Biometrics 71, 313–322
June 2015
DOI: 10.1111/biom.12268

Spatial Extended Hazard Model with Application to Prostate Cancer Survival

Li Li,[1,*] Timothy Hanson,[2] and Jiajia Zhang[3]

[1]Department of Mathematics and Statistics, University of New Mexico, Albuquerque, New Mexico, U.S.A.
[2]Department of Statistics, University of South Carolina, Columbia, South Carolina, U.S.A.
[3]Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, South Carolina, U.S.A.
*email: llis@unm.edu

Summary. This article develops a Bayesian semiparametric approach to the extended hazard model, with generalization to high-dimensional spatially grouped data. County-level spatial correlation is accommodated marginally through the normal transformation model of Li and Lin (2006, *Journal of the American Statistical Association*, **101**, 591–603), using a correlation structure implied by an intrinsic conditionally autoregressive prior. Efficient Markov chain Monte Carlo algorithms are developed, especially applicable to fitting very large, highly censored areal survival data sets. Per-variable tests for proportional hazards, accelerated failure time, and accelerated hazards are efficiently carried out with and without spatial correlation through Bayes factors. The resulting reduced, interpretable spatial models can fit significantly better than a standard additive Cox model with spatial frailties.

Key words: Censored data; Gaussian Copula; Intrinsic autoregressive prior; Normal transformation model.

1. Introduction

The extended hazard (EH) model (Etezadi-Amoli and Ciampi, 1987; Chen and Jewell, 2001) includes the proportional hazards (PH) model (Cox, 1972; Kalbfleisch, 1978), the accelerated failure time (AFT) model (Buckley and James, 1979; Komárek and Lessaffre, 2008), and the accelerated hazards (AH) model (Chen and Wang, 2000; Chen, Hanson, and Zhang, 2014) as special cases. Denote $\lambda_0(\cdot)$ as the baseline hazard function and $\mathbf{z}$ as a covariate vector. The EH model assumes the individual hazard function

$$\lambda(t|\mathbf{z}) = \lambda_0(te^{\boldsymbol{\beta}'\mathbf{z}})e^{\boldsymbol{\gamma}'\mathbf{z}}. \quad (1)$$

The more easily interpretable PH, AFT, and AH models occur as special cases when $\boldsymbol{\beta} = \mathbf{0}$, $\boldsymbol{\beta} = \boldsymbol{\gamma}$, and $\boldsymbol{\gamma} = 0$, respectively. Note that the model also allows for per-variable PH, AFT, or AH effects. For example, say $\mathbf{z} = (z_1, z_2)$ and consider the model $\lambda(t|\mathbf{z}) = \lambda_0(te^{\beta z_1})e^{\beta z_1 + \gamma z_2}$. Holding z_1 constant, z_2 has PH interpretation; holding z_2 constant, z_1 has AFT interpretation. Such reduced semiparametric models have enhanced interpretability, separating inference into easily interpretable parametric (regression coefficients) and nonparametric (baseline hazard) model components.

Our goal is to analyze large cancer registry data sets, which typically record each patient's location up to a district or county due to patient confidentiality. A common feature of these data is that the failure times are correlated. There are two main categories of methods to model spatially correlated areal failure times. One category introduces county-level frailties to the survival model, which gives conditional interpretations for covariate effects (e.g., Banerjee et al., 2003). The other category includes marginal methods, which produce "population-averaged" covariate effects, for example, the marginal method by Cai et al. (2007) through modifications of the hazard, and the normal transformation model by Li and Lin (2006), which is a Gaussian copula model. However, the normal transformation model by Li and Lin (2006) allows careful modeling of the spatial correlation. Since we seek to formally test whether simpler models are adequate relative to the EH model with spatial correlation, frailties complicate such tests, as two complete sets of frailties need be included, one for each linear predictor. For example, the EH model augmented with frailties is

$$\lambda(t_i|\mathbf{z}) = \lambda_0\{t_i e^{\boldsymbol{\beta}'\mathbf{z}_i + b_{c_i}}\}e^{\boldsymbol{\gamma}'\mathbf{z}_i + g_{c_i}},$$

where c_i is the county subject i belongs to, and for our data, $b_1, \ldots, b_{46}$ and $g_1, \ldots, g_{46}$ are county-level frailties for South Carolina. To test that PH is adequate, the hypotheses $H_0 : \boldsymbol{\beta} = \mathbf{0}, b_j = 0, j = 1, \ldots, 46$ where j refers to county, need to be considered; the per-variable tests are even more complex. In contrast, we show later in the article that the normal transformation model is more easily implemented and allows ready interpretation. Li and Lin (2006) consider estimation in the PH model for spatially correlated georeferenced data. Note that their georeferenced approach does not work for large areal data sets without significant modification. We generalize the georeferenced normal transformation PH model of Li and Lin (2006) to EH with a correlation structure suitable for areal data, and develop two novel Markov Chain Monte Carlo (MCMC) schemes for posterior updating. Since all three PH, AFT, and AH models are formally nested within

the EH model, Bayes factors are quickly computed using the Savage–Dickey ratio (Verdinelli and Wasserman, 1995).

Define $Y_i = \Phi^{-1}\left\{1 - e^{\Lambda_i(T_i)}\right\}$ where $\Phi(\cdot)$ is the standard normal cumulative distribution function, T_i the random failure time, and $\Lambda_i(\cdot)$ the cumulative hazard function for T_i. Let $\mathbf{Y} = (Y_1, \ldots, Y_n)'$. Under the normal transformation model in Li and Lin (2006), $\mathbf{Y}$ follows a joint multivariate normal distribution with mean zero and covariance $\boldsymbol{\Gamma}$. That is,

$$\mathbf{Y} \sim N(\mathbf{0}, \boldsymbol{\Gamma}). \tag{2}$$

The normal transformation model incorporates covariate effects in $\lambda_i(\cdot)$ through (1) and spatial dependence by $\boldsymbol{\Gamma}$. Details of $\boldsymbol{\Gamma}$ are presented in Section 3.2.

There has been renewed, recent interest in stably estimating the EH model. Both Tseng and Shu (2011) and Tong et al. (in press) consider a kernel-smoothed profile likelihood (KSPL) approach to fitting the EH model, and also propose tests to choose among EH, AFT, or PH. The KSPL approach uses a piecewise-constant baseline hazard function λ_0 (Sinha and Dey, 1997), with a fixed number of hazard jumps at fixed locations. It is difficult to generalize their optimization procedure for fitting the EH to the spatial case as the likelihood becomes much more complicated and hence cause problems for kernel-smoothing. Moreover, for big datasets, a large number of hazard jumps needs to be used because time accelerates or decelerates in the argument of the hazard function by a factor $e^{\boldsymbol{\beta}'\mathbf{z}}$ and it is not known a priori what the effective support of λ_0 is. For this reason, the baseline hazard should also have a scale factor to appropriately stretch or shrink $\lambda_0(\cdot)$ as necessary, depending on the effective support of the baseline survival. Several parametric families commonly used in survival analysis, generically $F_{\boldsymbol{\theta}}$, have such scale parameters, for example, log-logistic and gamma. We generalize these families via a penalized B-spline model that is centered at the parametric hazard in the sense that $E\{\lambda_0(t)\}$ approximates $\lambda_{\boldsymbol{\theta}}(t)$ over the positive support of the B-spline. The resulting model behaves like a blend of B-splines and a smoothed gamma process that is able to capture a wide variety of density/hazard shapes, yet remain anchored at a parametric family. Moreover, this penalized B-spline model easily accommodates the spatial generalization and greatly facilitates the MCMC computation for big data set.

We analyze a large prostate cancer data set from the South Carolina Central Cancer Registry (SCCCR) for the period 1996–2004; the SCCCR data are described in Hurley et al. (2009). The SCCCR is a population-based cancer registry covering the entire state of South Carolina that has data completeness in excess of 97.5%. Specifically, we investigate racial disparities in prostate cancer mortality accounting for county-level spatial dependence among subjects using interpretable refinements of an EH model.

This article is organized as follows. Section 2 presents the proposed method for fitting the EH model where the baseline hazard is flexibly modeled via a novel penalized B-spline centered at a given parametric family. Section 3 generalizes the EH model to incorporate spatial dependence on a lattice (county-level) while retaining marginal interpretation. An analysis of the SCCCR prostrate cancer data using the proposed models is presented in Section 4.

2. Extended Hazard Model

In the EH model (1), $\boldsymbol{\beta}$ characterizes the acceleration or deceleration of the hazard progression and $\boldsymbol{\gamma}$ characterizes the change in the relative hazards after adjusting the different hazard progressions. Let T and C be random failure and censoring times, respectively. Conditioning on a p-dimensional covariate vector $\mathbf{z}$, we assume T and C are independent. Consider n subjects in the study; each subject is observed with an event time t_i when $\delta_i = 1$ and is right-censored when $\delta_i = 0$. The likelihood based on data $\mathcal{D} = \{(t_i, \delta_i, \mathbf{z}_i)\}_{i=1}^n$ under model (1) is

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda_0(\cdot)) = \prod_{i=1}^{n} \left\{e^{\boldsymbol{\gamma}'\mathbf{z}_i}\lambda_0(e^{\boldsymbol{\beta}'\mathbf{z}_i}t_i)\right\}^{\delta_i} \times \exp\left\{-e^{\boldsymbol{\gamma}'\mathbf{z}_i}\int_0^{t_i}\lambda_0(te^{\boldsymbol{\beta}'\mathbf{z}_i})dt\right\}. \tag{3}$$

Proper priors are required to compute Bayes factors; the most common choice are normal priors with fixed means and covariances. We consider g-priors for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, that is,

$$\begin{aligned} p(\boldsymbol{\beta}) &\sim N_d\left(\boldsymbol{\mu}_{\boldsymbol{\beta}}, g_1 n(\mathbf{Z}'\mathbf{Z})^{-1}\right); \\ p(\boldsymbol{\gamma}) &\sim N_d\left(\boldsymbol{\mu}_{\boldsymbol{\gamma}}, g_2 n(\mathbf{Z}'\mathbf{Z})^{-1}\right), \\ g_1^{-1}; g_2^{-1} &\overset{\text{ind.}}{\sim} \text{Gamma}(a_g, b_g), \end{aligned} \tag{4}$$

where $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)'$ and $(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\mu}_{\boldsymbol{\gamma}})$ are prior means that use information on the range of acceleration factors and hazard ratios in the population. Here, b_g is the rate parameter of the Gamma distribution. Recently, the g-prior has been advocated for nonlinear regression models (e.g., Rathbun and Fei, 2006; Bové and Held, 2011). The g-prior takes into account the variability and correlation among predictors and can be quite flexible, compared to typical, diffuse normal priors for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. We have found the g-prior works well in simulations and data analyses. More on g-priors is discussed in Section 4.

2.1. *Baseline Hazard*

B-splines are now a standard tool in modeling hazard functions. They generalize the piecewise constant hazards (first order B-spline), which have been extensively used in Bayesian survival analysis (see, e.g., Ibrahim, Chen, and Sinha, 2001). Existing approaches to modeling hazard functions using B-splines (e.g., Gray, 1992; Hennerfeind et al., 2006; Sharef et al., 2010) choose either equispaced knots over the spread of the observed data or knots at the empirical quantiles of the observed event times. Since we intend to fit the EH model while accounting for spatial correlation, we develop a simpler, yet highly flexible approach to knot selection that borrows from Bayesian nonparametrics.

Assume the baseline hazard function to take the form

$$\lambda_0(t) = \sum_{j=1}^{J} b_j B_{kj}(t), \tag{5}$$

where $B_{k1}(\cdot), \ldots, B_{kJ}(\cdot)$ are kth order B-spline basis functions expanded over a knot sequence $\mathbf{s} = (s_{1-k}, \ldots, s_J)$ (De Boor,

2001) and $b_1, \ldots, b_J$ are positive B-spline coefficients. For an arbitrary vector $\mathbf{s}$, let $\mathbf{s}_{j:k}$ be the jth through kth elements of $\mathbf{s}$. Set the boundary knots $\mathbf{s}_{(1-k):0} = \mathbf{0}_k$, and $\mathbf{s}_{(J-k):J} = \mathbf{1}_k s_{J-k+1}$. Let $F_{\boldsymbol{\theta}}(\cdot)$ be the cumulative distribution function for $\lambda_{\boldsymbol{\theta}}(\cdot)$. Let $p_1, \ldots, p_{J-k+1}$ be probabilities between 0 and 1 in an increasing order. Our default choice is $p_j = jp_{\max}/(J-k+1)$, $j = 1, \ldots, J-k+1$, where $p_{\max}$ is a constant set to close to one; let $s_{\max} = F_{\boldsymbol{\theta}}^{-1}(p_{\max})$. Set $s_j = F_{\boldsymbol{\theta}}^{-1}(p_j)$, $j = 1, \ldots, J-k+1$. The proposed method automatically allocates more knots in regions of higher probability mass under the parametric family, and it works very well in simulations and our data analyses. To ensure a positive hazard over $(0, \infty)$, we assume $\lambda_0(t) = \lambda_0(s_{\max})$ for $t > s_{\max}$, implying a flat hazard past where the bulk of the data lie under the parametric model. Define $\tilde{s}_j = \sum_{l=j+1}^{j+k} s_l/(k-1)$. By Schoenberg's approximation theorem (Marsden, 1972), $\lambda_0(t) = \sum_{j=1}^{J} \lambda_{\boldsymbol{\theta}}(\tilde{s}_j) B_{kj}(t)$ approximates $\lambda_{\boldsymbol{\theta}}(t)$ over $t \in (0, s_{\max})$ with uniformly bounded error, that is, $\max_{0 \le t \le s_{\max}} ||\lambda_0(t) - \lambda_{\boldsymbol{\theta}}(t)|| \le 2\max\{|\lambda_{\boldsymbol{\theta}}(x) - \lambda_{\boldsymbol{\theta}}(y)| : |x-y| \le \min\{s_{\max}/\sqrt{2k-2}, \max_j\{s_{j+1} - s_j\}\sqrt{k/12}\}\}$.

To center $\lambda_0(t)$ at $\lambda_{\boldsymbol{\theta}}$ in our Bayesian framework, we take the prior mean $E(b_j) = \lambda_{\boldsymbol{\theta}}(\tilde{s}_j)$, specifically $b_j \sim \text{Gamma}\,(c\lambda_{\boldsymbol{\theta}}(\tilde{s}_j), c)$ where the scalar c controls the how stochastically "close" λ_0 is to $\lambda_{\boldsymbol{\theta}}$ under the prior and is assigned a prior $\Gamma(a_c, b_c)$. The distribution family $F_{\boldsymbol{\theta}}$ anchors the prior shape of λ_0. A data-driven prior for $\boldsymbol{\theta}$ is assumed by first obtaining maximum likelihood estimates $\widehat{\boldsymbol{\theta}}$ under the underlying parametric EH model, and its associated inverse information matrix $\mathbf{V}_{\boldsymbol{\theta}}$; $\boldsymbol{\theta}$ is then assigned Gaussian prior $N(\hat{\boldsymbol{\theta}}, a_{\boldsymbol{\theta}}\mathbf{V}_{\boldsymbol{\theta}})$ where $a_{\boldsymbol{\theta}} > 1$ is a scalar. The number of B-spline basis function J is typically chosen between 20 and 40; see Ruppert (2002) for a detailed discussion on selection of J. In summary,

$$\begin{aligned} b_j &\overset{\text{ind.}}{\sim} \text{Gamma}\,(c\lambda_{\boldsymbol{\theta}}(\tilde{s}_j), c)\,,\ j = 1, \ldots, J; \\ c &\sim \text{Gamma}\,(a_c, b_c);\ \boldsymbol{\theta} \sim N_2(\hat{\boldsymbol{\theta}}, a_{\boldsymbol{\theta}}\mathbf{V}_{\boldsymbol{\theta}}). \end{aligned} \tag{6}$$

Sharef et al. (2010) posit a hazard model that is a weighted sum of a parametric hazard and a penalized B-spline to induce shrinkage toward a specified parameter target. In contrast, the prior we suggest directly shrinks the B-spline toward a parametric target. The parametric target both centers inference and also guides knot locations.

2.2. *MCMC Sampling*

Denote S and S^c as the sets of observed and censored subjects, respectively. For notational simplicity, we omit k in B_{kj}. To facilitate the sampling of $\mathbf{b}$, we follow Lin and Wang (2011), and introduce binary latent variables $\{u_{ij} \in \{0,1\}, j = 1, \ldots, J\}$ for uncensored subjects. Let $\mathbf{u}_i = (u_{i1}, \ldots, u_{iJ})$ for $i \in S$ and $\mathbf{u} = (\mathbf{u}_i, i \in S)$. The augmented likelihood $L^A(\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{b}, \boldsymbol{\theta}, \mathbf{u})$ is

$$\begin{aligned} &\prod_{i \in S} \left\{ \mathrm{e}^{\boldsymbol{\gamma}'\mathbf{z}_i} \prod_{1}^{J} \left[b_j B_j(\mathrm{e}^{\boldsymbol{\beta}'\mathbf{z}_i} t_i) \right]^{u_{ij}} I\left(\sum_{j=1}^{J} u_{ij} = 1 \right) \right\} \\ &\times \exp\left\{ -\sum_{i=1}^{n} \mathrm{e}^{\boldsymbol{\gamma}'\mathbf{z}_i} \int_0^{t_i} \sum_{j=1}^{J} b_j B_j(t\mathrm{e}^{\boldsymbol{\beta}'\mathbf{z}_i}) \mathrm{d}t \right\} \end{aligned} \tag{7}$$

where $I(\cdot)$ is an indicator function. The joint posterior $p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{b}, \boldsymbol{\theta}, c, g_1, g_2 | \mathcal{D})$ following the augmented likelihood (7) and priors (4) and (6) is proportional to

$$L^A(\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{b}, \boldsymbol{\theta}, \mathbf{u}) p(\boldsymbol{\beta}|g_1) p(\boldsymbol{\gamma}|g_2) p(\boldsymbol{\theta}) p(c) p(g_1) p(g_2) \prod_{j=1}^{J} p(b_j | \boldsymbol{\theta}, c), \tag{8}$$

where $\mathbf{b} = (b_1, \ldots, b_J)$. One iteration of the MCMC algorithm follows.

Step 1: Update the blocks $\{\boldsymbol{\beta}, \boldsymbol{\gamma}\}$, $\boldsymbol{\theta}$, c separately using adaptive Metropolis–Hastings steps (Haario, Saksman, and Tamminen, 2005).

Step 2: Sample g_1^{-1} from $\text{Gamma}(a_g + p/2, \boldsymbol{\beta}'\mathbf{Z}'\mathbf{Z}\boldsymbol{\beta}/2n + b_g)$ and g_2^{-1} from $\text{Gamma}(a_g + p/2, \boldsymbol{\gamma}'\mathbf{Z}'\mathbf{Z}\boldsymbol{\beta}/2n + b_g)$.

Step 3: Sample the latent random vectors $\mathbf{u}_i$ from

$$\text{Multinomial}\left(\frac{b_1 B_1(\mathrm{e}^{\boldsymbol{\beta}'\mathbf{z}_i} t_i)}{\sum_{j=1}^{n} b_j B_j(\mathrm{e}^{\boldsymbol{\beta}'\mathbf{z}_i} t_i)}, \cdots, \frac{b_J B_J(\mathrm{e}^{\boldsymbol{\beta}'\mathbf{z}_i} t_i)}{\sum_{j=1}^{n} b_j B_j(\mathrm{e}^{\boldsymbol{\beta}'\mathbf{z}_i} t_i)} \right).$$

Step 4: Sample B-spline coefficient b_j from

$$\text{Gamma}\left(\sum_{i \in S} u_{ij} + c\lambda_{\boldsymbol{\theta}}(\tilde{s}_j), c + \sum_{i=1}^{n} \mathrm{e}^{\boldsymbol{\gamma}'\mathbf{z}_i} \int_0^{t_i} B_j(t\mathrm{e}^{\boldsymbol{\beta}'\mathbf{z}_i}) \mathrm{d}t \right).$$

Updating the baseline hazard using the augmented latent variable vector $\mathbf{u}$ and adaptive Metropolis–Hastings algorithms, coupled with a data-driven prior on $\boldsymbol{\theta}$ has given very efficient MCMC chains. More details on computation time are presented in the Supplementary Materials.

3. Spatial Correlation

The normal transformation model of Li and Lin (2006) is extended here to areal data. Relevant articles include Li and Rahman (2011) and Smith (2013). Both approaches use a multivariate normal for the transformed responses coupled with latent data; for them a latent continuous variable underlies a discrete response, for us a latent censored survival time.

3.1. *Likelihood of the Normal Transformation Model*

Under the the normal transformation model (2), Li and Lin's (2006) likelihood simplifies to

$$\begin{aligned} L_s(\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{b}, \boldsymbol{\theta}, \boldsymbol{\Gamma}) = \int &\left[\prod_{i \in S} \frac{f_i(t_i)}{\phi(y_i)} \right] \left[\prod_{i \in S^c} \frac{f_i(x_i)}{\phi(y_i)} I(x_i > t_i) \right] \\ &\times \phi(\mathbf{y}; \mathbf{0}, \boldsymbol{\Gamma}) \prod_{i \in S^c} \mathrm{d}x_i, \end{aligned} \tag{9}$$

where $\phi(\cdot)$ is standard normal density,

$$y_i = \begin{cases} \Phi^{-1}(F_i(t_i)), & i \in S \\ \Phi^{-1}(F_i(x_i)), & i \in S^c \end{cases},$$

$\mathbf{y} = (y_1, \ldots, y_n)$, and $\boldsymbol{\Gamma}$ is a positive definite matrix with diagonal elements being one. Note that this joint density is

identical to that obtained from a Gaussian copula model on the X_i. It is difficult to integrate out $\{x_i, i \in S^c\}$ both theoretically and numerically.

3.2. *CAR and ICAR Correlation Structures*

In this section, we construct the correlation matrix $\mathbf{\Gamma}$ so that the transformed survival times follow in the spatial transformation model (2) for areal data, based on intrinsic conditionally autoregressive (ICAR) model (Besag, York, and Mollie, 1991; Banerjee, Carlin, and Gelfand, 2004). The framework developed here allows immediate extension to conditionally autoregressive (CAR) model and exchangeable correlation structures as well.

Suppose subjects come from m areal units where some areal units share boundaries and some do not. Assume $\mathbf{W} = (w_{ij})$ where $w_{ij} = 1$ if areal unit i is adjacent to areal unit j, and $w_{ij} = 0$ if they are not adjacent. Customarily, $w_{ii} = 0$ for $i = 1, \ldots, m$. Let d_i be the total number units adjacent to unit i, that is, $d_i = \sum_{j=1}^{m} w_{ij}$. Denote $\mathbf{D} = \text{diag}(d_1, \ldots, d_m)$. Let $\widetilde{Y}_{ij}$ be a normal random variable for the jth individual in unit i. Let $\widetilde{\mathbf{Y}}_i = (\widetilde{Y}_{i1}, \ldots, \widetilde{Y}_{in_i})$ and $\widetilde{\mathbf{Y}} = (\widetilde{\mathbf{Y}}'_1, \ldots, \widetilde{\mathbf{Y}}'_m)$ where n_i is the number of observations in unit i. Let $\mathbf{Y}$ be the vector of transformed failure times $\{Y_i = \Phi^{-1}(F_i(T_i)), i = 1, \ldots, n\}$, sorted so that they correspont to the elements of $\widetilde{\mathbf{Y}}$. Assume $\mathbf{\Gamma} = cov(\mathbf{Y}) = cov(\widetilde{\mathbf{Y}})$.

To induce marginal ICAR correlation on $\widetilde{\mathbf{Y}}$, first consider the random effects model:

$$\widetilde{Y}_{ij} = \alpha_i + \epsilon_{ij};\ \boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)' \sim N(\mathbf{0}, \mathbf{B\Omega B});$$
$$\epsilon_{ij} \overset{iid}{\sim} N\left(0, \frac{\sigma^2}{\sqrt{\omega_{ii} + \sigma^2}}\right), \tag{10}$$

where $\alpha_i, i = 1, \ldots, m$ are the random effects, $\Omega = (\omega_{ij})$ introduces spatial dependence to $\boldsymbol{\alpha}$, $\mathbf{B} = \text{diag}(1/\sqrt{\omega_{11} + \sigma^2}, \ldots, 1/\sqrt{\omega_{mm} + \sigma^2})$, and ϵ_{ij} is the error term for the jth subject in the unit i, independent of other error terms and the spatial random effects. Note that $Var(\widetilde{Y}_{ii}) = 1$. Popular models for $\mathbf{\Omega}$ include independent normals (i.e., exchangeable), conditional autoregressive (CAR) models, ICAR, simultaneous autoregressive models (SAR), and many others. Assume $\boldsymbol{\eta}$ is a random vector with $\boldsymbol{\eta} \sim N(\mathbf{0}, \mathbf{\Omega})$. Under the proper CAR model, $\eta_j | \boldsymbol{\eta}_{-j}, \varphi \sim N(r \sum_{j=1}^{n} w_{ij}\eta_j / w_{j+}, 1/(\varphi w_{j+}))$, which implies $\mathbf{\Omega} = \varphi^{-1}(\mathbf{D} - r\mathbf{W})^{-1}$. When $r = 0$, η_i is independent of η_j if $j \neq i$. It is difficult to estimate r and φ simultaneously, as noticed by many authors. One way is to fix $r = 1$, which leads to ICAR model. However, the implied covariance matrix of $\boldsymbol{\eta}$ under ICAR model is improper as $\mathbf{D} - \mathbf{W}$ is singular. A common strategy to restore the propriety (without standardization) is imposing a constraint $\sum_{j=1}^{m} \eta_j = 0$ during Gibbs sampling. In Appendix, we derive the implied covariance matrix under this constraint and yield $cov(\boldsymbol{\eta}) = \varphi^{-1}\Omega^*$, where $\Omega^* = (\omega^*_{ij})$ only depends on matrices $\mathbf{W}$ and $\mathbf{D}$. Then replace ω_{ii} by ω^*_{ii} in $\mathbf{B}$. The implied covariance matrix $\mathbf{\Gamma}$ under the ICAR model only involves one unknown quantity $\varphi\sigma^2$, that is, $cov(\widetilde{Y}_{ij}, \widetilde{Y}_{kl}) = cov(\alpha_i, \alpha_k) = \omega^*_{ik}/\sqrt{(\omega^*_{ii} + \varphi\sigma^2)(\omega^*_{kk} + \varphi\sigma^2)}$. Denote $\varphi^* = \varphi\sigma^2$. A smaller value of φ^* corresponds to stronger spatial dependence within counties and across counties.

3.3. *Latent Survival Times Approach*

As mentioned in Section 3.1, it is difficult to evaluate the likelihood function (9). In this section, we introduce a latent failure time X_i for each censored observation. The augmented likelihood with sampled latent failure time variables $\{x_i, i \in S^c\}$ for the spatial model is

$$L^A_s(\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{b}, \boldsymbol{\theta}, \mathbf{\Gamma}, \{x_i, i \in S^c\}) = \left[\prod_{i \in S} \frac{f_i(t_i)}{\phi(y_i)}\right]\left[\prod_{i \in S^c} \frac{f_i(x_i)}{\phi(y_i)} I(x_i > t_i)\right] \phi(\mathbf{y}; \mathbf{0}, \mathbf{\Gamma}). \tag{11}$$

Note that each y_i is a function of t_i for $i \in S$ and a function of x_i for $i \in S^c$. Based on the augmented likelihood (11) and the priors (6) and (4), MCMC sampling steps 1–3 are similar as those in Section 2.2, except that now we introduce latent binary random vector $\mathbf{u}_i$ for all subjects. Sampling $\{\mathbf{b}, \varphi^*\}$ is accomplished as follows.

Step 4: Propose $b_{j(\text{new})}$ from

$$\text{Gamma}\left(\sum_{i=1}^{n} u_{ij} + c\lambda_{\boldsymbol{\theta}}(\tilde{s}_j), c + \sum_{i=1}^{n} e^{\boldsymbol{\gamma}'\mathbf{z}_i} \int_0^{t_i} B_j(te^{\boldsymbol{\beta}'\mathbf{z}_i})\text{d}t\right)$$

and accept it with probability

$$\min\left\{1, \frac{e^{-\mathbf{y}_{\text{new}}{}'\Gamma^{-1}\mathbf{y}_{\text{new}}/2} \prod_{i=1}^{n} \phi(y_i)}{e^{-\mathbf{y}'\Gamma^{-1}\mathbf{y}/2} \prod_{i=1}^{n} \phi(y_{i(\text{new})})}\right\},$$

where $\mathbf{y}_{\text{new}}$ is new transformed failure time vector corresponding to $b_{j(\text{new})}$. Evaluations of $\mathbf{\Gamma}^{-1}$ and $\mathbf{y}'\mathbf{\Gamma}^{-1}\mathbf{y}$ are efficiently carried out in Appendix.

Step 5: Sample $Y_i \sim N(y_i | \mathbf{y}_{-i}, \mathbf{\Gamma}) I(y_i > \Phi^{-1}(F_i(t_i)))$ (e.g., Geweke, 1991) where $\mathbf{y}_{-i} = (y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)$. Then set $x_i = F_i^{-1}(\Phi(y_i))$ using bisection or the Newton–Raphson algorithm.

Step 6: Update φ^* using adaptive Metropolis–Hastings method.

The latent survival approach is computationally straightforward and can accommodate large datasets. However, the imputation of latent failure becomes inefficient as the number of censored failure times increases. In the next section, we propose an alternative approach.

3.4. *Random-Effect Approach for Lattice Data*

Let $\widetilde{T}_{ij}$ be the failure time for the jth observation in county i and $\tilde{t}_{ij}$ be the observed event time for $\widetilde{T}_{ij}$. Conditional on random effect α_i, the likelihood contribution for a censored observation in areal unit i is

$$S_{ij}(\tilde{t}_{ij} | \alpha_i) = P(\widetilde{T}_{ij} > \tilde{t}_{ij} | \alpha_i) = 1 - \Phi\left(\frac{\Phi^{-1}(F_{ij}(\tilde{t}_{ij})) - \alpha_i}{\sqrt{(\sigma^2/(\omega_{ii} + \sigma^2))}}\right),$$

where $F_{ij}(\cdot)$ is the cumulative distribution function associated with $\widetilde{T}_{ij}$. For an observed observation the likelihood contribution is $f_{ij}(\tilde{t}_{ij} | \alpha_i) = -\partial S_{ij}(\tilde{t}_{ij} | \alpha_i)/\partial \tilde{t}_{ij}$ Under the ICAR prior, $\sigma^2/(\omega_{ii} + \sigma^2) = \varphi^*/(\omega^*_{ii} + \varphi^*)$. Survival probability S_{ij}

Table 1
Summary characteristics of prostate cancer patients in SC from 1996 to 2004

Covariate		n	Sample percentage
Race	Black	6483	0.32
	White	14,116	0.68
Marital status	Non-married	4525	0.22
	Married	16,074	0.78
Grade	well or moderately differentiated	15,309	0.74
	poorly differentiated or undifferentiated	5290	0.26
SEER summary stage	Localized or regional	19,792	0.96
	Distant	807	0.04

increases as α_i increases, holding φ^* constant. With the addition of $\boldsymbol{\alpha}$, the joint likelihood (9) is written as

$$L_s(\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{b}, \boldsymbol{\theta}, \boldsymbol{\Gamma}) = \int \prod_{(i,j)\in S} f_{ij}(\tilde{t}_{ij}|\alpha_i) \times \prod_{(i,j)\in S^c} S_{ij}(\tilde{t}_{ij}|\alpha_i) P(\alpha_1, \ldots, \alpha_m) \mathrm{d}\alpha_1 \ldots \mathrm{d}\alpha_m, \tag{12}$$

where $P(\alpha_1, \ldots, \alpha_m)$ is the joint density of $\boldsymbol{\alpha}$. The dimension of integration in (12) is typically much lower than that in (10) for highly censored data. The sampling steps for $\mathbf{b}$ and $\alpha_1, \ldots, \alpha_m$ are carried out through adaptive Metropolis–Hastings steps.

4. Data Analysis

In our supplementary materials, we include simulations validating our approach to the EH and spatial EH models. The latter is fitted using the latent survival times approach. For the simulated data, model parameters and the survival functions are estimated with low bias and coverage probabilities for model parameters are consistently close to the nominal levels. The 95% credible intervals for the estimated survival functions contain the true functions when choosing $F_{\boldsymbol{\theta}}$ to be either log-logistic or log-normal.

Using the proposed methods, we now analyze SCCCR prostate cancer data for the period 1996–2004. Covariates at diagnosis include county of residence, age (standaradized in the analyses), race, marital status, grade of tumor differentiation, and SEER summary stage. Table 1 provides summaries for the categorical covariates. There are $n = 20,599$ patients in the dataset after excluding subjects with missing information; 72.3% of the survival times are right-censored.

One purpose of the study is to quantify racial disparity in prostate cancer survival, adjusting for the remaining risk factors while accounting for the county where the subject lives in. We expect patients residing in the nearby counties to be correlated due to similarities in access to health care and socioeconomic factors. Mortality rates (percentages of death) for each county based on the SCCCR prostate cancer data for the period 1996–2004 are mapped in Figure 1, which suggests strong spatial patterns in the northwestern and eastern parts of South Carolina. A test for the PH assumption using methods proposed in Grambsch and Therneau (1994) yield a global p-value less than 0.01, that is, PH is rejected. We first fit the EH model with the following specifications: $F_{\boldsymbol{\theta}}$ is the log-logistic distribution, which provides the best fit to a parametric PH, AFT, or EH model compared to fits using other commonly used parametric hazard families, $\boldsymbol{\theta} \sim N(\widehat{\boldsymbol{\theta}}, a_{\boldsymbol{\theta}}\widehat{\Sigma}_{\hat{\theta}})$ where $\widehat{\boldsymbol{\theta}}$ and $\widehat{\Sigma}_{\hat{\theta}}$ are obtained under the underlying log-logistic model, $\boldsymbol{\beta} \sim N(\mathbf{0}, g_1 n(\mathbf{Z}'\mathbf{Z})^{-1})$, $\boldsymbol{\gamma} \sim N(\mathbf{0}, g_2 n(\mathbf{Z}'\mathbf{Z})^{-1})$, $g_1^{-1}, g_2^{-1} \overset{\text{ind.}}{\sim} \text{Gamma}(a_g, b_g)$, $a_g = b_g = 0.1$, $a_{\boldsymbol{\theta}} = 1000$, $a_c = b_c = 0.1$, and $J = 30$.

Hanson, Branscum, and Johnson (in press) explore the use of the g-prior in a generalized linear model setting in detail. They view the g-prior as a population-averaged prior and detail its use in providing informative prior information on ob-

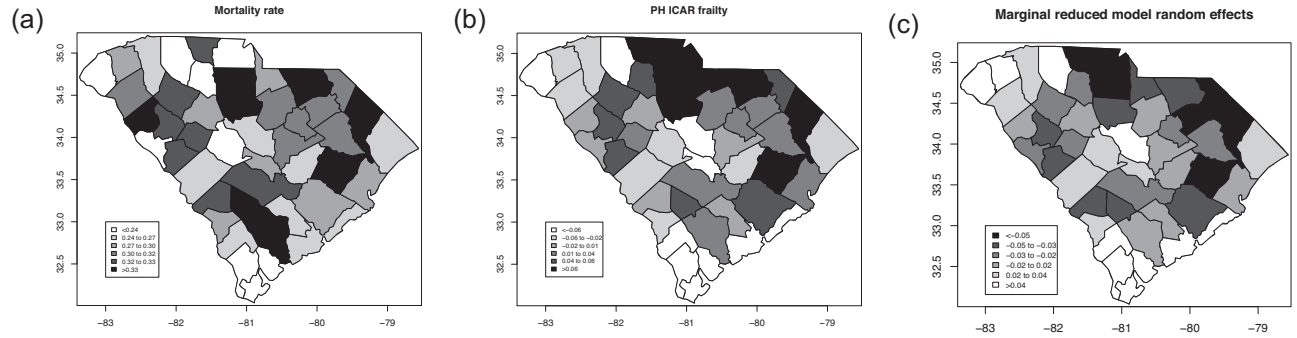

Figure 1. Map of (a) mortality rate, (b) ICAR frailties in the spatial PH model and (c) random effects in the marginal reduced model for the SCCCR data. Larger values of frailties in (b) correspond to higher risk of hazard function; larger values of random effects in (c) are related to higher survival probabilities.

Table 2
Summary of fitting EH, the reduced model, AFT, and PH; ∗ indicates LPML—21,000 and DIC—42,000

Covariates		EH	Reduced	AFT $\boldsymbol{\beta} = \boldsymbol{\gamma}$	PH $\boldsymbol{\beta} = 0$	PH + additive age $\boldsymbol{\beta} = 0$
Age	β_1	0.50 (0.48,0.52)	0.48 (0.46,0.50)	0.48 (0.45,0.51)	–	
	γ_1	0.45 (0.42,0.49)	$\gamma_1 = \beta_1$	–	0.65 (0.62,0.68)	–
Race	β_2	0.18 (0.15,0.21)	0.20 (0.16,0.21)	0.18 (0.15,0.22)	–	–
	γ_2	0.18 (0.12,0.24)	$\gamma_2 = \beta_2$	–	0.26 (0.21,0.32)	0.26 (0.20,0.31)
Marital status	β_3	−0.06 (−0.11, −0.02)	−0.05 (−0.09, −0.00)	0.26 (0.21,0.30)	–	–
	γ_3	0.35 (0.29,0.40)	0.33 (0.28,0.40)	–	0.33 (0.27,0.39)	0.31 (0.26,0.37)
Grade	β_4	0.03 (−0.02, 0.08)	$\beta_4 = 0$	0.27 (0.22,0.32)	–	–
	γ_4	0.36 (0.29,0.41)	0.37 (0.31,0.43)	–	0.38 (0.32,0.44)	0.37 (0.33,0.43)
SEER stage	β_5	3.19 (2.80,3.53)	3.27 (2.79,3.57)	1.50 (1.41,1.59)	–	–
	γ_5	1.02 (0.83,1.20)	1.00 (0.82,1.19)	–	1.56 (1.47,1.64)	1.57 (1.19,1.65)
$LPML^*$		−161.0	−162.0	−206.5	−242.5	−231.9
DIC^*		267.7	270.7	366.0	443.0	412.8

servables of interest. For us, these are the acceleration factors e^{β_j} and hazard ratios e^{γ_j}. The simulated 60% (20–80%) credible intervals for age, race, marital status, grade, and SEER stage using the g-prior are $(-24, 24)$, $(-50, 49)$, $(-52, 58)$, and $(-53, 50)$, and $(-117, 146)$ for our data. These prior ranges for the log-acceleration factors and log-hazard ratios under the g-prior allow for very weak or very strong covariate effects.

For all analyses we take a burn-in of 2000, followed by 6000 iterates after thinning every other five for the SCCCR data. The column under EH in Table 2 gives the fitted results. Tests on null hypotheses $H_0 : \boldsymbol{\beta} = 0$, $H_0 : \boldsymbol{\gamma} = 0$, and $H_0 : \boldsymbol{\beta} = \boldsymbol{\gamma}$ lead to global comparisons of EH to PH, AH and AFT model, respectively. Let M_1 denote the model under the null hypothesis with parameter Υ and M_2 denote EH under the alternative with parameters $(\boldsymbol{\tau}, \Upsilon)$. Our M_1 models are restricted versions of M_2, for example, when M_1 is PH, let $\boldsymbol{\tau} = \boldsymbol{\beta}$, $\Upsilon = (\boldsymbol{\gamma}, \mathbf{b}, \boldsymbol{\theta}, c, g_2)$ and then $f_1(\mathcal{D}|\Upsilon, M_1) = f_2(\mathcal{D}|\boldsymbol{\tau} = \mathbf{0}, \Upsilon, M_2)$ where $f_1(\mathcal{D}|\cdot)$ and $f_2(\mathcal{D}|\cdot)$ are the likelihoods of the data $\mathcal{D}$ under models M_1 and M_2, respectively. The Bayes factor for comparing models M_1 and M_2 is $BF_{12} = \int f_1(\mathcal{D}|\Upsilon, M_1)\pi_1(\Upsilon)\mathrm{d}\Upsilon / \int f_2(\mathcal{D}|\Upsilon, \boldsymbol{\tau}, M_2)\pi_2(\Upsilon, \boldsymbol{\tau})\mathrm{d}\Upsilon d\boldsymbol{\tau}$, where $\pi_1(\Upsilon)$ and $\pi_2(\Upsilon, \boldsymbol{\tau})$ are the prior densities. We assume $\pi_1(\Upsilon) = \pi_2(\Upsilon|\boldsymbol{\tau} = \mathbf{0})$, so the Bayes factor BF_{12} for comparing M_1 to M_2 is reduced to a Savage–Dicky ratio, $BF_{12} = \pi_2(\boldsymbol{\tau} = \mathbf{0}|\mathcal{D})/\pi_2(\boldsymbol{\tau} = \mathbf{0})$ (Verdinelli and Wasserman, 1995) where $\pi_2(\boldsymbol{\tau}|\mathcal{D})$ is the posterior distribution and $\pi_2(\boldsymbol{\tau})$ the marginal prior distribution of $\boldsymbol{\tau}$ under M_2. We use MCMC iterates to obtain the posterior ordinates based on normal approximations (Verdinelli and Wasserman, 1995). Similarly, we can also compute per-variable Bayes factors for testing $H_0 : \beta_j = 0$, $H_0 : \gamma_j = 0$, and $H_0 : \beta_j = \gamma_j$ which correspond to letting $\tau = \beta_j$, $\tau = \gamma_j$, and $\tau = \beta_j - \gamma_j$, respectively. The per-variable Bayes factors provide further guidance on choosing sub-models.

The Bayes factors for globally testing EH versus PH, AH, and AFT are much greater than 1000, indicating evidence against those commonly assumed models. Variable-specific Bayes factors in Table 3 under EH indicate evidence favoring a reduced model with AFT components for age and race, EH components for marital status and SEER stage, and a PH component for grade. We compared this reduced model, the general EH model, AFT, and PH models using the same prior specifications, except that for PH, knots are fixed and equally spaced—a common way for fitting PH model (Hennerfeind et al., 2006). The results are displayed in Table 2. The LPML and DIC statistics indicate that EH and the reduced model outperform AFT or PH. We carried out a prior sensitivity analysis for the EH model; there is very little difference in parameter estimates, LPML, and DIC for the alternative priors $g_1^{-1} \sim \Gamma(0.001, 0.001)$, $g_2^{-1} \sim \Gamma(0.001, 0.001)$. We also tried fitting vaguely flat priors on $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ with little difference. There is also very little difference when increasing J to 50. The model appears to be robust to different prior settings for these data.

We further fit the EH model with spatial dependence (2) via the random-effect approach described in Section 3.4, due to a high percentage of censored observations and a large sample size. The fitted results are presented in Table 4. The Bayes factors after taking into account spatial correlation are presented Table 3 under the column Spatial + EH, implying the same reduced model as that in Table 2 under independence. Taking into account the spatial correlation significantly improves model fit according to LPML and DIC.

We also fit what might be considered state-of-the-art, a PH model with ICAR frailties and a B-spline transformation for age, that is, a partially linear Cox model with spatial frailties (e.g., Kneib and Fahrmeir, 2007). The nonlinear transforma-

Table 3
Bayes factors for comparing EH to PH, AFT, and AH with and without spatial correlation

	EH			Spatial+EH		
Covariate	PH	AFT	AH	PH	AFT	AH
Age	>1000	0.08	>1000	>1000	0.01	>1000
Race	>1000	0.01	>1000	>1000	<0.01	>1000
Marital status	1.79	>1000	>1000	1.18	>1000	>1000
Grade	0.14	>1000	>1000	0.08	>1000	>1000
SEER stage	>1000	>1000	>1000	>1000	>1000	>1000

Table 4
Summary of spatial models; ∗ indicates LPML—21,000 and DIC—42,000

Covariates		Marginal EH	Marginal reduced	PH+ICAR+additive age $\boldsymbol{\beta} = 0$
Age	β_1	0.50 (0.47,0.52)	0.47 (0.46,0.49)	–
	γ_1	0.46 (0.43,0.49)	$\gamma_1 = \beta_1$	–
Race	β_2	0.18 (0.15,0.21)	0.20 (0.17,0.22)	–
	γ_2	0.17 (0.11,0.23)	$\gamma_2 = \beta_2$	0.24 (0.18,0.30)
Marital status	β_3	−0.06 (−0.10, −0.02)	−0.02 (−0.05, −0.00)	–
	γ_3	0.34 (0.28,0.41)	0.33 (0.27,0.39)	0.32 (0.25,0.38)
Grade	β_4	0.03 (−0.01, 0.07)	$\beta_4 = 0$	–
	γ_4	0.36 (0.30,0.42)	0.38 (0.32,0.43)	0.37 (0.32,0.44)
SEER stage	β_5	3.16 (2.86,3.34)	2.77 (2.72,2.82)	–
	γ_5	1.10 (0.94,1.26)	1.21 (1.01,1.33)	1.55 (1.46,1.64)
φ^*		50.1 (19.9,113.7)	54.6 (22.7,120.8)	33.08 (9.2,100.1)
$LPML^*$		−142.7	−143.2	−215.7
DIC^*		192.4	164.0	332.5

tion of age improves model fit beyond a linear age effect in PH, and the inclusion of ICAR frailties improves model fit beyond the assumption of independence according to LPML and DIC. However, the independent EH and EH-reduced models outperform the PH model, even augmented with a nonlinear transformation of age and spatial ICAR frailties. The ICAR-normal transformation model improves model fit of EH and EH-reduced even further. Our findings agree with Zhao, Hanson, and Carlin (2009) in that the most important aspect affecting model fit and prediction is the overarching model tying covariates to survival; of lesser importance is the spatial aspect of the model. Here, an EH model with linear effects (and a more interpretable EH-reduced model) vastly outperforms the PH model with a nonlinear effect.

The random effects in the marginal EH-reduced model and the ICAR frailties in the spatial PH model are mapped in Figure 1. Note that the random effects have opposite interpretation from the PH frailties that smaller random effect indicates poorer survival. Both plots suggest similar spatial patterns to the map of mortality rates, however, the latter two plots adjust for other factors such as race, age, and aspects of the diagnosed cancer.

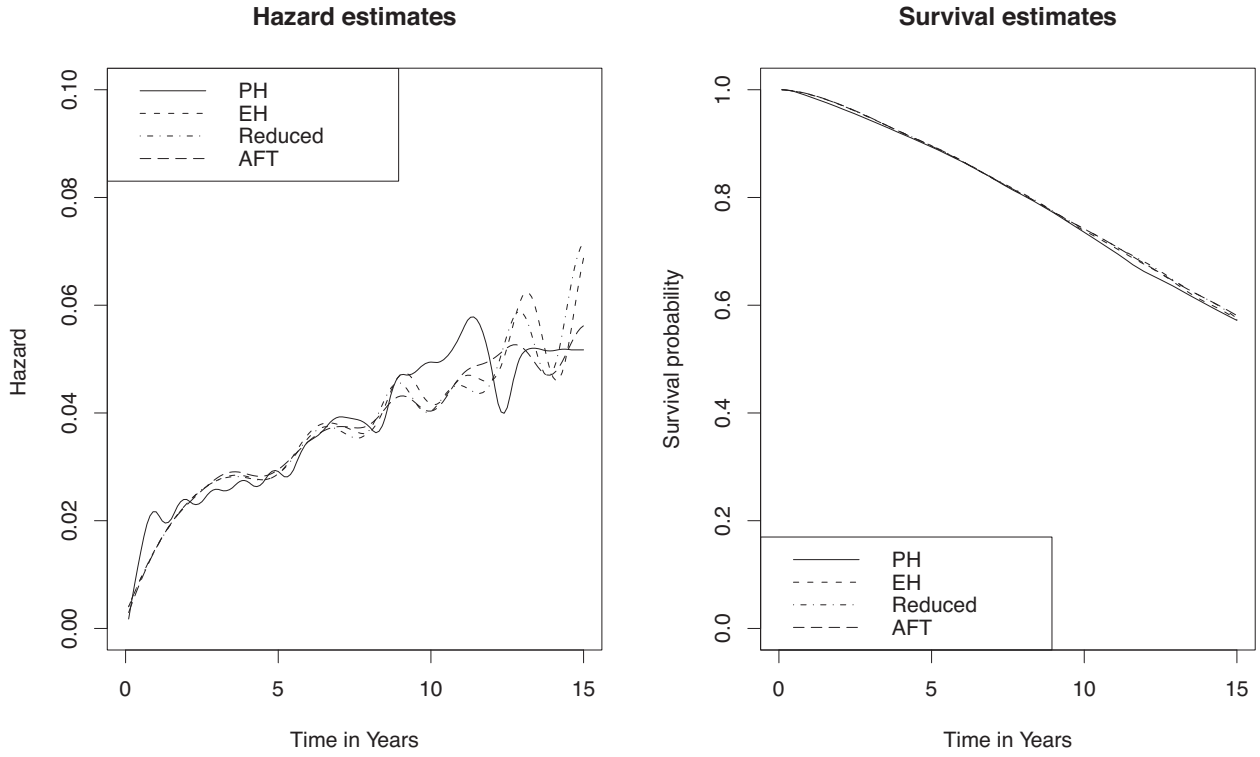

Figure 2. Baseline hazard (left) and survival function (right) estimates for the SCCCR data.

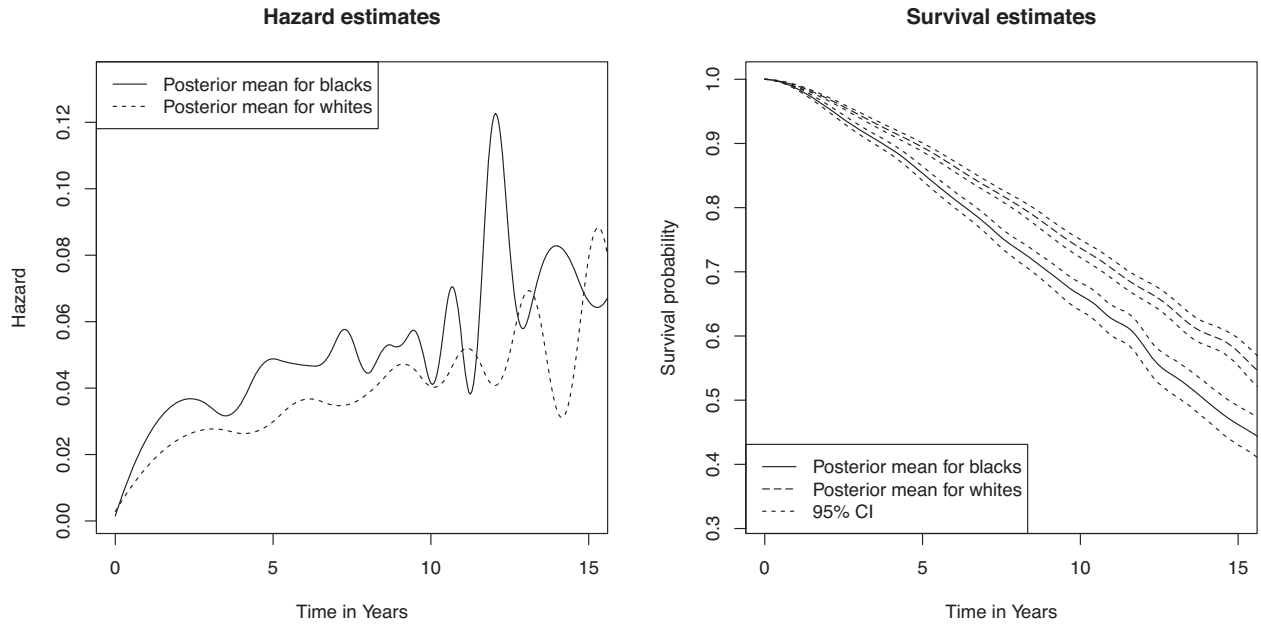

Figure 3. Hazard (left) and survival function (right) estimates for black patients (solid line) and white patients (long-dashed line). Short-dashed lines are 95% credible intervals.

Based on the fitted results of the reduced models with and without spatial dependence, white South Carolina subjects diagnosed with prostrate cancer in live 22% longer ($e^{0.20} \approx 1.22$) than black patients (95% CI is 18–25%), fixing age, stage, and SEER stage. Note that this interpretation pertains to individuals randomly selected from any South Carolina county, that is, is population-averaged. Cox–Snell residual plots (Cox and Snell, 1968) show major lack-of-fit of the PH model (not shown) while EH, the EH-reduced model, and AFT show no lack of fit. Finally, we plot the estimated baseline survival and hazard functions for PH, EH, AFT, and the reduced model in Figure 2. To compare the survival probabilities for white and black patients, in Figure 3 we plot the baseline hazard and survival function estimates for each race while setting age at the sample mean and other discrete covariates at the reference levels. Survival probabilities for black patients are significantly lower than those for white patients when other factors are fixed at the same levels. Note that the largest event time is only 12.2 years.

Since the sample standard deviation of age is 8.47 and $e^{0.47/8.47} \approx 1.054$, decreasing age by 1 year increases survival time by 5.4%. Taking $e^{0.38} \approx 1.46$ indicates that the hazard of dying increases 46% for poorly or undifferentiated grades vs. well or moderately differentiated, holding age, race, and SEER stage constant. For SEER stage, which has general EH effects, $e^{2.77} \approx 16$ (AH) and $e^{1.21} \approx 3.4$ (PH). Those with distant stage are at least three times worse in one-sixteenth of the time as those with localized or regional. Finally, in the reduced model marital status essentially has PH interpretation; single (including widowed or separated) subjects are $e^{0.33} \approx 1.39$ times more likely to die at any instant than married.

In summary, a Bayesian semiparametric method for fitting the EH model to data on South Carolina subjects diagnosed with prostate cancer is developed, and further generalized to include spatially correlated data through a normal transformation model. A novel B-spline prior on the baseline hazard is centered at a parametric scale-family, thus allowing baseline stretching or shrinking as necessary for the EH, AH, and AFT models. For lattice data, we introduce a marginal correlation matrix based on the ICAR prior to accommodate spatial correlation and construct two MCMC approaches for fitting the model. Our findings for the SCCCR data help further quantify racial differences in prostate cancer survival as well as indicate South Carolina counties with higher mortality for further etiologic research, adjusted for other risk factors.

5. Supplementary Materials

Web Appendix A, referenced in Section 4, and Fortran 90 codes to this Section, are available with this paper at the *Biometrics* website on Wiley Online Library.

Acknowledgements

This research is funded by Grant R03CA165110 and R03CA176739 from National Institutes of Health. The authors thank the South Carolina Central Cancer Registry for providing the data analyzed in this article: South Carolina Central Cancer Registry (SCCCR), Division of Public Health Statistics and Information Services, South Carolina Department of Health and Environmental Control. `http://www.scdhec.gov/administration/phsis/scccr/SCCCRmain.htm`

References

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, FL: Chapman & Hall/CRC Press.

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.

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

Bové, D. S. and Held, L. (2011). Hyper-g priors for generalized linear models. *Bayesian Analysis* **6**, 1–24.

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

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.

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

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

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

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

Cox, D. R. and Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society B* **30**, 248–275.

De Boor, C. (2001). *A Practical Guide to Splines. Applied Mathematical Sciences 27*. Berlin: Springer-Verlag.

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

Geweke, J. (1991). Efficient simulation from the multivariate normal and Student-t distributions subject to linear constraints and the evaluation of constraint probabilities. In *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, E. M. Keramidas (eds), 571–578. Fairfax Station, VA: Interface Foundation of North America, Inc.

Grambsch, P. and Therneau, T. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* **81**, 515–526.

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

Haario, H., Saksman, E., and Tamminen, J. (2005). Componentwise adaptation for high dimensional MCMC. *Computational Statistics* **20**, 265–273.

Hanson, T., Branscum, A., and Johnson, W. (in press). Informative *g*-priors for logistic regression. Bayesian Analysis **9**, 597–612.

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

Hurley, D. M., Ehlers, M. E., Mosley-Broughton, C. M., Bolick-Aldrich, S. W., and Savoy, J. E. (2009). *Cancer in South Carolina, USA 1996–2005: South Carolina Central Cancer Registry Ten Year Report*. Columbia, SC: South Carolina Department of Health and Environmental Control, Office of Public Health Statistics and Information Services, Central Cancer Registry.

Ibrahim, J. G., Chen, M. H., and Sinha, D. (2001). *Bayesian Survival Analysis*. New York: John Wiley & Sons, Ltd.

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

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

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

Li, P. and Rahman, M. A. (2011). Bayesian analysis of multivariate sample selection models using gaussian copulas. *Advances in Econometrics* **27**, 269–288.

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

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

Marsden, M. J. (1972). On uniform spline approximation. *Journal of Approximation Theory* **6**, 249–253.

Rathbun, S. L. and Fei, S. (2006). A spatial zero-inflated poisson regression model for oak regeneration. *Environmental and Ecological Statistics* **13**, 409–426.

Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics* **11**, 735–757.

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

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

Smith, M. S. (2011). Bayesian approaches to copula modelling. In P. Damien, P. Dellaportas, N. Polson, and D. Stephens (eds.), Hierarchical models and MCMC: A Tribute to Adrian Smith.

Tong, X., Zhu, L., Leng, C., Leisenring, W., and Robison, L. L. (in press). A general semiparametric hazards regression model: Efficient estimation and structure selection. *Statistics in Medicine* **32**, 4980–4994.

Tseng, Y.-K. and Shu, K.-N. (2011). Efficient estimation for a semiparametric extended hazards model. *Communications in Statistics—Simulation and Computation* **40**, 258–273.

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

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

Received January 2014. *Revised October* 2014.
Accepted October 2014.

Appendix

Covariance Matrix of ICAR Model under the Constraint

In the following, we derive the covariance matrix of $\boldsymbol{\eta}$ under the constraint $\sum_{j=1}^{m} \eta_j = 0$. Under the ICAR prior for $\boldsymbol{\eta}$,

$p(\boldsymbol{\eta}) \propto \exp(-\varphi\boldsymbol{\eta}'(\mathbf{D}-\mathbf{W})\boldsymbol{\eta}/2)$. Note that

$$\boldsymbol{\eta}'(\mathbf{D}-\mathbf{W})\boldsymbol{\eta} = \sum_{j=1}^{m} w_{j+}\eta_i^2 - \sum_{j=1}^{m}\sum_{i=1}^{m} w_{ji}\eta_i\eta_j. \quad (A.1)$$

Under the constraint $\sum_{j=1}^{m}\eta_j = 0$, let $\eta_m = -\eta_1 - \eta_2 - \cdots - \eta_{m-1}$ and plug it into (A.1), then

$$\boldsymbol{\eta}'(\mathbf{D}-\mathbf{W})\boldsymbol{\eta} = \sum_{j=1}^{m-1} w_{j+}\eta_i^2 - \sum_{j=1}^{m-1}\sum_{i=1}^{m-1}(w_{ji} - w_{jm} - w_{mi} - w_{m+})\eta_i\eta_j.$$

Let $\mathbf{D}^* = \mathrm{diag}(w_{1+}, \cdots, w_{(m-1)+})$, $\mathbf{W}^* = (w_{ij}^*)$ with $w_{ij}^* = w_{ji} - w_{jm} - w_{mi} - w_{m+}$. Let $\boldsymbol{\eta}^* = (\eta_1, \cdots, \eta_{m-1})$. Then under the constraint

$$\boldsymbol{\eta}'(\mathbf{D}-\mathbf{W})\boldsymbol{\eta} = \boldsymbol{\eta}^{*\prime}(\mathbf{D}^*-\mathbf{W}^*)\boldsymbol{\eta}^*. \quad (A.2)$$

If county m is adjacent to at least one county, $\mathbf{D}^*-\mathbf{W}^*$ is positive definite and hence $\mathrm{cov}(\boldsymbol{\eta}^*) = \varphi^{-1}(\mathbf{D}^*-\mathbf{W}^*)^{-1}$. Let $\boldsymbol{\Xi} = (\mathbf{D}^*-\mathbf{W}^*)^{-1}$ with elements (ξ_{ij}). Note that $\mathrm{cov}(\eta_m, \eta_i) = -\varphi^{-1}\sum_{j=1}^{m-1}\xi_{ij}$ and $\mathrm{var}(\eta_m) = -\varphi^{-1}\sum_{j=1}^{m-1}\sum_{i=1}^{m-1}\xi_{ij}$. Define

$$\Omega^* = \begin{pmatrix} \xi_{11} & \cdots & \xi_{1,m-1} & -\sum_{j=1}^{m-1}\xi_{1j} \\ \vdots & \vdots & \ddots & \vdots \\ \xi_{m-1,1} & \cdots & \xi_{m-1,m-1} & -\sum_{j=1}^{m-1}\xi_{m-1,j} \\ \sum_{j=1}^{m-1}\xi_{j1} & \cdots & \sum_{j=1}^{m-1}\xi_{j,m-1} & -\sum_{j=1}^{m-1}\sum_{i=1}^{m-1}\xi_{ij} \end{pmatrix}. \quad (A.3)$$

The covariance matrix of $\boldsymbol{\eta}$ under the constraint is $\phi^{-1}\Omega^*$.

Matrix Inversion

In the following, we find the inverse of $\boldsymbol{\Gamma} = cov(\widetilde{\mathbf{Y}})$. Let $\mathbf{J}_{n_i}$ and $\mathbf{J}_{n_in_j}$ be matrix of ones with dimension $n_i \times n_i$ and $n_i \times n_j$, respectively, $\mathbf{I}_{n_i}$ be an identity matrix with dimension $n_i \times n_i$, and $\mathbf{1}_{n_i}$ be a vector of ones with length n_i. Based on the random effects model (10), $\boldsymbol{\Gamma} = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}$ where $\mathbf{A} = \mathrm{blockdiag}\left(\sqrt{1/(\omega_{11}+\sigma^2)}\mathbf{I}_{n1}, \cdots, \sqrt{1/(\omega_{mm}+\sigma^2)}\mathbf{I}_{n_m}\right)$,

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{J}_{n_1}\omega_{11} + \mathbf{I}_{n_1}\sigma^2 & \mathbf{J}_{n_1n_2}\omega_{12} & \cdots & \mathbf{J}_{n_1n_m}\omega_{1m} \\ \mathbf{J}_{n_2n_1}\omega_{21} & \mathbf{J}_{n_2}\omega_{22} + \mathbf{I}_{n_2}\sigma^2 & \cdots & \mathbf{J}_{n_2n_m}\omega_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{J}_{n_mn_1}\omega_{m1} & \mathbf{J}_{n_mn_2}\omega_{m2} & \cdots & \mathbf{J}_{n_m}\omega_{mm} + \mathbf{I}_{n_m}\sigma^2 \end{pmatrix}. \quad (A.4)$$

Note that

$$\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Omega}\mathbf{P}' + \sigma^2\mathbf{I}_{n*}, \quad (A.5)$$

where $\mathbf{P} = \mathrm{blockdiag}\,(\mathbf{1}_{n_1}, \ldots, \mathbf{1}_{n_m})$ and $n* = \sum_{j=1}^{m} n_j$. Next we find a singular vector decomposition of $\mathbf{P}$. Define $l_{i1} = \sum_{j=1}^{i-1} n_j$, $l_{i2} = \sum_{j=i+1}^{m} n_j$, and $\tilde{\mathbf{u}}_i = (\mathbf{0}'_{l_{i1}}, \frac{1}{\sqrt{n_i}}\mathbf{1}'_{n_i}, \mathbf{0}'_{l_{i2}})'$ where $\mathbf{0}_{l_{i1}}$ is a vector of zeros with length l_{i1}. Define $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_{n^*})$ where $\mathbf{u}_i = \tilde{\mathbf{u}}_i$ for $i = 1, \ldots, m$ and $\mathbf{u}_{m+1}, \ldots, \mathbf{u}_{n^*}$ are the orthonormal expansion of $\mathbf{u}_1, \ldots, \mathbf{u}_m$. Define $\mathbf{U}_1 = (\mathbf{u}_1, \ldots, \mathbf{u}_m)$, $\mathbf{V} = \mathbf{I}_m$, $\mathbf{S}_0 = \mathrm{diag}(\sqrt{n_1}, \cdots, \sqrt{n_m})$, and $\mathbf{S} = (S_0, \mathbf{0}_{m\times n^*})'$ where $\mathbf{0}_{m\times n^*}$ is a matrix of zero with dimension $m \times n^*$. By singular vector decomposition, $\mathbf{P} = \mathbf{USV}'$.

Therefore, $\boldsymbol{\Sigma} = \mathbf{US}\boldsymbol{\Omega}\mathbf{SU}' + \sigma^2\mathbf{I}_{n*} = \mathbf{UKU}' + \sigma^2\mathbf{I}_{n*}$ where

$$\mathbf{K} = \begin{pmatrix} \mathbf{K}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \text{ and}$$

$$\mathbf{K}^* = \begin{pmatrix} n_1\omega_{11} & \sqrt{n_1n_2}\omega_{12} & \cdots & \sqrt{n_1n_m}\omega_{1m} \\ \sqrt{n_2n_1}\omega_{21} & n_2\omega_{22} & \cdots & \sqrt{n_2n_m}\omega_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{n_mn_1}\omega_{m1} & \sqrt{n_mn_2}\omega_{m2} & \cdots & n_m\omega_{mm} \end{pmatrix}.$$

Since $\mathbf{U}$ is orthonormal and $\boldsymbol{\Sigma} = \mathbf{U}\left(\mathbf{K} + \sigma^2\mathbf{I}_{n*}\right)\mathbf{U}'$, therefore,

$$\begin{aligned} \boldsymbol{\Sigma}^{-1} &= \mathbf{U}\left(\mathbf{K} + \sigma^2\mathbf{I}_{n*}\right)^{-1}\mathbf{U}' \\ &= \mathbf{U}\begin{pmatrix} (\mathbf{K}^* + \sigma^2\mathbf{I}_m)^{-1} & \mathbf{0} \\ \mathbf{0} & \sigma^{-2}\mathbf{I}_{(n^*-m)\times(n^*-m)} \end{pmatrix}\mathbf{U}' \\ &= \mathbf{U}\begin{pmatrix} (\mathbf{K}^* + \sigma^2\mathbf{I}_m)^{-1} - \sigma^{-2}\mathbf{I}_m & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{U}' + \sigma^{-2}\mathbf{I}_{n^*} \end{aligned}$$

and hence $\boldsymbol{\Sigma}^{-1} = \mathbf{U}_1\left((\mathbf{K}^* + \sigma^2\mathbf{I}_m)^{-1} - \sigma^{-2}\mathbf{I}_m\right)\mathbf{U}_1' + \sigma^{-2}\mathbf{I}_{n^*}$. Based on the definitions of $\mathbf{K}$ and $\mathbf{K}^*$, the determinant of $\boldsymbol{\Sigma}$ can be computed simply as $|\boldsymbol{\Sigma}| = |\mathbf{K}^* + \sigma^2\mathbf{I}_m|\sigma^{2(n^*-m)}$. Based on $\boldsymbol{\Sigma}^{-1}$,

$$\boldsymbol{\Gamma}^{-1} = \mathbf{A}^{-1}\mathbf{U}_1\left((\mathbf{K}^* + \sigma^2\mathbf{I}_m)^{-1} - \sigma^{-2}\mathbf{I}_m\right)\mathbf{U}_1'\mathbf{A}^{-1} + \sigma^{-2}\mathbf{A}^{-2}, \quad (A.6)$$

$$\mathbf{y}'\boldsymbol{\Gamma}^{-1}\mathbf{y} = \mathbf{x}'\left((\mathbf{K}^* + \sigma^2\mathbf{I}_m)^{-1} - \sigma^{-2}\mathbf{I}_m\right)\mathbf{x} + \sigma^{-2}\mathbf{y}'\mathbf{A}^{-2}\mathbf{y}, \quad (A.7)$$

where $x_i = \sqrt{(\omega_{ii}+\sigma^2)/n_i}\sum_{j=l_{i1}}^{l_{i2}} y_j$ and $\mathbf{x} = (x_1, \ldots, x_m)$; for our data $m = 46$. Note that $\mathbf{y}'\mathbf{A}^{-2}\mathbf{y}$ is a simple sum because $\mathbf{A}$ is diagonal.