$ thinned down to $10,000$ iterations. All these models are fitted using the `survregbayes` function available in the package `spBayesSurv`.

The LPML values under ICAR frailty PH, AFT, and PO are -2226, -2228, and -2210, respectively, while the corresponding LPMLs are -2230, -2224 and -2214 under exchangeable frailty models and are -2230, -2228, and -2214 under non-frailty models. We can observe that the ICAR frailty model has the best predictive ability within the context of either PH or PO, and the exchangeable frailty model performs best in terms of LPML under the AFT. Table 1.1 presents posterior means and equal-tailed 95% credible intervals (CI) for covariate effects under each of above model with ICAR frailties. All individual covariate effects are significant in each model. Higher age at diagnosis increases the hazard; e.g. a twenty-year increase in age is associated with an $\exp(0.019 \times 20) \approx 1.46$-fold increase in hazard. Using women with local stage of disease as the reference, the hazard rate of women of the same age

who live in the same county will be $\exp(0.26) \approx 1.30$ times larger if their cancer is detected at the regional stage, and $\exp(1.69) \approx 5.42$ times larger if detected at the distant stage. Under the AFT assumption, among patients living in the same county and having same age, a woman with local stage typically survives $\exp(0.22) \approx 1.25$ times longer than a woman with regional stage, and $\exp(1.51) \approx 4.53$ times longer than a woman with distant stage. Finally, for the PO model, after adjusting for the age at diagnosis and the RUCC, the odds of dying from breast cancer before any time t are $\exp(0.40) \approx 1.49$ greater for regional stage versus local stage, and are $\exp(2.59) \approx 13.33$ greater for distant stage versus local stage. These findings are confirmed in Figure 1.1, which shows the fitted survival functions for women aged at 68.8 years and living a county with RUCC at 5 for distant and local stages under the three competing models and assuming a spatial frailty of zero. Turning to the county-level RUCC effect, only the PO model provides a significant result at the 0.05 level; living in more urban counties is associated with poorer survival after a breast cancer diagnosis on average.

Zhou et al. (2015a, Chapter 2) fitted a PH model with LDTFP frailty terms using the package `DPpackage` and found more variability for frailties of rural counties. The resulting LPML is -2222 when RUCC is included into both the linear predictor and frailty terms. The pseudo Bayes factor for the LDTFP frailty model versus the ICAR frailty PH model is $\exp(2226 - 2222) \approx 55$, implying that the allowing frailties depending on RUCC improves the model's predictive ability about 55 times. Table 1.1 also shows the covariate effects under the LDTFP frailty PH model. An interesting finding is that now the RUCC effect becomes significant at the 0.05 level. This may be due to the fact that frailty distributions are covariate-dependent as shown in Figure 1.1(b). After controlling for individual covariates and county, the hazard rate of women living in urban counties (with RUCC $= 2$) will be $\exp(0.105 \times 7) \approx 2$ times larger than that of women in rural counties (with RUCC $= 9$).

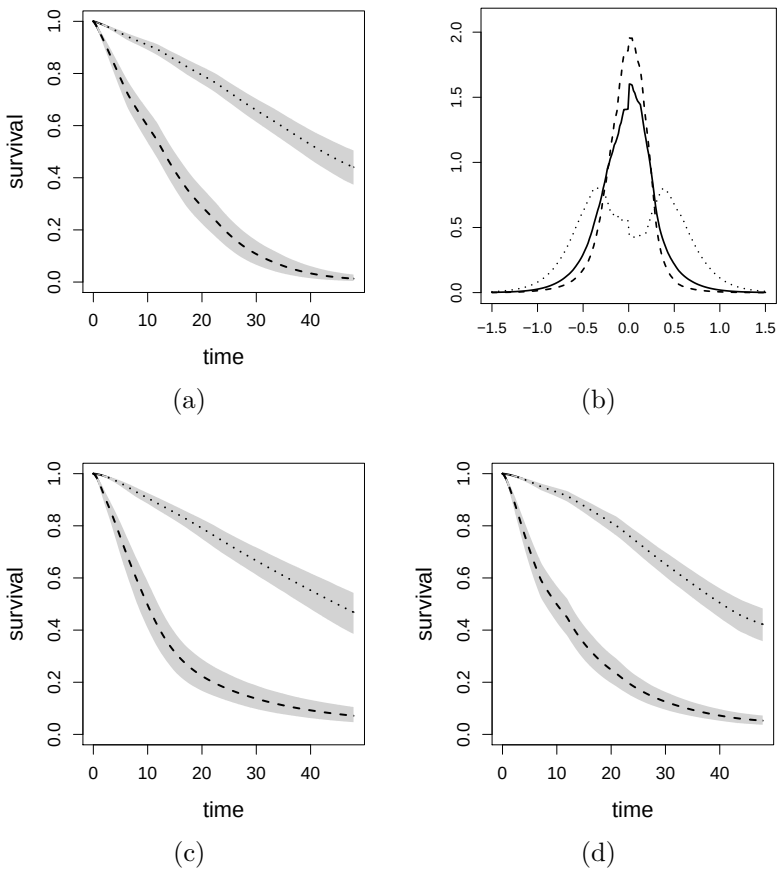


Figure 1.1 SEER breast cancer data. Panels (a), (c), and (d) show estimated survival curves for women aged at 68.8 years and living a county with RUCC at 5 for distant (dashed lines) and local (dotted lines) stages, under PH, AFT, and PO, respectively. The pointwise 95% credible bands are also displayed as grey areas. Panel (b) displays frailty densities for RUCC=2, 5, and 9, which are displayed as dashed, continuous, and dotted lines, respectively.

Leukemia data

We consider a dataset on the survival of acute myeloid leukemia in $n = 1,043$ patients, analyzed by Henderson et al. (2002) fitting a multivariate gamma frailty PH model. This dataset is available for access in Fahrmeir and Kneib (2011). It is of interest to investigate possible spatial variation in survival after accounting for known subject-specific prognostic factors, which include age, sex, white blood cell count (WBC) at diagnosis, and the Townsend score, for which higher values indicate less affluent areas. The censoring rate is 16%. Both exact residential locations of all patients and their

administrative districts (24 districts that make up the whole region) are available. Therefore, we can fit both geostatistical and lattice models.

For the geostatistical case, we fit the copula model (1.19) proposed by Zhou et al. (2015b, Chapter 3) using the function `spCopulaDDP`, where the marginal model $F_{\mathbf{x}}(\cdot)$ is defined via the LDDPM in (1.12) and the copula function C is specified through the Gaussian spatial copula in (1.20) assuming the exponential correlation function. We then use the function `spCopulaCoxph` to fit the copula model assuming a piecewise exponential PH model for $F_{\mathbf{x}}(\cdot)$, where the partition is based on $J = 20$ cut-points with each a_k defined as the $\frac{k}{J}$th quantile the empirical distribution of observed survival times (see Section 1.2). For comparison, standard non-spatial LDDPM and piecewise exponential PH models are also fitted using the functions `anovaDDP` and `indeptCoxph`, respectively. The default priors are considered for above models as suggested in Zhou et al. (2015b, Chapter 3). Regarding the lattice case, we fit each of the Polya trees PH, AFT, and PO models with ICAR frailties as in Section 1.5 using the function `survregbayes` and their corresponding non-frailty models, where the Polya trees are truncated at level $J = 5$. Finally, we fit the generalized AFT model (1.13) with and without ICAR frailties using the function `frailtyGAFT`, where ϵ_i is allowed to depend on age and WBC. We refer readers to Zhou et al. (2015c, Chapter 4) for discussion of prior specifications and posterior samplings. For all models, we retain $10,000$ scans thinned from $50,000$ after a burn-in period of $10,000$ iterations.

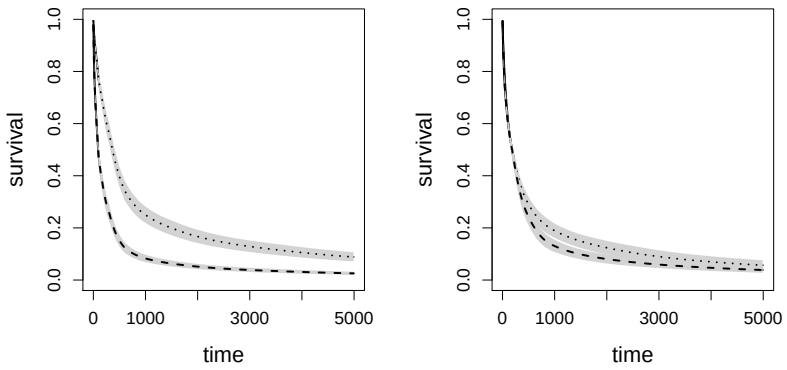
The LPML measures for the copula with LDDPM, copula with piecewise exponential PH, PH, AFT, and PO with Polya trees baselines and ICAR frailties, and generalized AFT with ICAR frailties are -5932, -5939, -5930, -5953, -5925, and -5936, respectively. Without spatial components, the above LPML values become -5934, -5941, -5934, -5950, -5925, and -5942. The PO models significantly outperform others from a predictive point of view regardless of whether spatial dependence is taken into account. Within the context of LDDPM and PH, the use of

the Gaussian spatial copula slightly improves the model's predictive ability, indicating that the spatial dependence is relatively weak in this dataset. Under the framework of PH, the Polya trees prior works much better than piecewise exponential prior for modeling baseline functions. The AFT models provide the worst LPML values, while allowing the baseline varying with covariates (i.e., generalized AFT) can significantly improve the models' predictive ability; the Bayes factors for age and WBC effects on the baseline survival are 124 and 23, respectively under the ICAR frailty model, and are 73 and 31 under the non-frailty model.

For the copula LDDPM model, the posterior median of the nugget effect parameter θ_1 is 0.051 with the 95% CI $(0.000, 0.176)$, indicating that only 5% of the heterogeneity variance is explained by spatial effect on average. The posterior median of θ_2 is 0.831 with the 95% CI $(0.001,\ 3.075)$ indicates that the correlation decays by $1 - e^{-0.831} \approx 56\%$ for every kilometer increase in distance on average. However, given such a small value θ_1, the spatial decay becomes less important. Figure 1.2(a) shows the survival curves under the PO ICAR frailty model for female patients aged at 49 (25%th quantile) and aged at 74 (75%th quantile) holding other covariates at population averages, where we see that higher age is associated with lower survival probability. Figure 1.2(b) shows the baseline survival curves under the generalized AFT ICAR frailty model for female patients aged at 49 and aged at 74 holding WBC at its population average, where we can see that the baseline varies with age which clearly violates the AFT assumption.

1.6 Concluding remarks

We have reviewed commonly-used priors on baseline functions, semiparametric and nonparametric Bayesian survival models, and recent approaches for accommodating spatial dependence, both frailty and copula. Many R packages are discussed for implementation including `DPpackage`, `spBayesSurv`, `R2BayesX`, and `spatsurv`. Two



(a) (b)

Figure 1.2 Leukemia data. Panel (a) shows estimated survival curves for women aged at 49 years (dotted lines) and aged at 75 years (dashed lines), holding other covariates at population averages and frailties at zeros, under the PO model with ICAR frailties. Panel (b) shows estimated baseline survival curves for women aged at 49 years (dotted lines) and aged at 75 years (dashed lines), holding WBC at its population average and frailties at zeros, under the generalized AFT with ICAR frailties. The pointwise 90% credible bands are also displayed as grey areas.

interesting data sets are illustrated, where both analyses show that PO models perform significantly better than all other models we considered including the PH, AFT, and two generalizations of AFT.

Acknowledgement

This work was supported by federal grants 1R03CA165110 and 1R03CA176739-01A1.

Chapter 2

Covariate-adjusted Frailty Proportional Hazards Model[1]

Understanding the factors that explain differences in survival times is an important issue for establishing policies to improve national health systems. Motivated by breast cancer data arising from the Surveillance Epidemiology and End Results program, we propose a covariate-adjusted proportional hazards frailty model for the analysis of clustered right-censored data. Rather than incorporating exchangeable frailties in the linear predictor of commonly-used survival models, we allow the frailty distribution to flexibly change with both continuous and categorical cluster-level covariates and model them using a dependent Bayesian nonparametric model. The resulting process is flexible and easy to fit using an existing R package. The application of the model to our motivating example showed that, contrary to intuition, those diagnosed during a period of time in the 1990s in more rural and less affluent Iowan counties survived breast cancer better. Additional analyses showed the opposite trend for earlier time windows. We conjecture that this anomaly has to be due to increased hormone replacement therapy treatments prescribed to more urban and affluent subpopulations.

[1]This is a reprint of the original article "Zhou, H., Hanson, T., Jara, A., and Zhang, J. (2015). Modeling county level breast cancer survival data using a covariate-adjusted frailty proportional hazards model. *The Annals of Applied Statistics*, 9, 43-68. Reprinted here with permission of the Institute of Mathematical Statistics.

Keywords: Clustered time-to-event data; Proportional hazards model; Spatial; Tailfree process.

2.1 Introduction

Based on data gathered for Iowa State in the Surveillance Epidemiology and End Results (SEER) program of the National Cancer Institute, we assess the effect of potential risk factors for womens' breast cancer. This involves the analysis of clustered time-to-event right-censored data, where event times of patients from the same county of residence are expected to be associated with each other, possibly due to sharing common unobserved characteristics, such as region-specific differences in environments, treatment resources or diagnosis of the patients. As is widely known, taking into account the clustered nature of the data is a must to obtain valid statistical inferences (see, e.g., Therneau and Grambsch, 2000, Chapter 8).

A standard way of modeling clustered survival data is to introduce a common random effect (frailty) into the survival model for each cluster, yielding shared frailty models. "Frailties," termed by Vaupel et al. (1979), were originally introduced to deal with possible heterogeneity due to unobserved covariates and are regarded as unobserved common characteristics for each cluster able to account for the dependence among event times. In the context of the proportional hazards (PH) model, as conventionally implemented, frailties are incorporated into the linear predictor, and the median or mean of the frailty distribution is constrained to be zero to avoid identifiability problems. Conditional on the frailty, the model retains its interpretation in terms of constants of proportionality of the hazards. Survival models with frailties have been extensively used in the statistical literature, especially when the comparison of event times within cluster is of interest.

A common assumption in shared frailty survival models is the one of homogeneity, where the frailties are assumed to be independent and identically distributed (*iid*)

random variables from a parametric or nonparametric distribution (see, e.g., Clayton and Cuzick, 1985; Gustafson, 1997; Qiou et al., 1999; Walker and Mallick, 1997). Although the nonparametric approach provides flexibility in capturing a frailty distribution's variance, skewness, shape and even modality, it essentially assumes that these frailty distributional aspects are the same across all the clusters, which may be restrictive for particular data sets (Noh et al., 2006). For example, in the kidney transplantation study, Liu et al. (2011) argue that the frailty distribution may be affected by some cluster-level covariates, since "*...urban transplant facilities may exhibit more uniform practices than rural transplant hospitals, corresponding to less heterogeneity (smaller variance) for frailties of urban centers...*". Ignoring such heterogeneity can drastically affect the inference for cluster-specific effects and prediction (McCulloch and Neuhaus, 2011).

As the process generating the frailty terms is on its own right of scientific interest, different extensions of the *iid* frailty modeling approach have been considered. Wassell and Moeschberger (1993) studied the impact of interventions in the Framingham Heart Study by introducing a modified gamma frailty with a pairwise covariate-dependent parameter. Yashin and Iachine (1999) considered the dependence between frailty and observed covariates (BMI and smoking) in Danish twins to investigate the heritability of susceptibility to death. Noh et al. (2006) verified frailty distribution heterogeneity in a well-known kidney infection data set by applying a dispersed normal model. Cottone (2008) assumed either Bernoulli or normal distributions for the frailties where the frailty distribution mean or variance depends on cluster-level covariates through specified link functions. Liu et al. (2011) proposed a covariate-dependent positive stable shared frailty model with an application to kidney transplantation data from the Scientific Registry of Transplant Recipients, and demonstrated the heterogeneity in facility performance. Wang and Louis (2004) studied a related approach for binary data that has both conditional and marginal

interpretation using the so-called bridge distribution instead of positive stable.

The previously described model extensions allow for particular and specific aspects of distributional shape to change with cluster-level covariates. However, a more thorough evaluation of the effect of the predictors should account for potential changes in characteristics of the frailty distribution other than just, for example, the location or scale. It is, for instance, useful to examine potential changes in the skewness, symmetry, and multimodality of the frailty distribution. Therefore, a nonparametric formulation that anticipates changes in shape, skew and modality beyond simple location models is of interest.

In this paper, we propose a practicable and general framework for modeling clustered survival data as a function of covariates, based on a predictor-dependent Bayesian nonparametric model for the frailties and the Cox's PH model. Under the proposed approach the frailty distribution flexibly changes with both continuous and categorical cluster-level covariates, thus allowing for full heterogeneity across clusters. We apply this modeling approach to a subset of the SEER county-level breast cancer data consisting of 1073 women diagnosed with malignant breast cancer during 1995-1998. Important patient-level covariates include age at diagnosis, race, county of residence and the stage of the disease. Additional county-level covariates potentially associated with breast cancer survival are also available from census data, including median household income, poverty level, education and a rurality measure. These area-level socioeconomic factors have been discovered to be associated with breast cancer by many researchers (e.g., Sprague et al., 2011). Women living in more affluent or less rural geographic areas tend to survive breast cancer better after a diagnosis than those living in regions with indicators of low socioeconomic status. Moreover, rural counties may present more heterogeneity in access to quality care and screening for breast cancer, leading to more variability for frailties of rural counties (Zhao and Hanson, 2011). This suggests to us that the frailty distribution could be potential-

ly affected by these county-level socioeconomic factors. The results show that the proposed model provides better goodness of fit to the data and is predictively superior to the traditional PH spatial frailty model, as well as helping to piece together a plausible story for the data in terms of the prescribing of hormone replacement therapy.

The paper is organized as follows. In Section 2 we introduce the proposed frailty PH model, including a detailed description of the dependent Bayesian nonparametric model and the Markov chain Monte Carlo (MCMC) implementation of the posterior computations. Section 3 provides a detailed analysis of the motivating data set. Section 4 presents the results of simulation studies to evaluate the performance of the proposed model. Some concluding remarks and a final discussion are given in Section 5.

2.2 Covariate-adjusted frailty proportional hazards model

The modeling approach

Suppose that right-censored survival data $(\mathbf{w}_{ij}, t_{ij}, \delta_{ij})$ are collected for the jth subject of the ith cluster, where $j = 1, \ldots, n_i$, $i = 1, \ldots, n$, $\mathbf{w}_{ij}$ is a p-dimensional vector of exogenous covariates, t_{ij} is the recorded event time, and δ_{ij} is the censoring indicator equaling 1 if t_{ij} is an observed event time and equaling 0 if the event time is right-censored at t_{ij}. Let T_{ij} and C_{ij} be the event and censoring times respectively for the jth subject in the ith cluster. To take into account the within-cluster association structure, a frailty PH model is assumed for T_{ij}. The conditional PH assumption implies that the hazard function of T_{ij} is given by

$$\lambda(t|\mathbf{w}_{ij}, e_i) = \lambda_0(t) \exp(\mathbf{w}'_{ij}\boldsymbol{\xi} + e_i), \tag{2.1}$$

where $\mathbf{e} = (e_1, \ldots, e_n)'$ is an unobserved vector of frailties, and $\lambda_0(t)$ is the baseline hazard function corresponding to the event time of a subject with covariates $\mathbf{w} = \mathbf{0}$

and $e = 0$. We additionally assume a conditionally independent censoring scheme, that is, C_{ij} and T_{ij} are independent given $\mathbf{w}_{ij}$ and e_i. Often the frailties are assumed to be exchangeable or *iid* from some parametric or nonparametric distribution G. For instance, Therneau et al. (2003) considered exchangeable Gaussian frailties and proposed an estimation procedure based on a Laplace approximation of the likelihood function leading to a penalized partial likelihood. This approach, referred to below as GF, will be compared with our method in the simulation studies.

Now consider a partition of the predictor vector $\mathbf{w}_{ij} = (\tilde{\mathbf{w}}'_{ij}, \mathbf{x}'_i)'$, where $\mathbf{x}_i \in \mathcal{X} \subseteq \mathbf{R}^q$ is a q-dimensional vector of cluster-level covariates and $\tilde{\mathbf{w}}_{ij}$ is a $(p-q)$-dimensional vector of subject-specific covariates, respectively, and the corresponding partition of the regression coefficient vector $\boldsymbol{\xi} = (\tilde{\boldsymbol{\xi}}', \boldsymbol{\xi}'_x)'$. On the scale of the linear predictor $\mathbf{w}'_{ij}\boldsymbol{\xi} + e_i$, the frailty e_i models the cluster-specific behavior and its distribution G is shifted by $\mathbf{x}'_i\boldsymbol{\xi}_x$. Therefore, the homogeneity assumption implies that the vector of cluster-level covariates $\mathbf{x}_i$ modifies only the location of the distribution of cluster-specific effects but not its shape. To relax this assumption, we consider a covariate-adjusted frailty PH model, where the frailty distribution depends on cluster-level covariates $\mathbf{x}_i$. That is,

$$e_i \mid G_{\mathbf{x}_i} \overset{ind.}{\sim} G_{\mathbf{x}_i},$$

where for every $\mathbf{x} \in \mathcal{X}$, $G_{\mathbf{x}}$ is a probability measure defined on $\mathbb{R}$; this specifies a probability model for the entire collection of probability measures $\mathcal{G}^{\mathcal{X}} = \{G_{\mathbf{x}} : \mathbf{x} \in \mathcal{X}\}$, such that its elements are allowed to smoothly vary with the cluster-level covariates $\mathbf{x}$. Specifically, we consider a mixture of linear dependent tailfree processes (LDTFP) prior (Jara and Hanson, 2011) for $\mathcal{G}^{\mathcal{X}}$, denoted as

$$\mathcal{G}^{\mathcal{X}} \mid J, h, \theta, c, \rho \sim \text{LDTFP}(h, \Pi^{J,\theta}, \mathcal{A}^{J,c,\rho}),$$

and

$$c \mid Q \sim Q,$$

where $J \in \mathbb{N}$ is the level of specification of the process, $c \in \mathbb{R}_+$ is a prior precision parameter controlling the prior variability of the process, $h(\cdot) = \frac{\exp\{\cdot\}}{1+\exp\{\cdot\}}$, $\Pi^{J,\theta}$ is a J-level sequence of binary partitions of $\mathbb{R}$, depending on the scale parameter $\theta \in \mathbb{R}_+$, $\mathcal{A}^{J,c,\rho} = \{2n/c\rho(1), \ldots, 2n/c\rho(J)\}$ is a collection of positive numbers depending on J, c and ρ, $\rho : \mathbb{N} \longrightarrow \mathbb{R}_+$ is an increasing function, and Q is a probability measure defined on $\mathbb{R}_+$.

The LDTFP is specified such that for every $\mathbf{x} \in \mathcal{X}$, the process $G_{\mathbf{x}}$ is centered around an $N(0, \theta)$ distribution, that is, $E(G_{\mathbf{x}}) = N(0, \theta)$, for every $\mathbf{x} \in \mathcal{X}$. Furthermore, the process is specified such that for every $\mathbf{x} \in \mathcal{X}$, $G_{\mathbf{x}}$ is almost surely a median-zero probability measure. The latter property is important to avoid identifiability problems. The LDTFP process includes as important special cases a nonparametric exchangeable frailty model where $G_{\mathbf{x}} = G_{\mathbf{x}'}$ for $\mathbf{x}' \neq \mathbf{x}$ as well as exchangeable normal frailties $G_{\mathbf{x}} = N(0, \theta)$ for all $\mathbf{x} \in \mathcal{X}$.

As shown by Jara and Hanson (2011), dependent tailfree processes have appealing theoretical properties, such as continuity as a function of the predictors, large support on the space of conditional density functions, straightforward posterior computation relying on algorithms for fitting generalized linear models, and the process closely matches conventional Polya tree priors (see, e.g., Hanson, 2006a) at each value of the predictor, which justify its choice here. Polya trees have been extensively studied in the literature and have desirable properties in terms of support and posterior consistency. Details on the trajectories of LDTFP$(h, \Pi^{J,\theta}, \mathcal{A}^{J,c,\rho})$, useful for a complete implementation of algorithms for exploring the corresponding posterior distributions, are given in Appendix A.1.

Other dependent processes could be considered for $\mathcal{G}^{\mathcal{X}}$, but a highly limiting requirement is that some aspect of the location, for example, mean or median, can be fixed. There are few examples where the process changes smoothly with covariates; one is the multivariate beta process of Trippa et al. (2011). Another approach using

Dirichlet process mixtures can be found in Reich et al. (2010), but this latter approach would have to be extended to allow the means or variances of the two mixture components to change with covariates.

Posterior computation

The conditional likelihood for $(\boldsymbol{\xi}, \lambda_0, \mathbf{e})$ is given by

$$\mathcal{L}(\boldsymbol{\xi}, \lambda_0, \mathbf{e}) = \prod_{i=1}^{n} \prod_{j=1}^{n_i} \left[\lambda_0(t_{ij}) \exp(\mathbf{w}_{ij}'\boldsymbol{\xi} + e_i)\right]^{\delta_{ij}} \exp\left\{-\Lambda_0(t_{ij}) \exp(\mathbf{w}_{ij}'\boldsymbol{\xi} + e_i)\right\},$$

where $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$ is the cumulative hazard function. The piecewise exponential model provides a flexible framework to deal with the baseline hazard (see, e.g., Walker and Mallick, 1997). We partition the time period $\mathbb{R}_+$ into K prespecified intervals, say $I_k = (a_{k-1}, a_k], k = 1, \ldots, K$, where $a_0 = 0$ and $a_K = \max\{t_{ij}\}$. The baseline hazard is assumed to be constant within each interval, that is,

$$\lambda_0(t) = \sum_{k=1}^{K} \lambda_k I\{t \in I_k\},$$

where $\lambda_1, \ldots, \lambda_K$ are unknown hazard values and $I\{A\}$ is the usual indicator function, that is, 1 when A is true, 0 otherwise. The prior hazard is specified by the hazard values $\{\lambda_k\}_{k=1}^K$ and cut-point vector $\mathbf{a} = (a_1, \ldots, a_K)$. If the prior on the λ_ks is taken to be independent gamma distributions and $\{I_k\}_{k=1}^K$ is a reasonably fine mesh, the gamma process (Kalbfleisch, 1978) is approximated. To determine the cut-point vector $\mathbf{a}$, one can set a_k to be the $\frac{k}{K}$th quantile of the empirical distribution of the t_{ij}s, or choose them based on other considerations (See Section 2.3). Some authors have considered random cut-points (see, e.g., Sahu and Dey, 2004). Regardless, the resulting model implies a Poisson likelihood (Laird and Olivier, 1981) as follows. Let $K(t) = \min\{k : a_k \geq t\}$, $\Delta_k(t) = \min\{a_k, t\} - a_{k-1}$, and $y_{ijk} = \delta_{ij} I\{k = K(t_{ij})\}$. Set $\mathbf{z}_{ijk} = (\boldsymbol{\iota}_k', \mathbf{w}_{ij}')'$ and $\boldsymbol{\gamma} = (\boldsymbol{\lambda}', \boldsymbol{\xi}')'$, where $\boldsymbol{\iota}_k$ is a K-dimensional vector of zeros except the kth element is 1 and $\boldsymbol{\lambda} = (\log(\lambda_1), \ldots, \log(\lambda_K))'$. Then the likelihood for $(\boldsymbol{\gamma}, \mathbf{e})$

becomes

$$\begin{aligned}\mathcal{L}(\boldsymbol{\gamma},\mathbf{e}) &= \prod_{i=1}^{n}\prod_{j=1}^{n_i}\left[\exp\left\{\log(\lambda_{K(t_{ij})})+\mathbf{w}_{ij}'\xi+e_i\right\}\right]^{\delta_{ij}}\left[\prod_{k=1}^{K(t_{ij})} e^{-\exp\left\{\log(\lambda_k)+\mathbf{w}_{ij}'\xi+e_i\right\}\Delta_k(t_{ij})}\right]\\ &= \prod_{i=1}^{n}\prod_{j=1}^{n_i}\prod_{k=1}^{K(t_{ij})}\left[\left(\exp\left\{\mathbf{z}_{ijk}'\boldsymbol{\gamma}+e_i\right\}\right)^{y_{ijk}} e^{-\exp\left\{\mathbf{z}_{ijk}'\boldsymbol{\gamma}+e_i+\log(\Delta_k(t_{ij}))\right\}}\right]\\ &\propto \prod_{i=1}^{n}\prod_{j=1}^{n_i}\prod_{k=1}^{K(t_{ij})} p(y_{ijk}|\boldsymbol{\gamma},e_i),\end{aligned}$$

where $\mu_{ijk}=\exp\left\{\mathbf{z}_{ijk}'\boldsymbol{\gamma}+e_i+\log(\Delta_k(t_{ij}))\right\}$ and $p(y_{ijk}|\boldsymbol{\gamma},e_i)$ is the probability mass function for a Poisson distribution with mean μ_{ijk}. For each $i=1,\ldots,n$, let $N_i=\sum_{j=1}^{n_i}K(t_{ij})$, $\mathbf{y}_i=(y_{ijk})$ be an $N_i\times 1$ vector with subscript ijk in lexicographical order.

Thus the proposed covariate-adjusted frailty PH model takes the following hierarchical structure:

$$\begin{aligned}\mathbf{y}_i|\boldsymbol{\gamma},e_i &\overset{ind}{\sim} \prod_{j=1}^{n_i}\prod_{k=1}^{K(t_{ij})} p(y_{ijk}|\boldsymbol{\gamma},e_i),\\ \boldsymbol{\gamma} &\sim N_{K+p}(\boldsymbol{\gamma}_0,\mathbf{S}_0),\\ e_i|G_{\mathbf{x}_i} &\overset{ind.}{\sim} G_{\mathbf{x}_i},\\ \mathcal{G}^{\mathcal{X}} \mid J,h,\theta,c,\rho &\sim \text{LDTFP}(h,\Pi^{J,\theta},\mathcal{A}^{J,c,\rho}),\\ \theta^{-2} &\sim \Gamma(\tau_1,\tau_2),\ c\sim\Gamma(a_c,b_c),\end{aligned}$$

which largely simplifies computations, where $N_p(\mathbf{m},\mathbf{S})$ refers to a p-variate normal distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{S}$. This forms the basis of an efficient Markov chain Monte Carlo (MCMC) scheme for obtaining posterior inference, which can be implemented using available software for generalized linear mixed models. A full description of the MCMC algorithm is given in Appendix A.2. Sample R code using the `LDTFPglmm` function available in `DPpackage` (Jara et al., 2011) is provided in Appendix A.3.

Time-dependent subject-specific covariates that are step-processes (Hanson et al., 2009) are naturally accommodated by including the times where the covariate values

change across all subjects into the cut-point vector $\mathbf{a}$. All that is changed above is $\mathbf{z}_{ijk} = (\boldsymbol{\iota}_k', \mathbf{w}_{ijk}')'$, that is, $\mathbf{w}_{ij}$ is replaced with its time-varying analogue $\mathbf{w}_{ijk}$. Similarly, time-varying regression effects can be included by replacing $\mathbf{z}_{ijk}'\boldsymbol{\gamma}$ with $\mathbf{z}_{ijk}'\boldsymbol{\gamma}_k$ in μ_{ijk}, yielding very general models. The proposed model implies exchangeable frailties for each subgroup with a unique $\mathbf{x} \in \mathcal{X}$. Time-dependent cluster-specific covariates are therefore naturally included in the model by simply allowing $\mathbf{x}$ to change with time. For example, in the SEER data set analyzed over a larger time window, for subjects living in the ith county, one could include into $\mathbf{x}_i$ the median house income of that county at their particular diagnosis year. Furthermore, the frailty distribution can itself evolve in time by simply including time as a covariate in $\mathbf{x}$, or a time-by-cluster covariate interaction could also be entertained.

2.3 Analysis of SEER county level breast cancer data

The Iowa SEER data

The SEER program of the National Cancer Institute (see `http://seer.cancer.gov/`) is an authoritative source of information on cancer incidence and survival in the US, providing county-level cancer data on an annual basis for particular states for public use. We fit our proposed covariate-adjusted frailty Cox's PH model to a subset of the Iowa SEER breast cancer survival data, which consists of a cohort of 1073 women from the 99 counties of Iowa, who were diagnosed with malignant breast cancer in 1995, with enrollment and follow-up continued through the end of 1998. The observed survival time, from 1 to 48, was calculated as the number of months from diagnosis to either death or the last follow-up. In our analysis, only deaths due to metastasis of cancerous nodes in the breast were considered to be events, while the deaths from other causes were censored at the time of death. That is, we consider cause-specific survival models assuming that all other deaths are independent of breast

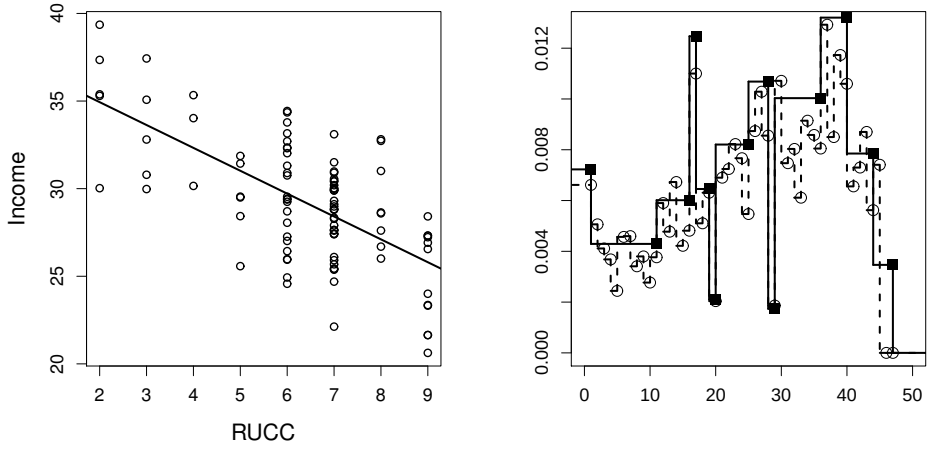
cancer. By the end of 1998, a total of 488 patients (45.5%) had died of breast cancer, while the remaining 585 patients were censored, either because they died of other causes or survived until the last follow-up.

For each patient, the observed survival time and county of residence at diagnosis are recorded. The data set also has individual-level covariates including age at diagnosis and the stage of the breast cancer: local (confined to the breast), regional (spread beyond the breast tissue), or distant (metastasis). We create two dummy variables for regional and distant, respectively, and treat the patients in the local group as the baseline. Although several individual-level covariates that affect breast cancer survival are not available (e.g., age at first child, age at menopause and breastfeeding), we are able to obtain county-level covariates potentially associated with breast cancer survival from census data, such as median household income (small area estimates in 1993), poverty level (percentage of families in poverty in 1990), education (percentage with Bachelor's degree or higher in 1990) and rurality (Rural-Urban Continuum Codes in 1993). The Economic Research Service Rural-Urban Continuum Codes (RUCC) vary from 1 to 9 (see `www.ers.usda.gov/data-_products/rural-_urban-_continuum-_codes`), distinguishing metropolitan counties by the population size of their metro area and nonmetropolitan counties by degree of urbanization and adjacency to a metro area. Higher RUCC indicates a more rural county. Other county-level covariates mentioned above are available at `data.iowadatacenter.org/browse/counties.html`. Since the effects of education and poverty on the survival times are not significant based on our initial model fitting by the proposed method, we exclude them in the analysis presented below. Thus, we have three-dimensional $\tilde{\mathbf{w}}_{ij}$ and two-dimensional $\mathbf{x}_i$. Table 2.1 presents several summary statistics for the data. As shown in Figure 2.1, median household income and RUCC are significantly, negatively correlated.

To get an initial feeling about the role that each county-level covariate is playing,

Table 2.1 Iowa SEER data: Summary statistics for follow-up times and both individual- and county-level covariates

Continuous variables	Minimum	Median	Maximum
follow-up time in months	1	19	47
Age in years	26	72	103
RUCC	2	7	9
Income (×1000)	20.627	29.110	39.356
Categorical variables	**Level**	**Count**	**Proportion (%)**
Status	Event	488	45.5
	Censored	585	54.5
Stage	Local	510	47.5
	Regional	355	33.1
	Distant	208	19.4



(a) (b)

Figure 2.1 Iowa SEER data: Panel (a) shows the scatter plot and simple linear regression line by regressing median household income on RUCC. Panel (b) shows the baseline hazards for **Model 1**. The dashed line corresponds to Breslow's estimate of $\lambda_0(t)$ obtained by the GF approach, where the circles represent the hazard values at each month; the solid line is the fitted baseline hazard by our approach, where the solid squares correspond to the cut-point values $\mathbf{a} = (1, 11, 16, 17, 19, 20, 25, 28, 29, 36, 40, 44, 47)$.

Table 2.2 provides the distribution of each county-level covariate stratified by the individual-level stage of disease. The gamma statistic (GK), originally proposed by Goodman and Kruskal (1954), is calculated to quantify the association between each county-level covariate and the stage of disease. The GK values range from -1 (100%

negative association) to 1 (100% positive association), where the value 0 indicates no association. We see that women with a distant-stage at diagnosis are much more likely than those with a local-stage to live in counties with a high degree of urbanization (GK = −0.11; 95% CI: from −0.20 to −0.01), while the association between stage and income is not significant (GK = 0.04; 95% CI: from −0.06 to 0.13). These associations roughly imply that women living in urban counties may suffer poorer survival, assuming that women in distant-stage are more likely to die than women in other stages. Next, we carefully examine both these individual-level and county-level covariates in relation to breast cancer survival, fitting the covariate-adjusted frailty proportional hazards model.

Table 2.2 Iowa SEER data: Distribution of each county-level covariate stratified by individual-level stage. The pattern of numbers is Number of women (%). Goodman and Kruskal's gamma statistics (95% confidence intervals) are −0.11 (−0.20, −0.01) and 0.04 (−0.06, 0.13) for RUCC and Income respectively.

		Stage		
Covariates	**All Women N=1073**	**Local N=510**	**Regional N=355**	**Distant N=208**
RUCC				
1-3	314 (29.3)	131 (25.7)	99 (27.9)	84 (40.4)
4-7	666 (62.1)	342 (67.1)	221 (62.3)	103 (49.5)
8-9	93 (8.6)	37 (7.2)	35 (9.8)	21 (10.1)
Income (×1000)				
20-27	163 (15.2)	79 (15.5)	51 (14.4)	33 (15.9)
27-34	651 (60.7)	312 (61.2)	223 (62.8)	116 (55.8)
> 34	259 (24.1)	119 (23.3)	81 (22.8)	59 (28.3)

Models and model comparison

We fitted the proposed covariate-adjusted frailty PH model for the Iowa SEER data with different county-level covariates, including models with RUCC only (**Model 1**), with median household income only (**Model 2**) and with both (**Model 3**). To see how the piecewise assumption of baseline hazard affects the predictive ability of

models, we considered three specifications of cut-point vector $\mathbf{a}$ as follows:

Case I: $\mathbf{a} = (1, 11, 16, 17, 19, 20, 25, 28, 29, 36, 40, 44, 47)$, which was determined by visually examining Breslow's estimate of $\lambda_0(t)$ using the GF approach, which is given in Panel (b) of Figure 2.1.

Case II: $\mathbf{a} = (3, 7, 12, 16, 19, 24, 29, 34, 41, 47)$, where a_k is the $\frac{k}{10}$th quantile of the empirical distribution of observed survival times.

Case III: $\mathbf{a} = (47)$, which yields an exponential baseline hazard.

In all cases, we set $J = 4$. We fitted all the models using the corresponding variants of the algorithm described in Appendix A.2 and similar prior specifications suggested in the simulation study. The Markov chain mixed reasonably well despite the high dimension of our models. For each version of our model and case, we ran a single Markov chain of $1,020,000$. A total number of $20,000$ were discarded as burn-in period and $10,000$ samples were retained for posterior inference. Moreover, we also considered another **Case II** with 13 cut-points and cut-point specifications based on the event time quantiles from the Kaplan-Meier curve in Appendix A.5. The results show that carefully choosing the cut-points is more important than simply increasing the number of cut-points.

For the sake of comparison, we further fitted the exchangeable MPT frailty Cox model and the Bayesian exchangeable Gaussian frailty Cox model. We compare the models using the log pseudo marginal likelihood (LPML) developed by Geisser and Eddy (1979) and the deviance information criterion (DIC) proposed by Spiegelhalter et al. (2002). In the context of the frailty Cox model, the LPML for model M is defined as LPML $= \sum_{i=1}^{n} \sum_{j=1}^{n_i} \log(\text{CPO}_{ij})$, where CPO_{ij}, the ijth conditional predictive ordinate, is given by $[\lambda(t_{ij})^{\delta_{ij}} e^{-\Lambda(t_{ij})} | \mathcal{D}_{(ij)}]$ with $\mathcal{D}_{(ij)}$ denoting the remaining data after excluding the ijth data point $\mathcal{D}_{ij}$. One can use the simple method suggested by Gelfand and Dey (1994) to estimate the CPO statistics from MCMC output.

A larger value of LPML indicates the corresponding model has better predictive ability. Furthermore, Geisser and Eddy (1979) discussed the exponentiated difference in LPML values from two models to obtain what they termed as a pseudo Bayes factor (PBF). The PBF is a surrogate for the more traditional Bayes factor, and can be interpreted similarly, but is more analytically tractable, much less sensitive to prior assumptions, and does not suffer from Lindley's paradox. Set $\mathbf{\Omega} = (\mathbf{e}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \theta)$ as the entire collection of model parameters. The DIC for model M is defined as $\text{DIC} = \bar{D} + p_D = E_{\mathbf{\Omega}|\mathcal{D}}\{D(\mathbf{\Omega})\} + p_D$, where $D(\mathbf{\Omega}) = -2\log\mathcal{L}(\boldsymbol{\gamma}, \mathbf{e})$ which is referred to as the deviance function, and $p_D = \bar{D} - D\left(E_{\mathbf{\Omega}|\mathcal{D}}\{\mathbf{\Omega}\}\right)$ which is a measure of model complexity. Note that the DIC is also readily computed from MCMC output.

Results

Table 2.3 shows the DIC and LPML for all models under consideration. All models under **Case I** provide significantly better prediction as measured by both DIC and LPML, with differences in the range of 20–55 for DIC and 10–25 for LPML, which indicates that the determination of the cut-point vector for the baseline hazard plays an important role on model prediction and fit. Comparing the frailty specifications in **Model 1** across all cases, the DIC and LPML show the same trend for goodness of fit, with the proposed model based on the LDTFP frailty model outperforming both the MPT and Gaussian models, although the differences are only in the range of 1–4. Comparing between **Model 2** and **Model 3**, the proposed model is always preferred in terms of LPML, while the MPT model is slightly better than others in term of DIC under **Model 2**. Comparing all the proposed models across **Model 1**–**Model 3**, the results indicate that **Model 1** always performs best. Overall, allowing the frailty distribution to change with county-level covariates (especially RUCC) does improve model prediction according to LPML. In what follows, we present the results under **Case I**.

Table 2.3 Iowa SEER data: Deviance information criteria (DIC) and log of the pseudo marginal likelihood (LPML) for models under consideration.

		Case I		**Case II**		**Case III**	
Model	Frailty	DIC	LPML	DIC	LPML	DIC	LPML
1	LDTFP	4436	-2222	4463	-2234	4495	-2247
	MPT	4441	-2225	4463	-2235	4496	-2248
	Gaussian	4444	-2225	4467	-2236	4497	-2248
2	LDTFP	4441	-2224	4465	-2235	4498	-2249
	MPT	4440	-2225	4462	-2236	4497	-2248
	Gaussian	4443	-2225	4465	-2235	4498	-2249
3	LDTFP	4438	-2223	4464	-2235	4496	-2248
	MPT	4441	-2225	4464	-2235	4498	-2249
	Gaussian	4445	-2226	4467	-2236	4498	-2248

Table 2.4 presents posterior medians and equal-tailed 95% credible intervals (CI) for main effects (components of $\boldsymbol{\xi}$) under **Model 1** – **Model 3**, with covariate-adjusted frailties, and compares the individual-level covariate effects, that is, (ξ_1, ξ_2, ξ_3), to those obtained by Zhao et al. (2009), under the standard nonfrailty Cox model and the Cox frailty model that has a MPT prior for the baseline survival, centered at the log-logistic family, and with conditionally autoregressive (CAR) county-level spatial frailties. The best fitting Cox model reported by Zhao et al. (2009) has an LPML of -2226. Therefore, the pseudo Bayes factor for the proposed model versus the CAR model is $e^{2226-2222} \approx 55$, implying that the proposed model predicts about 55 times better than the model with CAR frailties. In addition, the proposed model offers a unique interpretation. The posterior medians and 95% CIs for all individual-level effects change little across the different versions of the proposed model, indicating that the Cox regression estimates are reasonably stable for these data, except for the estimate of "Regional stage," for which the CAR model 95% CI is much wider than those under the considered versions of the proposed model. This may be partly due to the benefit of including county-level covariates. The best model according to LPML, **Model 1**, indicates that all the individual-level effects are significant at the 0.05 level. Higher age at diagnosis increases the hazard within each county. For in-

stance, women are about $e^{0.019\times 20} \approx 1.46$ times more likely to die from breast cancer than those twenty years younger who have the same disease stage and live in the same county. Compared with women having local stage of disease, women of the same age and living in the same county are $e^{0.27} \approx 1.31$ times more likely to die if their cancer is detected at the regional stage, and $e^{1.64} \approx 5.16$ times more likely to die if detected at the distant stage. We additionally present the fixed effects under the marginal PH model (i.e., using the R function `coxph` with option `cluster`) across **Model 1**–**Model 3** in Appendix A.5. Note that the coefficient estimates under the marginal PH model have population-averaged interpretations and cannot be directly compared with those fitted from the proposed frailty PH model due to different model structures.

Table 2.4 Iowa SEER data: Posterior medians (95% credible intervals) of fixed effects $\boldsymbol{\xi}$ from various models.

Predictor	**Model 1**	**Model 2**	**Model 3**	CAR	Cox
ξ_1 (Age)	0.019	0.020	0.020	0.018	0.019
	(0.013, 0.025)	(0.014, 0.026)	(0.014, 0.026)	(0.012, 0.025)	(0.013, 0.025)
ξ_2 (Regional)	0.27	0.27	0.27	0.22	0.30
	(0.03, 0.49)	(0.03, 0.47)	(0.05, 0.50)	(0.01, 0.49)	(0.08, 0.52)
ξ_3 (Distant)	1.64	1.67	1.65	1.65	1.64
	(1.43, 1.88)	(1.43, 1.89)	(1.43, 1.89)	(1.40, 1.93)	(1.42, 1.87)
ξ_{x_1} (RUCC)	-0.105		-0.082		
	(-0.185, -0.041)		(-0.179, 0.011)		
ξ_{x_2} (Income)		0.042	0.011		
		(0.003, 0.084)	(-0.040, 0.066)		

Regarding the effect of county-level covariates, living in a higher median household income or urban counties is associated with poorer survival after a breast cancer diagnosis. For example, the results under **Model 1** indicate that after controlling for individual covariates and county, the hazard rate of women living in urban counties (with RUCC = 2) will be $e^{0.105\times 7} \approx 2$ times larger than that of women in rural counties (with RUCC = 9). The results under **Model 2** imply that after controlling for individual covariates and frailties, women have about a 1.7 times larger hazard rate if they live in median household income counties of $\$35,301$ compared with median

household income of $\$23,354$ (see also Figure 2.2). Under **Model 3**, the results indicate that when both the county-level covariates are included simultaneously, their independent effects are attenuated, partly due to the multicollinearity between them (see the middle two plots in Figure 2.3).

We obtain the fitted predictive frailty densities for both $\mathbf{e}_i$ (median-zero) and $\mathbf{e}_i + \mathbf{x}_i'\boldsymbol{\xi}_x$ (full distribution) and survival curves for women with mean entry age 68.8 years and distant stage of disease who live in the counties with different levels of median household income or RUCC, under the different versions of the proposed model. The three levels are chosen from the 5%, 50% and 95% quantiles of each covariate value. The results are reported in Figures 2.2 and 2.3. Under our best fitting, **Model 1** (see left three plots in Figure 2.2), the results indicate that higher values of RUCC increase the frailty variance and suggest a non-Gaussian shape (upper); we also see overall higher frailty after mixing over the location shift $\mathbf{x}_i'\boldsymbol{\xi}_x$ (middle) and so poorer survival (lower) in urban counties. Increasing heterogeneity as ruralness increases under **Model 1** translates into increasing association among those living in more rural counties versus urban. In Appendix A.5, Kendall's tau is computed and plotted as a function of RUCC for individuals with mean entry age 68.8 years and distant stage. Kendall's tau increases *by a factor of three* as RUCC goes from 2 to 9. Note that under a traditional gamma frailty model the association is static.

Under **Model 2**, the frailty densities only slightly change compared with **Model 1**, but we do see poorer survival in counties with higher median household income. Figure 2.3 demonstrates that after adjusting individual covariates and median household income (right three plots), there is little effect of RUCC on either predictive frailty densities or survival curves; while after adjusting for RUCC (left three plots), the effect of median household income is almost negligible. In Appendix A.5, the survival curves are also compared with those obtained under the marginal PH model. Overall, the marginal PH model under-predicts survival time up to about 1 month,

for example, it gives estimates of median survival a month less, compared with our proposed model for patients with mean entry age 68.8 years and distant stage of disease who live in the same county. This may be partly due to the fact that the marginal PH model averages over the changing behavior of the frailty distribution over the ruralness measure.

It is widely known that access to quality care and screening for breast cancer is more readily available to those with greater financial means and/or those living in urban areas. Therefore, our findings of increased survival for poorer and more rural counties for this cohort are initially puzzling. However, hormone replacement therapy (HRT) increased about 150% in the 1990s (Wysowski and Governale, 2005), after several observational studies linked HRT to prevention of osteoporosis and protection from heart disease. However, this increasing use of HRT abated suddenly in 2002, when the Women's Health Initiative clinical trial linked HRT to aggressively invasive breast cancer (Rossouw et al., 2002). In fact, overall breast cancer incidence rates peaked in 1999. Between 2001 and 2004 overall invasive breast cancer incidence declined, but fell much more drastically among women living in urban versus rural counties, and among women living in low-poverty versus high-poverty counties. Hausauer et al. (2009) attribute this discrepancy to greater use of postmenopausal estrogen/progestin hormone replacement therapy among more affluent women and/or women living in urban counties up until 2002, when the Women's Health Initiative trial was stopped prematurely on May 31, 2002, according to Rossouw et al. (2002), "*...because the test statistic for invasive breast cancer exceeded the stopping boundary for this adverse effect and the global index statistics supported risks exceeding benefits.*" It is plausible that increased risk (i.e., stochastically larger frailties) in more affluent and more urban counties has to do with a larger proportion of women being prescribed HRT in the late 1980s and 1990s. Further exploratory analyses on other cohorts of SEER Iowan breast cancer data (1975–'79, '80–'84, '85–'89, and '90–'94)

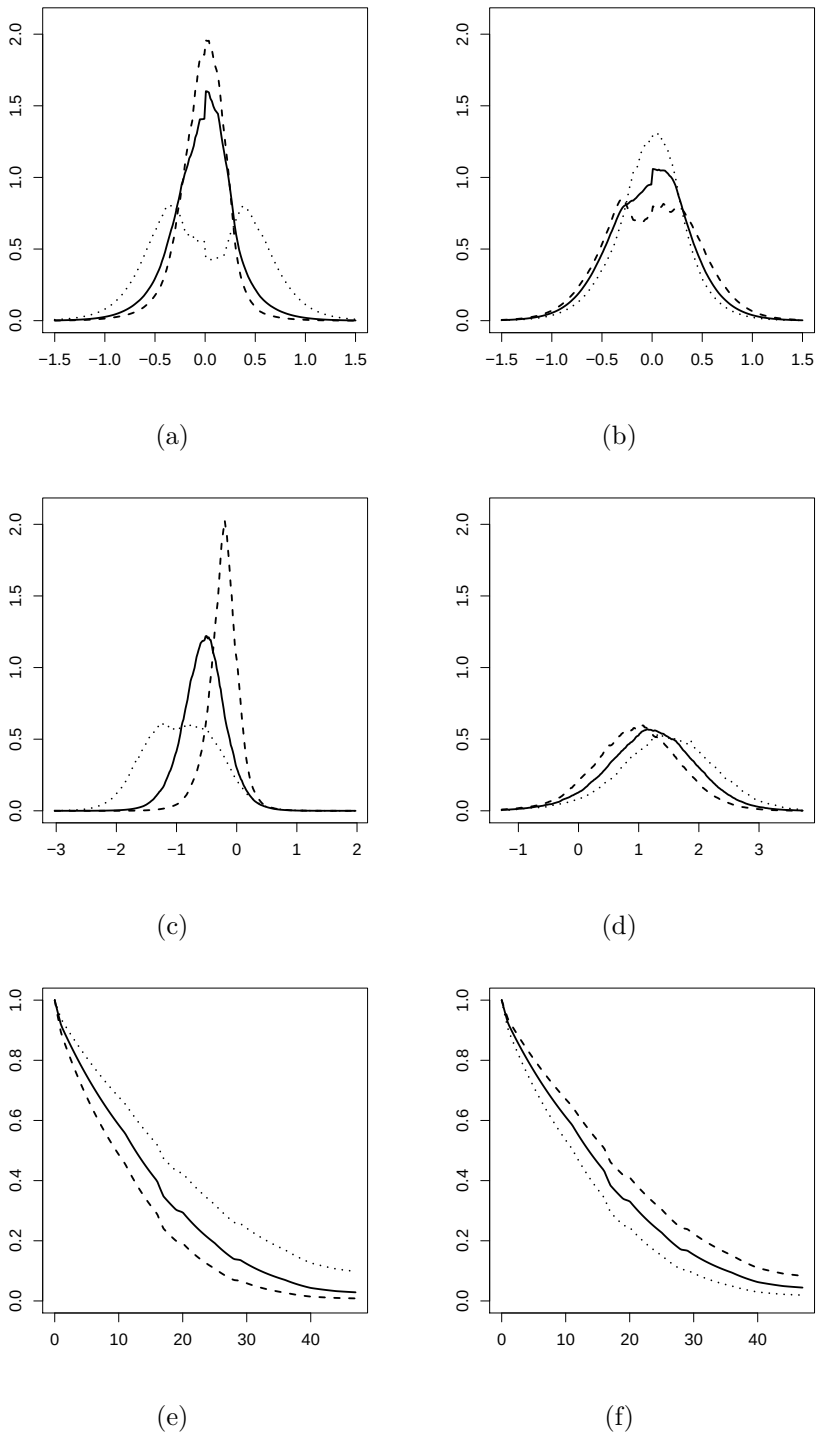


Figure 2.2 Iowa SEER data: Fitted predictive frailty densities [Panels (a) and (b)], frailty densities with location shifts [Panels (c) and (d)] and survival curves [Panels (e) and (f)] for women with mean entry age 68.8 years and distant stage of disease from different county covariate levels under Model 1 [Panels (a), (c) and (e)] and Model 2 [Panels (b), (d) and (f)]. In Panels (a), (c) and (e), the results for RUCC=2, 5 and 9 are displayed as dashed, continuous and dotted lines, respectively. In Panels (b), (d) and (f), the results for Income=23.354, 29.176 and 35.301 are displayed as dashed, continuous and dotted lines, respectively.

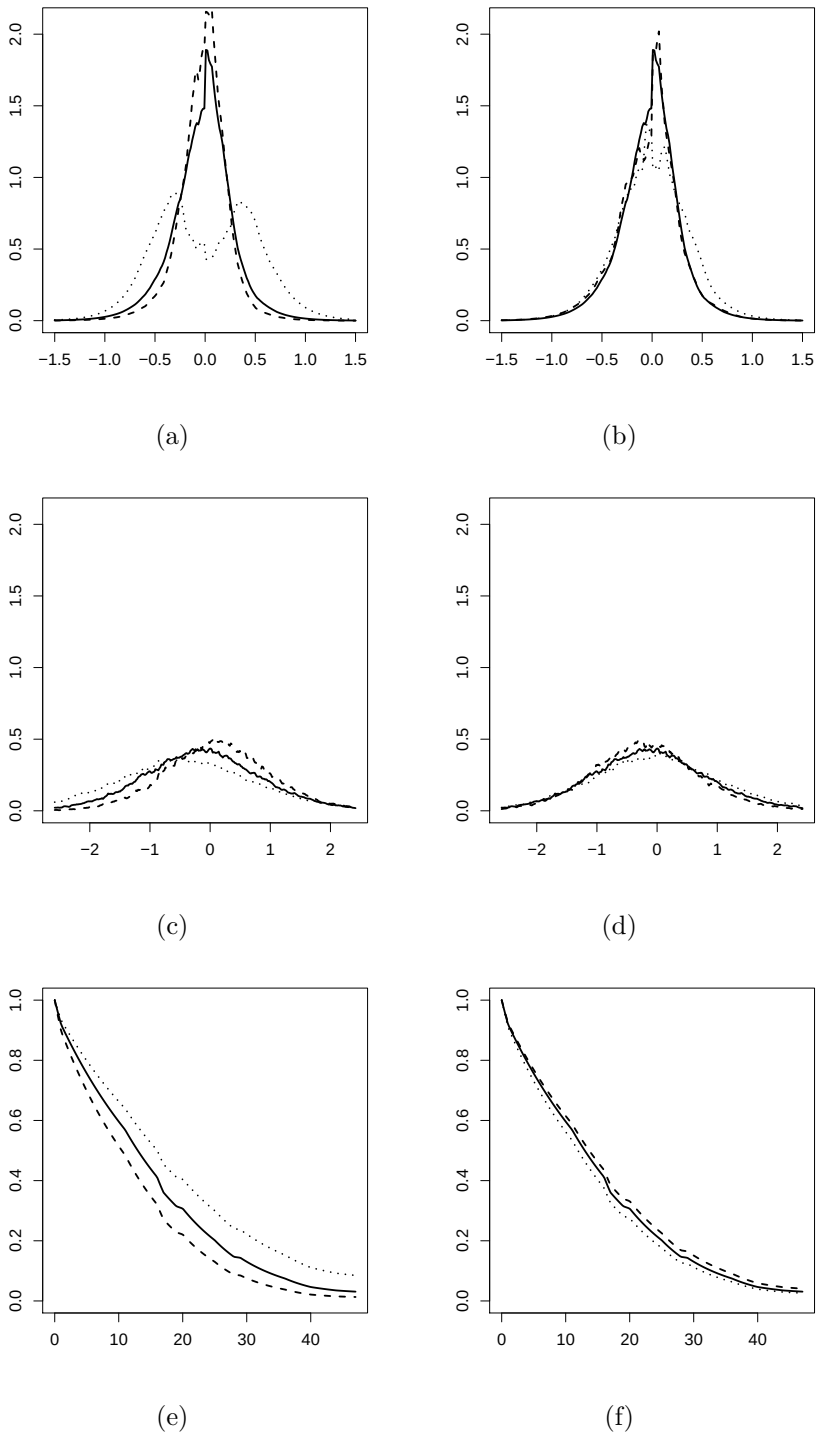


Figure 2.3 Iowa SEER data: Fitted predictive frailty densities [Panels (a) and (b)], frailty densities with location shifts [Panels (c) and (d)] and survival curves [Panels (e) and (f)] for women with mean entry age 68.8 years and distant stage of disease from different county covariate levels under Model 3. In Panels (a), (c) and (e), the results for RUCC=2, 5 and 9 are displayed as dashed, continuous and dotted lines, respectively. In Panels (b), (d) and (f), the results for Income=23.354, 29.176 and 35.301 are displayed as dashed, continuous and dotted lines, respectively.

show a reversal of the effects of income and ruralness, agreeing with intuition. Paralleling our study, Krieger et al. (2010) used county-level census data on income and found rising and falling breast cancer incidence rates for the SEER data over the range 1992–2005 for caucasian women living in high-income counties, which "*mirrored the social patterning of hormone therapy use.*"

In a longer follow-up study of the Women's Health Initiative trial, Chlebowski et al. (2010) found that those on estrogen plus progestin compared to placebo had about 25% higher incidence of invasive breast cancer. Among those diagnosed with breast cancer, the two treatment arms had similar histology, but the estrogen plus progestin group were 78% more likely to have cancers that had spread to lymph nodes than placebo, and the estrogen plus progestin group were about twice as likely to die from breast cancer versus placebo. It would appear that hormone replacement therapy fortified the the virulence of breast cancer, significantly increasing both incidence and mortality. This same study showed an impressive 7% one-year drop in incidence right after the Women's Health Initiative study was prematurely stopped and the medical community warned of a possible link between hormone replacement therapy and breast cancer.

2.4 Simulation studies

We performed a simulation study to assess the performance of the proposed model. The simulated data are also used to compare the proposed approach with existing models. Specifically, we consider the GF approach described in Section 2.1 and the positive stable frailty Cox model proposed by Liu et al. (2011). Under this latter model, the shape parameter is allowed to depend on cluster-level covariates. In terms of our notation, they assumed that the conditional hazard function of T_{ij} is

$$\lambda(t|\tilde{\mathbf{w}}_{ij}, \mathbf{x}_i, e_i) = \lambda_{0i}(t) \exp(\tilde{\mathbf{w}}'_{ij}\tilde{\boldsymbol{\xi}}_i + e_i), \tag{2.2}$$

where the baseline hazard functions $\lambda_{0i}(t)$ and regression parameters $\tilde{\boldsymbol{\xi}}_i$ are cluster-specific, and $\exp(e_i)$ follows a positive stable distribution with shape parameter $\alpha_i \in (0, 1)$, relying on the cluster-level covariates vector $\mathbf{x}_i$ through a logit link function, denoted by $PS(\alpha_i)$. They did not deal with this model directly, but rather derived the marginal model

$$\lambda(t|\tilde{\mathbf{w}}_{ij}, \mathbf{x}_i) = h_0(t) \exp(\tilde{\mathbf{w}}'_{ij}\boldsymbol{\eta}) \tag{2.3}$$

by imposing the restrictions $\boldsymbol{\eta} = \alpha_i \tilde{\boldsymbol{\xi}}_i$, $H_0(t) = \{\Lambda_{0i}(t)\}^{\alpha_i}$, where $H_0(t) = \int_0^t h_0(s)ds$ and $\Lambda_{0i} = \int_0^t \lambda_{0i}(s)ds$. In other words, they essentially fitted the above marginal Cox model by maximizing the pseudo partial likelihood under the working independence assumption (Wei et al., 1989), and then utilized the imposed constraints to estimate the parameters in the frailty model. Although they considered a more flexible conditional Cox model, they made many assumptions to get the marginal model, some of which are difficult to check in practice. Moreover, they faced a nonidentifiability problem when a cluster-level covariate was included in the conditional Cox model, so cluster-level covariates had to be excluded from the marginal model as well, leading to potentially poorer prediction of the marginal survival function. Their method, referred to below as PSF, will be compared with our approach focusing on the prediction of survival functions in the simulation studies. A comparison of the two methods for the fixed effect estimates cannot be conducted, since they have different model structures. We conducted the simulation study in `R`. The GF and PSF approaches were implemented by using the function `coxme` and `coxph` (with the option `cluster`), respectively, included in the R packages `coxme` and `survival`.

Simulation Settings

Two scenarios for the frailty distributions were considered. In the first case, referred to as **Scenario I**, a covariate-dependent family of distributions is considered, where the density shape evolves from one mode to two as the cluster-specific covariate x

increases its value; this mirrors the effect of RUCC in panel (a) of Figure 2.3 for **Model 1**. In the second case, referred to as **Scenario II**, a covariate-dependent positive stable distribution is considered. The specific distributional forms for each setting were the following:

Scenario I: $e_i|x_i \overset{ind.}{\sim} 0.5N(-e^{0.4x_i}, 1) + 0.5N(e^{0.4x_i}, 1)$, $x_i \overset{iid}{\sim} U(-3, 3)$.

Scenario II: $\exp(e_i)|x_i \overset{ind.}{\sim} PS(\alpha_i)$, $\alpha_i = 1/(1 + e^{-0.5-0.5x_i})$, $x_i \overset{iid}{\sim} U(0, 2)$.

Note that the first setting is not a particular case of the proposed model; the second setting, chosen from the simulation study of Liu et al. (2011), is included to evaluate the behavior of the proposed approach when the PSF model is correct.

Given the frailties, the data under **Scenario I** were simulated from the conditional PH model (2.1) with $\lambda_0(t) = 1$, $\mathbf{w}_{ij} = (w_{1ij}, w_{2ij}, x_i)'$ and $\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_x)' = (1.0, 0.5, 1.0)'$; the data under **Scenario II** were simulated from the PSF model (2.2) with $\tilde{\mathbf{w}}_{ij} = (w_{1ij}, w_{2ij})'$, $\boldsymbol{\eta} = (1, 0.5)'$ and $H_0(t) = t$. For each simulation scenario, 200 replicates of the data set were generated by assuming the following: $w_{1ij} \overset{i.i.d.}{\sim} N(0, 1)$ and $w_{2ij} \overset{i.i.d.}{\sim}$ Bernoulli(0.5), $i = 1, \ldots, 100$, $j = 1, \ldots, 10$. In each case, a noninformative censoring scheme was considered, where the censoring times were simulated from an $U(0.25, 4)$ distribution, so that the censoring rate is approximately 35% under **Scenario I** and 25% under **Scenario II** .

For each data set, the GF approach was employed, yielding point estimates of $\boldsymbol{\xi}$, $\text{var}(e_i)$ and e_i, which we denote by $\hat{\boldsymbol{\xi}}^{(0)}$, $\hat{\theta}^{2(0)}$ and $\hat{e}_i^{(0)}$ respectively. Based on these point estimates, the predictive survival function was calculated as follows:

$$\hat{S}_{GF}(t|\mathbf{w}) = n^{-1} \sum_{i=1}^{n} \exp\left\{-\hat{\Lambda}_0^{(0)}(t) \exp\left\{\mathbf{w}'\hat{\boldsymbol{\xi}}^{(0)} + \hat{e}_i^{(0)}\right\}\right\}, \tag{2.4}$$

where $\hat{\Lambda}_0^{(0)}(t)$, depending on $\hat{e}_i^{(0)}$s, denotes Breslow's estimator of $\Lambda_0(t)$ (see, e.g., Therneau et al., 2003, Section 2). We then fitted the proposed model, by considering $J = 4$, $K = 10$, $\tau_1 = 1.001$, $\tau_2 = 1.001\hat{\theta}^{2(0)}$, $a_c = 1$, $b_c = 1$, $\boldsymbol{\gamma}_0 = \mathbf{0}_{13}$ and $\mathbf{S}_0 =$

$10^3 \times \mathbf{I}_{13}$. For each data set a single Markov chain of length $55,000$ was obtained by using the algorithm described in Appendix A.2. A burn-in period of $5,000$ scans was considered, and $5,000$ samples were retained for posterior inferences. The posterior mean of the corresponding parameters are denoted by $\hat{\boldsymbol{\xi}}$, $\hat{\theta}^2$, $\hat{g}(e|\mathbf{x})$ and $\hat{S}(t|\mathbf{w})$. Finally, the PSF approach was considered but including the cluster-level covariates in the linear predictor, and the associated predictive survival function, based on Breslow's estimator of the underlying baseline hazard function, was obtained and is denoted by $\hat{S}_{PSF}(t|\mathbf{w})$.

The competing approaches were compared regarding the estimation of the regression coefficients and also compared by computing the weighted integrated squared error (ISE) for the estimated survival distributions, given by

$$\int_0^\infty \left\{\hat{S}_m(t|\mathbf{w}) - S(t|\mathbf{w})\right\}^2 f_T(t|\mathbf{w})dt,$$

where $\hat{S}_m(t|\mathbf{w})$, $S(t|\mathbf{w})$ and $f_T(t|\mathbf{w})$ are the estimated survival function, the true survival function and density function, respectively, for a subject with covariate vector $\mathbf{w}$.

Simulation results

The results for the regression coefficients using the proposed model and the GF approach under **Scenario I** are given in Table 2.5, where the bias of the corresponding point estimators, the Monte Carlo mean of the posterior standard deviation/standard error (MEAN-SD), the Monte Carlo standard deviation of the point estimates (SD-MEAN) and the Monte Carlo coverage probability (CP) of the 95% credible interval/confidence intervals are presented. The results suggest that the posterior means of $\boldsymbol{\xi}$ are almost unbiased estimators and that the observed bias for ξ_x under the proposed approach is much smaller than the corresponding value obtained under the GF approach. Moreover, under the proposed model, the MEAN-SD and the SD-MEAN

values are in fairly close agreement, indicating that the posterior standard deviation is an unbiased estimator of the frequentist standard error. Finally, the CPs are all around the nominal 95%. The same does not hold for GF, which substantially underestimates the standard error for ξ_x, leading to low coverage probabilities.

Table 2.5 Simulation data – Scenario I: True value, bias of the point estimator, mean (across Monte Carlo simulations) of the posterior standard deviations/standard errors (MEAN-SD), standard deviation (across Monte Carlo simulations) of the point estimator (SD-MEAN) and Monte Carlo coverage probability for the 95% credible interval/confidence interval (CP) for the regression parameters. The results are presented under the proposed model and under the GF approach.

		Proposed Model				GF Model			
Parameters	True	BIAS	MEAN-SD	SD-MEAN	CP	BIAS	MEAN-SD	SD-MEAN	CP
ξ_1	1.0	0.011	0.052	0.054	0.930	-0.011	0.051	0.059	0.910
ξ_2	0.5	0.008	0.088	0.090	0.945	-0.003	0.088	0.091	0.950
ξ_x	1.0	-0.009	0.141	0.126	0.965	-0.052	0.083	0.142	0.775

The average of the estimated frailty distributions and survival functions across simulated data sets for some specific covariate values are presented in Figure 2.4 for **Scenario I** and in Figure 2.5 for **Scenario II**. The results in **Scenario I** reveal that the proposed model roughly captures the modal behavior of the covariate-dependent frailty distributions. Although not perfect, the proposed model performs remarkably well given that only $n = 100$ imperfectly-observed observations were generated for each data set. The situation is much less favorable for the GF approach, which fails to correctly capture the shape of the frailty distributions, leading to poor estimated survival functions. This behavior is likely driving the underestimation of survival noted in the SEER analysis. As expected, the PSF approach also suffers from bad prediction since the underlying assumption for frailty distribution is violated. The results in **Scenario II** show that the proposed model is still able to capture the frailty distributional shape even when the data were truly generated from the PSF model. Regarding the estimated survival curves, the results suggest that essentially no differences among the three methods are observed; all estimated functions are close

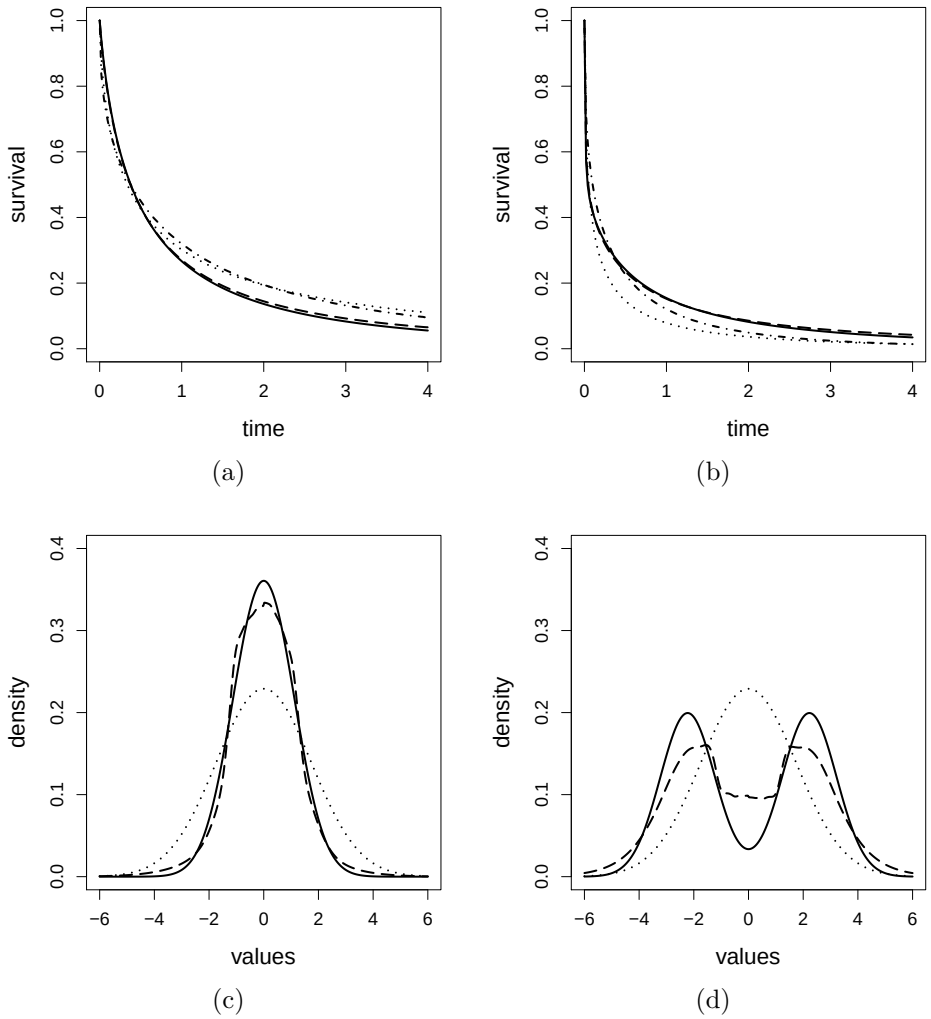


Figure 2.4 Simulated data – Scenario I: Mean, across simulations, of the posterior mean of the survival and frailty density functions under the proposed model. Panels (a) and (b) show the results for the survival functions. Panels (c) and (d) show the results for the frailty densities. Panels (a) and (c) show the results for covariate values $(2, 1, -2)$. Panels (b) and (d) show the results for covariate values $(0, 1, 2)$. The true curves are represented by continuous lines. The results under the proposed model are represented by dashed lines. The results under the exchangeable Gaussian frailty model are represented by dotted lines. In Panels (a) and (b) the results obtained under the PSF approach are represented by dot-dashed lines.

to the truth, indicating that there is little price to be paid when using the proposed model for the clustered survival data that were truly generated from the PSF model.

The results of the comparison of the estimated survival curves in terms of ISE are presented in Table 2.6, where the Monte Carlo mean and standard deviations for

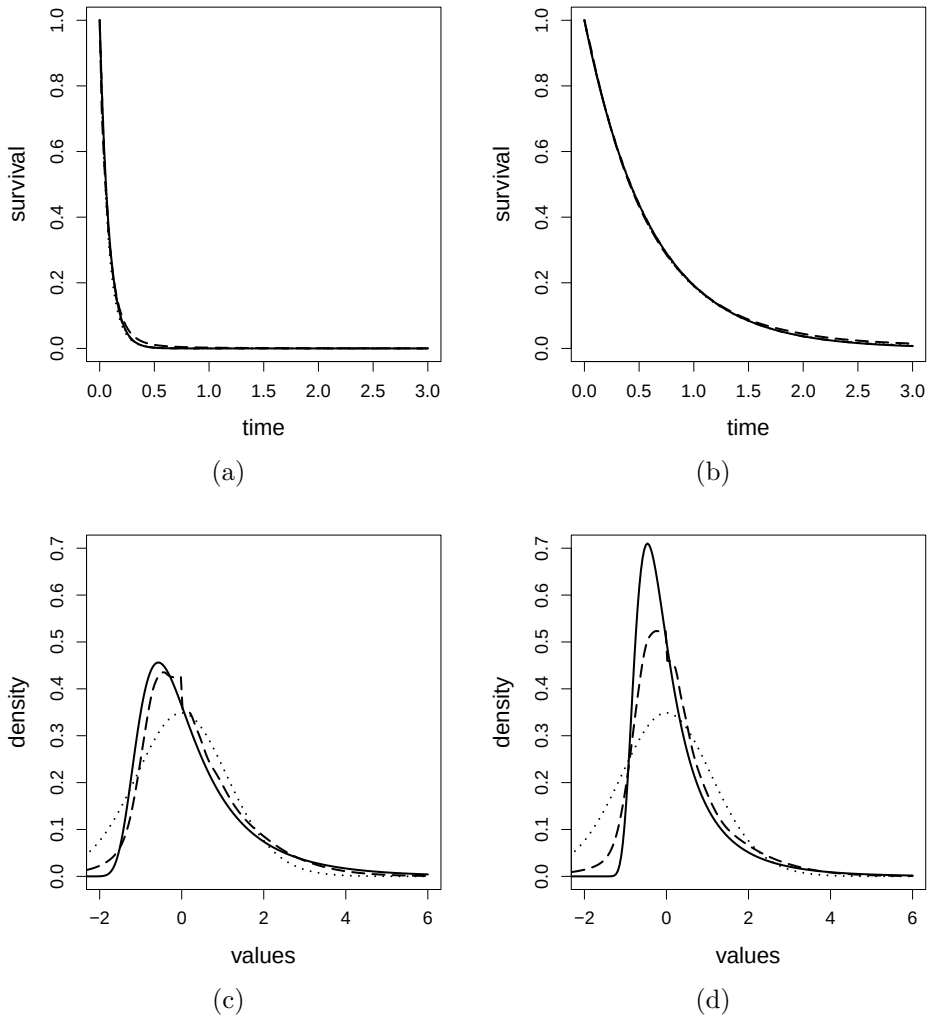


Figure 2.5 Simulated data – Scenario II: Mean, across simulations, of the posterior mean of the survival and frailty density functions under the proposed model. Panels (a) and (b) show the results for the survival functions. Panels (c) and (d) show the results for the frailty densities. Panels (a) and (c) show the results for covariate values $(2, 1, 0.5)$. Panels (b) and (d) show the results for covariate values $(0, 1, 1.5)$. The true curves are represented by continuous lines. The results under the proposed model are represented by dashed lines. The results under the exchangeable Gaussian frailty model are represented by dotted lines. In Panels (a) and (b) the results obtained under the PSF approach are represented by dot-dashed lines.

the ISE for two different predictor values are given. The results under **Scenario I** show a close agreement with the observed for the regression coefficients; the proposed model substantially outperforms the other two methods in terms of smaller means and standard deviations of the ISE. Even under **Scenario II**, the proposed model still provides almost the same results as the PSF model in terms of ISE.

Table 2.6 Simulated data – Scenario II: Monte Carlo mean (Monte Carlo standard deviation) for the ISE of the survival function for two different predictor values. The results for the different approaches under both simulation scenarios are presented. The numbers correspond to 10^3 times the original values.

Scenario	(w_1, w_2, x)	Proposed Model	GF Model	PSF Model
I	$(2, 1, -2)$	2.02 (2.48)	4.37 (3.46)	6.28 (3.49)
	$(0, 1, 2)$	1.94 (2.53)	10.5 (6.86)	14.3 (10.9)
II	$(2, 1, 0.5)$	3.17 (4.66)	3.13 (3.33)	2.19 (2.26)
	$(0, 1, 1.5)$	0.96 (1.18)	0.89 (1.22)	0.83 (1.10)

In Appendix A.4, additional simulation results are presented which show that, under **Scenario I**, for larger sample sizes better estimates of the frailty distributions are obtained and that the approach is not affected by the choice of J in the specification of the LDTFP model. For further comparison, we also fitted the exchangeable mixture of Polya trees (MPT) (Hanson, 2006b) frailty Cox's model using the function `PTglmm` available in `DPpackage` (Jara et al., 2011) under **Scenario I**, in which the results show that our approach outperforms the MPT, and considered a third scenario favorable to the GP approach, where the results show that our method pays little price for the extra generality when using the proposed model when normality and exchangeability are valid assumptions. Overall, the proposed approach provides a flexible way to capture the heterogeneity in the frailty distribution, provides superior prediction, and yields an essential improvement for the estimation of population effects, especially when the intra-cluster correlation (or variability in frailties) is relatively large. When the frailty variances are small across clusters, the proposed approach is still recommended due to its flexibility.

2.5 Concluding remarks

Very limited work has been done on covariate-adjusted frailty survival models for clustered time-to-event data. Liu et al. (2011) proposed a stratified Cox model with positive stable frailties, where the shape parameter of the frailty distribution is allowed to depend on cluster-level covariates. However, they essentially fitted a marginal Cox model, and then utilized the positive stable assumption and some imposed constraints to estimate the parameters in their proposed model. The model proposed in this paper cleanly separates population-level effects from the cluster-level effects, which determine the shape of the frailty distribution. Frailty density shape is modeled using a tractable median-zero LDTFP prior. Other nonparametric density regression approaches could also be considered; however, model identifiability requires a location constraint such as mean-zero or median-zero. The proposed model provides a natural generalization of the conventional PH model with parametric or nonparametric exchangeable frailties, and accommodates frailty distribution "evolution" over cluster-level covariates providing superior prediction, as shown in our simulation studies. When data are truly generated according to an exchangeable Gaussian frailty PH model or the model of Liu et al. (2011), our model does about the same as the underlying true model in terms of fixed effects and/or marginal survival estimations. We illustrate the usefulness of the proposed model with an analysis of a subset of the Iowa SEER breast cancer data, and demonstrate that higher degree of ruralness corresponds to a more bimodal frailty distributional shape with larger variance. In general, the proposed model is more flexible than currently existing frailty PH models, leading to more robust inferences, and thus is recommended. One drawback of the proposed model is that, as currently fit in R, obtaining inference takes longer.

For ease of computation, we have assumed a piecewise constant structure for the baseline hazard function $\lambda_0(t)$ and taken the independent normal prior distributions for $\log(\lambda_k)$'s, so that the baseline hazard heights λ_k and covariate effects $\boldsymbol{\xi}$ can be

updated simultaneously. The use of empirically-derived cut-points has permeated much of the Bayesian survival literature for over a decade. Use of Breslow's baseline estimate coupled with the GF approach led to a greatly increased LPML over the empirical approach. An obvious extension of our current work is to employ a smoothed baseline, for example, using penalized B-splines (Hennerfeind et al., 2006), the piecewise exponential with random cut-points (Sahu and Dey, 2004), MPT (Hanson, 2006b), etc. Any of these approaches could improve model fit and prediction, but cannot currently be fitted in the `R` software. We are currently working on extending the methodology in this paper to other survival models and smoothed baselines.

Acknowledgments

This work is supported by NCI grants R03CA165110 and 5R03CA176739, and ASPIRE grant from the University of South Carolina. A. Jara's research is supported by Fondecyt grant 1141193. We thank the referees and editors for numerous insightful suggestions which greatly enhanced the readability of this paper.

Supplementary Material

Supplement to this chapter is available in Appendix A, where we provide (A.1) technical details on the mixture of linear dependent tailfree processes, (A.2) detailed description of the MCMC algorithm, (A.3) sample R code to analyze the SEER data, (A.4) additional simulation studies and (A.5) additional analysis of the SEER data.

Chapter 3

Marginal Bayesian Nonparametric Spatial Survival Model[1]

The global emergence of *Batrachochytrium dendrobatidis* (Bd) has caused the extinction of hundreds of amphibian species worldwide. It has become increasingly important to be able to precisely predict time to Bd arrival in a population. The data analyzed herein present a unique challenge in terms of modeling because there is a strong spatial component to Bd arrival time and the traditional proportional hazards assumption is grossly violated. To address these concerns, we develop a novel marginal Bayesian nonparametric survival model for spatially correlated right-censored data. This class of models assumes that the logarithm of survival times marginally follow a mixture of normal densities with a linear dependent Dirichlet process prior as the random mixing measure, and their joint distribution is induced by a Gaussian copula model with a spatial correlation structure. To invert high-dimensional spatial correlation matrices, we adopt a full-scale approximation that can capture both large- and small-scale spatial dependence. An efficient Markov chain Monte Carlo algorithm with delayed rejection is proposed for posterior computation, and an R package `spBayesSurv` is provided to fit the model. This ap-

[1]Zhou, H., Hanson, T., and Knapp, R. (2015). Marginal Bayesian nonparametric model for time to disease arrival of threatened amphibian populations. *Biometrics*, accepted. Reprinted here with permission of Wiley.

proach is first evaluated through simulations, then applied to threatened frog populations in Sequoia-Kings Canyon National Park.

Keywords: Spatial survival data; Point-referenced; Bayesian nonparametric; Dependent Dirichlet process; Delayed rejection; Full-scale approximation; Copula; Proportional hazards.

3.1 Introduction

The Earth is currently experiencing the most severe mass extinction of species since the dinosaurs died off 65 million years ago. Scientists estimate that we are currently losing on the order of up to 50,000 species per year, 1,000 to 10,000 times greater than the fossil record (Chivian and Bernstein, 2008). The current mass extinction is almost entirely due to humankind in the form of destruction of natural habitats, but with disease being increasingly recognized as another important driver. The global emergence of *Batrachochytrium dendrobatidis* (Bd), a fungus that can kill frogs within a few weeks, has caused the extinction of hundreds of amphibian species worldwide (Wake and Vredenburg, 2008), including the recent rapid local extinction of many mountain yellow-legged frog populations in the Sierra Nevada mountains of California (Rachowicz et al., 2006; Vredenburg et al., 2010). These impacts of Bd have been described as "*...the most spectacular loss of vertebrate biodiversity due to disease in recorded history...*" (Skerratt et al., 2007). Once the most common amphibian in the region, mountain yellow-legged frogs now inhabit less than one-tenth of their range of one hundred years ago and continue to disappear at an alarming rate.

The mountain yellow-legged frog is a species complex of the southern mountain yellow-legged frog *Rana muscosa* and the Sierra Nevada yellow-legged frog *Rana sierrae* (Vredenburg et al., 2007). In part due to Bd-caused declines, these species were recently listed as "endangered" under the U.S. Endangered Species Act (Federal Register 2014). Bd often spreads in wavelike patterns (Cheng et al., 2011; Lips et al.,

2008). One of the authors (Knapp) collected data over a 12-year period by hiking large areas of Sequoia and Kings Canyon National Parks, one of the few areas in the Sierra Nevada that still contained some Bd-negative frog populations at the beginning of this study (Vredenburg et al., 2010). Bd-negative frog populations were discovered during the first park-wide survey of all suitable habitats (Knapp et al., 2003), and were revisited, typically every 1-2 years, over the study period to determine their Bd status over time.

It has become increasingly important to be able to predict when Bd is likely to arrive in a frog population, considering that interventions may be useful to prevent the extinction of frog populations following the Bd arrival (Vredenburg et al., 2010). We model the Bd arrival time of frog populations from discovery given two baseline measures of Bd proximity: (i) "bdwater," indicating whether Bd was present during the previous year ($= 1$) or not ($= 0$) in the watershed containing the frog population of interest, and (ii) "bddist," indicating the linear distance to the nearest Bd-positive location during the previous year. Once Bd arrived at a site, this site was assumed to always be Bd-positive in subsequent years (Vredenburg et al., 2010). The location was recorded as Universal Transverse Mercator (x, y)–coordinates in meters for each frog population. These data provide a unique challenge in terms of modeling: there is a strong spatial component to Bd arrival times across populations, and the proportional hazards assumption is rather severely violated.

Traditionally, most survival models incorporating spatial information have been semiparametric, conditional (so-called frailty) models; these include conditionally proportional hazards models (Banerjee et al., 2003; Hennerfeind et al., 2006), proportional odds models (Banerjee and Dey, 2005; Zhao et al., 2009), and accelerated failure time models (Zhao et al., 2009; Wang et al., 2012). Any of these models is preferable if they actually fit the data, as such semiparametric structure allows for easy conditional interpretation in terms of hazard ratios, odds ratios, or acceleration

factors, respectively. However, these models impose constraints on survival densities. For example, both proportional hazards and accelerated failure time models induce stochastically ordered survival times for populations with different covariates or frailties. For this reason, if the semiparametric structure is not appropriate, the addition of frailties may not improve model fit, or can even make it worse by adding noise. Instead of incorporating a spatial frailty term, Li and Lin (2006) proposed a marginal survival model, which they termed a semiparametric normal transformation model, where survival times are assumed to marginally follow the proportional hazards (PH) model and their joint distribution is specified by a Gaussian copula model with a spatial correlation structure. One advantage of their model over frailty models is that the regression coefficients have population-level interpretations. However, their model imposes constraints so that survival curves from different covariate levels are not allowed to cross, which is unrealistic in many practical applications (see De Iorio et al., 2009), including the mountain yellow-legged frog illustration used in the current study (e.g. see Kaplan-Meier estimates in Appendix B.6). For these data, a more flexible survival model is needed to quantify the risk factors associated with the Bd arrival time of frog populations while taking spatial correlation into account.

Flexible Bayesian nonparametric modeling techniques have been successfully developed for handling complex survival data that traditional semiparametric survival models fail to fit. One appealing feature about nonparametric approaches for estimating survival densities is their ability to avoid unrealistic constraints on how the variance, skewness, shape and even modality change with covariates. The linear dependent Dirichlet process mixture model (De Iorio et al., 2009; Jara et al., 2010, 2011), essentially a countable mixture of accelerated failure time models, provides a flexible way to capture crossing hazard and survival curves. However, it is unclear how to extend these Bayesian nonparametric models to a geostatistical setting for the analysis of spatially correlated survival data, and there has been virtually no related

literature so far.

In this article, we develop a marginal Bayesian nonparametric spatial survival model for the analysis of the Bd arrival time of yellow-legged frog populations given the baseline covariates bdwater and bddist. This model assumes that the logarithm of survival times marginally follow a linear dependent Dirichlet process mixture (LDDPM) model, and specifies their joint distribution via a Gaussian copula model. Two major features of the proposal are that the resulting survival curves are allowed to cross, or not, as dictated by the data, and that the inclusion of spatial information improves the prediction of Bd arrival time dramatically. To invert the high-dimensional spatial correlation matrices, we adopt a full scale approximation that can capture both large- and small-scale spatial dependence (Sang and Huang, 2012). To compare our proposal with the semiparametric normal transformation model proposed by Li and Lin (2006), we also present a Bayesian version of their model, denoted as a marginal PH spatial model. We develop efficient Markov chain Monte Carlo (MCMC) algorithms for both our model and the marginal PH spatial model. The analysis of the mountain yellow-legged frog data shows that our proposed spatial LDDPM model provides improvement over most traditional models, including the non-spatial LDDPM, PH, PH augmented with spatial frailties, and marginal PH spatial. For ease of implementation, we also developed an R package `spBayesSurv`, which is available for downloading at `http://cran.r-project.org/web/packages/spBayesSurv`, that can fit our proposed spatial LDDPM, the non-spatial LDDPM, PH, and the marginal PH spatial models.

The rest of this article is organized as follows. Section 2 describes the marginal LDDPM spatial survival model. Section 3 provides a recipe for efficient MCMC inference, discusses prediction, and proposes a cross-validated predictive model comparison criterion. Section 4 discusses simulation results, shows how ignoring spatial correlation grossly can bias inferences, and validates that our model selection crite-

rion works well. Section 5 presents the analysis of the time to Bd arrival data, and shows how posterior inference can be interpreted and put to practical use. The paper concludes with concluding remarks in Section 6.

3.2 Marginal LDDPM spatial survival model

Model specification

Suppose right-censored spatial survival data $\{(t_i^o, \delta_i, \mathbf{x}_i, \mathbf{s}_i) : i = 1, \ldots, n\}$ are collected in a spatial region of interest $\mathcal{D}$, where t_i^o is a recorded event time, δ_i is a censoring indicator equaling 1 if t_i^o is the observed event time and equaling 0 if the event time is right-censored at t_i^o, $\mathbf{x}_i$ is a p-dimensional vector of covariates including the intercept, and $\mathbf{s}_i$ records the spatial location. We denote by t_i the latent true (unobserved if $\delta_i = 0$) event time corresponding to t_i^o, and then the relationship between δ_i and t_i can be described by $\delta_i = I(t_i = t_i^o)$, where $\delta_i = 0$ implies $t_i > t_i^o$. Since event times are all positive, it remains appealing to work on the logarithms of event times, i.e. $y_i = \log(t_i)$, $y_i^o = \log(t_i^o)$. In addition, we assume an independent censoring scheme; that is, the event and censoring times are independent given the observed covariates.

We assume that $y_i|\mathbf{x}_i$ marginally follows a linear dependent Dirichlet process mixture model of De Iorio et al. (2009),

$$F_{\mathbf{x}_i}(y|G) = \int \Phi\left(\frac{y - \mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right) dG\{\boldsymbol{\beta}, \sigma^2\}, \tag{3.1}$$

where $\Phi(\cdot)$ is the cumulative distribution function (cdf) of the standard normal, and G follows the Dirichlet Process (DP) prior (Ferguson, 1973) with concentration parameter $\alpha > 0$ and base measure G_0 on $\mathbb{R}^p \times \mathbb{R}^+$, denoted by $G \sim DP(\alpha, G_0)$. This Bayesian nonparametric model treats the conditional distribution $F_{\mathbf{x}}$ as a function-valued parameter and allows its variance, skewness, modality and other features to flexibly vary with the $\mathbf{x}$ covariates. See Pati et al. (2013) for the sufficient conditions

for posterior consistency. Note that there is no clear approach to incorporate simple spatial frailties into this model, as there is a countable number of linear predictors.

To incorporate the spatial correlation among the event times, we first define a residual process $z(\mathbf{s}_i)$, after adjusting for covariate effects, as

$$z(\mathbf{s}_i) = \Phi^{-1}\left\{F_{\mathbf{x}_i}(y_i|G)\right\}, i = 1, \ldots, n. \tag{3.2}$$

Note that $z(\mathbf{s}_i)$ follows the standard normal distribution, providing a natural way of modeling spatial dependence. Specifically, we assume that $\mathbf{z} = (z(\mathbf{s}_1), \ldots, z(\mathbf{s}_n))'$ arises from a zero-mean Gaussian process (GP) $\{z(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}$ with a valid correlation function $C_{\boldsymbol{\theta}}(\mathbf{s}, \mathbf{s}')$; that is, $\mathbf{z}$ follows a multivariate Gaussian distribution as $\mathbf{z} \sim N_n(\mathbf{0}, \mathbf{C}_{\boldsymbol{\theta}})$, where $\mathbf{C}_{\boldsymbol{\theta}} = [C_{\boldsymbol{\theta}}(\mathbf{s}_i, \mathbf{s}_j)]_{i,j=1}^n$ is the $n \times n$ correlation matrix depending on a parameter vector $\boldsymbol{\theta}$. To allow for a nugget effect, we consider $C_{\boldsymbol{\theta}}(\mathbf{s}_i, \mathbf{s}_j) = \theta_1 \rho(\mathbf{s}_i, \mathbf{s}_j) + (1 - \theta_1) I(\mathbf{s}_i = \mathbf{s}_j)$, where $\rho(\cdot, \cdot)$ is a valid correlation function, and $\theta_1 \in [0, 1]$, also known as a "partial sill" in Waller and Gotway (2004), is a scale parameter measuring a local maximum correlation. The simplest parameterization for the correlation function $\rho(\cdot, \cdot)$ is an isotropic one, where the spatial correlation is assumed to be a function solely of the Euclidean distance d_{ij} between locations $\mathbf{s}_i$ and $\mathbf{s}_j$. In this paper, we consider the exponential correlation function $\rho(\mathbf{s}, \mathbf{s}') = \exp\{-\theta_2 ||\mathbf{s} - \mathbf{s}'||\}$, where θ_2 controls the spatial decay over distance. Other choices such as the spherical, Gaussian and Matérn correlation functions are also possible.

The above model specification is completely equivalent to Gaussian copula modeling (Song, 2000). In fact we have modeled the joint distribution of $\mathbf{y} = (y_1, \ldots, y_n)$ as a function of its marginal cdf, that is,

$$\mathbf{y} \sim \Phi_n\left(\Phi^{-1}\{F_{\mathbf{x}_1}(y_1|G)\}, \ldots, \Phi^{-1}\{F_{\mathbf{x}_n}(y_n|G)\}; \mathbf{C}_{\boldsymbol{\theta}}\right),$$

where $\Phi_n(\cdot; \boldsymbol{\Sigma})$ is the cdf of an n-dimensional normal with mean zero and covariance matrix $\boldsymbol{\Sigma}$. Following Song (2000), the likelihood function based upon the complete

data $\{(y_i, \mathbf{x}_i, \mathbf{s}_i), i = 1, \ldots, n\}$ is

$$\mathcal{L} = |\mathbf{C}_{\boldsymbol{\theta}}|^{-1/2} \exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{C}_{\boldsymbol{\theta}}^{-1} - \mathbf{I}_n)\mathbf{z}\right\} \prod_{i=1}^{n} f_{\mathbf{x}_i}(y_i|G), \tag{3.3}$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix, and $f_{\mathbf{x}_i}(y_i|G)$ is the density function corresponding to $F_{\mathbf{x}_i}(y_i|G)$.

Prior specification and hierarchical modeling

For the DP prior $G \sim DP(\alpha, G_0)$, we define the base measure G_0 through independent priors $N_p(\boldsymbol{\beta}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\text{Ga}(\sigma^{-2}|\nu_a, \nu_b)$. Here, $N_p(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\text{Ga}(\cdot|a, b)$ denote a p-dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and a gamma distribution with shape a and rate b, respectively. Note that it is critical to select an appropriate prior for the concentration parameter α, since its value controls the number of distinct components to which the data points are allocated. We assume a $\text{Ga}(\alpha|a_0, b_0)$ prior for α, which has been widely used by many researchers because of its tractability. As for the correlation parameters $\boldsymbol{\theta} = (\theta_1, \theta_2)$, they are typically not consistently estimable for a wide range of correlation functions, as demonstrated by Zhang (2004). This implies that increasing sample size does not necessarily obliterate the priors' impact, and thus weakly informative priors may be desirable to help identify the parameters $\boldsymbol{\theta}$, say, a $\text{Beta}(\theta_1|\theta_{1a}, \theta_{1b})$ for θ_1 and a $\text{Ga}(\theta_2|\theta_{2a}, \theta_{2b})$ for θ_2 with hyperparameters $\boldsymbol{\theta}_0 = (\theta_{1a}, \theta_{1b}, \theta_{2a}, \theta_{2b})$ being carefully choosen. Finally, we specify conjugate hyperpriors on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}^{-1}$ using a normal $N_p(\boldsymbol{\mu}|\mathbf{m}_0, \mathbf{S}_0)$ and a Wishart $W_p\left((\kappa_0\boldsymbol{\Sigma}_0)^{-1}, \kappa_0\right)$, respectively, where the Wishart has mean $\boldsymbol{\Sigma}_0^{-1}$ and degrees of freedom κ_0.

Following de Carvalho et al. (2013), we suggest reasonable default hyperpriors as follows: $a_0 = b_0 = 2$, $\nu_a = 3$, $\nu_b = \hat{\sigma}^2$, $\boldsymbol{\theta}_0 = \mathbf{1}$, $\mathbf{m}_0 = \hat{\boldsymbol{\beta}}$, $\mathbf{S}_0 = \hat{\boldsymbol{\Sigma}}$, $\boldsymbol{\Sigma}_0 = 30\hat{\boldsymbol{\Sigma}}$, and $\kappa_0 = 7$, where $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are the maximum likelihood estimates of $\boldsymbol{\beta}$ and σ^2 from fitting the log-normal accelerated failure time model $\log(t_i) = \mathbf{x}_i'\boldsymbol{\beta} + \sigma\epsilon_i, \epsilon_i \sim N(0, 1)$,

and $\hat{\boldsymbol{\Sigma}}$ is the asymptotic covariance estimate for $\hat{\boldsymbol{\beta}}$.

For ease of hierarchical modeling, we express the DP prior G in the stick-breaking form (Sethuraman, 1994) as

$$G = \sum_{k=1}^{\infty} w_k \delta_{(\boldsymbol{\beta}_k, \sigma_k^2)}, \quad w_k = V_k \prod_{j<k} (1 - V_j), \tag{3.4}$$

where δ_a is a Dirac probability measure concentrated at a, $V_k \overset{iid}{\sim} \text{Beta}(1, \alpha)$ and $(\boldsymbol{\beta}_k, \sigma_k^2) \overset{iid}{\sim} G_0$ are mutually independent for $k = 1, \ldots, \infty$. In practical implementations, either fixed (Ishwaran and James, 2001) or random stopping (Papaspiliopoulos and Roberts, 2008) approximation procedures of the infinite sum representation (3.4) can be considered. In this paper, we use the truncation approximation, replacing G with $G^N = \sum_{k=1}^{N} w_k \delta_{(\boldsymbol{\beta}_k, \sigma_k^2)}$, with N being pre-specified, where w_ks result from a truncated version of the stick-breaking construction: $w_1 = V_1$, $w_k = V_k \prod_{j=1}^{k-1}(1 - V_j), k = 2, \ldots, N$, $V_N = 1$. The truncation level N can be determined by considering the properties of the higher-order w_k values in the infinite sum representation (3.4), i.e. $U_N = \sum_{k=N+1}^{\infty} w_k$. Ishwaran and Zarepour (2000) demonstrated that $E(U_N|\alpha) = \alpha^N/(1+\alpha)^N$ and $\text{Var}(U_N|\alpha) = \alpha^N/(2+\alpha)^N - \alpha^{2N}/(1+\alpha)^{2N}$. Then for any given truncation level N, we can approximate these expressions by averaging over the gamma prior for α. For example, setting $N = 10$ and placing $\text{Ga}(2, 2)$ on α in our simulation study will result in $E(U_N) \approx 0.0055$ and $\text{Var}(U_N) \approx 0.0002$, which is more than adequate for data analyses.

In order to determine which component the ith data point is allocated, we introduce configuration variables K_i. Then the hierarchical model for the data, together

with the augmented latent true event-times, can be written as follows:

$$
\begin{aligned}
y_i|\mathbf{B},\boldsymbol{\sigma}^2,\mathbf{K} &\sim N(\mathbf{x}_i'\boldsymbol{\beta}_{K_i},\sigma^2_{K_i}) \\
z(\mathbf{s}_i) &= \Phi^{-1}\left\{\sum_{k=1}^{N} w_k \Phi\left(\frac{y_i-\mathbf{x}_i'\boldsymbol{\beta}_k}{\sigma_k}\right)\right\} \\
(z(\mathbf{s}_1),\ldots,z(\mathbf{s}_n))'|\boldsymbol{\theta} &\sim N_n(\mathbf{0},\mathbf{C}_{\boldsymbol{\theta}}) \\
P(K_i=k|\mathbf{V}) &= w_k,\ k=1,\ldots,N \\
(V_k|\alpha) &\overset{iid}{\sim} \text{Beta}(1,\alpha),\ k=1,\ldots,N-1 \\
(\boldsymbol{\beta}_k,\sigma_k^{-2})|\boldsymbol{\mu},\boldsymbol{\Sigma} &\overset{iid}{\sim} N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})\times \text{Ga}(\nu_a,\nu_b) \\
\alpha &\sim \Gamma(a_0,b_0) \\
(\theta_1,\theta_2) &\sim \text{Beta}(\theta_{1a},\theta_{1b})\times \text{Ga}(\theta_{2a},\theta_{2b}) \\
(\boldsymbol{\mu},\boldsymbol{\Sigma}^{-1}) &\sim N_p(\mathbf{m}_0,\mathbf{S}_0)\times W_p\left((\kappa_0\boldsymbol{\Sigma}_0)^{-1},\kappa_0\right)
\end{aligned}
\tag{3.5}
$$

where $\mathbf{B}=(\boldsymbol{\beta}_1',\ldots,\boldsymbol{\beta}_N')$, $\boldsymbol{\sigma}^2=(\sigma_1^2,\ldots,\sigma_N^2)$, $\mathbf{K}=(K_1,\ldots,K_n)$, and $\mathbf{V}=(V_1,\ldots,V_N)$.

3.3 Posterior inference

MCMC sampling

We develop an efficient MCMC algorithm for posterior sampling from the hierarchical model representation (3.5). The full conditionals are straightforward to derive, but most of them are not recognizable due to the incorporation of spatial dependence. A complete description and derivation of the updating steps are available in Appendix B.1 . The posterior samples for the model parameters are used for all inferences of interest.

Let $\boldsymbol{\Omega}=(\mathbf{K},\mathbf{y},\mathbf{B},\boldsymbol{\sigma}^2,\mathbf{V},\alpha,\boldsymbol{\theta})$ denote collectively the model parameters to be updated. Note that the likelihood function involves the inversion and determinant calculation of a very large global correlation matrix $\mathbf{C}_{\boldsymbol{\theta}}$ and these matrix operations have to be repeated for every MCMC iteration. For large values of the sample size n, e.g., $n \geq 500$, we suggest replacing $\mathbf{C}_{\boldsymbol{\theta}}$ with $\mathbf{C}^{\dagger}_{\boldsymbol{\theta}}$ based on the full scale approximation

(FSA) approach as described in Appendix B.2. Conditional on all other parameters, K_i is sampled from a multinomial distribution. For updating y_i, $\boldsymbol{\beta}_k$, σ_k^2 and V_k, we notice that the full conditional for each is proportional to a recognizable density multiplied by a common part $\exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{C}_{\boldsymbol{\theta}}^{-1} - \mathbf{I}_n)\mathbf{z}\right\}$. Thus we use Metropolis-Hastings (M-H) with independent proposals, where each proposal is based on the recognizable density. However, we observed for this inititial MCMC scheme that often some $\boldsymbol{\beta}_k$ got "stuck" for a long period, leading to poor MCMC mixing. As a remedy, we found that delayed rejection (Tierney and Mira, 1999) works very well. Upon a rejection in the M-H, instead of retaining the same position, a second-stage proposal corresponding to a random walk is proposed. The precision parameter α, and hyperparameters $(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1})$ are updated from their conjugate full conditionals. Finally, to update the correlation parameters $\boldsymbol{\theta}$, we first take transformations $\boldsymbol{\vartheta} = (\vartheta_1, \vartheta_2)'$ with $\vartheta_1 = \log\left(\frac{\theta_1}{1-\theta_1}\right)$ and $\vartheta_2 = \log(\theta_2)$, and then update $\boldsymbol{\vartheta}$ using adaptive Metropolis-Hastings algorithms (Haario et al., 2001).

Given a set of posterior samples $\{\boldsymbol{\Omega}^{(l)}, l = 1, \ldots, L\}$, the marginal conditional density of log event time y given the covariates $\mathbf{x}$ is estimated by

$$\hat{f}_{\mathbf{x}}(y) = \frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{N} w_k^{(l)} \frac{1}{\sigma_k^{(l)}} \phi\left(\frac{y - \mathbf{x}'\boldsymbol{\beta}_k^{(l)}}{\sigma_k^{(l)}}\right), \tag{3.6}$$

where $\phi(\cdot)$ is the density of the standard normal. The marginal conditional survival and hazard functions can be estimated similarly. Then all the marginal density, survival and hazard functions of an event time $t = \exp\{y\}$ given $\mathbf{x}$ can be easily obtained. An R package `spBayesSurv` accompanying this paper is provided to implement the MCMC algorithm and plot the estimated curves; see Appendix B.7 for sample R code.

Spatial prediction

In geostatistics, one major interest is predicting the survival t_0 at a new location $\mathbf{s}_0$ with associated covariate values $\mathbf{x}_0$. Given the parameters $\boldsymbol{\Omega}$, by noting that $\mathbf{z} \sim N_n(\mathbf{0}, \mathbf{C}_{\boldsymbol{\theta}})$, we can easily obtain that $z(\mathbf{s}_0) \sim N(\mu(\mathbf{s}_0), \tau^2(\mathbf{s}_0))$ with $\mu(\mathbf{s}_0) = \mathbf{h}(\mathbf{s}_0)'\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{z}$ and $\tau^2(\mathbf{s}_0) = 1 - \mathbf{h}(\mathbf{s}_0)'\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{h}(\mathbf{s}_0)$, where $\mathbf{h}(\mathbf{s}_0) = [\theta_1\rho(\mathbf{s}_0, \mathbf{s}_i)]_{i=1}^n$ is an $n \times 1$ vector. Note that both $\mathbf{C}_{\boldsymbol{\theta}}$ and ρ would be replaced by $\mathbf{C}_{\boldsymbol{\theta}}^{\dagger}$ and $\rho^{\dagger}$ respectively if the FSA is used. Based on the definition of $z(\cdot)$ in equation (3.2) and the N-level truncation of G, we have

$$z(\mathbf{s}_0) = \Phi^{-1}\left\{\sum_{k=1}^{N} w_k \Phi\left(\frac{\log t_0 - \mathbf{x}_0'\boldsymbol{\beta}_k}{\sigma_k}\right)\right\}. \tag{3.7}$$

It follows that the predictive density of t_0 is given by

$$\begin{aligned} f(t|\boldsymbol{\Omega}, \mathbf{x}_0, \mathbf{s}_0) =& \frac{1}{t \cdot \tau(\mathbf{s}_0)}\phi\left(\frac{z(\mathbf{s}_0) - \mu(\mathbf{s}_0)}{\tau(\mathbf{s}_0)}\right) \\ & \times \frac{\sum_{k=1}^{N} w_k \frac{1}{\sigma_k}\phi\left(\frac{\log t - \mathbf{x}_0'\boldsymbol{\beta}_k}{\sigma_k}\right)}{\phi(z(\mathbf{s}_0))}. \end{aligned} \tag{3.8}$$

Given a set of posterior samples $\{\boldsymbol{\Omega}^{(l)}, l = 1, \ldots, L\}$ obtained from Section 3.3, we first draw $z^{(l)}(\mathbf{s}_0)$ from $N(\mu(\mathbf{s}_0), \tau^2(\mathbf{s}_0))$ for $l = 1, \ldots, L$, and then make the transformation according to (3.7) to obtain a sample of predictive event-times $\{t^{(l)}(\mathbf{s}_0), l = 1, \ldots, L\}$ at the new location $\mathbf{s}_0$. The final predicted value of t_0 can be either the mean or median of $t^{(l)}(\mathbf{s}_0)$s. In practice, it is difficult to observe all the covariates at the whole study region $\mathcal{D}$, so it is not practically feasible to create a map for predictive event times. Alternatively, we may show a spatial map for the residual process, which can be interpreted in a manner similar to the spatial frailties in conditional survival models; that is, the higher the $z(\mathbf{s})$ is, the larger the event time $y(\mathbf{s})$ would be on average. Note that the predictive density of $t(\mathbf{s}_0)$ is simply (3.8) averaged over the MCMC iterates. It is worth highlighting here that (3.8) is different from the truncated version of the marginal density in (3.1). This is due to the fact that the event-times are spatially correlated so that the prediction at a new location will

borrow information from locations where the data have been collected. On the other hand, if we assume there is no spatial correlation, i.e. $\theta_1 = 0$, it is easy to see that the predictive density in (3.8) reduces to the N-level truncation version of (3.1).

Model comparison

To compare the predictive ability of competing models, we consider the conditional predictive ordinate (CPO) statistic as suggested by Geisser and Eddy (1979). Let $\mathcal{D}_{-i}$ denote the observed data excluding the ith data point. For a given model, the CPO statistic for the ith observation is defined as $\text{CPO}_i = f\left(t_i^o|\mathcal{D}_{-i}\right)^{\delta_i} S\left(t_i^o|\mathcal{D}_{-i}\right)^{1-\delta_i}$, where f and S denote the marginal posterior predictive density and survival functions of t_i^o given $\mathcal{D}_{-i}$, respectively. A higher value of CPO_i under one model implies a better fit of that model to the ith observation. Let $\mathbf{t} = (t_1, \ldots, t_n)$ be the vector of latent true survival times and $\mathbf{t}_{-i}$ be the corresponding vector with the ith element removed. According to the hierarchical model in (3.5), given all the model parameters $\boldsymbol{\Theta} = (\mathbf{B}, \boldsymbol{\sigma}^2, \mathbf{V}, \boldsymbol{\theta})$, we show that the CPO of Gelfand and Dey (1994) is generalized to

$$\text{CPO}_i = \left(E_{(\mathbf{t},\boldsymbol{\Theta}|\mathcal{D})} \left[\frac{1}{f(t_i^o|\mathbf{t}_{-i}, \boldsymbol{\Theta})^{\delta_i} S(t_i^o|\mathbf{t}_{-i}, \Theta)^{1-\delta_i}} \right] \right)^{-1}. \tag{3.9}$$

See Appendix B.3 for its derivation and the expressions of $f(t_i^o|\mathbf{t}_{-i}, \boldsymbol{\Theta})$ and $S(t_i^o|\mathbf{t}_{-i}, \boldsymbol{\Theta})$. To give an aggregate summary measure of a model's predictive ability, we define the log pseudo marginal likelihood (LPML) as $\text{LPML} = \sum_{i=1}^{n} \log(\text{CPO}_i)$. The LPML is a cross-validated predictive measure: the larger a model's LPML is, the better predictive ability the model has. From (3.9), one can easily compute LPML from the MCMC output.

3.4 Simulations

We conduct a simulation study to illustrate the proposed model (LDDPM-spatial) and assess its performance. We also compare it with the Bayesian version of Li and

Lin (2006) (PH-spatial, see Appendix B.4) and the model by De Iorio et al. (2009) (LDDPM-ind). All analyses can be run in `R` using the package `spBayesSurv`.

We randomly select 400 locations over a spatial region $[0, 40] \times [0, 100]$ (mirroring from the frog data) and hold out 100 of them for assessing the predictive performance, yielding a total sample of $n = 300$ subjects for estimation. The log event times $y(\mathbf{s}) = \log t(\mathbf{s})$ at these 400 locations are simulated from a mixture model $f(y|x) = 0.4N(3.5 + 0.5x, 1^2) + 0.6N(2.5 - x, 0.5^2)$ with the spatial dependence described in Section 3.2, where x is generated independently from a uniform distribution over $(-1.5, 1.5)$ and $\mathbf{C}_{\boldsymbol{\theta}}$ is specified with $\theta_1 = 0.98$ and $\theta_2 = 0.1$. The choice of this mixture model is based on a modification of the simulation study in De Iorio et al. (2009). The log censoring times are simulated from a uniform distribution on $(3, 4)$ so that the censoring rate is about $20\% \sim 50\%$. This simulation study is referred to as Scenario I, for which 100 Monte Carlo replicate datasets are generated.

First, we fit the LDDPM-spatial model using truncation level $N = 10$ under the default prior specifications. We also fit the LDDPM-spatial model with $\mathbf{C}_{\boldsymbol{\theta}}$ approximated using the FSA approach (denoted as LDDPM-spatial-FSA), where we experiment with $m = 10$ regularly spaced knots and $B = 10$ blocks taken as equally sized squares. Second, we fit the PH-spatial model with default priors. Finally, we fit the LDDPM-ind model using the same priors as the LDDPM-spatial model. For each MCMC, we retain $10,000$ scans thinned from $50,000$ after a burn-in period of $10,000$ iterations. To assess the prediction ability and accuracy, for each above model, we calculate the LPML and mean squared prediction error (MSPE), where MSPE $= \sum_{i=1}^{100}(y_i - \tilde{y}_i)^2/100$ with y_is being the held-out true log survival times and $\tilde{y}_i$s being the corresponding predicted values based on posterior means. The models are also compared by computing the integrated squared error (ISE) for estimated survival curves, given by ISE $= \int_0^\infty \left\{\hat{S}(y|x) - S(y|x)\right\}^2 dy$, where $\hat{S}(y|x)$ and $S(y|x)$ are estimated and true survival functions given x, respectively.

Table 3.1 Simulated data – Scenario I. True value, bias of the point estimator (posterior mean), mean (across Monte Carlo replicates) of the posterior standard deviations (MEAN-SD), standard deviation (across Monte Carlo replicates) of the point estimator (SD-MEAN), and Monte Carlo coverage probability for the 95% credible interval (CP) for the spatial correlation parameter $\boldsymbol{\theta}$. The averaged computing time is also presented.

Model	Parameters	True	BIAS	MEAN-SD	SD-MEAN	CP
LDDPM-spatial	θ_1	0.98	-0.026	0.028	0.028	0.91
(32 minutes)	θ_2	0.10	0.017	0.026	0.025	0.95
LDDPM-spatial-FSA	θ_1	0.98	-0.017	0.025	0.024	0.96
(24 minutes)	θ_2	0.10	0.016	0.025	0.023	0.93
PH-spatial	θ_1	0.98	-0.065	0.037	0.047	0.61
(24 minutes)	θ_2	0.10	-0.009	0.023	0.030	0.78

The posterior inferences for spatial correlation parameters $\boldsymbol{\theta} = (\theta_1, \theta_2)$ under each approach are presented in Table 3.1, where the bias of corresponding point estimates (i.e. posterior means), the Monte Carlo mean of posterior standard deviation estimates (MEAN-SD), the Monte Carlo standard deviation of point estimates (SD-MEAN), and the Monte Carlo coverage probability of 95% credible intervals (CP) are reported. The results suggest that the point estimates of $\boldsymbol{\theta}$ are almost unbiased, and that the observed biases under the LDDPM-spatial model are much smaller than the corresponding values under the PH-spatial model. The MEAN-SD and SD-MEAN values are fairly close indicating that the posterior standard deviation is an appropriate estimator of the frequentist standard error. The CPs are around the nominal 95% level. In contrast, the PH-spatial model provides substantially lower coverage probabilities. Furthermore, the posterior estimates with FSA are very close to those using the exact model, suggesting that FSA is a good approximation of the correlation matrix $\mathbf{C}_{\boldsymbol{\theta}}$. Table 3.1 also presents the Monte Carlo mean of computing times under each approach, where we see that FSA does speed up the computation as expected.

Figure 3.1 shows boxplots of the ISEs for estimated survival curves, LPMLs, and MSPEs under the considered models. The LDDPM-spatial models (with and without FSA) provide much smaller biases of the fitted survival functions on average,

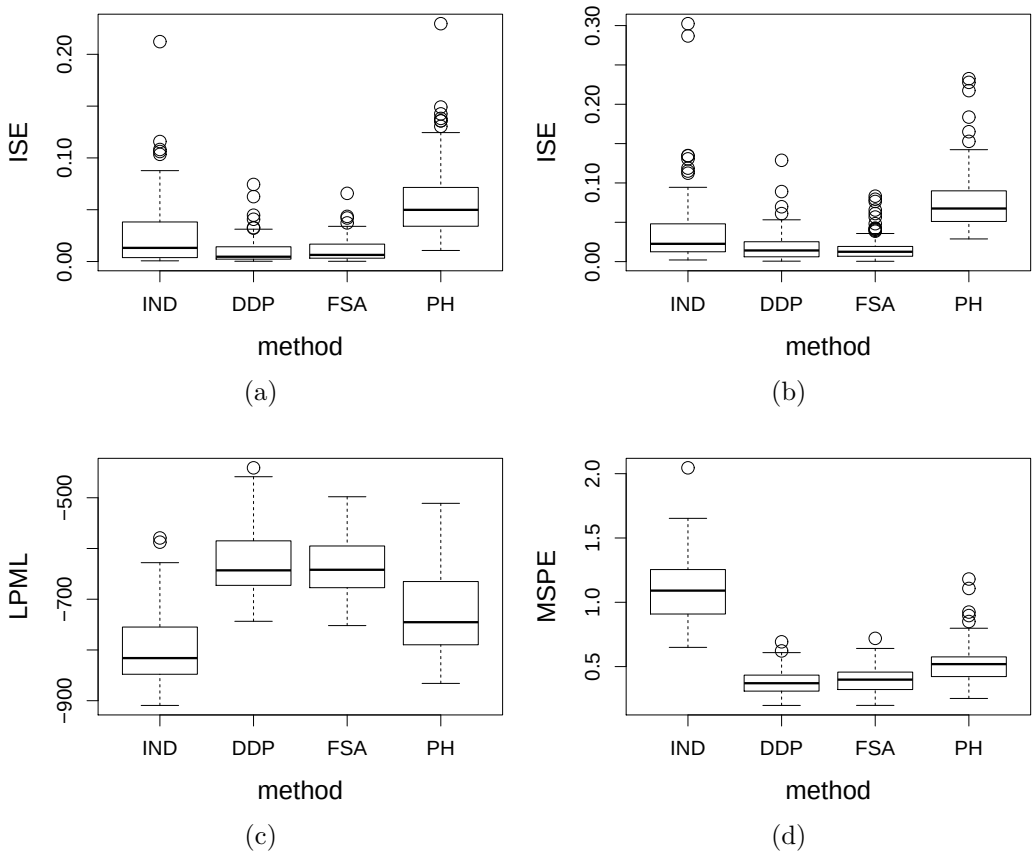


Figure 3.1 Simulated data – Scenario I. Panel (a) and (b): boxplots of ISEs for fitted survival curves when $x = -1$ and $x = 1$, respectively. Panel (c): boxplots of LPMLs. Panel (d): boxplots of MSPE. In each panel, the four models from left to right are LDDPM-ind, LDDPM-spatial, LDDPM-spatial-FSA, and PH-spatial, respectively.

compared with LDDPM-ind and PH-spatial, indicating that either violation of the PH assumption or ignorance of the spatial dependence could bias the inference. As for prediction ability and accuracy, the proposed models (with and without FSA) yield the best prediction performance as measured by both LPML and MSPE, compared with the PH-spatial and LDDPM-ind models. These simulations, i.e., when the truth is known, show that the LPML is consistent with the MSPE and hence validates its use for complex spatial models exhibiting dependence.

In Appendix B.5, we also tested the performance of LDDPM-spatial model when the PH assumption is satisfied and compared it with the PH-spatial model. The

Table 3.2 Frog data. Summary for event times, censoring status, bdwater, and bddist.

Time to Bd (yrs)	Bd status	bdwater	bddist (km)
5 (median)	33 (censored)	57 (bdwater=1)	1.811 (median)
1-11 (range)	276 (event)	252 (bdwater=0)	0.092-9.189 (range)

results show that two models provide almost the same boxplots of LPMLs and MSPEs, indicating that the LDDP-spatial model is quite competitive even when the PH assumption is satisfied.

3.5 Application to frog data

From 1997 to 2002, all mapped lentic water bodies in Sequoia and Kings Canyon National Parks were surveyed for mountain yellow-legged frogs. Starting in 2002 and continuing through 2011, nearly all the frog populations that were discovered during the initial surveys were visited every 1-2 years. A primary objective of these resurveys was to determine the Bd status of each frog population over time. Bd attacks the keratinized tissues of tadpole mouthparts and disrupts the normal black pigmentation of these structures. Therefore, early efforts to detect Bd in live amphibians often relied on visual inspections of tadpole mouthparts (see Table 4 in Knapp and Morgan, 2006). This method was replaced by a real-time quantitative PCR assay (qPCR) in the mid-2000s (Boyle et al., 2004), and remains the most reliable method for detecting Bd. In our data set, Bd status was determined using the inspections of tadpole mouthparts during resurveys conducted from 2002 to 2004 and using the qPCR after 2004. The data consist of $n = 309$ frog populations that were initially discovered during park-wide surveys conducted from 1997 to 2002, and then resurveyed regularly through 2011. The observed event time is calculated as the number of years from the initial survey to either Bd arrival or the last resurvey. By the end of the study, about 11% of the frog populations remained Bd-negative (censored). Table 3.2 presents a summary of the data.

Table 3.3 Frog data. Posterior statistics for θ_1 and θ_2 under the LDDPM-spatial model assuming the exponential correlation function. The computing time is also presented.

Model	Parameters	Mean	Median	Std. dev.	95% CI
LDDPM-spatial	θ_1	0.991	0.992	0.004	(0.982, 0.998)
(3.2 hours)	θ_2	0.133	0.130	0.040	(0.060, 0.216)

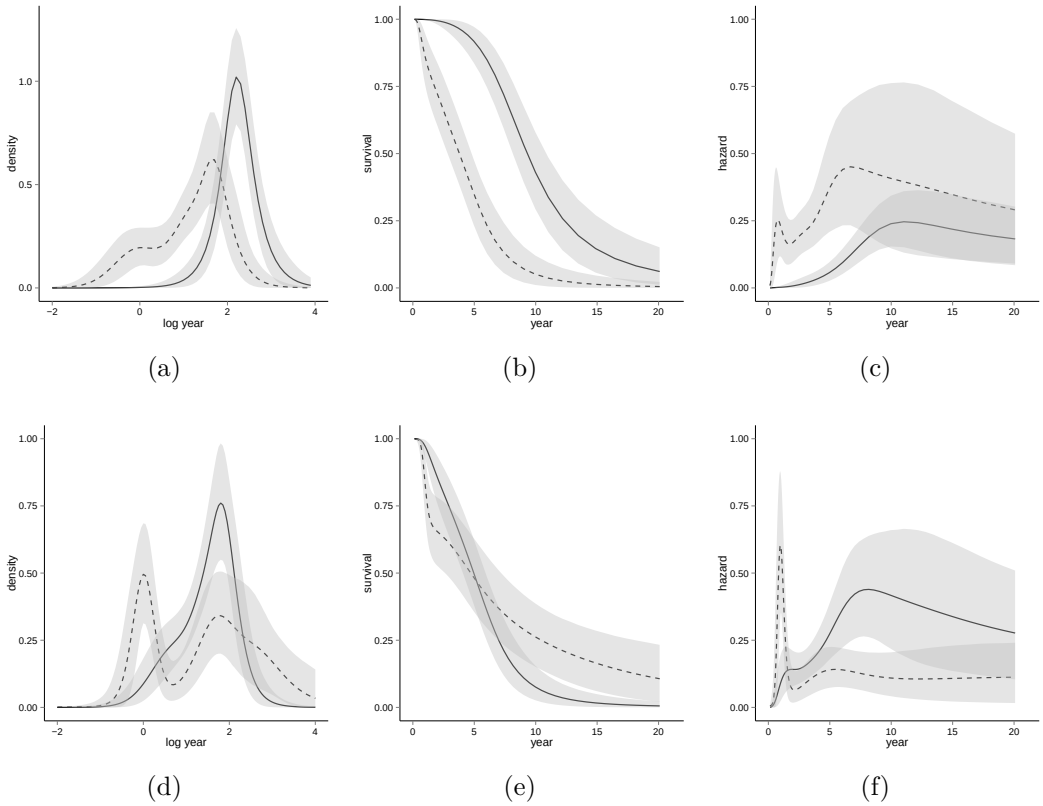
We fit the LDDPM-spatial and PH-spatial models to the data using the same prior specifications as previous simulations. We also fit the LDDPM-ind model, the standard PH model $\lambda_i(t) = \lambda_0(t)\exp\{\mathbf{x}_i'\boldsymbol{\beta}\}$ as well as the PH model with a point-referenced frailty term $\lambda_i(t) = \lambda_0(t)\exp\{\mathbf{x}_i'\boldsymbol{\beta}+w(\mathbf{s}_i)\}$ (PH-frailty), where $w(\mathbf{s})$ follows a Gaussian process (GP) with the exponential covariance function, and the baseline hazard $\lambda_0(t)$ is modeled in the same way as the PH-spatial model. Based upon examintion of trace plots for model parameters in both the simulations and real data analyses, for each MCMC run we retain $20,000$ scans thinned from $200,000$ after a burn-in period of $200,000$ iterations. These are grossly conservative numbers; burnin and thinning requirements may change considerably depending on the amount of spatial correlation and data. The Markov chains mix reasonably well for all fitted models. For the LDDPM models, the number of components with non-negligible mass ranged from two to five, indicating at most five components in the mixture. In Appendix B.6, we present the posterior trace plots for $\boldsymbol{\theta}$ under the LDDPM-spatial model.

We first obtain the LPML values for all models under consideration as follows: -276.7 for LDDPM-spatial, -304.4 for PH-spatial, -631.5 for LDDPM-ind, -705.3 for PH and -703.4 for PH-frailty. The LDDPM-spatial model provides significantly better prediction as measured by LPML, with differences ranging from 27 to 428. Interestingly, the PH model augmented with a GP frailty surface hardly improves inference over the standard PH model. In what follows, we only focus on interpreting results from the proposed model. Table 3.3 shows posterior estimates of the spatial

dependence parameters θ_1 and θ_2. The partial sill parameter θ_1 measures the maximum correlation between frog populations if they were located in the same location. Our analysis shows that such a correlation is very strong, almost equal to 1. The parameter θ_2 controls the decay of spatial dependence over distance measured by kilometers. For instance, the posterior mean of $\theta_2 = 0.133$ indicates that the correlation decays by $1 - e^{-0.133 \times 5} \approx 48.6\%$ for every 5-km increase in distance. This tells us that the spatial dependence still does not disappear even when the two frog populations are located 5-km apart.

Figure 3.2(b) shows, for example, for frogs living in a watershed that was Bd-positive during the previous year, the population-averaged median Bd arrival time is cut from around 9.8 years to 4.5 years, more than half, when the distance to nearest Bd-positive location goes from 7.731-km to 0.443-km. Interestingly, the *shapes* of densities also change, going from unimodal to bimodal in Figures 3.2(a) and 3.2(d). This figure clearly shows the invalidation of the standard AFT model for these data. Similarly, Figure 3.2(c) shows that the non-proportional change in hazards clearly invalidates the PH assumption. Comparing Bd-positive to Bd-negative in the watershed yields strikingly different outcomes. Populations in basins in which Bd is present, holding the distance to the nearest Bd-positive location constant, have a hazard spike near zero reaching up to around 0.7 (Figure 3.2(f)). This implies that the population-averaged probability of Bd arrival within one-year is about 70%. The corresponding survival curves cross (Figure 3.2(e)), invalidating most common semi-parametric models. These results are for any frog population randomly found in the study region. The simulations in Section 3.4 show that if the accompanying spatial information $\mathbf{s}_0$ is also used, prediction will be accurately refined.

We are also interested in predicting which areas have overall lower survival rates. Since the marginal distribution at each location cannot be predicted unless the associated baseline covariates are available, we instead show a spatial map (Figure 3.3)



(a) (b) (c)

(d) (e) (f)

Figure 3.2 Frog data. Fitted marginal densities (panels (a) and (d)), survival curves (panels (b) and (e)) and hazard curves (panels (c) and (f)) with 90% point-wise credible intervals for high versus low value of bddist when bdwater is equal to 0 (panels (a), (b) and (c)) and for bdwater=0 versus bdwater=1 when bddist is equal to its population mean of 2.717-km (panels (d), (e) and (f)). In panels (a), (b) and (c), the results for bddist=95% and bddist=5% quantiles are displayed as solid and dashed lines, respectively. In panels (d), (e) and (f), the results for bdwater=0 and bdwater=1 are displayed as solid and dashed lines, respectively.

by smoothing the predicted residual process $z(\mathbf{s})$ at $10,000$ randomly simulated new locations over the national park. One may interpret this map in a manner similar to the frailty map in the GP frailty PH model in the absence of covariate information; lower value of $z(\mathbf{s})$ indicates lower survival rate at location $\mathbf{s}$ on average. Overall, the frog populations living in darker regions became infected by Bd more earlier than lighter areas.

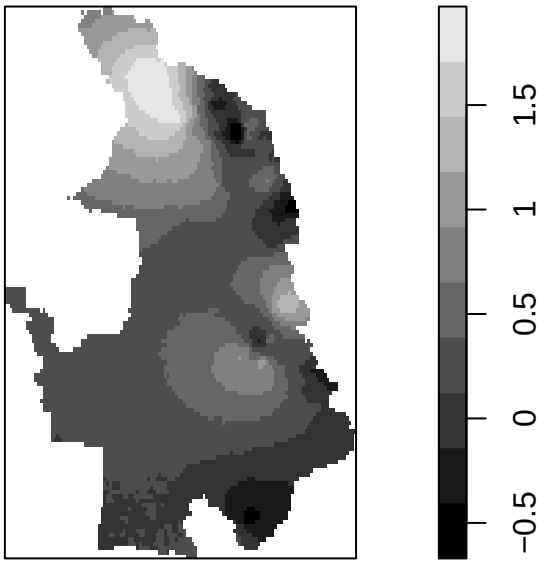


Figure 3.3 Frog data. Spatial map for the predicted residual process $z(\mathbf{s})$ at $10,000$ randomly simulated new locations. Higher value implies better survival overall.

3.6 Concluding remarks

This paper presents a unified approach to the nonparametric modeling of point-referenced survival data. The methodology is broadly illustrated on an analysis of Bd arrival time of yellow legged frog populations throughout Sequoia-Kings Canyon National Park. The frog data present a unique challenge and opportunity, as most common semiparametric survival models are grossly violated and there is a strong spatial component to the data. In the analysis, we considered two important measures of Bd proximity: bdwater and bddist; additional covariates could be easily incorporated into the vector $\mathbf{x}$ in our model (De Iorio et al., 2009; Jara et al., 2010). As expected, the closer a frog population is to a Bd-positive location, the less time that population has until Bd infects them. This analysis shows how modern, cutting edge statistical techniques can be used to understand a real ecological problem, here predicting the arrival time by Bd in Bd-negative frog populations.

One surprising result is just how poorly the PH model can do in terms of prediction when it is wrong, even when a traditional GP spatial frailty surface is added to the linear predictor. Zhao et al. (2009) note that in a conditional spatial survival setting, the survival model itself is the most important factor affecting prediction, in their case choosing among PO, PH, or AFT model. The survival model itself is more important than whether frailties are included, or whether the frailties are spatially varying or exchangeable. This observation also holds for the frog data: the assumptions that most semiparametric models imply are not met. In particular, the traditional PH model is inadequate relative to a more flexible nonparametric model. Once an appropriate nonparametric model has been chosen, the copula model can drastically improve prediction. It may be that, for many data sets, the marginal copula-based geostatistical modeling approach provides superior prediction over more traditional GP random-effects models, e.g., the types that can be fit in SAS `glimmix` and `mixed` procedures, the free `BayesX` program (Brezger et al., 2005). For this reason we have provided an efficient R package `spBayesSurv` for others to explore the use of these geostatistical copula models.

One drawback of the proposed model is that covariate effects are interpreted by examining the plots, which becomes challenging for high-dimensional covariates. Due to the inverse of high-dimensional correlation matrices during MCMC, the LDDPM model can also suffer longer computational times. In addition, the proposed model cannot handle time-dependent covariates. The traditional PH model can handle time-dependent covariates, and can have time-varying regression effects, which are not considered here nor were considered by Li and Lin (2006). However, once the PH model is augmented with time-varying effects, inference is also reduced to the examination of plots (e.g. survival curves and densities) and one might as well start with a purely nonparametric model. Future research will examine the fit of so-called additive spatial PH models, i.e. with time-varying regression effects.

An alternative way to incorporate spatial dependence into model (3.1) is to assume a multivariate latent stick-breaking process prior (Rodríguez et al., 2010) for the random mixing measure G. This approach uses a copula representation only on the random mixing measure G but requires constraints on the atoms in the mixing measure. The model is broadly developed for the univariate case, but multivariate cases are problematic in that the order constraint becomes complex even for two dimensions. The dimensionality of $(\boldsymbol{\beta}_k, \boldsymbol{\sigma}_k^2)$ in our approach is four, but likely to be much higher in general. By taking the full LDDPM model as the marginal in the copula representation, analysis is greatly simplified.

Supplementary material

Appendices referenced in Sections 3.1, 3.3, 3.4, and 3.5, and sample R code fitting the proposed model are available with this dissertation in Appendix B.

Chapter 4

Generalized Accelerated Failure Time Spatial Frailty Model[1]

Flexible incorporation of both geographical patterning and risk effects in cancer survival models is becoming increasingly important, due in part to the recent availability of large cancer registries. Most spatial survival models stochastically order survival curves from different subpopulations. However, it is common for survival curves from two subpopulations to cross in epidemiological cancer studies and thus interpretable standard survival models can not be used without some modification. Common fixes are the inclusion of time-varying regression effects in the proportional hazards model or fully nonparametric modeling, either of which destroys any easy interpretability from the fitted model. To address this issue, we develop a generalized accelerated failure time model which allows stratification on continuous or categorical covariates, as well as providing per-variable tests for whether stratification is necessary via novel approximate Bayes factors. The model is interpretable in terms of how median survival changes and able to capture crossing survival curves in the presence of spatial correlation. A detailed Markov chain Monte Carlo algorithm is presented for posterior inference and a freely-available function `frailtyGAFT` is provided to fit the model in the R package `spBayesSurv`.

[1]Zhou, H., Hanson, T., and Zhang, J. (2015). Generalized accelerated failure time spatial frailty model for arbitrarily censored data. *Lifetime Data Analysis*, revision submitted.

We apply our approach to a subset of the prostate cancer data gathered for Louisiana by the Surveillance, Epidemiology, and End Results program of the National Cancer Institute.

Keywords: Interval-censored data; Heteroscedastic survival; Linear dependent tailfree process; Median regression; Spatial data; Stratified AFT model.

4.1 Introduction

Spatially correlated survival data are commonly observed in biomedical and epidemiological studies. For example, in cancer data arising from the Surveillance Epidemiology and End Results (SEER) program of the National Cancer Institute, survival times of patients from the same or adjacent counties are often expected to be more alike than those from distant counties due to region-specific similarities in environments and treatment resources. Simultaneously modeling the risk factors and geographical pattern that explain the differences in survival probabilities is of particular importance for establishing policies to improve national health systems. Choosing an appropriate survival model that takes into account the spatial dependence across counties is a must for valid statistical inferences.

The use of parametric and semiparametric hierarchical frailty survival models has become quite popular for analyzing spatially correlated survival data. For example, Henderson et al. (2002) proposed a multivariate gamma frailty proportional hazards (PH) model for incorporating spatial dependence; Li and Ryan (2002) extended the ordinary frailty PH model by allowing frailties accommodating spatial correlation via spatial covariance functions; Banerjee et al. (2003) considered a parametric Weibull PH model with both lattice and point-referenced frailty specifications; Banerjee and Carlin (2003) studied semiparametric PH frailty models using a mixture of beta densities baseline; Banerjee and Dey (2005) developed a Bayesian hierarchical model

for capturing spatial heterogeneity within the framework of proportional odds (PO); Hennerfeind et al. (2006) extended the PH model with additive covariates including both time and spatial effects; Zhao et al. (2009) considered all the three commonly used survival models (i.e. accelerated failure time, PH and PO) with conditionally autoregressive (CAR) prior on frailties, where the baseline function is assumed to have a mixture of Polya trees prior; Zhang and Lawson (2011) and Wang et al. (2012) developed parametric and semiparametric spatial frailty accelerated failure time (AFT) models, respectively. However, all the above survival models assume a homogeneous baseline distribution $S_0(\cdot)$ (either parametric or nonparametric) coupled with a parametric part of the model (e.g. PO, AFT, PH).

Stratification is commonly used in semiparametric survival modeling when the shape of the baseline hazard changes dramatically with a discrete covariate, such as clinic identification. Stratified analyses allow a completely different baseline hazard for each strata. The partial likelihood of Cox (1975) is easily modified to accommodate stratified variables and this can be implemented, e.g. via the `STRATA` subcommand in SAS procedure `PHREG`. Unfortunately, tests concerning the stratification variable(s) cannot be carried out and continuous variables cannot be stratified on. Recently, Zhao and Hanson (2011) generalized the stratified PH model to allow spatially smooth baseline hazards over an areal map. Similarly, Hanson et al. (2012) considered longitudinally smoothed baseline hazards in a stratified PH model. In both cases, spatial or longitudinal smoothing improved predictive performance over completely independent baseline hazards or a common baseline hazard.

Semiparametric AFT models, the natural competitor to PH, have been widely studied, including Christensen and Johnson (1988), Kuo and Mallick (1997), Walker and Mallick (1999), Kottas and Gelfand (2001), Hanson and Johnson (2002, 2004) and Hanson (2006a), typically using Dirichlet process mixtures or Polya trees for the homogeneous baseline distribution. However, very limited work has been completed

to date on stratified AFT models. Marginal models (Chiou et al., 2015) only allow for discrete stratification and average over the strata, giving a population-averaged interpretation for the acceleration factors. Significance of semiparametric risk factors can be watered down and often the conditional interpretation is more relevant as, e.g. in the case of stratifying on gender, the gender is known and should not be averaged over. Conditionally stratified AFT has been considered, but termed "AFT with heteroscedastic error." In most of these models the error term in the log-linear model specifying the survival times is simply a single index model of the covariates times a residual, e.g. Pang et al. (2015). In such cases covariates affect the location and scale of the log-survival times, but the overall shape of the density remains static across covariates. To relax the assumption of common baseline distributional shape, Hanson and Jara (2013) discussed an AFT model with general heteroscedastic error terms. Their model can be viewed as a particular kind of censored quantile regression (Koenker, 2008), but with heteroscedastic error that changes with covariate levels. However, the finite sample performance of their model has not been well studied, and how to incorporate spatial dependence into the model remains a question.

In this paper we propose a generalized accelerated failure time spatial frailty model which allows stratification on continuous or categorical covariates, as well as per-variable tests for whether stratification is necessary via novel approximate Bayes factors. This model extends the linear dependent tailfree process of Jara and Hanson (2011) to the interval-censored data setting, and also incorporates exchangeable or spatially varying areal-level frailties. In contrast to the aforementioned index models (Pang et al., 2015), the entire shape of the residual density changes smoothly with strata covariates. Furthermore, the residual density is median-zero, providing greatly enhanced interpretation of non-strata variables in terms of how median survival changes. The proposed model includes the traditional parametric and semiparametric AFT spatial frailty models as special cases; therefore, it provides a means of testing

the fit of simpler, commonly-used models. A highly novel aspect of this work is the development of fast, approximate Bayes factors based only on a fit of the full model for testing the adequacy of many important, commonly-fit reduced models. Methods of obtaining posterior inference are carefully detailed and a freely-available function `frailtyGAFT` is provided in the R package `spBayesSurv` that provides relevant output including per-variable tests for stratification covariates. All Markov chain Monte Carlo (MCMC) relies on a thoroughly-tested self-contained algorithm; no MCMC tuning is required. In addition, functions for obtaining survival curves, etc. are available. The model accommodates general interval-censored data, including standard right-censored data as well as special cases such as Case I (current status data) and Case II interval-censoring.

The rest of the paper is organized as follows. Section 2 describes the proposed model together with the MCMC implementation of posterior inference and Bayesian hypothesis tests. Section 3 presents simulation studies to evaluate the performance of the proposed model. Section 4 provides a detailed analysis of the SEER prostate cancer data. The paper is concluded by a discussion in Section 5.

4.2 Generalized accelerated failure time spatial frailty model

Standard survival modeling

Let t_{ij} be a random event time associated with the jth subject in the ith county and $\mathbf{z}_{ij} = (z_{1ij}, \ldots, z_{pij})'$ be a related p-dimensional vector of covariates, $j = 1, \ldots, m_i, i = 1, \ldots, m$. Let δ_{ij} be a censoring indicator equaling 1 if t_{ij} is an observed event time and equaling 0 if the event time is censored to lie in the interval $(l_{ij}, u_{ij}]$ with $l_{ij} < u_{ij}$, where $l_{ij} = 0$ ($u_{ij} = \infty$) corresponds to left (right) censoring. The event times from the same county of residence are expected to be correlated due to sharing common unobserved characteristics, such as region-specific similarities in environments and

treatment resources. To incorporate spatial dependence, a traditional way is to introduce a random effect (frailty) into the linear predictor of semiparametric survival models. In this paper we consider the AFT spatial frailty model, specified as

$$S_{\mathbf{z}_{ij}}(t) = S_0\left(e^{-\mathbf{z}'_{ij}\boldsymbol{\beta}-v_i}t\right), \tag{4.1}$$

where v_i is an unobserved frailty associated with county i, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is a vector of regression coefficients, and $S_0(t)$ is the baseline survival function corresponding to $\mathbf{z}_{ij} = \mathbf{0}$ and $v_i = 0$. In practice, $\mathbf{z}_{ij}$s are usually normalized so that $S_0(t)$ can serve as a reference. Often $S_0(t)$ is assumed to be a static parametric or nonparametric survival function, free of covariates. However this assumption implies that the resulting survival curves are not allowed to cross for different covariates, which can be unrealistic in practical applications (De Iorio et al., 2009). In what follows, we present a generalized AFT spatial frailty model, where $S_0(t)$ is allowed to flexibly vary with covariates under some identifiability constraints, yielding a particular kind of stratified AFT model that allows stratification on continuous or categorical covariates. In comparison to standard AFT spatial survival models, the proposed model has increased flexibility while retaining interpretability of model parameters.

Spatial frailty modeling

For modeling the spatial frailties, we consider a version of conditionally autoregressive prior (Besag, 1974). Given frailties $v_1, \ldots, v_m$ associated with counties $1, \ldots, m$, we define an $m \times m$ symmetric *proximity matrix* W with the ijth entry w_{ij} representing some type of connection between counties i and j, where w_{ii} is customarily set to 0. Typically $w_{ij} = 1$ if i and j share a common boundary and zero otherwise; this is the measure used in this paper. Note that w_{ij} could instead reflect some fashion of meaningful "distance" between counties, e.g. Mahalanobis distance of median household income or a ruralness measure between counties i and j. The frailties are

assumed to follow the conditional independence assumption

$$v_i|\{v_j\}_{j\neq i}, \tau \sim N\left(\sum_{j=1}^{m} w_{ij}v_j/w_{i+},\ \tau^2/w_{i+}\right),\ i = 1, \ldots, m, \tag{4.2}$$

where $w_{i+} = \sum_{j=1}^{m} w_{ij}$ and τ is a scale parameter. A little algebra yields the joint density of frailties as

$$p(\mathbf{v}) = p(v_1, \ldots, v_m) \propto \left(\frac{1}{\tau^2}\right)^{\frac{m-1}{2}} \exp\left\{-\frac{1}{2\tau^2}\mathbf{v}'(D_w - W)\mathbf{v}\right\}, \tag{4.3}$$

where $\mathbf{v} = (v_1, \ldots, v_m)'$ and D_w is an $m \times m$ diagonal matrix with $(D_w)_{ii} = w_{i+}$. Note this joint density is improper since we can add any constant to all of the v_is and (4.3) is unaffected; that is, the v_is are not centered. We consider the constraint $\sum_{i=1}^{m} v_i = 0$, which provides the needed centering to avoid identifiability issues. Under this constraint, the conditionally autoregressive prior becomes

$$v_i|\{v_j\}_{j\neq i} \sim N\left(\sum_{j\neq i}^{m-1} w^*_{ij}v_j/w^*_{i+},\ \tau^2/w^*_{i+}\right),\ i = 1, \ldots, m-1, \tag{4.4}$$

where $w^*_{i+} = w_{i+} + w_{m+} + 2w_{mi}$ and $w^*_{ij} = w_{ij} - w_{m+} - w_{mj} - w_{mi}$. We prefer the incoporation of the sum-to-zero constraint directly into the prior and model to avoid the *ad hoc* adjustments typically made within the MCMC scheme itself.

Mixture of linear dependent tailfree processes prior for $S_0(t)$

We allow the baseline survival function $S_0(t)$ to depend on certain covariates, say a q-dimensional vector $\mathbf{x}_{ij}$ which is often a subset of $\mathbf{z}_{ij}$, yielding the generalized AFT (GAFT) spatial frailty model

$$S_{\mathbf{z}_{ij}}(t) = S_{0,\mathbf{x}_{ij}}\left(e^{-\mathbf{z}'_{ij}\beta - v_i}t\right), \tag{4.5}$$

where v_i has the CAR prior (4.3) with $\sum_{i=1}^{m} v_i = 0$. For ease of handling identifiability issues, we rewrite the model (4.5) as:

$$y_{ij} = \log(t_{ij}) = \tilde{\mathbf{z}}'_{ij}\tilde{\boldsymbol{\beta}} + v_i + \epsilon_{ij}, \tag{4.6}$$

where $\tilde{\mathbf{z}}_{ij} = (1, \mathbf{z}'_{ij})'$ includes an intercept, $\tilde{\boldsymbol{\beta}} = (\beta_0, \boldsymbol{\beta}')'$ is a vector of corresponding coefficients, ϵ_{ij} is a heteroscedastic error term independent with v_i, and $P(e^{\beta_0+\epsilon_{ij}} > t|\mathbf{x}_{ij}) = S_{0,\mathbf{x}_{ij}}(t)$. Here we assume

$$\epsilon_{ij}|G_{\mathbf{x}_{ij}} \overset{ind.}{\sim} G_{\mathbf{x}_{ij}},$$

where $G_{\mathbf{x}}$ is a probability measure defined on $\mathbb{R}$ for every $\mathbf{x} \in \mathcal{X}$; this defines a model for the entire collection of probability measures $\mathcal{G}_{\mathcal{X}} = \{G_{\mathbf{x}} : \mathbf{x} \in \mathcal{X}\}$ so that each element is allowed to smoothly change with the covariates $\mathbf{x}$. In particular, we consider a mixture of linear dependent tailfree processes (LDTFP) prior (Jara and Hanson, 2011) for $\mathcal{G}_{\mathcal{X}}$,

$$\mathcal{G}_{\mathcal{X}}|L, h, \alpha, \rho, \sigma^2 \sim \text{LDTFP}(h, \alpha, \rho, \Pi^{L,\sigma}), \sigma^2 \sim P(d\sigma^2), \tag{4.7}$$

where $L \in \mathbb{N}^+$ is the level of specification of the process, $h(\cdot)$ is a strictly increasing function linking the process with covariates, $\alpha \in \mathbb{R}^+$ is a precision parameter controlling the prior variability of the process, $\rho(\cdot)$ is an increasing function, $\Pi^{L,\sigma}$ is an L-level sequence of nested partitioning sets of $\mathbb{R}$ depending on a scale parameter $\sigma \in \mathbb{R}^+$, and P is a prior probability measure for σ^2. In this paper we consider an inverse gamma prior on σ^2, say $\sigma^{-2} \sim \Gamma(a_\sigma, b_\sigma)$, where $\Gamma(a, b)$ refers to the gamma distribution with shape a and rate b.

We focus on an LDTFP centered at a normal distribution Φ_σ with mean 0 and variance σ^2, that is, $E(G_{\mathbf{x}}) = N(0, \sigma^2)$ for every $\mathbf{x} \in \mathcal{X}$. The LDTFP successively partitions the real line into finer and finer partitions; each refinement of a partition produces the next level of sets in the process. At level l, the process partitions $\mathbb{R}$ into 2^l intervals $B_{l,k} = \left(\Phi_\sigma^{-1}((k-1)2^{-l}), \Phi_\sigma^{-1}(k2^{-l})\right]$, $k = 1, \ldots, 2^l$, with $B_{l,2^l}$ being right-open, so that an L-level set of nested partitions is defined as $\Pi^{L,\sigma} = \{B_{l,k} : k = 1, \ldots, 2^l, l = 1, \ldots, L\}$. Note that $B_{l,k} = B_{l+1,2k-1} \cup B_{l+1,2k}$. Given that an observation is in set k at level l, say $B_{l,k}$, it could then be in either $B_{l+1,2k-1}$ or $B_{l+1,2k}$ at level $l+1$ with conditional probability $Y_{l+1,2k-1}$ or $Y_{l+1,2k}$ respectively.

Clearly they must sum to one for every $\mathbf{x} \in \mathcal{X}$, and so we consider logistic regression for each of these probabilities, allowing the entire shape of the density to change with covariates. Specifically, we assume $Y_{l+1,2k-1}(\mathbf{x}) = h(\tilde{\mathbf{x}}'\boldsymbol{\gamma}_{l,k})$ and $Y_{l+1,2k}(\mathbf{x}) = 1 - h(\tilde{\mathbf{x}}'\boldsymbol{\gamma}_{l,k})$, where $\tilde{\mathbf{x}} = (1, \mathbf{x}')'$ includes an intercept, $\boldsymbol{\gamma}_{l,k} = (\gamma_{l,k,0}, \ldots, \gamma_{l,k,q})'$ is a vector of coefficients, and $h(\cdot) = \frac{\exp\{\cdot\}}{1+\exp\{\cdot\}}$. Finally, there are $2^L - 1$ regression coefficient vectors $\boldsymbol{\gamma} = \{\boldsymbol{\gamma}_{l,k}\}$, e.g. for $L = 3$, $\{\boldsymbol{\gamma}_{0,1}, \boldsymbol{\gamma}_{1,1}, \boldsymbol{\gamma}_{1,2}, \boldsymbol{\gamma}_{2,1}, \boldsymbol{\gamma}_{2,2}, \boldsymbol{\gamma}_{2,3}, \boldsymbol{\gamma}_{2,4}\}$. Let $\mathbf{X}$ be the $n \times (q+1)$ design matrix of covariates $\tilde{\mathbf{x}}_{ij}$s, where $n = \sum_{i=1}^{m} m_i$. Following Jara and Hanson (2011), each $\boldsymbol{\gamma}_{l,k}$ is assigned an independent normal g-prior (Zellner, 1983), $\boldsymbol{\gamma}_{l,k} \sim N_{q+1}\left(\mathbf{0}, \frac{2n}{\alpha\rho(l+1)}(\mathbf{X}'\mathbf{X})^{-1}\right)$, where $\rho(l) = l^2$. Furthermore, the LDTFP can be specified that for every $\mathbf{x} \in \mathcal{X}$, $G_{\mathbf{x}}$ is almost surely a median-zero probability measure by setting $\boldsymbol{\gamma}_{0,1} \equiv \mathbf{0}$. This is important to avoid identifiability issues.

The precision parameter $\alpha \in \mathbb{R}^+$ controls how closely the random distribution $G_{\mathbf{x}}$ follows Φ_σ in terms of L_1 distance (Hanson et al., 2008). Large values of α indicate a strong belief that ϵ_{ij}s are closely *iid* from Φ_σ. Smaller values of α, on the other hand, allow more pronounced deviations of $G_{\mathbf{x}}$ from Φ_σ. We consider a gamma prior on α, say, $\alpha \sim \Gamma(a_0, b_0)$, as suggested in Jara and Hanson (2011); the full conditional distribution for α is also a gamma distribution. As for the choice of L, we typically consider $L \approx \log_2(n/n_0)$, where n is the sample size and n_0 (usually from 5 to 10) is a "typical" number of observations falling into each set at level L (Hanson, 2006a). However, this choice is conservative for the LDTFP relative to Polya trees (Jara and Hanson, 2011). Hanson (2006a) observed that the LPML remains essentially unchanged or gets slightly worse after a certain level L; this is confirmed for the LDTFP in (Zhou et al., 2015a, Chapter 2).

Based on the above LDTFP specification, given $\boldsymbol{\gamma}$, σ and $\mathbf{x}$, $G_{\mathbf{x}}$ is known. Define the function $k_\sigma(x)$ to be the index $k \in \{1, \ldots, 2^L\}$ such that x falls into set $B_{L,k}$, i.e. $k_\sigma(x) = \lceil 2^L \Phi_\sigma(x) \rceil$, where $\lceil x \rceil$ is the ceiling function, the smallest integer greater

than or equal to x. Further define probability $p_{\mathbf{x}}(k)$ for $k = 1, \ldots, 2^L$ as

$$p_{\mathbf{x}}(k) = G_{\mathbf{x}}\{B_{L,k}|\boldsymbol{\gamma}, \sigma\} = \prod_{l=1}^{L} Y_{l,\lceil k2^{l-L}\rceil}(\mathbf{x}).$$

The resulting density of $\epsilon_{ij}|\mathbf{x}_{ij}$ is given by

$$g_{\mathbf{x}_{ij}}(e) = 2^L \phi_\sigma(e) p_{\mathbf{x}_{ij}}\{k_\sigma(e)\}, \tag{4.8}$$

where ϕ_σ is the density function corresponding to Φ_σ. The cumulative distribution function associated with $g_{\mathbf{x}_{ij}}(e)$ is given by

$$G_{\mathbf{x}_{ij}}(e) = p_{\mathbf{x}_{ij}}\{k_\sigma(e)\}\left\{2^L\Phi_\sigma(e) - k_\sigma(e)\right\} + \sum_{k=1}^{k_\sigma(e)} p_{\mathbf{x}_{ij}}(k). \tag{4.9}$$

As shown by Jara and Hanson (2011), the LDTFP has appealing theoretical properties such as continuity as a function of the covariates, large support on the space of conditional density functions, straightforward posterior computation relying on algorithms for fitting generalized linear models, and the process closely matches conventional Polya tree priors (see, e.g., Hanson, 2006a) at each value of the covariate, which justify its choice here.

Regarding the priors for regression coefficients $\tilde{\boldsymbol{\beta}}$ and τ^2, typically $\tilde{\boldsymbol{\beta}} \sim N_p(\mathbf{m}_0, \mathbf{S}_0)$ independent of $\tau^{-2} \sim \Gamma(a_\tau, b_\tau)$ are considered; these can both be chosen to relatively vague. In summary the proposed GAFT spatial frailty model takes the following hierarchical structure:

$$\begin{aligned}
(1-\delta_{ij})|t_{ij} &= I(l_{ij} < t_{ij} \leq u_{ij}), j = 1, \ldots, m_i, i = 1, \ldots, m \\
y_{ij} = \log(t_{ij}) &= \tilde{\mathbf{z}}_{ij}'\tilde{\boldsymbol{\beta}} + v_i + \epsilon_{ij}, j = 1, \ldots, m_i, i = 1, \ldots, m \\
\epsilon_{ij}|\boldsymbol{\gamma}, \sigma &\overset{ind.}{\sim} g_{\mathbf{x}_{ij}}(\cdot), j = 1, \ldots, m_i, i = 1, \ldots, m \\
\boldsymbol{\gamma}_{l,k}|\alpha &\overset{ind.}{\sim} N_{q+1}(\mathbf{0}, \frac{2n}{\alpha(l+1)^2}(\mathbf{X}'\mathbf{X})^{-1}), k = 1, \ldots, 2^l, l = 1, \ldots, L-1 \\
\alpha &\sim \Gamma(a_0, b_0), \quad \sigma^{-2} \sim \Gamma(a_\sigma, b_\sigma) \\
v_i|\{v_j\}_{j\neq i}, \tau &\sim N\left(\sum_{j=1}^{m} w_{ij}v_j/w_{i+},\ \tau^2/w_{i+}\right), i = 1, \ldots, m \\
\tau^{-2} &\sim \Gamma(a_\tau, b_\tau), \quad \tilde{\boldsymbol{\beta}} \sim N_{p+1}(\mathbf{m}_0, \mathbf{S}_0).
\end{aligned} \tag{4.10}$$

We next describe a Markov chain Monte Carlo (MCMC) scheme for obtaining posterior inference, which can be implemented using the function `frailtyGAFT` available in the R package `spBayesSurv`.

Posterior computation

In this section we describe an MCMC sampling algorithm for the proposed model, arrived at after a considerable amount of trial and error, in terms of what provides reasonable mixing and speed. By now, MCMC schemes are a standard part of the Bayesian statistician's toolbox, we refer the reader to Robert and Casella (2005) for an overview. Posterior MCMC samples for the model parameters are used for all inference of interest. A complete description of updating steps is available in the supplementary material.

Let $\Omega = (\mathbf{y}_c, \tilde{\boldsymbol{\beta}}, \mathbf{v}, \tau^2, \sigma^2, \boldsymbol{\gamma}, \alpha)$ denote collectively the model parameters to be updated, where $\mathbf{y}_c = \{y_{ij} : \delta_{ij} = 0\}$ are censored log-survival times. The $y_{ij} \in \mathbf{y}_c$ are sampled using the single-variable slice sampling method (Neal, 2003), where the "stepping out" procedure is performed for finding an interval around the current iteration value and "shrinkage" procedure is applied for sampling from the interval. The median regression coefficients $\tilde{\boldsymbol{\beta}}$ may be conveniently updated using adaptive Metropolis-Hastings (Haario et al., 2001) with multivariate Gaussian proposals. However this method suffers extremely low acceptance rates when the LDTFP is far from its centering normal distribution. We instead use a single-variable slice sampling step to update each component of $\tilde{\boldsymbol{\beta}}$. The updating of spatial effects $\mathbf{v}$ and baseline variance σ are both easily done again via single-variable slice sampling steps. For the LDTFP regression parameters $\gamma_{l,k}$, we utilize Metropolis-Hastings steps with Gaussian proposals based on iterative weighted least squares (Gamerman, 1997), recognizing that the $\gamma_{l,k}$ full conditionals are proportional to logistic regression likelihoods. The hyperparameter τ^2 and α are sampled according to their conjugate full

conditional distributions.

Given a set of posterior samples $\{\boldsymbol{\Omega}^{(s)}, s = 1, \ldots, S\}$, all the inference targets can be easily estimated. For example, the baseline survival function $S_{0,\mathbf{x}}(t) = P(e^{\beta_0+\epsilon} > t|\mathbf{x})$ given the covariate $\mathbf{x}$ is estimated by

$$S_{0,\mathbf{x}}(t) = \frac{1}{S}\sum_{s=1}^{S}\left\{1 - G_{\mathbf{x}}^{(s)}\left(\log t - \beta_0^{(s)}\right)\right\}, \tag{4.11}$$

where $G_{\mathbf{x}}^{(s)}(\cdot)$ is given in (4.9) with all unknown parameters replaced by corresponding posterior values in the sth iterate.

Bayesian hypothesis testing

The proposed GAFT spatial frailty model includes the following as important special cases: an AFT spatial frailty model with nonparametric baseline where $G_{\mathbf{x}} = G_{\mathbf{x}'}$ for all $\mathbf{x} = \mathbf{x}'$ and parametric baseline model $G_{\mathbf{x}} = N(0, \sigma^2)$ for all $\mathbf{x} \in \mathcal{X}$. Hypothesis tests can be constructed based on the LDTFP coefficients $\{\boldsymbol{\gamma}_{l,k} : k = 1, \ldots, 2^l, l = 1, \ldots, L-1\}$, where $\boldsymbol{\gamma}_{l,k} = (\gamma_{l,k,0}, \ldots, \gamma_{l,k,q})'$. Let $\boldsymbol{\gamma}_{l,k,-j}$ denote the subvector of $\boldsymbol{\gamma}_{l,k}$ without element $\gamma_{l,k,j}$ for $j = 0, \ldots, q$. Set $\boldsymbol{\Upsilon}_j = (\gamma_{l,k,j}, k = 1, \ldots, 2^l, l = 1, \ldots, L-1)'$, $\boldsymbol{\Upsilon}_{-j} = (\boldsymbol{\gamma}'_{l,k,-j}, k = 1, \ldots, 2^l, l = 1, \ldots, L-1)'$ and $\boldsymbol{\Upsilon} = (\boldsymbol{\gamma}'_{l,k}, k = 1, \ldots, 2^l, l = 1, \ldots, L-1)'$. Testing the hypotheses $H_0 : \boldsymbol{\Upsilon}_{-0} = \mathbf{0}$ and $H_0 : \boldsymbol{\Upsilon} = \mathbf{0}$ leads to global comparisons of the proposed model with the above two special cases respectively. Similarly, we may also test the null hypothesis $H_0 : \boldsymbol{\Upsilon}_j = \mathbf{0}$ for the jth covariate effect of $\mathbf{x}$ on the baseline survival, $j = 1, \ldots, q$. We use Bayes factors to accomplish these hypotheses.

Suppose we wish to test $H_0 : \boldsymbol{\Upsilon}_j = \mathbf{0}$ versus $H_1 : \boldsymbol{\Upsilon}_j \neq \mathbf{0}$, for fixed $j \in \{1, \ldots, q\}$. The Bayes factor between hypotheses H_1 and H_0 is defined as

$$BF_{10} = \frac{\int \mathcal{L}(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi}) p(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi}) d(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi})}{\int \mathcal{L}(\boldsymbol{\Upsilon}_j = \mathbf{0}, \boldsymbol{\psi}) p_0(\boldsymbol{\psi}) d\boldsymbol{\psi}}, \tag{4.12}$$

where $\boldsymbol{\psi}$ is the remaining model parameters under the alternative, $p_0(\boldsymbol{\psi})$ and $p(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi})$ are the prior probability densities under H_0 and H_1 respectively, $\mathcal{D}$ is the observed

data, and $\mathcal{L}(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi})$ is the likelihood function. According to the Savage-Dickey density ratio expression (Dickey, 1971), if

$$p(\boldsymbol{\psi}|\boldsymbol{\Upsilon}_j = 0) = p_0(\boldsymbol{\psi}), \tag{4.13}$$

then BF_{10} can be written as

$$BF_{10} = \frac{p(\boldsymbol{\Upsilon}_j = 0)}{p(\boldsymbol{\Upsilon}_j = 0|\mathcal{D})}, \tag{4.14}$$

where $p(\boldsymbol{\Upsilon}_j) = \int p(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi}) d\boldsymbol{\psi}$ and $p(\boldsymbol{\Upsilon}_j|\mathcal{D}) = \int p(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi}|\mathcal{D}) d\boldsymbol{\psi}$ are the marginal prior and posterior density of $\boldsymbol{\Upsilon}_j$ respectively under H_1. We show in Proposition 1 that the assumption (4.13) holds when precision parameter α is fixed.

Proposition 4.1. *Assume that* $\boldsymbol{\gamma}_{l,k}|\alpha \overset{ind.}{\sim} N_{q+1}\left(\mathbf{0}, \frac{2n}{\alpha\rho(l+1)}(\mathbf{X}'\mathbf{X})^{-1}\right)$ *under* H_1 *and* $\boldsymbol{\gamma}_{l,k,-j}|\alpha \overset{ind.}{\sim} N_q\left(\mathbf{0}, \frac{2n}{\alpha\rho(l+1)}(\mathbf{X}'_{-j}\mathbf{X}_{-j})^{-1}\right)$ *under* H_0, *where* α *is fixed and* $\mathbf{X}_{-j}$ *is the design matrix* $\mathbf{X}$ *excluding the* $(j+1)$*th column. Then the assumption (4.13) holds, and*

$$p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\alpha) = \prod_{l=1}^{L-1}\prod_{k=1}^{2^l} \phi\left(0\middle|0, \frac{2n}{\alpha\rho(l+1)}(\mathbf{X}'\mathbf{X})^{-1}_{jj}\right). \tag{4.15}$$

where $(\mathbf{X}'\mathbf{X})^{-1}_{jj}$ *is the* $(j+1, j+1)$*th element of* $(\mathbf{X}'\mathbf{X})^{-1}$, *and* $\phi(\cdot|\mu, \sigma^2)$ *denotes the normal density with mean* μ *and variance* σ^2.

Proof. Since $\boldsymbol{\gamma}_{l,k}|\alpha$ follows a multivariate normal, $(\boldsymbol{\gamma}_{l,k,-j}|\boldsymbol{\gamma}_{l,k,j} = 0, \alpha)$ still follows a multivariate normal distribution

$$\begin{aligned} p(\boldsymbol{\gamma}_{l,k,-j}|\boldsymbol{\gamma}_{l,k,j} = 0, \alpha) &\propto \exp\left\{-\frac{\alpha\rho(l+1)}{4n}\boldsymbol{\gamma}'_{l,k}(\mathbf{X}'\mathbf{X})\boldsymbol{\gamma}_{l,k}\right\} \\ &\propto \exp\left\{-\frac{\alpha\rho(l+1)}{4n}\boldsymbol{\gamma}'_{l,k,-j}(\mathbf{X}'_{-j}\mathbf{X}_{-j})\boldsymbol{\gamma}_{l,k,-j}\right\} \\ &\propto N_q\left(\mathbf{0}, \frac{2n}{\alpha\rho(l+1)}(\mathbf{X}'_{-j}\mathbf{X}_{-j})^{-1}\right). \end{aligned}$$

This implies that $p(\boldsymbol{\gamma}_{l,k,-j}|\boldsymbol{\gamma}_{l,k,j} = 0, \alpha) = p_0(\boldsymbol{\gamma}_{l,k,-j}|\alpha)$ and by independence assumption $p(\boldsymbol{\Upsilon}_{-j}|\boldsymbol{\Upsilon}_j = 0, \alpha) = p_0(\boldsymbol{\Upsilon}_{-j}|\alpha)$. In addition, α is fixed and $\boldsymbol{\Upsilon}_{-j}$ is independent with all other parameters in $\boldsymbol{\psi}$, thus the assumption (4.13) holds. It is easy to evaluate $p(\boldsymbol{\Upsilon}_j = 0)$ by noting the properties of multivariate normal. □

Now consider putting a prior π and π_0 on α under H_1 and H_0 respectively, assuming the same priors on $\boldsymbol{\gamma}_{l,k}$ as Proposition 4.1. Then we have

$$\begin{aligned} p(\boldsymbol{\Upsilon}_{-j}|\boldsymbol{\Upsilon}_j = 0) &= \int p(\boldsymbol{\Upsilon}_{-j}|\boldsymbol{\Upsilon}_j = 0, \alpha)\frac{p(\boldsymbol{\Upsilon}_j = 0|\alpha)\pi(\alpha)}{p(\boldsymbol{\Upsilon}_j = 0)}d\alpha \\ &= \int p_0(\boldsymbol{\Upsilon}_{-j}|\alpha)\frac{p(\boldsymbol{\Upsilon}_j = 0|\alpha)\pi(\alpha)}{p(\boldsymbol{\Upsilon}_j = 0)}d\alpha \end{aligned} \tag{4.16}$$

and $p_0(\boldsymbol{\Upsilon}_{-j}) = \int p_0(\boldsymbol{\Upsilon}_{-j}|\alpha)\pi_0(\alpha)d\alpha$. To satisfy (4.13), we need equation (4.16) equal to $p_0(\boldsymbol{\Upsilon}_{-j})$, which holds when

$$\pi_0(\alpha) = \frac{p(\boldsymbol{\Upsilon}_j = 0|\alpha)\pi(\alpha)}{p(\boldsymbol{\Upsilon}_j = 0)}.$$

Let $\Gamma(\cdot|a, b)$ denote the density of $\Gamma(a, b)$ distribution. Taking $\pi(\cdot) = \Gamma(\cdot|a_0, b_0)$ yields that $\pi_0(\cdot) = \Gamma(\cdot|a_0 + 2^{L-1} - 1, b_0)$. However, π_0 puts too much probability on H_0 against H_1, so that H_0 can be hardly rejected. To avoid this undesirable situation, we take $\pi_0 = \pi$ and apply the generalized Savage-Dickey density ratio approach, proposed by Verdinelli and Wasserman (1995), which does not rely on the assumption (4.13). We show in Proposition 4.2 that BF_{10} can be written as a product of two quantities and both can be estimated from posterior simulation.

Proposition 4.2. *Assume the same priors on $\boldsymbol{\gamma}_{l,k}$ as Proposition 4.1 and additional prior on α as $\pi(\alpha) = \Gamma(\alpha|a_0, b_0)$ under both H_1 and H_0. Then given existence of all involved expectations, BF_{10} can be written as*

$$BF_{10} = \{p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\mathcal{D})\}^{-1}\left\{E\left[\frac{1}{p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\alpha)}\right]\right\}^{-1}, \tag{4.17}$$

where the expectation is with respect to $p(\alpha|\boldsymbol{\Upsilon}_j = \mathbf{0}, \mathcal{D})$.

Proof. Note that $p(\boldsymbol{\Upsilon}_j = \mathbf{0}, \boldsymbol{\psi}|\mathcal{D})\int \mathcal{L}(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi})p(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi})d(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi}) = \mathcal{L}(\boldsymbol{\Upsilon}_j = \mathbf{0}, \boldsymbol{\psi}) \times$

$p(\boldsymbol{\Upsilon}_0, \boldsymbol{\psi})$ and $p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\boldsymbol{\psi}) = p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\alpha)$, we have

$$\begin{aligned}
BF_{10}^{-1} &= p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\mathcal{D}) \int \frac{\mathcal{L}(\boldsymbol{\Upsilon}_j = \mathbf{0}, \boldsymbol{\psi}) p(\boldsymbol{\psi})}{p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\mathcal{D}) \int \mathcal{L}(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi}) p(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi}) d(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi})} d\boldsymbol{\psi} \\
&= p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\mathcal{D}) \int \frac{\mathcal{L}(\boldsymbol{\Upsilon}_j = \mathbf{0}, \boldsymbol{\psi}) p(\boldsymbol{\psi}) p(\boldsymbol{\psi}|\boldsymbol{\Upsilon}_j = \mathbf{0}, \mathcal{D})}{p(\boldsymbol{\Upsilon}_j = \mathbf{0}, \boldsymbol{\psi}|\mathcal{D}) \int \mathcal{L}(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi}) p(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi}) d(\boldsymbol{\Upsilon}_j, \boldsymbol{\psi})} d\boldsymbol{\psi} \\
&= p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\mathcal{D}) \int \frac{p(\boldsymbol{\psi}) p(\boldsymbol{\psi}|\boldsymbol{\Upsilon}_j = \mathbf{0}, \mathcal{D})}{p(\boldsymbol{\Upsilon}_j = \mathbf{0}, \boldsymbol{\psi})} d\boldsymbol{\psi} \\
&= p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\mathcal{D}) \int \int \frac{1}{p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\boldsymbol{\psi})} p(\boldsymbol{\psi}_{-\alpha}, \alpha|\boldsymbol{\Upsilon}_j = \mathbf{0}, \mathcal{D}) d\boldsymbol{\psi}_{-\alpha} d\alpha \\
&= p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\mathcal{D}) \int \int \frac{1}{p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\alpha)} p(\boldsymbol{\psi}_{-\alpha}, \alpha|\boldsymbol{\Upsilon}_j = \mathbf{0}, \mathcal{D}) d\boldsymbol{\psi}_{-\alpha} d\alpha \\
&= p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\mathcal{D}) \int \frac{1}{p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\alpha)} p(\alpha|\boldsymbol{\Upsilon}_j = \mathbf{0}, \mathcal{D}) d\alpha,
\end{aligned}$$

where $(\boldsymbol{\psi}_{-\alpha}, \alpha) = \boldsymbol{\psi}$. □

As noted by Raftery (1996), only a crude approximation of Bayes factor is needed. Thus we estimate the marginal posterior density $p(\boldsymbol{\Upsilon}_j|\mathcal{D})$ by an $M = 2^L - 2$ dimensional multivariate normal distribution using MCMC posterior samples. That is, we assume $p(\boldsymbol{\Upsilon}_j|\mathcal{D}) \approx \phi_M(\boldsymbol{\Upsilon}_j; \hat{\mathbf{m}}_j, \hat{\mathbf{S}}_j)$, where $\phi_M(\cdot; \mathbf{m}, \mathbf{S})$ is a multivariate normal density with mean $\mathbf{m}$ and covariance matrix $\mathbf{S}$, and $\hat{\mathbf{m}}_j$ and $\hat{\mathbf{S}}_j$ are estimated by the sample mean and covariance of the posterior sample for $\boldsymbol{\Upsilon}_j$. To avoid of drawing a sample from $p(\alpha|\boldsymbol{\Upsilon}_j = \mathbf{0}, \mathcal{D})$, we assume that $p(\alpha|\boldsymbol{\Upsilon}_j = \mathbf{0}, \mathcal{D}) \approx p(\alpha|\mathcal{D})$ under H_0. We then approximate $\{E\left[1/p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\alpha)\right]\}^{-1}$ by $p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\hat{\alpha})$, where $\hat{\alpha}$ is the posterior mean of α. Thus, we can estimate the Bayes factor by

$$\hat{BF}_{10} = \frac{p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\hat{\alpha})}{\phi_M(\boldsymbol{\Upsilon}_j = \mathbf{0}; \hat{\mathbf{m}}_j, \hat{\mathbf{S}}_j)}, \tag{4.18}$$

where $p(\boldsymbol{\Upsilon}_j = \mathbf{0}|\alpha)$ is given in (4.15).

4.3 Simulation studies

We performed simulations to illustrate and assess the proposed approach using the provided `R` package `spBayesSurv`. The data were simulated from the GAFT model

(4.6), where $\tilde{\mathbf{z}}_{ij} = (1, z_{1ij}, z_{2ij})'$ with $z_{1ij} \overset{iid}{\sim} N(0,1)$ and $z_{2ij} \overset{iid}{\sim}$ Bernoulli(0.5), $\tilde{\boldsymbol{\beta}} = (\beta_0, \beta_1, \beta_2)' = (-1, 1, -0.5)'$, v_i follows the CAR model (4.3) with $\tau^2 = 0.1$ and W being the Louisiana proximity matrix used in the SEER data analysis, and $i = 1, \ldots, 64$, $j = 1, \ldots, 10$. We considered three different distribution settings for the error term ϵ_{ij}:

Scenario I: $\epsilon_{ij}|x_{ij} \overset{ind.}{\sim} \begin{cases} N(0, 0.8^2) & \text{if } x_{ij} = 0 \\ 0.5N(-1, 0.5^2) + 0.5N(1, 0.5^2) & \text{if } x_{ij} = 1 \end{cases}$

Scenario II: $\epsilon_{ij} \overset{i.i.d}{\sim} 0.5N(-1, 0.5^2) + 0.5N(1, 0.5^2)$

Scenario III: $\epsilon_{ij} \overset{i.i.d}{\sim} N(0, 0.8^2)$

where $x_{ij} = z_{2ij}$. Note that the first scenario is not a particular case of the proposed model; the second and third scenarios are included to examine the behaviour of the proposed approach when a standard parametric or semiparametric AFT model (with covariate-free error term) is correct. Non-informative right-censoring times were simulated from a Uniform$(0.5, 1)$ distribution so that the censoring rate is around 25%. For each simulation scenario, a total of 500 replicates of the dataset were generated. We then fitted the proposed model with both covariates included in the LDTFP modeling part, i.e. $\tilde{\mathbf{x}}_{ij} = \tilde{\mathbf{z}}_{ij}$, using the following prior settings: $L = 4$, $a_0 = 5$, $b_0 = 1$, $\mathbf{m}_0 = \mathbf{0}_3$, $\mathbf{S}_0 = 10^5\mathbf{I}_3$, $a_\sigma = b_\sigma = 2.001$, $a_\tau = b_\tau = 0.1$. For each MCMC algorithm, $10,000$ scans were thinned from $100,000$ after a burn-in period of $10,000$ iterations.

Table 4.1 presents the proportions of Bayes factor BF_{10} greater than 3, 10 and 30 for the hypotheses discussed in Section 4.2. The results demonstrate that the proposed Bayes factor is able to identify which covariate truly affects the shape of the error term and which one does not, with very low wrong-decision rates across all the scenarios. The cutpoint 3 gives around 0.05 type I error in Scenarios I and II.

Table 4.1 Simulation data. The proportions of BF_{10} greater than 3, 10 and 30, respectively, across Monte Carlo simulations.

Scenario	H_0	% $BF_{10} > 3$	% $BF_{10} > 10$	% $BF_{10} > 30$
I	$\boldsymbol{\Upsilon}_1 = \mathbf{0}$	0.056	0.004	0.004
	$\boldsymbol{\Upsilon}_2 = \mathbf{0}$	1	1	1
II	$\boldsymbol{\Upsilon}_1 = \mathbf{0}$	0.052	0.016	0.008
	$\boldsymbol{\Upsilon}_2 = \mathbf{0}$	0.018	0.008	0.004
	$\boldsymbol{\Upsilon}_{-0} = \mathbf{0}$	0.022	0.004	0.002
	$\boldsymbol{\Upsilon} = \mathbf{0}$	1	1	1
III	$\boldsymbol{\Upsilon}_1 = \mathbf{0}$	0.096	0.016	0
	$\boldsymbol{\Upsilon}_2 = \mathbf{0}$	0.080	0.010	0.002
	$\boldsymbol{\Upsilon}_{-0} = \mathbf{0}$	0.124	0.018	0
	$\boldsymbol{\Upsilon} = \mathbf{0}$	0.116	0.030	0.008

Table 4.2 Simulation data. True value, averaged bias (BIAS) and posterior standard deviation (PSD) of each point estimate (i.e. posterior mean), standard deviation (across Monte Carlo simulations) of the point estimate (SD-Est) and coverage probability (CP) for the 95% credible interval.

Scenario	Parameter	True	BIAS	PSD	SD-Est	CP
I	β_0	-1.0	0.011	0.069	0.061	0.966
	β_1	1.0	0.005	0.056	0.050	0.958
	β_2	-0.5	0.010	0.129	0.104	0.976
	τ^2	0.1	0.034	0.055	0.045	0.978
II	β_0	-1.0	0.019	0.126	0.105	0.966
	β_1	1.0	-0.005	0.078	0.068	0.970
	β_2	-0.5	0.005	0.171	0.150	0.966
	τ^2	0.1	0.029	0.052	0.041	0.982
III	β_0	-1.0	0.004	0.062	0.055	0.980
	β_1	1.0	0.004	0.048	0.046	0.960
	β_2	-0.5	-0.001	0.085	0.076	0.968
	τ^2	0.1	0.039	0.061	0.047	0.982

Table 4.2 summaries the inference results for regression parameters, including the averaged bias (BIAS) and posterior standard deviation (PSD) of each point estimate, the Monte Carlo standard deviation of the point estimate (SD-Est) and the coverage probability (CP) of the 95% highest posterior density interval. The results show that the point estimates (i.e. posterior means) of $\tilde{\boldsymbol{\beta}}$ are almost unbiased under all

three scenarios, while the point estimates of τ^2 are positively biased. This is not surprising, as the posterior distribution of τ^2 is grossly right-skewed, in which case the mode would be an ideal choice for the point estimate. By using the posterior mode (calculated from a kernel-smoothed density) as a point estimate of τ^2, the averaged biases reduce to 0.007, 0.006 and 0.010 under Scenarios I, II and III, respectively. The PSD values are all slightly greater than, but fairly close to the SD-Est values, suggesting that the posterior standard deviation is an appropriate estimator of the frequentist standard error. The CP values are all close to the nominal 95% level or slightly greater. Here all covariates are included in the LDTFP baseline survival function regardless of their significance. Note that the use of Bayes factors allows us to detect which covariates need to be stratified on and how significant they are. Based on the test results, we can always remove the covariates that have $BF_{10} < 1$ from the LDTFP modeling. For this reason, in the online material we also report the parameter inferences with the LDTFP specified according to the truth, i.e., $\tilde{\mathbf{x}}_{ij} = (1, z_{2ij})'$ under Scenario I and $\tilde{\mathbf{x}}_{ij} = 1$ under Scenarios II & III. The results reveal that the CP values for $\tilde{\boldsymbol{\beta}}$ are now around the nominal 95% except the CP for β_2 in Scenario I. It seems that our approach tends to slightly overestimate the standard deviation of the covariate coefficient when that covariate highly affects the baseline function. Figure 4.1 presents the average, across the 500 simulated data sets, of the fitted density and survival functions of log survival times for some specific covariate values. The results reveal that the proposed model is capable to capture the crossing behaviour of survival curves very well.

In Section 3 of online material, additional simulation results are presented for the sensitivity analyses on the prior of α and the choice of L. The parameter estimates and hypothesis tests are essentially not affected by the choices of hyperparameters in the prior of α, although we observe that the curve estimates under $\alpha \sim \Gamma(2, 2)$ are slightly closer to the truth than those under $\alpha \sim \Gamma(20, 2)$. Regarding the impact

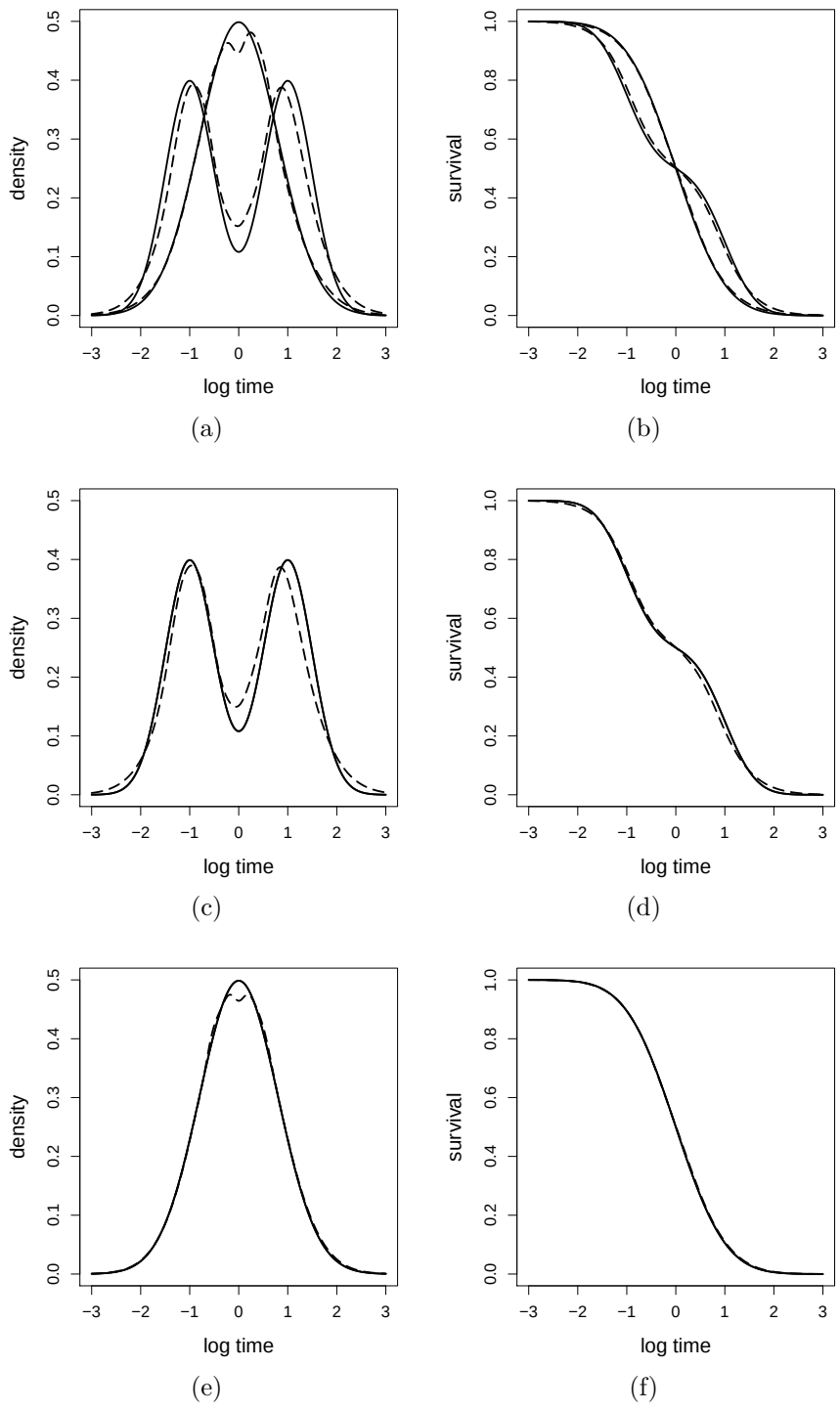


Figure 4.1 Simulated data. Mean, across simulations, of the posterior mean of the density functions (left three panels) and survival functions (right three panels) of log survival times under Scenario I (panel a and b), Scenario II (panel c and d) and Scenario III (panel e and f). The curves in each panel of a and b are for $(z, x) = (1.5, 1)$ (initially left curve) and $(z, x) = (1, 0)$. The other curves are for $(z, x) = (1, 0)$. The true curves are represented by continuous lines. The results under the proposed model are represented by dashed lines.

of L, we find that parameter and curve estimates are not sensitive to the choice of L, but hypothesis tests are more sensitive. For instance, when L increases from 4 to 5, the Bayes factor values become larger overall, especially for Scenario III (i.e. when the log-normal AFT is the truth), making our tests more conservative. For further comparison, we also fitted the censored quantile regression model (Portnoy, 2003) using the function `crq` available in the `quantreg` R package (Koenker, 2008) under Scenario I, but with the data generated without frailties. Note that `crq` does not allow spatial information. In comparison to our approach, the results show that `crq` provides almost two times greater standard deviation estimates for non-intercept coefficients, and more troubling, the coverage probability for estimating β_2 is much lower than 95%. These findings inform us that ignoring heteroscedastic errors could result in badly overestimated standard deviations and low coverage probabilities.

4.4 Application to SEER prostate cancer data

We apply the proposed model to the prostate cancer survival data from the SEER program of the National Cancer Institute (see `http://seer.cancer.gov/`). The data set we consider consists of a cohort of 2999 men from the 64 counties of Louisiana, who have been diagnosed with prostate cancer in 2002, with follow-up continued through the end of 2011. In our analysis, the observed survival time is calculated as the amount of years from diagnosis to either death or the last follow-up, where death can be from any cause. By the end of 2011, 62.1% of patients who survived until the last follow-up are treated as right-censored. The observed survival time in years and county of residence at diagnosis are available for each individual. The individual-specific covariates at diagnosis include: age, race (white and black), SEER summary stage (localized/regional and distant), marital status at diagnosis (married and other), grade of tumor differentiation (well/moderately differentiated and poorly/not differentiated), where the first category in each above parenthesis is treated

Table 4.3 SEER data. Summary characteristics for Louisiana prostate cancer patients diagnosed in 2002 with follow-up continued through the end of 2011.

Continuous	Mean	Median	Std. Dev.
follow-up (yrs)	7.63	9.17	2.87
Age (yrs)	67.52	68.00	9.45
Categorical	**Level**	**Count**	**Proportion**
Race	White	2159	0.72
	Black	840	0.28
Stage	Local/Regional	2875	0.96
	Distant	124	0.04
Marital	Married	2330	0.78
	Other	669	0.22
Grade	well/moderately differentiated	2220	0.74
	poorly/not differentiated	779	0.26

as reference. Table 4.3 presents several summary statistics for the data. Since the effects of stage, marital status and grade on baseline survival functions are not significant (per-variable Bayes factor is less than 1) based our initial model fitting via the proposed approach, we exclude them in modeling baseline functions presented below. Thus we have 5-dimensional $\mathbf{z}_{ij}$ and 2-dimensional $\mathbf{x}_{ij}$.

We fit the proposed GAFT spatial frailty model using the corresponding variants of the algorithm described in Section 4.2 and similar prior specifications used in the simulation study. Based upon examination of trace plots for model parameters, we run a single chain of 350,000, where 10,000 scans are thinned after a burn-in period of 50,000. In Section 4 of supplementary material, we present the posterior trace plots for $\tilde{\boldsymbol{\beta}}$, α and σ^2, and their autocorrelation function (ACF) plots together with effective samples sizes. The Markov chain mixed reasonably well regardless of the high dimension of parameters in our model.

For comparison, we further fit a semiparametric AFT spatial frailty model, where the baseline survival is fitted using the LDTFP but with intercept only, and a GAFT model without frailties. The three models are compared using the log pseudo marginal

Table 4.4 SEER data. Posterior means (95% credible intervals) of fixed effects $\tilde{\boldsymbol{\beta}}$ from fitting the proposed model and spatial frailty semiparametric AFT. The LPML is also shown for each model. Results are based on standardized ages.

Covariates	GAFT/CAR (LPML=-4110.5)	AFT/CAR (LPML=-4115.6)	GAFT (LPML=-4116.3)
Intercept	2.85(2.74, 2.96)	2.87(2.77, 2.96)	2.87(2.79, 2.96)
Age	-0.55(-0.62, -0.48)	-0.56(-0.62, -0.49)	-0.55(-0.61, -0.49)
Race	-0.23(-0.35, -0.11)	-0.29(-0.40, -0.19)	-0.23(-0.34, -0.14)
Marital	-0.33(-0.45, -0.22)	-0.32(-0.43, -0.20)	-0.30(-0.40, -0.21)
Grade	-0.23(-0.33, -0.12)	-0.25(-0.36, -0.14)	-0.21(-0.31, -0.12)
Stage	-1.55(-1.76, -1.34)	-1.52(-1.75, -1.29)	-1.57(-1.78, -1.37)

Table 4.5 SEER data. Bayes factors for testing each covariate effect on the baseline survival.

Covariates	Intercept	Age	Race	Overall
BF_{10}	> 1000	24.5	4.7	61.0

likelihood (LPML) developed by Geisser and Eddy (1979). Let $f_{\mathbf{z}_{ij}}(\cdot)$ be the density function corresponding to $S_{\mathbf{z}_{ij}}(\cdot)$. In the context of the proposed model, the LPML is defined as LPML $= \sum_{i=1}^{n}\sum_{j=1}^{n_i}\log(\text{CPO}_{ij})$, where CPO_{ij}, referred to as conditional predictive ordinate, is given by $[f_{\mathbf{z}_{ij}}(t_{ij})^{\delta_{ij}}\left\{S_{\mathbf{z}_{ij}}(l_{ij}) - S_{\mathbf{z}_{ij}}(u_{ij})\right\}^{1-\delta_{ij}}|\mathcal{D}_{(ij)}]$ with $\mathcal{D}_{(ij)}$ denoting the remaining data after excluding the ijth data point $\mathcal{D}_{ij}$. A larger value of LPML indicates better predictive ability for the corresponding model. Furthermore, Geisser and Eddy (1979) discussed the exponentiated difference in LPML values from two models to obtain the so called pseudo Bayes factor (PBF). The PBF is a surrogate for the traditional Bayes factor, and can be interpreted similarly, but is more analytically tractable, less sensitive to prior specifications, and does not suffer from Lindley's paradox. The method suggested by Gelfand and Dey (1994) can be used to estimate the CPO statistics from MCMC output.

Table 4.4 summarizes the results. The proposed GAFT model with CAR frailties has the largest LPML compared to the semiparametric AFT frailty model and non-frailty GAFT model, indicating that both allowing baseline survival function varying

with covariates and taking into account spatial correlation improve model fit according to LPML. Note that the pseduo Bayes factor for the proposed model versus the AFT model is $e^{4115.6-4110.5} \approx 164$, indicating a decisive win for the GAFT/CAR model. Regarding the regression coefficient estimates, we see that all the three models give very similar covariate effects except for race, where GAFT/CAR model gives an estimate that is 26% higher than that from AFT model. This can be explained by the moderate significance of race on the baseline survival function based on the Bayes factors for per-value tests, as shown in Table 4.5. The age also highly affects the baseline survival according to the Bayes factor value.

For sake of visualization, Figure 4.2 presents the posterior mean curves under the proposed model. Panel (a) shows the baseline survival curves for white patients at three different ages. We see that the baseline curves coincide up to 18 years and then become very different, indicating that patients diagnosed at different age tend to have different baseline survival curves. Panel (b) shows that white and black patients have crossing baseline survival curves although they are not that significant. Panels (c) and (d) present the corresponding baseline densities, where we see that the baseline density changes from one mode to two as age increases. These results indicate that the traditional AFT assumption is violated for this data set, and the proposed model provides more valid inference. Finally, the final covariate-adjusted survival curves are presented for patients with three different ages in panel (e) and for white versus black patients in panel (f), where we see that race has a significant impact on survival up to 20 years after the diagnosis date.

The posterior means of spatial frailties for each county are mapped in Figure 4.3. The map shows that northern counties have relatively lower frailties and several counties in the southeast also exhibit lower spatial frailties. Since the frailties are additive to the logarithm of survival times in the proposed model, survival times are expected to be shorter in the regions with lower frailties. The mortality rates

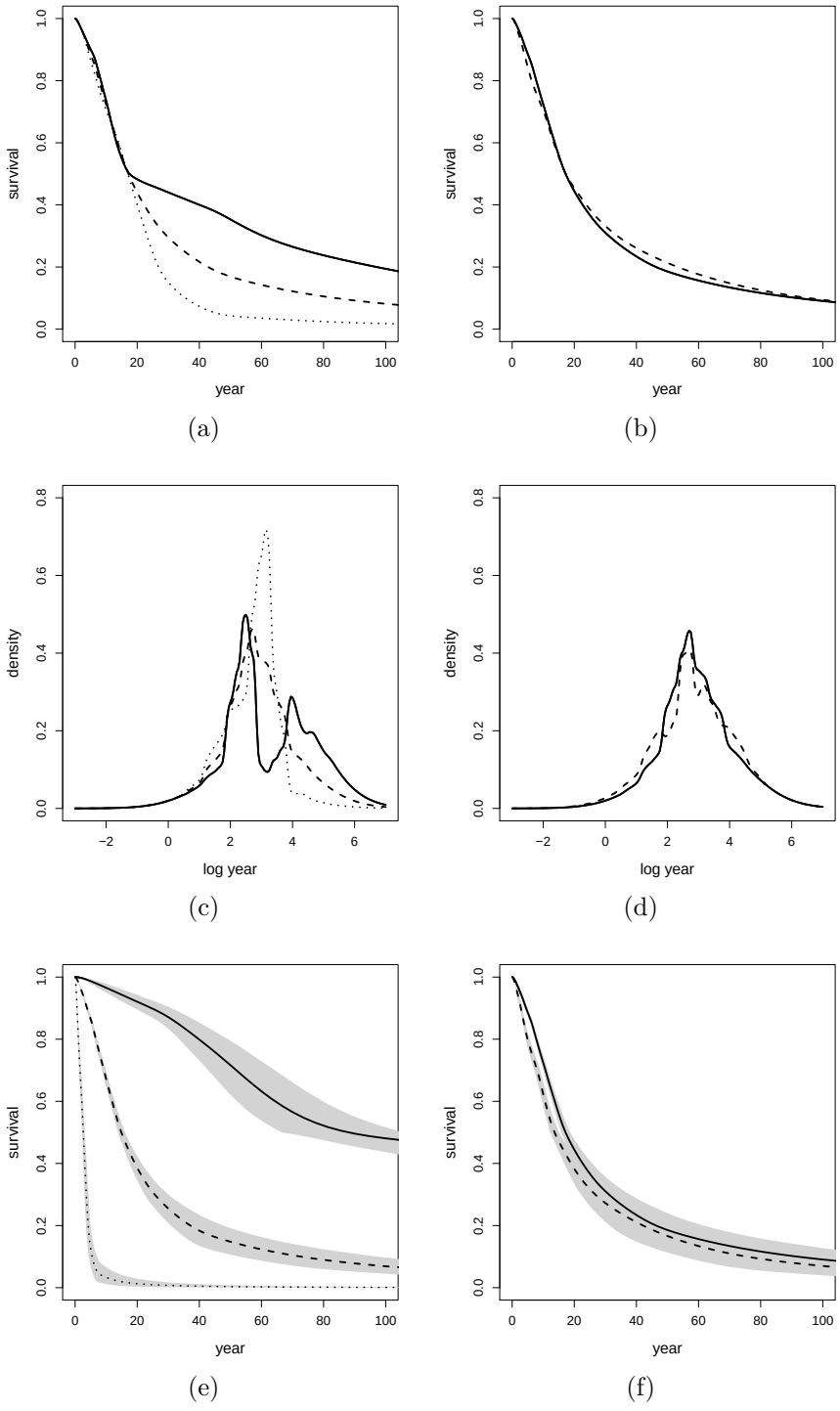
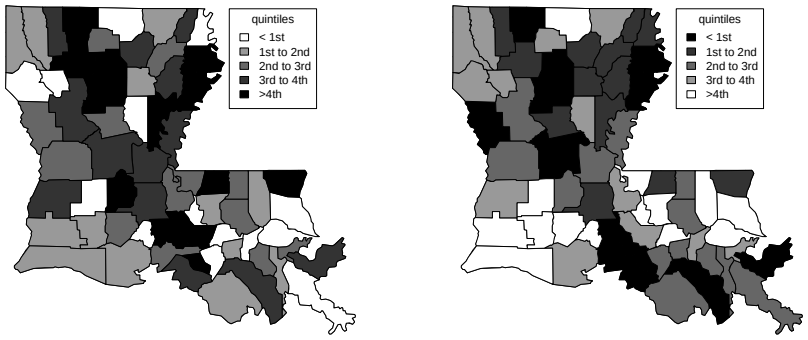


Figure 4.2 SEER data. Posterior mean curves under the proposed model. Panel (a) shows the baseline survival curves for white patients with age=40 (solid), age=70 (dashed) and age=100 (dotted). Panel (b) shows the baseline survival curves for white (solid) versus black (dashed) patients with age at the population mean (67.5). Panels (c) and (d) show their corresponding baseline densities. Panel (e) presents the final survival curves for patients with age=40 (solid), age=70 (dashed) and age=100 (dotted), holding other covariates at the reference levels. Panel (f) presents the final suvival curves for white (solid) versus black (dashed) patients, holding other categorical covariates at the reference levels and the age at the population mean (67.5). The 90% point-wise credible intervals are shown in gray areas.



(a) (b)

Figure 4.3 SEER data. Maps of the 2002 mortality rate (panel a) and CAR frailties (panel b) in the proposed model for Louisiana counties in 2002.

(percentages of death) based on our data set are also mapped for comparison, where we see that the counties showing lower frailties typically correspond higher mortality rates, providing support for the frailty modeling approach.

Based on the results from fitting the proposed model with CAR frailties as shown in Table 4.4, we see that all age, race, stage, marital status and grade are significant risk factors for survival of prostate cancer. Finally, note that $e^{-0.23} \approx 0.8$. The median lifetime for blacks is about 80% of that for whites, adjusting for other covariates and county of residence. The GAFT/CAR model retains the easy interpretability in terms of typical lifetime as traditional AFT models, yet allows for a smoothly changing, heteroscedastic baseline survival. In Section 4 of supplementary material, the covariate effects are also compared with those obtained under the censored quantile regression model (Portnoy, 2003), where we note that the standard deviation for race effect is five times greater than that under GAFT, and consequently race becomes insignificant.

4.5 Discussion

We have proposed a general median-regression survival model with CAR frailties that includes the traditional semiparametric and parametric AFT models as special cases. The overall strength of the proposed model is that robust, flexible modeling assumptions make it appropriate for use in most real data applications, leading to significant improvement of the prediction of cancer survival in epidemiological cancer studies. We have also developed an efficient `R` function `frailtyGAFT` in the package `spBayesSurv`, as well as plotting functions, LPML estimation, etc., so others can easily use the proposed model.

Along with a general, flexible spatial survival model that retains easy interpretability but allows for crossing survival curves, we also offer simple tests for adequacy of the traditional AFT model with static nonparametric S_0, as well as per-variable tests for whether covariates impact S_0, implemented in the `frailtyGAFT` function. Crossing or partially coinciding survival curves is fairly common (Bouliotis and Billingham, 2011; Logan et al., 2008) in clinical studies. Often survival curves coincide or are negligibly different during the initial period and differences start to occur after treatments take effect. Panel (a) in Figure 4.2 shows coinciding survivals in the baseline $S_{0,\mathbf{x}}(t)$ for about 18 years post-diagnosis in the Louisiana SEER data, then marked differences for white patients with different diagnosed ages, holding other covariates constant. Our analysis of the Louisiana SEER data shows a highly significant difference in survival between blacks and whites; adjusting for county and other covariates, whites have a median lifetime that is about 26% greater than blacks. However, the traditional censored quantile regression of Portnoy (2003) fails to detect such significant racial difference.

A referee has made the following observation. It is possible to consider a multivariate CAR on the $(p+1)$-dimensional $\mathbf{v}_i$ (Gelfand and Vounatsou, 2003), yielding the model $\log(t_{ij}) = \tilde{\mathbf{z}}'_{ij}(\tilde{\boldsymbol{\beta}} + \mathbf{v}_i) + \epsilon_{ij}$, thus allowing for county-level changes in how

predictors affect survival. The $\tilde{\boldsymbol{\beta}}$ are interpreted as overall regression effects and $\mathbf{v}_i$ are county-level deviations from $\tilde{\boldsymbol{\beta}}$.

Acknowledgements

This work is supported by NCI grant 5R03CA176739.

Supplementary material

Supplement to this chapter is available in Appendix C.

Bibliography

Aalen, O. O. (1980). A model for nonparametric regression analysis of counting processes. In *Mathematical Statistics and Probability Theory*, volume 2, pages 1–25. Springer-Verlag.

Aalen, O. O. (1989). A linear regression model for the analysis of life times. *Statistics in Medicine*, 8(8):907–925.

Alzola, C. and Harrell, F. (2006). *An Introduction to S and the Hmisc and Design Libraries*. Online manuscript available at http://biostat.mc.vanderbilt.edu/wiki/pub/Main/RS/sintro.pdf.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics*, 10(4):1100–1120.

Aslanidou, H., Dey, D., and Sinha, D. (1998). Bayesian analysis of multivariate survival data using Monte Carlo methods. *Canadian Journal of Statistics*, 26(1):33–48.

Banerjee, S. and Carlin, B. P. (2003). Semiparametric spatio-temporal frailty modeling. *Environmetrics*, 14(5):523–535.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data, Second Edition*. Chapman and Hall/CRC Press.

Banerjee, S. and Dey, D. K. (2005). Semiparametric proportional odds models for spatially correlated survival data. *Lifetime Data Analysis*, 11(2):175–191.

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(4):825–848.

Banerjee, S., Wall, M. M., and Carlin, B. P. (2003). Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota. *Biostatistics*, 4(1):123–142.

Bárdossy, A. (2006). Copula-based geostatistical models for groundwater quality parameters. *Water Resources Research*, 42(11):1–12.

Belitz, C., Brezger, A., Klein, N., Kneib, T., Lang, S., and Umlauf, N. (2015). BayesX - Software for Bayesian inference in structured additive regression models. Version 3.0. Available from http://www.bayesx.org.

Berger, J. O. and Guglielmi, A. (2001). Bayesian and conditional frequentist testing of a parametric model versus nonparametric alternatives. *Journal of the American Statistical Association*, 96(453):174–184.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B*, 36(2):192–236.

Bouliotis, G. and Billingham, L. (2011). Crossing survival curves: alternatives to the log-rank test. *Trials*, 12(Suppl 1):A137.

Boyle, D., Boyle, D., Olsen, V., Morgan, J., and Hyatt, A. (2004). Rapid quantitative detection of chytridiomycosis (Batrachochytrium dendrobatidis) in amphibian samples using real-time Taqman PCR assay. *Diseases of Aquatic Organisms*, 60:141–148.

Brezger, A., Kneib, T., and Lang, S. (2005). BayesX: Analyzing Bayesian structured additive regression models. *Journal of Statistical Software*, 14(11):1–22.

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika*, 66(3):429–436.

Burridge, J. (1981). Empirical Bayes analysis of survival time data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 43(1):65–75.

Cai, B., Lin, X., and Wang, L. (2011). Bayesian proportional hazards model for current status data with monotone splines. *Computational Statistics & Data Analysis*, 55(9):2644–2651.

Cai, B. and Meyer, R. (2011). Bayesian semiparametric modeling of survival data based on mixtures of B-spline distributions. *Computational Statistics & Data Analysis*, 55(3):1260–1272.

Carlin, B. P. and Hodges, J. S. (1999). Hierarchical proportional hazards regression models for highly stratified data. *Biometrics*, 55(4):1162–1170.

Chang, I.-S., Hsiung, C. A., Wu, Y.-J., and Yang, C.-C. (2005). Bayesian survival analysis using Bernstein polynomials. *Scandinavian Journal of Statistics*, 32(3):447–466.

Chen, Y., Hanson, T., and Zhang, J. (2014). Accelerated hazards model based on parametric families generalized with Bernstein polynomials. *Biometrics*, 70(1):192–201.

Chen, Y. Q. and Jewell, N. P. (2001). On a general class of semiparametric hazards regression models. *Biometrika*, 88(3):687–702.

Chen, Y. Q. and Wang, M.-C. (2000). Analysis of accelerated hazards models. *Journal of the American Statistical Association*, 95(450):608–618.

Cheng, S. C., Wei, L. J., and Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika*, 82(4):835–845.

Cheng, T. L., Rovito, S. M., Wake, D. B., and Vredenburg, V. T. (2011). Coincident mass extirpation of neotropical amphibians with the emergence of the infectious fungal pathogen Batrachochytrium dendrobatidis. *Proceedings of the National Academy of Sciences*, 108(23):9502–9507.

Chernoukhov, A. (2013). Bayesian Spatial Additive Hazard Model. *Electronic Theses and Dissertations*. Paper 4965.

Chiou, S. H., Kang, S., and Yan, J. (2015). Semiparametric accelerated failure time modeling for clustered failure times from stratified sampling. *Journal of the American Statistical Association*, 110(510):621–629.

Chivian, E. and Bernstein, A. (2008). *Sustaining life: how human health depends on biodiversity*. Oxford University Press.

Chlebowski, R. T., Anderson, G. L., Gass, M., Lane, D. S., Aragaki, A. K., Kuller, L. H., Manson, J. E., Stefanick, M. L., Ockene, J., Sarto, G. E., et al. (2010). Estrogen plus progestin and breast cancer incidence and mortality in postmenopausal women. *JAMA*, 304(15):1684–1692.

Christensen, R. (2011). *Plane answers to complex questions: the theory of linear models*. Springer Science+ Business Media.

Christensen, R. and Johnson, W. (1988). Modeling accelerated failure time with a Dirichlet process. *Biometrika*, 75(4):693–704.

Clayton, D. (1991). A monte carlo method for Bayesian inference in frailty models. *Biometrics*, 47(2):467–485.

Clayton, D. and Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model. *Journal of the Royal Statistical Society, Series A*, 148(2):82–117.

Cottone, F. (2008). *Covariate Dependent Random Effects in Survival Analysis*. PhD Dissertation, Universita degli Studi "Roma Tre".

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62(2):269–276.

Dabrowska, D. M. and Doksum, K. A. (1988). Estimation and testing in a two-sample generalized odds-rate model. *Journal of the American Statistical Association*, 83(403):744–749.

Darmofal, D. (2009). Bayesian spatial survival models for political event processes. *American Journal of Political Science*, 53(1):241–257.

Dasgupta, P., Cramb, S. M., Aitken, J. F., Turrell, G., and Baade, P. D. (2014). Comparing multilevel and Bayesian spatial random effects survival models to assess geographical inequalities in colorectal cancer survival: a case study. *International Journal of Health Geographics*, 13(1):36.

de Boor, C. (2001). *A Practical Guide to Splines*. Applied Mathematical Sciences, Vol. 27. Springer-Verlag: New York.

de Carvalho, V. I., Jara, A., Hanson, T., and de Carvalho, M. (2013). Bayesian nonparametric ROC regression modeling. *Bayesian Analysis*, 8(3):623–646.

De Iorio, M., Johnson, W. O., Müller, P., and Rosner, G. L. (2009). Bayesian nonparametric nonproportional hazards survival modeling. *Biometrics*, 65(3):762–771.

Devarajan, K. and Ebrahimi, N. (2011). A semi-parametric generalization of the Cox proportional hazards regression model: Inference and applications. *Computational Statistics & Data Analysis*, 55(1):667–676.

Dickey, J. M. (1971). The weighted likelihood ratio, linear hypotheses on normal location parameters. *The Annals of Mathematical Statistics*, 42(1):204–223.

Diva, U., Dey, D. K., and Banerjee, S. (2008). Parametric models for spatially correlated survival data for individuals with multiple cancers. *Statistics in Medicine*, 27(12):2127–2144.

Dukić, V. and Dignam, J. (2007). Bayesian hierarchical multiresolution hazard model for the study of time-dependent failure patterns in early stage breast cancer. *Bayesian Analysis*, 2:591–610.

Dunson, D. B. and Herring, A. H. (2005). Bayesian model selection and averaging in additive and proportional hazards. *Lifetime Data Analysis*, 11:213–232.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2):89–102.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90:577–588.

Fabius, J. (1964). Asymptotic behavior of Bayes' estimates. *The Annals of Mathematical Statistics*, 35(2):846–856.

Fahrmeir, L. and Kneib, T. (2011). *Bayesian Smoothing and Regression for Longitudinal, Spatial and Event History Data*. Oxford University Press.

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.

Finley, A. O., Sang, H., Banerjee, S., and Gelfand, A. E. (2009). Improving the performance of predictive process modeling for large datasets. *Computational statistics & data analysis*, 53(8):2873–2884.

Freedman, D. (1963). On the asymptotic behavior of Bayes' estimates in the discrete case. *The Annals of Mathematical Statistics*, 34(4):1386–1403.

Furrer, R., Genton, M. G., and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15(3):502–523.

Gamerman, D. (1997). Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing*, 7(1):57–68.

Geisser, S. and Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, 74(365):153–160.

Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society: Series B*, 56(3):501–514.

Gelfand, A. E. and Mallick, B. K. (1995). Bayesian analysis of proportional hazards models built from monotone functions. *Biometrics*, 51(3):843–852.

Gelfand, A. E. and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*, 4(1):11–15.

Goodman, L. A. and Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, 49(268):732–764.

Gray, R. J. (1992). Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association*, 87:942–951.

Griffin, J. (2010). Default priors for density estimation with mixture models. *Bayesian Analysis*, 5:45–64.

Gustafson, P. (1997). Large hierarchical Bayesian analysis of multivariate survival data. *Biometrics*, 53(1):230–242.

Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223–242.

Hanson, T., Johnson, W., and Laud, P. (2009). Semiparametric inference for survival models with step process covariates. *Canadian Journal of Statistics*, 37(1):60–79.

Hanson, T. and Johnson, W. O. (2002). Modeling regression error with a mixture of Polya trees. *Journal of the American Statistical Association*, 97(460):1020–1033.

Hanson, T. and Johnson, W. O. (2004). A Bayesian semiparametric AFT model for interval-censored data. *Journal of Computational and Graphical Statistics*, 13(2):341–361.

Hanson, T., Kottas, A., and Branscum, A. (2008). Modelling stochastic order in the analysis of receiver operating characteristic data: Bayesian nonparametric approaches. *Journal of the Royal Statistical Society: Series C*, 57(2):207–225.

Hanson, T. E. (2006a). Inference for mixtures of finite Polya tree models. *Journal of the American Statistical Association*, 101(476):1548–1565.

Hanson, T. E. (2006b). Modeling censored lifetime data using a mixture of gammas baseline. *Bayesian Analysis*, 1:575–594.

Hanson, T. E., Branscum, A., and Johnson, W. O. (2005). Bayesian nonparametric modeling and data analysis: An introduction. In Dey, D. and Rao, C., editors, *Bayesian Thinking: Modeling and Computation (Handbook of Statistics, volume 25)*, pages 245–278. Elsevier: Amsterdam.

Hanson, T. E., Branscum, A., and Johnson, W. O. (2011). Predictive comparison of joint longitudinal–survival modeling: a case study illustrating competing approaches. *Lifetime Data Analysis*, 17:3–28.

Hanson, T. E. and Jara, A. (2013). Surviving fully Bayesian nonparametric regression models. In *Bayesian Theory and Applications*, pages 592–615. Oxford University Press, Oxford.

Hanson, T. E., Jara, A., Zhao, L., et al. (2012). A Bayesian semiparametric temporally-stratified proportional hazards model with spatial frailties. *Bayesian Analysis*, 7(1):147–188.

Hanson, T. E. and Yang, M. (2007). Bayesian semiparametric proportional odds models. *Biometrics*, 63(1):88–95.

Hausauer, A. K., Keegan, T. H., Chang, E. T., Glaser, S. L., Howe, H., and Clarke, C. A. (2009). Recent trends in breast cancer incidence in us white women by county-level urban/rural and poverty status. *BMC Medicine*, 7(1):31.

Henderson, R., Shimakura, S., and Gorst, D. (2002). Modeling spatial variation in leukemia survival data. *Journal of the American Statistical Association*, 97(460):965–972.

Hennerfeind, A., Brezger, A., and Fahrmeir, L. (2006). Geoadditive survival models. *Journal of the American Statistical Association*, 101(475):1065–1075.

Hjort, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *The Annals of Statistics*, 18:1259–1294.

Hutton, J. L. and Monaghan, P. F. (2002). Choice of parametric accelerated life and proportional hazards models for survival data: Asymptotic results. *Lifetime Data Analysis*, 8:375–393.

Ibrahim, J. G., Chen, M. H., and Sinha, D. (2001). *Bayesian Survival Analysis*. Springer-Verlag.

Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453):161–173.

Ishwaran, H. and Zarepour, M. (2000). Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika*, 87(2):371–390.

Jara, A., Hanson, T., Quintana, F., Müller, P., and Rosner, G. (2011). DPpackage: Bayesian semi- and nonparametric modeling in R. *Journal of Statistical Software*, 40(5):1.

Jara, A. and Hanson, T. E. (2011). A class of mixtures of dependent tailfree processes. *Biometrika*, 98(3):553–566.

Jara, A., Hanson, T. E., and Lesaffre, E. (2009). Robustifying generalized linear mixed models using a new class of mixtures of multivariate Polya trees. *Journal of Computational and Graphical Statistics*, 18(4):838–860.

Jara, A., Lesaffre, E., De Iorio, M., and Quitana, F. (2010). Bayesian semiparametric inference for multivariate doubly-interval-censored data. *The Annals of Applied Statistics*, 4(4):2126–2149.

Johnson, W. O. and Christensen, R. (1989). Nonparametric bayesian analysis of the accelerated failure time model. *Statistics and Probability Letters*, 8:179–184.

Kalbfleisch, J. (1978). Non-parametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society, Series B*, 40(2):214–221.

Kaufman, C. G., Schervish, M. J., and Nychka, D. W. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association*, 103(484):1545–1555.

Kay, R. and Kinnersley, N. (2002). On the use of the accelerated failure time model as an alternative to the proportional hazards model in the treatment of time to event data: A case study in influenza. *Drug Information Journal*, 36:571–579.

Knapp, R. A., Matthews, K. R., Preisler, H. K., and Jellison, R. (2003). Developing probabilistic models to predict amphibian site occupancy in a patchy landscape. *Ecological Applications*, 13(4):1069–1082.

Knapp, R. A. and Morgan, J. A. (2006). Tadpole mouthpart depigmentation as an accurate indicator of chytridiomycosis, an emerging disease of amphibians. *Copeia*, 2006(2):188–197.

Kneib, T. (2006). *Mixed model based inference in structured additive regression*. Ludwig-Maximilians-Universität München.

Kneib, T. and Fahrmeir, L. (2007). A mixed model approach for geoadditive hazard regression. *Scandinavian Journal of Statistics*, 34(1):207–228.

Koenker, R. (2008). Censored quantile regression redux. *Journal of Statistical Software*, 27(6):1–25.

Koenker, R. and Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, 15:143–156.

Komárek, A. and Lesaffre, E. (2007). Bayesian accelerated failure time model for correlated censored data with a normal mixture as an error distribution. *Statistica Sinica*, 17:549–569.

Komárek, A. and Lesaffre, E. (2008). Bayesian accelerated failure time model with multivariate doubly-interval-censored data and flexible distributional assumptions. *Journal of the American Statistical Association*, 103:523–533.

Konomi, B. A., Sang, H., and Mallick, B. K. (2014). Adaptive Bayesian nonstationary modeling for large spatial datasets using covariance approximations. *Journal of Computational and Graphical Statistics*, 23:802–929.

Kottas, A. and Gelfand, A. E. (2001). Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association*, 96(456):1458–1468.

Krieger, N., Chen, J. T., and Waterman, P. D. (2010). Decline in us breast cancer rates after the women's health initiative: Socioeconomic and racial/ethnic differentials. *American Journal of Public Healthn*, 100(6):132–139.

Kuo, L. and Mallick, B. (1997). Bayesian semiparametric inference for the accelerated failure-time model. *Canadian Journal of Statistics*, 25(4):457–472.

Laird, N. and Olivier, D. (1981). Covariance analysis of censored survival data using log-linear analysis techniques. *Journal of the American Statistical Association*, 76(374):231–240.

Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics*, 13:183–212.

Lavine, M. (1992). Some aspects of Polya tree distributions for statistical modelling. *The Annals of Statistics*, 20:1222–1235.

Lavine, M. (1994). More aspects of Polya tree distributions for statistical modelling. *The Annals of Statistics*, 22:1161–1176.

Li, J. (2010). Application of copulas as a new geostatistical tool. *Dissertation*.

Li, J., Hong, Y., Thapa, R., and Burkhart, H. E. (2015a). Survival analysis of loblolly pine trees with spatially correlated random effects. *Journal of the American Statistical Association*, in press.

Li, L., Hanson, T., and Zhang, J. (2015b). Spatial extended hazard model with application to prostate cancer survival. *Biometrics*, in press.

Li, Y. and Lin, X. (2006). Semiparametric normal transformation models for spatially correlated survival data. *Journal of the American Statistical Association*, 101(474):591–603.

Li, Y. and Ryan, L. (2002). Modeling spatial survival data using semiparametric frailty models. *Biometrics*, 58(2):287–297.

Lin, X., Cai, B., Wang, L., and Zhang, Z. (2015). A Bayesian proportional hazards model for general interval-censored data. *Lifetime Data Analysis*, in press.

Lin, X. and Wang, L. (2011). Bayesian proportional odds models for analyzing current status data: univariate, clustered, and multivariate. *Communications in Statistics-Simulation and Computation*, 40(8):1171–1181.

Lips, K. R., Diffendorfer, J., Mendelson, J. R., and Sears, M. W. (2008). Riding the wave: reconciling the roles of disease and climate change in amphibian declines. *PLoS Biology*, 6(3):e72.

Liu, D., Kalbfleisch, J., and Schaubel, D. (2011). A positive stable frailty model for clustered failure time data with covariate-dependent frailty. *Biometrics*, 67(1):8–17.

Liu, Y. (2012). *Bayesian analysis of spatial and survival models with applications of computation techniques*. PhD thesis, University of Missouri–Columbia.

Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *Annals of Statistics*, 12:351–357.

Logan, B. R., Klein, J. P., and Zhang, M.-J. (2008). Comparing treatments in the presence of crossing survival curves: an application to bone marrow transplantation. *Biometrics*, 64(3):733–740.

Martinussen, T. and Scheike, T. H. (2006). *Dynamic Regression Models for Survival Data*. Springer-Verlag.

McCulloch, C. E. and Neuhaus, J. M. (2011). Misspecifying the shape of a random effects distribution: why getting it wrong may not matter. *Statistical Science*, 26(3):388–402.

McKinley, T. J. (2007). *Spatial survival analysis of infectious animal diseases*. PhD thesis, University of Exeter.

Müller, P., Quintana, F., Jara, A., and Hanson, T. (2015). *Bayesian Nonparametric Data Analysis*. Springer-Verlag: New York.

Müller, P. and Quintana, F. A. (2004). Nonparametric Bayesian data analysis. *Statistical Science*, 19(1):95–110.

Murphy, S. A., Rossini, A. J., and van der Vaart, A. W. (1997). Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association*, 92:968–976.

Neal, R. M. (2000). Markov chain sampling methods for dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9:249–265.

Neal, R. M. (2003). Slice sampling. *Annals of Statistics*, 31(3):705–767.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer, 2nd edition.

Nieto-Barajas, L. E. (2013). Lévy-driven processes in Bayesian nonparametric inference. *Boletín de la Sociedad Matemática Mexicana*, 19:267–279.

Noh, M., Ha, I. D., and Lee, Y. (2006). Dispersion frailty models and hglms. *Statistics in Medicine*, 25(8):1341–1354.

Ojiambo, P. and Kang, E. (2013). Modeling spatial frailties in survival analysis of cucurbit downy mildew epidemics. *Phytopathology*, 103(3):216–227.

Orbe, J., Ferreira, E., and Núñez Antón, V. (2002). Comparing proportional hazards and accelerated failure time models for survival analysis. *Statistics in Medicine*, 21(22):3493–3510.

Pan, C., Cai, B., Wang, L., and Lin, X. (2014). Bayesian semi-parametric model for spatial interval-censored survival data. *Computational Statistics & Data Analysis*, 74:198–209.

Pang, L., Lu, W., and Wang, H. J. (2015). Local Buckley-James estimation for heteroscedastic accelerated failure time model. *Statistica Sinica*, 25(3):863–877.

Papaspiliopoulos, O. and Roberts, G. O. (2008). Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. *Biometrika*, 95(1):169–186.

Pati, D., Dunson, D. B., and Tokdar, S. T. (2013). Posterior consistency in conditional distribution estimation. *Journal of Multivariate Analysis*, 116:456–472.

Petrone, S. (1999a). Bayesian density estimation using Bernstein polynomials. *The Canadian Journal of Statistics*, 27:105–126.

Petrone, S. (1999b). Random Bernstein polynomials. *Scandinavian Journal of Statistics*, 26:373–393.

Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association*, 98(464):1001–1012.

Qiou, Z., Ravishanker, N., and Dey, D. (1999). Multivariate survival analysis with positive stable frailties. *Biometrics*, 55(2):637–644.

Rachowicz, L. J., Knapp, R. A., Morgan, J. A., Stice, M. J., Vredenburg, V. T., Parker, J. M., and Briggs, C. J. (2006). Emerging infectious disease as a proximate cause of amphibian mass mortality. *Ecology*, 87(7):1671–1683.

Raftery, A. E. (1996). Hypothesis testing and model selection via posterior simulation. In *Markov Chain Monte Carlo in Practice*, pages 163–187. Springer.

Reich, B. J., Bondell, H. D., and Wang, H. J. (2010). Flexible Bayesian quantile regression for independent and clustered data. *Biostatistics*, 11(2):337–352.

Reid, N. (1994). A conversation with Sir David Cox. *Statistical Science*, 9:439–455.

Robert, C. and Casella, G. (2005). *Monte Carlo Statistical Methods*. Springer.

Roberts, G. O. and Smith, A. F. (1994). Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. *Stochastic Processes and Their Applications*, 49(2):207–216.

Rodríguez, A., Dunson, D. B., and Gelfand, A. E. (2010). Latent stick-breaking processes. *Journal of the American Statistical Association*, 105(490):647–659.

Rossouw, J., Anderson, G., Prentice, R., LaCroix, A., Kooperberg, C., Stefanick, M., Jackson, R., Beresford, S., Howard, B., Johnson, K., Kotchen, J., and Ockene, J. (2002). Risks and benefits of estrogen plus progestin in healthy postmenopausal women. *Journal of the American Medical Association*, 288:321–333.

Ryan, T. and Woodall, W. (2005). The most-cited statistical papers. *Journal of Applied Statistics*, 32(5):461–474.

Sahu, S. K. and Dey, D. K. (2004). On multivariate survival models with a skewed frailty and a correlated baseline hazard process. *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, pages 321–338.

Sang, H. and Huang, J. Z. (2012). A full scale approximation of covariance functions for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(1):111–132.

Scharfstein, D. O., Tsiatis, A. A., and Gilbert, P. B. (1998). Semiparametric efficient estimation in the generalized odds-rate class of regression models for right-censored time-to-event data. *Lifetime data analysis*, 4(4):355–391.

Schörgendorfer, A., Branscum, A. J., and Hanson, T. E. (2013). A Bayesian goodness of fit test and semiparametric generalization of logistic regression with measurement data. *Biometrics*, 69(2):508–519.

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

Sharef, E., Strawderman, R. L., Ruppert, D., Cowen, M., and Halasyamani, L. (2010). Bayesian adaptive B-spline estimation in proportional hazards frailty models. *Electronic Journal of Statistics*, 4:606–642.

Sinha, D. and Dey, D. K. (1997). Semiparametric Bayesian analysis of survival data. *Journal of the American Statistical Association*, 92:1195–1212.

Sinha, D., McHenry, M. B., Lipsitz, S. R., and Ghosh, M. (2009). Empirical bayes estimation for additive hazards regression models. *Biometrika*, 96(3):545–558.

Skerratt, L. F., Berger, L., Speare, R., Cashins, S., McDonald, K. R., Phillott, A. D., Hines, H. B., and Kenyon, N. (2007). Spread of chytridiomycosis has caused the rapid global decline and extinction of frogs. *EcoHealth*, 4(2):125–134.

Smith, M. S. (2013). Bayesian approaches to copula modelling. *Bayesian Theory and Applications*, pages 336–358.

Song, P. (2000). Multivariate dispersion models generated from Gaussian copula. *Scandinavian Journal of Statistics*, 27(2):305–320.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, 64(4):583–639.

Sprague, B. L., Trentham-Dietz, A., Gangnon, R. E., Ramchandani, R., Hampton, J. M., Robert, S. A., Remington, P. L., and Newcomb, P. A. (2011). Socioeconomic status and survival after an invasive breast cancer diagnosis. *Cancer*, 117(7):1542–1551.

Susarla, V. and Van Ryzin, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete observations. *Journal of the American Statistical Association*, 71:897–902.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag Inc.

Therneau, T. M., Grambsch, P. M., and Pankratz, V. S. (2003). Penalized survival models and frailty. *Journal of Computational and Graphical Statistics*, 12(1):156–175.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22(4):1701–1728.

Tierney, L. and Mira, A. (1999). Some adaptive Monte Carlo methods for Bayesian inference. *Statistics in Medicine*, 18(1718):2507–2515.

Trippa, L., Müller, P., and Johnson, W. (2011). The multivariate beta process and an extension of the polya tree model. *Biometrika*, 98(1):17–34.

Umlauf, N., Adler, D., Kneib, T., Lang, S., and Zeileis, A. (2015). Structured additive regression models: An R interface to BayesX. *Journal of Statistical Software*, 63(21):1–46.

Vaupel, J. W., Manton, K. G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3):439–454.

Verdinelli, I. and Wasserman, L. (1995). Computing Bayes factors using a generalization of the Savage-Dickey density ratio. *Journal of the American Statistical Association*, 90(430):614–618.

Vredenburg, V., Bingham, R., Knapp, R., Morgan, J., Moritz, C., and Wake, D. (2007). Concordant molecular and phenotypic data delineate new taxonomy and conservation priorities for the endangered mountain yellow-legged frog. *Journal of Zoology*, 271(4):361–374.

Vredenburg, V. T., Knapp, R. A., Tunstall, T. S., and Briggs, C. J. (2010). Dynamics of an emerging disease drive large-scale amphibian population extinctions. *Proceedings of the National Academy of Sciences*, 107(21):9689–9694.

Wake, D. B. and Vredenburg, V. T. (2008). Are we in the midst of the sixth mass extinction? A view from the world of amphibians. *Proceedings of the National Academy of Sciences*, 105(Supplement 1):11466–11473.

Walker, S. G. and Mallick, B. K. (1997). Hierarchical generalized linear models and frailty models with Bayesian nonparametric mixing. *Journal of the Royal Statistical Society, Series B: Methodological*, 59:845–860.

Walker, S. G. and Mallick, B. K. (1999). A Bayesian semiparametric accelerated failure time model. *Biometrics*, 55(2):477–483.

Waller, L. A. and Gotway, C. A. (2004). *Applied spatial statistics for public health data*. John Wiley & Sons.

Wang, L. and Dunson, D. B. (2011). Semiparametric Bayes' proportional odds models for current status data with underreporting. *Biometrics*, 67(3):1111–1118.

Wang, S., Zhang, J., and Lawson, A. B. (2012). A Bayesian normal mixture accelerated failure time spatial model and its application to prostate cancer. *Statistical Methods in Medical Research*, http://dx.doi.org/10.1177/0962280212466189.

Wang, Z. and Louis, T. A. (2004). Marginalized binary mixed-effects models with covariate-dependent random effects and likelihood inference. *Biometrics*, 60(4):884–891.

Wassell, J. T. and Moeschberger, M. (1993). A bivariate survival model with modified gamma frailty for assessing the impact of interventions. *Statistics in Medicine*, 12(3-4):241–248.

Wei, L.-J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association*, 84(408):1065–1073.

West, M. (1984). Outlier models and prior distributions in Bayesian linear regression. *Journal of the Royal Statistical Society, Series B*, 46(3):431–439.

Wysowski, D. K. and Governale, L. A. (2005). Use of menopausal hormones in the united states, 1992 through june, 2003. *Pharmacoepidemiology and Drug Safety*, 14(3):171–176.

Yang, S. (1999). Censored median regression using weighted empirical survival and hazard functions. *Journal of the American Statistical Association*, 94:137–145.

Yang, S. and Prentice, R. L. (1999). Semiparametric inference in the proportional odds regression model. *Journal of the American Statistical Association*, 94:125–136.

Yashin, A. I. and Iachine, I. A. (1999). What difference does the dependence between durations make? insights for population studies of aging. *Lifetime Data Analysis*, 5(1):5–22.

Yin, G. and Ibrahim, J. G. (2005). A class of Bayesian shared gamma frailty models with multivariate failure time data. *Biometrics*, 61:208–216.

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.

Zellner, A. (1983). Applications of Bayesian analysis in econometrics. *The Statistician*, 32(1/2):23–34.

Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99(465):250–261.

Zhang, J. and Lawson, A. B. (2011). Bayesian parametric accelerated failure time spatial model and its application to prostate cancer. *Journal of Applied Statistics*, 38(3):591–603.

Zhang, J., Peng, Y., and Zhao, O. (2011). A new semiparametric estimation method for accelerated hazard model. *Biometrics*, 67:1352–1360.

Zhang, M. and Davidian, M. (2008). "Smooth" semiparametric regression analysis for arbitrarily censored time-to-event data. *Biometrics*, 64(2):567–576.

Zhao, L. and Hanson, T. E. (2011). Spatially dependent Polya tree modeling for survival data. *Biometrics*, 67(2):391–403.

Zhao, L., Hanson, T. E., and Carlin, B. P. (2009). Mixtures of Polya trees for flexible spatial frailty survival modelling. *Biometrika*, 96(2):263–276.

Zhou, H., Hanson, T., Jara, A., and Zhang, J. (2015a). Modeling county level breast cancer survival data using a covariate-adjusted frailty proportional hazards model. *The Annals of Applied Statistics*, 9(1):43–68.

Zhou, H., Hanson, T., and Knapp, R. (2015b). Marginal Bayesian nonparametric model for time to disease arrival of threatened amphibian populations. *Biometrics*, in press.

Zhou, H., Hanson, T., and Zhang, J. (2015c). Generalized accelerated failure time spatial frailty model for arbitrarily censored data. *Lifetime Data Analysis*, in revision.

Appendix A

Supplement to Chapter 2

A.1 Mixtures of linear dependent Tailfree processes

A tailfree process is a stochastic process with trajectories on a given space of probability distributions. Tailfree processes generalize Polya tree processes, which further generalize Dirichlet processes. Freedman (1963) and Fabius (1964) introduced and developed tailfree processes. Ferguson (1974) summarized, extended, and made connections among these developments.

Let G_θ be a family of distribution functions on $\mathbb{R}$, indexed by $\theta \in \Theta$, serving to center the random probability measure G. A tailfree random distribution G with support on the real line is defined by allocations of conditional probabilities to increasingly refined partitions of $\mathbb{R}$. Let $E^m = \{\epsilon_1 \cdots \epsilon_m : \epsilon_i = 0, 1; i = 1 \ldots, m\}$ be the m-fold Cartesian product for all $m \in \mathbb{N}$ except $E^0 = \{\emptyset\}$, and set $E_j = \bigcup_{m=0}^{j} E^m$ for every $j = 1, 2, \ldots, J$, where J is a positive integer. For convenience, we use the convention that $\boldsymbol{\epsilon}0 = 0$ and $\boldsymbol{\epsilon}1 = 1$ for $\boldsymbol{\epsilon} = \emptyset$. At each level $j \in \{1, 2, \ldots, J\}$, we partition $\mathbb{R}$ into 2^j sets $\pi_\theta^j = \{B_\theta(\boldsymbol{\epsilon}) : \boldsymbol{\epsilon} \in E^j\}$, where $B_\theta(\boldsymbol{\epsilon}) = (G_\theta^{-1}(k/2^j), G_\theta^{-1}([k+1]/2^j)]$, with G_θ^{-1} being the quantile function of G_θ, and k being the decimal representation of $\boldsymbol{\epsilon} = \epsilon_1 \cdots \epsilon_j \in E^j$. For $\boldsymbol{\epsilon} = \emptyset$, define $B_\theta(\boldsymbol{\epsilon}) = \mathbb{R}$. These partition sets produce dyadic splits: $B_\theta(\boldsymbol{\epsilon}) = B_\theta(\boldsymbol{\epsilon}0) \cup B_\theta(\boldsymbol{\epsilon}1)$, for $\boldsymbol{\epsilon} \in E_{J-1}$. For each partition set at level j, indexed by the binary number $\boldsymbol{\epsilon} \in E^j$, a tailfree construction assigns two random conditional probabilities, $Y_{\boldsymbol{\epsilon}0} = G\{B_\theta(\boldsymbol{\epsilon}0)|B_\theta(\boldsymbol{\epsilon})\}$ and $Y_{\boldsymbol{\epsilon}1} = 1 - Y_{\boldsymbol{\epsilon}0} = G\{B_\theta(\boldsymbol{\epsilon}1)|B_\theta(\boldsymbol{\epsilon})\}$, to the offspring sets of $B_\theta(\boldsymbol{\epsilon})$. The collections $\{Y_0\}$, $\{Y_{00}, Y_{10}\}$,$\{Y_{000}, Y_{010}, Y_{100}, Y_{110}\}, \ldots,$

are mutually independent. Assume that G follows G_θ on sets in the finest partition π_θ^J and note that $G(B_\theta(\boldsymbol{\epsilon})) = \prod_{i=1}^{j} Y_{\epsilon_1 \cdots \epsilon_i}$ for every $\boldsymbol{\epsilon} \in E_J$. Given the J partitions $\Pi^{J,\theta} = \bigcup_{j=1}^{J} \pi_\theta^j$, a particular finite tailfree process is defined in Table A.1 for $j = 1, 2, 3$.

Table A.1 Principle scheme to define a tailfree process centered around G_θ

$\mathbb{R}$							
$B_\theta(0)$ $Y_0 = G\{B_\theta(0)\|\mathbb{R}\}$				$B_\theta(1)$ $Y_1 = G\{B_\theta(1)\|\mathbb{R}\}$			
$B_\theta(00)$ $Y_{00} = G\{B_\theta(0)\|B_\theta(0)\}$		$B_\theta(01)$ $Y_{01} = G\{B_\theta(01)\|B_\theta(0)\}$		$B_\theta(10)$ $Y_{10} = G\{B_\theta(10)\|B_\theta(1)\}$		$B_\theta(11)$ $Y_{11} = G\{B_\theta(11)\|B_\theta(1)\}$	
$B_\theta(000)$ Y_{000}	$B_\theta(001)$ Y_{001}	$B_\theta(010)$ Y_{010}	$B_\theta(011)$ Y_{011}	$B_\theta(100)$ Y_{100}	$B_\theta(101)$ Y_{101}	$B_\theta(110)$ Y_{110}	$B_\theta(111)$ Y_{111}

A Polya tree prior is defined by assigning to the conditional probabilities $\{Y_{\boldsymbol{\epsilon}0}\}_{\boldsymbol{\epsilon} \in E_{J-1}}$ independent beta distributions, that is, $Y_{\boldsymbol{\epsilon}0} \overset{ind.}{\sim} \text{Beta}(cd_{\boldsymbol{\epsilon}0}^2, cd_{\boldsymbol{\epsilon}0}^2)$, where $c > 0$ is the precision parameter and $d_{\boldsymbol{\epsilon}0}$ is the length of $\boldsymbol{\epsilon}0$. To allow the entire shape of the frailty distribution to vary smoothly with covariates, we instead consider logistic-normal conditional probabilities that closely follow beta distributions yielding a particular class of dependent tailfree processes (Jara and Hanson, 2011). Let $\mathbf{X} = [\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_n]'$ be the $n \times (q+1)$ design matrix, where each $\tilde{\mathbf{x}}_i = (1, \mathbf{x}_i')'$. Given the cluster-level covariate values $\mathbf{x}$ in cluster, a tailfree random distribution $G_{\mathbf{x}}$ is defined by replacing the above independent beta random probabilities $\{Y_{\boldsymbol{\epsilon}0}\}_{\boldsymbol{\epsilon} \in E_{J-1}}$ with the following logistic-normal random variables

$$Y_{\boldsymbol{\epsilon}0}(\mathbf{x}, \boldsymbol{\beta}_{\boldsymbol{\epsilon}0}) = h\left(\tilde{\mathbf{x}}'\boldsymbol{\beta}_{\boldsymbol{\epsilon}0}\right) = \frac{\exp(\tilde{\mathbf{x}}'\boldsymbol{\beta}_{\boldsymbol{\epsilon}0})}{1 + \exp(\tilde{\mathbf{x}}'\boldsymbol{\beta}_{\boldsymbol{\epsilon}0})}, \quad Y_{\boldsymbol{\epsilon}1}(\mathbf{x}, \boldsymbol{\beta}_{\boldsymbol{\epsilon}0}) = 1 - h\left(\tilde{\mathbf{x}}'\boldsymbol{\beta}_{\boldsymbol{\epsilon}0}\right) = \frac{1}{1 + \exp(\tilde{\mathbf{x}}'\boldsymbol{\beta}_{\boldsymbol{\epsilon}0})},$$

where $\boldsymbol{\beta}_{\boldsymbol{\epsilon}0} \overset{ind}{\sim} N_{q+1}\left(\mathbf{0}, 2n(\mathbf{X}'\mathbf{X})^{-1}/(c\rho(d_{\boldsymbol{\epsilon}0}))\right)$, with $\rho(d_{\boldsymbol{\epsilon}0}) = d_{\boldsymbol{\epsilon}0}^2$, following so called g-priors (Zellner, 1983). The resulting process $\{G_{\mathbf{x}} : \mathbf{x} \in \mathcal{X}\}$ is referred to as a *p*artially specified linear dependent tailfree process with parameters $(h, \Pi^{J,\theta}, \mathcal{A}^{J,c,\rho})$. It is easy to see that $E\{Y_{\boldsymbol{\epsilon}0}(\mathbf{x}, \boldsymbol{\beta}_{\boldsymbol{\epsilon}0})|\mathbf{x}\} = 0.5$, so $E(G_{\mathbf{x}}) = G_\theta$, namely the random distribution $G_{\mathbf{x}}$ is centered around G_θ for all $\mathbf{x} \in \mathcal{X}$.

For notational simplicity, set $Y_{\boldsymbol{\epsilon}0}(i) = Y_{\boldsymbol{\epsilon}0}(\mathbf{x}_i, \boldsymbol{\beta}_{\boldsymbol{\epsilon}0})$ and $\eta_{\boldsymbol{\epsilon}0}(i) = \tilde{\mathbf{x}}_i'\boldsymbol{\beta}_{\boldsymbol{\epsilon}0}$. Define $\boldsymbol{\eta}_{\boldsymbol{\epsilon}0} = (\eta_{\boldsymbol{\epsilon}0}(1), \ldots, \eta_{\boldsymbol{\epsilon}0}(n))'$. Based on the g-priors of $\boldsymbol{\beta}_{\boldsymbol{\epsilon}0}$, we have $\boldsymbol{\eta}_{\boldsymbol{\epsilon}0}|c \sim N_n\left(\mathbf{0}, g\mathbf{M}\right)$

where $g = 2/(c\rho(d_{\epsilon 0}))$, with $\rho(d_{\epsilon 0}) = d_{\epsilon 0}^2$, and $\mathbf{M} = n\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. This specification implies the logit conditional probabilities are positively correlated for the clusters with covariates that are close to each other. Let $\bar{\mathbf{x}} = \sum_{i=1}^n \mathbf{x}_i$ and $(n-1)\mathbf{S} = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$. Christensen (2011) notes that the ith diagonal element of $\mathbf{M}$ is given by $m_{ii} = 1 + nd_{ii}/(n-1)$, where $d_{ii} = (\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})$ is the sample squared Mahalanobis distance between $\mathbf{x}_i$ and $\bar{\mathbf{x}}$. For cluster-level covariate values $\mathbf{x}_{i_1}$, $\mathbf{x}_{i_2}$ reasonably close to the sample mean $\bar{\mathbf{x}}$ (in terms of the Mahalanobis distance), a first order approximation gives that $(Y_{\epsilon 0}(i_1), Y_{\epsilon 0}(i_2))'$ approximately follows $N_2(\mathbf{0}, g\mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0)$, where $\mathbf{X}_0 = (\mathbf{x}_{i_1}, \mathbf{x}_{i_2})'$. The larger c or $d_{\epsilon 0}$ are, the better this approximation becomes. This leads to

$$\text{corr}\{Y_{\epsilon 0}(i_1), Y_{\epsilon 0}(i_2)\} \approx \frac{1 + d_{i_1,i_2}}{\sqrt{1 + d_{i_1,i_1}}\sqrt{1 + d_{i_2,i_2}}},$$

where $d_{i,j} = (\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})$. Note that the locations $\mathbf{x}_{i_1}$ and $\mathbf{x}_{i_2}$ relative to the mean covariate vector $\bar{\mathbf{x}}$ play a role in the correlation, and not just the Mahalanobis distance between them. When the covariate vector for a cluster i is fixed at $\mathbf{x}_i$, Jara and Hanson (2011) show that assuming $\eta_{\epsilon 0}(i) \sim N(0, 2m_{ii}/cd_{\epsilon 0}^2)$ approximates $Y_{\epsilon 0}(i) \sim \text{Beta}(c_i d_{\epsilon 0}^2, c_i d_{\epsilon 0}^2)$, where $c_i = cm_{ii}^{-1}$. This approximation explains the similarity between a marginal realization $G_{\mathbf{x}_i}$ of the linear-dependent tailfree process and a "traditional" Polya tree prior.

The precision parameter $c \in \mathbb{R}^+$ controls how closely the random distribution $G_{\mathbf{x}}$ follows G_θ in terms of L_1 distance (Hanson et al., 2008). Large values of c indicate a strong belief that the frailties are closely *iid* from G_θ, since as c tends to ∞, the random distribution $G_{\mathbf{x}}$ is G_θ with probability 1. Smaller values of c, on the other hand, allow more pronounced deviations of G from G_θ. The choice $c = 1$ has been advocated by many authors, e.g. recently Schörgendorfer et al. (2013). Alternatively, we can also put a gamma prior on c, say, $c \sim \Gamma(a_c, b_c)$, as suggested in Jara and Hanson (2011); if that is the case, the full conditional distribution for c is also a gamma distribution. As for the choice of J, we typically consider $J \approx \log_2(n/n_0)$,

where n is the sample size and n_0 (usually from 5 to 10) is a "typical" number of observations falling into each set at level J (Hanson, 2006a). In addition, the linear dependent tailfree process depends on the partition $\Pi^{J,\theta}$ which is further determined by the centering distribution G_θ. If one simply fixes θ at a specific value, the posterior inferences may be affected unduly due to the partition dependence. In practice, one common strategy to mitigate this problem is to specify a mixture of LDTFP by allowing parameter θ of the centering distribution to be random, that is,

$$\{G_{\mathbf{x}} : \mathbf{x} \in \mathcal{X}\} \sim \text{LDTFP}(h, \Pi^{J,\theta}, \mathcal{A}^{J,c,\rho}), \quad \theta \sim P(d\theta),$$

where $P(d\theta)$ represents a prior for θ; in this article, we consider $\theta^{-2} \sim \Gamma(\tau_1, \tau_2)$, where Γ refers to the gamma distribution with shape τ_1 and rate τ_2.

Given θ and $\boldsymbol{\beta} = \{\boldsymbol{\beta}_{\boldsymbol{\epsilon}0} : \boldsymbol{\epsilon} \in E_{J-1}\}$, the conditional density of e_i, given cluster-level predictors $\mathbf{x}_i$, is given by

$$g(e_i|\theta, \boldsymbol{\beta}) = 2^J \phi_\theta(e_i) \prod_{j=1}^{J} Y_{\boldsymbol{\epsilon}_\theta(e_i,j)}(\mathbf{x}_i, \boldsymbol{\beta}_{\boldsymbol{\epsilon}_\theta(e_i,j-1)0}), \tag{A.1}$$

where $\phi_\theta(\cdot)$ is the density of a $N(0, \theta^2)$ variate and $\boldsymbol{\epsilon}_\theta(e_i, j) = \epsilon_1\epsilon_2 \cdots \epsilon_j$ is the set in π_θ^j that e_i is in. Therefore, the joint density of the frailty terms is given by

$$p(e_1, \ldots, e_n|\theta, \boldsymbol{\beta}) = \left[\prod_{i=1}^{n} 2^J \phi_\theta(e_i)\right] \left[\prod_{\boldsymbol{\epsilon} \in E_{J-1}} \prod_{i=1}^{n} \frac{\exp(\tilde{\mathbf{x}}_i' \boldsymbol{\beta}_{\boldsymbol{\epsilon}0})^{I\{e_i \in B_\theta(\boldsymbol{\epsilon}0)\}}}{[1 + \exp(\tilde{\mathbf{x}}_i' \boldsymbol{\beta}_{\boldsymbol{\epsilon}0})]^{I\{e_i \in B_\theta(\boldsymbol{\epsilon})\}}}\right], \tag{A.2}$$

where $I\{A\}$ is the indicator function for A. This expression has the form of $2^J - 1$ logistic regression kernels, one for each $\boldsymbol{\epsilon}0$, times the likelihood for θ obtained from fitting the standard parametric family $N(0, \theta^2)$ to data $\mathbf{e} = (e_1, \ldots, e_n)$. This forms the basis of an efficient MCMC scheme for obtaining posterior inference.

A.2 MCMC DETAILS

In this appendix, we describe an efficient MCMC algorithm for obtaining a posterior sample $\{(\boldsymbol{\gamma}^{(s)}, \mathbf{e}^{(s)}, \theta^{(s)}, \boldsymbol{\beta}^{(s)})\}_{s=1}^{S}$. Based on this random sample, we can obtain

any posterior target of inference of interest. For instance, the covariate effect $\boldsymbol{\xi}$ is estimated by the posterior mean $\hat{\boldsymbol{\xi}} = \mathcal{S}^{-1}\sum_{s=1}^{\mathcal{S}} \boldsymbol{\xi}^{(s)}$; the centering parameter θ^2 is estimated by the $\hat{\theta}^2 = \mathcal{S}^{-1}\sum_{s=1}^{\mathcal{S}} \theta^{2(s)}$; the predictive frailty density given some specific cluster-level covariate $\mathbf{x}$ is estimated by $\hat{g}(e|\mathbf{x}) = \mathcal{S}^{-1}\sum_{s=1}^{\mathcal{S}} g(e|\theta^{(s)}, \boldsymbol{\beta}^{(s)})$ with g given in (A.1); the predictive survival function given covariate vector $\mathbf{w}$ is estimated by

$$\hat{S}(t|\mathbf{w}) = \mathcal{S}^{-1}\sum_{s=1}^{\mathcal{S}}\sum_{k=1}^{\mathcal{K}} \exp\left\{-\sum_{k=1}^{K(t)} \lambda_k^{(s)}\Delta_k(t)\exp\left\{\mathbf{w}'\boldsymbol{\xi}^{(s)} + e_k^{(s)}\right\}\right\}, \tag{A.3}$$

where $\{e_k^{(s)}\}_{k=1}^{\mathcal{K}}$ is a random sample from the frailty density $g(\cdot|\theta^{(s)}, \boldsymbol{\beta}^{(s)})$.

Updating the frailties

Modifying Jara et al. (2009), the random effects e_i can be updated as follows. For $i = 1, \ldots, n$, let $V_i(\boldsymbol{\gamma}, e_i) = [\theta^{-2} + \mathbf{1}'_{N_i}\mathbf{W}_i(\boldsymbol{\gamma}, e_i)\mathbf{1}_{N_i}]^{-1}$, where $\mathbf{1}_{N_i}$ is an $N_i \times 1$ vector of 1s and $\mathbf{W}_i(\boldsymbol{\gamma}, e_i) = \text{diag}(\mu_{ijk})$ is an $N_i \times N_i$ diagonal matrix. Then for the $(s+1)$th scan of the posterior distribution, the candidate $e_i^* \sim N(e_i^{(s)}, V_i(\boldsymbol{\gamma}, e_i^{(s)}))$ is accepted with probability

$$\alpha(e_i^*|e_i^{(s)}) = 1 \wedge \frac{L_i(\boldsymbol{\gamma}, e_i^*)g(e_i^*|\theta, \boldsymbol{\beta})\phi(e_i^{(s)}|e_i^*, V_i(\boldsymbol{\gamma}, e_i^*))}{L_i(\boldsymbol{\gamma}, e_i^{(s)})g(e_i^{(s)}|\theta, \boldsymbol{\beta})\phi(e_i^*|e_i^{(s)}, V_i(\boldsymbol{\gamma}, e_i^{(s)}))}, \tag{A.4}$$

where $L_i(\boldsymbol{\gamma}, e_i) = \prod_{j=1}^{n_i}\prod_{k=1}^{K(t_{ij})} p(y_{ijk}|\boldsymbol{\gamma}, e_i)$ and $\phi(\cdot|\mu, \sigma^2)$ represents the density function of a $N(\mu, \sigma^2)$ random variable.

Updating fixed effects parameters

Set $N = \sum_{i=1}^n N_i$, $\mathbf{y} = (\mathbf{y}'_1, \ldots, \mathbf{y}'_n)'$, $\boldsymbol{\mu}(\boldsymbol{\gamma}, \mathbf{e}) = E[\mathbf{y}|\boldsymbol{\gamma}, \mathbf{e}]$. Let $\mathbf{Z} = (\mathbf{z}'_{ijk})$ be an $N \times (K + p)$ design matrix and $\mathbf{W}(\boldsymbol{\gamma}, \mathbf{e}) = \text{diag}(\mu_{ijk})$ be an $N \times N$ diagonal matrix, where all subscripts ijk are in lexicographical order. Assume a normal prior for the fixed effects $\boldsymbol{\gamma} \sim N_{K+p}(\boldsymbol{\gamma}_0, \mathbf{S}_0)$. Let $V(\boldsymbol{\gamma}, \mathbf{e}) = [\mathbf{S}_0^{-1} + \mathbf{Z}'\mathbf{W}(\boldsymbol{\gamma}, \mathbf{e})\mathbf{Z}]^{-1}$ and

$$\mathbf{m}(\boldsymbol{\gamma}, \mathbf{e}) = \mathbf{V}(\boldsymbol{\gamma}, \mathbf{e})[\mathbf{S}_0^{-1}\boldsymbol{\gamma}_0 + \mathbf{Z}'\mathbf{W}(\boldsymbol{\gamma}, \mathbf{e})\mathbf{Z}\boldsymbol{\gamma} + \mathbf{Z}'(\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\gamma}, \mathbf{e}))].$$

Then for the $(s+1)$th scan of the posterior distribution, the candidate $\boldsymbol{\gamma}^*$ is generated from an $N_{K+p}(\mathbf{m}(\boldsymbol{\gamma}^{(s)}, \mathbf{e}), \mathbf{V}(\boldsymbol{\gamma}^{(s)}, \mathbf{e}))$ distribution with acceptance proability

$$\alpha(\boldsymbol{\gamma}^*|\boldsymbol{\gamma}^{(s)}) = 1 \wedge \frac{L(\boldsymbol{\gamma}^*, \mathbf{e})\phi_{K+p}(\boldsymbol{\gamma}^*|\boldsymbol{\gamma}_0, \mathbf{S}_0)\phi_{K+p}(\boldsymbol{\gamma}^{(s)}|\mathbf{m}(\boldsymbol{\gamma}^*, \mathbf{e}), \mathbf{V}(\boldsymbol{\gamma}^*, \mathbf{e}))}{L(\boldsymbol{\gamma}^{(s)}, \mathbf{e})\phi_{K+p}(\boldsymbol{\gamma}^{(s)}|\boldsymbol{\gamma}_0, \mathbf{S}_0)\phi_{K+p}(\boldsymbol{\gamma}^*|\mathbf{m}(\boldsymbol{\gamma}^{(s)}, \mathbf{e}), \mathbf{V}(\boldsymbol{\gamma}^{(s)}, \mathbf{e}))}, \quad \text{(A.5)}$$

where $L(\boldsymbol{\gamma}, \mathbf{e}) = \prod_{i=1}^n L_i(\boldsymbol{\gamma}, e_i)$ and $\phi_k(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ represents the density function of a $N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ random vector.

Updating the Tailfree process coefficients

A proposal based on one step of the Newton-Raphson algorithm (West, 1984; Gamerman, 1997) efficiently updates the parameters $\boldsymbol{\beta}_{\epsilon 0}$. To maintain identifiability of the model we set the first level conditional probability $Y_0(\mathbf{x}_i, \boldsymbol{\beta}_0) = 0.5$, i.e. $\boldsymbol{\beta}_0 = \mathbf{0}$. It follows that we need to update $\boldsymbol{\beta}_{\epsilon 0}$ for all $\boldsymbol{\epsilon} \in E_{J-1} \backslash \emptyset$ yielding $2^J - 2$ in total. Define "pseudodata"

$$\widetilde{e}_i(\boldsymbol{\beta}_{\epsilon 0}) = \tilde{\mathbf{x}}_i'\boldsymbol{\beta}_{\epsilon 0} + \frac{I\{e_i \in B_{\boldsymbol{\theta}}(\boldsymbol{\epsilon} 0)\} - Y_{\epsilon 0}(\mathbf{x}_i, \boldsymbol{\beta}_{\epsilon 0})}{Y_{\epsilon 0}(\mathbf{x}_i, \boldsymbol{\beta}_{\epsilon 0})[1 - Y_{\epsilon 0}(\mathbf{x}_i, \boldsymbol{\beta}_{\epsilon 0})]},$$

and weights

$$w_{ii}(\boldsymbol{\beta}_{\epsilon 0}) = Y_{\epsilon 0}(\mathbf{x}_i, \boldsymbol{\beta}_{\epsilon 0})[1 - Y_{\epsilon 0}(\mathbf{x}_i, \boldsymbol{\beta}_{\epsilon 0})] I\{e_i \in B_\theta(\boldsymbol{\epsilon})\},$$

placed in the vector $\tilde{\mathbf{e}}(\boldsymbol{\beta}_{\epsilon 0}) = (\widetilde{e}_1(\boldsymbol{\beta}_{\epsilon 0}), \ldots, \widetilde{e}_n(\boldsymbol{\beta}_{\epsilon 0}))'$ and matrix

$$\mathbf{W}(\boldsymbol{\beta}_{\epsilon 0}) = \text{diag}(w_{11}(\boldsymbol{\beta}_{\epsilon 0}), \ldots, w_{nn}(\boldsymbol{\beta}_{\epsilon 0}))'.$$

Let $\mathbf{V}_{\epsilon 0} = \frac{2n}{cj^2}(\mathbf{X}'\mathbf{X})^{-1}$ for $\boldsymbol{\epsilon} 0 \in E^j$, under the g-prior (7 in the paper) for $\boldsymbol{\beta}_{\epsilon 0}$, the M-H proposal is $\boldsymbol{\beta}^*_{\epsilon 0} \sim N_{q+1}(\mathbf{m}(\boldsymbol{\beta}^{(s)}_{\epsilon 0}), \mathbf{C}(\boldsymbol{\beta}^{(s)}_{\epsilon 0}))$ where

$$\mathbf{m}(\boldsymbol{\beta}_{\epsilon 0}) = \mathbf{C}(\boldsymbol{\beta}_{\epsilon 0})[\mathbf{X}'\mathbf{W}(\boldsymbol{\beta}_{\epsilon 0})\tilde{\mathbf{e}}(\boldsymbol{\beta}_{\epsilon 0})] \text{ and } \mathbf{C}(\boldsymbol{\beta}_{\epsilon 0}) = [\mathbf{V}_{\epsilon 0}^{-1} + \mathbf{X}'\mathbf{W}(\boldsymbol{\beta}_{\epsilon 0})\mathbf{X}]^{-1}. \quad \text{(A.6)}$$

This proposal is accepted with probability $\alpha(\boldsymbol{\beta}^*_{\epsilon 0}|\boldsymbol{\beta}^{(s)}_{\epsilon 0})$ defined by

$$\alpha(\boldsymbol{\beta}^*_{\epsilon 0}|\boldsymbol{\beta}^{(s)}_{\epsilon 0}) = 1 \wedge \left\{ \frac{\phi_{q+1}(\boldsymbol{\beta}^*_{\epsilon 0}|\mathbf{0}, \mathbf{V}_{\epsilon 0})}{\phi_{q+1}(\boldsymbol{\beta}_{\epsilon 0}|\mathbf{0}, \mathbf{V}_{\epsilon 0})} \frac{\phi_{q+1}(\boldsymbol{\beta}^{(s)}_{\epsilon 0}|\mathbf{m}(\boldsymbol{\beta}^*_{\epsilon 0}), \mathbf{C}(\boldsymbol{\beta}^*_{\epsilon 0}))}{\phi_{q+1}(\boldsymbol{\beta}^*_{\epsilon 0}|\mathbf{m}(\boldsymbol{\beta}^{(s)}_{\epsilon 0}), \mathbf{C}(\boldsymbol{\beta}^{(s)}_{\epsilon 0}))} q(\boldsymbol{\beta}^*_{\epsilon 0}|\boldsymbol{\beta}^{(s)}_{\epsilon 0}) \right\} \quad \text{(A.7)}$$

where

$$q(\boldsymbol{\beta}^*_{\boldsymbol{\epsilon}0}|\boldsymbol{\beta}^{(s)}_{\boldsymbol{\epsilon}0}) = \prod_{i:e_i\in B_\theta(\boldsymbol{\epsilon})} \frac{\exp\left\{\tilde{\mathbf{x}}_i'(\boldsymbol{\beta}^*_{\boldsymbol{\epsilon}0} - \boldsymbol{\beta}^{(s)}_{\boldsymbol{\epsilon}0})\right\}^{I\{e_i\in B_\theta(\boldsymbol{\epsilon}0)\}}}{\left[1+\exp(\tilde{\mathbf{x}}_i'\boldsymbol{\beta}^*_{\boldsymbol{\epsilon}0})\right]/\left[1+\exp(\tilde{\mathbf{x}}_i'\boldsymbol{\beta}^{(s)}_{\boldsymbol{\epsilon}0})\right]},$$

and $\phi_q(\cdot|\boldsymbol{\mu},\boldsymbol{\Sigma})$ is the density of a q-variate normal distribution with mean and covariance matrix $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively. Note that the M-H proposal and full conditional distribution for $\boldsymbol{\beta}_{\boldsymbol{\epsilon}0}$ only depends on the observations $\{(\mathbf{x}_i, e_i) : e_i \in B_\theta(\boldsymbol{\epsilon})\}$. Computational speed can be greatly increased by making use of this fact, especially at higher tree levels.

Updating θ

The centering parameter θ can be updated via a random walk M–H step. The proposal is $\log(\theta^*) \sim N(\log(\theta), v_0)$ for some v_0. This candidate θ^* is accepted with probability

$$\alpha(\theta^*|\theta^{(s)}) = 1 \wedge \frac{p(e_1,\ldots,e_n|\theta^*,\boldsymbol{\beta})\pi(\theta^*)}{p(e_1,\ldots,e_n|\theta,\boldsymbol{\beta})\pi(\theta)}, \tag{A.8}$$

where $\pi(\cdot)$ is the prior density of θ.

Summary of the MCMC scheme

Putting all of these updating steps together yields an efficient sampling algorithm for approximating the joint posterior distribution $p(\mathbf{e},\boldsymbol{\gamma},\boldsymbol{\beta},\theta|\mathcal{D})$, where $\mathcal{D} = \{\mathcal{D}_{ij} : i = 1,\ldots,n,\ j = 1,\ldots,n_i\}$ with $\mathcal{D}_{ij}$ being the ijth observed data point. Let $U(a,b)$ denote the uniform distribution with support $[a,b]$. Given the current values $(\mathbf{e}^{(s)},\boldsymbol{\gamma}^{(s)},\boldsymbol{\beta}^{(s)},\theta^{(s)})$ from the sth scan of the Markov chain, we obtain new values as follows:

1. Update $\mathbf{e}$: for each $i \in \{i,\ldots,n\}$,

   a) Generate a candidate e_i^* from $N(e_i^{(s)}, V_i(\boldsymbol{\gamma}^{(s)}, e_i^{(s)}))$.

   b) Compute the acceptance probability $\alpha(e_i^*|e_i^{(s)})$ in (A.4) and sample $u \sim U(0,1)$. If $u < \alpha(e_i^*|e_i^{(s)})$ assign $e_i^{(s+1)} = e_i^*$, otherwise $e_i^{(s+1)} = e_i^{(s)}$.

2. Update $\boldsymbol{\gamma}$:

   a) Generate a candidate $\boldsymbol{\gamma}^*$ from $N_{K+p}(\mathbf{m}(\boldsymbol{\gamma}^{(s)}, \mathbf{e}^{(s+1)}), \mathbf{V}(\boldsymbol{\gamma}^{(s)}, \mathbf{e}^{(s+1)}))$.

   b) Compute the acceptance probability $\alpha(\boldsymbol{\gamma}^*|\boldsymbol{\gamma}^{(s)})$ in (A.5) and sample $u \sim U(0,1)$. If $u < \alpha(\boldsymbol{\gamma}^*|\boldsymbol{\gamma}^{(s)})$ assign $\boldsymbol{\gamma}^{(s+1)} = \boldsymbol{\gamma}^*$, otherwise $\boldsymbol{\gamma}^{(s+1)} = \boldsymbol{\gamma}^{(s)}$.

3. Update $\boldsymbol{\beta}$: for each $\boldsymbol{\epsilon} \in E_{J-1} \backslash \emptyset$,

   a) Generate a candidate $\boldsymbol{\beta}^*_{\boldsymbol{\epsilon}0}$ from $N_{q+1}(\mathbf{m}(\boldsymbol{\beta}^{(s)}_{\boldsymbol{\epsilon}0}), \mathbf{C}(\boldsymbol{\beta}^{(s)}_{\boldsymbol{\epsilon}0}))$.

   b) Compute the acceptance probability $\alpha(\boldsymbol{\beta}^*_{\boldsymbol{\epsilon}0}|\boldsymbol{\beta}^{(s)}_{\boldsymbol{\epsilon}0})$ in (A.7) and sample $u \sim U(0,1)$. If $u < \alpha(\boldsymbol{\beta}^*_{\boldsymbol{\epsilon}0}|\boldsymbol{\beta}^{(s)}_{\boldsymbol{\epsilon}0})$ assign $\boldsymbol{\beta}^{(s+1)}_{\boldsymbol{\epsilon}0} = \boldsymbol{\beta}^*_{\boldsymbol{\epsilon}0}$, otherwise $\boldsymbol{\beta}^{(s+1)}_{\boldsymbol{\epsilon}0} = \boldsymbol{\beta}^{(s)}_{\boldsymbol{\epsilon}0}$. Note that $\mathbf{m}(\boldsymbol{\beta}^{(s)}_{\boldsymbol{\epsilon}0})$, $\mathbf{C}(\boldsymbol{\beta}^{(s)}_{\boldsymbol{\epsilon}0})$ and $\alpha(\boldsymbol{\beta}^*_{\boldsymbol{\epsilon}0}|\boldsymbol{\beta}^{(s)}_{\boldsymbol{\epsilon}0})$ also depend on the updated $\mathbf{e}^{(s+1)}$ in step 1.

4. Update θ:

   a) Generate a candidate $\log(\theta^*)$ from $N(\log(\theta), v_0)$ for some v_0.

   b) Compute the acceptance probability $\alpha(\theta^*|\theta^{(s)})$ in (A.8) and sample $u \sim U(0,1)$. If $u < \alpha(\theta^*|\theta^{(s)})$ assign $\theta^{(s+1)} = \theta^*$, otherwise $\theta^{(s+1)} = \theta^{(s)}$.

The Markov chain achieves approximate stationarity after a large enough burn-in period of iterations; see Tierney (1994) and Roberts and Smith (1994) for some general convergence conditions. After the convergence is established, a random sample, say $\{(\mathbf{e}^{(s)}, \boldsymbol{\gamma}^{(s)}, \boldsymbol{\beta}^{(s)}, \theta^{(s)})\}_{s=1}^{\mathcal{S}}$, from the posterior distribution can be obtained by saving only every kth scan to reduce chain correlations.

A.3 Sample R code to analyze the Iowa SEER data

Subsets of the SEER database are obtained from
`https://seer.cancer.gov/seertrack/data/request/`.

```
#########################################################################
# Breast cancer patients;
# t      -- Follow-up time in Months
# delta  -- Status: 1 = death; 0 = alive(censored)
# ClusterID -- County: State County code (1-99)
# -------- Individual-level Covariates ---------------------------------
# w1     -- Age:    age of the patient at diagnosis in complete years
# w2     -- Stage:  1 = Regional; 0 = Distant or Local
# w3     -- Stage:  1 = Distant;  0 = Regional or Local
# -------- County-level Covariates -------------------------------------
# xtf.Income-- Median Household Income / 1000 in 1993
# xtf.Poverty-- Percentage of families in poverty in 1990
# xtf.Edu -- Percent of Bachelor's degree or higher in 1990
# xtf.RUCC -- Rural-Urban Continuum Codes in 1993
#########################################################################
library(survival)
library(DPpackage)
library(coxme)
library(MASS)

#############################################################
# read the data and create variables
#############################################################
# Individual Level:
data = read.table("SEER_BreastCancer_IA.txt", header = TRUE)
d = data[order(data$County),];
ni = as.vector(table(d$County))
n = length(ni)
N = sum(ni)
ClusterID = rep(1:n, ni)
t = d$t
delta = d$Status
w1 = d$Age;
w2 = d$Regional
w3 = d$Distance

# County Level:
d2 = read.table("CountyCovariates.txt", head=T)
xtf.Income = d2$Income_93/1000
```

```
xtf.Poverty = d2$Poverty_Family_90*100
xtf.Edu = d2$Education_90*100
xtf.RUCC = d2$RUCC_93

# Choose the county-level covariate that need to be included
Xindex = c(4) # Index for choosing RUCC only
Xtf = as.matrix( cbind(xtf.Income, xtf.Poverty, xtf.Edu, xtf.RUCC)[,Xindex])
xtf = cbind(rep(1,n), Xtf)
X = apply(Xtf, 2, function(x) rep(x, ni))
Xnames = c("Income", "Poverty", "Education", "RUCC")[Xindex]

Windex = c(1:3) # Index for choosing individual-level covariates
Ws = cbind(w1, w2, w3)[,Windex]
W = cbind(ClusterID, Ws, X)
Wnames = c("Age", "Regional", "Distance")[Windex]
pw  = length(Wnames); px=length(Xnames); p=pw+px+1; q=px+1;

WX = W[,-1]
colnames(W) = c("ClusterID", Wnames, Xnames)

############################################################
# Fit coxme
############################################################
fitcoxme = coxme(Surv(t,delta)~W[,-1] + (1|ClusterID))
fitcoxme

############################################################
# Fit LDTFP
############################################################
# Breslow estimate of the baseline hazard based coxme
lambdacoxme=function(time)
{
  pred.e=as.vector(fitcoxme$frail$ClusterID)
  n.pred.e=length(pred.e)
  dummy <- rep(0,n.pred.e)
  msurvival=rep(0,length(time))
  fitcoxme.coeff=as.vector(fitcoxme$coefficients)
  tf   = t[delta == 1];
  nf   = length(tf);
```

```
  rt   = matrix(t,N,nf);
  rtf  = matrix(tf,N,nf,byrow=TRUE);
  logic= (rt >= rtf);
  N.pred.e=rep(pred.e,ni)
  Sn0  = as.vector( exp(as.vector(WX%*%fitcoxme.coeff) + N.pred.e)%*%logic )+1e-10
  Lambda=rep(0,length(time))
  ntime = length(time)
  for (i in 1:length(time)){
Lambda[i]=sum( 1/Sn0*(tf<=time[i]) )
  }
  lambda = (Lambda[2:ntime]-Lambda[-ntime])/(time[2:ntime]-time[-ntime])
  lambda
}
# Plot the Breslow estimate for hazard values at each month
time0   = seq(0,47, 1)
lambda0 = lambdacoxme(time0)
sfun0   = stepfun(time0[-1], c(lambda0,0), right=T);
plot(sfun0)

# Determine the cut-points by examing the above plot
cutpoint = c(1, 11, 16, 17, 19, 20, 25, 28, 29, 36, 40, 44, 47);
intervals=length(cutpoint)

# Estimated hazards based on the above cut-point,
# which will be used as the starting values of log(lambda) for LDTFP
hazards0 = rep(0, intervals)
hazards0[1] = mean( lambda0[1:(cutpoint[1])] )
# if intervals>=2
for (i in 2:intervals){
hazards0[i] = mean( lambda0[(cutpoint[i-1]+1): (cutpoint[i])] )
}
sfun1=stepfun(cutpoint, c(hazards0,0), right=T);
lines(sfun1, lwd=2, col=2)

# Function to make a row with '1' at ind----
onv = function(ind,len){onv=rep(0,len); onv[ind]=1; onv}

# Creat new data structure ---------
y={}; Zmat={}; tot=0; off={}; nW=ncol(W)
```

```
for(i in 1:N)
{
   tot=tot+1
   Zmat=matrix(append(Zmat,c(W[i,1:nW],onv(1,intervals))),c(nW+intervals,tot))
   off=append(off,min(cutpoint[1],t[i]))
   if(t[i]<=cutpoint[1] && delta[i]==1)
   {
      y=append(y,1)
   } else
   {
      y=append(y,0)
   }
   if (intervals>1)
   {
   for(j in 1:(intervals-1))
   {
       if(t[i]>cutpoint[j])
       {
  off=append(off,min(cutpoint[j+1],t[i])-cutpoint[j])
  tot=tot+1
  Zmat=matrix(append(Zmat,c(W[i,1:nW],onv(j+1,intervals))),c(nW+intervals,tot))
  if(t[i] <= cutpoint[j+1] && delta[i]==1)
  {
     y=append(y,1)
  } else
  {
     y=append(y,0)
  }
       }
   }
   }
}
Zmat = t(Zmat);
id = Zmat[,1];
loghazard = Zmat[,-(1:nW)];
Z = Zmat[,-1]    # design matrix for fixed effects
if ((p-1)==pw) {
colnames(Z)=c( Wnames, paste("loghazard",1:intervals, sep="") )
} else {
```

```
colnames(Z)=c( Wnames, Xnames, paste("loghazard",1:intervals, sep="") )
}

#############################################################
# Fit LDTFP: prior specifications and initial state
#############################################################
# Design matrix for prediction of frailties
xtfpred1=xtf[77,-1]; xtfpred2=xtf[80,-1]
wxpred1=c( c(68.8, 0, 1),xtfpred1); wxpred2=c(c(68.8, 0, 1),xtfpred2)
xpred=rbind( c(rep(0,pw), rep(0,p-1-pw), rep(0,(intervals))) ,
 c(rep(0,pw), rep(0,p-1-pw), rep(0,(intervals))))
xtfpred=rbind(c(1,xtfpred1),c(1,xtfpred2))
prediction=list(xpred=xpred,xtfpred=xtfpred,quans=c(0.025,0.50,0.975))

# Initial based on coxme
loghazards0= log(hazards0)
gammacox   = c(as.vector(fitcoxme$coefficients), loghazards0)
sigma2bcox = as.vector(fitcoxme$vcoef[[1]])
frailcox   = as.vector(fitcoxme$frail$ClusterID)

# Prior information:
maxJ = 4 #
prior=list(maxm=maxJ,alpha=1,mub=rep(0,(p-1+intervals)),a0=1,b0=1,
Sb=diag(rep(10000,(p-1+intervals))),taub1=4,taub2=2*sigma2bcox)

# Initial state
betatf   = matrix(0,nrow=(2**maxJ-1),ncol=q)
gamma0   = gammacox
sigma2b0 = sigma2bcox
frail    = frailcox
state = list(alpha=1,beta=gamma0,b=frail,sigma2b=sigma2b0,betatf=betatf)

# MCMC parameters
nskip=30
mcmc=list(nburn=1000,nsave=500,nskip=nskip,ndisplay=500)

# Fitting the model
mingrid=-1.5; maxgrid=1.5; ngrid=200;
xgrid = seq(mingrid, maxgrid, length.out=ngrid);
```

```
fitLDTF=LDTFPglmm(y=y,x=Z,roffset=log(off),g=id,family=poisson(log),
                xtf=xtf,prior=prior,prediction=prediction,
                grid=seq(mingrid,maxgrid,len=ngrid),
      mcmc=mcmc,state=state,status=FALSE)
# Results
sLDTF=summary(fitLDTF)
sLDTF

#-----------functions to plot the densities for frailties -------------
## convert a binary vector to decimal
bintodeci = function (x) {
nx=length(x); index=nx:1
deci=sum(x*2^(index-1))
deci
}
## generate one frailty from G_xtf
gentf = function (xtf, beta, theta2) {
nxtf  = length(xtf)
nbeta = length(beta)/nxtf
J     = log(nbeta+2)/log(2)
e   = rep(0,J)
e[1]  = rbinom(1,1,0.5)
for (j in 2:J){
pos  = 2^(j-1)-2 + bintodeci(e[1:j])/2 +1
betaj= beta[(nxtf*pos-nxtf+1):(nxtf*pos)]
xbeta= (xtf%*%betaj)[1,1]
probj= exp(xbeta) / (1+exp(xbeta))
e[j] = rbinom(1,1,1-probj)
}
m=bintodeci(e)
ulow= m/2^J; uup=(m+1)/2^J
u=runif(1,ulow, uup)
qnorm(u,0, sqrt(theta2))
}
## frailty density evaluated at e
denfrail = function (e, xtf, beta, theta2) {
nxtf = length(xtf)
nbeta = length(beta)/nxtf
J = log(nbeta+2)/log(2)
```

```
tmp = e/sqrt(theta2)
loglik= dnorm(e, 0, sqrt(theta2), log=T)
if (tmp > 4.0) tmp2 = 0.9999683
else if (tmp < -4.0) tmp2 = 3.167124e-05
else tmp2 = pnorm(e, 0, sqrt(theta2))
for (j in 2:J) {
indx = floor( 2^j*tmp2+1 )
for (k in 1:2^(j-1)){
if(indx==(2*k-1) | indx==2*k) {
pos = 2^(j-1)-2 + k
betaj = beta[(nxtf*pos-nxtf+1):(nxtf*pos)]
xbeta = (xtf%*%betaj)[1,1]
Y0 = exp(xbeta)/( 1+exp(xbeta) )
if (indx==(2*k-1)) loglik = loglik + log(Y0)
else loglik = loglik - log( 1+exp(xbeta) )
}
}
}
loglik = loglik + (J-1)*log(2)
exp(loglik)
}
## posterior mean of frailty density
densm = function(xgrid, xtf) {
nwx    = pw + px
ngrid  = length(xgrid)
betatf = fitLDTF$save.state$tfpsave
theta2 = fitLDTF$save.state$thetasave[,(nwx+intervals+1)]
nsave  = length(theta2)
dummy  = rep(0, nsave)
denm   = rep(0, ngrid)
denu   = denm; denl = denm;
for (i in 1:ngrid){
for (j in 1:nsave){
dummy[j] = denfrail(xgrid[i], c(1,xtf), betatf[j,], theta2[j])
}
denm[i] = mean (dummy)
}
denm
}
```

```
## posterior mean of frailty density shifted by the main effect
densmshift = function(xgrid, xtf) {
nwx    = pw + px
ngrid  = length(xgrid)
betatf = fitLDTF$save.state$tfpsave
theta2 = fitLDTF$save.state$thetasave[,(nwx+intervals+1)]
gammax = as.matrix( fitLDTF$save.state$thetasave[,(pw+1):(nwx)] )
nsave  = length(theta2)
dummy  = rep(0, nsave)
denm   = rep(0, ngrid)
denu   = denm; denl = denm;
for (i in 1:ngrid){
for (j in 1:nsave){
evalue   = xgrid[i] - (gammax[j,]%*%xtf)[1,1]
dummy[j] = denfrail(evalue, c(1,xtf), betatf[j,], theta2[j])
}
denm[i] = mean (dummy)
}
denm
}

#----------------- functinos to plot survival curves for LDTFP ---------
ScLDTF = function(t,wx,gamma.h,frailty,intervals)
{
p = length(wx)
k = 1
temp = exp(gamma.h[p+1])*min(cutpoint[1],t)
while (t>cutpoint[k] && k<intervals)
{
   temp = temp+exp(gamma.h[p+k+1])*(min(cutpoint[k+1],t)-cutpoint[k])
   k=k+1
}
exp(-temp*exp((wx%*%gamma.h[1:p])[1,1] + frailty))
}
## Function to plot survival curve
surLDTF = function (t, wx)
{
  nwx    = length(wx)
```

```
  xtf    = c( 1, wx[-(1:pw)] )
  betatf = fitLDTF$save.state$tfpsave
  theta2 = fitLDTF$save.state$thetasave[,(nwx+intervals+1)]
  gamma.h= fitLDTF$save.state$thetasave[,1:(nwx+intervals)]
  nsave  = nrow(gamma.h)
  nt     = length(t)
  dummy  = matrix(0, nsave, nt)
  n.pred.e=500;
  for (j in 1:nsave)
  {
   tmp = matrix(0, n.pred.e, nt);
   for (i in 1:n.pred.e) {
   pred.e=gentf (xtf, betatf[j,], theta2[j]);
   tmp[i,] = as.vector( sapply(t, ScLDTF, wx, gamma.h[j,], pred.e, intervals))
   }
   remove(tmp);
   dummy[j,] = colMeans(tmp);
  }
  colMeans(dummy)
}

#-------------------------- LPML and DIC ----------------------------
# density function evaluated at each data point
densfun = function(t,delta,wx,gamma.h,frailty,intervals){
p = length(wx)
k = 1
temp = exp(gamma.h[p+1])*min(cutpoint[1],t)
while (t>cutpoint[k] && k<intervals)
{
   temp = temp+exp(gamma.h[p+k+1])*(min(cutpoint[k+1],t)-cutpoint[k])
   k=k+1
}
lambda.K= gamma.h[p+k]
temp2 = (wx%*%gamma.h[1:p])[1,1] + frailty
exp((lambda.K+temp2)*delta-temp*exp(temp2))
}
## calculate LPML and DIC
nwx = pw + px
frails= fitLDTF$save.state$randsave
```

```
betatf= fitLDTF$save.state$tfpsave
theta2 = fitLDTF$save.state$thetasave[,(nwx+intervals+1)]
gamma.h = fitLDTF$save.state$thetasave[,1:(nwx+intervals)]
nsave = nrow(gamma.h)
cpo1 = rep(0, N)
cpo2 = rep(0, N)
dummy = rep(0, nsave)
for (i in 1:N){
Ni = rep(1:n, ni)
for (j in 1:nsave){
pred.e   = frails[j,Ni[i]]
dummy[j] = densfun(t[i],delta[i],W[i,-1],gamma.h[j,],pred.e,intervals)
}
cpo1[i] = 1/mean(1/dummy)
cpo2[i] = mean(log(dummy))
}
LPML.LDTF = sum(log(cpo1))
LPML.LDTF

Dbar = -2*sum(cpo2)
tmp = rep(0, N)
mean.pred.e = colMeans(frails)
mean.gamma.h = as.vector(colMeans(gamma.h))
for (i in 1:N){
Ni = rep(1:n, ni)
pred.e = mean.pred.e[Ni[i]]
tmp[i] = densfun(t[i],delta[i],W[i,-1],mean.gamma.h,pred.e,intervals)
}
pD = Dbar + 2*sum(log(tmp))
DIC.LDTF = Dbar+pD
DIC.LDTF

############################################################
# Plots: frailty densities at xtfpred1, xtfpred0 and xtfpred2
#     survival curves at wxpred1, wxpred0, wxpred2
############################################################
# Take xtfpred0 as the mean of county-level covariate values
xtf.quan = apply(Xtf, 2, function(x) quantile(x, c(0.05, 0.95)))
xtfpred0 = apply(Xtf, 2, mean);
```

```
if (Xtf[1,px]==8) {xtfpred0[px]=5; xtf.quan[,px]=c(2,9)}

# plot frailty densities and shifted version
xgrid2 = seq(-2.5, 2.5, length=300) + ( xtfpred0%*%fitLDTF$coeff[(pw+1):(pw+px)] )[1,1]
ngrid2 = length(xgrid2)
densm0 = densm(xgrid, xtfpred0)
densmshift0 = densmshift(xgrid2, xtfpred0)
densm1 = matrix(0,px,ngrid)
densm2 = densm1
densmshift1 = matrix(0,px,ngrid2)
densmshift2 = densmshift1
for (i in 1:px){
pindx = i
xtfpred1 = xtfpred0; xtfpred1[pindx] = xtf.quan[1,pindx]
xtfpred2 = xtfpred0; xtfpred2[pindx] = xtf.quan[2,pindx]
densm1[i,] = densm(xgrid, xtfpred1)
densm2[i,] = densm(xgrid, xtfpred2)
densmshift1[i,] = densmshift(xgrid2, xtfpred1)
densmshift2[i,] = densmshift(xgrid2, xtfpred2)
}


time = seq(0, max(t), length.out=200); ntime=length(time)
wpred0 = c(mean(d$Age),0,1)
wxpred0 = c(wpred0,xtfpred0)
mLDTF0 =surLDTF(time, wxpred0)
survm1 = matrix(0,px,ntime)
survm2 = survm1
for (i in 1:px){
pindx = i
xtfpred1 = xtfpred0; xtfpred1[pindx] = xtf.quan[1,pindx]
xtfpred2 = xtfpred0; xtfpred2[pindx] = xtf.quan[2,pindx]
wxpred1 = c(wpred0,xtfpred1)
wxpred2 = c(wpred0,xtfpred2)
survm1[i,] = surLDTF(time, wxpred1)
survm2[i,] = surLDTF(time, wxpred2)
}

#save.image("RUCC.RData")
```

```
par(mfrow=c(3,px))
par(mar = c(3, 3, 2, 1)+0.2)
par(mgp = c(2.2, 1, 0))
# frailty density
for (i in 1:px) {
plot(xgrid, densm0, "l", lty="solid", xlab="values", ylab="density",
ylim=c(0,1.9), lwd=2, main=paste("Model 1:", Xnames[i], sep=" " ))
lines(xgrid, densm1[i,], "l", lty="dashed", lwd=2, col=2)
lines(xgrid, densm2[i,], "l", lty="dotted", lwd=2, col=4)
legend(0.52, 1.9, c(paste(Xnames[i], "2", sep="="), paste(Xnames[i], "5", sep="="),
        paste(Xnames[i], "9", sep="=")), col = c(2,1,4),
       lty = c("dashed", "solid", "dotted"), cex=1)
}
# shifted frailty density
for (i in 1:px) {
plot(xgrid2, densmshift0, "l", lty="solid", xlab="values", ylab="density",
ylim=c(0,1.9), lwd=2, main=paste("Model 1:", Xnames[i], sep=" " ))
lines(xgrid2, densmshift1[i,], "l", lty="dashed", lwd=2, col=2)
lines(xgrid2, densmshift2[i,], "l", lty="dotted", lwd=2, col=4)
legend(0.4, 1.9, c(paste(Xnames[i], "2", sep="="), paste(Xnames[i], "5", sep="="),
        paste(Xnames[i], "9", sep="=")), col = c(2,1,4),
       lty = c("dashed", "solid", "dotted"), cex=1)
}
# survival curve
for (i in 1:px) {
plot(time, mLDTF0, "l", lty="solid", xlab="values", ylab="survival",
ylim=c(0,1), lwd=2, main=paste("Model 1:", Xnames[i], sep=" " ))
lines(time, survm1[i,], "l", lty="dashed", lwd=2, col=2)
lines(time, survm2[i,], "l", lty="dotted", lwd=2, col=4)
legend(31.5, 0.99, c(paste(Xnames[i], "2", sep="="), paste(Xnames[i], "5", sep="="),
        paste(Xnames[i], "9", sep="=")), col = c(2,1,4),
       lty = c("dashed", "solid", "dotted"), cex=1)
}

# Add the fitted baseline hazards plot based on LDTFP
# to the previous plot based on coxme
plot(sfun0, xlim=c(0,50))
lambda2 = as.vector(exp( fitLDTF$coefficients[p:(p+intervals-1)] ));
```

```
sfun2=stepfun(cutpoint, c(lambda2,0), right=T); lines(sfun2, lwd=2, col=4)
```

A.4 Additional simulation results

Effect of level J for larger sample size under Scenario I

To assess the effect of the level of the LDTFP on the posterior inferences, simulated 15 dat sets from the model with $n = 1,000$, under **Scenario I**. For each data set, we fitted different versions of the proposed model by considering $J = 4, 5, 6, 7$. Figures A.1-A.4 present the fitted frailty densities and survival curves averaged over replicates. For $J = 4$ (Figure A.1), we do see increased accuracy of estimated frailty and survival curves from increasing the number of clusters to $n = 1000$. Increasing from $J = 4$ to $J = 5$ does lead to a slight improvement on the estimated shape of the frailty density. Further increasing J to 6 or 7 does not help much, but roughly doubles the computing time for each additional level.

Comparison with the MPT frailty Cox model under Scenario I

Under **Scenario I**, we also fitted the exchangeable mixture of Polya trees (MPT) (Hanson, 2006a) frailty Cox's model using the function `PTglmm` available in `DPpackage` (Jara et al., 2011). The results for regression coefficients under the proposed model and the MPT approach are given in Table A.2. The average of the estimated frailty distributions and survival functions across simulation data sets for some specific covariate values are presented in Figure A.5, and the corresponding Monte Carlo mean and standard deviations for the ISE are given in Table A.3. The MPT approach outperforms the GF and PSF methods when the cluster-level covariate $x = 2$, but performs worse when $x = -2$. This is not surprising, since the MPT gives us a non-

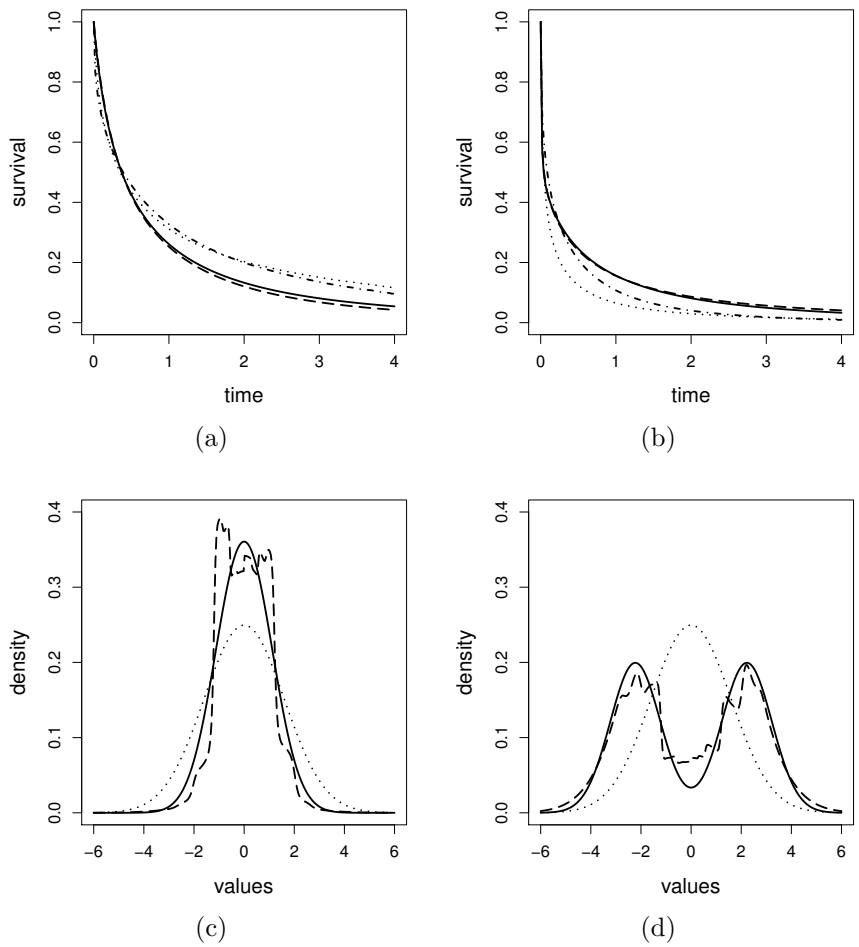


Figure A.1 Simulated data – Scenario III: Mean over simulations of the estimated curves under the proposed model with $J = 4$ and sample size $n = 1000$. Panels (a) and (b) show the results for the survival functions. Panels (c) and (d) show the results for the frailty densities. Panels (a) and (c) show the results for covariate values $(2, 1, -2)$. Panels (b) and (d) show the results for covariate values $(0, 1, 2)$. The true curves are represented by continuous lines. The results under the proposed model are represented by dashed lines. The results under the exchangeable Gaussian frailty model are represented by dotted lines. In Panels (a) and (b) the results obtained under of PSF approach are represented by a dot-dashed line.

parametric frailty distribution estimate which is a balance between one-mode density and two-mode density. Thus the performance of MPT estimates will depends on the value of cluster-level covariates.

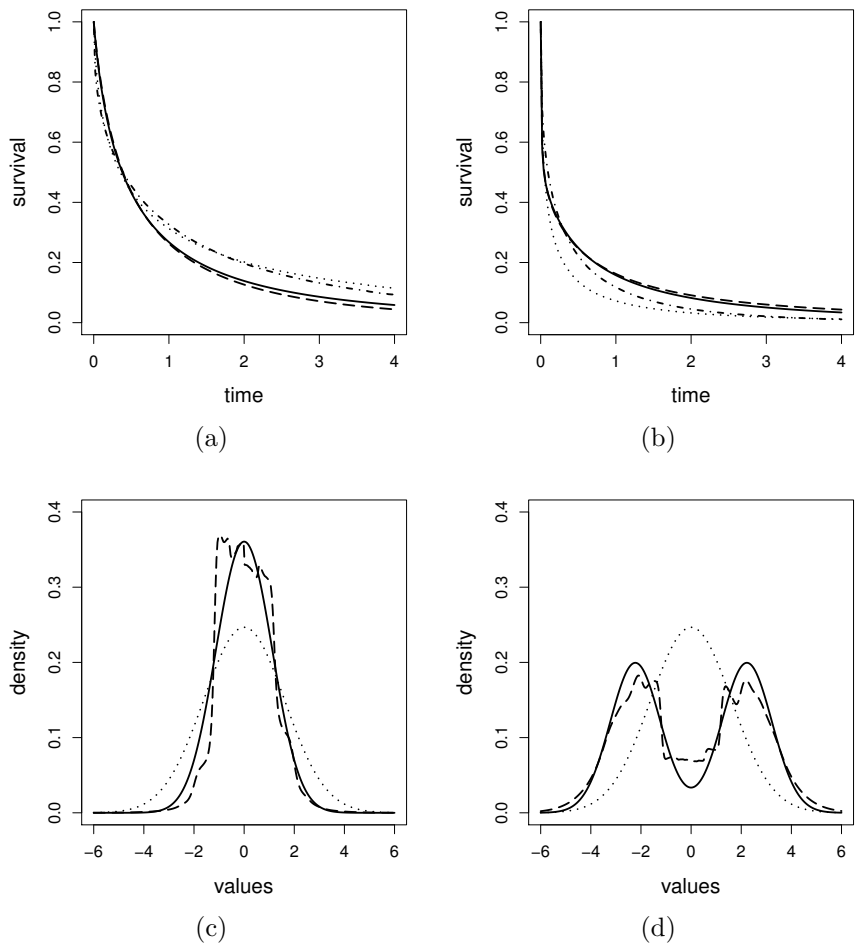


Figure A.2 Simulated data – Scenario III: Mean over simulations of the estimated curves under the proposed model with $J = 5$ and sample size $n = 1000$. Panels (a) and (b) show the results for the survival functions. Panels (c) and (d) show the results for the frailty densities. Panels (a) and (c) show the results for covariate values $(2, 1, -2)$. Panels (b) and (d) show the results for covariate values $(0, 1, 2)$. The true curves are represented by continuous lines. The results under the proposed model are represented by dashed lines. The results under the exchangeable Gaussian frailty model are represented by dotted lines. In Panels (a) and (b) the results obtained under of PSF approach are represented by a dot-dashed line.

Comparison with the GF approach under additional Scenario III

We also considered a third scenario favorable to the GF model to evaluate the behaviour of the proposed approach when a standard parametric exchangeable (covariate-free) frailty model is correct. Clustered failure time data were simulated in the same

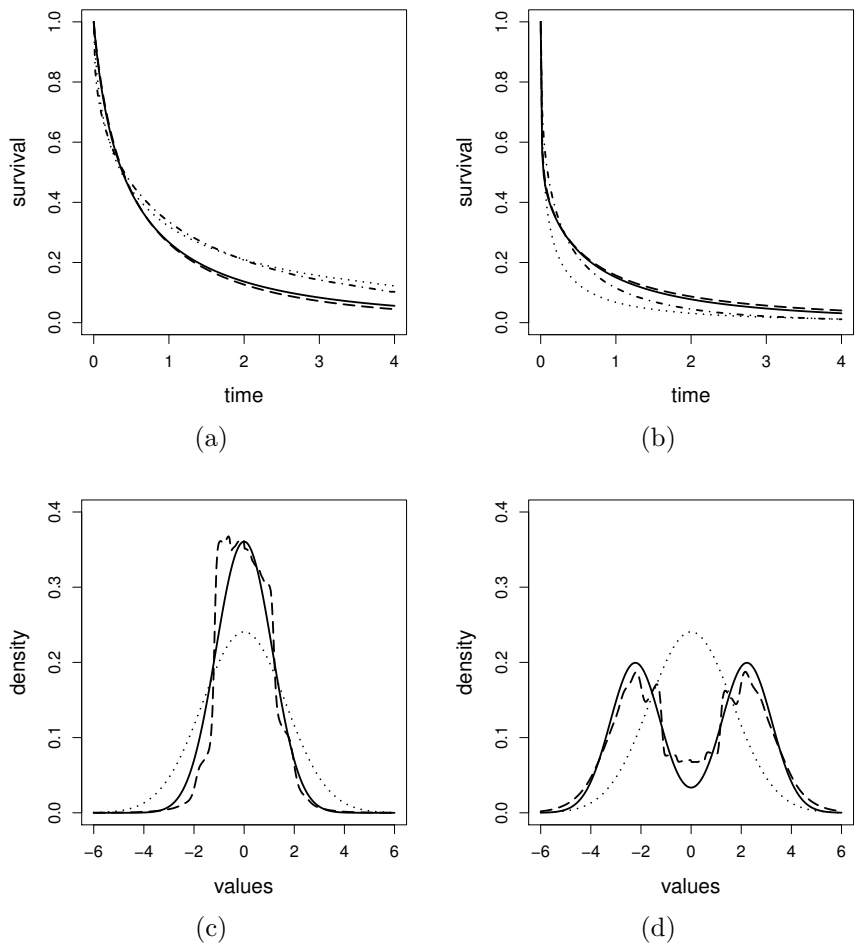


Figure A.3 Simulated data – Scenario III: Mean over simulations of the estimated curves under the proposed model with $J = 6$ and sample size $n = 1000$. Panels (a) and (b) show the results for the survival functions. Panels (c) and (d) show the results for the frailty densities. Panels (a) and (c) show the results for covariate values $(2, 1, -2)$. Panels (b) and (d) show the results for covariate values $(0, 1, 2)$. The true curves are represented by continuous lines. The results under the proposed model are represented by dashed lines. The results under the exchangeable Gaussian frailty model are represented by dotted lines. In Panels (a) and (b) the results obtained under of PSF approach are represented by a dot-dashed line.

way as Scenario I, but with frailties generated from standard normal distribution. The results for the regression coefficients under the proposed model and the GF approach are given in Table A.4. We can see that both approaches yielded unbiased estimates of $\boldsymbol{\xi}$ and θ^2, and almost the same values for MEAN-SD. However, the estimated standard error for ξ_x is severely underestimated by the GF approach, with CP of

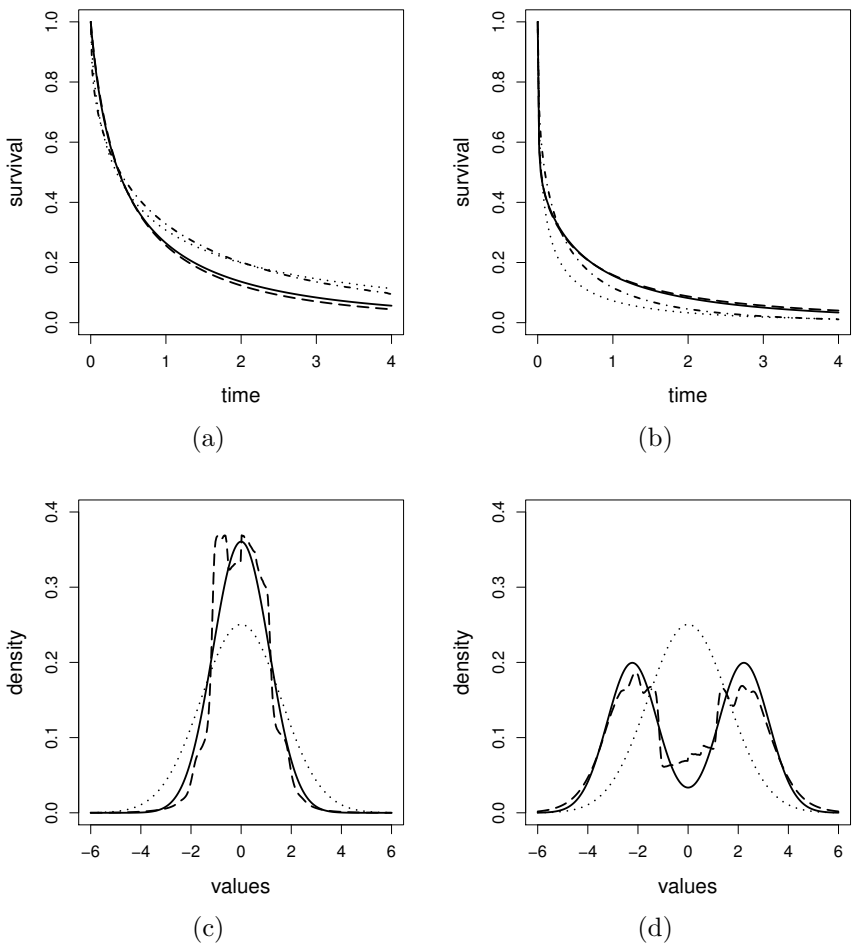


Figure A.4 Simulated data – Scenario III: Mean over simulations of the estimated curves under the proposed model with $J = 7$ and sample size $n = 1000$. Panels (a) and (b) show the results for the survival functions. Panels (c) and (d) show the results for the frailty densities. Panels (a) and (c) show the results for covariate values $(2, 1, -2)$. Panels (b) and (d) show the results for covariate values $(0, 1,\ 2)$. The true curves are represented by continuous lines. The results under the proposed model are represented by dashed lines. The results under the exchangeable Gaussian frailty model are represented by dotted lines. In Panels (a) and (b) the results obtained under of PSF approach are represented by a dot-dashed line.

86%. The average of the estimated frailty distributions and survival functions across simulated data sets for some specific covariate values are presented Figure A.4. The results suggest that essentially no differences among the three methods are observed; all estimated functions are close to the truth, indicating that there is little price to be paid for the extra generality when using the proposed model when normality and

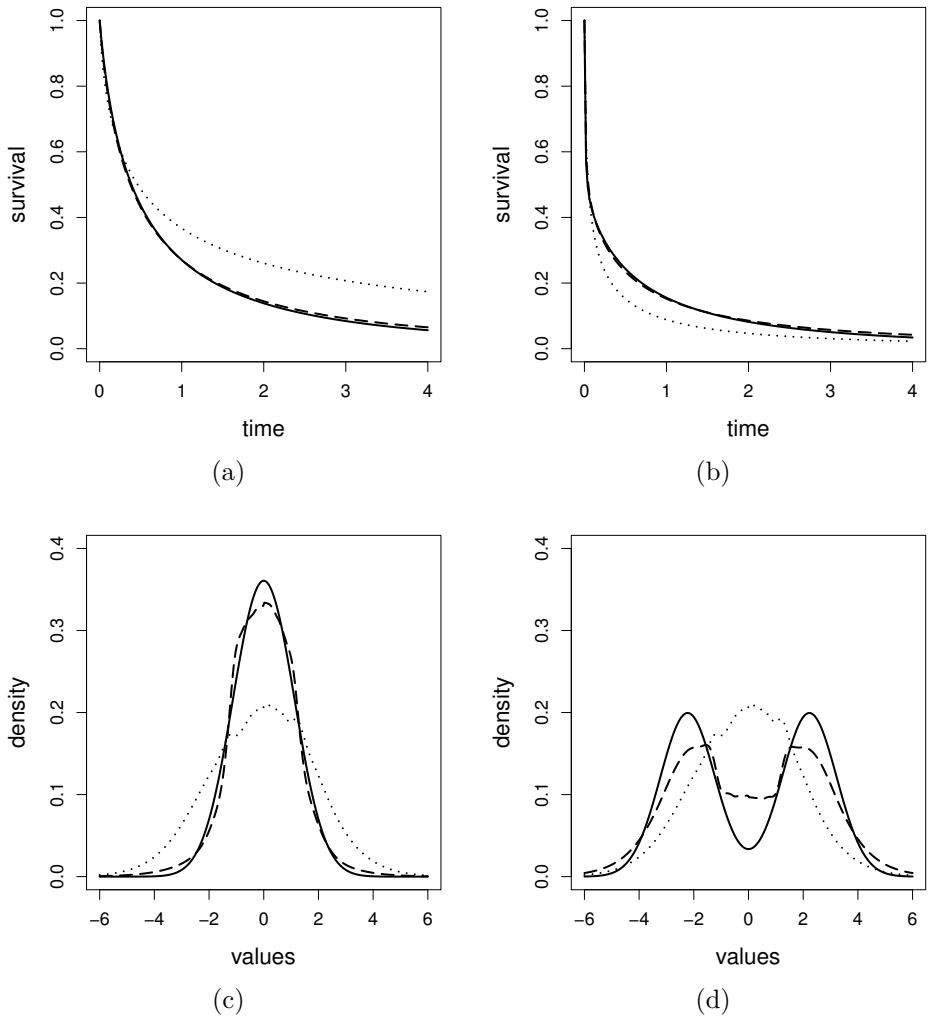


Figure A.5 Simulated data – Scenario I: Mean, across simulations, of the posterior mean of the survival and frailty density functions under the proposed model. Panels (a) and (b) show the results for the survival functions. Panels (c) and (d) show the results for the frailty densities. Panels (a) and (c) show the results for covariate values $(2, 1, -2)$. Panels (b) and (d) show the results for covariate values $(0, 1,\ 2)$. The true curves are represented by continuous lines. The results under the proposed model are represented by dashed lines. The results under the MPT frailty model are represented by dotted lines.

Table A.2 Simulation data – Scenario I: True value, bias of the point estimator, mean (across Monte Carlo simulations) of the posterior standard deviations/standard errors (MEAN-SD), standard deviation (across Monte Carlo simulations) of the point estimator (SD-MEAN) and Monte Carlo coverage probability for the 95% credible interval/confidence interval (CP) for the regression parameters. The results are presented under the proposed model and under the MPT approach.

		Proposed Model				MPT Model			
Parameters	True	BIAS	MEAN-SD	SD-MEAN	CP	BIAS	MEAN-SD	SD-MEAN	CP
ξ_1	1.0	0.011	0.052	0.054	0.930	0.009	0.046	0.041	0.985
ξ_2	0.5	0.008	0.088	0.090	0.945	0.005	0.087	0.087	0.945
ξ_x	1.0	-0.009	0.141	0.126	0.965	0.081	0.133	0.176	0.770

Table A.3 Simulated data – Scenario I: Monte Carlo mean (Monte Carlo standard deviation) for the ISE of the survival function for two different predictor values. The results for the different approaches under both simulation scenarios are presented. The numbers correspond to 10^3 times the original values.

(w_1, w_2, x)	Proposal	GF approach	PSF approach	MPT approach
$(2, 1, -2)$	2.02 (2.48)	4.37 (3.46)	6.28 (3.49)	7.71 (6.03)
$(0, 1, 2)$	1.94 (2.53)	10.5 (6.86)	14.3 (10.9)	8.19 (5.55)

exchangeability are valid assumptions.

Table A.4 Simulation data – Scenario III: True value, bias of the point estimator, mean (across Monte Carlo simulations) of the posterior standard deviations/standard errors (MEAN-SD), standard deviation (across Monte Carlo simulations) of the point estimator (SD-MEAN) and Monte Carlo coverage probability for the 95% credible interval/confidence interval (CP) for the regression parameters. The results are presented under the proposed model and under the GF approach.

		Proposed Model				GF Model			
Parameters	True	BIAS	MEAN-SD	SD-MEAN	CP	BIAS	MEAN-SD	SD-MEAN	CP
ξ_1	1.0	-0.002	0.044	0.045	0.940	-0.005	0.050	0.054	0.920
ξ_2	0.5	0.002	0.084	0.088	0.940	0.001	0.085	0.091	0.935
ξ_x	1.0	0.002	0.087	0.068	0.975	-0.018	0.055	0.070	0.860
θ^2	1.0	0.042		0.209		-0.036		0.185	

In addition, the results of the comparison of the estimated survival curves are presented in Table A.5, where the Monte Carlo mean and standard deviations for the ISE for two different predictor values are given. Even when an exchangeable frailty model with normal distribution applies, the proposed model is slightly more beneficial in estimating the survival functions. This may partly come from the differ-

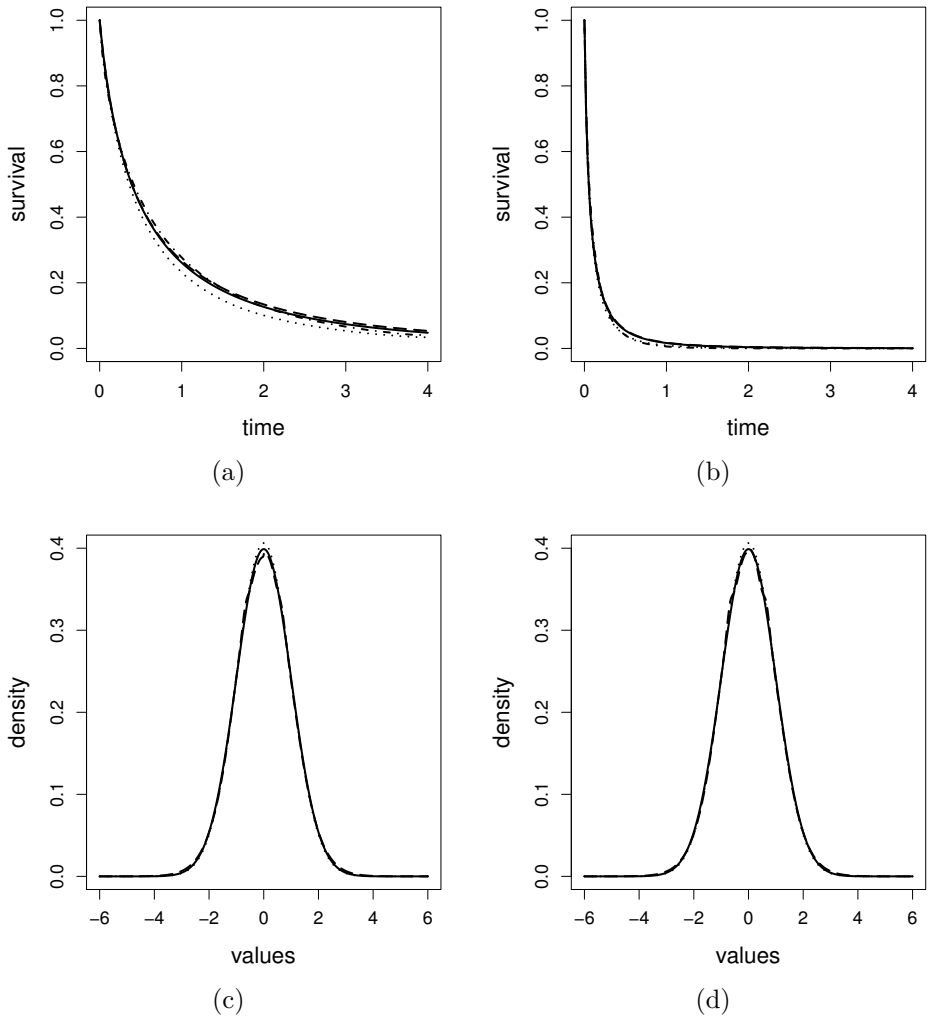


Figure A.6 Simulated data – Scenario III: Mean, across simulations, of the posterior mean of the survival and frailty density functions under the proposed model. Panels (a) and (b) show the results for the survival functions. Panels (c) and (d) show the results for the frailty densities. Panels (a) and (c) show the results for covariate values $(2, 1, -2)$. Panels (b) and (d) show the results for covariate values $(0, 1,\ 2)$. The true curves are represented by continuous lines. The results under the proposed model are represented by dashed lines. The results under the exchangeable Gaussian frailty model are represented by dotted lines. In Panels (a) and (b) the results obtained under of PSF approach are represented by a dot-dashed line.

ence between the Bayesian and frequentist methods for computation. The Bayesian approach averages over the posterior whereas the frequentist approach uses "plug-in" estimates which can underestimate variability.

Table A.5 Simulated data – Scenario III: Monte Carlo mean (Monte Carlo standard deviation) for the ISE of the survival function for two different predictor values. The results for the different approaches are presented. The numbers correspond to 10^3 times the original values.

(w_1, w_2)	Proposal	GF approach	PSF approach
$(2, 1)$	1.66 (2.66)	2.02 (2.89)	2.38 (3.46)
$(0, 1)$	1.53 (1.88)	1.70 (1.96)	2.16 (2.46)

A.5 Additional analysis of SEER data

Additional cut-point vector specifications

We consider additional cut-point vector specifications as follows:

Case IV: $\mathbf{a} = (2, 4, 7, 11, 15, 18, 22, 26, 32, 47)$, which is determined by the quantiles of the distribution of event times based on Kaplan-Meier curve so that each interval contains almost equal number of events.

Case IVb: $\mathbf{a} = (1, 3, 5, 8, 11, 13, 16, 18, 22, 25, 29, 34, 47)$, which is determined in the same way as **Case IV** but with size 13.

Case IIb: $\mathbf{a} = (2, 5, 9, 12, 15, 17, 21, 25, 28.2, 32.6, 37, 42.5, 47)$, where a_k is the $\frac{k}{13}$th quantile of the empirical distribution of observed survival times.

Table A.6 shows the DIC and LPML for the three models considered in the main article. We can see that **Case IV** gives a little better fit than the **Case II**, but still much worse than the **Case I**. It suggests that the Kaplan-Meier based specification of the cut-points provides slightly better fit than the empirical distribution based specification. When we increase the size of cut-point vector from 10 to 13 for **Case**

II and **Case IV**, the model fit only improves 1 unit for LPML and 4 units for DIC across the three models, compared with the Case II with 10 cut-points. It indicates that carefully choosing the cut-points is more important than simply increasing the number of cut-points.

Table A.6 Iowa SEER data: Deviance information criteria (DIC) and log of the pseudo marginal likelihood (LPML) for models under consideration.

	Case IV		**Case IVb**		**Case IIb**	
Model	DIC	LPML	DIC	LPML	DIC	LPML
1	4457	-2232	4456	-2232	4459	-2233
2	4459	-2233	4460	-2234	4461	-2234
3	4458	-2232	4459	-2234	4459	-2234

Figure A.7 presents the fitted predictive frailty densities for both $\mathbf{e}_i$ (median-zero) and $\mathbf{e}_i + \mathbf{x}_i'\boldsymbol{\xi}_x$ and survival curves for women with mean entry age 68.8 years and distant stage of disease who live in the counties with different levels of median household income or RUCC, under **Model 1** and **Model 2** for **Case IV**. We can see that all the estimated curves are very similar to those obtained under **Case I** considered in the main article.

Additional comparison between ours and marginal PH model

We additionally presented the fixed effects under the marginal non-frailty PH model (i.e. using the R function `coxph` with the option `cluster`) across **Model 1**–**Model 3** in Table A.7. Note that the marginal PH model is equivalent to the PSF model from the marginal model perspective, so these fixed effects are exactly the same as those obtained under the PSF model. Note that the coefficient estimates under the marginal PH model have population-averaged interpretations, and cannot be directly compared with those fitted from the proposed frailty PH model due to different model structures.

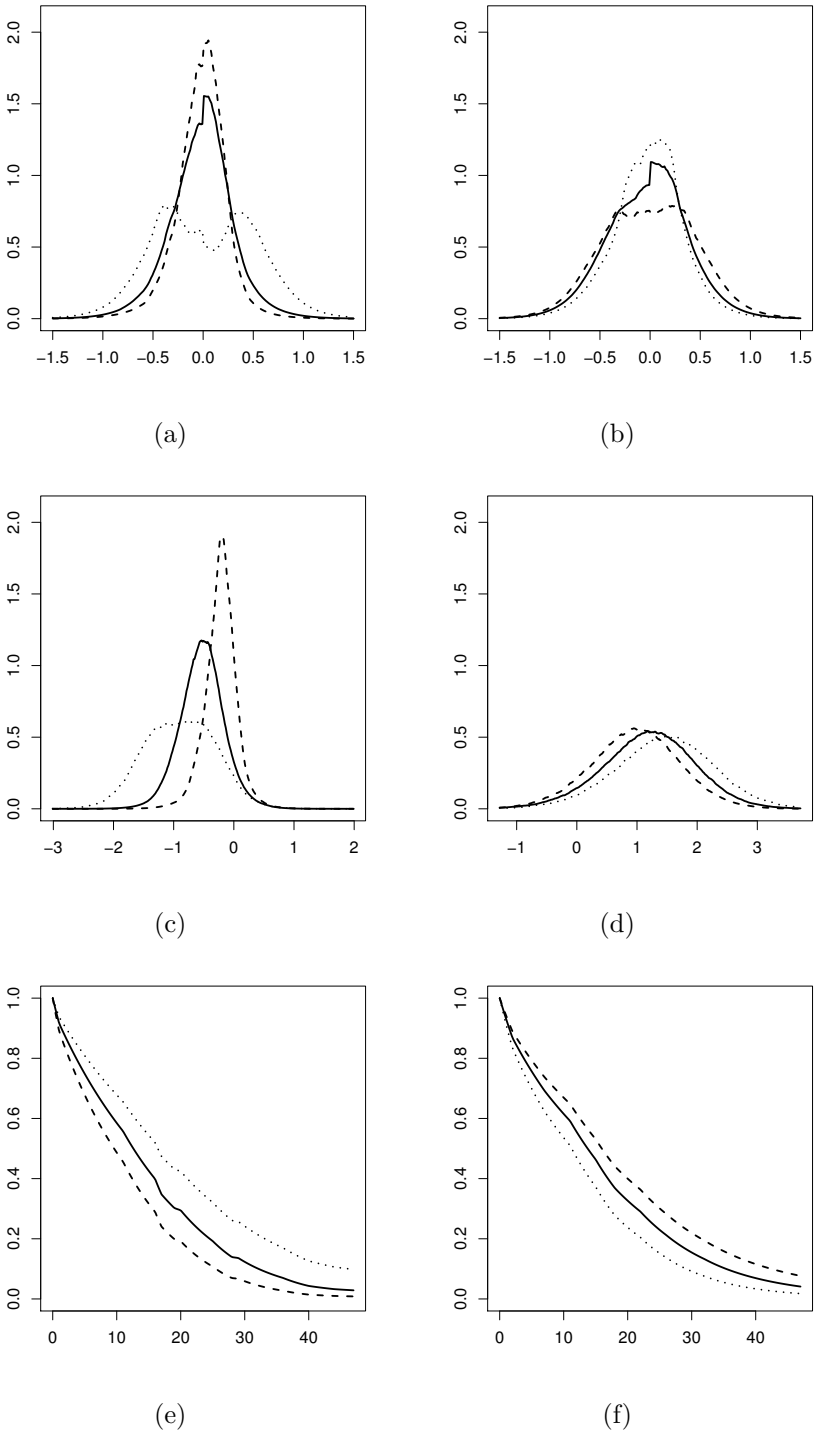


Figure A.7 Iowa SEER data: Fitted predictive frailty densities (Panels (a) and (b)), frailty densities with location shifts (Panels (c) and (d)) and survival curves (Panels (e) and (f)) for women with mean entry age 68.8 years and distant stage of disease who live in the counties with different county covariate levels under Model 1 (Panels (a), (c) and (e)) and Model 2 (Panels (b), (d) and (f)) under **Case IV**. In Panels (a), (c) and (e), the results for RUCC=2, 5 and 9 are displayed as dashed, continuous and dotted lines, respectively. In Panels (b), (d) and (f), the results for Income=23.354, 29.176 and 35.301 are displayed as dashed, continuous and dotted lines, respectively.

Table A.7 Iowa SEER data: Point estimates (95% confidence intervals) of fixed effects $\boldsymbol{\xi}$ from various models under the marginal PH model.

Predictor	**Model 1**	**Model 2**	**Model 3**	**Model 0**
ξ_1 (Age)	0.02	0.021	0.021	0.019
	(0.014, 0.026)	(0.015, 0.027)	(0.015, 0.027)	(0.013, 0.025)
ξ_2 (Regional)	0.30	0.30	0.30	0.30
	(0.08, 0.51)	(0.09, 0.52)	(0.08, 0.52)	(0.09, 0.50)
ξ_3 (Distant)	1.63	1.68	1.64	1.64
	(1.35, 1.92)	(1.39, 1.97)	(1.36, 1.93)	(1.34, 1.95)
ξ_{x_1} (RUCC)	-0.106		-0.087	
	(-0.154, -0.058)		(-0.152, -0.022)	
ξ_{x_2} (Income)		0.049	0.014	
		(0.019, 0.079)	(-0.025, 0.054)	

Although the fixed effects are not appropriate to be compared between the proposed model and the marginal PH model, we carefully compared the fitted survival curves across **Model 1**–**Model 3** for different covariates levels in Figure A.8 and A.9. Overall, the marginal PH model under-predicts survival time up to about 1 month compared with our proposed model for patients with mean entry age 68.8 years and distant stage of disease who live in the same county. It may be partly due to the fact that the marginal PH model and the PSF model fail to detect the bi-modal behavior of the frailty distribution for rural counties.

Measures of dependence within cluster

Kendall's tau is widely used to measure the overall dependence of a pair of subjects over the entire lifespan by integrating over time, which is based on a form of dependence known as *c*oncordance (Nelsen, 2006). As suggested by one of the referees, we calculated the Kendall's tau for different level of county-specific covariates under the proposed model. Suppose a posterior sample $\{(\boldsymbol{\gamma}^{(s)}, \mathbf{e}^{(s)}, \theta^{(s)}, \boldsymbol{\beta}^{(s)})\}_{s=1}^{S}$ has been obtained. For the sth iteration, we first draw a random sample, say $\{e_k^{(s)}\}_{k=1}^{\mathcal{K}}$, from the posterior frailty density $g_{\mathbf{x}}(\cdot|\theta^{(s)}, \boldsymbol{\beta}^{(s)})$, then draw a pair of survival times $(t_{1,k}^{(s)}, t_{2,k}^{(s)})$ independently from the predictive survival function $S_{\mathbf{w}}(t|\boldsymbol{\gamma}^{(s)}, e_k^{(s)})$, where $\mathbf{w} = (\tilde{\mathbf{w}}', \mathbf{x}')'$

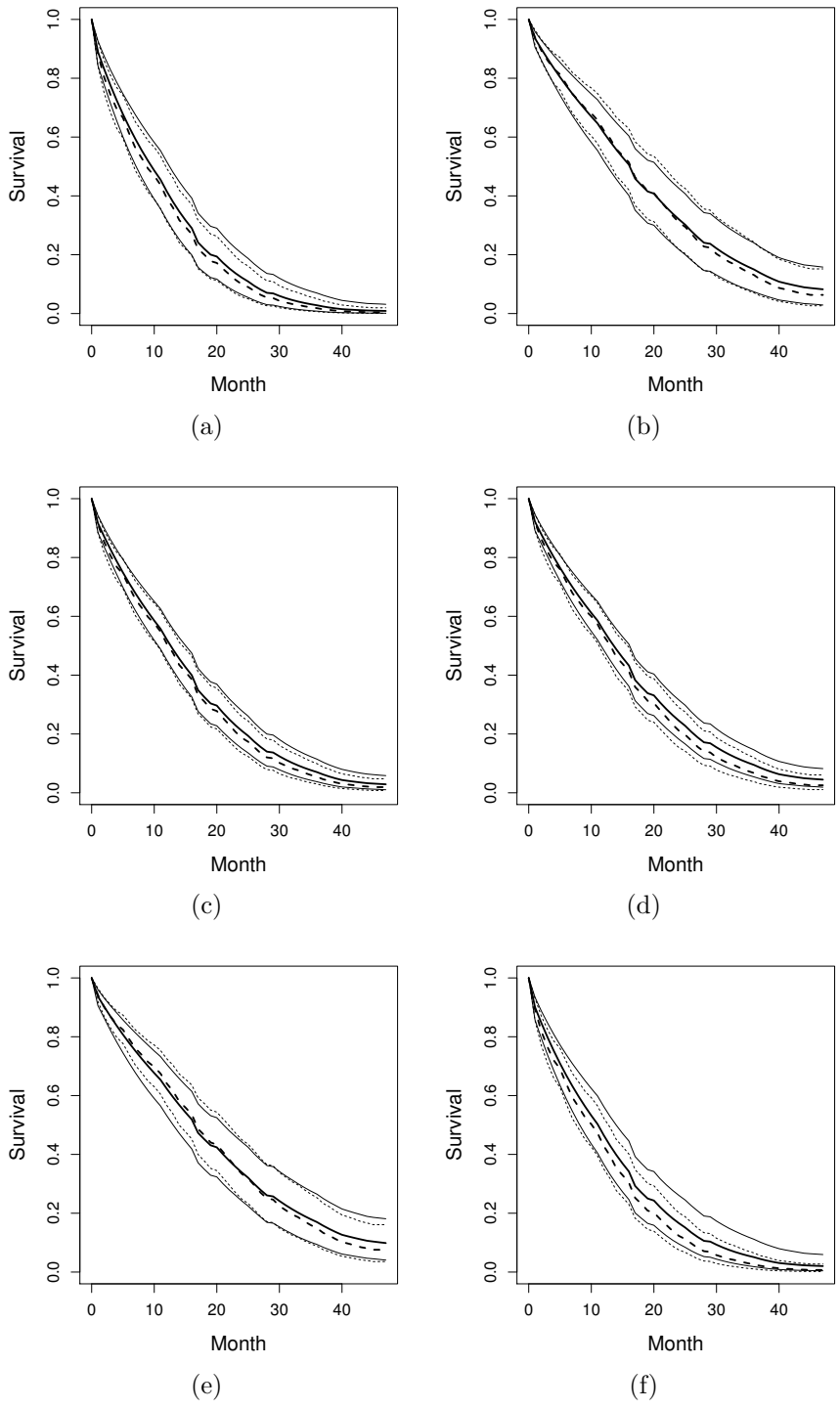


Figure A.8 Iowa SEER data: Fitted predictive survival curves (thicker lines) with 95% confidence/credible intervals (thinner lines) for women with mean entry age 68.8 years and distant stage of disease who live in the counties with different county covariate levels under the proposed model (solid lines) and under the marginal PH model (dashed lines). Panels (a), (c) and (e) are for RUCC=2, 5 and 9, respectively under Model 1. Panels (b), (d) and (f) are for Income=23.354, 29.176 and 35.301, respectively under Model 2.

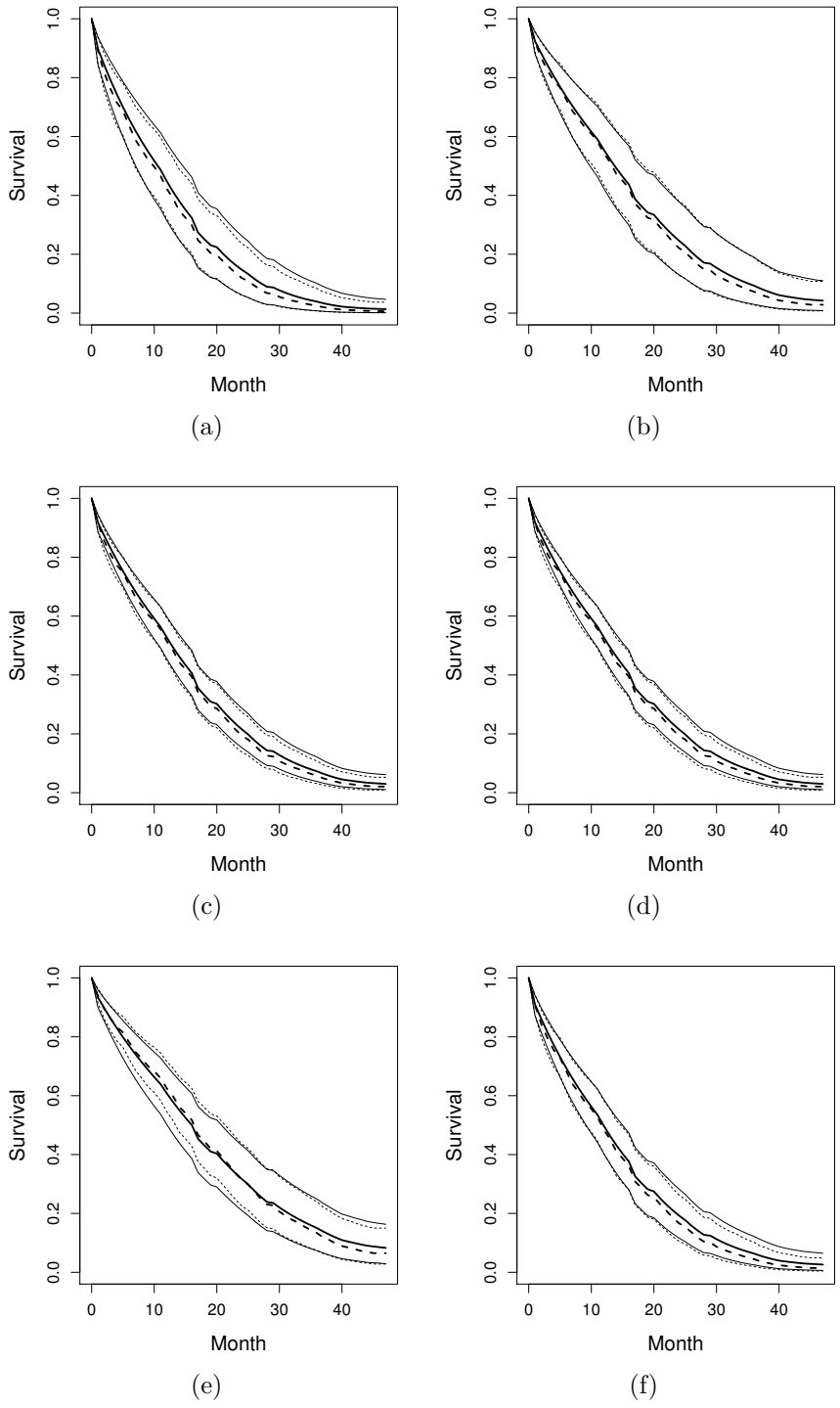


Figure A.9 Iowa SEER data: Fitted predictive survival curves (thicker lines) with 95% confidence/credible intervals (thinner lines) for women with mean entry age 68.8 years and distant stage of disease who live in the counties with different county covariate levels under the proposed model (solid lines) and under the marginal PH model (dashed lines). Panels (a), (c) and (e) are for RUCC=2, 5 and 9. Panels (b), (d) and (f) are for Income=23.354, 29.176 and 35.301.

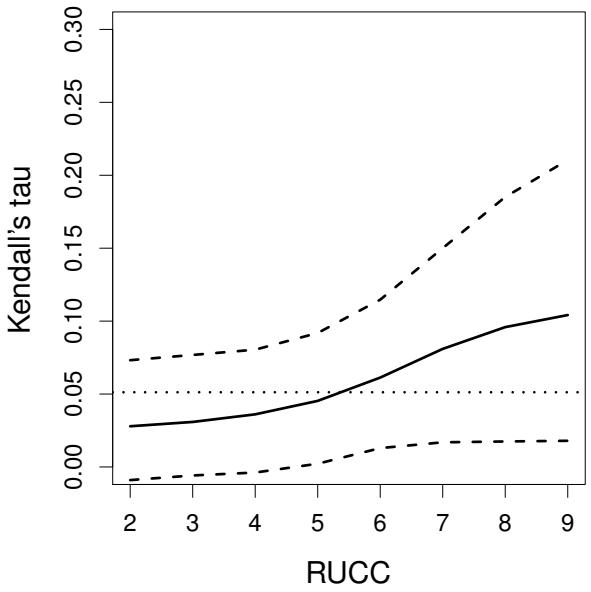


Figure A.10 Iowa SEER data: Estimated Kendall's tau as a function of RUCC (solid line) with point-wise 95% credible interval (dashed line) for individuals with mean entry age 68.8 years and distant stage of disease under Model 1. The Kendall's tau under the gamma frailty PH model is also presented in dotted line.

with $\tilde{\mathbf{w}}$ being any fixed vector of individual level covariates. Now there are a total of $\mathcal{K}(\mathcal{K}-1)/2$ pairs of survival times $(t_{1,k}^{(s)}, t_{2,k}^{(s)})$ and $(t_{1,\ell}^{(s)}, t_{2,\ell}^{(s)})$ among the random sample $\{(t_{1,k}^{(s)}, t_{2,k}^{(s)})\}_{k=1}^{\mathcal{K}}$. We then denote by $C_{\mathbf{x}}^{(s)}$ the total number of concordant pairs, that is, if $(t_{1,k}^{(s)} - t_{1,\ell}^{(s)})(t_{2,k}^{(s)} - t_{2,\ell}^{(s)})$ is strictly positive. It follows that the Kendall's tau for the sth iteration can be estimated by

$$\tau_s(\mathbf{x}) = \frac{2C_{\mathbf{x}}^{(s)}}{\mathcal{K}(\mathcal{K}-1)/2} - 1.$$

Thus we could estimate $\tau(\mathbf{x})$ and its credible interval based on the posterior sample $\{\tau_s(\mathbf{x})\}_{s=1}^{\mathcal{S}}$.

For example, Figure A.10 shows the estimated Kendall's tau for each possible RUCC level (from 2 to 9) with point-wise 95% credible interval under the **Model I** considered in the main body of paper. We can see that the within cluster correlation increases with RUCC, which agrees perfectly with the finding that frailty variance increases with RUCC. For comparison, we also calculated the Kendall's tau for the

gamma frailty PH model, which is static at a constant value, as shown in Figure A.10 with dotted line.

Appendix B

Supplement to Chapter 3

B.1 MCMC sampling

The full likelihood function for the data is given by

$$\begin{aligned}
\mathcal{L}(\mathbf{y}, \mathbf{B}, \boldsymbol{\sigma}, \mathbf{K}, \mathbf{V}, \alpha, \boldsymbol{\mu}, \boldsymbol{\theta}) \propto & \prod_{i=1}^{n} (\sigma^2_{K_i})^{-1/2} \exp\left\{-\frac{1}{2\sigma^2_{K_i}}(y_i - \mathbf{x}_i'\boldsymbol{\beta}_{K_i})^2\right\} \\
& \times \{\delta_i I(y_i = y_i^o) + (1-\delta_i) I(y_i > y_i^o)\} \\
& \times |\mathbf{C}_{\boldsymbol{\theta}}|^{-1/2} \exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{C}_{\boldsymbol{\theta}}^{-1} - \mathbf{I}_n)\mathbf{z}\right\} \\
& \times \prod_{i=1}^{n}\left[V_{K_i} \prod_{k<K_i}(1-V_k)\right] \\
& \times \left\{\prod_{k=1}^{N-1} \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)}(1-V_k)^{\alpha-1}\right\} \times \alpha^{a_0-1}\exp\{-b_0\alpha\} \\
& \times \prod_{k=1}^{N} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}_k - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}_k - \boldsymbol{\mu})\right\} \\
& \times \prod_{k=1}^{N} \left(\sigma_k^{-2}\right)^{\nu_a - 1} \exp\left\{-\nu_b \sigma_k^{-2}\right\} \\
& \times |\mathbf{S}_0|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\mu} - \mathbf{m}_0)'\mathbf{S}_0^{-1}(\boldsymbol{\mu} - \mathbf{m}_0)\right\} \\
& \times |\boldsymbol{\Sigma}|^{-(\kappa_0 - p - 1)/2} \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\kappa_0 \boldsymbol{\Sigma}_0 \boldsymbol{\Sigma}^{-1}\right)\right\} \\
& \times \theta_1^{\theta_{1a}-1}(1-\theta_1)^{\theta_{1b}-1} \times \theta_2^{\theta_{2a}-1} e^{-\theta_{2b}\theta_2}
\end{aligned} \tag{B.1}$$

Step 1: Update K_i for $i = 1, \ldots, n$.

The full conditional distribution for K_i is

$$\begin{aligned}
f(K_i|\text{else}) &\propto \omega_{K_i} (2\pi\sigma^2_{K_i})^{-1/2} \exp\left\{-\frac{1}{2\sigma^2_{K_i}}(y_i - \mathbf{x}_i'\boldsymbol{\beta}_{K_i})^2\right\} \\
&\propto \omega_{K_i} \phi(y_i | \mathbf{x}_i'\boldsymbol{\beta}_{K_i}, \sigma^2_{K_i}).
\end{aligned}$$

It follows that

$$P(K_i = k) = \frac{\omega_k \phi(y_i|\mathbf{x}_i'\boldsymbol{\beta}_k, \sigma_k^2)}{\sum_{k=1}^{N} \omega_k \phi(y_i|\mathbf{x}_i'\boldsymbol{\beta}_k, \sigma_k^2)}.$$

Step 2: Update y_i for $i = 1, \ldots, n$.

The full conditional distribution for y_i is

$$f(y_i|\text{else}) \propto \delta_i I(y_i = y_i^o) + (1-\delta_i)\phi(y_i|\mathbf{x}_i'\boldsymbol{\beta}_{K_i}, \sigma_{K_i}^2) \exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{C}_{\boldsymbol{\theta}}^{-1} - \mathbf{I}_n)\mathbf{z}\right\} I(y_i > y_i^o).$$

If $\delta_i = 1$, update $y_i = y_i^o$. If $\delta_i = 0$, propose y_i^* from $N(\mathbf{x}_i'\boldsymbol{\beta}_{K_i}, \sigma_{K_i}^2)$ distribution truncated above at y_i^o and accept it with probability

$$\min\left\{1, \frac{\exp\left\{-\frac{1}{2}\mathbf{z}^{*\prime}(\mathbf{C}_{\boldsymbol{\theta}}^{-1} - \mathbf{I}_n)\mathbf{z}^*\right\}}{\exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{C}_{\boldsymbol{\theta}}^{-1} - \mathbf{I}_n)\mathbf{z}\right\}}\right\},$$

where $\mathbf{z}^* = (z_i^*, \ldots, z_n^*)'$ is the new transformed vector corresponding to y_i^*.

Step 3: Update $\boldsymbol{\beta}_k$ for $k = 1, \ldots, N$ using Metropolis-Hastings algorithms with delayed rejection (Tierney and Mira, 1999).

The full conditional distribution for $\boldsymbol{\beta}_k$ is

$$\begin{aligned}
f(\boldsymbol{\beta}_k|\text{else}) &\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}_k - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}_k - \boldsymbol{\mu}) - \sum_{\{i:K_i=k\}} \frac{1}{2\sigma_k^2}(y_i - \mathbf{x}_i'\boldsymbol{\beta}_k)^2\right\} \\
&\quad \times \exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{C}_{\boldsymbol{\theta}}^{-1} - \mathbf{I}_n)\mathbf{z}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\boldsymbol{\beta}_k'\left(\boldsymbol{\Sigma}^{-1} + \sigma_k^{-2}\sum_{\{i:K_i=k\}}\mathbf{x}_i\mathbf{x}_i'\right)\boldsymbol{\beta}_k + \boldsymbol{\beta}_k'\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \sigma_k^{-2}\sum_{\{i:K_i=k\}}\mathbf{x}_i y_i\right)\right\} \\
&\quad \times \exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{C}_{\boldsymbol{\theta}}^{-1} - \mathbf{I}_n)\mathbf{z}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta}_k - \boldsymbol{\mu}_k^*\right)'\left(\boldsymbol{\Sigma}_k^*\right)^{-1}\left(\boldsymbol{\beta}_k - \boldsymbol{\mu}_k^*\right)\right\} \exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{C}_{\boldsymbol{\theta}}^{-1} - \mathbf{I}_n)\mathbf{z}\right\} \\
&\propto \phi_p\left(\boldsymbol{\beta}_k|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right) \exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{C}_{\boldsymbol{\theta}}^{-1} - \mathbf{I}_n)\mathbf{z}\right\},
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{\Sigma}_k &= \left(\boldsymbol{\Sigma}^{-1} + \sigma_k^{-2}\sum_{\{i:K_i=k\}}\mathbf{x}_i\mathbf{x}_i'\right)^{-1} \\
\boldsymbol{\mu}_k &= \boldsymbol{\Sigma}_k^*\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \sigma_k^{-2}\sum_{\{i:K_i=k\}}\mathbf{x}_i y_i\right).
\end{aligned}$$

Propose $\boldsymbol{\beta}_k^*$ from $N_p\left(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)$ and accept it with probability

$$\alpha_1(\boldsymbol{\beta}_k, \boldsymbol{\beta}_k^*) = \min\left\{1, \frac{\exp\left\{-\frac{1}{2}\mathbf{z}^{*\prime}(\mathbf{C}_{\boldsymbol{\theta}}^{-1} - \mathbf{I}_n)\mathbf{z}^*\right\}}{\exp\left\{-\frac{1}{2}\mathbf{z}^{\prime}(\mathbf{C}_{\boldsymbol{\theta}}^{-1} - \mathbf{I}_n)\mathbf{z}\right\}}\right\}.$$

If $\boldsymbol{\beta}_k^*$ is rejected, we further propose $\boldsymbol{\beta}_k^{**}$ from $N_p\left(\boldsymbol{\beta}_k, \boldsymbol{\Sigma}_k\right)$ and accept it with probability $\alpha_2(\boldsymbol{\beta}_k, \boldsymbol{\beta}_k^*, \boldsymbol{\beta}_k^{**})$ as

$$\begin{aligned}
&\min\left\{1, \frac{f(\boldsymbol{\beta}_k^{**}|\text{else})[1-\alpha_1(\boldsymbol{\beta}_k^{**}, \boldsymbol{\beta}_k^*)]}{f(\boldsymbol{\beta}_k|\text{else})[1-\alpha_1(\boldsymbol{\beta}_k, \boldsymbol{\beta}_k^*)]}\right\} \\
=&\min\left\{1, \frac{\phi_p\left(\boldsymbol{\beta}_k^{**}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)\left[\exp\left\{-\frac{1}{2}\mathbf{z}^{**\prime}(\mathbf{C}_{\boldsymbol{\theta}}^{-1} - \mathbf{I}_n)\mathbf{z}^{**} + \frac{1}{2}\mathbf{z}^{*\prime}(\mathbf{C}_{\boldsymbol{\theta}}^{-1} - \mathbf{I}_n)\mathbf{z}^*\right\} - 1\right]}{\phi_p\left(\boldsymbol{\beta}_k|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)\left[\exp\left\{-\frac{1}{2}\mathbf{z}^{\prime}(\mathbf{C}_{\boldsymbol{\theta}}^{-1} - \mathbf{I}_n)\mathbf{z} + \frac{1}{2}\mathbf{z}^{*\prime}(\mathbf{C}_{\boldsymbol{\theta}}^{-1} - \mathbf{I}_n)\mathbf{z}^*\right\} - 1\right]}\right\}.
\end{aligned}$$

Step 4: Update σ_k^2 for $k = 1, \ldots, N$.

The full conditional distribution for σ_k^{-2} is

$$\begin{aligned}
f(\sigma_k^{-2}|\text{else}) &\propto \left(\sigma_k^{-2}\right)^{\nu_a - 1}\exp\left\{-\nu_b\sigma_k^{-2}\right\}\prod_{\{i:K_i=k\}}(\sigma_k^2)^{-1/2}\exp\left\{-\frac{1}{2\sigma_k^2}(y_i - \mathbf{x}_i'\boldsymbol{\beta}_k)^2\right\} \\
&\quad\times\exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{C}_{\boldsymbol{\theta}}^{-1} - \mathbf{I}_n)\mathbf{z}\right\} \\
&\propto \left(\sigma_k^{-2}\right)^{\nu_a + n_k/2 - 1}\exp\left\{-\left(\nu_b + \frac{1}{2}\sum_{\{i:K_i=k\}}(y_i - \mathbf{x}_i'\boldsymbol{\beta}_k)^2\right)\sigma_k^{-2}\right\} \\
&\quad\times\exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{C}_{\boldsymbol{\theta}}^{-1} - \mathbf{I}_n)\mathbf{z}\right\} \\
&\propto \text{Ga}\left(\sigma_k^{-2}|\nu_a + n_k/2, \nu_{b,k}^*\right)\exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{C}_{\boldsymbol{\theta}}^{-1} - \mathbf{I}_n)\mathbf{z}\right\},
\end{aligned}$$

where

$$n_k = \sum_{i=1}^n I(K_i = k) \quad\text{and}\quad \nu_{b,k}^* = \nu + \frac{1}{2}\sum_{\{i:K_i=k\}}(y_i - \mathbf{x}_i'\boldsymbol{\beta}_k)^2.$$

Propose $\sigma_k^{-2(*)}$ from $\text{Ga}\left(\nu_a + n_k/2, \nu_{b,k}^*\right)$ and accept it with probability

$$\min\left\{1, \frac{\exp\left\{-\frac{1}{2}\mathbf{z}^{*\prime}(\mathbf{C}_{\boldsymbol{\theta}}^{-1} - \mathbf{I}_n)\mathbf{z}^*\right\}}{\exp\left\{-\frac{1}{2}\mathbf{z}^{\prime}(\mathbf{C}_{\boldsymbol{\theta}}^{-1} - \mathbf{I}_n)\mathbf{z}\right\}}\right\}.$$

Step 5: Update V_k for $k = 1, \ldots, N-1$.

Let $n_k = \sum_{i=1}^n I(K_i = k)$. Note that $\mathbf{z}$ depends on the values of V_k's through the

weiths w_k's. Thus, the full conditional distribution V_i is

$$f(V_k|\text{else}) \propto \prod_{i=1}^{n}\left[V_{K_i}\prod_{k<K_i}(1-V_k)\right]\times\prod_{k=1}^{N-1}(1-V_k)^{\alpha-1}\exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{C}_{\boldsymbol{\theta}}^{-1}-\mathbf{I}_n)\mathbf{z}\right\}$$
$$\propto V_k^{n_k}\left(1-V_k\right)^{n_{k+1}+n_{k+2}+\ldots+n_N+\alpha-1}\exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{C}_{\boldsymbol{\theta}}^{-1}-\mathbf{I}_n)\mathbf{z}\right\}.$$

Propose V_k^* from Beta $\left(1+n_k, \alpha+\sum_{j=k+1}^{N} n_j\right)$ and accept it with probability

$$\min\left\{1, \frac{\exp\left\{-\frac{1}{2}\mathbf{z}^{*\prime}(\mathbf{C}_{\boldsymbol{\theta}}^{-1}-\mathbf{I}_n)\mathbf{z}^*\right\}}{\exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{C}_{\boldsymbol{\theta}}^{-1}-\mathbf{I}_n)\mathbf{z}\right\}}\right\}.$$

Step 6: Update α.

The full conditional distribution for α is

$$f(\alpha|\text{else}) \propto \left\{\prod_{k=1}^{N-1}\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)}(1-V_k)^{\alpha}\right\}\times\alpha^{a_0-1}\exp\{-b_0\alpha\}$$
$$\propto \alpha^{a_0+N-2}\exp\left\{-\alpha\left(b_0-\sum_{k=1}^{N-1}\log(1-V_k)\right)\right\}$$
$$\propto \text{Ga}\left(a_0+N-1, b_0-\sum_{k=1}^{N-1}\log(1-V_k)\right).$$

Step 7: Update $\boldsymbol{\mu}$.

The full conditional distribution for $\boldsymbol{\mu}$ is

$$f(\boldsymbol{\mu}|\text{else}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\mu}-\mathbf{m}_0)'\mathbf{S}_0^{-1}(\boldsymbol{\mu}-\mathbf{m}_0)-\frac{1}{2}\sum_{k=1}^{N}(\boldsymbol{\mu}-\boldsymbol{\beta}_k)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\beta}_k)\right\}$$
$$\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}-\mathbf{m}_0^*\right)'\left(\mathbf{S}_0^*\right)^{-1}\left(\boldsymbol{\mu}-\mathbf{m}_0^*\right)\right\}$$
$$\propto N_p\left(\mathbf{m}_0^*, \mathbf{S}_0^*\right),$$

where

$$\mathbf{S}_0^* = \left(\mathbf{S}_0^{-1}+N\boldsymbol{\Sigma}^{-1}\right)^{-1} \quad \text{and} \quad \mathbf{m}_0^* = \mathbf{S}_0^*\left(\mathbf{S}_0^{-1}\mathbf{m}_0+\boldsymbol{\Sigma}^{-1}\sum_{k=1}^{N}\boldsymbol{\beta}_k\right).$$

Step 8: Update $\boldsymbol{\Sigma}$.

The full conditional distribution for $\boldsymbol{\Sigma}^{-1}$ is

$$f(\boldsymbol{\Sigma}^{-1}|\text{else}) \propto |\boldsymbol{\Sigma}|^{-(\kappa_0+1-p-1)/2}\exp\left\{-\frac{1}{2}\sum_{k=1}^{N}(\boldsymbol{\mu}-\boldsymbol{\beta}_k)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\beta}_k)-\frac{1}{2}\text{tr}\left(\kappa_0\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}^{-1}\right)\right\}$$
$$\propto W_p\left(\left(\kappa_0\boldsymbol{\Sigma}_0+\sum_{k=1}^{N}(\boldsymbol{\mu}-\boldsymbol{\beta}_k)(\boldsymbol{\mu}-\boldsymbol{\beta}_k)'\right)^{-1}, \kappa_0+N\right).$$

Step 9: Update $\boldsymbol{\theta} = (\theta_1, \theta_2)'$ using adaptive Metropolis-Hastings algorithms (Haario et al., 2001).

Let $\boldsymbol{\vartheta} = c(\vartheta_1, \vartheta_2)'$ with $\vartheta_1 = \log\left(\frac{\theta_1}{1-\theta_1}\right)$ and $\vartheta_2 = \log(\theta_2)$. Then, the full conditional distribution for $\boldsymbol{\vartheta}$ is

$$\begin{aligned} f(\boldsymbol{\vartheta}|\text{else}) &\propto |\mathbf{C}_{\boldsymbol{\theta}}|^{-1/2}\exp\left\{-\frac{1}{2}\mathbf{z}'\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{z}\right\} \times \theta_1^{\theta_{1a}-1}(1-\theta_1)^{\theta_{1b}-1}\theta_1^2 e^{-\vartheta_1} \times \theta_2^{\theta_{2a}-1}e^{-\theta_{2b}\theta_2}e^{\vartheta_2} \\ &\propto \exp\Big\{ -\frac{1}{2}\log|\mathbf{C}_{\boldsymbol{\theta}}| - \frac{1}{2}\mathbf{z}'\mathbf{C}_{\boldsymbol{\theta}}^{-1}\mathbf{z} \\ &\qquad + (\theta_{1a}+1)\log(\theta_1) + (\theta_{1b}-1)\log(1-\theta_1) - \vartheta_1 + \theta_{2a}\log\theta_2 - \theta_{2b}\theta_2\Big\}. \end{aligned}$$

Suppose we are currently in time l and have sampled the states $\boldsymbol{\vartheta}_0, \boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_{l-1}$. We select an index $l_0 > 0$ for the length of an initial period and define

$$S_l = \begin{cases} S_0, & l \le l_0 \\ \frac{(2.4)^2}{2}(\mathcal{C}_l + 0.05 I_2) & l > l_0. \end{cases}$$

Here $\mathcal{C}_l$ is the sample variance of $\boldsymbol{\vartheta}_0, \boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_{l-1}$. Set $\bar{\boldsymbol{\vartheta}}_l = \frac{1}{l}\sum_{i=0}^{l-1}\boldsymbol{\vartheta}_i$. Following Haario et al. (2001), we can use the recursive equations

$$\bar{\boldsymbol{\vartheta}}_{l+1} = \frac{l}{l+1}\bar{\boldsymbol{\vartheta}}_l + \frac{1}{l+1}\boldsymbol{\vartheta}_l$$

and

$$\begin{aligned} \mathcal{C}_{l+1} = \text{cov}(\boldsymbol{\vartheta}_0, \boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_l) &= \frac{1}{l}\sum_{i=0}^{l}\boldsymbol{\vartheta}_i\boldsymbol{\vartheta}_i' - \frac{l+1}{l}\bar{\boldsymbol{\vartheta}}_{l+1}\bar{\boldsymbol{\vartheta}}_{l+1}' \\ &= \frac{l-1}{l}\mathcal{C}_l + \bar{\boldsymbol{\vartheta}}_l\bar{\boldsymbol{\vartheta}}_l' - \frac{l+1}{l}\bar{\boldsymbol{\vartheta}}_{l+1}\bar{\boldsymbol{\vartheta}}_{l+1}' + \boldsymbol{\vartheta}_l\boldsymbol{\vartheta}_l'/l. \end{aligned}$$

It follows that for $l > l_0$

$$S_{l+1} = \frac{l-1}{l}S_l + \frac{(2.4)^2}{2l}\left(l\bar{\boldsymbol{\vartheta}}_l\bar{\boldsymbol{\vartheta}}_l' - (l+1)\bar{\boldsymbol{\vartheta}}_{l+1}\bar{\boldsymbol{\vartheta}}_{l+1}' + \boldsymbol{\vartheta}_l\boldsymbol{\vartheta}_l' + 0.05 I_p\right).$$

We propose $\boldsymbol{\vartheta}^*$ from $N_2(\boldsymbol{\vartheta}_{l-1}, S_l)$ and accept it with probability

$$\min\left\{1, \frac{f(\boldsymbol{\vartheta}^*|\text{else})}{f(\boldsymbol{\vartheta}_{l-1}|\text{else})}\right\},$$

in which case we set $\boldsymbol{\vartheta}_l = \boldsymbol{\vartheta}^*$, and otherwise $\boldsymbol{\vartheta}_l = \boldsymbol{\vartheta}_{l-1}$.

B.2 The full scale approximation

A computational bottleneck of the MCMC sampling scheme is inverting the $n \times n$ matrix $\mathbf{C}_{\boldsymbol{\theta}}$, which typically has computational cost $O(n^3)$. In this section, we introduce a full scale approximation (FSA) approach proposed by Sang and Huang (2012), which provides a high quality approximation to the correlation function ρ at both the large and the small spatial scales, such that the inverse of $\mathbf{C}_{\boldsymbol{\theta}}$ can be substantially sped up for large value of n, e.g., $n \geq 500$.

Consider a fixed set of "knots" $\mathcal{S}^* = \{\mathbf{s}_1^*, \ldots, \mathbf{s}_m^*\}$ chosen from the study region. The FSA approach approximates the correlation function $\rho(\mathbf{s}, \mathbf{s}')$ with

$$\rho^\dagger(\mathbf{s}, \mathbf{s}') = \rho_l(\mathbf{s}, \mathbf{s}') + \rho_s(\mathbf{s}, \mathbf{s}'). \tag{B.2}$$

The $\rho_l(\mathbf{s}, \mathbf{s}')$ in (B.2) is the reduced-rank part capturing the long-scale spatial dependence, defined as $\rho_l(\mathbf{s}, \mathbf{s}') = \rho'(\mathbf{s}, \mathcal{S}^*)\rho_{mm}^{-1}(\mathcal{S}^*, \mathcal{S}^*)\rho(\mathbf{s}', \mathcal{S}^*)$, where $\rho(\mathbf{s}, \mathcal{S}^*) = [\rho(\mathbf{s}, \mathbf{s}_i^*)]_{i=1}^m$ is an $m \times 1$ vector, and $\rho_{mm}(\mathcal{S}^*, \mathcal{S}^*) = [\rho(\mathbf{s}_i^*, \mathbf{s}_j^*)]_{i,j=1}^m$ is an $m \times m$ correlation matrix at knots $\mathcal{S}^*$. However, $\rho_l(\mathbf{s}, \mathbf{s}')$ cannot well capture the short-scale dependence due to the fact that it discards entirely the residual part $\rho(\mathbf{s}, \mathbf{s}') - \rho_l(\mathbf{s}, \mathbf{s}')$. The idea of FSA is to add a small-scale part $\rho_s(\mathbf{s}, \mathbf{s}')$ as a sparse approximate of the residual part, defined by $\rho_s(\mathbf{s}, \mathbf{s}') = \{\rho(\mathbf{s}, \mathbf{s}') - \rho_l(\mathbf{s}, \mathbf{s}')\}\, \Delta(\mathbf{s}, \mathbf{s}')$, where $\Delta(\mathbf{s}, \mathbf{s}')$ is a modulating function, which is specified so that the $\rho_s(\mathbf{s}, \mathbf{s}')$ can well capture the local residual spatial dependence while still permits efficient computations. Motivated by Konomi et al. (2014), we first partition the total input space into B disjoint blocks, and then specify $\Delta(\mathbf{s}, \mathbf{s}')$ in a way such that the residuals are independent across input blocks, but the original residual dependence structure within each block is retained. Specifically, the function $\Delta(\mathbf{s}, \mathbf{s}')$ is taken to be 1 if $\mathbf{s}$ and $\mathbf{s}'$ belong to the same block and 0 otherwise.The approximated correlation function $\rho^\dagger(\mathbf{s}, \mathbf{s}')$ in (B.2) provides an exact recovery of the true correlation within each block, and the approximation errors are $\rho(\mathbf{s}, \mathbf{s}') - \rho_l(\mathbf{s}, \mathbf{s}')$ for locations $\mathbf{s}$ and $\mathbf{s}'$ in different blocks. Those errors are expected

to be small for most entries because most of these location pairs are farther apart.

Applying the above FSA approach to approximate the correlation function $\rho(\mathbf{s}, \mathbf{s}')$, we can approximate the correlation matrix $\boldsymbol{\rho}_{nn}$ with

$$\boldsymbol{\rho}^{\dagger}_{nn} = \boldsymbol{\rho}_l + \boldsymbol{\rho}_s = \boldsymbol{\rho}_{nm}\boldsymbol{\rho}^{-1}_{mm}\boldsymbol{\rho}'_{nm} + \left(\boldsymbol{\rho}_{nn} - \boldsymbol{\rho}_{nm}\boldsymbol{\rho}^{-1}_{mm}\boldsymbol{\rho}'_{nm}\right) \circ \boldsymbol{\Delta}, \tag{B.3}$$

where $\boldsymbol{\rho}_{nm} = [\rho(\mathbf{s}_i, \mathbf{s}^*_j)]_{i=1:n, j=1:m}$, $\boldsymbol{\rho}_{mm} = [\rho(\mathbf{s}^*_i, \mathbf{s}^*_j)]^m_{i,j=1}$, and $\boldsymbol{\Delta} = [\Delta(\mathbf{s}_i, \mathbf{s}_j)]^n_{i,j=1}$. Here, the notation "$\circ$" represents the element-wise matrix multiplication. It follows from equation (B.3) that the covariance matrix $\mathbf{C}_{\boldsymbol{\theta}}$ can be approximated by

$$\mathbf{C}^{\dagger}_{\boldsymbol{\theta}} = \theta_1\boldsymbol{\rho}^{\dagger}_{nn} + (1-\theta_1)\mathbf{I}_n = \theta_1\boldsymbol{\rho}_{nm}\boldsymbol{\rho}^{-1}_{mm}\boldsymbol{\rho}'_{nm} + \mathbf{C}_s,$$

where $\mathbf{C}_s = \theta_1\boldsymbol{\rho}_s + (1-\theta_1)\mathbf{I}_n$. Applying the Sherman-Woodbury-Morrison formula for inverse matrices, we can approximate $\mathbf{C}^{-1}_{\boldsymbol{\theta}}$ by $\left(\mathbf{C}^{\dagger}_{\boldsymbol{\theta}}\right)^{-1}$ which is given by

$$\mathbf{C}^{-1}_s - \theta_1\mathbf{C}^{-1}_s\boldsymbol{\rho}_{nm}\left(\boldsymbol{\rho}_{mm} + \theta_1\boldsymbol{\rho}'_{nm}\mathbf{C}^{-1}_s\boldsymbol{\rho}_{nm}\right)^{-1}\boldsymbol{\rho}'_{nm}\mathbf{C}^{-1}_s. \tag{B.4}$$

In addition, the determinant of $\mathbf{C}_{\boldsymbol{\theta}}$ can be approximated by $\det\left(\mathbf{C}^{\dagger}_{\boldsymbol{\theta}}\right)$ given as

$$\det\left\{\boldsymbol{\rho}_{mm} + \theta_1\boldsymbol{\rho}'_{nm}\mathbf{C}^{-1}_s\boldsymbol{\rho}_{nm}\right\}\det(\boldsymbol{\rho}_{mm})^{-1}\det(\mathbf{C}_s). \tag{B.5}$$

Since the $n \times n$ matrix $\mathbf{C}_s$ is a block matrix, the right-hand sides of equations (B.4) and (B.5) involve only inverses and determinants of $m \times m$ low-rank matrices and $n \times n$ block diagonal matrices. Thus the computational complexity can be greatly reduced relative to the expensive computational cost of using original correlation function for large value of n.

B.3 Derivation of the CPO statistic

Let $f_{\mathbf{x}_i}$ and $F_{\mathbf{x}_i}$ be the density and distribution functions of T_i given $\mathbf{x}_i$, respectively. According to the hierarchical representation in Section 2.2, given all the model

parameters $\Theta = (\mathbf{B}, \boldsymbol{\sigma}^2, \mathbf{V}, \boldsymbol{\theta})$, we have

$$f_{\mathbf{x}_i}(t_i|\Theta) = \frac{1}{t_i}\sum_{k=1}^{N} w_k \frac{1}{\sigma_k}\phi\left(\frac{\log t_i - \mathbf{x}_i'\boldsymbol{\beta}_k}{\sigma_k}\right),$$
$$F_{\mathbf{x}_i}(t_i|\Theta) = \sum_{k=1}^{N} w_k \frac{1}{\sigma_k}\Phi\left(\frac{\log t_i - \mathbf{x}_i'\boldsymbol{\beta}_k}{\sigma_k}\right),$$

According to the definition of CPO, we have $\text{CPO}_i = f\left(t_i^o|\mathcal{D}_{-i}\right)^{\delta_i} S\left(t_i^o|\mathcal{D}_{-i}\right)^{1-\delta_i}$. Note that, for $\delta_i = 1$,

$$\begin{aligned} E_{(\mathbf{t},\Theta|\mathcal{D})}\left\{\frac{1}{f(t_i^o|\mathbf{t}_{-i},\Theta)}\right\} &= \frac{\int \frac{1}{f(t_i^o|\mathbf{t}_{-i},\Theta)} f(\mathbf{t}|\Theta)\pi(\Theta)d\Theta \prod_{\{j:\delta_j=0\}} I(t_j > t_j^o)dt_j}{\int f(\mathbf{t}|\Theta)\pi(\Theta)d\Theta \prod_{\{j:\delta_j=0\}} I(t_j > t_j^o)dt_j} \\ &= \frac{\int f(\mathbf{t}_{-i}|\Theta)\pi(\Theta)d\Theta \prod_{\{j:\delta_j=0\}} I(t_j > t_j^o)dt_j}{\int f(t_i^o,\mathbf{t}_{-i}|\Theta)\pi(\Theta)d\Theta \prod_{\{j:\delta_j=0\}} I(t_j > t_j^o)dt_j} \\ &= \frac{f(\mathcal{D}_{-i})}{f(t_i^o,\mathcal{D}_{-i})} = \frac{1}{f(t_i^o|\mathcal{D}_{-i})} \end{aligned} \tag{B.6}$$

and for $\delta_i = 0$,

$$\begin{aligned} E_{(\mathbf{t},\Theta|\mathcal{D})}\left\{\frac{1}{S(t_i^o|\mathbf{t}_{-i},\Theta)}\right\} &= \frac{\int \frac{1}{S(t_i^o|\mathbf{t}_{-i},\Theta)} f(\mathbf{t}|\Theta)\pi(\Theta)d\Theta \prod_{\{j:\delta_j=0\}} I(t_j > t_j^o)dt_j}{\int f(\mathbf{t}|\Theta)\pi(\Theta)d\Theta \prod_{\{j:\delta_j=0\}} I(t_j > t_j^o)dt_j} \\ &= \frac{\int \frac{f(t_i|\mathbf{t}_{-i},\Theta)}{S(t_i^o|\mathbf{t}_{-i},\Theta)} f(\mathbf{t}_{-i}|\Theta)\pi(\Theta)d\Theta \prod_{\{j:\delta_j=0\}} I(t_j > t_j^o)dt_j}{\int f(t_i^o|\mathbf{t}_{-i},\Theta) f(\mathbf{t}_{-i}|\Theta)\pi(\Theta)d\Theta \prod_{\{j:\delta_j=0\}} I(t_j > t_j^o)dt_j} \\ &= \frac{\int f(\mathbf{t}_{-i}|\Theta)\pi(\Theta)d\Theta \prod_{\{j\neq i:\delta_j=0\}} I(t_j > t_j^o)dt_j}{\int S(t_i^o|\mathbf{t}_{-i},\Theta) f(\mathbf{t}_{-i}|\Theta)\pi(\Theta)d\Theta \prod_{\{j\neq i:\delta_j=0\}} I(t_j > t_j^o)dt_j} \\ &= \frac{1}{S(t_i^o|\mathcal{D}_{-i})} \end{aligned} \tag{B.7}$$

Thus it follows that

$$\text{CPO}_i = \left(E\left[\frac{1}{f(t_i^o|\mathbf{t}_{-i},\Theta)^{\delta_i} S(t_i^o|\mathbf{t}_{-i},\Theta)^{1-\delta_i}}\right]\right)^{-1}, \tag{B.8}$$

where the expectation E is taken with respect to the joint posterior of $\{\mathbf{t}, \Theta|\mathcal{D}\}$. Here $f(t_i^o|\mathbf{t}_{-i},\Theta)$ and $S(t_i^o|\mathbf{t}_{-i},\Theta)$ are given by

$$\begin{aligned} f(t_i^o|\mathbf{t}_{-i},\Theta) &= \frac{1}{\sigma_{-i}}\phi\left(\frac{\Phi^{-1}\left\{F_{\mathbf{x}_i}(t_i^o|\Theta)\right\} - \mu_{-i}}{\sigma_{-i}}\right)\frac{f_{\mathbf{x}_i}(t_i^o|\Theta)}{\phi\left(\Phi^{-1}\left\{F_{\mathbf{x}_i}(t_i^o|\Theta)\right\}\right)}, \\ S(t_i^o|\mathbf{t}_{-i},\Theta) &= 1 - \Phi\left\{\frac{\Phi^{-1}\left\{F_{\mathbf{x}_i}(t_i^o|\Theta)\right\} - \mu_{-i}}{\sigma_{-i}}\right\}, \end{aligned} \tag{B.9}$$

where $\mu_{-i} = -\sum_{j\neq i} C_{ij}^{-}\Phi^{-1}\left\{F_{\mathbf{x}_i}(t_i|\Theta)\right\}/C_{ii}^{-}$ and $\sigma_{-i}^2 = 1/C_{ii}^{-}$ with C_{ij}^{-} being the ijth element of $\mathbf{C}^{-1}$.

Model specification

Assume that $T_i|\mathbf{x}_i$ marginally follows the Cox proportional hazard model with cumulative distribution function (cdf)

$$F_{\mathbf{x}_i}(t) = 1 - \exp\left\{-\Lambda_0(t)e^{\mathbf{x}_i'\boldsymbol{\beta}}\right\}$$

and probability density function (pdf)

$$f_{\mathbf{x}_i}(t) = \exp\left\{-\Lambda_0(t)e^{\mathbf{x}_i'\boldsymbol{\beta}}\right\}\lambda_0(t)e^{\mathbf{x}_i'\boldsymbol{\beta}},$$

where $\boldsymbol{\beta}$ is a $p\times 1$ vector of regression coefficients, $\lambda_0(t)$ is the baseline hazard function and $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$ is the cumulative baseline hazard function. The piecewise exponential model provides a flexible framework to deal with the baseline hazard (e.g. Walker and Mallick, 1997; Aslanidou et al., 1998; Qiou et al., 1999). We partition the time period $\mathbb{R}^+$ into M intervals, say $I_k = (d_{k-1}, d_k], k = 1, \ldots, M$, where $d_0 = 0$ and $d_M = \infty$. The baseline hazard is then assumed to be constant within each interval, i.e.

$$\lambda_0(t) = \sum_{k=1}^{M} h_k I\{t \in I_k\},$$

where $h_k \overset{iid}{\sim} \text{Ga}(\nu_0 h, \nu_0)$ are unknown hazard values and $I\{A\}$ is the usual indicator function, i.e. 1 when A is true, 0 otherwise. Consequently, the cumulative baseline hazard function can be written as

$$\Lambda_0(t) = \sum_{k=1}^{M(t)} h_k \Delta_k(t),$$

where $M(t) = \min\{k : d_k \geq t\}$ and $\Delta_k(t) = \min\{d_k, t\} - d_{k-1}$. In fact, the above piecewise exponential model centers $\lambda_0(t)$ at the exponential hazard family $\lambda_h(t) \equiv h$ indexed by h. However the resulting predictive density is not differentiable at the jump points among the M intervals, which is not desirable for many applications and

especially for prediction purposes. We propose a mixture of piecewise exponential model by taking the index h to be random; the resulting mixture model yields a differentiable, i.e. smooth, density. Specifically, we set $d_k = F_h^{-1}(k/M), k = 0, \ldots, M$, where $F_h(\cdot)$ is the cdf of exponential distribution with rate parameter h, and put a prior on h, say $h \sim N(h_0, V_0)$. Regardless, after incorporating spatial dependence as described in Section 2 of the main article, we consider the following hierarchical model for the data together with the augmented latent true event-times:

$$\begin{aligned}
\delta_i | t_i &= I(t_i = t_i^o),\ i = 1, \ldots, n \\
t_i | \boldsymbol{\beta}, \mathbf{h} &\sim F_{\mathbf{x}_i}(t) = 1 - \exp\left\{ - \sum_{k=1}^{M(t)} h_k \Delta_k(t) e^{\mathbf{x}_i' \boldsymbol{\beta}} \right\},\ i = 1, \ldots, n \\
\mathbf{z} = (z_1, \ldots, z_n)' | \mathbf{t}, \boldsymbol{\beta}, \boldsymbol{\theta} &\sim N_n(\mathbf{0}, \mathbf{C}),\ z_i = \Phi^{-1}\left\{ F_{\mathbf{x}_i}(t_i) \right\}, i = 1, \ldots, n \\
h_k &\overset{iid}{\sim} \mathrm{Ga}(r_0 h, r_0), k = 1, \ldots, M, \quad h \sim N(h_0, v_0^2) \\
\boldsymbol{\beta} | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0 &\overset{iid}{\sim} N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \\
(\theta_1, \theta_2) &\sim \mathrm{Beta}(\theta_{1a}, \theta_{1b}) \times \mathrm{Ga}(\theta_{2a}, \theta_{2b})
\end{aligned} \tag{B.10}$$

where $\mathbf{t} = (t_1, \ldots, t_n)'$, $\mathbf{h} = (h_1, \ldots, h_M)$. We consider following default hyper-prior values: $M = 10$, $r_0 = 1$, $h_0 = \hat{h}$, $\boldsymbol{\mu}_0 = \mathbf{0}$, $\boldsymbol{\Sigma}_0 = 10^5 \mathbf{I}_p$, $\theta_{1a} = \theta_{1b} = \theta_{2a} = \theta_{2b} = 1$, where $\hat{h}$ is the maximum likelihood estimate of the rate parameter h from fitting an exponential PH model. An R function `spCopulaCoxph` calling compiled `C++` to fit this model is provided in the R package `spBayesSurv` accompanying this paper. We also provide a function `indeptCoxph` to fit the non-spatial standard PH model with above baseline specification.

Remark: The function `spCopulaCoxph` provides an option to determine whether the centering parameter h is random or not. For random h, the `spCopulaCoxph` fails to work for certain data sets especially when spatial correlation is large; while the function `indeptCoxph` works without any problem. However, when h is fixed, both functions work very well.

MCMC

The full likelihood function for the data is given by

$$\begin{aligned}\mathcal{L}(\mathbf{t},\mathbf{h},\boldsymbol{\beta},\boldsymbol{\theta}) \propto & \prod_{i=1}^{n} \exp\left\{-\Lambda_0(t_i)e^{\mathbf{x}_i'\boldsymbol{\beta}}\right\} \lambda_0(t_i)e^{\mathbf{x}_i'\boldsymbol{\beta}} \left\{\delta_i I(t_i = t_i^o) + (1-\delta_i)I(t_i > t_i^o)\right\} \\ & \times |\mathbf{C}|^{-1/2} \exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{C}^{-1} - \mathbf{I}_n)\mathbf{z}\right\} \\ & \times \prod_{k=1}^{M} \frac{r_0^{r_0 h}}{\Gamma(r_0 h)} h_k^{r_0 h - 1} \exp\{-r_0 h_k\} \times \exp\left\{-\frac{1}{2v_0^2}(h - h_0)^2\right\} \\ & \times \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_0)\right\} \\ & \times \theta_1^{\theta_{1a}-1}(1-\theta_1)^{\theta_{1b}-1} \times \theta_2^{\theta_{2a}-1} e^{-\theta_{2b}\theta_2}\end{aligned} \tag{B.11}$$

The MCMC sampling steps are as follow:

Step 1: Update h using adaptive Metropolis-Hastings.

The full conditional distribution for h is

$$\begin{aligned} f(h|\text{else}) &\propto \prod_{k=1}^{M} \frac{r_0^{r_0 h}}{\Gamma(r_0 h)} h_k^{r_0 h - 1} I(h > 0) \exp\left\{-\frac{1}{2v_0^2}(h - h_0)^2\right\} \\ &\propto \exp\left\{M r_0 h \log r_0 - M \log\Gamma(r_0 h) + r_0 h \sum_{k=1}^{M} \log(h_k) - \frac{1}{2v_0^2}(h - h_0)^2\right\} I(h > 0). \end{aligned}$$

Step 2: Update t_i for $i = 1, \ldots, n$.

The full conditional distribution for t_i is

$$f(t_i|\text{else}) \propto \delta_i I(t_i = t_i^o) + (1-\delta_i) f_{\mathbf{x}_i}(t_i) \exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{C}^{-1} - \mathbf{I}_n)\mathbf{z}\right\} I(t_i > t_i^o).$$

If $\delta_i = 1$, update $t_i = t_i^o$. If $\delta_i = 0$, propose $t_i^* = F_{\mathbf{x}_i}^{-1}(u_i)$ with u_i from Unif $(F_{\mathbf{x}_i}(t_i^o), 1)$, and then accept it with probability

$$\min\left\{1, \frac{\exp\left\{-\frac{1}{2}\mathbf{z}^{*\prime}(\mathbf{C}^{-1} - \mathbf{I}_n)\mathbf{z}^*\right\}}{\exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{C}^{-1} - \mathbf{I}_n)\mathbf{z}\right\}}\right\},$$

where $\mathbf{z}^* = (z_i^*, \ldots, z_n^*)'$ is the new transformed vector corresponding to t_i^*.

Step 3: Update h_k for $k = 1, \ldots, M$.

The full conditional distribution for h_k is

$$\begin{aligned} f(h_k|\text{else}) &\propto \exp\left\{-\sum_{i=1}^{n}\sum_{k=1}^{M(t_i)} h_k \Delta_k(t_i) e^{\mathbf{x}_i'\boldsymbol{\beta}}\right\}\left\{\prod_{\{i:M(t_i)=k\}} h_{M(t_i)}\right\} h_k^{r_0 h-1}\exp\{-r_0 h_k\} \\ &\quad\times \exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{C}^{-1}-\mathbf{I}_n)\mathbf{z}\right\} \\ &\propto \exp\left\{-h_k \sum_{\{i:M(t_i)\geq k\}} \Delta_k(t_i) e^{\mathbf{x}_i'\boldsymbol{\beta}}\right\}\left\{h_k^{\sum_{i=1}^{n} I\{M(t_i)=k\}}\right\} h_k^{r_0 h-1}\exp\{-r_0 h_k\} \\ &\quad\times \exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{C}^{-1}-\mathbf{I}_n)\mathbf{z}\right\} \\ &\propto h_k^{r_0 h+m_k-1}\exp\left\{-\left(r_0+l_k\right)h_k\right\}\exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{C}^{-1}-\mathbf{I}_n)\mathbf{z}\right\}, \end{aligned}$$

where

$$m_k = \sum_{i=1}^{n} I\{M(t_i)=k\} \quad \text{and} \quad l_k = \sum_{\{i:M(t_i)\geq k\}} \Delta_k(t_i) e^{\mathbf{x}_i'\boldsymbol{\beta}}.$$

Propose h_k^* from $\text{Ga}\left(r_0 h + m_k, r_0 + l_k\right)$ and accept it with probability

$$\min\left\{1, \frac{\exp\left\{-\frac{1}{2}\mathbf{z}^{*\prime}(\mathbf{C}^{-1}-\mathbf{I}_n)\mathbf{z}^*\right\}}{\exp\left\{-\frac{1}{2}\mathbf{z}'(\mathbf{C}^{-1}-\mathbf{I}_n)\mathbf{z}\right\}}\right\}.$$

Step 4: Update $\boldsymbol{\beta}$ using adaptive Metropolis-Hastings.

$$f(\boldsymbol{\beta}|\text{else}) \propto \exp\left\{\sum_{i=1}^{n}\left(-\Lambda_0(t_i)e^{\mathbf{x}_i'\boldsymbol{\beta}} + \mathbf{x}_i'\boldsymbol{\beta}\right) - \frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\mu}_0)'\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_0) - \frac{1}{2}\mathbf{z}'(\mathbf{C}^{-1}-\mathbf{I}_n)\mathbf{z}\right\}$$

Let $\hat{\boldsymbol{\beta}}$ be the maximum likelihood estimate of $\boldsymbol{\beta}$ from fitting an exponential Cox model and $\hat{S}_0$ be its estimated covariance matrix. Suppose we are currently in time l and have sampled the states $\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{l-1}$. We select an index $l_0 > 0$ for the length of an initial period and define

$$S_l = \begin{cases} \hat{S}_0, & l \leq l_0 \\ \frac{(2.4)^2}{p}(\mathcal{C}_l + 0.05 I_p) & l > l_0. \end{cases}$$

where $\mathcal{C}_l$ is the sample covariance matrix of $\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{l-1}$. Set $\bar{\boldsymbol{\beta}}_l = \frac{1}{l}\sum_{i=0}^{l-1}\boldsymbol{\beta}_i$. Following (Haario et al., 2001), we can use the recursive equations

$$\bar{\boldsymbol{\beta}}_{l+1} = \frac{l}{l+1}\bar{\boldsymbol{\beta}}_l + \frac{1}{l+1}\boldsymbol{\beta}_l$$

and

$$\mathcal{C}_{l+1} = \text{cov}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_l) = \frac{1}{l}\sum_{i=0}^{l} \boldsymbol{\beta}_i\boldsymbol{\beta}_i' - \frac{l+1}{l}\bar{\boldsymbol{\beta}}_{l+1}\bar{\boldsymbol{\beta}}_{l+1}'$$
$$= \frac{l-1}{l}\mathcal{C}_l + \bar{\boldsymbol{\beta}}_l\bar{\boldsymbol{\beta}}_l' - \frac{l+1}{l}\bar{\boldsymbol{\beta}}_{l+1}\bar{\boldsymbol{\beta}}_{l+1}' + \boldsymbol{\beta}_l\boldsymbol{\beta}_l'/l.$$

It follows that for $l > l_0$

$$S_{l+1} = \frac{l-1}{l}S_l + \frac{(2.4)^2}{2l}\left(l\bar{\boldsymbol{\beta}}_l\bar{\boldsymbol{\beta}}_l' - (l+1)\bar{\boldsymbol{\beta}}_{l+1}\bar{\boldsymbol{\beta}}_{l+1}' + \boldsymbol{\beta}_l\boldsymbol{\beta}_l' + 0.05I_p\right).$$

We propose $\boldsymbol{\beta}^*$ from $N_p(\boldsymbol{\beta}_{l-1}, S_l)$ and accept it with probability

$$\min\left\{1, \frac{f(\boldsymbol{\beta}^*|\text{else})}{f(\boldsymbol{\beta}_{l-1}|\text{else})}\right\}.$$

Step 5: Update $\boldsymbol{\theta} = (\theta_1, \theta_2)'$ in the same way as Step 9 in Appendix B.1.

B.5 Additional simulations

Supplements of simulation – Scenario II

We test the performance of LDDPM-spatial model when the PH assumption is satisfied and compare it with the PH-spatial model. Similarly to Scenario I, we randomly select 400 locations over the spatial region $[0, 40] \times [0, 100]$ and hold out 100 of them for assessing the prediction performance. We then simulate the event times $T(\mathbf{s})$ at these 400 locations from a PH model $F_x(t) = 1 - \exp\{te^{-x}\}$ with the same sample spatial dependence and distribution on x as described in Scenario I. The noninformative censoring times are simulated from a uniform distribution on $(1, 3)$ so that the censoring rate is about $15\% \sim 35\%$.

Table B.1 presents the posterior inferences for spatial correlation parameters $\boldsymbol{\theta} = (\theta_1, \theta_2)$, where the bias of corresponding point estimates (i.e. posterior means), the Monte Carlo mean of posterior standard deviation estimates (MEAN-SD), the Monte Carlo standard deviation of point estimates (SD-MEAN), and the Monte Carlo

Table B.1 Simulated data – Scenario II. True value, bias of the point estimator (posterior mean), mean (across Monte Carlo replicates) of the posterior standard deviations (MEAN-SD), standard deviation (across Monte Carlo replicates) of the point estimator (SD-MEAN), and Monte Carlo coverage probability for the 95% credible interval (CP) for the spatial correlation parameter $\boldsymbol{\theta}$. The averaged computing time is also presented.

Model	Parameters	True	BIAS	MEAN-SD	SD-MEAN	CP
LDDPM-spatial	θ_1	0.98	-0.020	0.026	0.019	0.98
	θ_2	0.10	0.018	0.027	0.028	0.93
PH-spatial	θ_1	0.98	-0.015	0.023	0.016	0.99
	θ_2	0.10	0.009	0.019	0.023	0.90

coverage probability of 95% credible intervals (CP) are reported. The results suggest that the point estimates of $\boldsymbol{\theta}$ are almost unbiased under both the LDDPM-spatial and PH-spatial models. The MEAN-SD and SD-MEAN values under the PH-spatial model are fairly close indicating that the posterior standard deviation is an appropriate estimator of the frequentist standard error. The CPs are around the nominal 95% level. Overall, even when data are generated from the PH-spatial model, our model still performs reasonably well.

Figure B.1 shows boxplots of the ISEs for estimated survival curves, LPMLs, and MSPEs under the considered models. The PH-spatial model provides slightly smaller biases of the fitted survival functions on average, compared with our model. As for prediction ability and accuracy, the results show that two models provide almost the same boxplots of LPMLs and MSPEs, indicating that the LDDP-spatial model is quite competitive even when the PH assumption is satisfied.

B.6 Additional results to the analysis of frog data

Table B.2 shows posterior estimates of the spatial dependence parameters θ_1 and θ_2 under both the LDDPM-spatial and PH-spatial models. Figure B.2 presents the Kaplan-Meier survival curves for bdwater=0 versus bdwater=1. The results show that standard parametric or semi-parametric spatial models may be inadequate due

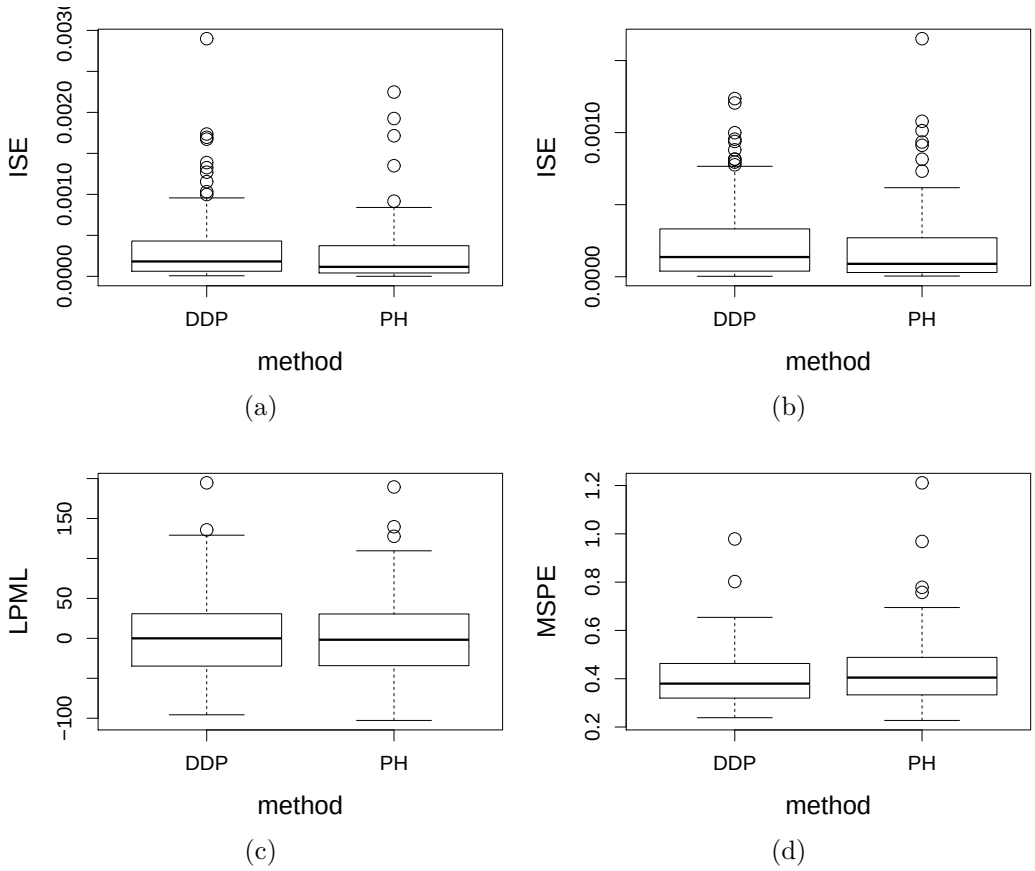


Figure B.1 Simulated data – Scenario II. Panel (a) and (b): boxplots of ISEs for fitted survival curves when $x = -1$ and $x = 1$, respectively. Panel (c): boxplots of LPMLs. Panel (d): boxplots of MSPE. In each panel, the two models from left to right are LDDPM-spatial and PH-spatial, respectively.

to the presence of crossing survivals. Figure B.3 presents the trace plots for the correlation parameters $\boldsymbol{\theta}$, which mix reasonably well.

Table B.2 Frog data. Posterior statistics for θ_1 and θ_2 under the LDDPM-spatial and PH-spatial models assuming the exponential correlation function. The computing time is also presented.

Model	Parameters	Mean	Median	Std. dev.	95% CI
LDDPM-spatial	θ_1	0.991	0.992	0.004	(0.982, 0.998)
(3.2 hours)	θ_2	0.133	0.130	0.040	(0.060, 0.216)
PH-spatial	θ_1	0.995	0.995	0.002	(0.991, 0.999)
(2.8 hours)	θ_2	0.081	0.080	0.013	(0.059, 0.109)

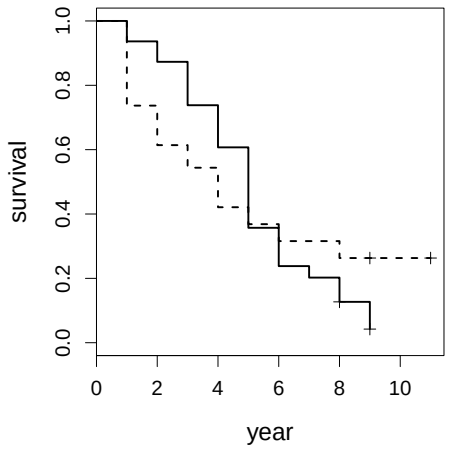
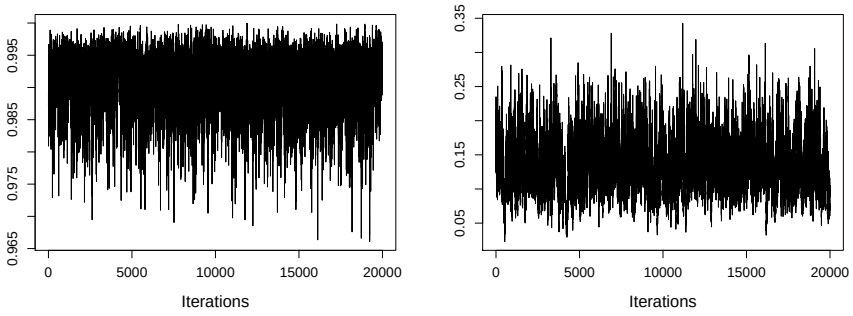


Figure B.2 Frog data. Kaplan-Meier survival curves for bdwater=0 (solid) versus bdwater=1 (dashed).



(a) (b)

Figure B.3 Frog data. Trace plots for θ_1 (panel a) and θ_2 (panel b).

B.7 Sample R code for simulated data

The R package `spBayesSurv` is available at the website

`http://cran.r-project.org/web/packages/spBayesSurv`.

```
################################################################
# Sample R code for implementing the marginal LDDPM spatial
# survival model proposed by Zhou, Hanson, and Knapp (2015)
# based on a simulated data: mixture of two normals (see
# Section 4 in the paper).
# Provided by Haiming Zhou on 4/2/2015
################################################################
```

```
##-------------Load libraries-------------------##
rm(list=ls())
library(MASS)
library(Rcpp)
library(RcppArmadillo)
library(coda)
library(survival)
library(spBayesSurv)

##-------------Set the true models--------------##
## True parameters
betaT = cbind(c(3.5, 0.5), c(2.5, -1));
wT = c(0.4, 0.6);
sig2T = c(1^2, 0.5^2);
theta1 = 0.98; theta2 = 0.1;

## The pdf of Ti:
fi = function(y, xi, w=wT){
nw = length(w);
ny = length(y);
res = matrix(0, ny, nw);
Xi = c(1,xi);
for (k in 1:nw){
res[,k] = w[k]*dnorm(y, sum(Xi*betaT[,k]), sqrt(sig2T[k]) )
}
apply(res, 1, sum)
}
## The CDF of Ti:
Fi = function(y, xi, w=wT){
nw = length(w);
ny = length(y);
res = matrix(0, ny, nw);
```

```
Xi = c(1,xi);
for (k in 1:nw){
res[,k] = w[k]*pnorm(y, sum(Xi*betaT[,k]), sqrt(sig2T[k]) )
}
apply(res, 1, sum)
}
## The inverse for CDF of Ti
Finvsingle = function(u, xi) {
res = uniroot(function (x) Fi(x, xi)-u, lower=-1e50,
upper=1e50, tol=.Machine$double.eps^0.5);
res$root
}
Finv = function(u, xi) {sapply(u, Finvsingle, xi)};

##-------------Generate data-------------------##
## generate coordinates:
## npred is the # of locations for prediction
n = 300; npred = 30; ntot = n + npred;
ldist = 100; wdist = 40;
s1 = runif(ntot, 0, wdist); s2 = runif(ntot, 0, ldist);
s = rbind(s1,s2);
#plot(s[1,], s[2,]);
## divide them into blocks
nldist=5; nwdist=2;
nb=nldist*nwdist; nb; # number of blocks;
coor = matrix(0, nb, 4); ## four edges for each block;
tempindex=1; lstep=ldist/nldist; wstep=wdist/nwdist;
for(i in 1:nwdist){
for(j in 1:nldist){
coor[tempindex,] = c((i-1)*wstep, i*wstep, (j-1)*lstep, j*lstep );
tempindex = tempindex + 1;
}
}
```

```
## Assign block id for each location
blockid = rep(NA,ntot);
for(i in 1:nb){
blockid[((s1>coor[i,1])*(s1<=coor[i,2])*(s2>coor[i,3])*(s2<=coor[i,4]))==1]=i;
}
## Choose knots S*
nldist=5; nwdist=2;
m=nldist*nwdist; m; # number of knots;
ss = matrix(0, m, 2);
tempindex=1; lstep=ldist/nldist; wstep=wdist/nwdist;
for(i in 1:nwdist){
for(j in 1:nldist){
ss[tempindex,] = c( (i-1)*wstep+wstep/2, (j-1)*lstep+lstep/2);
tempindex = tempindex + 1;
}
}
## Covariance matrix
dnn = .Call("DistMat", s, s, PACKAGE = "spBayesSurv");
corT = theta1*exp(-theta2*dnn)+(1-theta1)*diag(ntot);

## Generate x
x = runif(ntot,-1.5,1.5);
X = cbind(rep(1,ntot), x);
p = ncol(X); # number of covariates + 1
## Generate transformed log of survival times
z = mvrnorm(1, rep(0, ntot), corT);
## Generate log of survival times y
u = pnorm(z);
y = rep(0, ntot);
for (i in 1:ntot){
y[i] = Finv(u[i], x[i]);
}
#plot(x,y);
```

```
yTrue = y;

## Censoring scheme
Centime = runif(ntot, 3,4); #Centime = 10000;
delta = (y<=Centime) +0 ;
sum(delta)/ntot; #non-censoring rate
cen = which(delta==0);
y[cen] = Centime[cen];

## make a data frame
dtotal = data.frame(s1=s1, s2=s2, y=y, x=x, delta=delta,
yTrue=yTrue, id=blockid, t=exp(y));
## Hold out npred=30 for prediction purpose
predindex = sample(1:ntot, npred);
dpred = dtotal[predindex,];
dtrain = dtotal[-predindex,];

# rename the variables
d = dtrain; n=nrow(d); n;
d = d[order(d$id),];
s = cbind(d$s1, d$s2);

# FSA settings
knots = list(ss=ss, blockid=d$id);

# Prediction settings
xpred = dpred$x;
s0 = cbind( dpred$s1, dpred$s2 );
prediction = list(spred=s0, xpred=xpred, predid=dpred$id);

#################################################################
# fit the model using default priors
#################################################################
```

```
# MCMC parameters
nburn <- 2000
nsave <- 2000
nskip <- 0
ndisplay <- 500
mcmc <- list(nburn=nburn,
nsave=nsave,
nskip=nskip,
ndisplay=ndisplay)

# Prior information
prior = list(N = 10,
a0 = 2, b0 = 2);

# current state values
state <- NULL;

# Fit the model
ptm <- proc.time()
res = spCopulaDDP( y=d$y, delta=d$delta, x=d$x, s = s, prediction=prediction,
prior=prior, mcmc=mcmc,state=state,FSA=TRUE,knots=knots);
systime = sum((proc.time() - ptm)[1:2]);
# trace plots
par(mfrow = c(3,2))
w.save2 = res$w;
Kindex = which.max(rowMeans(w.save2));
traceplot(mcmc(w.save2[Kindex,]), main="w")
sig2.save2 = res$sigma2;
traceplot(mcmc(sig2.save2[Kindex,]), main="sig2")
beta.save2 = res$beta;
alpha.save2 = res$alpha;
traceplot(mcmc(beta.save2[2,Kindex,]), main="beta")
traceplot(mcmc(alpha.save2), main="alpha")
```

```
theta1.save2 = res$theta1;
theta2.save2 = res$theta2
traceplot(mcmc(theta1.save2), main="theta1")
traceplot(mcmc(theta2.save2), main="theta2")

## LPML
LPML2 = sum(log(res$cpo)); LPML2;
## MSPE
mean((dpred$yTrue-apply(res$Ypred, 1, median))^2);

## Proportions for number of clusters
gg=gg=apply(res$w, 2, function(x) length(which(x>0.001)) );
table(gg)/length(gg);

## plots
par(mfrow = c(2,2));
xnew = c(-1, 1);
xpred = cbind(xnew);
nxpred = nrow(xpred);
ygrid = seq(0,6.0,0.05); tgrid = exp(ygrid);
ngrid = length(ygrid);
estimates = GetCurves(res, xpred, ygrid, CI=c(0.05, 0.95));
fhat = estimates$fhat;
Shat = estimates$Shat;
## density in y
plot(ygrid, fi(ygrid, xnew[1]), "l", lwd=2, ylim=c(0, 0.8),
xlim=c(0,6), main="density in y")
for(i in 1:nxpred){
lines(ygrid, fi(ygrid, xnew[i]), lwd=2)
lines(ygrid, fhat[,i], lty=2, lwd=2, col=4);
}
## survival in y
plot(ygrid, 1-Fi(ygrid, xnew[1]), "l", lwd=2, ylim=c(0, 1),
```

```
xlim=c(0,6), main="survival in y")
for(i in 1:nxpred){
lines(ygrid, 1-Fi(ygrid, xnew[i]), lwd=2)
lines(ygrid, Shat[,i], lty=2, lwd=2, col=4);
lines(ygrid, estimates$Shatup[,i], lty=2, lwd=1, col=4);
lines(ygrid, estimates$Shatlow[,i], lty=2, lwd=1, col=4);
}
## density in t
plot(tgrid, fi(ygrid, xnew[1])/tgrid, "l", lwd=2, ylim=c(0, 0.15),
xlim=c(0,100), main="density in t")
for(i in 1:nxpred){
lines(tgrid, fi(ygrid, xnew[i])/tgrid, lwd=2)
lines(tgrid, fhat[,i]/tgrid, lty=2, lwd=2, col=4);
}
## survival in t
plot(tgrid, 1-Fi(ygrid, xnew[1]), "l", lwd=2, ylim=c(0, 1),
xlim=c(0,100), main="survival in t")
for(i in 1:nxpred){
lines(tgrid, 1-Fi(ygrid, xnew[i]), lwd=2)
lines(tgrid, Shat[,i], lty=2, lwd=2, col=4);
lines(tgrid, estimates$Shatup[,i], lty=2, lwd=1, col=4);
lines(tgrid, estimates$Shatlow[,i], lty=2, lwd=1, col=4);
}
```

B.8 Measures of dependence

In this section, we explore dependence relations between the original event times in the framework of copula models. Kendall's tau and Spearman's rho are the two most widely used scale-invariant measures for the overall dependence of a pair of subjects over the entire lifespan by integrating over time, both of which are based on a form of dependence known as *concordance* (Nelsen, 2006). In the context of survival data, we say a pair of random event times are concordant if large (small)

values of one tend to be associated with large (small) values of the other, otherwise they are discordant. Specifically, when $\mathbf{Y} = (Y_1, \ldots, Y_n) = (\log T_1, \ldots, \log T_n)$ follows the marginal LDDPM spatial survival model, the Kendall's tau and Spearman's rho of the original event times T_i and T_j (also the same as those of Y_i and Y_j based on their definitions) can be expressed as

$$\tau_{i,j}^K = r\left(\int_0^1 \int_0^1 C_{i,j}(u_i, u_j; \boldsymbol{\theta}) dC_{i,j}(u_i, u_j; \boldsymbol{\theta})\right) - 1 = 4E[C_{i,j}(U_i, U_j; \boldsymbol{\theta})] - 1 \quad \text{(B.12)}$$

and

$$\rho_{i,j}^S = 12 \int_0^1 \int_0^1 u_i u_j dC_{i,j}(u_i, u_j; \boldsymbol{\theta}) - 3 = 12E[U_i U_j] - 3, \quad \text{(B.13)}$$

where $C_{i,j}(u_i, u_j; \boldsymbol{\theta})$ is a bivariate marginal of the n-dimensional Gaussian copula function and $(U_i, U_j) \sim C_{i,j}(u_i, u_j; \boldsymbol{\theta})$. Thus the Kendall's tau and Spearman's rho are uniquely determined by the copula function which further depends on the spatial correlation parameters $\boldsymbol{\theta}$. The range of these two measures are between -1 and 1, where the higher the value is, the more concordant the two event times are. Although we don't have closed forms for both τ_{ij}^K and $\rho_{i,j}^S$, we can easily evaluate the expectations in equations (B.12) and (B.13) via posterior simulation (Smith, 2013). Given a set of posterior samples $\{\boldsymbol{\theta}^{(l)}, l = 1, \ldots, L\}$, we generate an iterate $(U_i^{(l)}, U_j^{(l)})$ from the bivariate marginal $C_{i,j}(u_i, u_j; \boldsymbol{\theta}^{(l)})$, and then estimate the Kendall's tau and Spearman's rho by

$$\hat{\tau}_{i,j}^K = \frac{4}{L} \sum_{l=1}^{L} C_{i,j}(U_i^{(l)}, U_j^{(l)}; \boldsymbol{\theta}^{(l)}) - 1, \quad \text{(B.14)}$$

and

$$\hat{\rho}_{i,j}^S = \frac{12}{L} \sum_{l=1}^{L} U_i^{(l)} U_j^{(l)} - 1. \quad \text{(B.15)}$$

For a given copula, the Kendall's tau and Spearman's rho between a pair of random variables are not necessarily the same. In fact they are often quite different for many families of copulas; see Nelsen (2006) for further illustrations.

Appendix C

Supplement to Chapter 4

C.1 MCMC Sampling

The joint posterior distribution for all parameters is given by

$$\begin{aligned}\mathcal{L}(\mathbf{y},\boldsymbol{\gamma},\tilde{\boldsymbol{\beta}},\mathbf{v},\sigma,\tau,\alpha) \propto & \prod_{i=1}^{m}\prod_{j=1}^{m_i} g_{\mathbf{x}_{ij}}(y_{ij}-\tilde{\mathbf{z}}_{ij}'\tilde{\boldsymbol{\beta}}-v_i)I(\log l_{ij}<y_{ij}<\log u_{ij}) \\ & \times \prod_{l=1}^{L-1}\prod_{k=1}^{2^l}\left(\frac{\alpha(l+1)^2}{2n}\right)^{(q+1)/2}|\mathbf{X}'\mathbf{X}|^{1/2}\exp\left\{-\frac{\alpha(l+1)^2}{4n}\boldsymbol{\gamma}_{l,k}'(\mathbf{X}'\mathbf{X})\boldsymbol{\gamma}_{l,k}\right\} \\ & \times \exp\left\{-\frac{1}{2}(\tilde{\boldsymbol{\beta}}-\mathbf{m}_0)'\mathbf{S}_0^{-1}(\tilde{\boldsymbol{\beta}}-\mathbf{m}_0)\right\} \\ & \times \left(\tau^{-2}\right)^{\frac{m-1}{2}}\exp\left\{-\frac{1}{2\tau^2}\mathbf{v}'(D_w-W)\mathbf{v}\right\} \\ & \times \left(\sigma^2\right)^{-a_\sigma-1}\exp\left\{-b_\sigma/\sigma^2\right\} \\ & \times \left(\tau^{-2}\right)^{a_\tau-1}\exp\left\{-b_\tau\tau^{-2}\right\} \\ & \times (\alpha)^{a_0-1}\exp\left\{-b_0\alpha\right\}\end{aligned} \tag{C.1}$$

Step 1: Update latent log survival time y_{ij} when $\delta_{ij}=0$ for $j=1,\ldots,m_i, i=1,\ldots,m$.

The full conditional distribution for y_{ij} is

$$p(y_{ij}|\text{else}) \propto g_{\tilde{\mathbf{z}}_{ij}}(y_{ij}-\tilde{\mathbf{z}}_{ij}'\tilde{\boldsymbol{\beta}}-v_i)I(\log l_{ij}<y_{ij}<\log u_{ij}).$$

The single-variable slice sampling method (Neal, 2003) is used.

Step 2: Update $\boldsymbol{\gamma}_{l,k}$ for $k=1,\ldots,2^l, l=1,\ldots,L-1$.

The full conditional distribution for $\boldsymbol{\gamma}_{l,k}$ is

$$p(\boldsymbol{\gamma}_{l,k}|\text{else}) \propto \exp\left\{-\frac{\alpha(l+1)^2}{4n}\boldsymbol{\gamma}'_{l,k}(\mathbf{X}'\mathbf{X})\boldsymbol{\gamma}_{l,k}\right\}\prod_{i=1}^{m}\prod_{j=1}^{m_i} g_{\mathbf{x}_{ij}}(\epsilon_{ij})$$
$$\propto \exp\left\{-\frac{\alpha(l+1)^2}{4n}\boldsymbol{\gamma}'_{l,k}(\mathbf{X}'\mathbf{X})\boldsymbol{\gamma}_{l,k}\right\}\prod_{i,j\in\mathcal{S}_k} Y_{l+1,\lceil 2^{l+1}\Phi_\sigma(\epsilon_{ij})\rceil}(\mathbf{x}_{ij}),$$

where $\mathcal{S}_k = \left\{i,j : \lceil 2^{l+1}\Phi_\sigma(\epsilon_{ij})\rceil \in \{2k, 2k-1\}\right\}$. Metropolis-Hastings steps with Gaussian proposals based on iterative weighted least squares (Gamerman, 1997) are used.

Step 3: Update $\tilde{\boldsymbol{\beta}}$.

The full conditional distribution for $\tilde{\boldsymbol{\beta}}$ is

$$p(\tilde{\boldsymbol{\beta}}|\text{else}) \propto \exp\left\{-\frac{1}{2}(\tilde{\boldsymbol{\beta}}-\mathbf{m}_0)'\mathbf{S}_0^{-1}(\tilde{\boldsymbol{\beta}}-\mathbf{m}_0)\right\}\prod_{i=1}^{m}\prod_{j=1}^{m_i} g_{\mathbf{x}_{ij}}(y_{ij}-\tilde{\mathbf{z}}'_{ij}\tilde{\boldsymbol{\beta}}-v_i).$$

Single-variable slice sampling is used for updating each component of $\tilde{\boldsymbol{\beta}}$ separately.

Step 4: Update v_i for $i = 1,\ldots,m-1$ under the constraint $\sum_{i=1}^m v_i = 0$.

Under the constraint $\sum_{i=1}^m v_i = 0$, letting $v_m = -\sum_{i=1}^{m-1} v_i$, one can easily show that

$$\begin{aligned}\mathbf{v}'(D_w - W)\mathbf{v} &= \sum_{i=1}^{m} w_{i+}v_i^2 - \sum_{i=1}^{m}\sum_{j=1}^{m} v_i w_{ij} v_j \\ &= \sum_{i=1}^{m-1} w_{i+}v_i^2 - \sum_{i=1}^{m-1}\sum_{j=1}^{m-1} v_i w_{ij} v_j + w_{m+}\left(\sum_{i=1}^{m-1} v_i\right)^2 \\ &\quad + 2\left(\sum_{i=1}^{m-1} v_i\right)\sum_{j=1}^{m-1} w_{mj}v_j\end{aligned}$$

It follows that the prior for v_i, $i = 2,\ldots,m$, is

$$\begin{aligned}p(v_i|\{v_j\}_{j\neq i}) &\propto \exp\left\{-\frac{1}{2\tau^2}\mathbf{v}'(D_w - W)\mathbf{v}\right\} \\ &\propto \exp\left\{-\frac{1}{2\tau^2}\Big(v_i^2[w_{i+} + w_{m+} + 2w_{mi}]\right. \\ &\qquad \left. - 2v_i\sum_{j\neq i}^{m-1} v_j\left[w_{ij} - w_{m+} - w_{mj} - w_{mi}\right]\Big)\right\} \\ &\propto \exp\left\{-\frac{w^*_{i+}}{2\tau^2}(v_i - \mu_i^*)^2\right\}\end{aligned}$$

where $w_{i+}^* = [w_{i+} + w_{m+} + 2w_{mi}]$ and $\mu_i^* = \sum_{j\neq i}^{m-1} v_j \left[w_{ij} - w_{m+} - w_{mj} - w_{mi}\right] / w_{i+}^*$.

The full conditional distribution for v_i is

$$p(v_i|\text{else}) \propto \exp\left\{-\frac{w_{i+}^*}{2\tau^2}\left(v_i - \mu_i^*\right)^2\right\} \prod_{j=1}^{m_i} g_{\mathbf{x}_{ij}}(y_{ij} - \tilde{\mathbf{z}}_{ij}'\tilde{\boldsymbol{\beta}} - v_i).$$

Single-variable slice sampling is used.

Step 5: Update σ^2.

The full conditional distribution for σ^2 is

$$p(\sigma^2|\text{else}) \propto \left(\sigma^2\right)^{-a_\sigma - 1} \exp\left\{-b_\sigma/\sigma^2\right\} \prod_{i=1}^{m}\prod_{j=1}^{m_i} g_{\mathbf{x}_{ij}}(y_{ij} - \tilde{\mathbf{z}}_{ij}'\tilde{\boldsymbol{\beta}} - v_i).$$

Single-variable slice sampling is used.

Step 6: Update τ^2.

The full conditional distribution for τ^{-2} is

$$p(\tau^{-2}|\text{else}) \propto \left(\tau^{-2}\right)^{a_\tau + \frac{m-1}{2} - 1} \exp\left\{-\left[b_\tau + \frac{1}{2}\mathbf{v}'(D_w - W)\mathbf{v}\right]\tau^{-2}\right\}.$$

Thus τ^2 is sampled from $\Gamma(a_\tau^*, b_\tau^*)$, where $a_\tau^* = a_\tau + \frac{m-1}{2} - 1$ and $b_\tau^* = b_\tau + \frac{1}{2}\mathbf{v}'(D_w - W)\mathbf{v}$.

Step 7: Update α.

The full conditional distribution for α is

$$\begin{aligned} p(\alpha|\text{else}) &\propto (\alpha)^{a_0 - 1} \exp\left\{-b_0\alpha\right\} \prod_{l=1}^{L-1}\prod_{k=1}^{2^l} (\alpha)^{(q+1)/2} \exp\left\{-\frac{\alpha(l+1)^2}{4n}\boldsymbol{\gamma}_{l,k}'(\mathbf{X}'\mathbf{X})\boldsymbol{\gamma}_{l,k}\right\} \\ &\propto \alpha^{a_0^* - 1} \exp\left\{-b_0^*\alpha\right\}, \end{aligned}$$

where $a_0^* = a_0 + (q+1)(2^{L-1} - 1)$ and

$$b_0^* = b_0 + \sum_{l=1}^{L-1}\sum_{k=1}^{2^l} \frac{(l+1)^2}{4n}\boldsymbol{\gamma}_{l,k}'(\mathbf{X}'\mathbf{X})\boldsymbol{\gamma}_{l,k}.$$

Thus α is sampled from $\Gamma(a_0^*, b_0^*)$.

C.2 Implementation Using R

We present how to use the `R` package `spBayesSurv` to fit the proposed model based on simulated data. We take Scenario I as an example, but we instead randomly generate the adjacency matrix W. The following code is used to generate data:

```
## True densities
Finvsingle = function(u, F) {
res = uniroot(function (x) F(x)-u, lower=-1000, upper=1000,
tol=.Machine$double.eps^0.5);
res$root
}
Finv = function(u, F) {sapply(u, Finvsingle, F)};
f0 = function(x) dnorm(x, 0, 0.8);
F0 = function(x) pnorm(x, 0, 0.8);
shift=1
f1 = function(x) 0.5*dnorm(x, -shift, 0.5) + 0.5*dnorm(x, shift, 0.5)
F1 = function(x) 0.5*pnorm(x, -shift, 0.5) + 0.5*pnorm(x, shift, 0.5);
ff = function(x, xtf=0) {
if(xtf==0) {res=f0(x)} else res=f1(x)
res
}
FF = function(x, xtf=0){
if(xtf==0) {res=F0(x)} else res=F1(x)
res
}
# Simulation settings;
betaT = c(-1, 1, -0.5);
tau2T = 0.1;
m = 64; # blocks
```

```
mi = 10;
mis = rep(mi, m);
id = rep(1:m,mis);
n = length(id); # Total number of subjects
# Generate symmetric adjaceny matrix, W
wi = rep(0, m)
while(any(wi==0)){
W = matrix(0,m,m)
W[upper.tri(W,diag=FALSE)]<-rbinom(m*(m-1)/2,1,.1)
W = W+t(W)
wi = apply(W,1,sum)  # No. neighbors
}
# Spatial effects, v
Wstar = matrix(0, m-1, m-1);
Dstar = diag(wi[-m]);
for(i in 1:(m-1)){
for(j in 1:(m-1)){
Wstar[i,j] = W[j,i]-W[j,m]-W[m,i]-wi[m]
}
}
Qstar = Dstar-Wstar;
covT = tau2T*solve(Qstar);
v0 = c(rmvnorm(1,sigma=covT));
v = c(v0,-sum(v0));
vn = rep(v, mis);
# responses
x1 = rnorm(n, 0, 1);
x2 = rbinom(n, 1, 0.5);
xtf = x2; ptf = 2;
```

```
X = cbind(1,x1,x2); pce = ncol(X);
u = runif(n, 0, 1)
y = rep(0, n);
for(i in 1:n) {
if(x2[i]==1) {
y[i] = sum(betaT*X[i,]) + vn[i] + Finv(u[i], F1)
}else{
y[i] = sum(betaT*X[i,]) + vn[i] + Finv(u[i], F0)
}
}
# generate responses
Cen = runif(n, 0.5, 1)
delta = (exp(y)<=Cen)+0;
sum(delta)/n
tTrue = exp(y);
tobs = cbind(tTrue, tTrue);
tobs[which(delta==0),] = cbind(Cen[which(delta==0)], NA);
dtotal = data.frame(tleft=tobs[,1], tright=tobs[,2], x1=x1,
x2=x2, xtf=x2, ID=id, tTrue=tTrue, censor=delta);
## sort the data by ID
d = dtotal[order(dtotal$ID),];
```

Note that the data have to be sorted by cluster ID before model fitting. The prior settings in the simulation are: $L = 4$, $a_0 = 5$, $b_0 = 1$, $\mathbf{m}_0 = \mathbf{0}_3$, $\mathbf{S}_0 = 10^5\mathbf{I}_3$, $a_\sigma = b_\sigma = 2.001$, $a_\tau = b_\tau = 0.1$. The following code illustrates how these priors are specified:

```
prior = list(maxL = 4, a0 = 5, b0 = 1, m0 = rep(0,3), S0 = diag(rep(1e5,3)),
siga0 = 2.001, sigb0 = 2.001, taua0 = 0.1, taub0 = 0.1);
```

For each MCMC algorithm, the chain was subsampled every 20 iterates to get a total of $10,000$ scans after a burn-in period of $50,000$ iterations. The following code is for MCMC specification:

```
mcmc=list(nburn=50000, nsave=10000, nskip=19, ndisplay=1000);
```

Finally the following code is used to fit the model:

```
res=frailtyGAFT(Surv(tleft, tright, type="interval2")~x1+x2+baseline(x1,x2),
data=d, mcmc=mcmc, prior=prior, frailty="CAR", ID=d$ID,
Proximity=W);
```

The above function can also fit a GAFT with independent Gaussian frailties by setting `frailty="normal"`, a GAFT without frailties by setting `frailty=NULL`, a semiparametric AFT by removing `baseline()` and a lognormal AFT by specifying `a0` at a negative value and adding an additional argument `state=list(alpha=1e30)`. The output is given below:

```
Generalized accelerated failure time frailty model:
Call:
frailtyGAFT.default(formula = Surv(tleft, tright, type = "interval2") ~
x1 + x2 + baseline(x1, x2), data = d, mcmc = mcmc, prior = prior,
frailty = "CAR", ID = d$ID, Proximity = W)

Posterior inference of regression coefficients
Mean      Median    Std. Dev.  95%HPD-Low  95%HPD-Upp
intercept  -1.03250  -1.03313   0.05876   -1.14984    -0.91519
x1          0.97998   0.98166   0.05274    0.87558     1.07728
x2         -0.54970  -0.54752   0.11725   -0.78024    -0.30805

Posterior inference of scale parameter
```

```
Mean     Median   Std. Dev.  95%HPD-Low  95%HPD-Upp
scale  0.84356  0.84095  0.04627    0.75335     0.93333

Posterior inference of precision parameter of LDTFP
Mean   Median  Std. Dev.  95%HPD-Low  95%HPD-Upp
[1,]  2.775  2.621   1.053      1.049       4.943

Posterior inference of conditional CAR frailty variance
Mean     Median   Std. Dev.  95%HPD-Low  95%HPD-Upp
variance  0.14578  0.12899  0.08171    0.02296     0.30270

Bayes factors for LDTFP covariate effects:
intercept         x1         x2    overall  normality
2.325e-01  3.930e-01  2.467e+08  5.764e+07  1.119e+09

Log pseudo marginal likelihood: LPML=161.7173
Deviance information criterion: DIC
Number of subjects:=640
```

The following code is used to show the trace plots in Figure C.1:

```
par(cex=1.2,mar=c(4.1,4.1,1,1),cex.lab=1.3,cex.axis=1.1)
par(mfrow = c(3,2))
traceplot(mcmc(res$beta[1,]), main="intercept", xlab="");
traceplot(mcmc(res$beta[2,]), main="x1", xlab="");
traceplot(mcmc(res$beta[3,]), main="x2", xlab="");
traceplot(mcmc(res$tau2), main="tau^2", xlab="");
traceplot(mcmc(res$alpha), main="alpha", xlab="");
traceplot(mcmc(res$sigma2), main="sigma^2", xlab="");
```

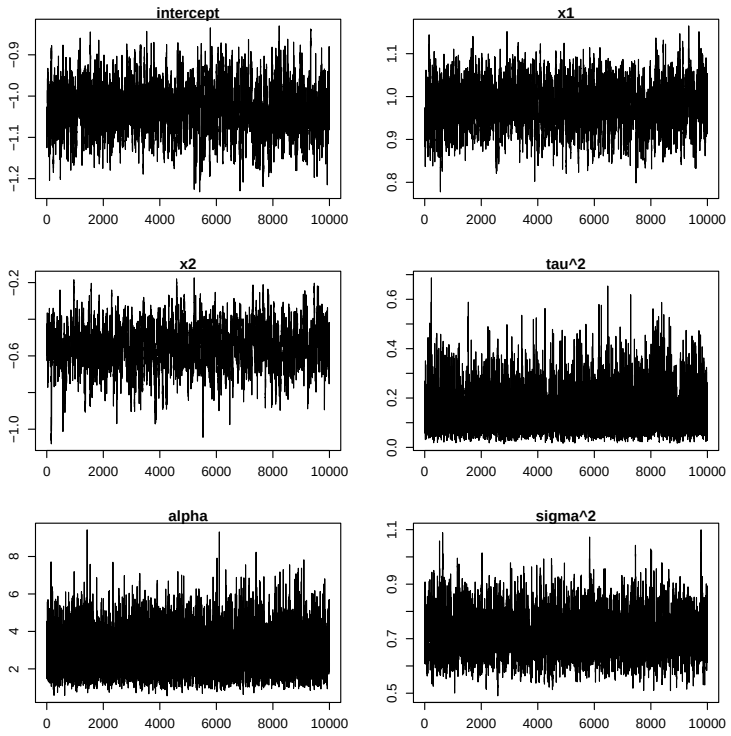


Figure C.1 Trace plot for a simulated data set.

The following code is used to plot autocorrelation functions (ACF) along with effective samples sizes:

```
par(cex=1.2,mar=c(4.1,4.1,3,1),cex.lab=1.3,cex.axis=1.1)
par(mfrow = c(3,2))
plot(acf(mcmc(res$beta[1,]), plot=FALSE),
main=paste("intercept:", "effective size=",
round(effectiveSize(mcmc(res$beta[1,])))), xlab="");
plot(acf(mcmc(res$beta[2,]), plot=FALSE),
main=paste("x1:", "effective size=",
round(effectiveSize(mcmc(res$beta[2,])))), xlab="");
plot(acf(mcmc(res$beta[3,]), plot=FALSE),
main=paste("x2:", "effective size=",
round(effectiveSize(mcmc(res$beta[3,])))), xlab="");
```

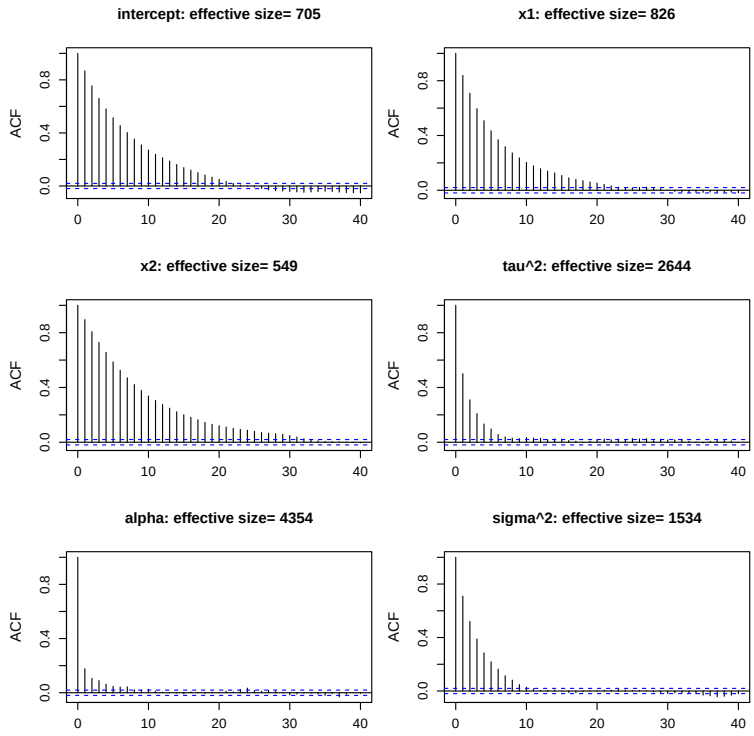


Figure C.2 ACF plots along with effective sample sizes for a simulated data set.

```
plot(acf(mcmc(res$tau2), plot=FALSE),
main=paste("tau^2:", "effective size=",
round(effectiveSize(mcmc(res$tau2)))), xlab="");
plot(acf(mcmc(res$alpha), plot=FALSE),
main=paste("alpha:", "effective size=",
round(effectiveSize(mcmc(res$alpha)))), xlab="");
plot(acf(mcmc(res$sigma2), plot=FALSE),
main=paste("sigma^2:", "effective size=",
round(effectiveSize(mcmc(res$sigma2)))), xlab="");
```

The following code is used to plot survival and density curves in Figure C.3, where `xpred` are fixed at zeros if baseline curves are needed:

```
ygrid = seq(-3,3,0.03); tgrid = exp(ygrid);
ngrid = length(ygrid);
```

```
xnew = c(0, 1)
xpred = cbind(c(1,1.5), xnew);
xtfpred = cbind(c(1,1.5), xnew);
nxpred = nrow(xpred);
estimates = plot(res, xpred, ygrid, xtfpred=xtfpred, CI=0.9, PLOT=F);
## survival function when xtf=(1,0)
pdf(file ="supp_sample_surv1.pdf", paper="special", width=7, height=7);
i=1
par(cex=1.5,mar=c(4.1,4.1,1,1),cex.lab=1.4,cex.axis=1.1)
plot(ygrid, 1-FF(ygrid-sum(c(1,xpred[i,])*betaT),xnew[i]), "l", lwd=3,
xlim=c(-3,3), ylim=c(0,1), main=NULL, xlab="log time", ylab="survival")
polygon(x=c(rev(ygrid),ygrid),
y=c(rev(estimates$Shatlow[,i]),estimates$Shatup[,i]),
border=NA,col="lightgray");
lines(ygrid, 1-FF(ygrid-sum(c(1,xpred[i,])*betaT),xnew[i]), "l", lwd=3,
xlim=c(-3,3), ylim=c(0,1))
lines(ygrid, estimates$Shat[,i], lty=3, lwd=3);
dev.off();
## density function when xtf=(1,0)
pdf(file ="supp_sample_dens1.pdf", paper="special", width=7, height=7);
i=1
par(cex=1.5,mar=c(4.1,4.1,1,1),cex.lab=1.4,cex.axis=1.1)
plot(ygrid, ff(ygrid-sum(c(1,xpred[i,])*betaT),xnew[i]), "l", lwd=3,
xlim=c(-3,3), ylim=c(0,1), main=NULL, xlab="log time", ylab="density")
polygon(x=c(rev(ygrid),ygrid),
y=c(rev(estimates$fhatlow[,i]),estimates$fhatup[,i]),
border=NA,col="lightgray");
lines(ygrid, ff(ygrid-sum(c(1,xpred[i,])*betaT),xnew[i]), "l", lwd=3,
xlim=c(-3,3), ylim=c(0,1))
```

```
lines(ygrid, estimates$fhat[,i], lty=3, lwd=3);
dev.off();
## survival function when xtf=(1.5,1)
pdf(file ="supp_sample_surv2.pdf", paper="special", width=7, height=7);
i=2
par(cex=1.5,mar=c(4.1,4.1,1,1),cex.lab=1.4,cex.axis=1.1)
plot(ygrid, 1-FF(ygrid-sum(c(1,xpred[i,])*betaT),xnew[i]), "l", lwd=3,
xlim=c(-3,3), ylim=c(0,1), main=NULL, xlab="log time", ylab="survival")
polygon(x=c(rev(ygrid),ygrid),
y=c(rev(estimates$Shatlow[,i]),estimates$Shatup[,i]),
border=NA,col="lightgray");
lines(ygrid, 1-FF(ygrid-sum(c(1,xpred[i,])*betaT),xnew[i]), "l", lwd=3,
xlim=c(-3,3), ylim=c(0,1))
lines(ygrid, estimates$Shat[,i], lty=3, lwd=3);
dev.off();
## density function when xtf=(1.5,1)
pdf(file ="supp_sample_dens2.pdf", paper="special", width=7, height=7);
i=2
par(cex=1.5,mar=c(4.1,4.1,1,1),cex.lab=1.4,cex.axis=1.1)
plot(ygrid, ff(ygrid-sum(c(1,xpred[i,])*betaT),xnew[i]), "l", lwd=3,
xlim=c(-3,3), ylim=c(0,1), main=NULL, xlab="log time", ylab="density")
polygon(x=c(rev(ygrid),ygrid),
y=c(rev(estimates$fhatlow[,i]),estimates$fhatup[,i]),
border=NA,col="lightgray");
lines(ygrid, ff(ygrid-sum(c(1,xpred[i,])*betaT),xnew[i]), "l", lwd=3,
xlim=c(-3,3), ylim=c(0,1))
lines(ygrid, estimates$fhat[,i], lty=3, lwd=3);
dev.off();
```

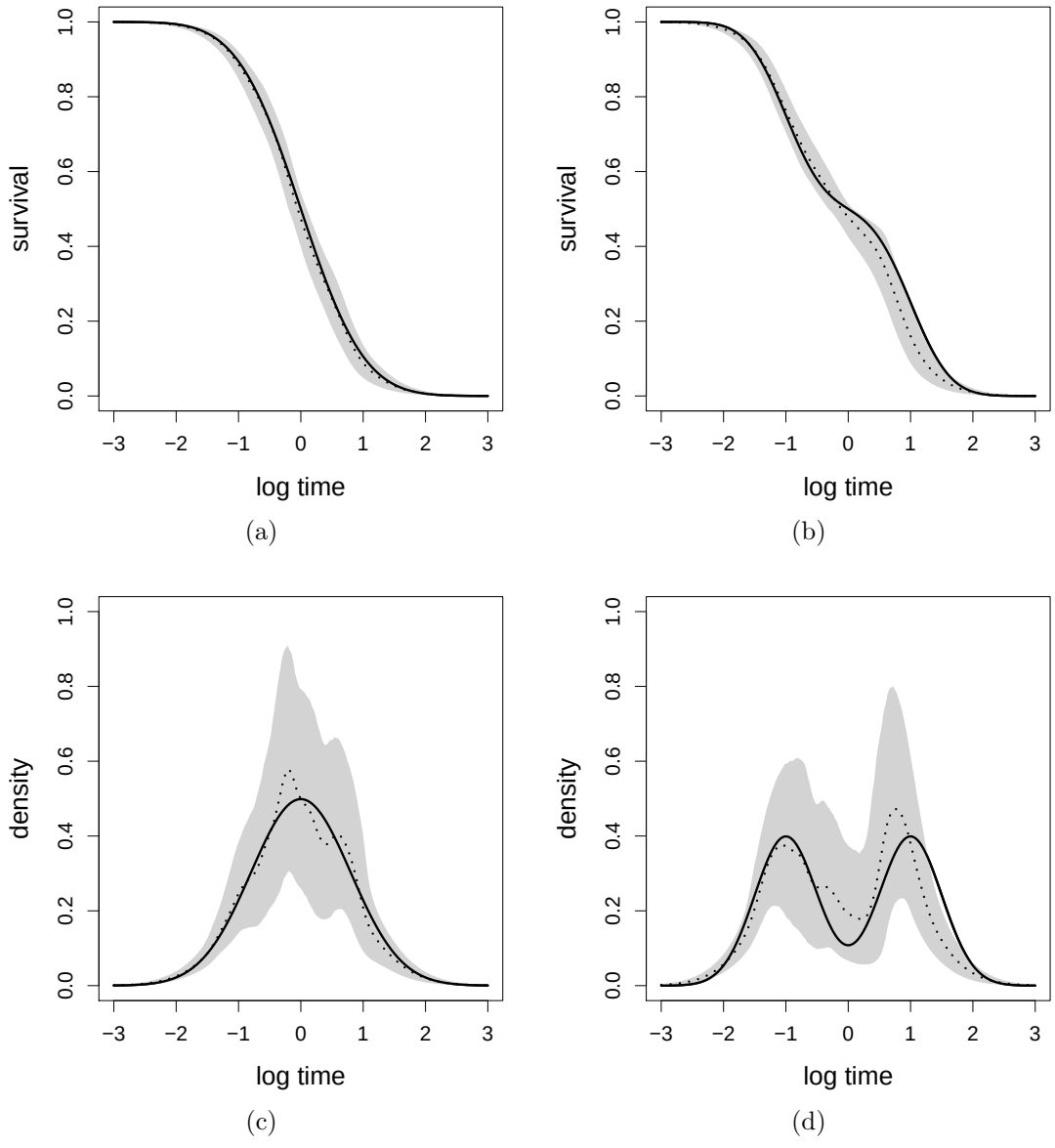


Figure C.3 Estimated curves with 90% point-wise credible intervals (gray areas) for $(x_1, x_2) = (1, 0)$ (panel a and c) and $(x_1, x_2) = (1.5, 1)$ (panel b and d) under the proposed model. The true curves are represented by continuous lines. The results under the proposed model are represented by dashed lines.

Table C.1 Simulation data. True value, averaged bias (BIAS) and posterior standard deviation (PSD) of each point estimate (i.e. posterior mean), standard deviation (across Monte Carlo simulations) of the point estimates (SD-Est) and coverage probability (CP) for the 95% credible intervals.

Scenario	Parameter	True	BIAS	PSD	SD-Est	CP
I	β_0	-1.0	0.007	0.068	0.061	0.970
	β_1	1.0	0.003	0.036	0.037	0.938
	β_2	-0.5	0.010	0.133	0.096	0.986
	τ^2	0.1	0.034	0.055	0.045	0.974
II	β_0	-1.0	0.031	0.110	0.085	0.956
	β_1	1.0	-0.002	0.033	0.034	0.944
	β_2	-0.5	0.000	0.059	0.060	0.942
	τ^2	0.1	0.025	0.050	0.038	0.988
III	β_0	-1.0	0.005	0.057	0.056	0.960
	β_1	1.0	0.004	0.038	0.040	0.932
	β_2	-0.5	-0.004	0.068	0.072	0.932
	τ^2	0.1	0.034	0.059	0.046	0.982

C.3 Additional simulation studies

Additional results

The use of Bayes factors allow us to tell which covariates may affect the baseline survival function and how significant they are. Based on the test results, we can always remove the covariates that have $BF_{10} < 1$ from the LDTFP modeling. For this reason, Table C.1 reports the parameter inferences with the LDTFP specified according to the truth, i.e., $\tilde{\mathbf{x}}_{ij} = (1, z_{2ij})'$ under Scenario I and $\tilde{\mathbf{x}}_{ij} = 1$ under Scenarios II & III. The results reveal that the CP values for $\tilde{\boldsymbol{\beta}}$ are now around the nominal 95% except the CP for β_2 in Scenario I. It seems that our approach tends to slightly overestimate the standard deviation of the covariate coefficient when that covariate highly affects the baseline function.

Sensitivity analysis to the prior of α

We consider two additional priors $\alpha \sim \Gamma(2,2)$ and $\alpha \sim \Gamma(20,2)$ with $L = 4$ to check how the results change compared with the results shown in the main article. The parameter estimates and hypothesis tests are not essentially affected by the choices of hyperparameters in the prior of α, although we observe that the survival curve estimates under $\alpha \sim \Gamma(2,2)$ are slightly better than those under $\alpha \sim \Gamma(20,2)$.

Table C.2 Simulation data. The proportions of BF_{10} greater than 3, 10 and 30, respectively, across Monte Carlo simulations under the prior $\alpha \sim \Gamma(20,2)$.

Scenario	H_0	% $BF_{10} > 3$	% $BF_{10} > 10$	% $BF_{10} > 30$
I	$\boldsymbol{\Upsilon}_1 = \mathbf{0}$	0.048	0.01	0.002
	$\boldsymbol{\Upsilon}_2 = \mathbf{0}$	1	1	1
II	$\boldsymbol{\Upsilon}_1 = \mathbf{0}$	0.144	0.018	0.002
	$\boldsymbol{\Upsilon}_2 = \mathbf{0}$	0.060	0.008	0.002
	$\boldsymbol{\Upsilon}_{-0} = \mathbf{0}$	0.098	0.022	0.006
	$\boldsymbol{\Upsilon} = \mathbf{0}$	1	1	1
III	$\boldsymbol{\Upsilon}_1 = \mathbf{0}$	0.038	0.004	0
	$\boldsymbol{\Upsilon}_2 = \mathbf{0}$	0.054	0.010	0.002
	$\boldsymbol{\Upsilon}_{-0} = \mathbf{0}$	0.068	0.014	0.004
	$\boldsymbol{\Upsilon} = \mathbf{0}$	0.070	0.016	0.004

Table C.3 Simulation data. The proportions of BF_{10} greater than 3, 10 and 30, respectively, across Monte Carlo simulations under the prior $\alpha \sim \Gamma(2,2)$.

Scenario	H_0	% $BF_{10} > 3$	% $BF_{10} > 10$	% $BF_{10} > 30$
I	$\boldsymbol{\Upsilon}_1 = \mathbf{0}$	0.028	0.006	0
	$\boldsymbol{\Upsilon}_2 = \mathbf{0}$	1	1	1
II	$\boldsymbol{\Upsilon}_1 = \mathbf{0}$	0.032	0.014	0.004
	$\boldsymbol{\Upsilon}_2 = \mathbf{0}$	0.022	0.012	0.006
	$\boldsymbol{\Upsilon}_{-0} = \mathbf{0}$	0.010	0.008	0.006
	$\boldsymbol{\Upsilon} = \mathbf{0}$	1	1	1
III	$\boldsymbol{\Upsilon}_1 = \mathbf{0}$	0.052	0.006	0.002
	$\boldsymbol{\Upsilon}_2 = \mathbf{0}$	0.038	0.004	0
	$\boldsymbol{\Upsilon}_{-0} = \mathbf{0}$	0.018	0.002	0.002
	$\boldsymbol{\Upsilon} = \mathbf{0}$	0.006	0.002	0

Table C.4 Simulation data. True value, averaged bias (BIAS) and posterior standard deviation (PSD) of each point estimate (i.e. posterior mean), standard deviation (across Monte Carlo simulations) of the point estimates (SD-Est) and coverage probability (CP) for the 95% credible intervals under the prior $\alpha \sim \Gamma(20, 2)$.

Scenario	Parameter	True	BIAS	PSD	SD-Est	CP
I	β_0	-1.0	0.012	0.065	0.058	0.970
	β_1	1.0	-0.002	0.052	0.046	0.972
	β_2	-0.5	0.000	0.110	0.094	0.978
	τ^2	0.1	0.025	0.054	0.040	0.990
II	β_0	-1.0	0.018	0.109	0.086	0.970
	β_1	1.0	-0.009	0.067	0.053	0.978
	β_2	-0.5	0.003	0.148	0.113	0.986
	τ^2	0.1	0.020	0.050	0.039	0.986
III	β_0	-1.0	0.004	0.060	0.056	0.968
	β_1	1.0	0.008	0.045	0.041	0.952
	β_2	-0.5	0.000	0.082	0.079	0.954
	τ^2	0.1	0.037	0.060	0.047	0.988

Table C.5 Simulation data. True value, averaged bias (BIAS) and posterior standard deviation (PSD) of each point estimate (i.e. posterior mean), standard deviation (across Monte Carlo simulations) of the point estimates (SD-Est) and coverage probability (CP) for the 95% credible intervals under the prior $\alpha \sim \Gamma(2, 2)$.

Scenario	Parameter	True	BIAS	PSD	SD-Est	CP
I	β_0	-1.0	0.017	0.072	0.065	0.952
	β_1	1.0	0.002	0.059	0.057	0.946
	β_2	-0.5	0.002	0.143	0.121	0.970
	τ^2	0.1	0.050	0.059	0.054	0.944
II	β_0	-1.0	0.035	0.131	0.114	0.933
	β_1	1.0	0.006	0.079	0.077	0.911
	β_2	-0.5	-0.009	0.181	0.159	0.929
	τ^2	0.1	0.043	0.054	0.046	0.966
III	β_0	-1.0	0.009	0.068	0.063	0.956
	β_1	1.0	0.007	0.052	0.050	0.960
	β_2	-0.5	-0.001	0.093	0.085	0.970
	τ^2	0.1	0.050	0.063	0.053	0.958

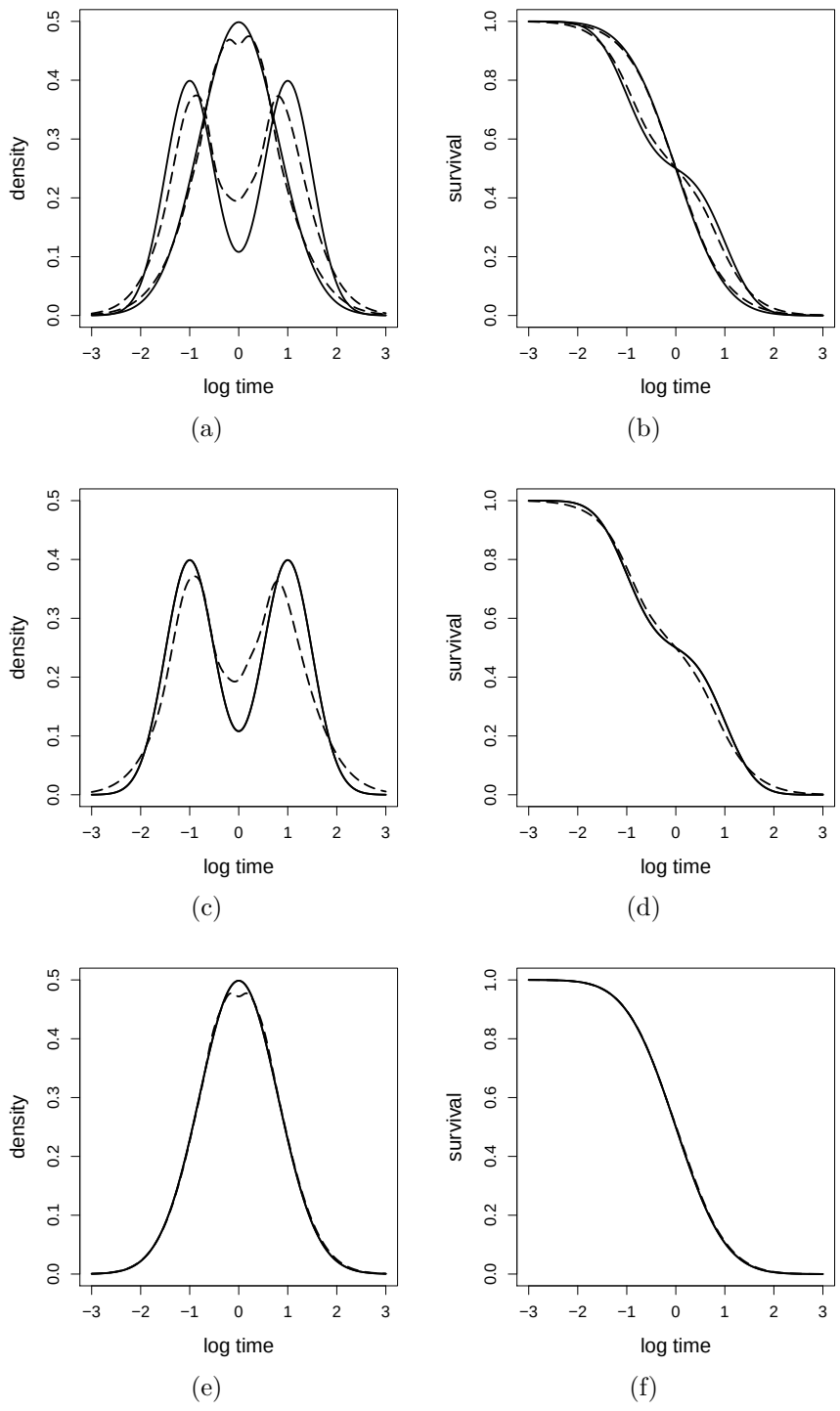


Figure C.4 Simulated data. Mean, across simulations, of the posterior mean of the density functions (left three panels) and survival functions (right three panels) of log survival times for a particular county under Scenario I (panel a and b), Scenario II (panel c and d) and Scenario III (panel e and f). The curves in each panel of a and b are for $(z, x) = (1.5, 1)$ (initially left curve) and $(z, x) = (1, 0)$. The other curves are for $(z, x) = (1, 0)$. The true curves are represented by continuous lines. The results under the proposed model are represented by dashed lines. Prior $\alpha \sim \Gamma(20, 2)$ is used.

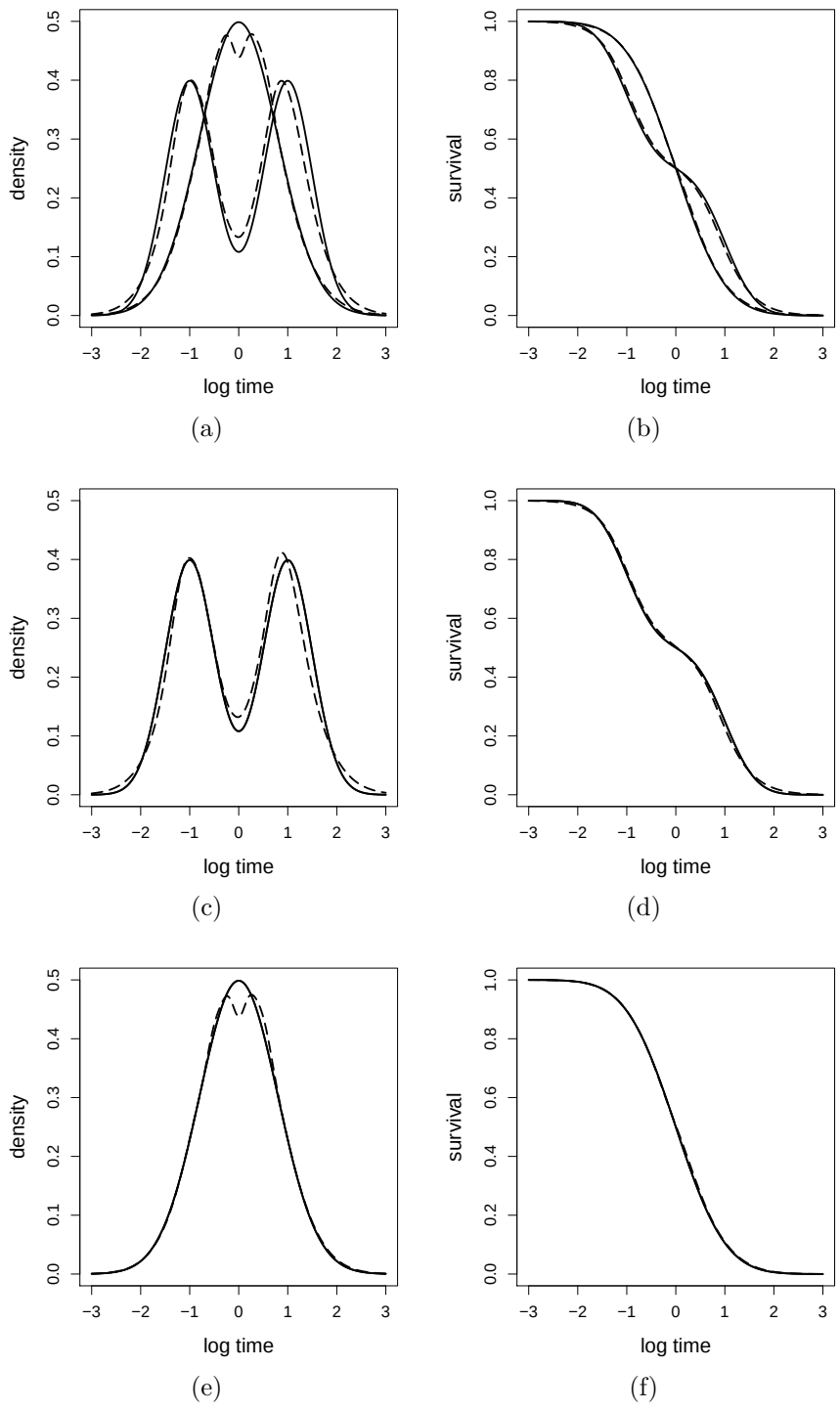


Figure C.5 Simulated data. Mean, across simulations, of the posterior mean of the density functions (left three panels) and survival functions (right three panels) of log survival times for a particular county under Scenario I (panel a and b), Scenario II (panel c and d) and Scenario III (panel e and f). The curves in each panel of a and b are for $(z, x) = (1.5, 1)$ (initially left curve) and $(z, x) = (1, 0)$. The other curves are for $(z, x) = (1, 0)$. The true curves are represented by continuous lines. The results under the proposed model are represented by dashed lines. Prior $\alpha \sim \Gamma(2, 2)$ is used.

Sensitivity analysis to L

We additionally consider $L = 5$ with $\alpha \sim \Gamma(5, 1)$ to check how the results change In comparison to the results shown in the main article. We can see that parameter and curve estimates are not sensitive to the choice of L, but hypothesis tests are more sensitive. Our tests become more conservative when L increases from 4 to 5 as indicated in comparison between Table 1 of main article and Table C.6. For instance, when L increases from 4 to 5, the Bayes factor values become larger overall, especially for Scenario III (i.e. when the log-normal AFT is the truth).

Table C.6 Simulation data. The proportions of BF_{10} greater than 3, 10 and 30, respectively, across Monte Carlo simulations. The LDTFP is truncated at $L = 5$.

Scenario	H_0	% $BF_{10} > 3$	% $BF_{10} > 10$	% $BF_{10} > 30$
I	$\boldsymbol{\Upsilon}_1 = \mathbf{0}$	0.19	0.032	0.004
	$\boldsymbol{\Upsilon}_2 = \mathbf{0}$	1	1	1
II	$\boldsymbol{\Upsilon}_1 = \mathbf{0}$	0.176	0.034	0.006
	$\boldsymbol{\Upsilon}_2 = \mathbf{0}$	0.113	0.026	0.006
	$\boldsymbol{\Upsilon}_{-0} = \mathbf{0}$	0.113	0.038	0.012
	$\boldsymbol{\Upsilon} = \mathbf{0}$	1	1	1
III	$\boldsymbol{\Upsilon}_1 = \mathbf{0}$	0.488	0.058	0.004
	$\boldsymbol{\Upsilon}_2 = \mathbf{0}$	0.318	0.048	0.010
	$\boldsymbol{\Upsilon}_{-0} = \mathbf{0}$	0.824	0.306	0.064
	$\boldsymbol{\Upsilon} = \mathbf{0}$	0.956	0.614	0.264

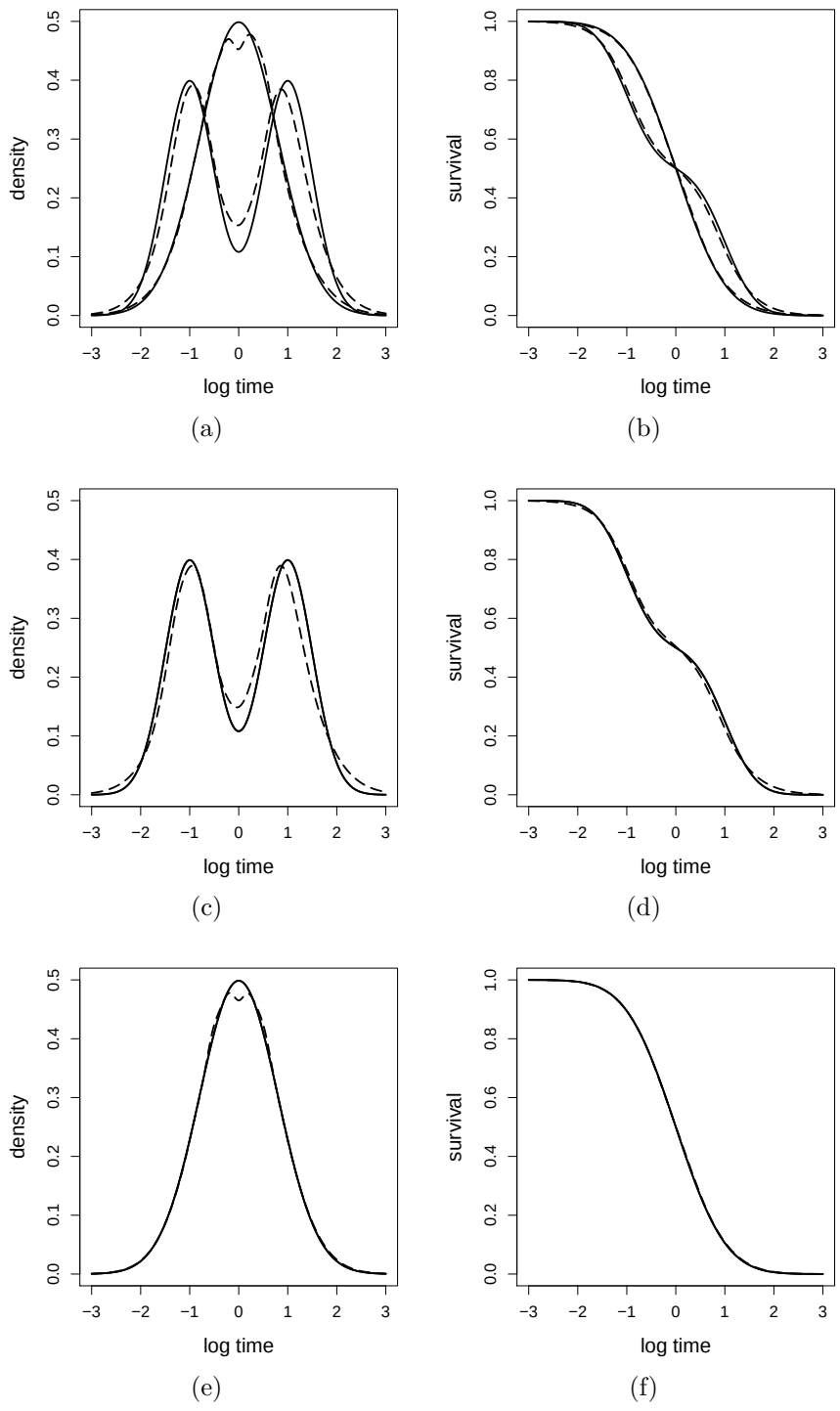


Figure C.6 Simulated data. Mean, across simulations, of the posterior mean of the density functions (left three panels) and survival functions (right three panels) of log survival times for a particular county under Scenario I (panel a and b), Scenario II (panel c and d) and Scenario III (panel e and f). The curves in each panel of a and b are for $(z, x) = (1.5, 1)$ (initially left curve) and $(z, x) = (1, 0)$. The other curves are for $(z, x) = (1, 0)$. The true curves are represented by continuous lines. The results under the proposed model are represented by dashed lines. The LDTFP is truncated at $L = 5$.

Table C.7 Simulation data. True value, averaged bias (BIAS) and posterior standard deviation (PSD) of each point estimate (i.e. posterior mean), standard deviation (across Monte Carlo simulations) of the point estimates (SD-Est) and coverage probability (CP) for the 95% credible intervals. The LDTFP is truncated at $L = 5$.

Scenario	Parameter	True	BIAS	PSD	SD-Est	CP
I	β_0	-1.0	0.012	0.068	0.064	0.948
	β_1	1.0	0.002	0.056	0.052	0.966
	β_2	-0.5	0.005	0.128	0.103	0.980
	τ^2	0.1	0.037	0.056	0.044	0.980
II	β_0	-1.0	0.039	0.126	0.107	0.941
	β_1	1.0	-0.003	0.076	0.067	0.953
	β_2	-0.5	-0.020	0.171	0.142	0.955
	τ^2	0.1	0.035	0.053	0.044	0.970
III	β_0	-1.0	0.004	0.062	0.056	0.970
	β_1	1.0	0.002	0.047	0.042	0.962
	β_2	-0.5	0.001	0.085	0.074	0.978
	τ^2	0.1	0.042	0.062	0.050	0.978

Comparison with the censored quantile regression model under Scenario I

For further comparison, we also fitted the censored quantile regression model (Portnoy, 2003) using the function `crq` available in the `quantreg` R package (Koenker, 2008) under Scenario I, but with the data generated without frailties. Note that `quantreg` does not allow spatial information. In comparison to our approach, the results show that `crq` provides almost two times greater standard deviation estimates for non-intercept coefficients, and more troubling, the coverage probability for estimating β_2 is much lower than 95%. These findings inform us that ignoring heteroscedastic errors could result in badly overestimated standard deviations and low coverage probabilities.

Table C.8 Simulation data – Scenario I. True value, averaged bias (BIAS) and posterior standard deviation (PSD) of each point estimate (i.e. posterior mean), standard deviation (across Monte Carlo simulations) of the point estimates (SD-Est) and coverage probability (CP) for the 95% credible intervals for the regression coefficients. The results are presented under the proposed model (via `frailtyGAFT`) and under the censored quantile regression model (via `crq`).

		`frailtyGAFT`				`crq`			
Parameters	True	BIAS	PSD	SD-Est	CP	BIAS	PSD	SD-Est	CP
β_0	-1.0	0.011	0.061	0.060	0.952	0.007	0.070	0.067	0.916
β_1	1.0	-0.001	0.034	0.040	0.900	0.003	0.089	0.084	0.916
β_2	-0.5	0.013	0.124	0.110	0.944	0.011	0.222	0.227	0.866

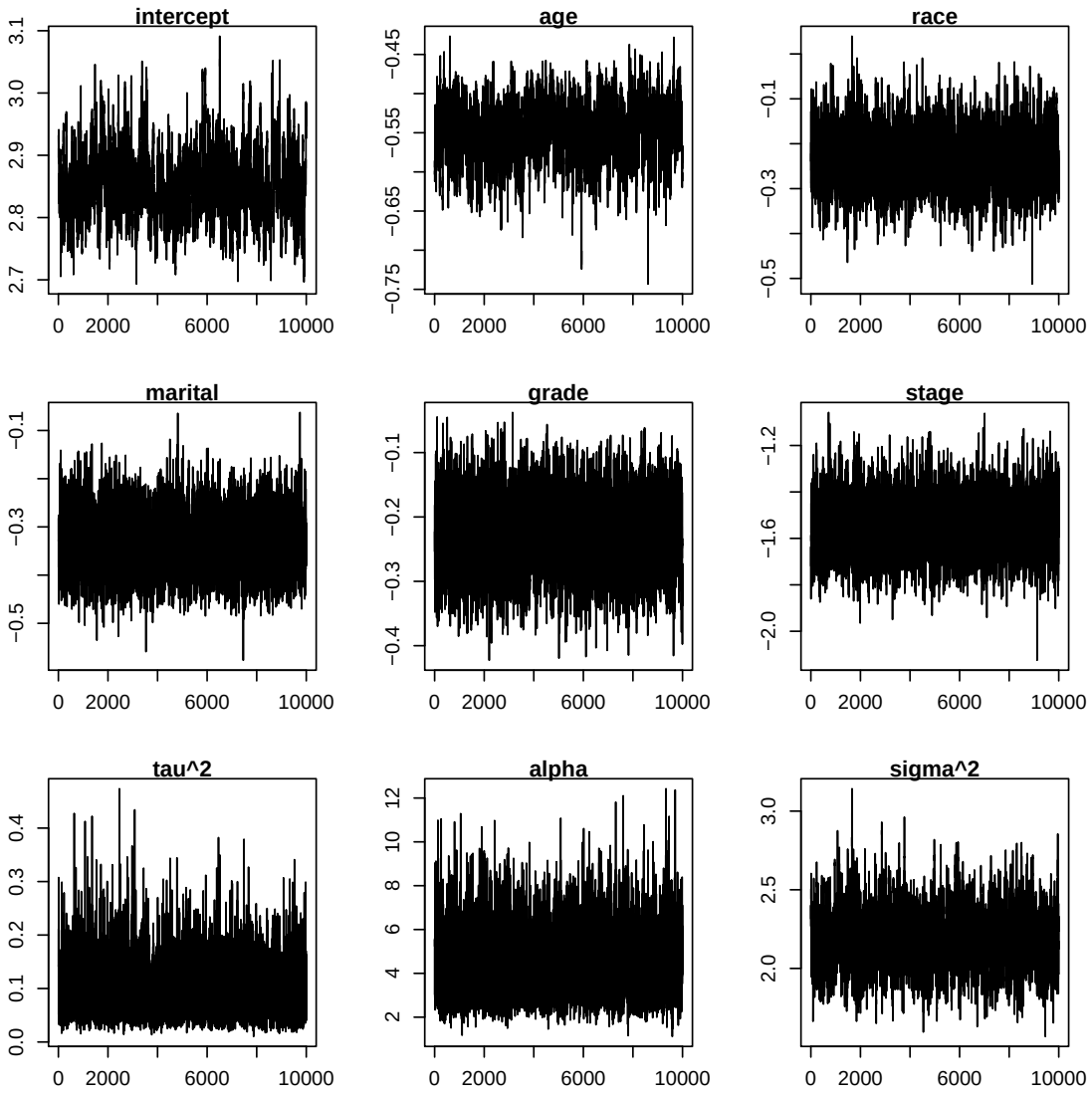


Figure C.7 SEER data. Trace and posterior density plots for model parameters.

C.4 Additional analysis of SEER prostate cancer data

MCMC chain analysis and convergence diagnostics

We present the posterior trace plots for $\tilde{\boldsymbol{\beta}}$, α and σ^2 in Figure C.7, and their autocorrelation function (ACF) plots together with effective samples sizes in Figure C.8. The Markov chain mixed reasonably well regardless of the high dimension of parameters in our model.

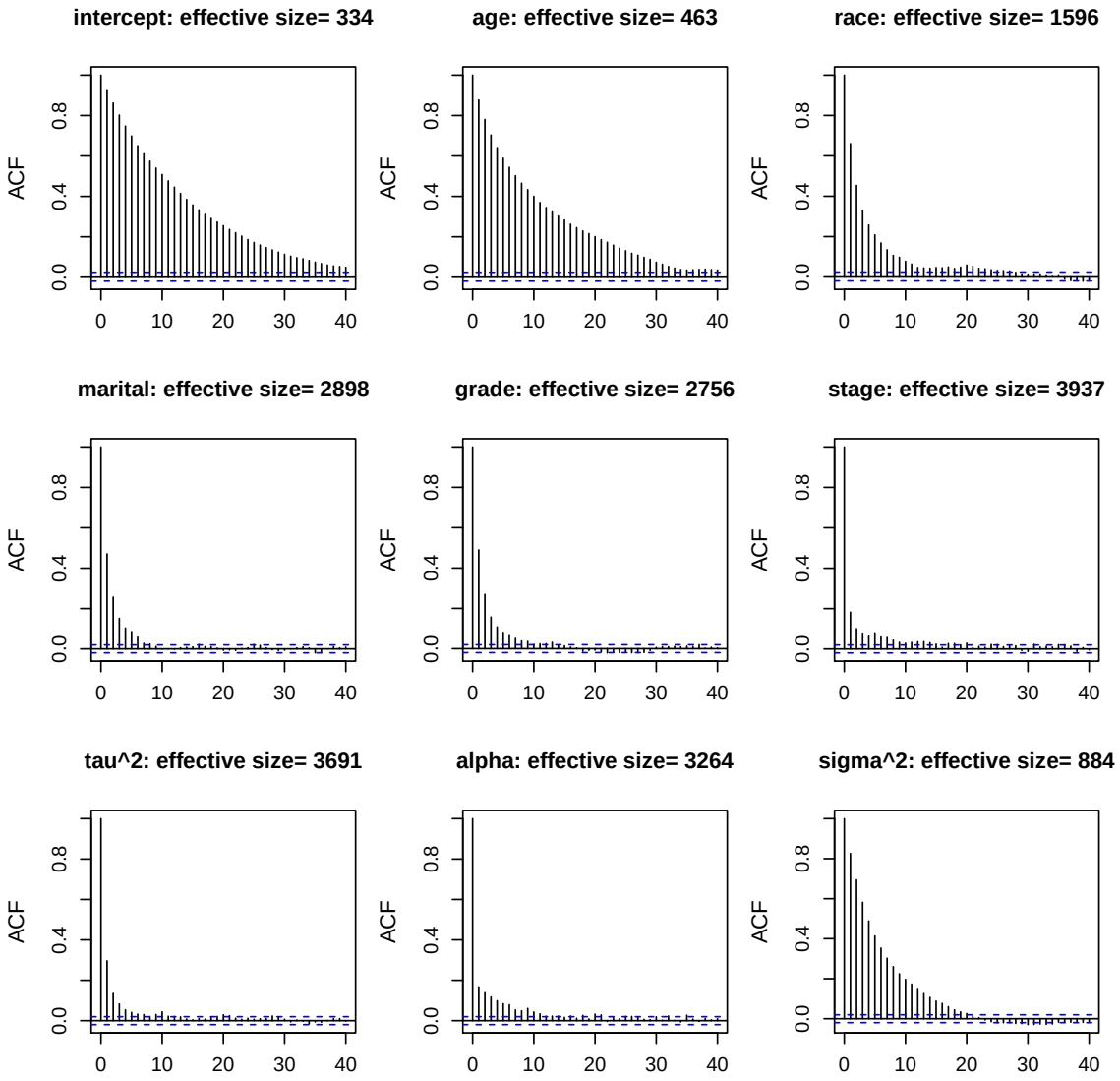


Figure C.8 SEER data. ACF plots and effective sample sizes for model parameters.

Table C.9 SEER data. Posterior means (95% credible intervals) of fixed effects $\tilde{\boldsymbol{\beta}}$ under the GAFT model (via `frailtyGAFT`) and under the censored quantile regression model (via `crq`). Results are based on standardized ages.

	GAFT			Quantile regression		
Covariates	Mean	Std. Dev.	CI	Mean	Std. Dev.	CI
Intercept	2.87	0.04	(2.79, 2.96)	2.78	0.08	(2.63, 2.94)
Age	-0.55	0.03	(-0.61, -0.49)	-0.51	0.04	(-0.59, -0.42)
Race	-0.23	0.05	(-0.34, -0.14)	-0.21	0.20	(-0.61, 0.19)
Marital	-0.30	0.05	(-0.40, -0.21)	-0.29	0.05	(-0.39, -0.18)
Grade	-0.21	0.05	(-0.31, -0.12)	-0.18	0.05	(-0.28, -0.08)
Stage	-1.57	0.10	(-1.78, -1.37)	-1.63	0.11	(-1.85, -1.42)

Comparison with the censored quantile regression model

The covariate effects are also compared with those obtained under the censored quantile regression model (Portnoy, 2003), where we note that the standard deviation for race effect is five times greater than that under GAFT, and consequently age becomes insignificant. This finding is consistent with the results shown in simulations, where `crq` tends to severely overestimate the standard deviations of covariate effects when these covariates may potentially affect the baseline survival function.

Appendix D

Copyright Permissions to Reprint

D.1 Permission to reprint Chapter 1

Dear Dr.Zhou,

I was informed about your intention to use your chapter for dissertation. You can use the chapter in your dissertation as long as you give credits to Springer. Your right is outlined in the Consent to Publish form which was signed by your co-author. So I suggest you to inform this to your co-author.

Thanks,

Simbu

—

Silembarasan P

Springer, SPi Global

Production Editor

—

DLF – SEZ IT Park | Manapakkam | Chennai | India - 600 089

Tel.:+ 914443950500; Ext: 32828

Silembarasan.Panneerselvam@springer.com

D.2 Permission to reprint Chapter 2

Dear Haiming:

As an author on the original paper, you do not need permission from the IMS to

reprint it. You retain this right.

Let me know if you have further questions.

Best-

Elyse

Elyse Gustafson

Executive Director

Institute of Mathematical Statistics

PO Box 22718

Beachwood, OH 44122 USA

T: 216-295-2340

T: 877-557-4674

F: 216-295-5661

E: erg@imstat.org

W: http://imstat.org

D.3 Permission to reprint Chapter 3

Dear Haiming Zhou:

Thank you for your request. Permission is granted to use the submitted version of your article as part of your dissertation. Please reference the article as pending publication.

Sincerely,

Paulette Goldweber

Associate Manager, Permissions

Wiley

pgoldweb@wiley.com

T +1 201-748-8765

F +1 201-748-6008

111 River Street, MS 4-02

Hoboken, NJ 07030-5774

U.S.

permissions@wiley.com