Bayesian Analysis (2012) **7**, Number 1, pp. 147–188

A Bayesian Semiparametric Temporally–Stratified Proportional Hazards Model with Spatial Frailties

Timothy E. Hanson*, Alejandro Jara† and Luping Zhao‡

Abstract. Incorporating temporal and spatial variation could potentially enhance information gathered from survival data. This paper proposes a Bayesian semiparametric model for capturing spatio–temporal heterogeneity within the proportional hazards framework. The spatial correlation is introduced in the form of county–level frailties. The temporal effect is introduced by considering the stratification of the proportional hazards model, where the time–dependent hazards are indirectly modeled using a probability model for related probability distributions. With this aim, an autoregressive dependent tailfree process is introduced. The full Kullback–Leibler support of the proposed process is provided. The approach is illustrated using simulated data and data from the Surveillance Epidemiology and End Results database of the National Cancer Institute on patients in Iowa diagnosed with breast cancer.

Keywords: Spatio–temporal modeling, Dependent processes, Tailfree processes, Breast cancer

1 Introduction

The proportional hazards (PH) model has been widely used in the analysis of censored survival data. In the usual PH regression model the hazard function for an individual with covariate vector $\mathbf{x}$, $\lambda(\cdot \mid \mathbf{x})$, is specified as $\lambda(\cdot \mid \mathbf{x}) = \lambda_0(\cdot)\exp\{\mathbf{x}'\beta\}$, where β is a vector of regression coefficients and $\lambda_0(\cdot)$ is a baseline hazard function for individuals with $\mathbf{x} = \mathbf{0}$. The PH assumption has several advantages: (i) the effect of a factor can be easily summarized by the relative risk, and (ii) a mathematical simplicity which has been exploited by Cox (1972) to produce a simple semi–parametric approach via the partial likelihood. The PH is, however, a strong assumption which may not be appropriate for all datasets (see, e.g., Hanson and Yang 2007; De Iorio et al. 2009; Jara et al. 2010).

One way to relax the PH assumption is by including strata (Kalbfleisch and Prentice 1980). In the stratified PH model, the hazard function for the jth stratum, $\lambda^j(\cdot \mid \mathbf{x})$, $j \in \mathcal{J} \subset I\!N$, is specified such that

$$\lambda^j(\cdot \mid \mathbf{x}) = \lambda_{0j}(\cdot)\exp\{\mathbf{x}'\beta\},$$

where the $\{\lambda_{0j} : j \in \mathcal{J}\}$ are unknown baseline hazards. Such a generalization is useful, for instance, if some discrete explanatory variables do not appear to have a multiplicative

*Department of Statistics, University of South Carolina, Columbia, SC, hansont@stat.sc.edu
†Department of Statistics, Pontificia Universidad Católica de Chile, Santiago, Chile atjara@uc.cl
‡Eli Lilly and Company, Indianapolis, IN, zhao_luping@lilly.com

 DOI:10.1214/12-BA705

effect on the hazard function. The primary interest of the stratified model is usually confined to the within stratum comparison of specific risk factors while attempting to adjust for the strata effect. In standard implementations of the stratified PH model the elements in the set $\{\lambda_{0j} : j \in \mathcal{J}\}$ are assumed unrelated, which may be inefficient, incurring some loss of information, especially for highly stratified data. Furthermore, the classical semiparametric approach based on the partial likelihood does not allow for prediction of new data, which is of interest, for instance, when the stratifying variable is a time indicator related to the definition of cohorts.

In this paper a Bayesian semiparametric temporally–stratified PH model that accounts for spatial association is proposed. The spatial association is taken into account by considering spatial frailty terms in the model. In order to borrow strength across the temporal strata, a Bayesian nonparametric model for the set of stratum–specific hazards $\{\lambda_{0j} : j \in \mathcal{J}\}$ is proposed. We are unaware of any approaches, Bayesian or otherwise, allowing for evolution of baseline survival in time or space in the PH model context. He and Lawless (2003) consider a likelihood–based approach for bivariate survival data. Hennerfeind et al. (2006) consider flexible models for geostatistical or lattice frailties incorporating additive effects, but enforce a static baseline hazard function $\lambda_0(t)$. The marginal transformation approach of Li and Lin (2006) also considers a static hazard. Cai et al. (2007) allow for varying hazards $\lambda_{0j}(t)$, but consider a marginal approach, similar to a stratified analysis on j.

Recently, dependent priors have been explored for modeling spatially or longitudinally dependent distributions based on stick–breaking priors, most famously the Dirichlet process (DP) prior (Ferguson 1974), Polya trees (Trippa et al. 2011) and more general tailfree processes (Jara and Hanson 2011). A random probability measure G, defined on a measurable space $(\Theta, \mathcal{B})$, follows a DP with parameters (M, G_0), where $M \in R_0^+$ and G_0 is a probability measure on $(\Theta, \mathcal{B})$, written as $G \mid M, G_0 \sim DP(MG_0)$, if the trajectories of the process can be almost surely (a.s.) represented by the stick–breaking construction (Sethuraman 1994), $G(B) = \sum_{i=1}^{\infty} w_i \delta_{\theta_i}(B)$, $B \in \mathcal{B}$, where $\delta_\theta(\cdot)$ is the Dirac measure at θ, $w_i = V_i \prod_{j<i}(1 - V_j)$, with $V_i \mid M \overset{iid}{\sim} Beta(1, M)$, and $\theta_i \mid G_0 \overset{iid}{\sim} G_0$. Motivated by the results of MacEachern (1999, 2000), Gelfand et al. (2005) generalize the support points θ_i to stochastic processes $\{\theta_i(\mathbf{s}) : \mathbf{s} \in \mathcal{S}\}$. Kottas et al. (2008) consider an extension of this model for a spatial epidemiological application. De Iorio et al. (2004) consider a related model with ANOVA–like rather than spatial or temporal dependence. Griffin and Steel (2006) consider an alternative model where also the weights $w_i(\mathbf{s})$ vary in space or time. Reich and Fuentes (2007) and Dunson et al. (2007) extend the stick–breaking priors to the multivariate spatial and regression settings, respectively, developing related approaches based on augmenting the weights $\{w_i\}_{i=1}^{\infty}$ with spatially– or covariate–weighted kernel functions. Other approaches can be found, for instance, in Dunson and Park (2008), Chung and Dunson (2009), Duan et al. (2007), Gelfand et al. (2007) and Taddy (2008).

As implemented, the dependent generalizations of stick–breaking priors have focused more on regression with a spatially varying component, $y(\mathbf{s}) = \mu(\mathbf{s}) + \mathbf{x}(\mathbf{s})'\beta + e(\mathbf{s})$, where $\mu(\mathbf{s})$ follows some development of a dependent stick–breaking prior, the $\mathbf{x}(\mathbf{s})$ are

covariates, and the $e(\mathbf{s})$ are *iid* white noise, typically $N(0, \sigma^2)$; see e.g. expressions (1) in Kottas et al. (2008), (6) in Griffin and Steel (2006), (1) in Dunson et al. (2007), and (4) in Reich and Fuentes (2007). The inclusion of $e(\mathbf{s})$ is actually required to use modifications of the now well–developed computational theory available for fitting Dirichlet process mixture (DPM) models. In these applications and elsewhere, the DP is usually convolved with a smooth, typically normal kernel $\phi(\cdot \mid \mu, \sigma^2)$, resulting in a continuous measure with density $f(\cdot \mid G, \sigma^2) = \int_R \phi(\cdot \mid \mu, \sigma^2) dG(\mu)$. This essentially amounts to requiring a nugget effect in the spatial or covariate process, but practically translates into a process that behaves similarly to a finite mixture model with component variances fixed at σ^2 and locations or weights varying smoothly in space or time. This built–in linear structure implies that dependent stick–breaking priors lend themselves naturally to accelerated failure time models in a survival analysis context, but not necessarily to PH, proportional odds, or other types of survival models that relate survival to covariates in a complex way (Hanson 2006; Zhao et al. 2009). An early (non–dependent) approach that took advantage of such structure was the DPM treatment of the accelerated failure time model by Kuo and Mallick (1997), where they viewed their model as "smoothing" the discrete realization G with a known continuous kernel ($e(\mathbf{s})$ above). An approach to PH models based on the extended gamma process (Ishwaran and James 2004) takes advantage of the fact that the DP is simply a normalized gamma process. With considerable effort, this could lead to dependent stick–breaking versions of the PH model, but has not been pursued as of yet.

In this paper, we propose a temporally–stratified PH model with spatial frailties. The time–dependent baseline hazards $\{\lambda_{0j} : j \in \mathcal{J}\}$ are indirectly modeled using a dependent tailfree process prior on the set of associated baseline distributions $\{F_{0j} : j \in \mathcal{J}\}$. The model includes several interesting special cases (e.g. purely stratified, temporally static) that can be tested versus the alternative model allowing for temporal baseline evolution, and allows for the estimation of survival and hazard curves for future years. The remainder of the paper is organized as follows. Section 2 gives a detailed description of the proposed model, including theoretical properties of the proposal such as its full Kullback–Leibler support. Section 3 provides the computational details related to MCMC implementation of the model. Section 4 offers the analysis of simulated data and a detailed analysis of data from the Surveillance Epidemiology and End Results (SEER) database of the National Cancer Institute on patients in Iowa diagnosed with breast cancer. A final discussion concludes the article.

2 The proposal

2.1 The stratified conditional proportional hazards model

Let $T_{ijk} \in I\!R^+$ be the time to the event of interest for the kth subject, diagnosed in the jth year in the ith county, with $i = 1, \ldots, I$, $j = 1, \ldots, J$, and $k = 1, \ldots, n_{ij}$. Let $\mathbf{x}_{ijk}$ be a p–dimensional vector of explanatory covariates associated with the kth individual, diagnosed in the jth year in the ith county. Finally, let γ_i be the frailty term associated with the ith county. We assume that, given the spatial effect $\gamma_i \in I\!R$, the observations

T_{ijk} are independent, following a temporally–stratified proportional hazards model with conditional density $f(\cdot \mid \mathbf{x}_{ijk}, j, \gamma_i)$, that is,

$$T_{ijk} \mid \mathbf{x}_{ijk}, \beta, j, \gamma_i \overset{ind.}{\sim} f(\cdot \mid \mathbf{x}_{ijk}, \beta, j, \gamma_i), \tag{1}$$

where

$$f(t \mid \mathbf{x}_{ijk}, \beta, j, \gamma_i) = \exp\{\mathbf{x}'_{ijk}\beta + \gamma_i\}\lambda_{0j}(t) \exp\left[-\exp\{\mathbf{x}'_{ijk}\beta + \gamma_i\} \int_0^t \lambda_{0j}(u)du\right]. \tag{2}$$

Here, $\lambda_{0j}(\cdot) = f_{0j}(\cdot)/S_{0j}(\cdot)$ denotes the baseline hazard for the jth year, with f_{0j} and $S_{0j} = 1 - F_{0j}$ being the year–specific baseline density and survival function, respectively. We assume common county–specific frailty terms across time and that

$$\gamma = (\gamma_1, \ldots, \gamma_I) \mid H \sim H,$$

where H is a probability distribution supported on R^I. Several parametric and nonparametric options could be considered for H (see, e.g. Jara et al. 2009). As the modeling of the H distribution is not the main aspect of the present proposal, we restrict ourselves to Gaussian alternatives. Following Banerjee et al. (2003), we focus on a version of the commonly used intrinsic conditionally autoregressive (CAR) prior of Besag et al. (1991). Here, frailty terms are conditionally specified as

$$\gamma_i|\lambda, \{\gamma_j\}_{j\neq i} \overset{ind.}{\sim} N(\bar{\gamma}_i, (\lambda d_i)^{-1}),$$

where d_i denotes the number of counties adjacent to county i, and $\bar{\gamma}_i$ is the sample mean of the d_i county effects in $\{\gamma_j\}_{j\neq i}$ adjacent to county i. Alternatively, an exchangeable model is also considered by assuming

$$\gamma_i|\lambda \overset{iid}{\sim} N(0, \lambda^{-1}),$$

where $N(\mu, \sigma^2)$ stands for the normal distribution with mean μ and variance σ^2.

The model assumes a common baseline distribution F_{0j} across the counties within a year. Several strategies could be considered to model the set of baseline distributions $\mathcal{F} = \{F_{0j} : j \in \mathcal{J}\}$, where $\mathcal{J} = \{1, \ldots, J\}$. One extreme model strategy would be to assume a common baseline distribution $F_{0j} \equiv F_0$, $j \in \mathcal{J}$. At the other extreme, the unknown probability distributions could be considered as independent parameters and to place independent priors $F_{0j} \mid H_j \sim H_j$, $j \in \mathcal{J}$, related at the level of the hyper–parameters. For many applications, the first case is too restrictive by assuming essentially one baseline population, and the latter could be too inefficient by borrowing little strength from potentially related information sources. Instead, we consider a Bayesian nonparametric dependent model which allows linking the elements in $\mathcal{F}$ at an intermediate level and the degree of dependence among the random distributions is governed by the year indicator. Specifically, we assume

$$\mathcal{F} = \{F_{0j} : j \in \mathcal{J}\} \mid h, \Pi^\theta, c, \kappa, \rho \sim \text{ARDTFP}\left(h, \Pi^\theta, \mathcal{A}^{c,\kappa,\rho}\right), \tag{3}$$

where ARDTFP refers to an autoregressive dependent tailfree process with parameters $(h, \Pi^\theta, \mathcal{A}^{c,\kappa,\rho})$. To better explain our proposal, we start with a review of tailfree and dependent tailfree processes next.

2.2 Tailfree and dependent tailfree processes

Tailfree processes are stochastic processes that can be defined to have trajectories on the space of probability distributions (see, e.g. Freedman 1963; Fabius 1964; Ferguson 1974). A tailfree random probability measure F with support on the positive real line, $\mathbb{R}^+$, is defined by allocations of random probabilities to increasingly refined partitions of $\mathbb{R}^+$. Let $E = \{0, 1\}$ and E^m be the m–fold Cartesian product $E \times \cdots \times E$. Further, set $E^* = \bigcup_{m=0}^{\infty} E^m$. Consider the sequence of partitions of $\mathbb{R}^+$ given by $\pi_1 = \{B_0, B_1\}$, $\pi_2 = \{B_{00}, B_{01}, B_{10}, B_{11}\}$, ..., such that, for every $m \in \mathbb{N}$ and every $\epsilon = \varepsilon_1 \cdots \varepsilon_m \in E^*$, $B_\epsilon = B_{\epsilon 0} \cup B_{\epsilon 1}$ and $B_{\epsilon 0} \cap B_{\epsilon 1} = \emptyset$. Assume that $B_{\epsilon 0}$ lies below $B_{\epsilon 1}$ and that B_ϵ is a left–open right–closed interval unless ϵ is a string of ones only. Throughout the paper, we use the convention that $\epsilon = \varepsilon_1 \cdots \varepsilon_{m-1} 0 = 0$ and $\epsilon = \varepsilon_1 \cdots \varepsilon_{m-1} 1 = 1$, if $m = 1$. Let $\Pi = \{\pi_1, \pi_2, \ldots\}$ and we further assume that the partitions form a rich class in the sense that $\Pi = \cup_{l=1}^{\infty} \pi_l$ is a generator of the Borel σ–field of $\mathbb{R}^+$.

Definition 1. *Let Π be a sequence of binary partitions as before. A random probability measure F on $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ is said to be a tailfree random probability measure with respect to the set of partitions Π, if there exists a collection $\mathcal{Y} = \{Y_\epsilon : \epsilon \in E^*\}$ of $[0, 1]$–valued random variables such that the following hold:*

(i) The vectors (Y_0, Y_1), $(Y_{00}, Y_{01}, Y_{10}, Y_{11})$, ... , are mutually independent.

(ii) For every $\epsilon = \varepsilon_1 \cdots \varepsilon_m \in E^*$, $Y_{\varepsilon_1 \cdots \varepsilon_{m-1} 0} + Y_{\varepsilon_1 \cdots \varepsilon_{m-1} 1} = 1$ a.s.

(iii) For every $\epsilon = \varepsilon_1 \cdots \varepsilon_m \in E^*$, the random probability measure F is related to $\mathcal{Y}$ through the relations

$$F(B_\epsilon) = \prod_{l=1}^{m} Y_{\varepsilon_1 \cdots \varepsilon_l}.$$

Tailfree processes admit Polya trees and Dirichlet processes as important special cases. Polya trees assign independent beta distributions to the elements in the collection $\{Y_{\epsilon 0}\}_{\epsilon \in E^*}$. The Dirichlet process is tailfree with respect to any sequence of partitions, the only prior process that has this property. A tailfree prior can be constructed such that the random probability measures generated by it are absolutely continuous with respect to a given finite measure with probability one (Dubins and Freedman 1967; Kraft 1964; Metivier 1971).

Jara and Hanson (2011) proposed a class of dependent tailfree processes by modeling the tailfree conditional probabilities as transformed mean–zero Gaussian processes with continuous index set $\mathcal{Z} \subset \mathbb{R}^q$. We briefly provide some detail important to the exposition that follows. The following definition is a modification of the one of Jara and Hanson (2011). Let $\Pi^{\mathcal{Z}} = \{\Pi_{\mathbf{z}} : \mathbf{z} \in \mathcal{Z}\}$ be a set of dyadic partitions, where, as before, $\Pi_{\mathbf{z}}$ is a sequence of binary partitions of $\mathbb{R}^+$, for every $\mathbf{z} \in \mathcal{Z}$.

Definition 2. *Let $\Pi^{\mathcal{Z}}$ be a set of sequence of binary partitions as before, $h : \mathbb{R} \longrightarrow [0,1]$ be a strictly increasing continuous function, $\mathcal{A} = \{V_{\epsilon 0} : \epsilon \in E^*\}$ be a set of covariance functions, and $\mathcal{P}(\mathbb{R}^+)$ be the set of Borel probability measures on $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$. Let $\{F(\mathbf{z}, \cdot) : \mathbf{z} \in \mathcal{Z}\}$ be a $\mathcal{P}(\mathbb{R}^+)$–valued stochastic process on an appropriate probability space $(\Omega, \mathcal{C}, P)$ such that, for almost every $\omega \in \Omega$, the following hold:*

(i) The sets $\{\eta_{\epsilon 0}(\mathbf{z}, \omega) : \mathbf{z} \in \mathcal{Z}\}$, for every $\epsilon = \varepsilon_1 \cdots \varepsilon_{m-1} 0 \in E^*$, are realizations of mutually independent mean–zero Gaussian processes with associated covariance functions $V_{\epsilon 0}$.

(ii) For every $\mathbf{z} \in \mathcal{Z}$ and every $\epsilon 0 = \varepsilon_1 \cdots \varepsilon_{m-1} 0 \in E^*$, $Y_{\epsilon 0}(\mathbf{z}, \omega) = h\{\eta_{\epsilon 0}(\mathbf{z}, \omega)\}$ and $Y_{\epsilon 1}(\mathbf{z}, \omega) = 1 - Y_{\epsilon 0}(\mathbf{z}, \omega)$.

(iii) For every $\mathbf{z} \in \mathcal{Z}$ and every $\epsilon = \varepsilon_1 \cdots \varepsilon_m \in E^*$,

$$F(\mathbf{z}, \omega)(B_\epsilon^{\mathbf{z}}) = \prod_{l=1}^{m} Y_{\varepsilon_1 \cdots \varepsilon_l}(\mathbf{z}, \omega).$$

Such a process $\{F_{\mathbf{z}} = F(\mathbf{z}, \cdot) : \mathbf{z} \in \mathcal{Z}\}$ will be referred to as a dependent tailfree process with parameters $\left(h, \Pi^{\mathcal{Z}}, \mathcal{A}\right)$, and denoted DTFP $\left(h, \Pi^{\mathcal{Z}}, \mathcal{A}\right)$.

Equivalent to the case of tailfree processes for single probability measures, the DTFP can be centered at a collection of parametric distributions $\mathcal{F}^\theta = \{F_{\theta_{\mathbf{z}}} : \mathbf{z} \in \mathcal{Z}\}$. For instance, the elements in $\mathcal{F}^\theta$ can be of the form $\text{LL}(\mu(\mathbf{z}), \sigma^2(\mathbf{z}))$, where $\text{LL}(\mu, \sigma^2)$ stands for the log–logistic distribution with location parameter μ and scale parameter σ^2. In fact, if h is taken to be the cumulative distribution function of any absolutely continuous random variable with distribution symmetric around 0 and, for every $\mathbf{z} \in \mathcal{Z}$, the partition sets in $\Pi^{\mathbf{z}}$ are of the form $B_\epsilon^{\mathbf{z}} = \left[F_{\theta_{\mathbf{z}}}^{-1}(k/2^m), F_{\theta_{\mathbf{z}}}^{-1}([k+1]/2^m)\right]$, with $F_{\theta_{\mathbf{z}}}^{-1}(\cdot)$ being the quantile function of $F_{\theta_{\mathbf{z}}}$ and k the decimal representation of $\epsilon = \varepsilon_1 \cdots \varepsilon_m \in E^*$, then the process $\{F_{\mathbf{z}} : \mathbf{z} \in \mathcal{Z}\}$ is centered around $\mathcal{F}^\theta$, that is, $E(F_{\mathbf{z}}) = F_{\theta_{\mathbf{z}}}$, for every $\mathbf{z} \in \mathcal{Z}$. In order to make explicit the dependence of the set of sequences of partitions on the centering parameters $\theta = \{\theta_{\mathbf{z}} : \mathbf{z} \in \mathcal{Z}\}$, the notation $\Pi^{\mathcal{Z},\theta}$ is used.

Similar to tailfree processes for single probability measures, if h is taken as before, the dependent tailfree process can easily be constrained in order to have trajectories such that, for every $\mathbf{z} \in \mathcal{Z}$, $F_{\mathbf{z}}^{-1}(0.5) = F_{\theta_{\mathbf{z}}}^{-1}(0.5)$ a.s., that is, such that every random probability distribution in $\{F_{\mathbf{z}} : \mathbf{z} \in \mathcal{Z}\}$ has the same median as the corresponding centering distribution $F_{\theta_{\mathbf{z}}}$ with probability one, facilitating the interpretation of the centering parameters in regression models. This can easily be done by letting $\eta_0(\mathbf{z}, \cdot) = 0$ a.s., for every $\mathbf{z} \in \mathcal{Z}$.

The form of the covariance functions in $\mathcal{A}$ completely defines the process and determines important properties. For instance, Jara and Hanson (2011) showed that its specification determines whether the trajectories of the process have a density with respect to Lebesgue measure, the continuity properties of the process and the support of the process. We discuss these relevant properties in the context of the temporally–dependent tailfree process in the next section.

2.3 The autoregressive dependent tailfree process

We define a temporally–dependent tailfree process with discrete index set $\mathcal{J} \subseteq \mathbb{N}$, by considering first–order autoregressive mean–zero Gaussian processes in the definition of the DTFP. For every $\epsilon 0 = \varepsilon_1 \cdots \varepsilon_{m-1} 0 \in E^*$, let $\kappa_{\epsilon 0} \in (-1, 1)$ be an autoregressive parameter and set $\kappa = \{\kappa_{\epsilon 0} : \epsilon 0 \in E^*\}$. Let $\rho : \mathbb{N} \longrightarrow \mathbb{R}^+$ be a monotonic increasing function and c a positive constant. For every $\epsilon 0 = \varepsilon_1 \cdots \varepsilon_{m-1} 0 \in E^*$, let $V_{\epsilon 0}^{c,\kappa_{\epsilon 0},\rho}(j, j') = 2\kappa_{\epsilon 0}^{|j-j'|}[c\rho(m_{\epsilon 0})]^{-1}$, $\forall j, j' \in \mathcal{J}$, be an autoregressive covariance function, where $m_{\epsilon 0}$ is the length of the string $\epsilon 0$.

Definition 3. *Let $\Pi^{\mathcal{J},\theta}$ be a set of sequence of binary partitions as before, $h : \mathbb{R} \longrightarrow [0, 1]$ be a strictly increasing continuous function, $\rho : \mathbb{N} \longrightarrow \mathbb{R}^+$ be a monotonic increasing function, $\mathcal{A}^{c,\kappa,\rho} = \{V_{\epsilon 0}^{c,\kappa_{\epsilon 0},\rho} : \epsilon 0 \in E^*\}$ be the set of autoregressive covariance functions, as previously defined, and $\mathcal{P}(\mathbb{R}^+)$ be the set of Borel probability measures on $(\mathbb{R}+, \mathcal{B}(\mathbb{R}^+))$. Let $\{F(j, \cdot) : j \in \mathcal{J}\}$ be a $\mathcal{P}(\mathbb{R}^+)$–valued stochastic process on an appropriate probability space $(\Omega, \mathcal{C}, P)$ such that, for almost every $\omega \in \Omega$, the following hold:*

(i) The sets $\{\eta_{\epsilon 0}(j, \omega) : j \in \mathcal{J}\}$, for every $\epsilon 0 = \varepsilon_1 \cdots \varepsilon_{m-1} 0 \in E^*$, are realizations of mutually independent mean–zero first–order autoregressive Gaussian processes with associated covariance functions $V_{\epsilon 0}^{c,\kappa_{\epsilon 0},\rho}$.

(ii) For every $j \in \mathcal{J}$ and every $\epsilon 0 = \varepsilon_1 \cdots \varepsilon_{m-1} 0 \in E^*$, $Y_{\epsilon 0}(j, \omega) = h\{\eta_{\epsilon 0}(j, \omega)\}$ and $Y_{\epsilon 1}(j, \omega) = 1 - Y_{\epsilon 0}(j, \omega)$.

(iii) For every $j \in \mathcal{J}$ and $\epsilon = \varepsilon_1 \cdots \varepsilon_m \in E^*$,

$$F(j, \omega)\left(B_\epsilon^j\right) = \prod_{l=1}^{m} Y_{\varepsilon_1 \cdots \varepsilon_l}(j, \omega).$$

Such a process $\{F_j = F(j, \cdot) : j \in \mathcal{J}\}$ will be referred to as the autoregressive dependent tailfree process with parameters $\left(h, \Pi^{\mathcal{J},\theta}, \mathcal{A}^{c,\kappa,\rho}\right)$, and denoted ARDTFP $\left(h, \Pi^{\mathcal{J},\theta}, \mathcal{A}^{c,\kappa,\rho}\right)$.

From Proposition 4 in Jara and Hanson (2011), it immediately follows that if, for every $c \in \mathbb{R}^+$, the function ρ is taken such that $\sum_{l=1}^{\infty} \rho(l)^{-1} < \infty$, then with probability one the trajectories of the ARDTFP are in the space of the probability measures supported on $\mathbb{R}^+$ and are absolutely continuous w.r.t. Lebesgue measure. Under the assumption of the almost sure continuity of the Gaussian process involved in the definition of the DTFP, Jara and Hanson (2011) show its large Kullback–Leibler support. Since the assumption of almost sure continuity does not hold for the autoregressive Gaussian processes with discrete index $\mathcal{J}$, we provide a proof of the full Kullback–Leibler support of the ARDTFP. The following theorem is proved in Appendix A.

Theorem 1. *Let $\{F_j : j \in \mathcal{J}\}$ be an* ARDTFP$\left(h, \Pi^{\mathcal{J},\theta}, \mathcal{A}^{c,\kappa,\rho}\right)$, *where $h(\cdot) = \exp\{\cdot\}/(1 + \exp\{\cdot\})$ is the logistic transformation and the set of sequences of partitions $\Pi^{\mathcal{J},\theta}$*

is created, as before, based on the collection of centering probability distributions $\mathcal{F}^\theta = \{F_{\theta_j} : j \in \mathcal{J}\}$. *If the function* $\rho(\cdot)$ *is taken such that* $\sum_{l=1}^{\infty} \rho(l)^{-1/2} < \infty$, *then for any set of probability distributions* $\{G_j : j \in \mathcal{J}\}$ *such that* $G_j \in \mathcal{P}(\mathbb{R}^+)$ *and* $\mathrm{D}_{\mathrm{KL}}(G_j, F_{\theta_j}) < \infty$, $\forall j \in \mathcal{J}$, *with* $\mathrm{D}_{\mathrm{KL}}(G_j, F_{\theta_j})$ *being the Kullback–Leibler divergence between* G_j *and* F_{θ_j},

$$P\left\{\omega \in \Omega : \mathrm{D}_{\mathrm{KL}}(G_j, F(\omega, j)) < \delta, \forall j \in \mathcal{J}\right\} > 0,$$

for every $\delta > 0$.

Since the predictive distribution of the ARDTFP at each $j \in \mathcal{J}$ does not have a closed form, exact calculations cannot be performed with a fully specified ARDTFP. This justifies the use of a partially specified or finite version we consider in the applications, where the tailfree process is terminated at level M, which we denote $\{F_j : j \in \mathcal{J}\} \mid M, h, \Pi^{\mathcal{J},\theta}, \mathcal{A}^{c,\kappa,\rho} \sim \mathrm{ARDTFP}^M\left(h, \Pi^{\mathcal{J},\theta}, \mathcal{A}^{c,\kappa,\rho}\right)$. Following, Jara and Hanson (2011), we typically consider $M \approx \log_2(n/N)$, where n is the sample size and N is 5 to 10. As is usually assumed for partially specified Polya tree priors, on sets in the finest partition π_M^j the random F_j is assumed to follow the parametric distribution F_{θ_j} (see, e.g. Hanson 2006).

It may be difficult in practice to specify a single set of distributions with which to center the ARDTFP and, once specified, this set may affect inferences due to the partition dependence. One way to mitigate both problems is to specify a mixture of prior distributions. A mixture of ARDTFP is induced for $\{F_j : j \in \mathcal{J}\}$ by allowing parameters of the centering distributions θ and/or the precision parameter c to be random, that is,

$$\{F_j : j \in \mathcal{J}\} \mid M, h, \Pi^{\mathcal{J},\theta}, \mathcal{A}^{c,\kappa,\rho} \sim \int \mathrm{ARDTFP}^M\left(h, \Pi^{\mathcal{J},\theta}, \mathcal{A}^{c,\kappa,\rho}\right) dQ(\theta, c),$$

where Q refers to the joint prior for θ and c.

3 Posterior computation

In this section we present results useful for the posterior computation in the temporally–stratified PH model using a partially specified ARDTFP. Some notation is needed to simplify computation. Set $E_M = \{\varepsilon_1 \cdots \varepsilon_{l-1} 0 \in E^l : l = 1, \ldots, M\}$. For every $j \in \mathcal{J}$, set

$$\mathcal{Y}_j = \left\{\eta_{\epsilon 0}(j) \equiv \eta_{\epsilon 0}(j, \cdot) : \epsilon 0 \in E_M\right\},$$

and set $\mathcal{Y} = \{\mathcal{Y}_j : j \in \mathcal{J}\}$. Further, for every $j \in \mathcal{J}$, set

$$\begin{aligned} p_{\theta,M,\mathcal{Y},h}[k, j] = F_{0j}(B^j_{\varepsilon_1 \cdots \varepsilon_M}) &= \prod_{l=1}^{M} Y_{\varepsilon_1 \cdots \varepsilon_l}(j, \cdot), \\ &= \prod_{l=1}^{M} \left[h\left(\eta_{\varepsilon_1 \cdots \varepsilon_{l-1} 0}(j)\right)\right]^{(1-\varepsilon_l)} \left[1 - h\left(\eta_{\varepsilon_1 \cdots \varepsilon_{l-1} 0}(j)\right)\right]^{\varepsilon_l}, \end{aligned}$$

where $\varepsilon_1 \cdots \varepsilon_M$ is the binary representation of $k-1$, $k \in \{1, \ldots, 2^M\}$. Finally, set $\nu_{M,\theta,j}(t) = \text{Int}\left(2^M F_{\theta_j}(t) + 1\right)$.

From expressions (1), (2) and (3), it follows that, under the assumption of non–informative censoring, the conditional likelihood function is given by

$$\begin{aligned} \mathcal{L}(\beta, \gamma, \mathcal{Y}, \theta) &= \prod_{i=1}^{I} \prod_{j=1}^{J} \prod_{k=1}^{n_{ij}} \mathcal{L}_{ijk}(\beta, \gamma_i, \mathcal{Y}_j, \theta_j), \\ &= \prod_{i=1}^{I} \prod_{j=1}^{J} \prod_{k=1}^{n_{ij}} \left[\left(\exp\{\mathbf{x}'_{ijk}\beta + \gamma_i\} \frac{f_{0j}(t_{ijk})}{1 - F_{0j}(t_{ijk})} \right)^{\delta_{ijk}} \times \right. \\ &\quad \left. \{1 - F_{0j}(t_{ijk})\}^{\exp\{(\mathbf{x}'_{ijk}\beta + \gamma_i)\}} \right], \end{aligned} \tag{4}$$

where $\delta_{ijk} = 0$ if the time of the event of interest for the kth subject, diagnosed in the jth year and in the ith county, is right-censored and $\delta_{ijk} = 1$ otherwise, and

$$\begin{aligned} f_{0j}(t) &= 2^M f_{\theta_j}(t) \left\{ \prod_{l=1}^{M} Y_{\varepsilon_1 \cdots \varepsilon_l}(j, \cdot) \right\}, \\ &= 2^M f_{\theta_j}(t) \left\{ \prod_{l=1}^{M} \left[h\left(\eta_{\varepsilon_1 \cdots \varepsilon_{l-1} 0}(j)\right) \right]^{(1-\varepsilon_l)} \left[1 - h\left(\eta_{\varepsilon_1 \cdots \varepsilon_{l-1} 0}(j)\right) \right]^{\varepsilon_l} \right\}, \\ &= 2^M f_{\theta_j}(t) \left\{ p_{\theta,M,\mathcal{Y},h}\left[\nu_{M,\theta,j}(t), j\right] \right\}, \end{aligned}$$

and

$$F_{0j}(t) = \left\{ \sum_{l=1}^{\nu_{M,\theta,j}(t)-1} p_{\theta,M,\mathcal{Y},h}[l, j] \right\} + p_{\theta,M,\mathcal{Y},h}\left[\nu_{M,\theta,j}(t), j\right] \left\{ 2^M F_{\theta_j}(t) - \nu_{M,\theta,j}(t) + 1 \right\}.$$

To explore the posterior distribution of $\beta, \gamma, \lambda, \mathcal{Y}, c, \kappa, \theta$, a Metropolis within Gibbs sampling approach (Tierney 1994) is considered, based on sampling from the full conditional distributions $p(\beta \mid \gamma, \lambda, \mathcal{Y}, c, \kappa, \theta, \mathcal{D})$, $p(\gamma \mid \lambda, \beta, \mathcal{Y}, c, \kappa, \theta, \mathcal{D})$, $p(\lambda \mid \gamma, \beta, \mathcal{Y}, c, \kappa, \theta, \mathcal{D})$, $p(\mathcal{Y} \mid \gamma, \lambda, \beta, c, \kappa, \theta, \mathcal{D})$, $p(\kappa \mid \gamma, \lambda, \beta, \mathcal{Y}, c, \theta, \mathcal{D})$, $p(c \mid \gamma, \lambda, \beta, \mathcal{Y}, \kappa, \theta, \mathcal{D})$ and $p(\theta \mid \beta, \gamma, \mathcal{Y}, c, \kappa, \mathcal{D})$, where $\mathcal{D} = \{(t_{ijk}, \delta_{ijk}) : i = 1, \ldots, I, j = 1, \ldots, J, k = 1, \ldots, n_{ij}\}$. In the rest of this section, we provide some detail on the distinctive full conditional distributions and the implementation of the MCMC algorithm. A description of the omitted full conditionals and algorithms is given in Appendix B.

3.1 Updating the autoregressive process parameters

From the properties of autoregressive processes, it follows that the conditional prior for $\eta_{\epsilon 0}(j) \mid \{\eta_{\epsilon 0}(k) : k \neq j\}, c, \kappa_{\epsilon 0}$, $\epsilon 0 \in E_M$, is a normal distribution with mean and variance depending on whether j is an extreme or interior point of the index set $\mathcal{J}$,

$$\eta_{\epsilon 0}(j) \mid \{\eta_{\epsilon 0}(k) : k \neq j\}, c, \kappa_{\epsilon 0} \sim N(\tilde{\mu}_j, \tilde{\sigma}_j^2), \tag{5}$$

where

$$\tilde{\mu}_j = \begin{cases} \kappa_{\epsilon 0}\eta_{\epsilon 0}(j+1) & \text{if } j = 1, \\ \frac{\kappa_{\epsilon 0}}{1-\kappa_{\epsilon 0}^2}\left[\eta_{\epsilon 0}(j-1) + \eta_{\epsilon 0}(j+1)\right] & \text{if } j \neq 1 \text{ and } j \neq J, \\ \kappa_{\epsilon 0}\eta_{\epsilon 0}(j-1) & \text{if } j = J, \end{cases}$$

and

$$\tilde{\sigma}_j^2 = \begin{cases} \frac{2}{c\rho(m_{\epsilon 0})}(1-\kappa_{\epsilon 0}^2) & \text{if } j = 1 \text{ or } j = J, \\ \frac{2}{c\rho(m_{\epsilon 0})}\left[\frac{1-\kappa_{\epsilon 0}^2}{1+\kappa_{\epsilon 0}^2}\right] & \text{if } j \neq 1 \text{ and } j \neq J. \end{cases}$$

Therefore, by the conditional independence assumption and Bayes' rule, the full conditional distribution for $\eta_{\epsilon 0}(j)$ only depends on data at time j, $j \in \mathcal{J}$, and is given by

$$p\left(\eta_{\epsilon 0}(j) \mid \cdots\right) \propto \prod_{i=1}^{I}\prod_{k=1}^{n_{ij}} \mathcal{L}_{ijk}\left(\beta, \gamma_i, \mathcal{Y}_j, \theta_j\right) \phi\left(\eta_{\epsilon 0}(j) \mid \tilde{\mu}_j, \tilde{\sigma}_j^2\right),$$

where $\phi(\cdot \mid \mu, \sigma^2)$ stands for the density of a normal distribution with mean μ and variance σ^2, respectively. Since the full conditional for $\eta_{\epsilon 0}(j)$ is not recognizable, a Metropolis–Hastings (MH) step (Tierney 1994) or slice sampling (Neal 2003) may be used to update $\eta_{\epsilon 0}(j)$. In our implementation, a random walk MH step is performed, where the candidates are generated from a normal distribution $\eta_{\epsilon 0}^*(j) \sim N(\eta_{\epsilon 0}(j), s^2)$, with s^2 being an appropriate positive constant. The candidate is accepted with probability given by

$$\min\left\{1, \frac{\prod_{i=1}^{I}\prod_{k=1}^{n_{ij}} \mathcal{L}_{ijk}\left(\beta, \gamma_i, \mathcal{Y}_j^*, \theta_j\right) \phi\left(\eta_{\epsilon 0}^*(j) \mid \tilde{\mu}_j, \tilde{\sigma}_j^2\right)}{\prod_{i=1}^{I}\prod_{k=1}^{n_{ij}} \mathcal{L}_{ijk}\left(\beta, \gamma_i, \mathcal{Y}_j, \theta_j\right) \phi\left(\eta_{\epsilon 0}(j) \mid \tilde{\mu}_j, \tilde{\sigma}_j^2\right)}\right\},$$

where $\mathcal{Y}_j^*$ is equal to $\mathcal{Y}_j$ but with $\eta_{\epsilon 0}(j)$ replaced by $\eta_{\epsilon 0}^*(j)$. In our experience, $s = 0.5$ is a good starting point yielding acceptance rates ranging from 40% to 60%.

3.2 Updating the autocorrelation parameters

From the definition of the ARDTFP, it follows that the joint distribution for $\eta_{\epsilon 0} = \{\eta_{\epsilon 0}(j) : j \in \mathcal{J}\}$, given $\kappa_{\epsilon 0}$, c and ρ, is the following multivariate normal distribution

$$p\left(\eta_{\epsilon 0} \mid \kappa_{\epsilon 0}, c, \rho\right) = (2\pi)^{-J/2}\left|\mathbf{C}_{\epsilon 0}\right|^{-1/2} \exp\left\{-\frac{1}{2}\eta_{\epsilon 0}'\mathbf{C}_{\epsilon 0}^{-1}\eta_{\epsilon 0}\right\}, \tag{6}$$

where $\mathbf{C}_{\epsilon 0}^{-1}$ is a tridiagonal matrix given by

$$\mathbf{C}_{\epsilon 0}^{-1} = \frac{c\rho(m_{\epsilon 0})}{2}\begin{pmatrix} d_{1,\epsilon 0} & -b_{\epsilon 0} & & & & & \\ -b_{\epsilon 0} & d_{2,\epsilon 0} & -b_{\epsilon 0} & & & & \\ & -b_{\epsilon 0} & d_{2,\epsilon 0} & -b_{\epsilon 0} & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & -b_{\epsilon 0} & d_{2,\epsilon 0} & -b_{\epsilon 0} & \\ & & & & -b_{\epsilon 0} & d_{1,\epsilon 0} \end{pmatrix},$$

with $d_{1,\epsilon 0} = 1 + \frac{\kappa_{\epsilon 0}^2}{(1-\kappa_{\epsilon 0}^2)}$, $b_{\epsilon 0} = \frac{\kappa_{\epsilon 0}}{(1-\kappa_{\epsilon 0}^2)}$, and $d_{2,\epsilon 0} = 1 + 2\frac{\kappa_{\epsilon 0}^2}{(1-\kappa_{\epsilon 0}^2)}$, respectively, and

$$|\mathbf{C}_{\epsilon 0}| = \left(\frac{2}{c\rho(m_{\epsilon 0})}\right)^J \left(1-\kappa_{\epsilon 0}^2\right)^{J-1}.$$

Notice that the tridiagonal form of the inverse of $\mathbf{C}_{\epsilon 0}$ implies that the quadratic form in expression (6) can be written as follows

$$\eta_{\epsilon 0}' \mathbf{C}_{\epsilon 0}^{-1} \eta_{\epsilon 0} = \frac{c\rho(m_{\epsilon 0})}{2\left(1-\kappa_{\epsilon 0}^2\right)} \left\{ \sum_{j=1}^{J} \left[\eta_{\epsilon 0}(j) - \kappa_{\epsilon 0}\eta_{\epsilon 0}(j-1)\right]^2 \right\},$$

where the convention $\eta_{\epsilon 0}(0) = 0$ is used. It follows that if independent uniform priors are placed on $\kappa_{\epsilon 0}$, $\epsilon 0 \in E_M$, $\kappa_{\epsilon 0} \sim U(-1,1)$, then its full conditional distribution is given by

$$p\left(\kappa_{\epsilon 0} \mid \cdots\right) \propto \left(1-\kappa_{\epsilon 0}^2\right)^{-(J-1)/2} \exp\left\{ -\frac{c\rho(m_{\epsilon 0})}{(1-\kappa_{\epsilon 0}^2)} \left(\sum_{j=1}^{J} \left[\eta_{\epsilon 0}(j) - \kappa_{\epsilon 0}\eta_{\epsilon 0}(j-1)\right]^2 \right) \right\}. \quad (7)$$

Similarly, if independent normal $N\left(\mu_{\kappa_{\epsilon 0}}, \sigma_{\kappa_{\epsilon 0}}^2\right)$ priors are placed on the Fisher transformation of $\kappa_{\epsilon 0}$, $\epsilon 0 \in E_M$, $(1/2)\log\left(\frac{1+\kappa_{\epsilon 0}}{1-\kappa_{\epsilon 0}}\right)$, then its full conditional distribution is given by

$$\begin{aligned} p\left(\kappa_{\epsilon 0} \mid \cdots\right) \quad \propto \quad & \frac{\left(1-\kappa_{\epsilon 0}^2\right)^{-(J-1)/2-1}}{\left(1+\kappa_{\epsilon 0}^2\right)} \times \\ & \exp\left\{ -\frac{c\rho(m_{\epsilon 0})}{(1-\kappa_{\epsilon 0}^2)} \left(\sum_{j=1}^{J} \left[\eta_{\epsilon 0}(j) - \kappa_{\epsilon 0}\eta_{\epsilon 0}(j-1)\right]^2 \right) \right\} \times \\ & \exp\left\{ -\frac{\left[\log\left(\frac{1+\kappa_{\epsilon 0}}{1-\kappa_{\epsilon 0}}\right) - 2\mu_{\kappa_{\epsilon 0}}\right]^2}{8\sigma_{\kappa_{\epsilon 0}}^2} \right\}. \end{aligned} \quad (8)$$

If in the ARDTFP specification it is assumed that $\kappa_{\epsilon 0} = \kappa$, for every $\epsilon 0 \in E_M$, the previous expressions (7) and (8) are replaced by

$$\begin{aligned} p\left(\kappa \mid \cdots\right) \quad \propto \quad & \left(1-\kappa^2\right)^{-\frac{(J-1)\left(2^M-1\right)}{2}} \times \\ & \exp\left\{ -\sum_{\epsilon 0 \in E_M} \left[\frac{c\rho(m_{\epsilon 0})}{(1-\kappa^2)} \left(\sum_{j=1}^{J} \left(\eta_{\epsilon 0}(j) - \kappa\eta_{\epsilon 0}(j-1)\right)^2 \right) \right] \right\}, \end{aligned} \quad (9)$$

and

$$p(\kappa \mid \cdots) \propto \left(1-\kappa^2\right)^{-\frac{(J-1)\left(2^M-1\right)}{2}-1}\left(1+\kappa^2\right)^{-1} \times \exp\left\{-\sum_{\epsilon 0 \in E_M}\left[\frac{c\rho(m_{\epsilon 0})}{(1-\kappa^2)}\left(\sum_{j=1}^{J}\left(\eta_{\epsilon 0}(j)-\kappa\eta_{\epsilon 0}(j-1)\right)^2\right)\right]\right\} \times \exp\left\{-\frac{\left(\log\left(\frac{1+\kappa}{1-\kappa}\right)-2\mu_\kappa\right)^2}{8\sigma_\kappa^2}\right\}, \tag{10}$$

respectively. Since expressions (7), (8), (9) and (10) are not recognizable, an MH step or slice sampling may be used to update the autocorrelation parameters. We have found that a Gaussian random walk on the scale of the Fisher transformation of $\kappa_{\epsilon 0}$, using a standard deviation of 0.5, yields an adequate mixing of the chain in several applications of the model. Notice that if the median function of the process is fixed to be the median of the centering distributions, τ_0 is not updated, and expressions (9) and (10) still apply but with $E_M = \{\varepsilon_1 \cdots \varepsilon_{l-1} 0 \in E^l : l = 2, \ldots, M\}$. Notice also that similar expressions arise when a common autocorrelation parameter is assumed for all the Gaussian processes assigned to the same level of the partition $\kappa_{\epsilon 0} = \kappa_{m_{\epsilon 0}}$, where, as before, $m_{\epsilon 0}$ denotes the length of the string $\epsilon 0$.

3.3 Updating the centering parameters

Different specifications of the ARDTFP regarding the centering distributions can be considered. In the more general model used in the applications, an unstructured collection of log–logistic distributions is considered, such that $\mathcal{F}^\theta = \{F_{\theta_j} : j \in \mathcal{J}\}$, where the density of F_{θ_j} is given by

$$f_{\theta_j}(t) = \exp\{\alpha_j + \vartheta_j\} t^{\exp\{\alpha_j\}-1}\left[1+\exp\{\vartheta_j\} t^{\exp\{\alpha_j\}}\right]^{-2}, \tag{11}$$

with $\theta_j = (\alpha_j, \vartheta_j) \in I\!R^2$. If, for instance, independent normal priors are placed on α_j and ϑ_j, $\alpha_j \mid \mu_\alpha, \sigma_\alpha^2 \overset{iid}{\sim} N(\mu_\alpha, \sigma_\alpha^2)$ and $\vartheta_j \mid \mu_\vartheta, \sigma_\vartheta^2 \overset{iid}{\sim} N(\mu_\vartheta, \sigma_\vartheta^2)$, then the full conditional distribution is not recognizable and is given by

$$p(\alpha_j, \vartheta_j \mid \cdots) \propto \left\{\prod_{i=1}^{I}\prod_{k=1}^{n_{ij}} \mathcal{L}_{ijk}(\beta, \gamma_i, \mathcal{Y}_j, \theta_j)\right\} \phi\left(\alpha_j \mid \mu_\alpha, \sigma_\alpha^2\right) \phi\left(\vartheta_j \mid \mu_\vartheta, \sigma_\vartheta^2\right). \tag{12}$$

A similar expression arises if a common centering distribution is assumed, that is, if $\theta_j = \theta$, for every $j \in \mathcal{J}$. In the latter case, all observations are involved in the likelihood contribution to the full conditionals. Random walk MH steps were used to update the centering parameters using appropriate normal candidate generating distributions. As the centering parameters (α_j, ϑ_j) are typically highly correlated, we suggest a bivariate Gaussian random-walk update with initial covariance matrix obtained from a parametric log–logistic fit of the data.

3.4 Updating the precision parameter

If a gamma prior is placed on the precision parameter of the ARDTFP, $c \mid \tau_1, \tau_2 \sim \Gamma(\tau_1/2, \tau_2/2)$, its full conditional distribution has a closed form given by

$$c \mid \cdots \sim \Gamma\left(\frac{J\left(2^M - a\right) + \tau_1}{2}, \frac{\Delta + \tau_2}{2}\right), \tag{13}$$

where

$$\Delta = \sum_{\epsilon 0 \in E_M} \left\{ \frac{\rho(m_{\epsilon 0})}{(1 - \kappa_{\epsilon 0}^2)} \left(\sum_{j=1}^{J} \left[\eta_{\epsilon 0}(j) - \kappa_{\epsilon 0}\eta_{\epsilon 0}(j-1)\right]^2 \right) \right\},$$

and $a = 1$ for the regular model and $a = 2$ for the ARDTFP where the median is fixed to the median of the centering distributions. In the latter case, $E_M = \{\varepsilon_1 \cdots \varepsilon_{l-1} 0 \in E^l : l = 2, \ldots, M\}$.

4 Illustrations

We consider simulated and real–life examples that emphasize the application of the semi-parametric temporally–stratified PH model. We use simulated datasets to illustrate the performance of the proposed approach. We consider data from nonstandard and evolving baseline distributions. Through the real–life data analyses, we emphasize different aspects of the inferential problem and show the different particular cases that arise from our proposal. The models considered in the next sections were fitted using compiled FORTRAN 90 functions, which are available upon request to the authors. Functions implementing the ARDTFP will be incorporated into the library "DPpackage" (Jara et al. 2011) of the R program (R Development Core Team 2011).

4.1 Analysis of simulated data

In order to illustrate the behavior of the ARDTFP, we consider simulated data for one county $\mathcal{I} = \{1\}$ over ten years $\mathcal{J} = \{1, \ldots, 10\}$. Each baseline survival function was assumed to be a mixture of two normals with weights changing linearly in time. Specifically, we assume that $S_{0j}(t) = 1 - (j/11)\Phi(t|5, 1) - (1 - j/11)\Phi(t|10, 1)$, where Φ stands for the CDF of a standard normal distribution. We generated $n_{1j} = 100$ observations for each time point, using quantiles of equally–spaced probability, that is,

$$t_{1jk} = F_{0j}^{-1}\left(k/(n_{1j} + 1)\right),$$

where $j \in \mathcal{J}$ and $k = 1, \ldots, n_{1j}$. The data thus approximate their expected order statistics and are a convenient "perfectly representative" sample.

We fitted the ARDTFP model, by assuming $\kappa_{\epsilon 0} = \kappa$, for every $\epsilon 0 \in E^*$, and place an $N(0, 3)$ prior on the Fisher transformation of κ. We further considered a common centering distribution, not indexed by time, that is, $F_{\theta_j} \equiv F_\theta$, for every $j \in \mathcal{J}$. The

common centering distribution was the log–logistic distribution with density given by expression (11). We further assume $h(\cdot) = \exp\{\cdot\}/(1+\exp\{\cdot\})$, $\rho(l) = l^2$, $c \sim \Gamma(10,1)$, $\alpha \sim N\left(0, 10^3\right)$ and $\vartheta \sim N\left(0, 10^3\right)$. We also considered an improper prior for the centering parameters $p(\alpha, \vartheta) \propto 1$ and obtained similar results to the ones reported here. We fitted two versions of the ARDTFP by considering $M = 4$ and $M = 5$.

In order to illustrate the advantage of the model that borrows strength across sub–populations, we also fitted a purely stratified model by assuming independent tailfree priors for each of the baseline distributions, that is, an ARDTFP with $\kappa_{\epsilon 0} = 0$, for every $\epsilon 0 \in E^*$. Each model was fitted using the MCMC updating scheme described in Section 3. In each case, a burn–in period of 2,000 was followed by a run of 100,000 iterations, on which posterior inference was based. Time series plots of the stored output and standard tests (not shown), as implemented in the BOA R library (Smith 2007), suggested convergence of the chains. The models were compared for predictive ability using the so called log pseudo marginal likelihood (LPML) developed by Geisser and Eddy (1979). In the context of the temporally–stratified PH model, the LPML for model M is defined as $\text{LPML}_M = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{n_{ij}} p_M\left(\mathcal{D}_{ijk} \mid \mathcal{D}^{-(i,j,k)}\right)$, where $\mathcal{D}^{-(i,j,k)}$ is the data matrix that results after excluding the kth observation of the jth year in the ith county, $\mathcal{D}_{ijk}$, and $p_M\left(\mathcal{D}_{ijk} \mid \mathcal{D}^{-(i,j,k)}\right)$ is the predictive density based on all of the data except the ijkth observation, evaluated at the observed $\mathcal{D}_{ijk}$, referred to as conditional predictive ordinate (CPO). The simple method suggested by Gelfand and Dey (1994) to obtain estimates of the CPO statistics from MCMC output was used here.

Table 1 shows the LPML statistics for the the two versions of the ARDTFP model and for the purely stratified versions of the model. The results show that the number of levels M does not have an important effect in terms of prediction. In all cases, the ARDTFP model does significantly better than the purely stratified model. Borrowing strength across these slowly–changing densities improves prediction markedly. The ARDTFP does a good job in estimating these evolving densities, especially considering that only 100 observations from each stratum were considered. The posterior mean closely approximated the true density of each stratum, which was entirely enclosed in 95% pointwise credible intervals. The posterior means and 95% pointwise credible intervals under the partially specified ARDFP with $M = 5$, are shown in Figure 1. In all cases, the posterior mean (dashed line) is almost indistinguishable from the true density (solid line).

4.2 Analysis of the Iowa SEER data

The data

The Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute may be the most authoritative source of information on cancer incidence and survival in the United States. SEER provides county–level cancer data on an annual basis for public use. We apply our proposed model to the analysis of 1989–1998 SEER Iowa breast cancer survival data. The data include information on a cohort of 20,315

	Independent tailfree		ARDTFP	
Strata	$M = 4$	$M = 5$	$M = 4$	$M = 5$
1	−182.1	−183.9	−175.7	−173.9
2	−199.6	−201.6	−191.3	−189.6
3	−210.7	−213.0	−200.3	−200.6
4	−217.8	−220.2	−207.7	−207.6
5	−220.5	−223.2	−210.2	−210.4
6	−220.3	−223.0	−209.9	−210.0
7	−216.8	−219.5	−206.6	−206.3
8	−208.7	−211.2	−199.7	−199.7
9	−196.6	−198.9	−189.0	−188.8
10	−178.7	−180.3	−172.8	−173.0
Global	−2051.9	−2074.7	−1963.1	−1959.9

Table 1: Simulated data: Log pseudo marginal likelihood (LPML) values for the ARDTFP and independent tailfree models. The LPML statistic is given for $M = 4$ and $M = 5$, and for each strata and the complete data in each case.

women in Iowa, who were diagnosed with malignant breast cancer starting in 1989 and ending in 1998, with follow–up continued through the end of 2003. Only deaths due to metastasis of cancerous nodes in the breast were considered to be events, while the rest (including death from metastasis of other types of cancer, or from other causes) were considered to be censored observations. By the end of 2003, 8,396 of the patients had died of breast cancer, while the remaining 11,919 women were censored, either because they survived till the end of the study period, died of other causes, or were lost to follow–up.

For each individual, the survival time in months (1 to 180) and county of residence at diagnosis is recorded. Several individual–level covariates were also available, including race (white or black), age at diagnosis, number of primaries (physiologically independent cancers diagnosed), and the stage of the disease: local (confined to the breast), regional (spread beyond the breast tissue), or distant (metastasis). We created a dummy variable "cpri", which equals 1 for women who have more than one independent cancer diagnosed, and equals 0 for those who have breast cancer only. We treated "local" as baseline, and created two dummy variables for "regional" and "distant," respectively.

Table 2 shows the summary statistics for follow–up time and covariates across the 10 years. The diagnosis age, number of primaries, and two stage dummies are roughly balanced across years. The follow–up time decreases and the proportion of censoring increases for the recent years' cohorts. Table 2 also shows the year–specific 25th percentiles of survival time in months, which indicate the estimated time in months past which 75% will survive for each year. The 25th survival percentiles increase across years, from 52 months in 1989 to 69 months in 1998.

Variable	Level	Statistic	Year									
			1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
Status	Event	Count	1135	1032	988	976	889	811	755	688	574	547
		%	60.2	53.6	50.1	47.3	44.0	40.4	37.1	34.0	27.2	24.1
	Censored	Count	752	895	983	1087	1131	1198	1282	1334	1535	1722
		%	39.9	46.5	49.9	52.7	56.0	59.6	62.9	66.0	72.8	75.9
Number of primaries	1	Count	1353	1415	1465	1528	1486	1482	1503	1519	1624	1723
		%	71.7	73.4	74.3	74.1	73.6	73.8	73.8	75.1	77.0	75.9
	>1	Count	534	512	506	535	534	527	534	503	485	546
		%	28.3	26.6	25.7	25.9	26.4	26.2	26.2	24.9	23.0	24.1
Stage	Local	Count	1148	1195	1274	1381	1378	1332	1373	1375	1438	1509
		%	60.8	62.0	64.6	66.9	68.2	66.3	67.4	68.0	68.2	66.5
	Regional	Count	612	630	571	565	545	556	546	513	539	621
		%	32.4	32.7	29.0	27.4	27.0	27.7	26.8	25.4	25.6	27.4
	Distant	Count	127	102	126	117	97	121	118	134	132	139
		%	6.7	5.3	6.4	5.7	4.8	6.0	5.8	6.6	6.3	6.1
Follow-up time (months)		Mean	112.6	111.8	106.7	101.0	94.6	89.0	81.0	72.4	66.0	57.2
		St. Dev.	61.6	57.4	52.2	47.2	41.8	36.2	32.0	27.7	21.9	17.6
Age (years)		Mean	65.8	64.6	65.4	65.0	64.7	64.8	65.1	65.0	64.7	64.7
		St. Dev.	14.4	15.2	14.6	14.6	14.7	14.5	14.6	14.9	14.5	14.4
25th survival percentiles			52	56	59	59	60	64	62	58	70	69

Table 2: SEER data: Summary statistics for follow-up time and covariates by year.

The models

Several versions of the temporally–stratified PH model given by expressions (1), (2) and (3) were considered for the analysis of the Iowa SEER data. Because ultimately we analyze lattice data (counties in Iowa), each model was augmented with frailty terms, spatial or exchangeable, and assuming $\lambda \sim \Gamma(0.1, 0.1)$. Models considered cover the complete range of dependence in the baseline survival functions. On one extreme, in Model Class 0, the temporal cohort effect is ignored by assuming a common baseline survival function $S_{0j} \equiv S_0$, $\forall j \in \mathcal{J}$ and a single tailfree prior on S_0. On the other extreme, in Model Class 1, a completely stratified model was considered by assuming independent tailfree priors for the baseline survival function of each strata S_{0j}, $j \in \mathcal{J}$. In this case, independent centering parameters for each group $\theta_j = (\alpha_j, \vartheta_j)$ were assumed. Note that the latter model is equivalent to the ARDTFP with $\kappa_{\epsilon 0} = 0$, for every $\epsilon 0 \in E^*$.

Three intermediate ARDTFP priors were considered varying in the complexity of the centering family. Model Class 2 considers an ARDTFP with common log–logistic centering distribution. We assume the autocorrelation parameter to vary with the level of the partition by considering $\kappa_{\varepsilon_1 \cdots \varepsilon_{l-1} 0} = \kappa_l$, $l = 1, \ldots, M$, and further assume $\kappa_1, \ldots, \kappa_M \overset{iid}{\sim} U(-1, 1)$. Model Class 3 considers an ARDTFP where the log–logistic centering distributions have different and independent parameters, that is, $\{F^{\theta_j} : j \in \mathcal{J}\}$, where the θ_j's are independent vectors. Finally, Model Class 4 considers an ARDTFP where the log–logistic centering distributions have different but dependent parameters following an autoregressive process, that is, $\{F_{\theta_j} : j \in \mathcal{J}\}$, $\theta_j = (\alpha_j, \vartheta_j)$, where $\{\alpha_j : j \in \mathcal{J}\}$ and $\{\vartheta_j : j \in \mathcal{J}\}$ are independent autoregressive processes with covariance function given by $\tau_\alpha^{-1} \kappa_\alpha^{|j-j'|}$ and $\tau_\vartheta^{-1} \kappa_\vartheta^{|j-j'|}$, respectively. We extended the

model by assuming $\tau_\alpha, \tau_\vartheta \overset{iid}{\sim} \Gamma(3,1)$, $\tau_\alpha, \tau_\vartheta \overset{iid}{\sim} \Gamma(0.1,0.1)$ or $\tau_\alpha, \tau_\vartheta \overset{iid}{\sim} \Gamma(0.001,0.001)$, and $\kappa_\alpha, \kappa_\vartheta \overset{iid}{\sim} U(-1,1)$.

For all models, we fixed the median of the process to coincide with the median of the set of centering probability distributions. In all cases, we assume $M = 4$, $h(\cdot) = \exp\{\cdot\}/(1+\exp\{\cdot\})$, $\rho(l) = l^2$ and $c \sim \Gamma(5,1)$. We considered both the improper prior $p(\alpha_1,\ldots,\alpha_J,\vartheta_1,\ldots,\vartheta_J) \propto 1$, as well as proper but vague prior $(\alpha_j,\vartheta_j) \overset{iid}{\sim} N_2\left(\mathbf{0}_2, 10^3\mathbf{I}_2\right)$ for the centering parameters. Independent normal $N(0,10^3)$ priors were assumed for the regression coefficients, as well as an improper prior $p(\beta) \propto 1$. The models were fitted using the corresponding variant of the algorithm described in Section 3. Despite the high dimension of the models, the MCMC chains mixed reasonably well. For each model, 40,000 iterations were retained for posterior estimation following a burn–in of 20,000 iterations. Models were compared using the LPML and the deviance information criterion (DIC), as presented by Spiegelhalter et al. (2002).

The results

The results were robust to the prior specification across models. Table 3 shows the DIC and LPML for the models under consideration. Both criteria indicate that the Model Class 0 has the poorest model fit. Thus, introducing temporally varying baseline survival functions improves goodness of fit and prediction. Furthermore, DIC and LPML favor the temporally dependent structure the most (Model Classes 2 – 4). The LPML differs on whether temporally evolving (Model Class 4) versus exchangeable (Model Class 3) underlying log–logistic parameters is beneficial; either is better than one overall log–logistic baseline (Model Class 0 and 2) or stratified types of analyses assuming no temporal dependence (Model Class 1). Both DIC and LPML indicate the same

Model Class	Frailty distribution	DIC	LPML
0 (Regular TFP, $F_{0j} \equiv F_0 \sim \text{TFP}, j \in \mathcal{J}$)	CAR	101259.5	–50637.7
	i.i.d.	101286.4	–50653.9
1 (Independent TFP, $F_{0j} \overset{ind.}{\sim} \text{TFP}, j \in \mathcal{J}$)	CAR	101087.7	–50556.7
	i.i.d.	101062.3	–50542.3
2 (ARDTFP with common centering)	CAR	101318.0	–50668.6
($F_j^\theta = F^\theta, j \in \mathcal{J}$)	*i.i.d.*	101290.2	–50651.9
3 (ARDTFP with unstructured centering)	CAR	101074.5	–50546.5
	i.i.d.	101053.0	–50534.6
4 (ARDTFP with AR centering)	CAR	101083.3	–50561.2
	i.i.d.	101062.1	–50551.1

Table 3: SEER data: Deviance information criteria (DIC) and log-pseudo marginal likelihood (LPML) for the models under consideration.

trends for goodness of fit, with the models with *iid* log–logistic parameters and autoregressive conditional probability parameters (Model Class 3) outperforming the models

with *iid* log–logistic parameters and *iid* conditional probability parameters (Model Class 1), which outperform the models with autoregressive log–logistic parameters and autoregressive conditional probability parameters (Model Class 4), which outperform the models with common log–logistic parameters and autoregressive conditional probability parameters (Model Class 2). Allowing *iid* log–logistic parameters on the baseline survival functions across years seems important to model fit. Furthermore, temporally dependent conditional probability parameters contribute more to goodness of fit, followed by temporally independent centering parameters.

Within each model, LPML and DIC indicate that exchangeable frailty terms provide better prediction than CAR frailties. The frailty terms offer a contribution to improve model prediction and fit, while the spatial arrangement of these frailty terms seems not helpful in improving model fit over the exchangeable pattern. This may be because the variance among data is mainly captured by temporal cohort effects; after adjusting for the temporal effect captured in temporally evolving baseline survival functions, there is little spatial story left. Therefore, although the spatial pattern is important in improving goodness of fit in the simple frailty models (Model Class 0), it is not helpful after allowing temporally different baseline survival (Model Classes 1–4) for this dataset. Overall, the *iid* frailty model with *iid* log-logistic parameters and autoregressive conditional probability parameters (the *iid* frailty model in Model Class 3) has the best model fit according to LPML and DIC.

Since $U(-1,1)$ priors were considered for all autocorrelation parameters, $(\kappa_\alpha, \kappa_\eta)$ and $(\kappa_2, \ldots, \kappa_J)$, we may formally test Model Class 3 versus Model Class 1 via the hypothesis $H_0 : \kappa_2 = \cdots = \kappa_J = 0$. Although the 95% credible intervals of κ_3 and κ_4 from Model 3 are centered at 0, the 95% credible interval for κ_2 is $(0.61, 0.95)$ indicating a significant difference from 0. The second level of the tailfree divides $\mathbb{R}^+$ into four pieces, and significant κ_2 indicates more gross, large-scale *positive* dependence in baseline survival than κ_l, $l > 2$. These data require the temporal evolution allowed by Model Class 3 over the purely independent stratification of Model Class 1. When autocorrelation structure is introduced to the log–logistic parameters along with the tailfree conditional probability parameters (Model Class 4), all 95% credible intervals of $\{\kappa_\alpha, \kappa_\eta, \kappa_2, \cdots, \kappa_M\}$ cover 0. Therefore we can neither reject $H_0 : \kappa_\alpha = \kappa_\eta = 0$ (Model Class 4 can be reduced to Model Class 3), nor reject $H_0 : \kappa_\alpha = \kappa_\eta = \kappa_2 = \cdots = \kappa_M = 0$ (Model 4 can be reduced to Model 1). Consistently, LPML favors Model Class 3 over Model Class 1, which outperforms Model Class 4.

Table 4 compares posterior medians and equal–tailed 95% credible intervals for main effects (components of β) under Model 3 assuming exchangeable frailties to those obtained under the standard semiparametric partial likelihood–based PH model with *iid* frailties. The standard results were obtained using the `survival` package in `R` (R Development Core Team 2011). As is often the case with main effects (which are typically well–identified), the estimates change little across models with the possible exception of the "centered age" variable. The best fit ARDTFP model indicates that all of the predictors are significant at the 0.05 level. Higher age at diagnosis increases the hazard; e.g., a twenty-year increase in age is associated with an $e^{0.0444\times 20} \approx 2.43$–fold increase in hazard rate. After adjusting for the age at diagnosis, disease stage and the county

of residence, the hazard of dying from breast cancer before any time t is $e^{0.19} \approx 1.21$ times the hazard for women having more than one independent cancer diagnosed versus women having only breast cancer. Using women with local stage of disease as the reference, the hazard of women of the same age who live in the same county and have the same number of independent cancers diagnosed will be $e^{0.68} \approx 1.97$ times the hazard if their cancer is detected at the regional stage, and $e^{2.13} \approx 8.41$ times the hazard if detected at the distant stage. Figures 2–11 show fitted survival densities for women

Coefficient	ARDTFP PH	Partial likelihood PH
β_1 (centered age)	0.0444 (0.0440, 0.0449)	0.0478 (0.0460, 0.0495)
β_2 (number of primaries)	0.19 (0.14, 0.24)	0.20 (0.15, 0.25)
β_3 (regional stage)	0.68 (0.63, 0.73)	0.68 (0.63, 0.73)
β_4 (distant stage)	2.13 (2.06, 2.20)	2.15 (2.08, 2.22)

Table 4: SEER data: Posterior medians (95% equal–tailed credible intervals) for the regression coefficients in the ARDTFP PH Model Class 3. The point estimates (95% confidence intervals) under the partial likelihood–based PH model with *iid* frailties are also presented.

aged 65 years at study entry (the mean in our dataset), with one primary, for three disease stages under the exchangeable frailty version of Model Class 3 and assuming a 0 value for the spatial frailty term across 10 years. These fitted densities are overlaid on histograms of all observed survival times for study participants with entry ages 55 to 75. Since 11,919 of our 20,315 observations are censored, to incorporate both the censored and uncensored observations we take the Kaplan–Meier survival function estimates and convert them back to an approximate histogram (Huzurbazar 2005). In all these plots, the predicted density curve shapes mirror the data appropriately, even though the predictive densities correspond to a frailty of $\gamma = 0$, whereas the histograms are averaged over all of Iowa. This confirms that the proposed temporally–stratified PH model using an ARDTFP correctly captures the variability of the data, and thus should have good predictive ability.

Figures 12 and 13 provide the fitted predictive densities and survival curves for mean age and three disease stages. At each stage, the curves show similar patterns across the 10 years. Iowan breast cancer mortality is decreasing in time, with the darkest lines (year 1989-1992) having the lowest survival rates, and the two lightest lines (years 1997 and 1998) having the best survival rates. This also confirms the dependence of the DTFP parameters in our final model.

In simpler frailty–based models, Banerjee and Carlin (2003) note fairly level mortality until about 1990, when a rather steep decline is apparent (their Figure 4). They suggest that this coincides with the recent rise in the use of mammography by American women. This decline also coincides with Iowa's *Care for Yourself* program, a part of the larger federal government's National Breast and Cervical Cancer Early Detection Program, established in 1991. The model they fit is based on independent CAR frailties,

one for each year, entering into the linear predictor of the Cox model. This model forces the overall shape of the hazard curve to be similar across time. In Figure 12 we see that although the overall shape is similar across time, there are notable differences in modal location, which would not be captured by conventional approaches. Predictive density and/or hazard plots were not included in either Banerjee et al. (2003) or Banerjee and Carlin (2003).

5 Concluding remarks

We have proposed a method to improve the precision in the estimation of a temporally–stratified PH model with spatial frailties. The proposal borrows strength across temporal strata by considering a dependent tailfree process prior to model temporally evolving baseline survival functions. The proposed model includes at the extremes purely stratified and common baseline survival functions as special cases.

The choice of the AR process in the definition of the dependent model is primarily based on its computational simplicity. However, a theoretical result related to the large support of the proposed prior shows that the resulting model covers a wide range of dependence to capture temporal cohort effects, including a stochastic order of the baseline survival functions in time. If strong prior information regarding the behavior of the baseline survival functions is available, like the one of stochastic order, a more focussed probability model for related probability distributions could be considered (see, e.g., Gelfand and Kottas 2001; Karabatsos and Walker 2007). However, these approaches cannot be naturally extended to make predictions for future time points.

Although our experience with the proposed MCMC algorithm is satisfactory, the study of more efficient and automatic algorithms for the proposed model is of interest. Specifically, the calibration of the MH proposals can be avoided by using local (Haario et al. 2001; Atchadé and Rosenthal 2005; Haario et al. 2006) or global (Gasemyr 2003; Cai et al. 2008; Giordani and Kohn 2010; Keith et al. 2008; Holden et al. 2009) adaptive algorithms that update the proposal distribution "on the fly" as the chain evolves. The evaluation of these algorithms is the subject of ongoing research.

Beyond developing and implementing useful new statistical models, our primary goal was to find spatial–longitudinal structures that can best explain the Iowa SEER data, and best predict mortality of patients. Temporally dynamic baseline survival anticipates decreasing mortality trends due to early detection and prevention, e.g. possibly stemming from Iowa's *Care for Yourself* program. Note also that the proportion of primaries roughly increases across the decade (see, Table 2), as does the proportion of local stage diagnoses – both are indicators of earlier detection and diagnosis.

Finally, several extensions of the proposed model can be considered. The extension to handle multivariate outcomes is of interest. The extension of the model to handle both temporal and spatial stratification is also the subject of ongoing research. We are currently working on a version of the model that takes into account potential measurement error of the failure and/or censoring times.

References

Atchadé, Y. F. and Rosenthal, J. S. (2005). "On adaptive Markov chain Monte Carlo algorithms." *Bernoulli*, 11: 815–828. 166

Banerjee, S. and Carlin, B. P. (2003). "Semiparametric spatio-temporal frailty modeling." *Environmetrics*, 14: 523–535. 165, 166

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: 123–142. 150, 166

Barron, A. R. (1985). "The strong ergodic theorem fron densities: generalized Shannon-McMillan-Breiman theorem." *The Annals of Probability*, 13: 1292–1393. 173

Besag, J., York, J., and Mollié, A. (1991). "Bayesian image restoration, with two applications in spatial statistics." *Annals of the Institute of Statistical Mathematics*, 43: 1–20. 150

Cai, B., Meyer, R., and Perron, F. (2008). "Metropolis-Hastings algorithms with adaptive proposals." *Statistics and Computing*, 18: 421–433. 166

Cai, J., Fan, J., Jiang, J., and Zhou, H. (2007). "Partially linear hazard regression for multivariate survival data." *Journal of the American Statistical Association*, 102: 538–551. 148

Chung, Y. and Dunson, D. B. (2009). "Nonparametric Bayes conditional distribution modeling with variable selection." *Journal of the American Statistical Association*, 104: 1646–1660. 148

Cox, D. R. (1972). "Regression models and life-tables (with Discussion)." *Journal of the Royal Statistical Society, Series B*, 34: 187–220. 147

De Iorio, M., Johnson, W. O., Müller, P., and Rosner, G. L. (2009). "Bayesian nonparametric non-proportional hazards survival modelling." *Biometrics*, 65: 762–771. 147

De Iorio, M., Müller, P., Rosner, G. L., and MacEachern, S. N. (2004). "An ANOVA model for dependent random measures." *Journal of the American Statistical Association*, 99: 205–215. 148

Duan, J. A., Guindani, M., and Gelfand, A. E. (2007). "Generalized spatial Dirichlet process models." *Biometrika*, 94: 809–825. 148

Dubins, L. E. and Freedman, D. A. (1967). "Random distribution functions." *Proceedings of the Fifth Berkeley Symposium on Mathematics, Statistics and Probability*, 2: 183–214. 151

Dunson, D. B. and Park, J. H. (2008). "Kernel stick-breaking processes." *Biometrika*, 95: 307–323. 148

Dunson, D. B., Pillai, N., and Park, J. H. (2007). "Bayesian density regression." *Journal of the Royal Statistical Society, Series B*, 69: 163–183. 148, 149

Fabius, J. (1964). "Asymptotic behavior of Bayes' estimates." *The Annals of Mathematical Statistics*, 35: 846–856. 151

Ferguson, T. S. (1974). "Prior distribution on the spaces of probability measures." *The Annals of Statistics*, 2: 615–629. 148, 151

Freedman, D. (1963). "On the asymptotic distribution of Bayes' estimates in the discrete case." *Annals of Mathematical Statistics*, 34: 1386–1403. 151

Gasemyr, J. (2003). "On an adaptive version of the Metropolis-Hastings algorithm with independent proposal distribution." *Scandinavian Journal of Statistics*, 30: 159–173. 166

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

Gelfand, A. E. and Dey, D. (1994). "Bayesian model choice: asymptotics and exact calculations." *Journal of the Royal Statistical Society, Series B*, 56: 501–514. 160

Gelfand, A. E., Guindani, M., and Petrone, S. (2007). "Bayesian nonparametric modelling for spatial data using Dirichlet processes (with Discussion)." In Bernadro, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Hecherman, D., Smith, A. F. M., and West, M. (eds.), *Bayesian Statistics 8*, 1–26. Oxford, UK: Oxford University Press. 148

Gelfand, A. E. and Kottas, A. (2001). "Nonparametric Bayesian modeling for stochastic order." *Annals of the Institute of Statistical Mathematics*, 53: 865–876. 166

Gelfand, A. E., Kottas, A., and MacEachern, S. N. (2005). "Bayesian nonparametric spatial modeling with Dirichlet process mixing." *Journal of the American Statistical Association*, 100: 1021–1035. 148

Ghosal, S., Ghosh, J. K., and Ramamoorthi, R. V. (2006). "Consistent semiparametric Bayesian inference about a location parameter." *Journal of Statistical Planning and Inference*, 77: 181–193. 171, 172

Giordani, P. and Kohn, R. (2010). "Adaptive independent Metropolis-Hastings by fast estimation of mixtures of normals." *Journal of Computational and Graphical Statistics*, 19(2): 243–259. 166

Griffin, J. E. and Steel, M. F. J. (2006). "Order-based dependent Dirichlet processes." *Journal of the American Statistical Association*, 101: 179–194. 148, 149

Haario, H., Laine, M., Mira, A., and Saksman, E. (2006). "DRAM: Efficient adaptive MCMC." *Statistics in Computing*, 16: 339–354. 166

Haario, H., Saksman, E., and Tamminen, J. (2001). "An adaptive Metropolis algorithm." *Bernoulli*, 7: 223–242. 166

Hanson, T. (2006). "Inference for mixtures of finite Polya tree models." *Journal of the American Statistical Association*, 101: 1548–1565. 149, 154

Hanson, T. and Yang, M. (2007). "Bayesian semiparametric proportional odds models." *Biometrics*, 63(1): 88–95. 147

He, W. and Lawless, J. F. (2003). "Flexible maximum likelihood methods for bivariate proportional hazards models." *Biometrics*, 59: 837–848. 148

Hennerfeind, A., Brezger, A., and Fahrmeir, L. (2006). "Geoadditive survival models." *Journal of the American Statistical Association*, 101: 1065–1075. 148

Holden, L., Hauge, R., and Holden, M. (2009). "Adaptive independent Metropolis-Hastings." *The Annals of Applied Probability*, 19: 395–413. 166

Huzurbazar, A. V. (2005). "A censored data histogram." *Communications in Statistics, Simulation and Computation*, 34: 113–120. 165

Ishwaran, H. and James, L. F. (2004). "Computational methods for multiplicative intensity models using weighted gamma processes: proportional hazards, marked point processes and panel count data." *Journal of the American Statistical Association*, 99: 175–190. 149

Jara, A. and Hanson, T. (2011). "A class of mixtures of dependent tail-free processes." *Biometrika*, 98: 553–566. 148, 151, 152, 153, 154, 171, 172

Jara, A., Hanson, T., 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: 838–860. 150

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): 1–30.
URL http://www.jstatsoft.org/v40/i05/ 159

Jara, A., Lesaffre, E., De Iorio, M., and Quintana, F. A. (2010). "Bayesian semiparametric inference for multivariate doubly-interval-censored data." *The Annals of Applied Statistics*, 4: 2126–2149. 147

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*. New York, USA: Wiley. 147

Karabatsos, G. and Walker, S. G. (2007). "Bayesian nonparametric inference of stochastically ordered distributions, with Polya trees and Bernstein polynomials." *Statistics and Probability Letters*, 77: 907–913. 166

Keith, J. M., Kroese, D. P., and Sofronov, G. Y. (2008). "Adaptive independence samplers." *Statistics and Computing*, 18: 409–420. 166

Kottas, A., Duan, J. A., and Gelfand, A. E. (2008). "Modeling disease incidence data with spatial and spatio–temporal Dirichlet process mixtures." *Biometrical Journal*, 50: 29–42. 148, 149

Kraft, C. M. (1964). "A class of distribution function processes which have derivatives." *Journal of Applied Probability*, 1: 385–388. 151, 171

Kuo, L. and Mallick, B. (1997). "Bayesian semiparametric inference for the accelerated failure–time model." *Canadian Journal of Statistics*, 25: 457–472. 149

Li, Y. and Lin, X. (2006). "Semiparametric normal transformation models for spatially

correlated survival data." *Journal of the American Statistical Association*, 101: 591–603. 148

MacEachern, S. N. (1999). "Dependent nonparametric processes." In *ASA Proceedings of the Section on Bayesian Statistical Science, Alexandria, VA*. American Statistical Association. 148

— (2000). "Dependent Dirichlet processes." Technical report, Department of Statistics, The Ohio State University. 148

Metivier, M. (1971). "Sur la construction de mesures aléatoires presque sûrement absolument continues par rapport à une mesure donnée." *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 20: 332–344. 151

Neal, R. (2003). "Slice sampling." *The Annals of Statistics*, 31: 705–767. 156

R Development Core Team (2011). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL http://www.R-project.org 159, 164

Reich, B. J. and Fuentes, M. (2007). "A multivariate nonparametric Bayesian spatial framework for hurricane surface wind fields." *Annals of Applied Statistics*, 1: 249–264. 148, 149

Sethuraman, J. (1994). "A constructive definition of Dirichlet priors." *Statistica Sinica*, 2: 639–650. 148

Smith, B. J. (2007). "BOA: An R package for MCMC output convergence assessment and posterior inference." *Journal of Statistical Software*, 21: 1–37. 160

Spiegelhalter, S. D., 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: 583–639. 163

Taddy, M. (2008). "Bayesian nonparametric analysis of conditional distributions and inference for Poisson point processes." Unpublished doctoral thesis, Department of Applied Mathematics and Statistics, University of California Santa Cruz. 148

Tierney, L. (1994). "Markov chains for exploring posterior distributions." *The Annals of Statistics*, 22: 1701–1762. 155, 156, 173

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. 148

Zhao, L., Hanson, T., and Carlin, B. P. (2009). "Flexible spatial frailty modeling via mixtures of Polya trees." *Biometrika*, 96: 263–276. 149

Appendix A: Proof of Theorem 1

Notice that, for every $j \in \mathcal{J}$ and every $\omega \in \Omega$, $G_j^0 = G_j \circ F_{\theta_j}^{-1}$ and $G(j,\omega) = F(j,\omega) \circ F_{\theta_j}^{-1}$ are probability measures defined on $[0,1]$, and that the partitions induced by every F_{θ_j}, $j \in \mathcal{J}$, correspond to the same canonical binary partition of $[0,1]$. Since D_{KL} is invariant under one–to–one transformations of the sample space, it follows that

$$P\left\{\omega \in \Omega : \mathrm{D}_{\mathrm{KL}}(G_j, F(\omega,j)) < \delta, \forall j \in \mathcal{J}\right\} = P\left\{\omega \in \Omega : \mathrm{D}_{\mathrm{KL}}(G_j^0, G(\omega,j)) < \delta, \forall j \in \mathcal{J}\right\}.$$

Further, notice that

$$\left\{\omega \in \Omega : \mathrm{D}_{\mathrm{KL}}(G_j^0, G(\omega,j)) < \alpha, \forall j \in \mathcal{J}\right\} \supseteq \left\{\omega \in \Omega : \sum_{j\in\mathcal{J}} \mathrm{D}_{\mathrm{KL}}(G_j^0, G(\omega,j)) < \delta\right\},$$

and, therefore,

$$P\left\{\omega \in \Omega : \mathrm{D}_{\mathrm{KL}}(F_j^0, F(\omega,j)) < \delta, \forall j \in \mathcal{J}\right\} \geq P\left\{\omega \in \Omega : \sum_{j\in\mathcal{J}} \mathrm{D}_{\mathrm{KL}}(G_j^0, G(\omega,j)) < \delta\right\}.$$

Now, by the martingale convergence theorem, for every $j \in \mathcal{J}$, there exists a collection of numbers $\{y^0_{\varepsilon_1\cdots\varepsilon_l}(j) = F_j^0(B^j_{\varepsilon_1\cdots\varepsilon_l} \mid B^j_{\varepsilon_1\cdots\varepsilon_{l-1}}) : \varepsilon_1\cdots\varepsilon_l \in E^*, l \in \mathbb{N}\}$ from $[0,1]$ such that, with probability one, the density of G_j^0 w.r.t. Lebesgue measure is given by

$$g_j^0(z) = \lim_{m\to\infty}\left[2^m \prod_{l=1}^{m} y^0_{\varepsilon_1\cdots\varepsilon_l}(j)\right],$$

where the limit is taken through a sequence $\varepsilon_1\varepsilon_2\cdots$ which corresponds to the dyadic expansion of $z \in [0,1]$. Since $\sum_{l=1}^{\infty}\rho(l)^{-1/2} < \infty$ by assumption, Proposition 4 in Jara and Hanson (2011) and the result by Kraft (1964) imply the existence of the density of $G(j,\omega)$ w.r.t. Lebesgue measure, for almost every $\omega \in \Omega$. Therefore, it similarly follows that

$$g(j,\omega)(z) = \lim_{m\to\infty}\left[2^m \prod_{l=1}^{m} Y_{\varepsilon_1\cdots\varepsilon_l}(j,\omega)\right],$$

for almost every ω. Borrowing the trick from Ghosal et al. (2006), used to show the large support of Polya trees, we note that for any $N \geq 1$,

$$D_{KL}\left(G_j^0, G(j,\omega)\right) = A_N^{j,\omega} + B_N^j - C_N^{j,\omega},$$

for almost every $\omega \in \Omega$, where

$$A_N^{j,\omega} = E_{g_j^0}\left[\log\left\{\prod_{l=1}^{N}\frac{y^0_{\varepsilon_1\cdots\varepsilon_l}(j)}{Y_{\varepsilon_1\cdots\varepsilon_l}(j,\omega)}\right\}\right],$$

$$B_N^j = E_{g_j^0}\left[\log\left\{\prod_{l=N+1}^{\infty} 2y^0_{\varepsilon_1\cdots\varepsilon_l}(j)\right\}\right],$$

and

$$C_N^{j,\omega} = E_{g_j^0}\left[\log\left\{\prod_{l=N+1}^{\infty} 2Y_{\varepsilon_1\cdots\varepsilon_l}(j,\omega)\right\}\right],$$

with $E_{g_j^0}$ being the conditional expectation with respect to the distribution of $(\varepsilon_1\varepsilon_2\cdots)$, the dyadic expansion of z, which is distributed according to the density g_j^0. By the definition of the dependent ARDTFP, $\{A_N^{j,\cdot} : j \in \mathcal{J}\}$ and $\{C_N^{j,\cdot} : j \in \mathcal{J}\}$ are independent random functions for every integer $N \geq 1$.

Now notice that, by the Markov inequality,

$$\begin{aligned}
P\left\{\omega \in \Omega : \sum_{j\in\mathcal{J}}\left|C_N^{j,\omega}\right| \geq \frac{\delta}{3}\right\} &\leq 3\delta^{-1}\sum_{j\in\mathcal{J}} E\left|C_N^{j,\cdot}\right|, \\
&\leq \delta^{-1}\sum_{j\in\mathcal{J}} E\left[E_{g_j^0}\left(\sum_{l=N+1}^{\infty}\left|\log\left(2Y_{\varepsilon_1\cdots\varepsilon_l}(j,\cdot)\right)\right|\right)\right], \\
&\leq \delta^{-1}\sum_{j\in\mathcal{J}} E_{g_j^0}\left[\sum_{l=N+1}^{\infty}\max_{\varepsilon_l\in\{0,1\}}\left\{E\left|\log\left(2Y_{\varepsilon_1\cdots\varepsilon_l}(j,\cdot)\right)\right|\right\}\right], \\
&\leq \delta^{-1}\sum_{j\in\mathcal{J}}\left[\sum_{l=N+1}^{\infty}\max_{\varepsilon_1\cdots\varepsilon_{l-1}\in\{0,1\}^{l-1}}\left\{E\left|\log\left(2Y_{\varepsilon_1\cdots\varepsilon_{l-1}0}(j,\cdot)\right)\right|\right\}\right], \\
&\leq \delta^{-1}\sum_{j\in\mathcal{J}}\left[\sum_{l=N+1}^{\infty}\varphi\left(c\rho(l)\right)\right],
\end{aligned}$$

where $\varphi\left(c\rho(l)\right) = E\left|\log\left(2Y_{\varepsilon_1\cdots\varepsilon_{l-1}0}(j,\cdot)\right)\right|$. By Proposition 3 in Jara and Hanson (2011), for every $j \in \mathcal{J}$, $Y_{\varepsilon_1\cdots\varepsilon_{l-1}0}(j,\cdot) \xrightarrow{\mathcal{L}} U$, $l \longrightarrow \infty$, where U is a $\text{Beta}(c\rho(j), c\rho(j))$ random variable. It follows that, by Lemma A.1 in Ghosal et al. (2006), $\varphi\left(c\rho(l)\right) = \text{O}\left(\left[c\rho(l)\right]^{-1/2}\right)$ as $l \longrightarrow \infty$. Since $\sum_{l=1}^{\infty}\rho(l)^{-1/2} < \infty$ by assumption, $\sum_{l=1}^{\infty}\varphi\left(c\rho(l)\right)$ is a convergent series and, therefore, there exists an integer N_C such that

$$P\left\{\omega \in \Omega : \sum_{j\in\mathcal{J}}\left|C_{N_C}^{j,\omega}\right| \geq \frac{\delta}{3}\right\} \leq 3\delta^{-1}\sum_{j\in\mathcal{J}}\left[\sum_{l=N_C+1}^{\infty}\varphi\left(c\rho(l)\right)\right] < 1,$$

i.e.,

$$P\left\{\omega \in \Omega : \sum_{j\in\mathcal{J}}\left|C_{N_C}^{j,\omega}\right| < \frac{\delta}{3}\right\} > 0. \tag{A.1}$$

By Lemma 2 in Barron (1985), for every $j \in \mathcal{J}$, there exists an integer N_B^j such that $\left|B^j_{N_B^j}\right| < \frac{\delta}{3J}$. It follows that

$$\sum_{j\in\mathcal{J}} \left|B_N^j\right| < \frac{\delta}{3}, \tag{A.2}$$

where $N = \max\left(N_C, \{N_B^j, j \in \mathcal{J}\}\right)$. Finally, the set

$$\left\{\left(Y_{\varepsilon_1\cdots\varepsilon_N}(j,\cdot), j \in \mathcal{J}, \varepsilon_1\cdots\varepsilon_N \in \{0,1\}^N\right) : \sum_{j\in\mathcal{J}} A_N^{j,\cdot} < \delta/3\right\},$$

is a nonempty open set in $\mathbb{R}^{2^{JN}-1}$; it is open by the continuity of the relevant map while it is nonempty as $\left(y^0_{\varepsilon_1\cdots\varepsilon_N}(j), j \in \mathcal{J}, \varepsilon_1\cdots\varepsilon_N \in \{0,1\}^N\right)$ belongs to this set. Thus, by the nonsingularity of the multivariate logistic normal distribution, which is induced by the nonsingularity of the multivariate normal distribution of the AR Gaussian processes and by the continuity of the logistic transformation, it follows that

$$P\left\{\omega \in \Omega : \sum_{j\in\mathcal{J}} A_N^{j,\omega} < \frac{\delta}{3}\right\} > 0. \tag{A.3}$$

Thus, from relations (A.1), (A.2) and (A.2), it follows that for any $\delta > 0$, there is some $N > 1$ such that

$$\begin{aligned}
&P\left\{\omega \in \Omega : \mathrm{D}_{\mathrm{KL}}\left(G_j^0, G(j,\omega)\right) < \delta, \forall j \in \mathcal{J}\right\} \\
&\qquad \geq P\left\{\omega \in \Omega : \sum_{j\in\mathcal{J}} \mathrm{D}_{\mathrm{KL}}\left(G_j^0, G(j,\omega)\right) < \delta\right\}, \\
&\qquad = P\left\{\omega \in \Omega : \sum_{j\in\mathcal{J}} \left[A_N^{j,\omega} + B_N^j - C_N^{j,\omega}\right] < \delta\right\}, \\
&\qquad \geq P\left\{\omega \in \Omega : \sum_{j\in\mathcal{J}} A_N^{(j,\omega)} < \frac{\delta}{3}\right\} P\left\{\omega \in \Omega : \sum_{j\in\mathcal{J}} \left|C_N^{j,\omega}\right| < \frac{\delta}{3}\right\}, \\
&\qquad > 0,
\end{aligned}$$

which completes the proof of the theorem.

Appendix B: Complementary MCMC details

Updating the regression coefficients

We updated β using a random walk MH step (Tierney 1994). The covariance matrix of the candidate generating distribution is initially taken from the asymptotic variance of

the regression coefficients, obtained fitting the parametric non–frailty model. We then "refined" the covariance matrix by running a crude MH random walk sampler for the full model for 5000 steps. The resulting empirical covariance matrix $\mathbf{V}$ was then used to generate the candidates from a normal distribution $\beta^* \sim N_p(\beta, k\mathbf{V})$, where k is an appropriate positive constant, which were accepted with probability given by

$$\min\left\{1, \frac{\mathcal{L}\left(\beta^*, \gamma, \mathcal{Y}\right)}{\mathcal{L}\left(\beta, \gamma, \mathcal{Y}\right)}\right\}.$$

Updating the frailty terms

The full conditional distribution for each frailty term is not recognizable. Therefore, the γ_i's can be updated using slice sampling or an MH step. In our implementation, a random walk MH algorithm was considered, where the candidates were generated from a normal distribution $\gamma_i^* \sim N(\gamma_i, s^2_{\gamma_i})$, with s_{γ_i} being a given constant which can be adjusted at each county to obtain an appropriate acceptance rate. We accept γ_i^* with probability

$$\min\left\{1, \frac{\left[\prod_{j=1}^{J}\prod_{k=1}^{n_{ij}}\mathcal{L}_{ijk}\left(\beta, \gamma_i^*, \mathcal{Y}_j, \theta_j\right)\right] p\left(\gamma_i^* \mid \lambda, \gamma_l, l \neq i\right)}{\left[\prod_{j=1}^{J}\prod_{k=1}^{n_{ij}}\mathcal{L}_{ijk}\left(\beta, \gamma_i, \mathcal{Y}_j, \theta_j\right)\right] p\left(\gamma_i^* \mid \lambda, \gamma_l, l \neq i\right)}\right\},$$

where

$$p\left(\gamma_i \mid \lambda, \gamma_l, l \neq i\right) = \begin{cases} \phi\left(\gamma_i \mid 0, \lambda^{-1}\right), & \text{if an exchangeable prior is assumed,} \\ \phi\left(\gamma_i \mid \bar{\gamma}_i, [d_i\lambda]^{-1}\right) & \text{if a CAR prior is assumed} \end{cases},$$

d_i is the number of neighbors of county i and $\bar{\gamma}_i$ is the mean of the d_i adjacent frailties. Since there are a lot of frailties, a common standard deviation is chosen $s_{\gamma_i} \equiv s_\gamma$ for all i, that balances overall frailty acceptance rates to be around 40%.

Updating the scale parameter of the frailty distribution

Under $\gamma_1 \ldots, \gamma_I \mid \lambda \sim \text{CAR}(\lambda)$, if a gamma prior is placed on λ, $\lambda \mid \tau_{1,\lambda}, \tau_{2,\lambda} \sim \Gamma\left(\tau_{1,\lambda}/2, \tau_{2,\lambda}/2\right)$, then the full conditional distribution for λ has a closed form and is given by

$$\lambda \mid \cdots \sim \Gamma\left(\frac{\tau_{1,\lambda} + I - 1}{2}, \frac{\tau_{2,\lambda} + \sum_{i=1}^{I} d_i\gamma_i(\gamma_i - \bar{\gamma}_i)}{2}\right),$$

where, as before, d_i is the number of neighbors of county i and $\bar{\gamma}_i$ is the mean of the d_i adjacent frailties.

Similarly, if the same prior is used for λ, under $(\gamma_1, \ldots, \gamma_I) \mid \lambda \sim N_n(\mathbf{0}, \lambda^{-1}\mathbf{I}_I)$, then the full conditional distribution is also a gamma distribution given by

$$\lambda \mid \cdots \Gamma\left(\frac{\tau_{1,\lambda} + I}{2}, \frac{\tau_{2,\lambda} + \sum_{i=1}^{I}\gamma_i^2)}{2}\right).$$

Acknowledgments

The work of the first and third authors was supported in part by NIH grant 2–R01–CA95955–05. The first author was also supported by NSF grant CMMI-0855329. The second author was supported by Fondecyt 11100144 grant. The authors are grateful to the associate editor and reviewers for many detailed suggestions and helpful comments that have led to a considerable improvement in the manuscript.

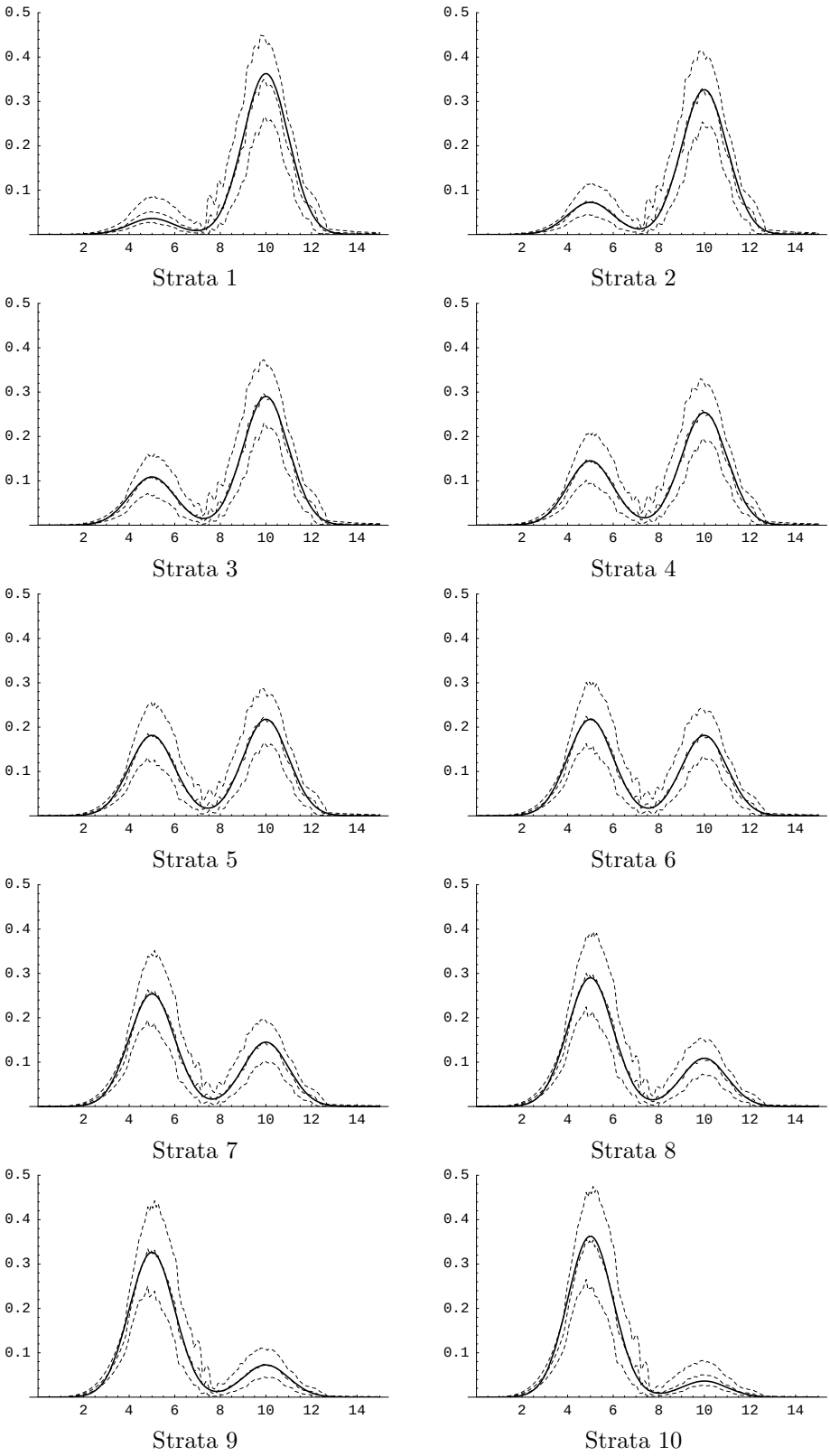

Figure 1: Simulated data: The posterior mean and 95% pointwise credible interval under the partially specified ARDFP with $M = 5$ are displayed as dashed lines for each strata. The true density in each case is displayed as a solid line.

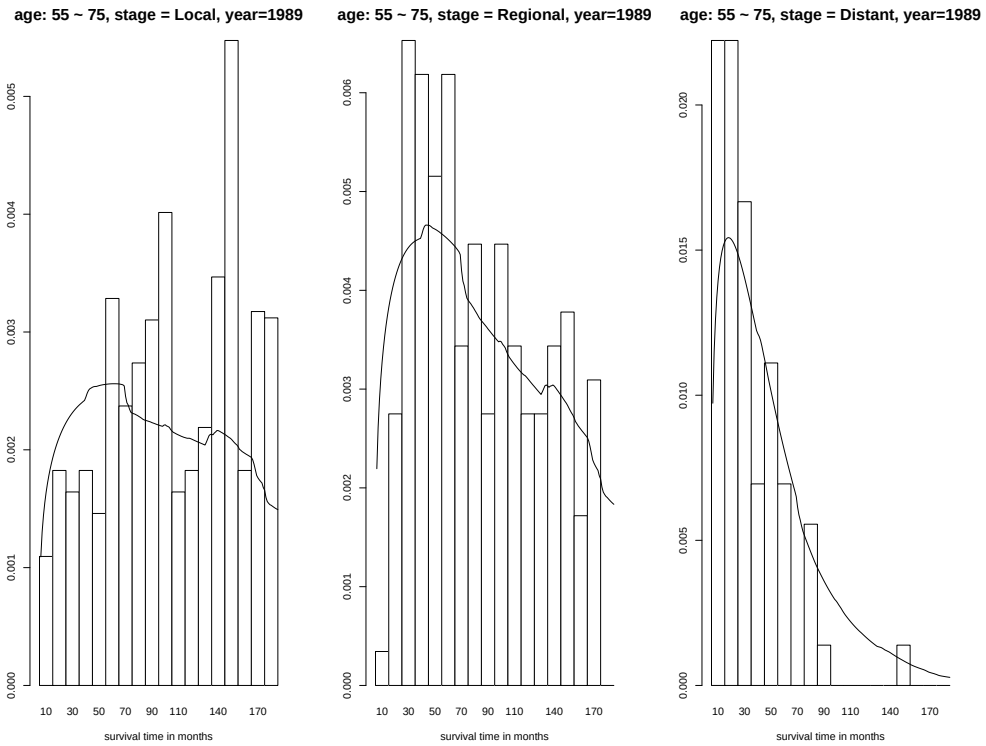

Figure 2: SEER data: Censored histograms of raw survival in months for diagnosis ages 55–75, one diagnosed independent cancer, and "local" stage (left), "regional" (middle), or "distant" (right), with fitted predictive density curves from the best fitting ARDTFP model overlaid for 1989.

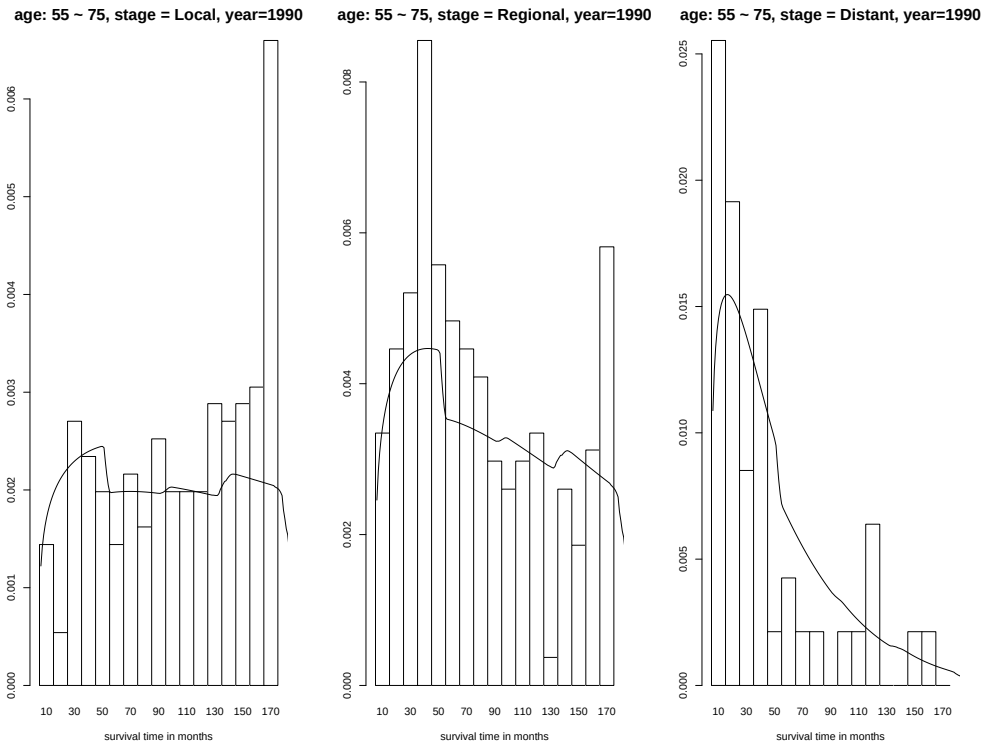

Figure 3: SEER data: Censored histograms of raw survival in months for diagnosis ages 55–75, one diagnosed independent cancer, and "local" stage (left), "regional" (middle), or "distant" (right), with fitted predictive density curves from the best fitting ARDTFP model overlaid for 1990.

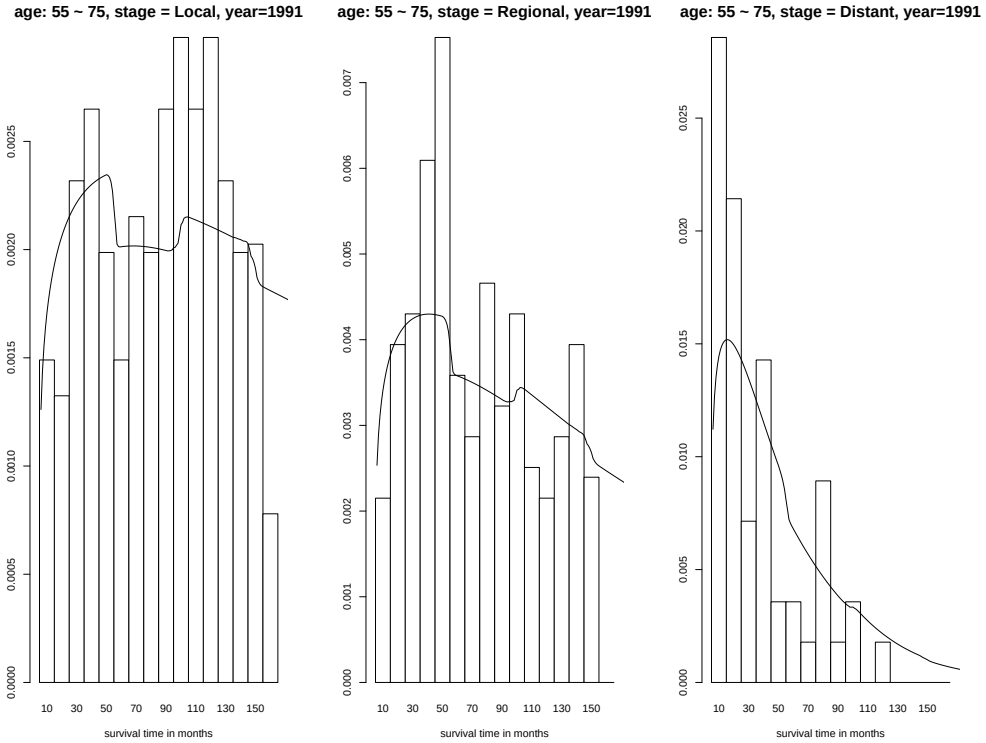

Figure 4: SEER data: Censored histograms of raw survival in months for diagnosis ages 55–75, one diagnosed independent cancer, and "local" stage (left), "regional" (middle), or "distant" (right), with fitted predictive density curves from the best fitting ARDTFP model overlaid for 1991.

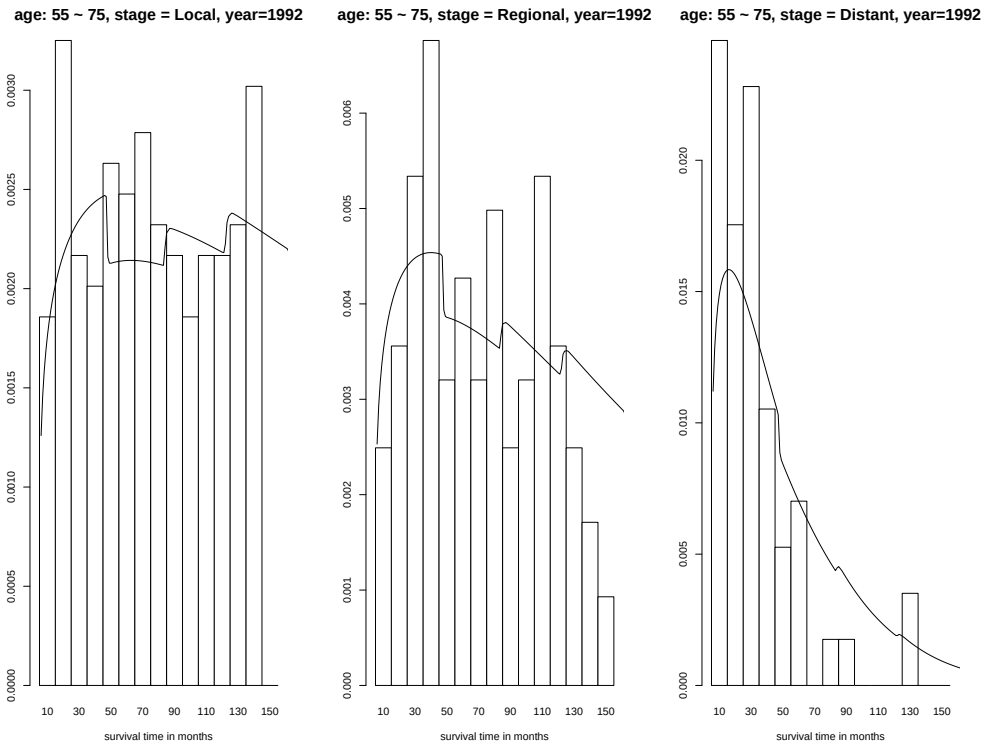

Figure 5: SEER data: Censored histograms of raw survival in months for diagnosis ages 55–75, one diagnosed independent cancer, and "local" stage (left), "regional" (middle), or "distant" (right), with fitted predictive density curves from the best fitting ARDTFP model overlaid for 1992.

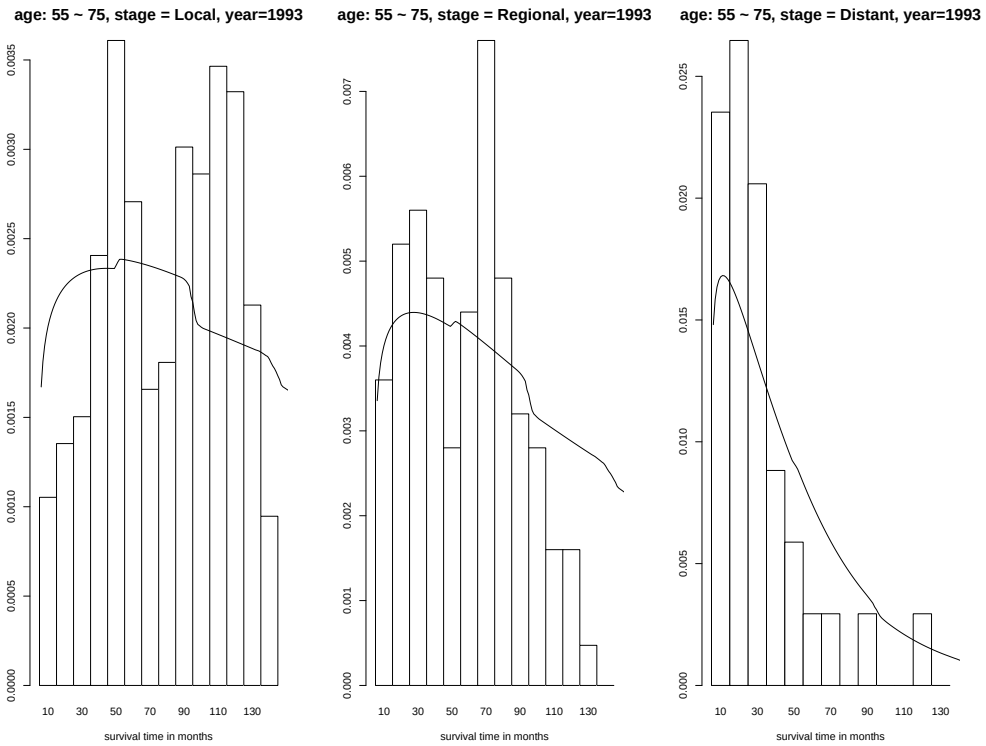

Figure 6: SEER data: Censored histograms of raw survival in months for diagnosis ages 55–75, one diagnosed independent cancer, and "local" stage (left), "regional" (middle), or "distant" (right), with fitted predictive density curves from the best fitting ARDTFP model overlaid for 1993.

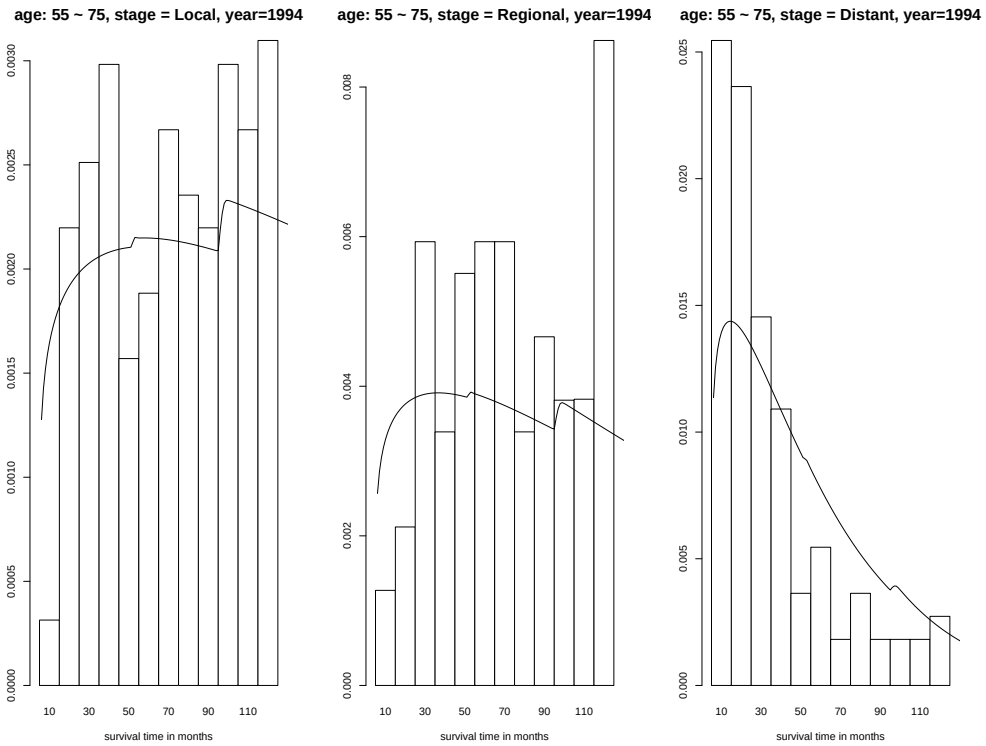

Figure 7: SEER data: Censored histograms of raw survival in months for diagnosis ages 55–75, one diagnosed independent cancer, and "local" stage (left), "regional" (middle), or "distant" (right), with fitted predictive density curves from the best fitting ARDTFP model overlaid for 1994.

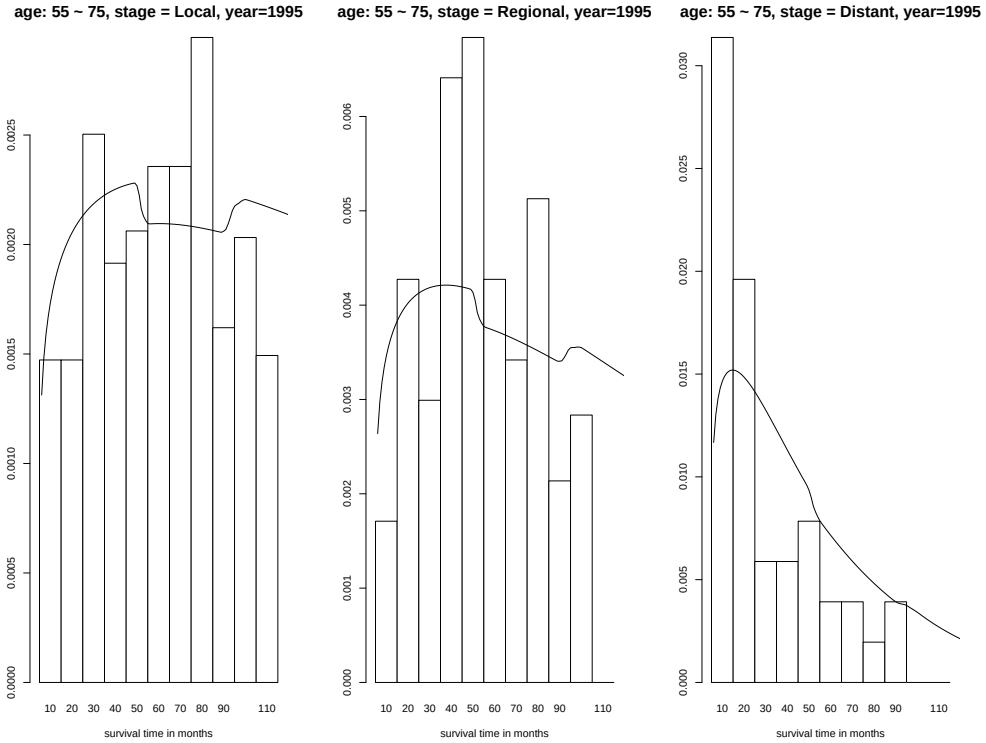

Figure 8: SEER data: Censored histograms of raw survival in months for diagnosis ages 55–75, one diagnosed independent cancer, and "local" stage (left), "regional" (middle), or "distant" (right), with fitted predictive density curves from the best fitting ARDTFP model overlaid for 1995.

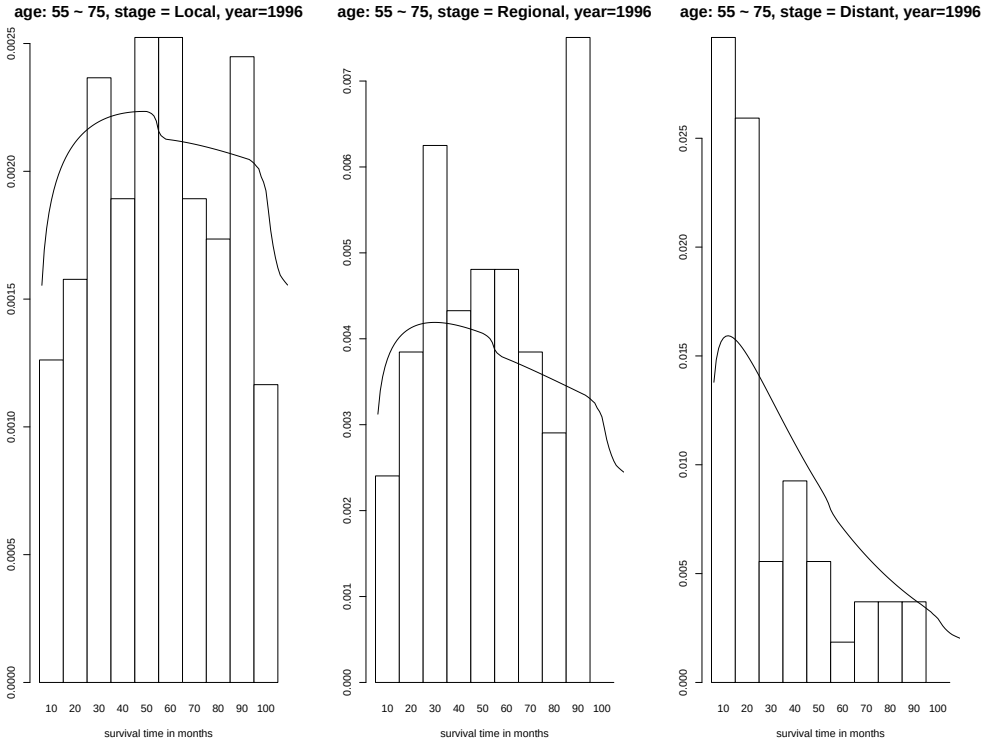

Figure 9: SEER data: Censored histograms of raw survival in months for diagnosis ages 55–75, one diagnosed independent cancer, and "local" stage (left), "regional" (middle), or "distant" (right), with fitted predictive density curves from the best fitting ARDTFP model overlaid for 1996.

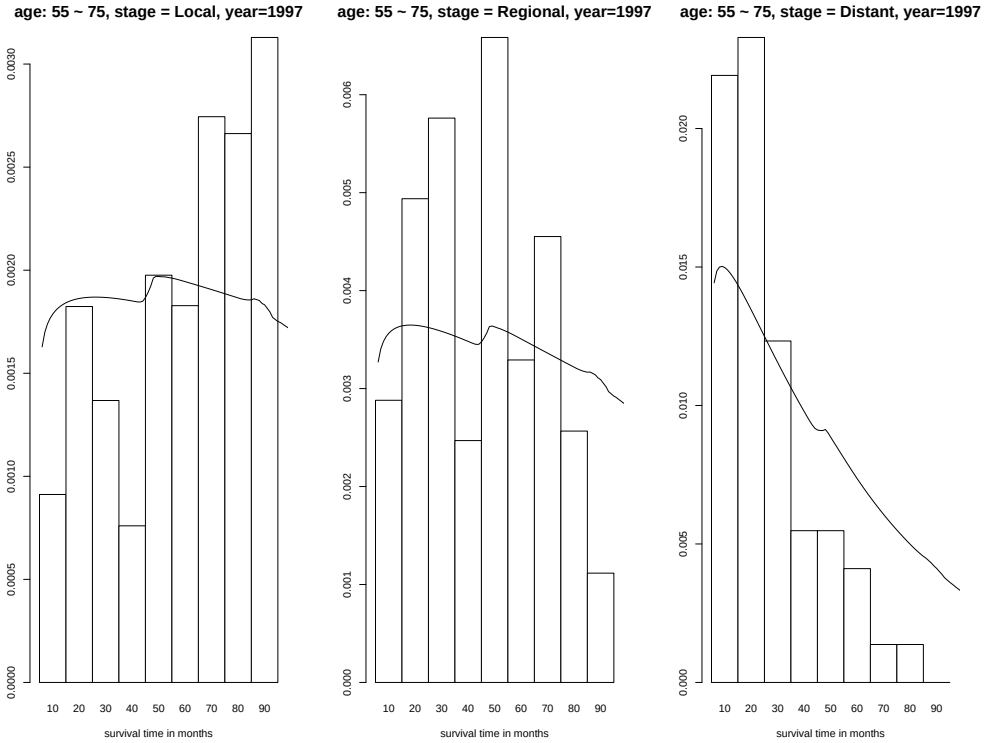

Figure 10: SEER data: Censored histograms of raw survival in months for diagnosis ages 55–75, one diagnosed independent cancer, and "local" stage (left), "regional" (middle), or "distant" (right), with fitted predictive density curves from the best fitting ARDTFP model overlaid for 1997.

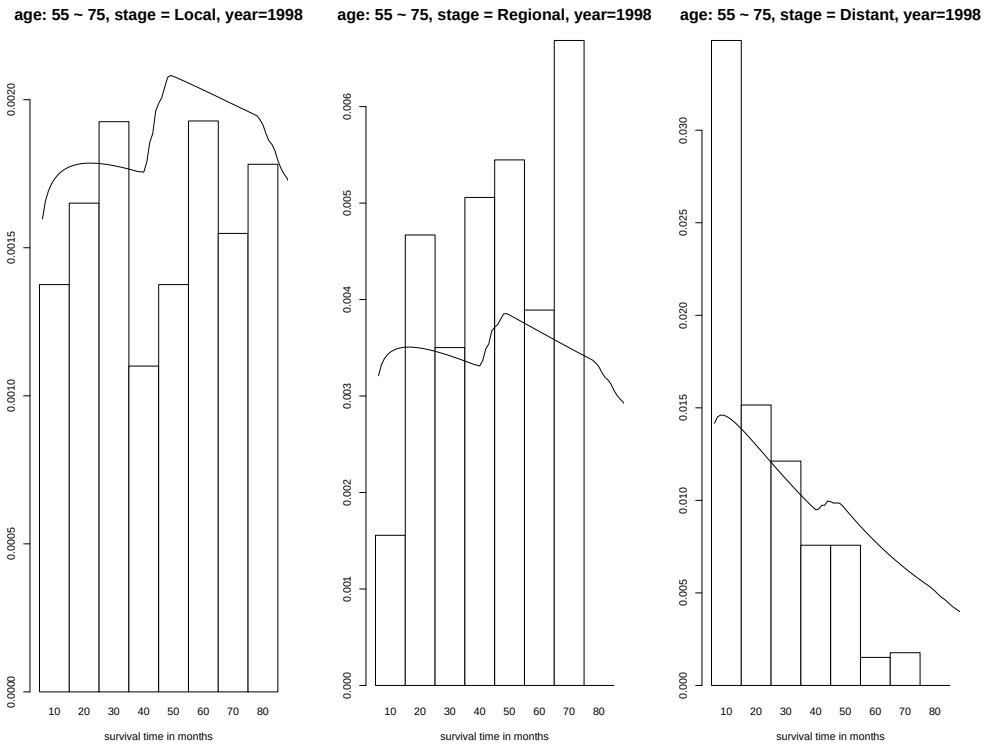

Figure 11: SEER data: Censored histograms of raw survival in months for diagnosis ages 55–75, one diagnosed independent cancer, and "local" stage (left), "regional" (middle), or "distant" (right), with fitted predictive density curves from the best fitting ARDTFP model overlaid for 1998.

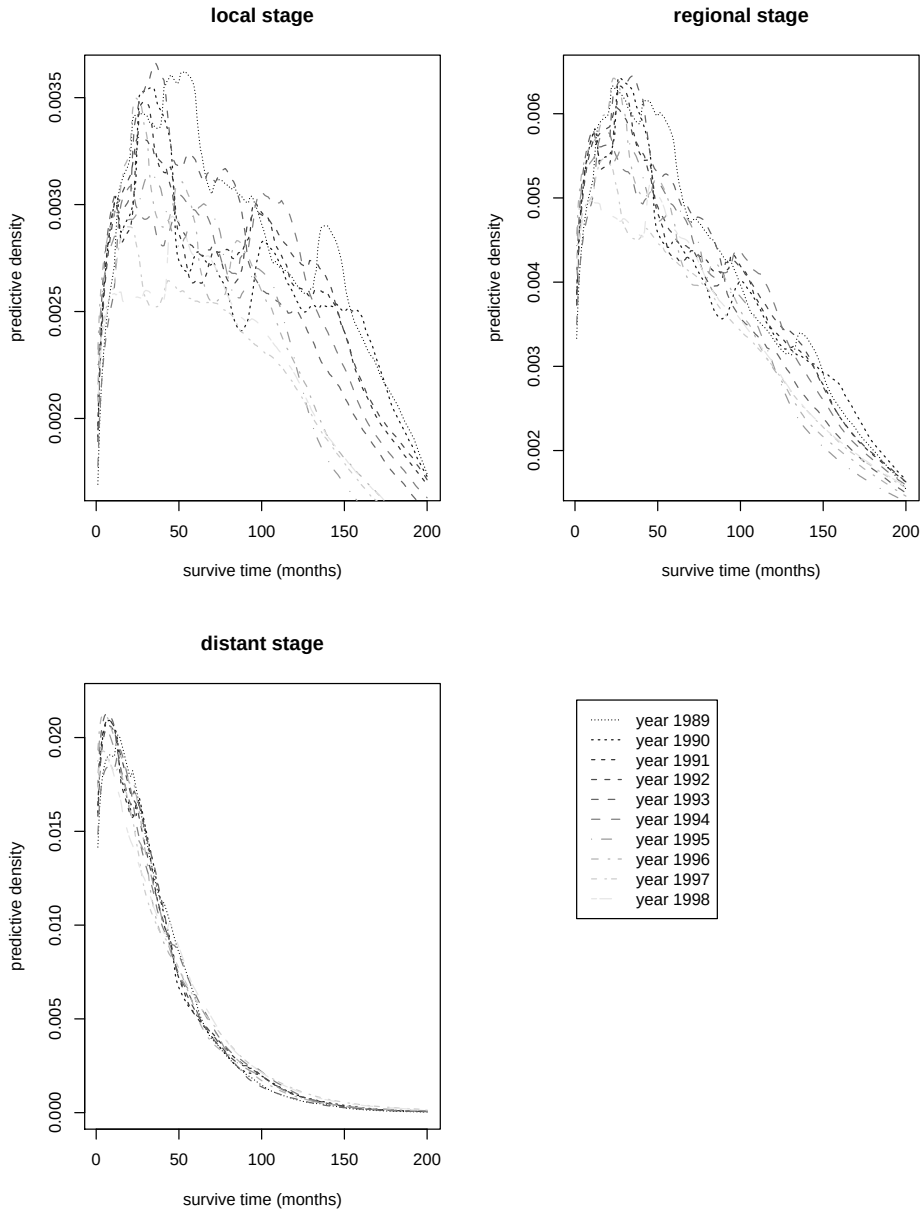

Figure 12: SEER data: Predictive densities for age 65, breast cancer only, and disease stages "local" (upper left), "regional" (upper right), and "distant" (lower left) across 200 months.

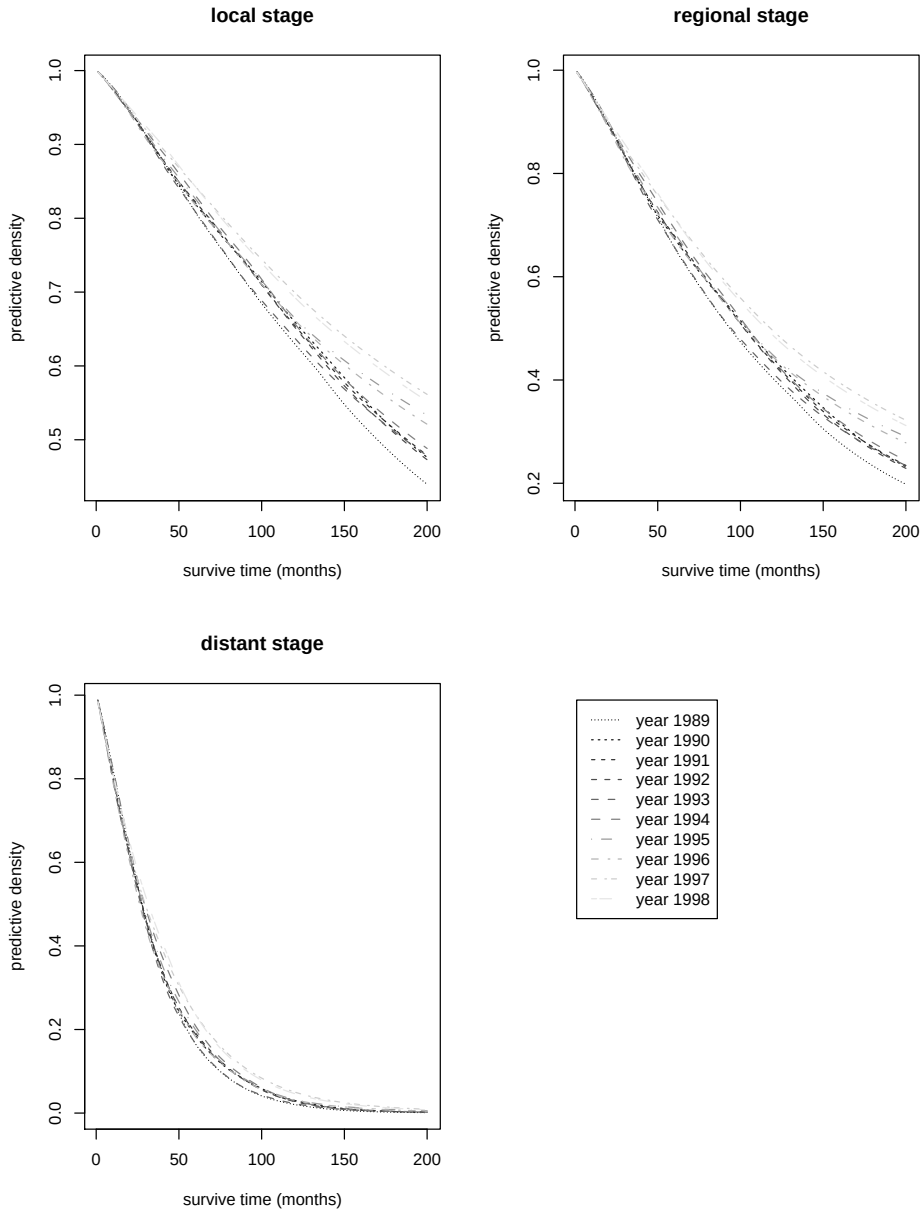

Figure 13: SEER data: Predictive survival curves of age 65, breast cancer only, and disease stages "local" (upper left), "regional" (upper right), and "distant" (lower left) across 200 months.