

2014

Applications of Bayesian Nonparametrics to Reliability and Survival Data

Li Li

University of South Carolina - Columbia

Follow this and additional works at: <https://scholarcommons.sc.edu/etd>



Part of the [Statistics and Probability Commons](#)

Recommended Citation

Li, L. (2014). *Applications of Bayesian Nonparametrics to Reliability and Survival Data*. (Doctoral dissertation). Retrieved from <https://scholarcommons.sc.edu/etd/2670>

This Open Access Dissertation is brought to you by Scholar Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact digres@mailbox.sc.edu.

APPLICATIONS OF BAYESIAN NONPARAMETRICS TO RELIABILITY AND SURVIVAL
DATA

by

Li Li

Bachelor of Science
Wuhan University, 2009

Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in
Statistics
College of Arts and Sciences
University of South Carolina
2014

Accepted by:

Timothy Hanson, Major Professor

Edsel Peña, Committee Member

Jiajia Zhang, Committee Member

Lianming Wang, Committee Member

Lacy Ford, Vice Provost and Dean of Graduate Studies

© Copyright by Li Li, 2014
All Rights Reserved.

ACKNOWLEDGMENTS

I want to give whole-hearted thanks to my advisor Timothy Hanson. He is not only a great academic advisor, who passionately guided and inspired me in doing research, but also a good friend, who faithfully supported me in pursuing a professional career that makes the best of my talents. I also want to thank my committee members Edsel Pena, Jiajia Zhang and Lianming Wang for their support. I give special thanks to Dr. Pena; the classes he taught are among the best. I also thank many other people in the department for their support.

This dissertation would not have been possible without the love and blessings bestowed on me from the LORD who gave me a new life. I thank Him for giving me talent, perseverance, and strength to finish the dissertation. I also thank him for being my stronghold and vision forever. Because of Him, I am also blessed with the love from Ron and Marianne Parker, who are as dear as my natural parents. I want to thank them for their faithful support in times of good and bad. Also because of Him, I have many friends in the fellowship. I thank my friends for encouraging and helping me.

Finally, I gave my whole-hearted thanks to my parents and sister. They always show unconditional love to me and I am indebted to them. My parents poured whatever they have to help with my education and my happiness while asking for nothing. Their many encouragements are life-time treasures to me. I thank my sister for the joy she brought me every time we talked to each other.

ABSTRACT

Reliability and survival data are widely encountered across many common settings. Subjects under investigation often include machines, bioassays, patients, etc.; their reliability or survival distribution, and its association with covariate processes, are commonly of interest. Within this dissertation, the first two chapters focus on reliability data where repairable systems fail and get interventions, e.g. repairs in the event process. It begins with a nonparametric test for the commonly assumed “good as old” assumption for minimal repair models and then a semi-parametric regression model is introduced for reliability data using Kijima’s effective age. The third chapter focuses on survival data observed with potential spatial correlation. We first develop a Bayesian semi-parametric approach to the extended hazard model and then extend this framework to allow for spatial correlation among survival times. In contrast to widely used frailty models, our approach preserves marginal interpretations. Flexible modeling approaches in the Bayesian context are used for baseline failure rate or hazard and Markov chain Monte Carlo techniques to obtain the posterior inferences. The proposed tests and models are examined in several simulation studies and applications.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	iii
ABSTRACT	iv
LIST OF TABLES	vii
LIST OF FIGURES	ix
CHAPTER 1 INTRODUCTION	1
CHAPTER 2 A BAYESIAN NONPARAMETRIC TEST FOR MINIMAL RE- PAIR	17
2.1 Introduction	17
2.2 Model development	22
2.3 Simulations	34
2.4 Data analysis	40
2.5 Discussion	44
CHAPTER 3 A BAYESIAN SEMIPARAMETRIC REGRESSION MODEL FOR RELIABILITY DATA USING EFFECTIVE AGE	47
3.1 Introduction	47
3.2 Model development	51
3.3 Posterior inferences	56

3.4	Simulations	60
3.5	Data analysis	63
3.6	Discussion	71
CHAPTER 4 SPATIAL EXTENDED HAZARD MODEL WITH APPLICATION TO PROSTATE CANCER SURVIVAL		72
4.1	Introduction	72
4.2	Extended hazard model	76
4.3	Spatial correlation	79
4.4	Hypothesis tests using Bayes factors	83
4.5	Simulations	84
4.6	Data analysis	87
4.7	Conclusion	92
BIBLIOGRAPHY		97
APPENDIX A COPYRIGHT PERMISSION TO REPRINT CHAPTER 2 & 3		106
A.1	Chapter 2	106
A.2	Chapter 3	107

LIST OF TABLES

Table 2.1	Counts of perfect/minimal by response to “failure” / “censored” for the air conditioners.	22
Table 2.2	Type I error and power for testing H_0 vs. H_1 for simulation I; 6.0 – 2.0 are nine values of γ_2 in defining f_1 ; tabled values are the proportion out of 200 replications where H_0 is rejected.	36
Table 2.3	Type I error and power for testing H_0 vs. H_1 for simulation II; 0.2 – 1.0 represent nine choices of q and 1.0 represents no departure of minimal repair; tabled values are the proportion out of 200 replications where H_0 is rejected.	37
Table 2.4	Type I error for testing H_0 vs. H_1 for simulation III; 1 – 3 represents three choices of f_0 described in the text.	38
Table 2.5	Type I error and power for testing H_0 vs. H_1 with two sets of prior for c for simulation IV; tabled values are the proportion out of 100 simulated data sets where H_0 is rejected.	38
Table 2.6	Counts of perfect/minimal by response to “failure” / “censored” for each chiller.	41
Table 3.1	Summary of simulation studies: $f=0.5\text{Weibull}(2, 2)+0.5\text{Weibull}(2, 4)$; link function is logistic.	61
Table 3.2	Summary of simulation studies: $r(t)=\exp(t^2/3 + t/3)$; link function is logistic.	62
Table 3.3	Summary of simulation studies: $f=0.5\text{Weibull}(2, 2)+0.5\text{Weibull}(2, 4)$; link function is exponential.	62
Table 3.4	Summary of simulation studies: $r(t)=\exp(t^2/3 + t/3)$; $h(w_{ij2}) = \sum_{l=1}^5 b_l B_{l3}(w_{ij2})$	63
Table 3.5	Summaries of β_0 for Kijima type I and type II models for the valve seats maintenance data.	66

Table 3.6	Summary of the coefficients for Kijima type I and type II models for Syringe-driver maintenance data; $J = 5$, link function is exponential, estimates are posterior means, and 95% CIs are credible intervals.	68
Table 3.7	Goodness-of-fit measures for Kijima type I and type II model for Syringe-driver maintenance data; $J = 5$, and link function is exponential.	68
Table 3.8	Summary of the coefficients for Kijima type I model for Syringe-driver maintenance data; $h(w_2)$ is approximated by a B-spline, estimates are posterior means, and 95% CI are credible intervals.	70
Table 4.1	Summaries for Simulation I: $n = 300$ and censoring rates are 30% and 0%; baseline distribution is $0.5\text{lognormal}(1, 0.2)+0.5\text{lognormal}(2, 0.2)$; SSD: standard error of the posterior means; ESE: average of sample standard errors.	86
Table 4.2	Summaries for Simulation II: sample size $n = 500$ with 30% censored observations; baseline distribution is $0.5\text{lognormal}(1, 0.2)+0.5\text{lognormal}(2, 0.2)$; SSD: standard error of the posterior means; ESE: average of sample standard errors.	86
Table 4.3	Summary characteristics of prostate cancer patients in SC from 1996-2004.	87
Table 4.4	Summary of fitting the extended hazard model EH, the reduced model, AFT, and PH; * indicates $LPML - 21000$ and $DIC - 42000$	88
Table 4.5	Bayes factors for comparing EH to PH, AFT, and AH with and without spatial correlation.	88
Table 4.6	Summary of spatial models; * indicates $LPML - 21000$ and $DIC - 42000$	90

LIST OF FIGURES

Figure 2.1	(a) Weibull (α, γ) with $\alpha = 4$ and $\gamma = 4$; (b–g) tailfree densities, centered at (a) with conditional probabilities specified up to $J = 3$; (h) mixture of tailfree processes assuming $\alpha, \gamma \stackrel{ind.}{\sim} N(4, 0.05^2)$. 29
Figure 2.2	Left panel plots the hazard h_0 (solid and thick line) and 8 choice for h_1 (dashed and thin lines) versus time t for simulation I; right panel plots the intensity of the system versus time t when failures occur at $\{3, 6, 9\}$ for all q from 1 (solid and thick line) and $0.2 - 0.9$ (dashed and thin lines) for simulation II. 35
Figure 2.3	Results of a simulated sample of $n = 1000$ events under H_1 ; true (left) and estimated (right) survival and hazard estimates versus time t ; solid lines correspond to F_0 and short-dashed lines correspond to F_1 ; long-dashed lines correspond to 95% credible intervals. 39
Figure 2.4	Results of two simulated samples of $n = 1000$ interfailure times under Kijima type II model with $q = 0.2$ (left) and 0.5 (right); hazard estimates of h_0 (solid black) and h_1 (dashed black) versus time t ; dashed gray lines are intensities of ten systems. 40
Figure 2.5	Calendar times of events for the six chillers (AC); for each chiller, vertical bars represent observed failures on the top line and censored events on the bottom line; big (small) vertical bar denotes perfect (minimal) repair at the event time. 42
Figure 2.6	Kaplan-Meier estimates for survival functions using first failures after perfect repairs of the essential chillers system; group 1 and 2 (top left) ; AC 1 to 3 in group 1 (top right); AC 4 to 6 in group 2 (bottom left); bottom right panel plots M_1 estimates for S_0 for group 1 and 2. 43

Figure 2.7	This figure contains estimates of the survivor and hazard functions for group 1 (top panels) and 2 (bottom panels) essential chillers system when both parametric and nonparametric models are fitted for M_1 . Left panels plot nonparametric estimates of the survivor functions corresponding to F_0 (solid) and F_1 (short-dashed) and their 95% credible intervals (long-dashed). Right panels plot the parametric (smooth) and nonparametric (less smooth) estimates of the hazard functions corresponding to F_0 (solid) and F_1 (short-dashed).	45
Figure 3.1	Density and survival estimates for the simulated data sets with (a) $0.5\text{Weibull}(2, 2) + 0.5\text{Weibull}(2, 4)$ and (b) $h_0(t) = \exp(t^2/3 + t/3)$ based on Kijima type I model; the dark lines are the true density or survival functions and the gray lines are the point-wise posterior means.	63
Figure 3.2	Mean (gray-solid lines), 2.5% and 97.5% quantiles (gray-dashed lines) of the estimates for $h(w)$ based on the simulated datasets from the type I model with logistic (left) and exponential (right) links. The black solid lines are the true function $h(w) = -w^2$	64
Figure 3.3	Plots for valve seats maintenance data; left panel is the MCF plot for data where ‘+’ is the empirical point estimate and ‘×’ is its 95% confidence interval, overlaid with estimates of the mean function for NHPP (solid line), Kijima type I (short-dashed line) and Kijima type II (long-dashed line); right panel is the estimated baseline hazard function for Kijima type I (solid line), Kijima type II (dashed line) and NHPP (dotted line).	65
Figure 3.4	Plots for Syringe-driver maintenance data for Kijima type I (left) and type II (right) model; solid lines are baseline density (hazard) estimates and dashed lines are 95% credible intervals. Smooth estimates (dotted lines) are fitted from the parametric Weibull fit.	69
Figure 3.5	Estimate (solid) of $h(w_2)$ and its 95% credible intervals (dashed) for type I model with logistic (exponential) link on the left (right) panel; dotted lines are linear functions, i.e. $h(w_2) = -0.2w_2$ on the left panel.	70

Figure 4.1	Estimates and 95% credible intervals of the baseline survival, density, and hazard functions based on scenario (1) in simulation I with 0% censored subjects. F_{θ} is log-normal distribution for the left panels and log-logistic distribution for the right panels. Bold solid lines are the true functions; solid lines are the mean estimates; dashed lines are the 95% credible intervals; Bold dashed lines on the bottom panels are the mean functions of λ_{θ}	91
Figure 4.2	Map of (a) Mortality rate, (b) ICAR frailties in the PH model and (c) random effects in the marginal reduced model for SC counties. Larger values of frailties in (b) corresponds to higher risk of hazard function; larger values of random effects in (c) are related to higher survival probabilities.	92
Figure 4.3	Baseline hazard (left) and survival probabilities (right) estimates.	92
Figure 4.4	Baseline hazard (left) and survival probabilities (right) estimates for black patients (solid line) and white patients (long-dashed line). Short-dashed lines are 95% credible intervals.	93

CHAPTER 1

INTRODUCTION

Life-time event modeling is widely applied in epidemiology, clinical trials, engineering, economics and many other fields. Let T be the time to some specified event for an experimental unit, e.g. death of cancer, failure of a machine, development of a disease, etc. Let $F(t)$ be the cumulative distribution function for T . The survival function associated with F is then

$$S(t) = Pr(T > t) = 1 - F(t). \quad (1.1)$$

The hazard function quantifies the probability of an individual of age t experiencing an event in next instant which is formally defined as

$$h(t) = \lim_{\Delta t \rightarrow 0^+} \frac{Pr\{t \leq T \leq t + \Delta t | T \geq t\}}{\Delta t}. \quad (1.2)$$

Event times are often subject to censoring. A commonly observed censoring type is right-censoring. Denote C as a random censoring time. If $C < T$, then the failure time T is right-censored at time C . Define the censoring indicator $\delta = 1$ if $T < C$ and $\delta = 0$ if $T > C$. The smaller of these two $X = \min\{T, C\}$ is observed, along with δ . I assume C is independent of T in this dissertation. Other commonly seen censoring types include left-censoring and interval-censoring.

Subjects in a study may experience a single event or multiple events. When the multiple events are a number of repeated events of the same type for the subjects, the events are often referred as recurrent events. Recurrent events data, arising as multivariate data, involve an underlying stochastic process generating the stochastically dependent event times. For both single event and recurrent event data modeling, the

event times in the population of interest are typically assumed to be independent. However, many lifetime modeling techniques also take into account the correlations among the subjects. One important area of work deals with geographically clustered subjects where the spatial correlation receives careful consideration in modeling.

Common objectives of the analysis include estimating the failure rate or survival probability for each individual, describing the effects of covariates, and quantifying the variations among heterogeneous sub-populations or dependences among stochastic events. This chapter is organized as follows: section 1.1 introduces some models for recurrent event data, section 1.2 presents a review for spatially correlated survival data, and section 1.3 gives a brief introduction to B-splines.

Recurrent event models

Let $0 \leq T_1 < T_2 < \dots$ be a vector of random recurrent event failure times. Let $X_j = T_j - T_{j-1}$ be the gap time between the successive events. The associated counting process $\{N(t), t \geq 0\}$ records the cumulative number of failures over time, i.e. the number of $\{T_i\}$ less than t . Let $H(t) = \{N(s) : 0 \leq s < t\}$ describe the history of the process at time t . Models for recurrent events can be specified through the probability distribution for the number of events in short intervals $[t, t + \Delta t]$, given the history of event occurrence before time t . Suppose events occur in continuous time and two events can not occur simultaneously. Then the intensity function gives the instantaneous probability of an event occurring at t , conditional on the process history. The intensity function is defined as

$$\phi(t|H(t)) = \lim_{\Delta t \rightarrow 0^+} \frac{Pr\{N(t + \Delta t) - N(t) = 1|H(t)\}}{\Delta t}. \quad (1.3)$$

Depending on intensity function assumptions, processes for describing the recurrent events $\{T_1, T_2, \dots, \}$ can be divided into categories. Commonly seen processes include the Poisson process, the renewal process, multi-state processes, effective age

processes, and so on. The following sections include some details of the above mentioned processes.

Assume that event occurrences are recorded over $[0, \tau]$ for a specific subject. The time τ is the termination time of the study. The termination event may be related or not related to the event process; the former usually involves another type of event that ends the main event process and the latter often occurs when the termination of observation is due to a study or follow-up ending. Assume τ to be fixed. For an individual with n events at times $0 \leq t_1 < t_2 < \dots < t_n \leq \tau$, the likelihood is

$$L = \prod_{j=1}^n \phi(t_j | H(t_j)) \cdot \exp\left\{-\int_0^{\tau} \phi(u | H(u)) du\right\}. \quad (1.4)$$

When an intervention (repair) is performed (independent of the failure process) without a corresponding failure, then the system failure time is right-censored. One can obtain the resulting likelihood after small modifications of the above formula.

Poisson process

The Poisson process is one of the most important random processes because of its wide applications in time and space. The intensity function of a Poisson process is assumed to be of form

$$\phi(t | H(t)) = h(t), \quad t > 0 \quad (1.5)$$

An equivalent definition based on the counting process $\{N(t) : t \geq 0\}$ needs to satisfy that (i) $N(0) = 0$; (ii) For each $t > 0$, $N(t)$ has a Poisson distribution with mean $m(t) = \int_0^t h(s) ds$; (iii) For each $0 \leq t_1 \leq t_2 \leq \dots \leq t_m$, $N(t_1)$, $N(t_2) - N(t_1), \dots$, $N(t_m) - N(t_{m-1})$ are independent random variables. That is, the Poisson process describes situations where events occur randomly such that numbers of events in non-overlapping time intervals are statistically independent. The process is called homogeneous if $h(t) = \rho$ and non-homogeneous if $h(t)$ changes over time. When the process is homogeneous, the gap times are independently and identically distributed as $\text{Exp}(\rho)$.

Example 1: consider a system placed in service. Once it fails, it is restored to the “good as old” state and the process is repeated. Assume that the repair time is immediate. Then the times for the failures are the Poisson events.

Example 2: consider the the arrival times of guests to a fast food restaurant. The arrival rate to a fast food restaurant varies with the time of day and increases to a local maximum during meal times. Then the number of customers can be viewed as a nonhomogenous Poisson process.

A covariate process $\mathbf{w}(t)$ can be incorporated into the intensity function proportionally

$$\phi(t|H(t)) = h_0(t)\exp(\boldsymbol{\beta}'\mathbf{w}(t)), \quad t > 0. \quad (1.6)$$

When $h_0(\cdot)$ is assumed to be a parametric function, indexed by parameter $\boldsymbol{\alpha}$, inferences for $\boldsymbol{\alpha}$ and regression parameters $\boldsymbol{\beta}$ can be obtained by the maximum likelihood method. A more flexible method assumes $h_0(\cdot)$ to be piece-wise constant over prespecified intervals. In particular, let $a_1 < a_2, \dots, < a_K$ be K cut-points such as $a_1 > 0$ and $a_K = \tau$. The baseline hazard function $h_0(t; \boldsymbol{\alpha}) = \alpha_k$ for $a_{k-1} \leq t \leq a_k$. Cook and Lawless (2007) recommend three to ten pieces with cut-points evenly distributed over the event times. Maximum likelihood method can again be used to obtain inferences. There are many other alternatives to the piece-wise constant method, including Gamma process and Beta process priors on the cumulative hazard function, and Dirichlet process and Polya tree priors on the cumulative distribution function. Ibrahim et al. (2001) gives a comprehensive review of flexible methods for modeling $h_0(\cdot)$. Methods that do not make a parametric assumption about the baseline hazard are appealing. In addition to the methods of assuming flexible model for $h_0(\cdot)$, a semiparametric method which is based on the score function considers $h_0(\cdot)$ as an arbitrary positive-valued function (Andersen and Gill 1982) and can use profile likelihood approach to make inferences.

Random effects are also often incorporated into the model to accommodate het-

erogeneity across individuals. For example, assume the conditional “subject-specific” intensity function for subject i to be in the form

$$\phi_i(t|H_i(t), u_i) = u_i h_0(t) \exp(\boldsymbol{\beta}' \mathbf{w}(t)), \quad t > 0, \quad (1.7)$$

where the u_i s are nonnegative independent random variables and are often assumed to be independent of the covariates. A commonly used distribution for u_i is the Gamma distribution. The EM algorithm offers a convenient approach in fitting the random effect model.

There are also robust semi-parametric methods assessing the effects of the covariates on the mean function $m(t)$ and a marginal method of analyzing multivariate failure time data, see Cook and Lawless (2007) for more details. Typical tests for Poisson processes are for trend, proportionality assumptions of specific covariates, extra-Poisson variation, and comparisons among two or more groups. However, there are few methods for testing whether the event process is Poisson, for example, whether repairs are “good as old”. Martingale residuals might indicate associations in the successive event counts, suggesting violation of the Poisson assumption. I consider a Bayesian nonparametric test for the Poisson assumption which is presented in Chapter 2.

Renewal process

Renewal processes are commonly used if an individual is “renewed” after each event occurrence. Renewal processes assume that the gap times (X_1, X_2, \dots) are independent and identically distributed. Equivalently, the intensity function is assumed as

$$\phi(t|H(t)) = h(t - T_{N(t-)}), \quad t > 0, \quad (1.8)$$

where $h(\cdot)$ is the hazard function (failure rate) of X_i and $T_{N(t-)}$ is the time of last event occurrence.

Example 1: consider customers arriving at a service station, assuming that customers come randomly. Then the arrivals of the customers are renewal events.

Example 2: consider a device placed in service that eventually fails. It is replaced by a device of the same type and the process is repeated. Assume that the replacement is immediate. Then the times for the failures are the renewal events.

Renewal processes have a very rich and interesting mathematical structure and can be used as a foundation for building more realistic models. The modulated renewal models (Cox 1972b, Lawless 2003a) assume that the intensity function given a covariate process $\mathbf{w}(t)$ and baseline hazard function $h_0(\cdot)$ is of the form

$$\phi(t|H(t)) = h_0(t - T_{N(t-)})\exp(\boldsymbol{\beta}'\mathbf{w}(t)). \quad (1.9)$$

The covariate process can include fixed effects and components of prior event history such as gap times and number of events. In that case, the gap times are not identically distributed. Standard methods for the Cox survival model can be used to obtain inferences. A more general framework is formulated through the sequence of conditional distributions for the gap times

$$F_j(x|x^{(j-1)}, \mathbf{w}_j) = Pr(X_j < x|x^{(j-1)}) \quad (1.10)$$

where $x^{(j-1)} = (x_1, \dots, x_{j-1})'$ and \mathbf{w}_j is the vector of covariates associated with the j th gap time. The general framework allows $X_j, j = 1, 2, \dots$ to depend on previous gap times. The marginal distributions of the gap times, conditional on covariates but not on previous gap times are often of interest to analysts.

Multi-state process and effective age process

In absence of covariates, direct extensions of renewal process and Poisson process include the multi-state processes. The multi-state extension of a renewal process assumes

$$\phi(t|H(t)) = h_k(t - T_{N(t-)}), \quad t > 0, N(t-) = k. \quad (1.11)$$

The distributions of the gap times are hence non-identical. Similarly, the multi-state extension of a Poisson process assumes

$$\phi(t|H(t)) = h_k(t), \quad t > 0, N(t-) = k \quad (1.12)$$

where the intensity function depends on the state occupied and the time since the start of the process.

For repairable systems, a renewal process implies that all repairs bring the system to a “good as new” state (system’s effective age is reset to zero) and a Poisson process assumes that all repairs bring the system to a “good as old” state (system’s effective age is set to the time right before the repair/failure). Both processes introduce the notion of “effective age” of the system after each repair. Let $\varepsilon(t)$ be the effective age of an event process and assume that the intensity function is directly related to the effective age through $\phi(t|H(t)) = h(\varepsilon(t))$. Therefore, $\varepsilon(t) = t - T_{N(t-)}$ (backward occurrence time) for renewal processes and $\varepsilon(t) = t$ for Poisson processes. More general effective age processes have been studied by Kijima (1989), Dorado, Hollander, and Sethuraman (1997), Peña et al. (2007), and others.

Kijima models define a class of effective age processes by assuming $\varepsilon(t) = \varepsilon(T_{N(t-)} + t - T_{N(t-)})$ where $\varepsilon(t_i) = \varepsilon(t_{i-1}) + (t_i - t_{i-1})q$ (Kijima type I) or $\varepsilon(t_i) = (\varepsilon(t_{i-1}) + t_i - t_{i-1})q$ (Kijima type II). Here q is a positive scalar describing the effectivenesses of the repairs. When q equals zero, both types imply renewal processes; when q equals one, both types imply Poisson processes. The effective age characterizes a spectrum of effectivenesses of the interventions and hence is very flexible in modeling the dependences among the events. In Chapter 3, I propose a Bayesian semi-parametric model for recurrent events data by modeling q as a function of some covariates.

Mean cumulative function

A useful tool to examine the rate of occurrence of events is the mean cumulative function (MCF) for $N(t)$:

$$M(t) = E\{N(t)\}. \quad (1.13)$$

A nonparametric estimator for $M(t)$ is proposed by Nelson (1995) and examined by many others.

Let $I(A)$ be indicator function which takes 1 if A is true and 0 otherwise. Define $Y(t) = I(\text{process is observed at time } t)$. When a subject is observed within time period $[\tau_0, \tau]$, then $Y(t) = I(\tau_0 \leq t \leq \tau)$. Suppose m subjects are under investigation. Let $d\bar{N}.(s) = \sum_{i=1}^m Y_i(s)dN_i(s)$ be the total number of events observed over $[s, s+ds]$ and $Y.(s) = \sum_{i=1}^m Y_i(s)$ be the total number of subjects at risk over $[s, s+ds]$. Assume k systems are mutually independent and the termination times have random censoring, independent of the event processes. Further we assume that mean cumulative function $M(t)$ exists and is continuous. An unbiased estimator, proposed by Nelson (1995) is

$$\widehat{M}(t) = \sum_{h:t(h) \leq t} \frac{d\bar{N}.(t(h))}{Y.(t(h))}. \quad (1.14)$$

Spatial survival analysis

Environmental and epidemiological life-time data often involve spatially correlated subjects, since geographically close subjects tend to be exposed to similar environmental and social conditions. The spatial dependence needs to be taken into account into the analysis since it may affect inferences on other model parameters, see Cressie (1993). The spatial correlation pattern itself may also be of interest, in addition to inferential tasks on describing the relationship between the survivorship and the covariates (demographic information, treatments, disease stage and etc.) and estimating the survival/hazard function.

Example 1 (East Boston asthma data, Li and Lin 2006): a questionnaire data for subjects enrolled at community health clinics where age at onset of childhood asthma (response), low respiratory index, maternal asthma status, and maternal cotinine levels were recorded. Since asthma is believed to be strongly associated with environmental triggers and children living adjacent locations might be exposed to similar physical and social environments, their ages at onset of childhood asthma are likely to have spatial dependence.

Example 2 (South Carolina prostate cancer registration data): a registration data for prostate cancer patients where time of death since diagnoses with prostate cancer (response), age, race, marital status, grade of tumor, SEER summary stage were recorded. Since cancer patients' survival is affected by treatment conditions (hospital), income levels, and some other environmental and social factors, their survival times are likely to be subject to spatial dependence.

The relationships between an individual's survival and observed risk factors are quantified through basic survival models. The spatial dependence is typically incorporated into the model through spatial frailties (Li and Ryan 2002, Hennerfeind et al. 2006). More recently, Li and Lin (2006) proposed a marginal approach through normally transformed survival functions. The following sections will briefly present the aforementioned models.

Basic survival models

Commonly seen survival models include the proportional hazards model (PH) (Cox, 1972) and the accelerated failure time (AFT) (Kalbfleisch and Prentice 2002). There is also increasing attention given to the accelerated hazards model (AH) (Chen and Wang 2000) and the extended hazard model (EH) (Etezadi-Amoli and Ciampi, 1987; Chen and Jewell, 2001).

Denote the covariate vector as \mathbf{w} . The PH model assumes

$$h(t|\mathbf{w}) = h_0(t)\exp(\boldsymbol{\beta}'\mathbf{w}), \quad (1.15)$$

where $\boldsymbol{\beta}$ describes the covariates' effects. The PH model parameterizes the treatment's effect as a constant of proportionality between the hazard functions. The model parameter $\boldsymbol{\beta}$ can be estimated by maximizing the partial likelihood and asymptotic properties of the estimators have been developed (Anderson and Gill 1982). The model can also be fitted using Bayesian semiparametric methods. For example, assuming the baseline hazard $h_0(\cdot)$ to be piecewise constant, or the cumulative hazard function to be a gamma process, or the density corresponding to $h_0(\cdot)$ to be placed with a Dirichlet process prior. Bayesian inferences for both $\boldsymbol{\beta}$ and $h_0(\cdot)$ can be easily obtained through MCMC iterations. Ibrahim et al. (2001) gives a comprehensive review for the aforementioned Bayesian semiparametric models.

The AFT model provides an direct extension of classic linear models. Denote $Y = \log(T)$. The AFT model assumes a linear relationship between Y and covariate vector \mathbf{w} ,

$$Y = \mu + \boldsymbol{\beta}'\mathbf{w} + \sigma\epsilon, \quad (1.16)$$

where ϵ is an error term and σ is a scalar. In terms of the hazard function of T , this can be expressed as

$$h(t|\mathbf{w}) = h_0(te^{\boldsymbol{\beta}'\mathbf{w}})\exp(\boldsymbol{\beta}'\mathbf{w}). \quad (1.17)$$

Tsiatis (1990) proposed a class of linear rank estimates for estimating $\boldsymbol{\beta}$. Jin et al. (2003) provided simple and reliable methods for implementing the aforementioned rank estimators. Christensen and Johnson (1988) gave a semi-Bayesian analysis of the model where the error term is assigned a Dirichlet process prior (Ferguson 1973). Walker and Mallick (1999) proposed a Bayesian semiparametric approach for fitting the AFT model where the error distribution is assigned a Polya tree prior (Lavine 1992). More recently, Zeng and Lin (2007) proposed an approximate nonparametric

maximum likelihood method by maximizing a kernel-smoothed profile likelihood function. When the covariates are time-dependent and possibly with measurement errors, Tseng et al. (2005) proposed a joint modeling approach.

Both the PH model and the AFT model assume differing hazard functions at time zero for different covariates. Instead, the AH model quantifies the relative hazard through time progression in the form of

$$h(t|\mathbf{w}) = h_0(te^{\beta'\mathbf{w}}). \quad (1.18)$$

A general class of models which include the above three models as subclasses is the EH model where the relationship between individual hazard and baseline hazard is specified as

$$h(t|\mathbf{w}) = h_0(te^{\beta'\mathbf{w}})\exp(\gamma'\mathbf{w}). \quad (1.19)$$

The parameter β describes the relative hazard progressions and γ characterizes the relative hazard after adjusting for the different progressions. The EH model can be used to compare the three subclasses of models. Chapter 4 gives an extensive review of methods in literature for fitting the EH model and also presents my method in a Bayesian semi-parametric framework.

Spatial correlation

Spatial data records the location for each subject. Depending on how the locations are documented, the spatial data can be divided into categories and different categories may have quite distinguished methodologies or similar methods but with some modifications. Two main types of data are geostatistical (point-referenced) data, where the locations vary continuously over a subset of a Euclidean space, and areal data, where the locations are finite number of areal units with well-defined boundaries. Suppose s is either a spatial index for m areal units, i.e. $s \in \{1, \dots, m\}$, or an exact spatial coordinate $s_i = (x_i, y_i)$, for example, longitude and latitude. Let \mathcal{D} be the set of areal

units or the finite array of sites. Assume $\{Y(s), s \in \mathcal{D}\}$ to be a spatial stochastic process. Commonly used processes include stationary Gaussian random field (GRF), Markov random fields (MRF), and more recently proposed two dimensional tensor product P-spline priors (Lang and Brezger 2004, Hennerfeind et al. 2006). GRF and two dimensional P-spline priors are for geostatistical data. MRF, e.g. the conditional autoregressive prior (CAR), is widely used for areal data.

MRF priors assume the spatial effect of an area is based on its “neighbors”. Consider the spatial process $\mathbf{Y} = (Y_1, \dots, Y_m)$ defined on the lattice $\mathcal{D} = \{1, \dots, m\}$. In general, \mathbf{Y} is characterized by its joint distribution. Define $N(s)$ to be the collection of all other areal units such that

$$p(Y_s | \mathbf{Y}_{-s}) = p(Y_s | \mathbf{Y}_{N(s)}), s = 1, \dots, m$$

where $p(Y_s | \cdot)$ denotes a conditional distribution for Y_s , $\mathbf{Y}_{N(s)} = \{Y_j, j \in N(s)\}$, and $\mathbf{Y}_{-s} = (Y_1, \dots, Y_{s-1}, Y_{s+1}, \dots, Y_m)$. When the conditional probabilities define the joint probability of \mathbf{Y} , the process \mathbf{Y} is called a Markov random field. An illustration of an MRF is the Conditional autoregressive model (CAR) model for continuous data. The CAR model assumes the conditional distributions $p(Y_s | \mathbf{Y}_{-s})$ are Gaussian with mean and variance

$$\begin{aligned} E(Y_s | \mathbf{Y}_{-s}) &= \frac{1}{d_s} \sum_{j \in N(s)} Y_j, \\ \text{Var}(Y_s | \mathbf{Y}_{-s}) &= \tau^2 / d_s, \end{aligned}$$

where d_s is the number of neighbors of area unit s and τ^2 is the variance parameter which controls the amount of spatial smoothness. The Markov-type property in space is defined in terms of the neighborhood $N(s)$.

A GRF assumes $Y(s)$ to have mean μ and variance τ^2 , and use a isotropic covariance function where $\text{cov}(Y(s), Y(s')) = C(\|s - s'\|)$ with $\|\cdot\|$ being the Euclidean distance. For a finite array $\mathbf{Y} = (Y(s_1), \dots, Y(s_n))$, the prior for \mathbf{Y} can be expressed

as

$$p(\mathbf{Y}) \propto \exp\{-\mathbf{Y}'\mathbf{K}\mathbf{Y}/2\}$$

with $\mathbf{K} = \mathbf{C}^{-1}$ where $C_{ij} = C(\|s_i - s_j\|)$. A typical choice in practice for the function $C(\cdot)$ is the Matèrn family of covariance functions, which is given by

$$C(h) = \begin{cases} \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(2\zeta\sqrt{\nu}h)^\nu \mathcal{K}_\nu(2\zeta\sqrt{\nu}h), & \text{if } h > 0, \\ 0, & \text{otherwise,} \end{cases}$$

where ζ, ν, σ^2 are parameters, $\Gamma(\cdot)$ is the conventional Gamma function, and $\mathcal{K}_\nu(\cdot)$ is the modified Bessel function of order ν . Here ν is the parameter controlling the smoothness of the realized random field, ζ is a spatial scale parameter controlling how fast the covariances die out, and σ^2 is the variance of $Y(s)$. When $\nu = 3/2$, $C(h) = \sigma^2(1 + \zeta h)\exp(-\zeta h)$. Note that the dimension of the penalty matrix \mathbf{K} corresponds to the number of distinct locations. When the number of locations is large, the computation is very heavily burdened. To reduce this computational burden, a ‘‘low-rank’’ kriging approximation using a representative subset of knots obtained from a space-filling algorithm have been given by Kammann and Wand (2003) and applied for spatial survival models in Kneib and Fahrmeir (2006).

A two dimensional P-spline modeling approach assumes that the spatial process $Y(s)$ can be approximated by the tensor product of one-dimensional B-splines, that is,

$$Y(s) = Y(x_s, y_s) = \sum_{m_1=1}^{d_1} \sum_{m_2=1}^{d_2} \beta_{m_1 m_2}^{spat} B_{m_1}(x_s) B_{m_2}(y_s),$$

where $\beta^{spat} = (\beta_{11}^{spat}, \dots, \beta_{1d_2}^{spat}, \dots, \beta_{d_1 1}^{spat}, \dots, \beta_{d_1 d_2}^{spat})$ are random B-spline coefficients, $B_{m_1}(\cdot)$ and $B_{m_2}(\cdot)$ are B-spline basis functions. MRF priors for areal data can be used to construct priors for β^{spat} . For example, a prior based on four nearest neighbors can be defined by

$$\beta_{m_1 m_2}^{spat} | \cdot \sim N\left(\frac{1}{4}(\beta_{m_1-1, m_2}^{spat} + \beta_{m_1+1, m_2}^{spat} + \beta_{m_1, m_2-1}^{spat} + \beta_{m_1, m_2+1}^{spat}), \frac{\tau^2}{4}\right).$$

Two dimensional P-splines give a smooth surface for the spatial effect and MRF allows rather heterogenic neighbors. A choice between the two depends on the characteristics of the spatial effect.

Assume $f_{spat}(s_i)$ to be a (structured) spatial effect for individual i . A typical extension of PH model to the spatial case assume

$$\lambda_i(t) = \lambda_0(t)\exp(\beta' \mathbf{w}_i + f_{spat}(s_i)). \quad (1.20)$$

For geo-statistical data, $f_{spat}(s_i) = f_{spat}(x_{s_i}, y_{s_i})$, and for areal data $f_{spat}(s_i) = \beta_s^{spat}$ where $s_i = s$. Let r_1, \dots, r_M be the distinct values of $\{s_i, i = 1, \dots, n\}$. Assume that the censoring time is independent of the failure random variable. The likelihood for observing $\{t_i, \mathbf{w}_i, s_i, \delta_i\}_{i=1}^n$ is

$$L = \int \prod_{i=1}^n [\lambda_i(t_i|\mathbf{r})]^{\delta_i} S_i(t_i|\mathbf{r}) p(\mathbf{r}) d\mathbf{r}$$

where $\mathbf{r} = (r_1, \dots, r_M)$ and $S_i(\cdot)$ is the survival function corresponding to $\lambda_i(\cdot)$. Since the integral in evaluating the full likelihood is analytically intractable, simulation methods are often used, including direct Monte Carlo approximation by sampling \mathbf{r} from $p(\mathbf{r})$ for an equivalent marginal likelihood based on ranks (Li and Ryan 2002), and likelihood augmentation by updating \mathbf{r} through MCMC iterations (Banerjee et al. 2003, Hennerfeind et al. 2006). Approximation approach has been very limited for spatial survival data. Laplace approximation for the marginal rank likelihood was used in Li and Ryan (2002). As the dimension of \mathbf{r} gets high, The computation burden for the simulation based methods is heavy. More recently, Martino et al. (2011) used approximate Bayesian inference for survival models based on integrated nested Laplace approximations (INLA).

Very few marginal approaches exist in the literature. Li and Lin (2006) proposed a class of normal transformation models assuming PH for the marginal survival model. Compared to frailty approaches, marginal approaches allow population-averaged interpretation for the regression coefficients. Chapter 4 introduces an extension to Li

and Lin (2006) by allowing the basic model to be the EH model and accounting for the spatial correlation implied by a large lattice data set.

B-splines

B-splines are widely used for curve-fitting. A B-spline basis function $B_{j,K}(t)$ is a piece-wise polynomial function of t and of given degree K . It is defined over a range $t_j \leq t \leq t_{j+K+2}$. The points where $t = t_i$ are known as knots, joining the pieces of polynomial functions together. A B-spline is a continuous function at the knots and its derivatives are also continuous up to the derivative of order $K - 1$ at distinct knots. Any spline function of given degree K , can be expressed as a linear combination of B-splines of that degree, i.e. $(t - \tau)^{K-1} = \sum_{j=1}^J b_j B_{j,K}(t)$ where $b_j, j = 1, \dots, J$ are B-spline coefficients. Expressions for B-splines can be derived by means of a recursive formula:

$$B_{j,0}(t) = \begin{cases} 1 & \text{if } t_j \leq t < t_{j+1} \\ 0 & \text{otherwise} \end{cases}$$

$$B_{i,K}(t) = \frac{t - t_i}{t_{i+K-1} - t_i} B_{i,K-1}(t) + \frac{t_{i+K} - t}{t_{i+K} - t_{i+1}} B_{i+1,K-1}(t).$$

Consider the regression of m data points (x_i, y_i) assuming $E(Y) = \sum_{j=1}^J b_j B_{j,K}(x)$. The least squares of the objective function to minimize is $S = \sum_{i=1}^m \{y_i - \sum_{j=1}^J b_j B_{j,K}(x_i)\}^2$.

As the number of knots increases, the fitted curve tends to show more variation than is justified by the data. It is common to introduce a penalty on the fitted curve.

Eilers and Marx (1996) proposed a penalty method on the B-spline coefficients under which the objective function is

$$S = \sum_{i=1}^m \{y_i - \sum_{j=1}^J b_j B_{j,K}(x_i)\}^2 + \lambda \sum_{j=k+1}^J (\Delta^k b_j)^2$$

where k is the order of penalty, λ is the smoothing parameter, $\Delta b_j = b_j - b_{j-1}$, and $\Delta^k b_j = \Delta^{k-1} b_j - \Delta^{k-1} b_{j-1}$. It is typical to use first-order and second-order penalties.

Lang and Brezger (2004) proposed a Bayesian version of the penalized B-spline by introducing a structured prior distribution for $\mathbf{b} = (b_1, \dots, b_J)$. Let \mathbf{D}_1 be a $(J-1) \times J$ matrix and \mathbf{D}_2 be a $(J-2) \times J$ matrix with entries specified as

$$\mathbf{D}_1 = \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \dots & & \\ & & & -1 & 1 \end{pmatrix}; \mathbf{D}_2 = \begin{pmatrix} -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & \dots & & \\ & & & -1 & 2 & -1 \end{pmatrix}.$$

The prior equivalent to k -order penalty ($k = 1, 2$) assumes

$$\mathbf{b}|\tau^2 \propto \exp\left(-\frac{1}{2\tau^2}\mathbf{b}'(\mathbf{D}'_k\mathbf{D}_k)\mathbf{b}\right)$$

where τ^2 controls the amount of smoothness. Full Bayesian inference also consider hyper-priors on τ^2 . The joint distribution is improper as the rank of $\mathbf{D}'_k\mathbf{D}_k$ is $k-1$. The joint distribution for the first-order penalty also implies $b_j|\mathbf{b}_{-j} \sim N(\frac{1}{2}(b_{j-1} + b_{j+1}), \frac{\tau^2}{2})$ where $\mathbf{b}_{-j} = (b_1, \dots, b_{j-1}, b_{j+1}, \dots, b_J)$.

CHAPTER 2

A BAYESIAN NONPARAMETRIC TEST FOR MINIMAL REPAIR

The paper develops a Bayesian nonparametric reliability model for recurrent events where failure and truncated time-to-failure density shape is regressed on past maintenance decisions: perfect repair and minimal repair. By comparing the system interfailure lifetime distributions after minimal and perfect repair, we are able to test the minimal repair assumption of “good as old.” Interfailure hazard functions after perfect and minimal repairs are estimated, shedding light on departures from minimal repair. The method is illustrated both on simulated data as well as failure time data from air-conditioning units at the South Texas Nuclear Operating Company near Bay City, Texas. This article has supplementary material online.

Keywords: Effective age; Repairable system; Tailfree process; Truncated data.

2.1 INTRODUCTION

Repairable systems have been widely studied in the reliability literature. Systems fail and upon each failure, a system gets repaired. The distribution of interfailure times between system failures is commonly of interest. In general, recurrent event modeling methods can be divided into categories based on the type of maintenance a

The content in this chapter is a reprint by permission of Taylor & Francis LLC for “ Li Li, Timothy Hanson, Paul Damien, and Elmira Popova. (2014+). A Bayesian nonparametric test for minimal repair. *Technometrics* ”.

system receives. Renewal processes are commonly used if all the maintenance repairs bring the system to a “good as new” state (commonly known as perfect repair). One example of this kind of repair would be a complete overhaul of the system. Non-homogeneous Poisson processes are used if all repairs bring the system to the “good as old” state (commonly known as minimal repair), e.g., replacing a failed sub-component of a system. Some authors have proposed models that allow for a combination of perfect and minimal repairs, see Block et al. (1985) and Whitaker and Samaniego (1989). However, the basic assumption of a consistently “minimal” repair is questionable; usually several types of maintenance, with varying degrees of repair, are undertaken throughout the lifetime of the system. Kijima (1989), proposes a general model that includes perfect, minimal, and in-between repairs by introducing the “effective age” of the system after each repair, essentially measuring how successful the repair was. Following Kijima (1989), Dorado, Hollander, and Sethuraman (1997), allow for repairs of varying degree by including so-called “life supplements” – numbers between zero and one indicating the degree of the repair between perfect and minimal – and assume the life supplements are *known*. Recently, Veber et al. (2008) assume one life supplement that is unknown, i.e. each repair reduces the effective age of the system by the same fraction q , and propose an EM-algorithm to estimate q and the unknown failure distribution F . As an extension to a common q , Rigdon and Pan (2009) allow the repair effectiveness parameter to vary from system to system. Presnell et al. (1994) proposed a test for the minimal repair assumption in a particular model that Block et al. (1985) propose; if the null hypothesis where minimal repair assumption holds is rejected, the question remains as to whether “minimal repair” brings the system better or worse than minimal; in many application scenarios this distinction is crucial. If one ignores maintenance decisions, Cooper et al. (2006) point out that decisions based on the incorrect assumption of minimal repairs can lead to a so-called “spiral down” effect, where system reliability gets worse than the predicted

level after repair cycles (i.e. more failures than predicted); this happens because the assumed minimal repairs are actually often worse than “good as old”.

Consider a brand new system; denote the cumulative distribution function for the first failure s as $F_0(s)$. After a perfect repair at failure time s , which sets the system clock back to zero, the distribution governing the next failure is $F_0(t - s)$, where $t > s$. A minimal repair brings the system to the exact state it was in right before failure; this implies the intensity function, formally defined below in (2.1), does not change after a minimal repair. After successive minimal repairs, if a minimal repair is newly performed at failure time s (time since new condition), the cumulative distribution function for the next failure is F_0 , but truncated to be larger than s , i.e. $[F_0(t) - F_0(s)]/S_0(s)$, where $S_0(s) = 1 - F_0(s)$ is the reliability (also termed survival) function and $t > s$. We propose to relax this assumption by allowing the intensity to change after the minimal repair: the distribution for the next failure is instead $[F_1(t) - F_1(s)]/S_1(s)$. The assumption of a static intensity function is given by $H_0 : F_0 = F_1$, providing an intuitive test of the minimal repair assumption. If H_0 is rejected in favor of $H_1 : F_0 \neq F_1$, estimated hazard functions $h_0(t) = f_0(t)/S_0(t)$ and $h_1(t) = f_1(t)/S_1(t)$ enable us to find when system performs actually worse (or better) than the expected condition under the minimal repair assumption. Our framework can be easily generalized to include known life supplements (as in Dorado et al. 1997) and subsequently test for this assumption.

The hypothesis testing in this paper involves two unknown distributions. A parametric approach assumes particular distribution families for F_0 and F_1 , e.g. Weibull is commonly used for non-homogeneous Poisson or renewal process models. We propose a Bayesian nonparametric model that generalizes the Weibull assumption on F_0 and F_1 , termed a ‘tailfree prior.’ The tailfree approach we use augments the standard Weibull family indexed by θ with additional parameters $\{\pi(\epsilon)\}$ that change the shape of the Weibull density in successive layers. These additional parameters add

flexibility beyond the Weibull shape, much like adding detail to an initially washed canvas; each new layer or ‘level’ allows more refined detail to be accommodated. The Bayesian approach simply places a prior on the additional parameters $\{\pi(\epsilon)\}$.

In general, Bayesian nonparametric methods model distributions as random cumulative distribution functions (CDFs) $F(s)$, either directly or indirectly (e.g. through the hazard). Technically, a random CDF $F(s)$ is a stochastic process indexed by s , so $\{F(s) : s > 0\}$ describes a random function from \mathbb{R}^+ to $[0, 1]$, and for any fixed s , $F(s)$ is a random variable. These processes include the Dirichlet process (Ferguson 1973), Polya tree priors (Lavine 1992), Dirichlet process mixtures (Escobar and West 1995), and neutral to the right processes (Ferguson and Phadia 1979). Taddy and Kottas (2012) use Dirichlet process mixtures for the interfailure density in Poisson process models. Priors on the space of cumulative hazard functions include gamma processes, weighted gamma processes, beta processes; see Lo (1992), Kuo and Ghosh (1997) and Ishwaran and James (2004). Often, the random CDF $F(s)$ is centered at a parametric distribution G_θ in the sense that $E\{F(s)\} = G_\theta(s)$ for all $s > 0$; i.e. G_θ is the ‘prior mean’ of F . Our proposed framework uses tailfree priors (Fabius 1964, Ferguson 1974, Jara and Hanson 2011) to model F centered at the Weibull family, $E\{F(s)\} = 1 - e^{-(s/\gamma)^\alpha}$ given (α, γ) , but allows for substantial data-driven deviations from Weibull. Our approach naturally tests whether Weibull is adequate, as well as incorporating maintenances where no failure has actually occurred (i.e. censored failures). Few existing nonparametric approaches make use of information from censored system failure times, although in practice, maintenance schedules are common.

The proposed estimation procedure is applied to historical data from the South Texas project nuclear operating company located in Bay city, Texas. The system of interest is the essential chillers which is a group of six 300-ton air conditioners, 3 for each nuclear reactor unit. They provide chilled water for air handling units to

provide a suitable environment for personnel and equipment located in the electrical auxiliary building, mechanical auxiliary building, and fuel handling building. An essential chiller provides chilled water for the cooling coils of various safety related air handling units during normal, faulted and upset conditions. All three chilled water system trains are automatically started up if particular emergency situations are detected, such as safety injection signal, loss of offsite power from the switchyard, or a combination of both, to supply cooling to many essential safety systems. Those air conditioners are repairable systems. Maintenances to the air conditioners include replacement of subcomponents (oil pump, vane controller, solenoid valves, etc.) in response to failures, and overhauls, typically upon inspection, which involve a major rework on parts, e.g. compressor vane, and renewing soft materials (gaskets, refrigerant, lubrication – grease, oil) when excessive wear or other degraded conditions are noted. In our analysis, overhauls are categorized into perfect repairs while replacement of subcomponents are grouped into minimal repairs. For repairs that are not in response to a failure, yielding right-censored failure times, we do not differentiate scheduled repairs and responses to apparent degradation (but not failure), and further assume that the times for those repairs are independent of the system failure processes. The data set is comprised of two groups of observations for the two nuclear reactor units with the first group of 1274 events and the second group of 1092 events. All air conditioners are assumed to work independently. Each observation consists of an event time, associated maintenance decision, and indicator of censoring for whether a failure occurred at the event time, i.e. (t_i, d_i, δ_i) . Most minimal repairs were in response to failure and perfect repairs were performed without an accompanying failure (Table 2.1). It is assumed in this data analysis that those perfect repairs bring the system to the “good as new” condition and our main interest is that whether those minimal repairs bring the system to the “good as old” state.

Section 2 describes the model, introduces tailfree priors, and outlines the Markov

Table 2.1: Counts of perfect/minimal by response to “failure” / “censored” for the air conditioners.

	Failure	Censored
Perfect	86	1175
Minimal	1085	14

chain Monte Carlo (MCMC) algorithm used to fit the model. Section 3 presents simulation results for testing the minimal repair assumption and accompanying density estimation. Section 4 applies the method to the South Texas project data. Section 5 concludes the paper with a discussion.

2.2 MODEL DEVELOPMENT

Consider a general repairable system framework: up to the present time t_{max} we observe a series of repairs and maintenance decisions made at each repair. The times for repairs are recorded as $0 = t_0 < t_1 < t_2 < \dots < t_n = t_{max}$. The corresponding repair at event time t_i is denoted as d_i with $d_i = 1$ if minimal repair was performed and $d_i = 0$ if perfect repair was performed; we assume $d_0 = 0$. Denote the last perfect repair time prior to decision d_i as $t_i^* = \max\{t_j : j < i, d_j = 0\}$. If maintenance (random or planned) is performed at time t_i *without* an accompanying failure, the failure time stemming from the previous decision is censored, indicated by $\delta_i = 0$ and 1 otherwise. For simplicity, we assume that δ_i is independent of the failure process. Since repairs must occur after failures, and can also occur without a failure event, the set of failure times is a subset of $\{t_1, \dots, t_n\}$. Full data are $\mathcal{D} = \{(t_i, \delta_i, d_i)\}_{i=1}^n$. For data observed over the window $[0, t_{max}]$, the event time $t_n = t_{max}$ is the time at which data collection stops and $\delta_n = 0$. We assume that maintenance decisions are observable where perfect repairs bring the system to “good as new” state and minimal repairs otherwise. Furthermore we assume that the time for repair is negligible, i.e.

there is no “down” time during the repair.

Let the counting process $\{N(t), t \geq 0\}$ record the cumulative number of failures over time and $H(t) = \{N(s) : 0 \leq s < t\}$ denote the history of the process at time t . Then the intensity function is defined as

$$\phi(t|H(t)) = \lim_{\Delta \rightarrow 0^+} \frac{Pr\{N(t + \Delta) - N(t) = 1|H(t)\}}{\Delta}. \quad (2.1)$$

The intensity function describes the instantaneous probability of a failure occurring at t , conditioning on the process history. Let $F_0(t)$ be the probability that the system lasts less than t time units since a perfect repair and $S_0(t) = 1 - F_0(t)$ be the survival probability; denote the density as $f_0(t)$ and hazard as $h_0(t)$. A previous perfect repair $d_{i-1} = 0$ brings the system to “good as new” status, i.e. resets the system clock to zero. A failure t_i right after perfect repair $d_{i-1} = 0$ at t_{i-1} has likelihood contribution $f_0(t_i - t_{i-1})$. If instead, a minimal repair $d_{i-1} = 1$ restores the system to the exact state it was in right before failure at t_{i-1} , then the system has aged $t_i - t_i^*$ units since the last perfect repair, truncated at $t_{i-1} - t_i^*$, yielding the likelihood contribution $f_0(t_i - t_i^*)/S_0(t_{i-1} - t_i^*)$. This above assumption is commonly referred as “minimal repair assumption” and it implies that the underlying intensity function for the recurrent events does not change after minimal repairs, $\phi(t|H(t)) = h_0(t - t_i^*)$ over $[t_i^*, t_i)$, regardless of the minimal repairs preceding t_i . In our framework, we do not make the minimal repair assumption, and simply allow the intensity function to change after the first minimal repair on the renewed system. The intensity function is then $\phi(t|H(t)) = h_1(t - t_i^*)$, for hazard $h_1(t) = f_1(t)/S_1(t)$ and $S_1(t) = \int_t^\infty f_1(s)ds$, and the likelihood contribution is $f_1(t_i - t_i^*)/S_1(t_{i-1} - t_i^*)$. Note that subsequent minimal repairs do not further change the intensity function. The resulting model can be viewed as a generalization to a two-state Poisson process. Given substantially more data, a multi-state Poisson process (Cook and Lawless 2007) could be fit, assuming $\phi(t|H(t)) = h_k(t)$, $N(t-) = k$, $k \in \{0, 1, 2, \dots\}$ following each perfect repair.

Denote $H_0 : F_0 = F_1$ as the hypothesis assuming the minimal repair assumption holds and $H_1 : F_0 \neq F_1$ as the hypothesis allowing departure from this assumption. Under H_0 we put one tailfree prior (to be elaborated in Section 2.1) on F_0 ; under H_1 we place two conditionally independent priors on F_0 and F_1 . Under H_0 the likelihood is

$$\mathcal{L}(f_0) = \prod_{i=1}^n \left[f_0(t_i - t_{i-1})^{\delta_i} S_0(t_i - t_{i-1})^{1-\delta_i} \right]^{1-d_{i-1}} \left[\frac{f_0(t_i - t_i^*)^{\delta_i} S_0(t_i - t_i^*)^{1-\delta_i}}{S_0(t_{i-1} - t_i^*)} \right]^{d_{i-1}}. \quad (2.2)$$

Under H_1 the likelihood is

$$\mathcal{L}(f_0, f_1) = \prod_{i=1}^n \left[f_0(t_i - t_{i-1})^{\delta_i} S_0(t_i - t_{i-1})^{1-\delta_i} \right]^{1-d_{i-1}} \left[\frac{f_1(t_i - t_i^*)^{\delta_i} S_1(t_i - t_i^*)^{1-\delta_i}}{S_1(t_{i-1} - t_i^*)} \right]^{d_{i-1}}. \quad (2.3)$$

In terms of interfailure hazard functions h_0 , the likelihood under H_1 is

$$\begin{aligned} \mathcal{L}(h_0, h_1) &= \prod_{i=1}^n \left[h_0(t_i - t_{i-1})^{\delta_i} \exp \left\{ - \int_0^{t_i - t_{i-1}} h_0(s) ds \right\} \right]^{1-d_{i-1}} \\ &\times \left[h_1(t_i - t_i^*)^{\delta_i} \exp \left\{ - \int_{t_{i-1}}^{t_i} h_1(s - t_i^*) ds \right\} \right]^{d_{i-1}}. \end{aligned}$$

It is straightforward to interpret F_0 as the failure time distribution of a new system. Let t^* be the time at which the last perfect repair was made. Since our alternative model assumes $\phi(t|H(t)) = h_1(t - t^*)$ after minimal repairs, an estimate of h_1 averages the intensities over time after the first minimal repair in each renewed cycle. When the system performs better (or worse) than an expected level at time t under the minimal repair assumption, then $h_1(t)$ will be lower (higher) than $h_0(t)$. We perform a simulation in Section 3 to illustrate this point.

Tailfree process priors on F_0 and F_1

We place tailfree process priors on F_0 and F_1 . The use of the term ‘tailfree’ dates to Freedman (1963), who considered conditions for consistency of Bayesian probability measures on the positive integers; the main condition has to do with the shape of the

tail of the density, i.e. the integers stretching off to infinity. Fabius (1964) extended Freedman’s notion of ‘tailfree’ to continuous measures with densities; however the requirements for consistency no longer deal with the tails of the distribution. The construction that follows is a modest reworking of Fabius (1964).

Let $G_{\boldsymbol{\theta}}$ denote the Weibull cumulative distribution function parameterized as $G_{\boldsymbol{\theta}}(t) = 1 - \exp\{-(t/\gamma)^\alpha\}$ for $t \geq 0$, where $\boldsymbol{\theta} = (\log(\alpha), \log(\gamma))'$. Let $g_{\boldsymbol{\theta}}(t)$ be the corresponding density. The tailfree prior augments the Weibull family indexed by $\boldsymbol{\theta}$ with additional parameters $\{\pi(\boldsymbol{\epsilon})\}$ ($\boldsymbol{\epsilon}$ is a binary number, described below) that change the shape of the Weibull density on successive levels; this ‘more flexible Weibull’ CDF is denoted $F(t)$. Before delving into the definition, we can get the flavor of the approach through a preliminary look at Figure 2.1. Panel (a) shows a Weibull density for a particular $\boldsymbol{\theta}$; let T be drawn from this Weibull density. Panel (b) adjusts the shape of the density by adding one parameter $\pi(0)$ that changes the probability of T being less than the median of the Weibull density from 0.5 to 0.45, but leaves the shape of the density the same – this is the first level. Panels (c) and (d) add two more parameters, $\pi(00)$ and $\pi(10)$ successively, that modify the shape of the density on smaller sets in the second level, but leave the density shape the same on these smaller sets. Panels (e), (f), and (g) add four more parameters on the third level. The Bayesian approach simply places priors on the parameters $\{\pi(\boldsymbol{\epsilon})\}$, in addition to $\boldsymbol{\theta}$, yielding a random CDF F and corresponding density f . Let $T \sim F$. The prior is chosen so that, given $\boldsymbol{\theta}$, the probability $P_F(a < T < b) = \int_a^b f(s)ds$ has expectation $\int_a^b g_{\boldsymbol{\theta}}(s)ds$; for example $E\{F(s)\} = G_{\boldsymbol{\theta}}(s)$ for any $s > 0$. In this sense, the ‘prior mean’ of F is $G_{\boldsymbol{\theta}}$.

We now present a technical specification of the nonparametric prior. Let $\epsilon_1 \cdots \epsilon_j$ be a j -digit binary number where $\epsilon_i \in \{0, 1\}$ for $i = 1, 2, \dots, j$. Each $\boldsymbol{\epsilon} = \epsilon_1 \cdots \epsilon_j$ indexes a set $B_{\boldsymbol{\theta}}(\boldsymbol{\epsilon}) \subset [0, \infty)$. Following Lavine (1992), these sets are intervals with endpoints that are quantiles of the centering family: if m is the base-10 rep-

representation of the binary number $\epsilon = \epsilon_1 \cdots \epsilon_j \in \{0, 1\}^j$, then $B_\theta(\epsilon)$ is the interval $(G_\theta^{-1}(m/2^j), G_\theta^{-1}((m+1)/2^j)]$. Note then that at each level j , the class $\{B_\theta(\epsilon) : \epsilon \in \{0, 1\}^j\}$ forms a partition of the positive reals and furthermore $B_\theta(\epsilon) = B_\theta(\epsilon 0) \cup B_\theta(\epsilon 1)$. Figure 2.1(a) shows the first three partitions for a Weibull(4,4) centering distribution, e.g. $\theta = (\log(4), \log(4))$. Note that $[0, \infty) = B_\theta(0) \cup B_\theta(1)$, $[0, \infty) = B_\theta(00) \cup B_\theta(01) \cup B_\theta(10) \cup B_\theta(11)$, etc. For a specific $\epsilon 0$, the parameter $\pi(\epsilon 0)$ approximately follows a beta(cj^2, cj^2) density where j is the number of digits in $\epsilon 0$; more details are presented below. Walker et al. (1999) suggest thinking of a “...*particle cascading through these partitions.*” The particle, say $T \sim F$, initially moves into $B_\theta(0)$ with probability $\pi(0)$ or into $B_\theta(1)$ with probability $\pi(1) = 1 - \pi(0)$. From then on, at any level j with index $\epsilon = \epsilon_1 \cdots \epsilon_j$, if the particle is in $B_\theta(\epsilon)$, it moves into $B_\theta(\epsilon 0)$ with probability $\pi(\epsilon 0)$ or into $B_\theta(\epsilon 1)$ with probability $\pi(\epsilon 1) = 1 - \pi(\epsilon 0)$. When the particle finally makes its way into a set $B_\theta(\epsilon_1 \cdots \epsilon_J)$ in the finest partition at level J , it simply follows the base CDF G_θ restricted to $B_\theta(\epsilon_1 \cdots \epsilon_J)$ – this does not depend on the $\{\pi(\epsilon)\}$. That is, for an interval $(a, b) \subset B_\theta(\epsilon_1 \cdots \epsilon_J)$ and $T \sim F$,

$$P\{a < T < b | T \in B_\theta(\epsilon_1 \cdots \epsilon_J)\} = \frac{\int_a^b g_\theta(s) ds}{\int_{B_\theta(\epsilon_1 \cdots \epsilon_J)} g_\theta(s) ds}. \quad (2.4)$$

The definition of a tailfree prior uses a binary partitioning tree. Although most authors have used binary splits, other partitioning schemes could be implemented, e.g. Mauldin, Sudderth, and Williams (1992).

If all of the conditional probabilities are equal to one-half, i.e. $\pi(\epsilon) = 0.5$ for all ϵ , then the density $f(s)$ is simply $g_\theta(s)$, the corresponding Weibull density. The tailfree prior simply takes the *expectation* of these conditional probabilities to be one-half, $E\{\pi(\epsilon)\} = 0.5$ for all ϵ ; then $E\{f(s)\} = g_\theta(s)$. For a given set of conditional probabilities $\{\pi(\epsilon)\}$ this construction builds a density $f(s)$ that has jumps at the quantiles of G_θ , $G_\theta^{-1}(m/2^J)$, and the values of $\{\pi(\epsilon)\}$ determine the jump size. Figure 2.1(a) takes all $\pi(\epsilon) = 0.5$. Figure 2.1(b) then sets $\pi(0) = 0.45$. Figure 2.1(c) further sets $\pi(00) = 0.7$, then Figure 2.1(d) sets $\pi(10) = 0.6$. Panels (e), (f), and

(g) successively set $\pi(000) = 0.8$, $\pi(010) = 0.7$, and $\pi(100) = 0.4$ & $\pi(110) = 0.55$. Already with only three levels, we obtain quite interesting possibilities. Typically, the number of levels is higher, usually $5 \leq J \leq 8$, allowing for more refined shapes. The original Fabius (1964) construction deals with $J = \infty$. Figure (h) averages tailfree densities which have these conditional probabilities over the prior $\alpha, \gamma \stackrel{ind.}{\sim} N(4, 0.05^2)$, yielding a smooth mixture of tailfree densities.

Tailfree prior densities are essentially a weighted average between a parametric density and a histogram, with bin locations coming from the parametric density. The histogram takes the shape of the parametric density over bin intervals, and there are jumps at the bin endpoints as usual. By taking $\boldsymbol{\theta}$ to be random, as in Figure 2.1(h), the bin locations are ‘jittered’ or shifted, and the resulting density is smoothed, and is in fact differentiable (Prop. 1; Hanson, 2006). The resulting density model is similar to the ‘averaged shifted histogram’ of Scott (1985). However, Scott’s approach does not make use of a parametric family. The tailfree density has a pronounced nonparametric flavor where data are plentiful and unlike a Weibull density (e.g. multimodal), but retains the shape of the centering Weibull density where data are sparse and/or data approximately follow a Weibull distribution.

Define $\mathbf{p} = (p(1), \dots, p(2^J))'$ to be the vector of random probabilities of the 2^J sets in the finest partition at level J . Pairs of conditional probabilities $\{(\pi(\boldsymbol{\epsilon}0), \pi(\boldsymbol{\epsilon}1))\}$ are assumed to be mutually independent, implying

$$p(l+1) = P\{T \in B_{\boldsymbol{\theta}}(\epsilon_1 \cdots \epsilon_J)\} = \prod_{i=1}^J \pi(\epsilon_1 \cdots \epsilon_i), \quad (2.5)$$

where $\epsilon_1 \cdots \epsilon_J$ is the base-2 representation of l , $l = 0, \dots, 2^J - 1$. For example, say $J = 3$. Then to obtain $P\{T \in B_{\boldsymbol{\theta}}(110)\}$ one computes

$$\begin{aligned} P\{T \in B_{\boldsymbol{\theta}}(110)\} &= P\{T \in B_{\boldsymbol{\theta}}(110) | T \in B_{\boldsymbol{\theta}}(11)\} P\{T \in B_{\boldsymbol{\theta}}(11) | T \in B_{\boldsymbol{\theta}}(1)\} \\ &\times P\{T \in B_{\boldsymbol{\theta}}(1)\} \\ &= \pi(110)\pi(11)\pi(1). \end{aligned}$$

We require the survival function $S(t) = 1 - F(t)$. Let $T \sim F$. For a given $t > 0$ let t_l and t_r be the left and right endpoints of the partition interval at level J that contains t . That is, $t_l < t < t_r$ where $t_l = G_{\theta}^{-1}(m/2^J)$, $t_r = G_{\theta}^{-1}((m+1)/2^J)$, and m is such that $G_{\theta}^{-1}(m/2^J) < t < G_{\theta}^{-1}((m+1)/2^J)$. Then $P(T > t) = P(t < T \leq t_r) + P(T > t_r)$. Using (2.4) and (2.5), $P(t < T \leq t_r) = p_m \int_t^{t_r} g_{\theta}(s) ds / \int_{t_l}^{t_r} g_{\theta}(s) ds = p_m [G_{\theta}(t_r) - G_{\theta}(t)] / 2^{-J}$ and $P(T > t_r) = \sum_{j=s_{\theta}(t)+1}^{2^J} p(j)$ where $s_{\theta}(t) = m = \lceil 2^J G_{\theta}(t) \rceil$ and $\lceil \cdot \rceil$ is the ceiling function. These results imply that the survival function with respect to F is

$$S(t) = 1 - F(t) = p\{s_{\theta}(t)\} \{s_{\theta}(t) - 2^J G_{\theta}(t)\} + \sum_{l=s_{\theta}(t)+1}^{2^J} p(l), \quad (2.6)$$

where $p(l)$ is given by (2.5). By differentiating (2.6), the density with respect to F is given by

$$f(t) = \sum_{l=1}^{2^J} 2^J p(l) g_{\theta}(t) I_{B_{\theta}\{\epsilon_J(l-1)\}}(t) = 2^J p\{s_{\theta}(t)\} g_{\theta}(t), \quad (2.7)$$

where $\epsilon_J(i)$ is the binary representation $\epsilon_1 \cdots \epsilon_J$ of the integer i . Recall that Figure 2.1(b–g) plots the density (2.7) centered at $G_{\theta} = \text{Weibull}(4,4)$, $J = 3$, for different sets of $\{\pi(\epsilon)\}$.

Now introduce the subscript k to make clear we are defining two tailfree processes F_k where $k = 0, 1$ for perfect and minimal repair, respectively. Let the random variable $\lambda_k(\epsilon 0)$ be the logit transformation of $\pi_k(\epsilon 0)$, i.e.

$$\lambda_k(\epsilon 0) = \text{logit}\{\pi_k(\epsilon 0)\}. \quad (2.8)$$

The priors on $\{(\lambda_0(\epsilon 0), \lambda_1(\epsilon 0))\}$ are given by

$$\lambda_0(\epsilon 0), \lambda_1(\epsilon 0) \stackrel{\text{ind.}}{\sim} N\left(0, \frac{2}{c\rho(j)}\right), \quad (2.9)$$

where j is the number of digits in $\epsilon 0$. The $N(0, 2/c\rho(j))$ prior on $\lambda_k(\epsilon 0)$ mimics a $\text{beta}(c\rho(j), c\rho(j))$ prior for Polya tree conditional probabilities $\{\pi_k(\epsilon 0)\}$ (Jara and Hanson, 2011). A common choice which we adopt is $\rho(j) = j^2$. The parameter c

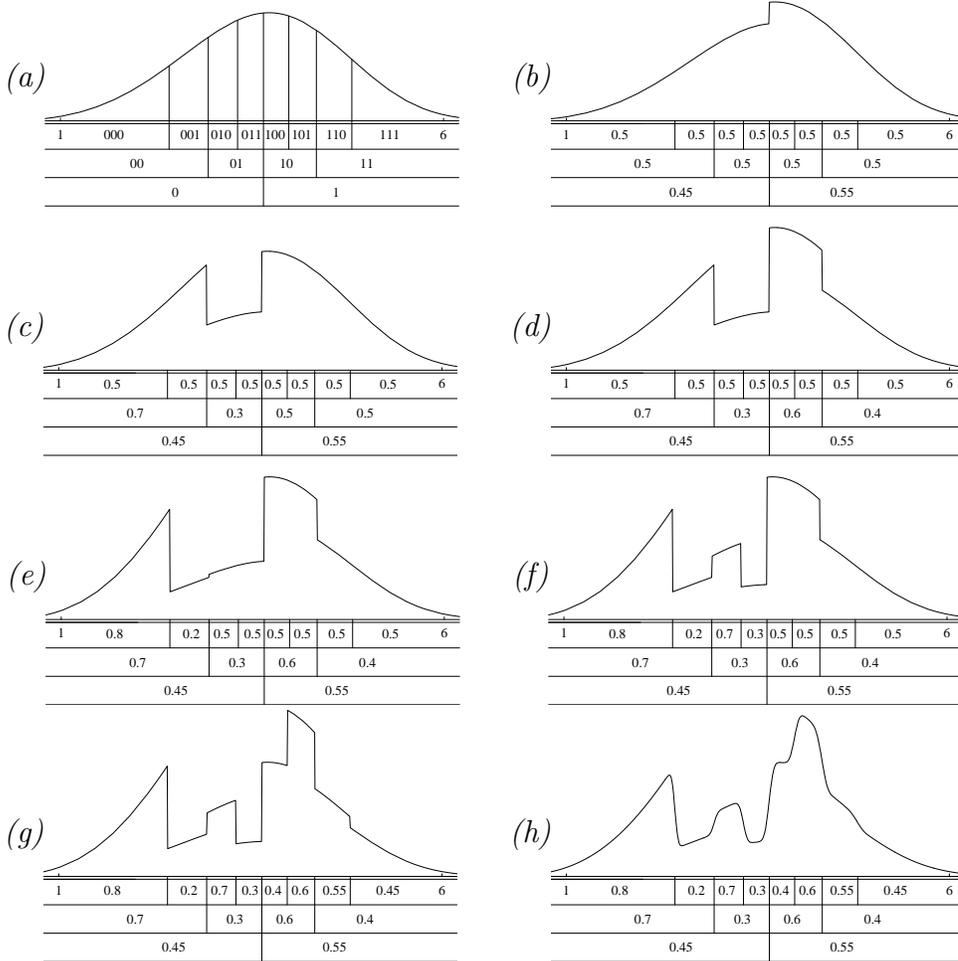


Figure 2.1: (a) Weibull (α, γ) with $\alpha = 4$ and $\gamma = 4$; (b–g) tailfree densities, centered at (a) with conditional probabilities specified up to $J = 3$; (h) mixture of tailfree processes assuming $\alpha, \gamma \stackrel{ind.}{\sim} N(4, 0.05^2)$.

acts much like the precision in a Dirichlet process (Ferguson, 1973). As $c \rightarrow 0^+$, $E\{F_k(\cdot)\}$ tends to the empirical CDF of the data (Hanson and Johnson 2002); as $c \rightarrow \infty$, all conditional probabilities $\pi_k(\epsilon)$ go to 0.5 and hence $F_k(t) \rightarrow G_{\theta_k}(t)$ with probability one for all $t > 0$. We assign c a gamma prior $c \sim \Gamma(a, b)$; typically $a = 10$ or 5 and $b = 1$; motivation for these priors are provided in Hanson, Kottas, and Branscum (2008) using the prior L_1 distance between F_k and G_{θ_k} . For $c \sim \Gamma(5, 1)$ the median L_1 distance of the random tailfree density from the centering distribution is 0.28 with 95% probability interval (0.11, 0.76); for $c \sim \Gamma(10, 1)$ these values are

0.19 and (0.08,0.51). So $\Gamma(5, 1)$ typically allows about 30% more mass to be moved than $\Gamma(10, 1)$, as we would expect. The model under the alternative hypothesis H_1 is summarized in terms of interfailure times as

$$\begin{aligned}
t_i - t_{i-1} | d_{i-1} = 0 &\stackrel{ind.}{\sim} F_0(\cdot), \\
t_i - t_i^* | d_{i-1} = 1 &\stackrel{ind.}{\sim} \frac{F_1(\cdot)}{S_1(t_{i-1} - t_i^*)}, \\
F_0 | \boldsymbol{\theta}_0, c &\sim TF^J(c, \rho, G_{\boldsymbol{\theta}_0}), \\
F_1 | \boldsymbol{\theta}_1, c &\sim TF^J(c, \rho, G_{\boldsymbol{\theta}_1}),
\end{aligned}$$

where $TF^J(c, \rho, G_{\boldsymbol{\theta}_k})$ is shorthand for the random tailfree F_k given through (2.6), (2.7), (2.8) and (2.9) up to level J . The model under null hypothesis H_0 simply replaces F_1 by F_0 and $\boldsymbol{\theta}_1$ by $\boldsymbol{\theta}_0$ above. The two models are referred as M_1 and M_0 with respect to H_1 and H_0 .

The model with a common Weibull centering distribution $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_1$ is a linear dependent tailfree process (Jara and Hanson, 2011) regressed on a binary predictor (maintenance decisions), albeit with a likelihood involving truncated observations, e.g. $F_1(\cdot)/S_1(t_{i-1} - t_i^*)$ for $d_{i-1} = 1$ under H_1 . This model generalizes the Polya tree in the same spirit as De Iorio et al. (2004) generalize the celebrated Dirichlet process through an ANOVA-type structure. Under this model the $e^{\lambda_1(\boldsymbol{\epsilon}) - \lambda_0(\boldsymbol{\epsilon})}$ are interpreted as how the odds of failing in the time interval $B_{\boldsymbol{\theta}}(\boldsymbol{\epsilon})$ change from minimal to perfect repair; this information can be useful for finding time intervals $B_{\boldsymbol{\theta}}(\boldsymbol{\epsilon})$ where minimal repair fixes the problem in a manner substantially worse than F_0 would allow. Under the model where $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_1$, if each pair of $\lambda_0(\boldsymbol{\epsilon}_0), \lambda_1(\boldsymbol{\epsilon}_0)$ are assigned identical and independent priors, then $E\{F_0(t)\} = E\{F_1(t)\} = G_{\boldsymbol{\theta}}(t)$ for all $t > 0$ and hence the null model M_0 is formally nested in the alternative model M_1 . From many simulations (beyond what is included in this paper), allowing distinct $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ increases discriminatory ability, but also inflates type I error.

Testing H_0 versus H_1

As stated in the introduction, the test for the assumption of “minimal repair” is of interest, and so it is important to choose a measure to compare the models. As H_0 is formally nested in H_1 , a likelihood-ratio type test could be considered, or a Bayes factor (the Bayesian equivalent). However, computing the Bayes factor with the truncated data likelihoods (2.2) and (2.3) is challenging and existing methods are unstable (Hanson, 2006). Instead we consider an alternative measure, termed the log pseudo-marginal likelihood (LPML) (Geisser and Eddy, 1979), a measure of a model’s predictive ability. The LPML is easy to compute based on MCMC output (Gelfand and Dey, 1994). By definition,

$$\text{LPML} = \sum_{i=1}^n \log\{f_i(t_i|\mathbf{t}_{-i})\}.$$

Here, $f_i(t_i|\mathbf{t}_{-i})$ is the predictive density for t_i based on the remaining data $\mathbf{t}_{-i} = \{t_j : j \neq i\}$, $f_i(\cdot|\mathbf{t}_{-i})$, evaluated at t_i . This is called the i th conditional predictive ordinate (CPO) statistic, and measures how well t_i is predicted from the remaining \mathbf{t}_{-i} through the model. In our context, we compute the predictive density ($\delta_i = 1$) or survival ($\delta_i = 0$) at t_i based on the failure times, repair times and repair decisions during $[0; t_{i-1}]$ and $[t_{i+1}, t_{max}]$, plus partial information during (t_{i-1}, t_{i+1}) that a certain type of repair was performed at t_i . The LPML simply aggregates the log of these. The difference in LPML measures between H_1 and H_0 can be exponentiated giving the pseudo Bayes factor BF_{10} for the two models. Common interpretations for Bayes factors apply, e.g. $3 < BF_{10} < 20$ indicates “positive” evidence toward H_1 ; $20 < BF_{10} < 150$ indicates “strong” evidence and $BF_{10} > 150$ indicates “very strong” evidence (Kass and Raftery, 1995). Under mild conditions the LPML converges to the posterior score and so the pseudo Bayes factor is related to Aitkin’s posterior Bayes factor (Aitkin, 1991) as well. In simulations we find the LPML to work well in differentiating H_1 from H_0 .

The LPML is approximated by

$$\text{LPML} = - \sum_{i=1}^n \log \left\{ \frac{1}{s} \sum_{k=1}^s \frac{1}{p_i(t_i | \mathcal{D}, \boldsymbol{\tau}^k)} \right\}, \quad (2.10)$$

where p_i is the likelihood contribution of event at time t_i , \mathcal{D} is the observed data and (2.2) and (2.3) define the model under H_0 and H_1 respectively; $\{\boldsymbol{\tau}^k, k = 1, 2, \dots, s\}$ are iterates from MCMC outputs of all the parameters, i.e $\{\boldsymbol{\lambda}_0^k, \boldsymbol{\lambda}_1^k, c^k, k = 1, 2, \dots, s\}$ under H_1 .

Model fitting

Following the discussion at the end of Section 2.1, for the purposes of testing $H_0 : F_0 = F_1$ we suggest that F_0 and F_1 have the same prior mean Weibull distribution in fitting M_1 ; e.g. $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_1 = \boldsymbol{\theta}$. In simulations, we fix $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_1 = \hat{\boldsymbol{\theta}}$, where $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimate assuming minimal repair holds and Weibull reliability, i.e. $F_0(t) = F_1(t) = G_{\boldsymbol{\theta}}(t)$ (the Weibull is obtained under the tailfree prior when $c \rightarrow \infty$). A similar practice is recommended by Berger and Guglielmi (2001) and Hanson, Branscum, and Gardner (2008) in simpler situations.

Upon rejecting H_0 , we suggest refitting M_1 , allowing distinct $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$. It is known in the literature that fixing $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ results in “jumpy” densities as each of f_0 and f_1 have discontinuities at each partition interval endpoint, as Figure 2.1 shows. For the purposes of estimating the interfailure hazard functions $\hat{h}_0(t)$ and $\hat{h}_1(t)$ we suggest an empirical Bayes approach: place normal priors on $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ derived from their large sample asymptotic distributions under the underlying Weibull assumption ($c \rightarrow \infty$):

$$\boldsymbol{\theta}_k \stackrel{ind.}{\sim} N_2(\hat{\boldsymbol{\theta}}_k, \boldsymbol{\Sigma}_k), \quad (2.11)$$

where $\boldsymbol{\Sigma}_k$ is the large-sample covariance matrix under a frequentist Weibull fit ($\hat{\boldsymbol{\theta}}_k$ and $\boldsymbol{\Sigma}_k$ are easily obtained using optimization procedures in R or SAS); noninformative priors ($p(\boldsymbol{\theta}_k) \propto 1$) for $\boldsymbol{\theta}_k$ can also be used. Placing priors on $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ smooths out

the estimated density and hazard curves, yielding a mixture of tailfree processes for F_0 and F_1 .

MCMC computing requires full specification of the likelihoods and priors. The likelihoods (2.2) and (2.3) under H_0 and H_1 , are functions of f_0 and (f_0, f_1) respectively. Let $\mathcal{E} = \{\boldsymbol{\epsilon} = \epsilon_1 \cdots \epsilon_j, j = 1, \dots, J-1\}$. Conditioning on $\{\boldsymbol{\pi}_k(\boldsymbol{\epsilon}0)\}$ for $\boldsymbol{\epsilon} \in \mathcal{E}$, the densities (f_0 and f_1) and reliability functions (S_0 and S_1) are given by (2.7) and (2.6) in terms of probability vectors \mathbf{p}_k , functions of $\boldsymbol{\pi}_k(\boldsymbol{\epsilon}0)$ defined in (2.5). Note that $\boldsymbol{\pi}_k(\boldsymbol{\epsilon}0)$ is a function of $\boldsymbol{\lambda}_k$ through (2.8). The posterior under H_1 is proportional to

$$p(\boldsymbol{\lambda}, \boldsymbol{\theta}_0, \boldsymbol{\theta}_1, c | \mathcal{D}) \propto \mathcal{L}(f_0, f_1) p(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1) \Gamma(c | a, b) \prod_{k=0}^1 \prod_{\boldsymbol{\epsilon} \in \mathcal{E}} N\left(\lambda_k(\boldsymbol{\epsilon}0) | 0, \frac{2}{c_j^2}\right),$$

where $\boldsymbol{\lambda} = \{\boldsymbol{\lambda}_0(\boldsymbol{\epsilon}0), \boldsymbol{\lambda}_1(\boldsymbol{\epsilon}0)\}$, $\mathcal{L}(f_0, f_1)$ is defined in equation (2.3) and $p(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1)$ are product of independent priors for $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$. The posterior under H_0 is similar.

Parameters $\{\boldsymbol{\lambda}, \boldsymbol{\theta}_0, \boldsymbol{\theta}_1\}$ are updated using random-walk Metropolis–Hastings updates (Tierney, 1994). Gaussian random-walk proposals are used for each element of $\{\lambda_k(\boldsymbol{\epsilon}0) : k = 0, 1; \boldsymbol{\epsilon} \in \mathcal{E}\}$,

$$\lambda_k(\boldsymbol{\epsilon}0)^* \sim N(\lambda_k(\boldsymbol{\epsilon}0), v_k(\boldsymbol{\epsilon}0)),$$

where $\lambda_k(\boldsymbol{\epsilon}0)^*$ is the latest accepted value for $\lambda_k(\boldsymbol{\epsilon}0)$ and $v_k(\boldsymbol{\epsilon}0)$ is tuned to get acceptance rates in the 20% to 50% range. Similarly, $\boldsymbol{\theta}_k \sim N_2(\boldsymbol{\theta}_k^*, V_k)$ where V_k needs to be tuned. We have found automatic tuning of $v_k(\boldsymbol{\epsilon})$ and V_k to proceed quickly (Haario, Saksman, and Tamminen, 2005). Specifically, let the sequence $\lambda_k^1(\boldsymbol{\epsilon}0), \lambda_k^2(\boldsymbol{\epsilon}0), \dots$ be the states of the Markov chain for $\lambda_k(\boldsymbol{\epsilon}0)$. When deciding the t -th state $\lambda_k^t(\boldsymbol{\epsilon}0)$, we sample $\lambda_k(\boldsymbol{\epsilon}0)^* \sim N(\lambda_k^{t-1}(\boldsymbol{\epsilon}0), v_k^t(\boldsymbol{\epsilon}0))$ with

$$v_k^t(\boldsymbol{\epsilon}0) = \begin{cases} v^0(\boldsymbol{\epsilon}0), & t < t_0 \\ s \text{Var} \{ \lambda_k^1(\boldsymbol{\epsilon}0), \dots, \lambda_k^{t-1}(\boldsymbol{\epsilon}0) \} + s_0, & t > t_0 \end{cases}$$

where s is recommended to be 2.4, s_0 is a small constant and $v_k^0(\boldsymbol{\epsilon}0)$ is an initial variance of the proposal distribution. A similar automatic procedure applies to $\boldsymbol{\theta}_k$

with V_k^t being the empirical covariance matrix after t_0 . The parameter c is updated through posterior

$$p(c|\boldsymbol{\lambda}, \boldsymbol{\theta}, \mathcal{D}) \sim \Gamma \left\{ (a + 2^J - 1), b + \sum_{\epsilon_1 \epsilon_2 \dots \epsilon_j \in \mathcal{E}} \sum_{k=0}^1 \boldsymbol{\lambda}_k (\epsilon_1 \epsilon_2 \dots \epsilon_j)^2 j^2 / 4 \right\}.$$

FORTRAN 90 code for fitting the data analysis in Section 4 is included in the on-line supplementary material for this paper.

2.3 SIMULATIONS

We conducted four simulations to see how well the pseudo Bayes factor can discriminate between H_0 and H_1 and one simulation to illustrate the estimation of the reliability functions. Simulation I involves a sequence of increasing departures of f_1 from f_0 according to our alternative model M_1 . Simulation II considers a sequence of departures from H_0 using the effective age models; details are presented in simulation II. Simulation III investigates type I error. Simulation IV examines how the prior on c affects the test. Simulation V estimates reliability functions F_0 and F_1 . Sample sizes for simulated data are the total number of interfailure times after perfect or minimal repair. Each dataset is comprised of one third interfailure times after perfect repair and two thirds interfailure times after minimal repair truncated from the accumulated age since the most recent perfect repair. For simplicity, all repairs occur in response to failures. For the hypothesis tests, the unknown distributions are assigned finite tailfree priors with the following specifications. We fix $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ for F_0 under M_0 and $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_1 = \hat{\boldsymbol{\theta}}$ for F_0, F_1 under M_1 where $\hat{\boldsymbol{\theta}}$ is an estimate for $\boldsymbol{\theta}$ under the Weibull null model; $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ each contain the log of the Weibull shape and scale parameters; the level of the partition tree is fixed at $J = 5$, and c is considered with prior $\Gamma(5, 1)$ for simulations I-III and priors $\Gamma(5, 1)$ and $\Gamma(10, 1)$ for simulation IV. Based on our simulation experience, $J = 5$ is sufficient for the sample sizes in our simulations and increasing J changes the LPML negligibly. For each dataset, we run 4000 MCMC

iterations and use the last 3000 MCMC samples for inferences. We reject H_0 in favor of H_1 if the LPML for H_1 is greater than that for H_0 by 3.5, otherwise we choose H_0 . For estimating the reliability functions in simulation V, we place the empirical Bayes priors as detailed in Section 2.3 on θ_0 and θ_1 , and assume $c \sim \Gamma(10, 1)$. After a burn-in of 10,000 iterates, 400,000 iterates were thinned to a sample of 4,000 for inference.

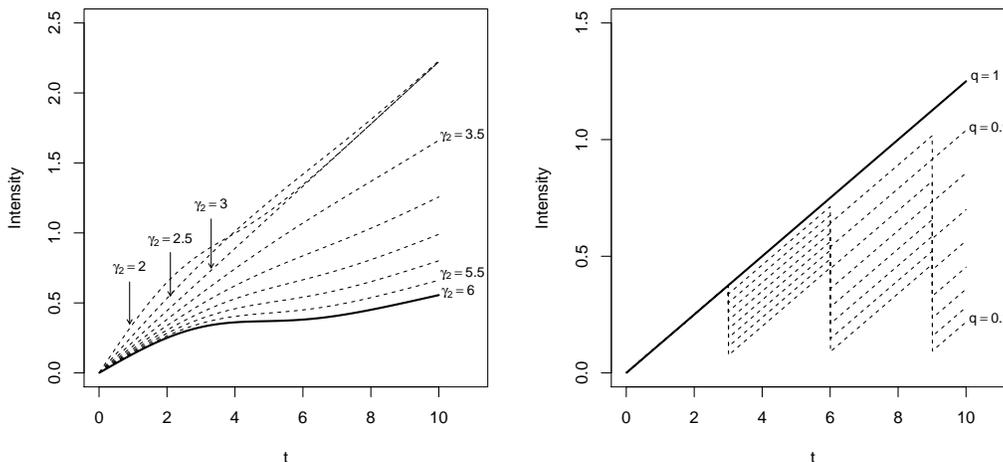


Figure 2.2: Left panel plots the hazard h_0 (solid and thick line) and 8 choice for h_1 (dashed and thin lines) versus time t for simulation I; right panel plots the intensity of the system versus time t when failures occur at $\{3, 6, 9\}$ for all q from 1 (solid and thick line) and 0.2 – 0.9 (dashed and thin lines) for simulation II.

Simulation I: Let $W(w, \alpha_1, \gamma_1, \alpha_2, \gamma_2)$ be a mixture of two Weibull distributions with weights w and $1 - w$, shape parameters α_1, α_2 and scale parameters γ_1, γ_2 . Table 2.2 reports the results of simulation I testing H_0 versus increasing departures H_1 according to our alternative model; $f_0 = W(0.5, 2, 3, 2, 6)$, $f_1 = W(0.5, 2, 3, 2, \gamma_2)$ with nine densities corresponding to $\gamma_2 = \{6, 5.5, 5, 4.5, 4, 3.5, 3, 2.5, 2\}$. The simulation involves two sample sizes $n = 200$ or $n = 500$. For each condition, 200 datasets are simulated. The hazard functions for f_0 (solid thick line) and f_1 (dashed thin lines) are plotted in the left panel of Figure 2.2; Table 2.2 values are proportions rejecting H_0 . The power is reasonably good in detecting this sequence of departures.

Table 2.2: Type I error and power for testing H_0 vs. H_1 for simulation I; 6.0 – 2.0 are nine values of γ_2 in defining f_1 ; tabled values are the proportion out of 200 replications where H_0 is rejected.

γ_2	6.0	5.5	5.0	4.5	4.0	3.5	3.0	2.5	2.0
Sample event									
$n = 200$	0.00	0.01	0.04	0.14	0.39	0.58	0.85	0.94	0.96
$n = 500$	0.04	0.05	0.14	0.50	0.93	1.00	1.00	1.00	1.00

Simulation II: We use the notion of imperfect repairs and effective age to introduce increasing departures from the minimal repair assumption but using our proposed models M_0 and M_1 for hypothesis testing. Effective age modeling in reliability has received a lot of attention since introduced by Kijima (1989). A spectrum of imperfect repairs can be modeled through effective ages, generating event processes which include renewal processes and Poisson processes as special cases. Following the notation introduced in Section 2.1, the times for repairs are recorded as $0 = t_0 < t_1 < t_2 < \dots < t_n = t_{max}$. The repair at t_i is denoted as d_i with $d_i = 0$ if perfect repair was performed and $d_i = 1$ otherwise. Define $z(t)$ as the effective age of the system at time t . We still assume perfect repairs reset the effective age to zero but now repairs recorded as $d_i = 1$ multiply the effective age right before the repair by a fraction q (known as Kijima type II model). That is, $z(t_i) = 0$ if $d_i = 0$, $z(t_i) = \{z(t_{i-1}) + t_i - t_{i-1}\} q$ if $d_i = 1$ and $z(t) = z(t_i) + t - t_i$ for $t_i < t < t_{i+1}$. This is only one departure from the null model M_0 that is different from our alternative model M_1 (Presnell et al. 1994). Note that $q = 1$ implies that the minimal repair assumption holds and $q < 1$ indicates repairs being better than “good as old”. Suppose F_0 defines the cumulative distribution function of the first failure time for a system; after repair at t_i , the distribution for the time to next failure is $F_0(z(t_i) + t)/S(z(t_i))$, $t > 0$. In the following simulation, we examine the power of the proposed test for a sequence of q using two sample sizes $n = 200$ or $n = 500$. For each condition, 200

datasets are simulated.

Table 2.3 reports the results of testing H_0 versus H_1 ; q takes values from 1 to 0.2 by 0.1; $f_0 = \text{Weibull}(2, 4)$; the intensities for a system with failures at $\{3, 6, 9\}$ are plotted in the right panel of Figure 2.2; the solid thick line corresponds to $q = 1$ and dashed lines correspond to $q < 1$ from 0.9 to 0.2; the tabled values are percentages of times rejecting H_0 . Even though the Kijima departure is not in the realm of our model, our test performs satisfactorily with power increasing to one as q gets small for $n = 500$. According to additional simulations, not included here, the power of our test increases when the slope of the hazard increases in the Kijima models.

Table 2.3: Type I error and power for testing H_0 vs. H_1 for simulation II; 0.2 – 1.0 represent nine choices of q and 1.0 represents no departure of minimal repair; tabled values are the proportion out of 200 replications where H_0 is rejected.

q	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2
Sample event									
$n = 200$	0.01	0.02	0.03	0.04	0.11	0.20	0.29	0.53	0.72
$n = 500$	0.03	0.04	0.06	0.21	0.53	0.81	0.94	0.99	1.00

Simulation III: We perform a simulation to investigate type I error using three sample sizes $n \in \{1000, 1500, 2000\}$, and three choices for f_0 . Table 2.4 reports the results of the third simulation based on 200 datasets where data are simulated only from M_0 ; the three densities are $W(0.5, 2, 3, 2, 6)$, $\text{Weibull}(2, 4)$, and $\text{Weibull}(1, 4)$; the tabled values are proportions rejecting H_0 . The type I errors appear to be stable and are always less than 0.05 for these distributions and sample sizes. Tables 1 and 2 shows that the LPML cutoff of 3.5 may be conservative for smaller sample sizes.

Simulation IV: We also conduct a simulation to see how the prior on c affects the test. 100 different data sets for each of three sample sizes $n = 200, 500, 1000$ (300 data sets total) were generated from model M_0 with $f_0 = W(0.5, 2, 3, 2, 5)$ where the minimal repair assumption holds, as well as model M_1 where F_0 and F_1 are different

Table 2.4: Type I error for testing H_0 vs. H_1 for simulation III; 1 – 3 represents three choices of f_0 described in the text.

Density type	1	2	3
Sample event			
$n = 1000$	0.03	0.04	0.04
$n = 1500$	0.03	0.05	0.03
$n = 2000$	0.04	0.05	0.04

Table 2.5: Type I error and power for testing H_0 vs. H_1 with two sets of prior for c for simulation IV; tabled values are the proportion out of 100 simulated data sets where H_0 is rejected.

Sample event	$H_0 : F_0 = F_1$		$H_1 : F_0 \neq F_1$	
	$c \sim \Gamma(5, 1)$	$c \sim \Gamma(10, 1)$	$c \sim \Gamma(5, 1)$	$c \sim \Gamma(10, 1)$
$n = 200$	0.00	0.00	0.75	0.60
$n = 500$	0.00	0.00	0.81	0.70
$n = 1000$	0.05	0.02	1.00	1.00

with $f_0 = W(0.2, 2, 0.7, 2, 5)$, $f_1 = W(0.5, 2, 0.7, 2, 3)$. Now c is considered with two priors, $\Gamma(5, 1)$ and $\Gamma(10, 1)$. Table 2.5 reports the results of testing H_0 versus H_1 ; the tabled values are percentages of times rejecting H_0 . The $\Gamma(5, 1)$ prior favors smaller values of c , yielding more modeling flexibility, and hence increasing differentiability. This effect is more obvious for smaller sample sizes.

Simulation V: The last simulation illustrates our approach by estimating reliability functions after perfect and minimal repairs for three simulated data sets. One data set of 1000 events was simulated from the above setting of M_1 . Two datasets of 1000 events were simulated from the Kijima type II effective age model with $q = 0.2, 0.5$ and $f_0 = \text{Weibull}(2, 4)$. For the first dataset, the true reliability functions S_0 and S_1 , and hazards h_0 and h_1 are displayed in the left panels of Figure 2.3 with F_0 plotted using solid lines and F_1 plotted using short-dashed lines. The estimated survival and hazard (pointwise posterior means) are plotted in the right panels of Figure 2.3, along with

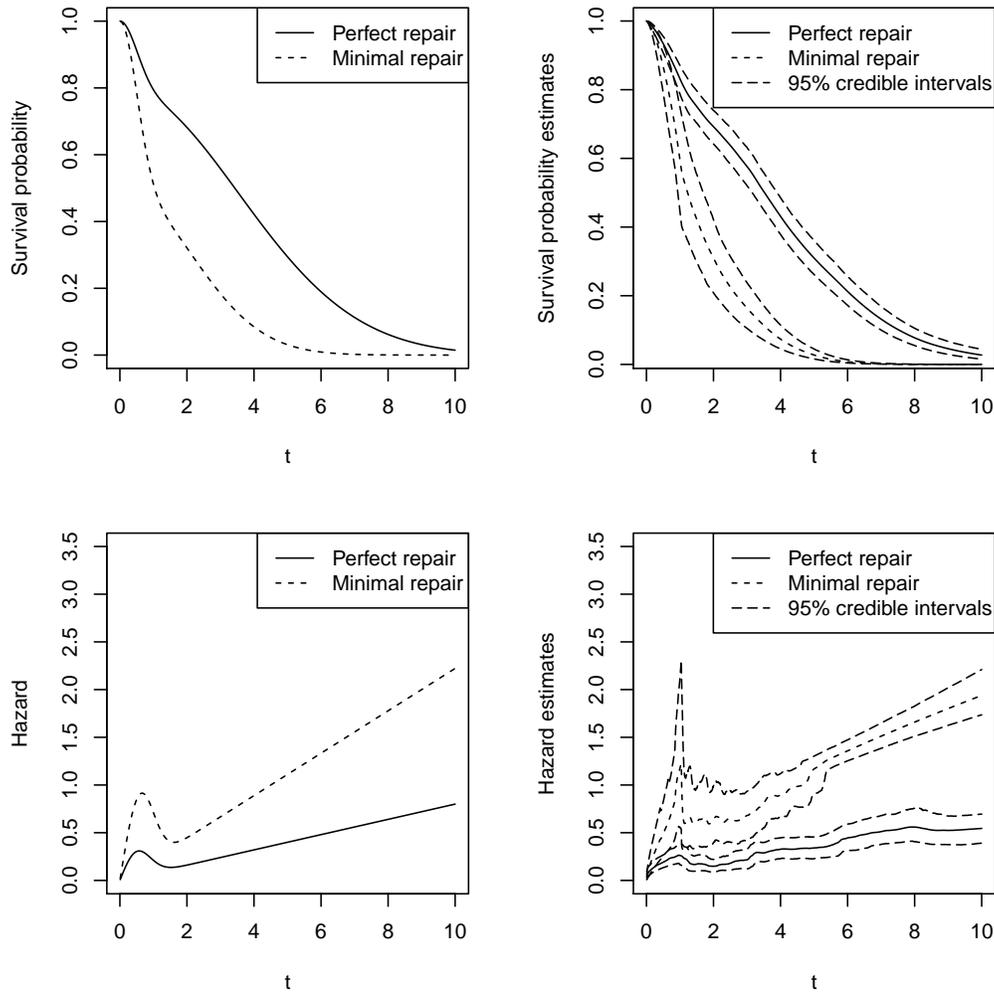


Figure 2.3: Results of a simulated sample of $n = 1000$ events under H_1 ; true (left) and estimated (right) survival and hazard estimates versus time t ; solid lines correspond to F_0 and short-dashed lines correspond to F_1 ; long-dashed lines correspond to 95% credible intervals.

the 95% credible intervals for the estimates (long-dashed lines); we can see that local features of the distributions are well captured. For the other two datasets, we plot the estimated hazard functions for h_0 (solid black) and h_1 (dashed black) in Figure 2.4, overlaid with intensities (dashed gray) of ten systems. We plot each intensity function over time since the first minimal repair for the corresponding system. For both datasets, our model's h_1 estimates essentially average the true intensities of the ten systems. We can see a larger difference between h_1 and h_0 estimates for $q = 0.2$

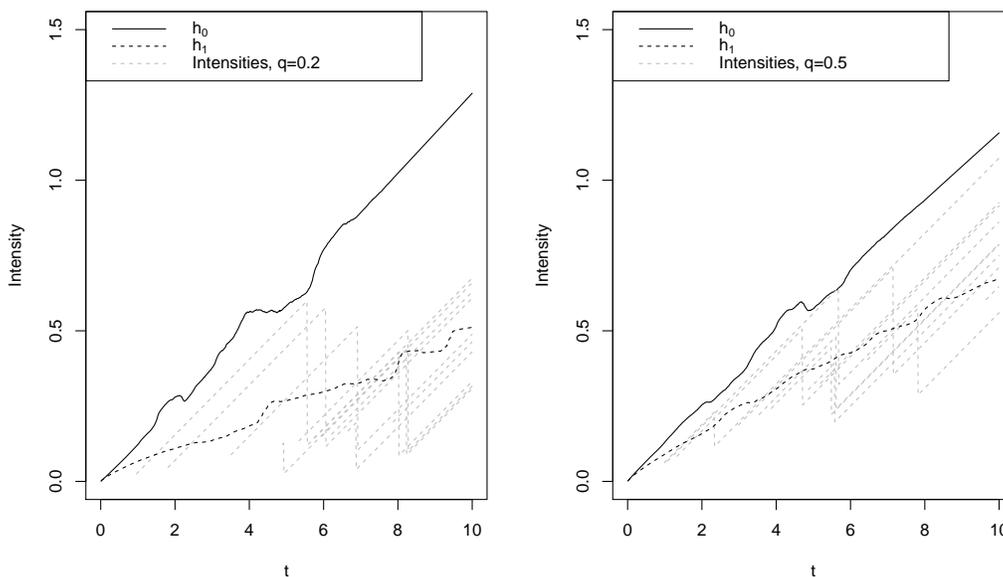


Figure 2.4: Results of two simulated samples of $n = 1000$ interfailure times under Kijima type II model with $q = 0.2$ (left) and 0.5 (right); hazard estimates of h_0 (solid black) and h_1 (dashed black) versus time t ; dashed gray lines are intensities of ten systems.

than that for $q = 0.5$, indicating better performances of system when $q = 0.2$.

The computing time for running the above simulations mainly depends on the number of MCMC iterates, sample size and the level of the partition tree J . For $J = 5$, 4000 MCMC iterates and a 3.00 GHz processor, it may take a few seconds for small sample sizes (e.g. 200, 500) to a couple of minutes for large sample sizes (i.e. 2000). The computing times for longer chains or higher levels of J are longer.

2.4 DATA ANALYSIS

We studied the data set provided by South Texas Project Electric Generating Station for the essential chillers system. Details on the chillers and maintenances are presented in the introduction. The original calendar time was recorded in days and we divided the time by 30.4 to transform the units to months. Calendar times of maintenance events in months are plotted in Figure 2.5 by chillers where vertical bars

represent censored failure times on the top line and observed failure times on the bottom line. The size of the vertical bar indicates perfect or minimal repair and the number of each type of repair per chiller is given in Table 2.6. We firstly investigate whether the six chillers are identical in new condition. Chillers (AC 4 to 6) in group 2 tend to last longer than Chillers (AC 1 to 3) in group 1 based on the Kaplan-Meier estimates (Figure 2.6) using the first failures after perfect repairs. Within each group, there appears no significant difference. The log-rank tests for homogeneity of survival curves for the first failures give a significant p-value of 0.03 across the two groups, but non-significant 0.26 for chillers within the first group, and 0.78 for chillers within the second group. Therefore we pool observations across chillers within each group and perform separate analyses for the two groups. From now on, the first group is referred as “group 1” and the second group as “group 2”.

Table 2.6: Counts of perfect/minimal by response to “failure” / “censored” for each chiller.

	AC1		AC2		AC3	
	Failure	Censored	Failure	Censored	Failure	Censored
Perfect	17	217	18	199	15	202
Minimal	232	4	179	1	184	3
	AC4		AC5		AC6	
	Failure	Censored	Failure	Censored	Failure	Censored
Perfect	14	189	9	184	13	184
Minimal	160	1	167	2	163	3

It is of interest to test the minimal repair assumption, i.e. whether there is a significant difference between the reliability distributions for the two types of maintenance decisions. We first fit the proposed nonparametric test. For group 1, the LPML for H_0 is -770 and for H_1 is -755 . For group 2, the LPML for H_0 is -713 and for H_1 is -695 . Exponentiating the LPML differences in the two groups (> 150) leads to strongly rejecting H_0 in both. We also fit parametric tests with Weibull

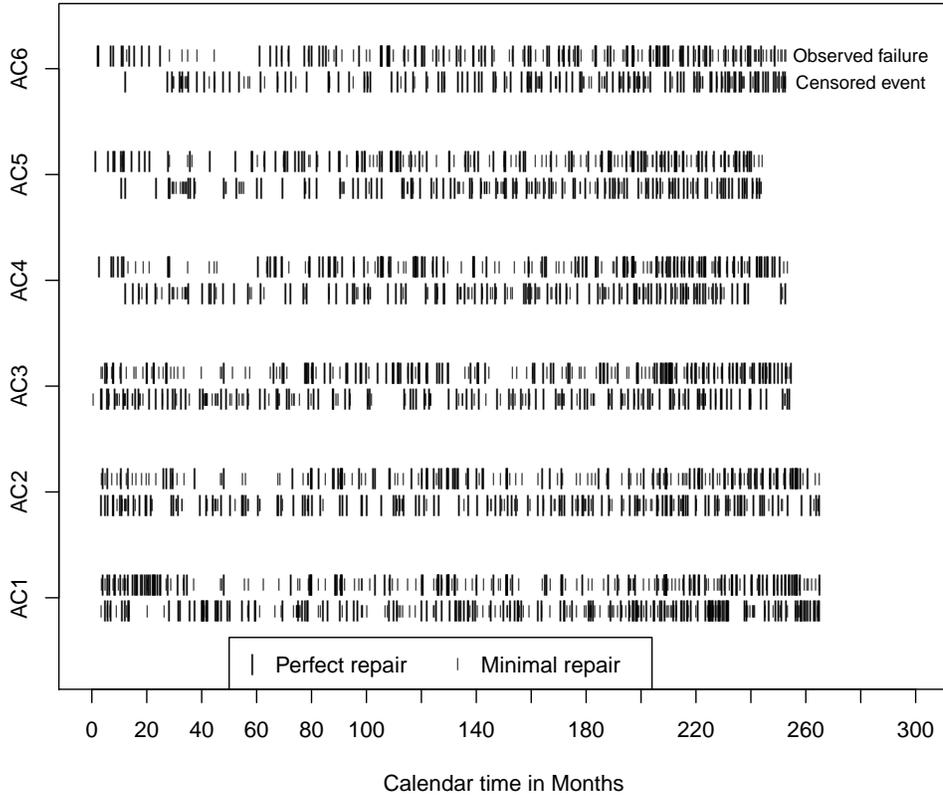


Figure 2.5: Calendar times of events for the six chillers (AC); for each chiller, vertical bars represent observed failures on the top line and censored events on the bottom line; big (small) vertical bar denotes perfect (minimal) repair at the event time.

family assumption for F_0 and F_1 . For group 1, the LPML for H_0 is -770 and for H_1 is -760 . For group 2, the LPML for H_0 is -714 and for H_1 is -694 . Note that the nonparametric method yields greater difference in LPMLs for group 1 than parametric method does. For group 2, there is not a significant difference between the parametric and nonparametric method. For estimating F_0 and F_1 , we refit M_1 using the nonparametric method presented in Section 2.3, place noninformative priors on θ_0 and θ_1 ($p(\theta_k) \propto 1$) and assume $c \sim \Gamma(10, 1)$ for the two groups. After a burn-in of 50,000 iterates, 4,000 MCMC samples were thinned from a total of 400,000 iterates. The computing time was about 30 seconds for hypotheses testing and a few

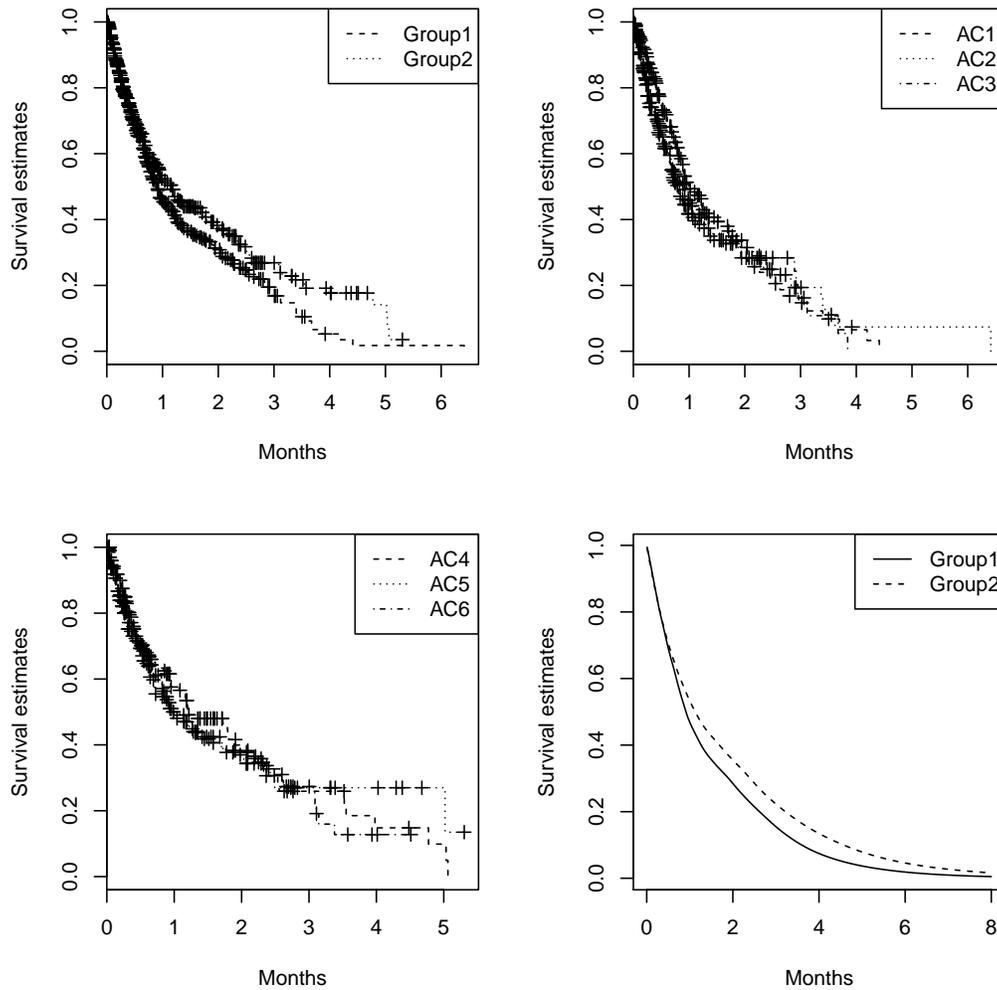


Figure 2.6: Kaplan-Meier estimates for survival functions using first failures after perfect repairs of the essential chillers system; group 1 and 2 (top left) ; AC 1 to 3 in group 1 (top right); AC 4 to 6 in group 2 (bottom left); bottom right panel plots M_1 estimates for S_0 for group 1 and 2.

minutes for estimation in M_1 . We plot the estimated point-wise posterior mean survival functions for F_0 (solid lines) and F_1 (short-dashed lines) on the left panel of Figure 2.7. The 95% credible intervals for the survival functions are plotted with long-dashed lines. The right panel of Figure 2.7 are the estimated point-wise posterior mean hazard functions for F_0 (solid lines) and F_1 (short-dashed lines) from both the nonparametric (less smooth) and parametric (smooth) approach. The nonparametric estimates for h_0 and h_1 for group 1 (top right panel) are close to each other

in the first one month, but after h_1 is larger than h_0 . This implies that the system performs “good as old” after minimal repairs at a young cumulative age but worse at an older cumulative age (the cumulative age is the time since the latest perfect repair). The posterior mean estimate for h_1 is always greater than h_0 for group 2 (bottom right panel) indicating the system performs worse than expected after minimal repairs. Compared to the estimates of h_0 and h_1 from the parametric method, the nonparametric estimates exhibit much more flexibility in the shape during the first few months where data are more plentiful, but follow a Weibull shape as time increases and data are scarcer.

2.5 DISCUSSION

We proposed a flexible Bayesian nonparametric framework to model recurrent events in a repairable system for the purpose of generalizing and testing the common “minimal repair” assumption. Upon system failure either a perfect or a minimal repair is performed. Tailfree priors are assumed for the unknown distributions F_0 and F_1 centered at the Weibull distribution. The Weibull serves to anchor inference and guide density shape where data are scarce, but tailfree probabilities change the Weibull shape when necessary in locations where data are plentiful. The typical assumption that a minimal repair brings the system back to the exact state it was in right before failure is tested by via pseudo Bayes factors. In simulations, the test was found to have good power, and appropriate Type I error. If the alternative model $H_1 : F_0 \neq F_1$ is preferred, we further compare the estimated hazard functions, shedding light on how minimal repairs perform relative to perfect repairs at different ages of the system. This is particularly useful for managers to schedule maintenance. If the null model is preferred, our model becomes a Bayesian nonparametric generalization of Weibull for modeling the failure times from non-homogeneous Poisson processes. It is then typically of interest to obtain smooth estimates for the density, hazard and survival

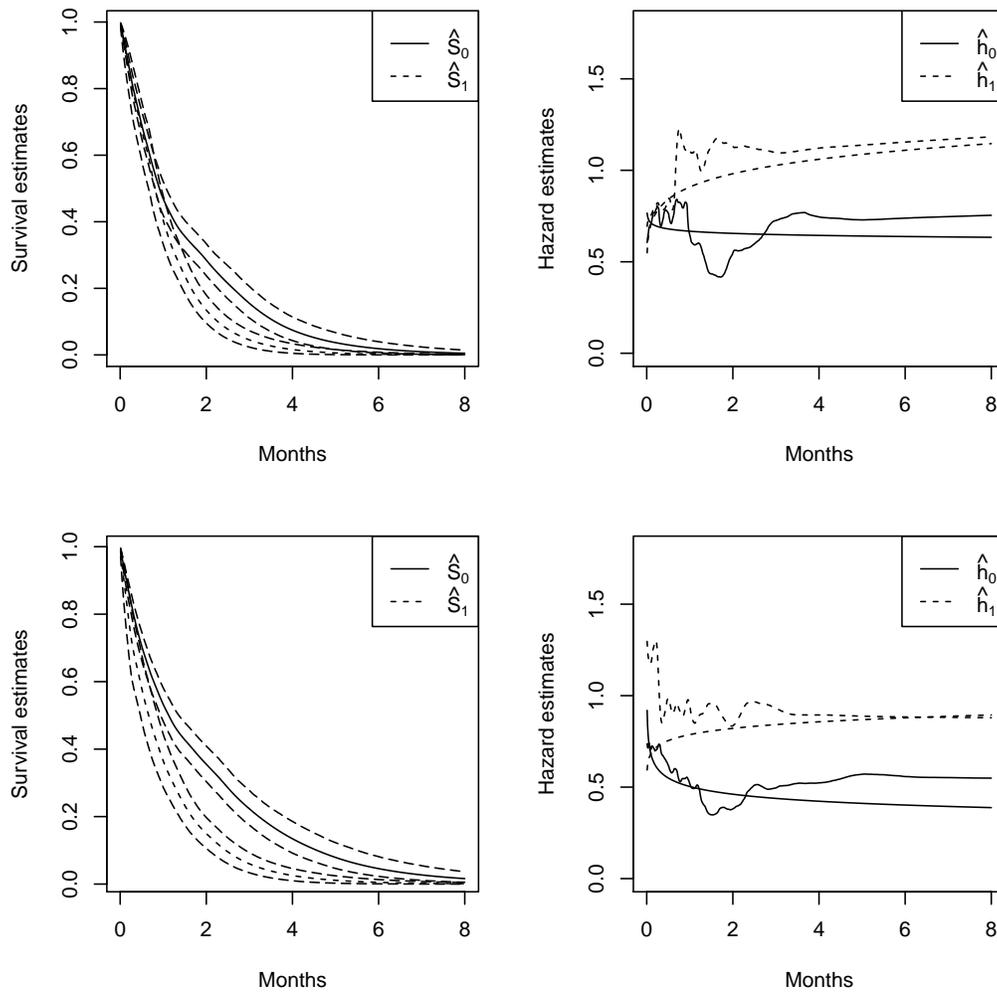


Figure 2.7: This figure contains estimates of the survivor and hazard functions for group 1 (top panels) and 2 (bottom panels) essential chillers system when both parametric and nonparametric models are fitted for M_1 . Left panels plot nonparametric estimates of the survivor functions corresponding to F_0 (solid) and F_1 (short-dashed) and their 95% credible intervals (long-dashed). Right panels plot the parametric (smooth) and nonparametric (less smooth) estimates of the hazard functions corresponding to F_0 (solid) and F_1 (short-dashed).

function. With slight changes in the likelihood, our method can also be used to test other repair assumptions, e.g. a known life supplement for a type of repair. We note that it is straightforward to include time-dependent covariates into the model, such as operating settings, the identity of the person making repairs, etc.

We stress that perfect repairs are indeed assumed to bring the system to as “good

as new.” In practice, there may be several types of maintenances pooled together which are close to “perfect repairs”. If several identical systems are maintained in the same way, maintenance records may be combined since identical systems have the same contribution as old systems which have just received an overhaul. However, combining the records from systems which are very different could result in confounding between the actual effects of the maintenances and the reliability of the system.

We also assume that the system after each minimal repair depends on the preceding minimal repairs only through the accumulated age (time since last perfect repair). The minimal repairs are “good as old” repairs with respect to F_1 . That is, the hazard function remains $h_1(t)$ over time after the first minimal repair in each cycle. This simplification facilitates the testing of H_0 versus H_1 and also allows comparison of maintenance decisions over time. However, if $H_1 : F_0 \neq F_1$ is concluded, the estimated h_1 may not be the dynamic hazard for the system after minimal repairs. The reason is that when minimal repairs are not “good as old” repairs, the actual effects of maintenances could aggregate, changing the hazard function after each repair. The “effective age” modeling of the system (Kijima, 1989) captures a dynamically changing hazard. However there is difficulty in determining the degree of each repair and hence the effective age for the system. It is one of our interests to model the dynamic hazard using Bayesian nonparametric methods.

CHAPTER 3

A BAYESIAN SEMIPARAMETRIC REGRESSION MODEL FOR RELIABILITY DATA USING EFFECTIVE AGE

A new regression model for recurrent events from repairable systems is proposed. The effectiveness of each repair in Kijima models I and II is regressed on repair-specific covariates. By modeling effective age in a flexible way, the model allows a spectrum of heterogeneous repairs besides “good as new” and “good as old” repairs. The density for the baseline hazard is modeled nonparametrically with a tailfree process prior which is centered at Weibull and yet allows substantial data-driven deviations from the centering family. Linearity in the predictors is relaxed using a B-spline transformation. The method is illustrated using simulations as well as two real data analyses.

Keywords: Effective age, Repairable system, Tailfree process, Truncated data.

3.1 INTRODUCTION

Repairable systems have been widely studied in the literature. Systems fail, get repaired upon failure, and these recurrent events (failures, repairs) are observed. The event process generating the repeated events is closely related to the intensity function, denoted as $\lambda(t|H(t))$ and formally defined in Section 2.1, which describes the probability of an instantaneous new failure, given the history of maintenances

The content in this chapter is a reprint by permission of Elsevier for “ Li, L. and Hanson, T. (2014). A Bayesian semiparametric regression model for reliability data using effective age. Computational Statistics and Data Analysis, 73, 177–188”.

and failures $H(t)$. In general, recurrent event modeling methods can be divided into categories based on the type of repairs a system receives. Renewal processes are used if all the repairs bring the system to the “good as new” state and Poisson processes are used if all the maintenances bring the system to a “good as old” state. Kijima (1989) introduced two classes of models using the notion of “effective age” (also known as “virtual age”) of the system to allow for a spectrum of repairs between “good as old” and “good as new”. Consider a system observed over $[0, \tau]$. Assume the repair times for the system are $0 < t_1 < t_2 < \dots < t_n$ and denote $\varepsilon(t)$ as the effective age of the system at time t . Suppose that the intensity $\lambda(t|H(t))$ is related to the unknown hazard, or failure rate, of a new system $r(t)$ through $\lambda(t|H(t)) = r\{\varepsilon(t)\}$. Poisson models assume $\varepsilon(t) = t$ and renewal models assume $\varepsilon(t) = t - s_{N(t-)}$ where $s_{N(t-)}$ is the time at which the last repair occurred. Kijima models introduce an age reduction factor D_i for each repair, occurring at calendar time t_i . Define $\varepsilon(t_i) = \varepsilon(t_{i-1}) + [t_i - t_{i-1}]D_i$ for the Kijima type I model and $\varepsilon(t_i) = [\varepsilon(t_{i-1}) + t_i - t_{i-1}]D_i$ for the Kijima type II model. Assume $\varepsilon(t) = \varepsilon(t_{i-1}) + t - t_{i-1}$ for $t \in (t_{i-1}, t_i)$. Note that $D_i = 1$ implies a Poisson process in models I and II, and $D_i = 0$ implies a renewal process in type II.

Lindqvist (2006) provides a review of the modeling of effective age. Dorado, Hollander, and Sethuraman (1997) generalize Kijima’s models that allow for repairs of varying degree by including known “life supplements” – numbers between zero and one indicating the degrees of the repairs. There is very limited literature dealing with unknown effective age processes. Doyen and Gaudoin (2004) studied a class of Kijima’s models where the repairs reduce the effective ages by one overall effectiveness scalar q . Recently, Veber et al. (2008) propose an EM-algorithm to estimate q and use Weibull mixtures for the baseline failure time distribution. Using one scalar is inappropriate for systems where repairs of varying effectiveness occur. For example, different maintenance types, or levels of experience in those carrying out

the repairs, can have drastic differences in repair effectiveness. Very recently, Yuan and Uday (2012) extend the single scalar parameter q to a time-dependent function, e.g. $q(t) = \exp(-et)$ where e is estimated, and assume the baseline distribution to be a parametric power-law distribution. In this work, we regress the effectiveness of each repair on covariates, e.g. materials used or the technician, and relax the parametric assumption of the baseline distribution using nonparametric priors in a Bayesian framework. Time trends in the effectiveness of repairs, characteristics of each repair, and association among repairs within each system can be flexibly coded into the covariate process. Specifically, the effectiveness measure D_i is regressed on a vector of covariates \mathbf{w}_i ; let $D_i = \exp\{\beta' \mathbf{w}_i\} / \{1 + \exp(\beta' \mathbf{w}_i)\}$ or $D_i = \exp(\beta' \mathbf{w}_i)$. The associations between the covariates and the effective age reduction are characterized by β . When the hazard of the system is monotone increasing, a repair with covariates resulting in a smaller age reduction factor D_i tends to be more effective than other repairs performed at the same effective age of the system.

Other generalizations of renewal and Poisson processes allowing for covariates also assume the effective age process $\varepsilon(t)$ is known, including for example, modulated renewal processes (Cox 1972b), point-process models incorporating renewals and time trends (Lawless and Thiagarajah 1996), and a general class of semiparametric models (Peña et al. 2007) which simultaneously accommodates the effects of increasing numbers of events, covariates, interventions (repairs), and association among the interevent times within a system. This literature encompasses a rich and widely used family of reliability models. However, it is difficult to assume that the effective age process is known. There might even be inter-play among the effective age process and history-dependent covariates and the baseline hazard function, as noted in Peña et al. (2007). Moreover, understanding the performance of repairs is often crucial to decision-making and even predictions.

A parametric analysis of our proposed model can be performed by choosing an

appropriate distribution family, e.g. Weibull, for $r(t)$. In this work, we seek a more flexible approach where the entire density, the cumulative hazard function, or the hazard is assigned a nonparametric prior distribution. Bayesian nonparametric priors have achieved prominent success due to their flexibility in modeling unknown distributions; examples include the Dirichlet process (Ferguson 1973), Polya tree priors (Lavine 1992), Dirichlet process mixtures (Escobar and West 1995), etc. However, the use of these nonparametric priors in recurrent event models has been quite limited. Very recently, Taddy and Kottas (2012) used Dirichlet process mixtures for the interfailure density in Poisson process models. Priors on the cumulative hazard $R(t) = \int_0^t r(s)ds$ include the beta and gamma processes (Lo 1992, Kuo and Ghosh 1997) which are discrete and not readily used in our context. The weighted gamma process (Ishwaran and James 2004) is centered at one unique baseline intensity and is also not appropriate for a model that involves a factor in the argument of the intensity. Our proposed framework uses tailfree priors (Freedman 1963, Ferguson 1974, Jara and Hanson 2011), on the space of densities, centered at the Weibull family, but allows for substantial data-driven deviations from the centering families. A special case of the tailfree prior, the Polya tree prior, has been widely used for models that warp the baseline r ; see Hanson(2006), Walker and Mallick (1999), and Hanson and Yang (2007) for applications involving the accelerated failure time model and the proportional odds model. Like the Dirichlet process, tailfree priors also have desirable consistency and large support properties (Jara and Hanson 2011). The general framework proposed herein allows model comparisons using the goodness-of-fit measures LPML and DIC so that comparisons among renewal processes, Poisson processes and Kijima models are readily made. We develop a full, automated MCMC sampling scheme to fit our proposed model and illustrate our method using simulations as well as on real data.

This paper is organized as follows: Section 2 presents a description of our model

and an introduction to tailfree priors. Section 3 provides the MCMC algorithm and an approach to relax linearity in the linear predictor, and Section 4 presents simulation results. Section 5 summarizes the results for two real dataset analyses and in Section 6 we provide some concluding remarks.

3.2 MODEL DEVELOPMENT

Likelihood construction

Consider a system starting from new. Suppose the system gets repaired at times t_i , $i = 1, \dots, n$ and $0 < t_1 < t_2 < \dots < t_n < \tau$ where τ is the time when data collection stops. We assume τ is independent of the failure process. If a repair is performed without an accompanying failure, the observation of event time is right censored. Let the indicator δ_i take the value 1 if the system fails at time t_i and 0 otherwise. Further we assume a d -dimensional covariate vector for each repair, independent of the failure process, i.e. $\mathbf{w}_i = (w_{i0}, w_{i1}, \dots, w_{i,d-1})$ for the repair at time t_i . This vector may incorporate information concerning technician skills, repair type, materials used, time trend, etc. Let the counting process $\{N(t), t \geq 0\}$ record the cumulative number of failures over time and $H(t) = \{N(s) : 0 \leq s < t\}$ be the history of the process at time t . The intensity function for an event process is defined as

$$\lambda(t|H(t)) = \lim_{\Delta \rightarrow 0} \frac{P\{N(t + \Delta) - N(t) = 1 | H(t)\}}{\Delta^+}. \quad (3.1)$$

The Kijima models for the event data assume $\lambda(t|H(t)) = r\{\varepsilon(t)\}$ where $\varepsilon(t)$ is the effective age. A Kijima type I model has $\varepsilon(t_i) = \varepsilon(t_{i-1}) + [t_i - t_{i-1}]D_i$ and the type II model has $\varepsilon(t_i) = [\varepsilon(t_{i-1}) + t_i - t_{i-1}]D_i$ where $t_i - t_{i-1}$ is the time since last repair. Denote the effective age right before t_i as $\varepsilon(t_{i-})$. The i th repair at t_i reduces the effective age right before t_i by a fraction of the time since last repair, that is, $(t_i - t_{i-1})(1 - D_i)$ in the type I model and a proportion of the effective age,

i.e. $\varepsilon(t_{i-}) (1 - D_i)$ in the type II model. Note that $D_i = 0$ sets the clock back to the status right after last repair in the type I model and to a new status in the type II model.

We propose to model D_i as a function of \mathbf{w}_i through the regression coefficient $\boldsymbol{\beta}$. Let $D_i = \exp(\boldsymbol{\beta}' \mathbf{w}_i)$ or $D_i = \text{logit}(\boldsymbol{\beta}' \mathbf{w}_i)$. Let $w_{i0} = 1$ be an intercept. When the link function is the CDF of a logistic distribution, $D_i \in (0, 1)$ for all repairs, i.e. all repairs are between “good as new” and “bad as old”. If the link is exponential; then $D_i \in (0, +\infty)$. An interesting case would be $D_i > 1$ where the system actually gets worse than “bad as old” after the repair. When the baseline hazard is monotone nondecreasing or nonincreasing, $\boldsymbol{\beta}$ can be interpreted directly with respect to effectiveness of repairs, i.e. if the baseline hazard is nondecreasing and β_j is positive, one may conclude that an increase in w_j results in less effective repairs overall.

We refer to Lindqvist (2006) in deriving the likelihood of observing a system with failures at $0 < t_1 < t_2 < \dots < t_n < \tau$ ($\delta_i = 1, i = 1, \dots, n$):

$$L = \prod_{i=1}^n r(\varepsilon(t_{i-1}) + x_i) \exp \left\{ - \sum_{i=1}^n \int_0^{x_i} r(\varepsilon(t_{i-1} + u)) du - \int_0^{\tau - t_n} r(\varepsilon(t_n + u)) du, \right\} \quad (3.2)$$

where $x_i = t_i - t_{i-1}$. The likelihood is equivalent to

$$L = \prod_{i=1}^n \frac{f(\varepsilon(t_{i-1}) + x_i)}{S(\varepsilon(t_{i-1}))} \cdot \frac{S(\varepsilon(t_n) + \tau - t_n)}{S(\varepsilon(t_n))},$$

where S and f are the unique survival and density functions corresponding to r . Denote F as the cumulative distribution function for r . Now suppose we observe m identical systems. Denote t_{ij} as the event time for the i th repair of system j , \mathbf{w}_{ij} as the covariate vector and δ_{ij} as the censoring indicator. Let τ_j be the termination time for observing system j . Conditional on the collection of observables $data = \{t_{ij}, \tau_j, \mathbf{w}_{ij}, \delta_{ij}, i = 1, 2, \dots, n_j, j = 1, 2, \dots, m\}$, the likelihood of observing m independent event processes is then

$$L = \prod_{j=1}^m \prod_{i=1}^{n_j} \frac{\{f(\varepsilon(t_{i-1,j}) + x_{ij})\}^{\delta_{ij}} \{S(\varepsilon(t_{i-1,j}) + x_{ij})\}^{1-\delta_{ij}}}{S(\varepsilon(t_{i-1,j}))} \cdot \prod_{j=1}^m \frac{S(\varepsilon(t_{n_j,j}) + \tau_j - t_{n_j,j})}{S(\varepsilon(t_{n_j,j}))}, \quad (3.3)$$

where $x_{ij} = t_{ij} - t_{i-1,j}$.

Prior specifications

Tailfree process prior on F

We place a tailfree process prior on F , centered at the Weibull family. Denote $G_{\boldsymbol{\theta}}$ as the cumulative distribution function for Weibull, $G_{\boldsymbol{\theta}}(t) = 1 - \exp(-(t/\eta)^\alpha)$ for $t \geq 0$ and $\boldsymbol{\theta} = (\log(\alpha), \log(\eta))'$. Let $\Pi_j = \{B_{\epsilon_1 \dots \epsilon_j} : \epsilon_i \in \{0, 1\}\}$ be a partition of the positive reals \mathbb{R}^+ and each set in Π_j be split into two sets in Π_{j+1} , e.g. $\{B_0, B_1\}$ at the first level; $\{B_{00}, B_{01}, B_{10}, B_{11}\}$ at the second level, and so on. Following Lavine (1992), the sets are given by quantiles of the centering family; if m is the base-10 representation of the binary number $\epsilon_1 \dots \epsilon_j$, then $B_{\epsilon_1 \dots \epsilon_j}$ is the interval $(G_{\boldsymbol{\theta}}^{-1}(m/2^j), G_{\boldsymbol{\theta}}^{-1}((m+1)/2^j)]$. Let $\Pi = \{\Pi_j, j = 1, 2, \dots\}$ be the sequence of partitions. We also refer to Π as the partition tree and $j = 1, 2, \dots$ as the tree levels.

Define $F(A)$ to be the probability of any set A for distribution F ; note that $F(A)$ is a random variable. The tailfree prior for F is constructed from the sequence of partitions Π and their associated pairwise conditional probabilities $(Y_{\epsilon_1 \dots \epsilon_{j-1}0}, Y_{\epsilon_1 \dots \epsilon_{j-1}1})$, assuming $Y_{\epsilon_1 \dots \epsilon_{j-1}0} = 1 - Y_{\epsilon_1 \dots \epsilon_{j-1}1} = F\{B_{\epsilon_1 \dots \epsilon_{j-1}0} | B_{\epsilon_1 \dots \epsilon_{j-1}}\}$. Let $\mathcal{Y} = \{Y_{\epsilon_1 \dots \epsilon_{j-1}0}, j = 1, 2, \dots\}$. Further, the tailfree prior assumes the random probabilities in \mathcal{Y} are mutually independent, and the random measure F is related to the probabilities through the relation: $F\{B_{\epsilon_1 \dots \epsilon_j}\} = \prod_{i=1}^j Y_{\epsilon_1 \dots \epsilon_i}$. Let $\lambda_{\epsilon_1 \dots \epsilon_{j-1}0}$ be the logit transformation of $Y_{\epsilon_1 \dots \epsilon_{j-1}0}$. By assuming $\lambda_{\epsilon_1 \dots \epsilon_{j-1}0}$ has the normal prior $N(0, 2/[c\rho(j)])$, $Y_{\epsilon_1 \dots \epsilon_{j-1}0}$ approximately follows the beta($c\rho(j), c\rho(j)$) distribution (Jara and Hanson 2011). That is,

$$\text{logit}\{Y_{\epsilon_1 \dots \epsilon_{j-1}0}\} = \lambda_{\epsilon_1 \dots \epsilon_{j-1}0}, \quad \lambda_{\epsilon_1 \dots \epsilon_{j-1}0} \sim N\left(0, \frac{2}{c\rho(j)}\right). \quad (3.4)$$

The sequence of partitions Π forms a generator of the Borel σ -field of \mathbb{R}^+ and hence for any measurable set $A \in \mathbb{R}^+$, $F(A)$ is defined.

The infinite number of levels in the partition tree Π is usually capped off by some fixed level J , typically $4 \leq J \leq 8$, which yields partitions up to level J , say Π^J . Furthermore, on partition sets $B_{\epsilon_1 \dots \epsilon_J} \in \Pi^J$ at level J we assume F follows the base measure G_θ , i.e. for all measurable $A \subset B_{\epsilon_1 \dots \epsilon_J}$,

$$F\{A|B_{\epsilon_1 \dots \epsilon_J}\} = G_\theta(A)/G_\theta\{B_{\epsilon_1 \dots \epsilon_J}\}. \quad (3.5)$$

We use $TF^J(c, \rho(\cdot), G_\theta)$ to denote this finite tailfree prior on F with level J . For $F \sim TF^J(c, \rho(\cdot), G_\theta)$, the survival function $S(t) = 1 - F(t)$ is given by

$$S(t) = p\{s(t)\} \left\{ s(t) - 2^J G_\theta(t) \right\} + \sum_{j=s(t)+1}^{2^J} p(j), \quad (3.6)$$

where $s(t) = \lceil 2^J G_\theta(t) \rceil$, $\lceil \cdot \rceil$ is the ceiling function. Here $p(j), j = 1, \dots, 2^J$ is defined as

$$p(j+1) = F\{B_{\epsilon_1 \dots \epsilon_J}\} = \prod_{i=1}^J Y_{\epsilon_1 \dots \epsilon_i}, \quad (3.7)$$

where $\epsilon_1 \dots \epsilon_J$ is the base-2 representation of j . Formula (3.6) can be obtained from (3.5) and (3.7) and

$$F(A) = F\{B_{\epsilon_1 \dots \epsilon_J}\} G_\theta(A) / G_\theta\{B_{\epsilon_1 \dots \epsilon_J}\}$$

for $A \subset B(\epsilon_1 \dots \epsilon_J)$. By differentiating (3.6), the density with respect to F is given by

$$f(t) = 2^J p\{s(t)\} g_\theta(t), \quad (3.8)$$

where $g_\theta(\cdot)$ is the density corresponding to G_θ .

A common choice for $\rho(j)$ is j^2 . The parameter c is a precision parameter; lower values of c allow mass of F to move easily from the centering distribution G_θ . As $c \rightarrow 0^+$, $E\{F(\cdot)\}$ tends to the empirical CDF of the data (Hanson and Johnson 2002); as $c \rightarrow \infty$, all conditional probabilities $\pi(\epsilon)$ go to 0.5 and hence $F(A) \rightarrow G_\theta(A)$ a.s. for all measurable sets. We assign c a gamma prior $c \sim \Gamma(a_c, b_c)$; typically $a = 5$ or 10 and $b = 1$. Alternatively, some authors simply set c as small values, e.g. $c = 1$.

It is well known that fixing $\boldsymbol{\theta}$ results in “jumpy” densities as f defined in (3.8) has discontinuities at each partition interval endpoint. Placing a continuous prior on $\boldsymbol{\theta}$ smooths out the posterior density and hazard curves, yielding a mixture of tailfree processes for F (Jara and Hanson 2011). For the Kijima models, we suggest an empirical approach: an easily-fit special case of the model, e.g. a renewal process or a Poisson process, coupled with the underlying parametric Weibull family $G_{\boldsymbol{\theta}}$ is fitted to obtain the maximum likelihood estimate $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ and the inverse information matrix $\mathbf{V}_{\boldsymbol{\theta}}$ associated with $\boldsymbol{\mu}_{\boldsymbol{\theta}}$. A Gaussian prior $N_2(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \mathbf{V}_{\boldsymbol{\theta}})$ is placed on $\boldsymbol{\theta}$. For example, in the first data analysis in Section 5, on the reliability of valve seats, many authors have fit Poisson processes; a Poisson process could be used to center $\boldsymbol{\theta}$. Without such prior knowledge, the first failures of all systems (i.i.d samples) can be used for a parametric inference on $\boldsymbol{\theta}$.

Note that there is little difference between choosing the standard Polya tree prior and the tailfree process prior for the distribution function of the “baseline” hazard. Since we use adaptive updating of the logit-transformed conditional probabilities, as presented below in Section 3.1, it is slightly easier to fit the tailfree version rather than the Polya tree version.

Priors on $\boldsymbol{\beta}$

We recommend Zellner’s g-prior (Zellner,1983) on $\boldsymbol{\beta}$, a “reference informative prior”. g-Prior can be used to take into account the correlation among the predictor covariates and has many advantages, as commonly seen in variable selection and linear or nonlinear regressions (Bové and Held, 2011, Marin and Robert 2007, Fouskakis et al. 2009). Let $\mathbf{W}_j = (\mathbf{w}'_{j1}, \dots, \mathbf{w}'_{jn_j})$ and $\mathbf{W}_{m^* \times d} = (\mathbf{W}'_1, \dots, \mathbf{W}'_m)'$ where $m^* = \sum_{j=1}^m n_j$. g-prior for $\boldsymbol{\beta}$ is then

$$\pi(\boldsymbol{\beta}) \sim N_d\left(0, gm^*(\mathbf{W}'\mathbf{W})^{-1}\right).$$

To avoid choosing g , one can assign g^{-1} a gamma prior $\Gamma(a_g, b_g)$. When $a_g = b_g = 1/2$, the prior on $\boldsymbol{\beta}$ is a multivariate Cauchy distribution (Zellner and Siow 1980).

In our simulations, we use a g-prior for the logistic link and obtained excellent performance. We found that the g-prior improves the overall mixing of MCMC chain for both links, but an uninformative prior ($\pi(\boldsymbol{\beta}) \propto 1$) for the logistic link leads to extremely poor MCMC mixing in many data sets. For the exponential link function, We found that uninformative prior shows pretty good and stable performance.

3.3 POSTERIOR INFERENCES

MCMC computing

MCMC is used to obtain posterior inferences. The likelihood L is defined in (3.3) and the prior on $\boldsymbol{\beta}$ is discussed in Section 2.2.2. Recall that we propose a mixture of tailfree processes prior on F with partitions capped off by J . The prior on $\boldsymbol{\theta}$ is defined at the end of Section 2.2.1; the prior on c is $\Gamma(a_c, b_c)$, and the prior on g^{-1} is $\Gamma(a_g, b_g)$. Let $E_0 = \{\boldsymbol{\epsilon} = \epsilon_1 \cdots \epsilon_{j-1} 0, j = 1, \dots, J\}$. Each $\lambda_{\boldsymbol{\epsilon}}$ is assigned a normal prior as detailed in (3.4). The posterior is then proportional to

$$\pi(\boldsymbol{\beta}, \lambda_{\boldsymbol{\epsilon}}, c, g, \boldsymbol{\theta} | data) \propto L \cdot \pi(\boldsymbol{\beta}) \Gamma(c | a_c, b_c) \Gamma(g^{-1} | a_g, b_g) \pi(\boldsymbol{\theta}) \prod_{\boldsymbol{\epsilon} \in E_0} N\left(\lambda_{\boldsymbol{\epsilon}} | 0, \frac{2}{c j^2}\right). \quad (3.9)$$

Parameters $\{\boldsymbol{\beta}, \lambda(\boldsymbol{\epsilon}), \boldsymbol{\epsilon} \in E_0, \boldsymbol{\theta}\}$ are updated using random-walk Metropolis-Hastings updates (Tierney, 1994). We build two blocks to update these parameters. Let \mathbf{b}_1 be a vector of all $\{\lambda_{\boldsymbol{\epsilon}}, \boldsymbol{\epsilon} \in E_0\}$ with dimension $2^J - 1$ and $\mathbf{b}_2 = (\boldsymbol{\beta}, \boldsymbol{\theta})$. Gaussian random-walk proposals are used for the two blocks

$$\mathbf{b}'_1 \sim N(\mathbf{b}_1^*, \mathbf{V}_1) \text{ and } \mathbf{b}'_2 \sim N(\mathbf{b}_2^*, \mathbf{V}_2),$$

where \mathbf{b}_1^* and \mathbf{b}_2^* are the latest accepted values for \mathbf{b}_1 and \mathbf{b}_2 . We have found automatic tuning of \mathbf{V}_1 and \mathbf{V}_2 to work very well in practice (Haario, Saksman, and

Tamminen, 2005) leading to proposal acceptance rates in the 20% to 50% range as typically desired. Specifically, let the sequence $\mathbf{b}_1^{(1)}, \mathbf{b}_1^{(2)}, \dots$ be the states of the Markov chain for \mathbf{b}_1 . When deciding the t -th state \mathbf{b}_1 , we sample $\mathbf{b}_1^* \sim N(\mathbf{b}_1^{(t-1)}, \mathbf{V}_1^{(t)})$ with

$$\mathbf{V}_1^{(t)} = \begin{cases} \mathbf{V}_1^{(0)}, & t < t_0 \\ s\text{Var}\{\mathbf{b}_1^{(1)}, \dots, \mathbf{b}_1^{(t-1)}\} + s_0\mathbf{I}_p, & t > t_0 \end{cases}$$

where p is the dimension of \mathbf{b}_1 , s is recommended to be $2.4^2/p$, s_0 is a small constant, $\mathbf{V}_1^{(0)}$ is the initial covariance of the proposal distribution and \mathbf{I}_p is an identity matrix. A similar automatic tuning procedure applies to \mathbf{b}_2 . The parameter c is updated through the full conditional distribution

$$p(c|\boldsymbol{\lambda}) \sim \Gamma\left\{(a_c + 2^{J-1} - 1/2), b_c + \sum_{\epsilon_1 \epsilon_2 \dots \epsilon_j \in E_0} \boldsymbol{\lambda}_{\epsilon_1 \epsilon_2 \dots \epsilon_j}^2 j^2/4}\right\}.$$

The full conditional distribution for g^{-1} given the remaining parameters is $\Gamma(a_g + 1, b_g + \boldsymbol{\beta}'\mathbf{W}'\mathbf{W}\boldsymbol{\beta}/2m^* + b_g)$. FORTRAN 90 codes for fitting the models in this paper are available from the first author, upon request.

Model comparison

We compare models using log pseudo-marginal likelihood (LPML) (Geisser and Eddy, 1979), a measure of a model's predictive ability and the deviance information criterion (DIC) (Spiegelhalter et al. 2002), a model selection criterion related to AIC but for use with Bayesian models. Both are easy to compute based on the MCMC output.

Let $\Theta = (\boldsymbol{\lambda}, \boldsymbol{\theta}, \boldsymbol{\beta})$ and $t_{n_j+1,j} = \tau_j$. By definition,

$$\text{LPML} = \sum_{j=1}^m \sum_{i=1}^{n_j+1} \log\{p(t_{ij}|\mathbf{t}_{-ij})\},$$

where $p(t_{ij}|\mathbf{t}_{-ij})$ is the predictive density ($\delta_{ij} = 1$) or survival probability ($\delta_{ij} = 0$) for t_{ij} based on the remaining data, $p(\cdot|\mathbf{t}_{-ij})$, evaluated at t_{ij} . This is called the ij -th conditional predictive ordinate (CPO) statistic, and measures how well t_{ij} is predicted from the remaining \mathbf{t}_{-ij} through the model. For system j that has events

at $0 < t_{1j} < t_{2j} < \dots < t_{n_j+1,j}$, we compute the predictive density or survival at t_{ij} based on failure and maintenance history for this system during time periods $(0, t_{i-1,j}]$ and $[t_{i+1,j}, \tau_j]$, plus partial information during $(t_{i-1,j}, t_{i+1,j})$ that a certain repair was performed at t_{ij} , plus the information from other systems. That is, to predict $p(t_{ij}|\mathbf{t}_{-ij})$, a repair is still assumed to be done at t_{ij} . The LPML simply aggregates the log of these. For this type of prediction, we are able to share the same form of computing as recommended by Gelfand and Dey (1994). As stated in Section 3.1, the likelihood contribution of failure at t_{ij} depends on repair times and their effectivenesses before t_{ij} for system i , i.e. $t_{ik}, D_{ik}, k = 1, \dots, j-1$. Conditional on Θ , the joint likelihood is $\prod_{j=1}^m \prod_{i=1}^{n_j+1} p(t_{ij}|\mathbf{t}_{1:i-1,j}, \Theta)$.

Following Gelfand and Dey (1994), we have

$$\begin{aligned} p(t_{ij}|\mathbf{t}_{-ij}) &= \int p(t_{ij}|\mathbf{t}_{-ij}, \Theta)\pi(\Theta|\mathbf{t}_{-ij})d\Theta \\ &= \int p(t_{ij}|\mathbf{t}_{1:i-1,j}, \Theta) \cdot \frac{\prod_{\{l,k\} \in A} p(t_{kl}|\mathbf{t}_{1:k-1,l}, \Theta)\pi(\Theta)}{\int \prod_{\{l,k\} \in A} p(t_{kl}|\mathbf{t}_{1:k-1,l}, \Theta)\pi(\Theta)d\Theta} d\Theta \\ &= \left\{ \int \frac{1}{p(t_{ij}|\mathbf{t}_{1:i-1,j}, \Theta)} \pi(\Theta|\mathbf{t})d\Theta \right\}^{-1}, \end{aligned}$$

where $A = \{l, k : l \neq j \text{ or } k \neq i\}$. The LPML is then estimated from the MCMC iterates by

$$\text{LPML} = - \sum_{j=1}^m \sum_{i=1}^{n_j+1} \log \left\{ \frac{1}{s} \sum_{k=1}^s \frac{1}{p(t_{ij}|\mathbf{t}_{1:i-1,j}, \Theta^{(k)})} \right\}, \quad (3.10)$$

where $\Theta^{(k)} = \{\boldsymbol{\lambda}^{(k)}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\beta}^{(k)}, k = 1, 2, \dots, s\}$ are iterates from MCMC outputs of all the parameters.

By definition,

$$\text{DIC} = 2E[D(\Theta|y)] - D(\widehat{\Theta}),$$

where $D(\Theta) = -2\log[L(\Theta)] + C$, $L(\Theta)$ is the likelihood and C is a constant canceled in model comparison. Conditional expectation $E[D(\Theta|y)]$ is typically estimated by averages of $D(\Theta)$ over posterior samples of Θ ; $\widehat{\Theta}$ in $D(\widehat{\Theta})$ is commonly chosen as the posterior mean of Θ .

Relaxing the linearity assumption

In this section, we generalize the linear predictor to a flexible additive structure. For simplicity, consider three covariates in the regression: an intercept, a discrete covariate w_{ij1} , and a continuous covariate w_{ij2} . A generalized additive model with the exponential link assumes

$$\log(D_{ij}) = \beta_0 + \beta_1 w_{ij1} + h(w_{ij2}), \quad i = 1, \dots, n_i, j = 1, \dots, m.$$

We approximate the unknown function $h(x)$ using B-splines, i.e.

$$h(x) = \sum_{l=1}^M b_l B_l(x),$$

where $\{B_l(\cdot)\}$ are quadratic B-spline basis functions, defined in De Boor (2001), with support (a, b) , the observed range of w_{ij2} . Since the space spanned by these functions includes the constant term, we let one B-spline coefficient be zero (Gray 1992) – we choose $b_M = 0$ – under which $h(x)$ equals the constant zero if and only if all the B-spline coefficients are equal to zero. In the following, define $\mathbf{b} = (b_1, \dots, b_{M-1})'$. Note then $h(x) = \sum_{l=1}^{M-1} b_l B_l(x)$. Define $\boldsymbol{\beta} = (\beta_0, b_1, \dots, b_{M-1})'$. The g-prior on $\boldsymbol{\beta}$ is $\boldsymbol{\beta} \sim N_M(\mathbf{0}, ng(\mathbf{X}'\mathbf{X})^{-1})$ where \mathbf{X} is the design matrix. The above extension can be fit using the algorithm developed in Section 3.

For equally-spaced knots, $\sum_{l=1}^{M-1} b_l B_l(x) = \beta_1 x$ for some β_1 when $b_{l-1} + b_{l+1} - 2b_l = 0$ for $l = 2, \dots, M-2$. Define $\boldsymbol{\Delta} = (b_1 + b_3 - 2b_2, b_2 + b_4 - 2b_3, \dots, b_{M-3} + b_{M-1} - 2b_{M-2}) = \boldsymbol{\Phi}\mathbf{b}$ where $\boldsymbol{\Phi}$ is a $(M-3) \times (M-1)$ matrix. Suppose MCMC iterates for \mathbf{b} are $\mathbf{b}^{(k)}, k = 1, \dots, s$. To test whether $h(x)$ is linear in x is equivalent to test the point null $H_0 : \boldsymbol{\Phi}\mathbf{b} = \mathbf{0}$. Bayes factors against the null hypothesis can be computed using the Savage-Dickey ratio (Verdinelli and Wasserman 1995),

$$BF \approx \frac{N_{M-3}(\mathbf{0} | \mathbf{0}, ng\boldsymbol{\Phi}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\Phi}')}{N_{M-3}(\mathbf{0} | \mathbf{m}, \mathbf{V})},$$

where $\mathbf{m} = s^{-1} \sum_{k=1}^s \boldsymbol{\Phi}\mathbf{b}^{(k)}$ and $\mathbf{V} = s^{-1} \sum_{k=1}^s (\boldsymbol{\Phi}\mathbf{b}^{(k)} - \mathbf{m})(\boldsymbol{\Phi}\mathbf{b}^{(k)} - \mathbf{m})'$. A larger BF value indicates stronger evidence against the null hypothesis.

3.4 SIMULATIONS

We perform simulations to examine the proposed models and the Bayesian non-parametric method. Suppose m systems are included in each simulated sample and each system is maintained up to its 5th failure, yielding a total number of events $m^* = 5m$. The associated event (failure) times are recorded as $t_{ij}, i = 1, 2, \dots, 5, j = 1, 2, \dots, m$. At each event time, a type of repair is performed with effectiveness according to the Kijima type I or type II model. The degree of effectiveness D_{ij} is $\text{logit}(\boldsymbol{\beta}'\mathbf{w}_{ij})$ or $\exp(\boldsymbol{\beta}'\mathbf{w}_{ij})$ where \mathbf{w}_{ij} includes $w_{ij0} = 1, w_{ij1}$ a Bernoulli (0.5) and w_{ij2} simulated so that D_{ij} follows uniform (0, 1). The true baseline distribution is $0.5\text{Weibull}(2, 2) + 0.5\text{Weibull}(2, 4)$ for simulations in Table 3.1 and 3.3 and has a corresponding hazard $\exp(t^2/3 + t/3)$ for simulations in Table 3.2. Coefficients are set to $\boldsymbol{\beta} = (-1, 1, 1)$ or $(1, -1, 1)$ and the sample size is $m^* = 300$ or 500. For each setup, 300 datasets are simulated and fitted with the following model and prior specifications. The baseline distribution F is given the tailfree prior with $J = 5, c \sim \Gamma(5, 1)$, and $\boldsymbol{\theta} \sim N(\boldsymbol{\mu}_\theta, \mathbf{V}_\theta)$ where $\boldsymbol{\mu}_\theta$ is the maximum likelihood estimate of $\boldsymbol{\theta}$ based on a parametric Weibull fit of the first failures of all systems and \mathbf{V}_θ the inverse information matrix associated with $\boldsymbol{\mu}_\theta$. Optimization routines in R or SAS give $\boldsymbol{\theta}$ and \mathbf{V}_θ . For the logistic link, regression parameter $\boldsymbol{\beta}$ is given the g-prior $N_3(0, gn(\mathbf{W}'\mathbf{W})^{-1})$ with $g^{-1} \sim \text{Exp}(1)$. For the exponential link, the flat prior $\pi(\boldsymbol{\beta}) \propto 1$ is used. Following algorithms in Section 3.1, we run 30000 iterations for each MCMC chain and thin the posterior samples by taking every fifth of them after a burn-in of 10000 iterates. Each chain takes a few minutes with a 3.00 GHz processor.

Simulation results are presented in Tables 1–3 including the average of the posterior means over 300 datasets, the sample standard deviation SSD of the posterior means, the average of the estimated standard deviations ESE and 95% the coverage probability CP. Based on the simulation results, the true parameters $\boldsymbol{\beta}$ are estimated

with little bias. As the sample size increases, both SSD and SSE decrease. We also get coverage probabilities close to the nominal level 0.95. For one simulation setup posterior means for the baseline density and survival functions (gray lines) are plotted in Figure 3.1, overlaid with the true density or survival functions in black lines.

Table 3.1: Summary of simulation studies: $f=0.5\text{Weibull}(2, 2)+0.5\text{Weibull}(2, 4)$; link function is logistic.

True	Point	n=300			n = 500			
		SSD	ESE	95% CP	Point	SSD	ESE	95% CP
Type I								
$\beta_0 = -1$	-0.92	1.10	1.17	0.95	-1.04	0.75	0.82	0.95
$\beta_1 = 1$	1.05	1.44	1.69	0.95	1.04	1.03	1.19	0.96
$\beta_2 = 1$	0.97	0.62	0.73	0.95	0.97	0.43	0.52	0.95
$\beta_0 = 1$	1.06	1.24	1.56	0.95	1.02	1.07	1.22	0.93
$\beta_1 = -1$	-0.95	1.30	1.72	0.96	-0.96	1.16	1.28	0.94
$\beta_2 = -1$	-0.94	0.55	0.74	0.98	-1.01	0.51	0.60	0.95
Type II								
$\beta_0 = -1$	-0.98	0.81	0.89	0.94	-1.02	0.60	0.60	0.92
$\beta_1 = 1$	0.97	1.19	1.26	0.95	1.00	0.89	0.87	0.93
$\beta_2 = 1$	1.00	0.48	0.56	0.95	1.01	0.39	0.40	0.95
$\beta_0 = 1$	1.02	1.01	1.14	0.95	0.96	0.85	0.80	0.93
$\beta_1 = -1$	-1.08	1.11	1.27	0.96	-0.96	0.89	0.85	0.93
$\beta_2 = -1$	-1.01	0.50	0.58	0.94	-0.99	0.41	0.40	0.94

We also perform simulations to examine the additive model described in Section 3.3. Assume $D_{ij} = \exp(\beta_0 + \beta_1 w_{ij1} - w_{ij2}^2)$ or $\text{logit}(\beta_0 + \beta_1 w_{ij1} - w_{ij2}^2)$ where w_{ij1} is sampled from Bernoulli (0.5) and w_{ij2} from uniform $(-1, 1)$. Data are simulated from the Kijima type I model with baseline hazard $r(t)=\exp(t^2/3 + t/3)$. We consider a sample size $m^* = 1000$ and each setup has 300 replications. We take $M = 6$ equally spaced quadratic B-splines to model the effect of w_{ij2} and one B-spline coefficient is set to be zero. Table 3.4 summarizes estimates for coefficients β_0, β_1 and Figure 3.2 plots the point-wise mean, 2.5% and 97.5% quantiles of the estimates for the true function $h(w) = -w^2$.

Table 3.2: Summary of simulation studies: $r(t)=\exp(t^2/3 + t/3)$; link function is logistic.

True	Point	n=300			n = 500			
		SSD	ESE	95% CP	Point	SSD	ESE	95% CP
Type I								
$\beta_0 = -1$	-1.00	0.37	0.37	0.96	-1.01	0.29	0.29	0.94
$\beta_1 = 1$	1.02	0.50	0.53	0.96	1.04	0.39	0.40	0.94
$\beta_2 = 1$	1.03	0.27	0.28	0.96	1.02	0.20	0.20	0.95
$\beta_0 = 1$	1.10	0.51	0.54	0.94	1.05	0.38	0.39	0.96
$\beta_1 = -1$	-1.04	0.53	0.55	0.93	-1.02	0.38	0.40	0.95
$\beta_2 = -1$	-1.05	0.26	0.29	0.95	-1.03	0.19	0.21	0.97
Type II								
$\beta_0 = -1$	-1.00	0.36	0.34	0.93	-1.02	0.29	0.27	0.93
$\beta_1 = 1$	0.99	0.45	0.43	0.94	1.02	0.35	0.33	0.93
$\beta_2 = 1$	1.01	0.22	0.22	0.95	1.02	0.18	0.17	0.94
$\beta_0 = 1$	1.04	0.36	0.37	0.93	1.01	0.27	0.27	0.95
$\beta_1 = -1$	-1.04	0.43	0.44	0.94	-1.03	0.31	0.33	0.96
$\beta_2 = -1$	-1.03	0.21	0.22	0.96	-1.01	0.17	0.17	0.94

Table 3.3: Summary of simulation studies: $f=0.5\text{Weibull}(2, 2)+0.5\text{Weibull}(2, 4)$; link function is exponential.

True	Point	n=300			n = 500			
		SSD	ESE	95% CP	Point	SSD	ESE	95% CP
Type I								
$\beta_0 = -1$	-1.09	0.51	0.49	0.92	-1.01	0.37	0.34	0.92
$\beta_1 = 1$	1.02	0.66	0.75	0.96	1.01	0.57	0.52	0.92
$\beta_2 = 1$	1.09	0.45	0.52	0.96	1.08	0.41	0.38	0.92
$\beta_0 = 1$	0.94	0.81	0.86	0.95	0.92	0.58	0.62	0.94
$\beta_1 = -1$	-1.01	0.71	0.73	0.94	-0.99	0.47	0.51	0.96
$\beta_2 = -1$	-1.07	0.45	0.50	0.96	-1.02	0.34	0.36	0.94
Type II								
$\beta_0 = -1$	-1.10	0.35	0.35	0.93	-1.01	0.28	0.27	0.91
$\beta_1 = 1$	0.94	0.51	0.56	0.96	0.98	0.43	0.43	0.92
$\beta_2 = 1$	1.04	0.34	0.40	0.95	1.04	0.30	0.31	0.94
$\beta_0 = 1$	0.91	0.62	0.69	0.95	1.02	0.55	0.52	0.93
$\beta_1 = -1$	-1.00	0.55	0.62	0.97	-1.02	0.50	0.46	0.92
$\beta_2 = -1$	-1.09	0.40	0.43	0.95	-1.07	0.35	0.33	0.93

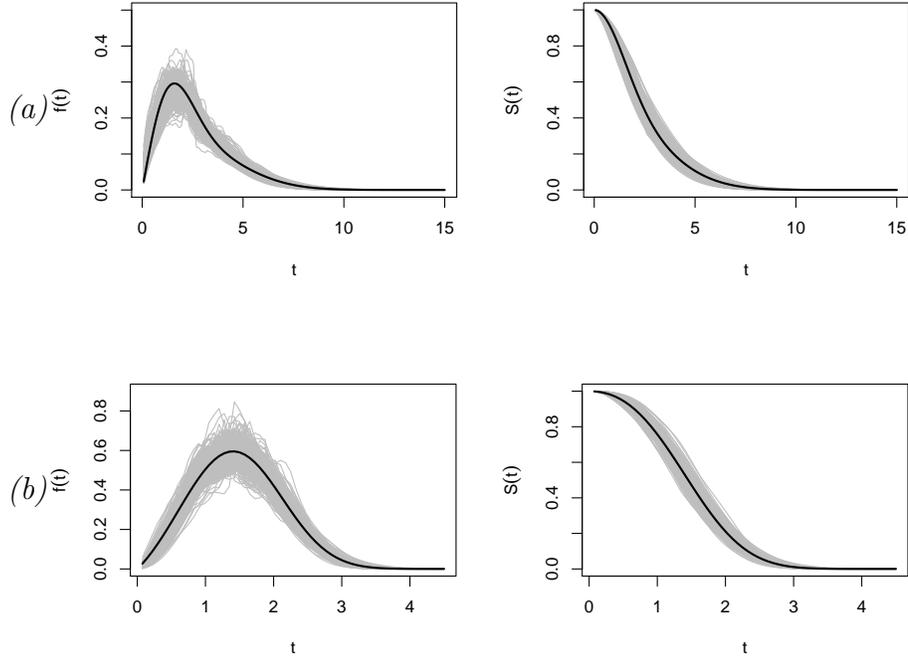


Figure 3.1: Density and survival estimates for the simulated data sets with (a) $0.5\text{Weibull}(2, 2) + 0.5\text{Weibull}(2, 4)$ and (b) $h_0(t)=\exp(t^2/3 + t/3)$ based on Kijima type I model; the dark lines are the true density or survival functions and the gray lines are the point-wise posterior means.

Table 3.4: Summary of simulation studies: $r(t)=\exp(t^2/3 + t/3)$; $h(w_{ij2}) = \sum_{l=1}^5 b_l B_{l3}(w_{ij2})$.

True	Logistic				Exponential			
	Point	SSD	ESE	95% CP	Point	SSD	ESE	95% CP
$\beta_0 = -1$	-1.09	0.38	0.37	0.93	-1.06	0.23	0.24	0.96
$\beta_1 = 1$	1.07	0.23	0.21	0.91	-1.07	0.17	0.17	0.92

3.5 DATA ANALYSIS

We first consider the dataset analyzed in Lawless and Nadeau (1995) that gives the times of replacing valve seats on 41 diesel engines in a service fleet. A few successive repairs recorded on the same day are deleted. We assume the end of history-time is independent of the event process, as concluded in Lawless and Nadeau (1995). On the left panel of Figure 3.3, a nonparametric estimate for the mean cumulative function

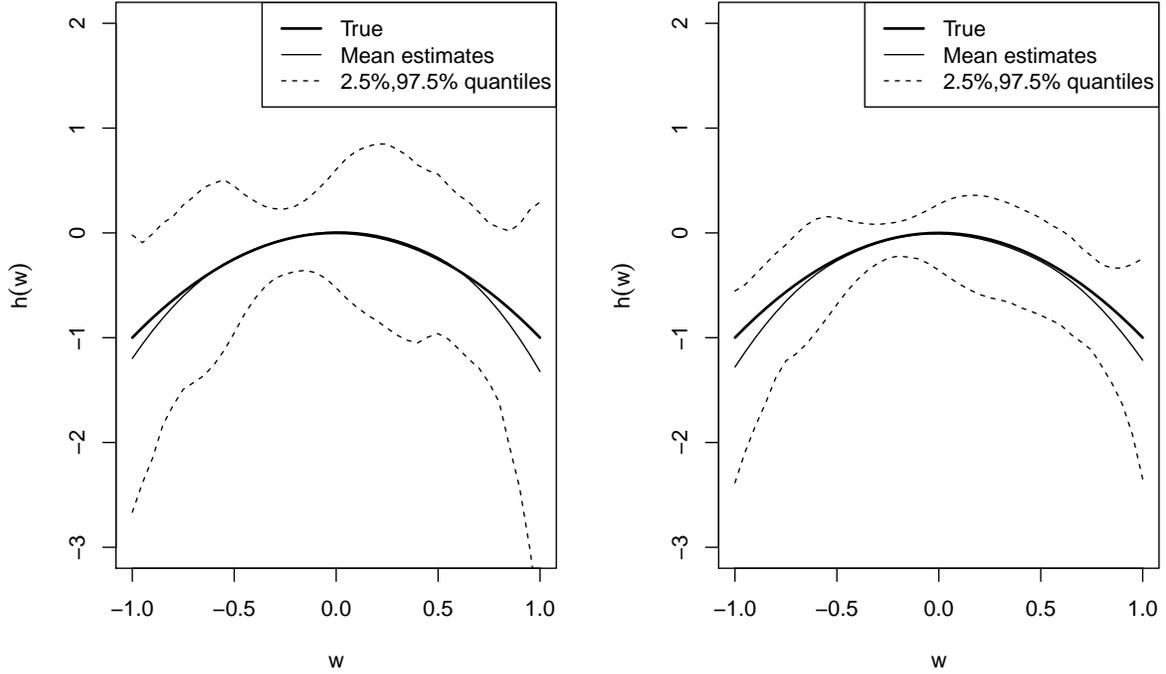


Figure 3.2: Mean (gray-solid lines), 2.5% and 97.5% quantiles (gray-dashed lines) of the estimates for $h(w)$ based on the simulated datasets from the type I model with logistic (left) and exponential (right) links. The black solid lines are the true function $h(w) = -w^2$.

(MCF) (Nelson 1995) is plotted, where each ‘+’ is a point estimate of MCF at its corresponding time and ‘×’s are its associated 95% confidence interval. Note that MCF is defined as $E[N(t)]$. When the event process is NHPP, MCF equals the mean function $\int_0^t r(s)ds$. The dataset has been fitted many times in the literature assuming NHPP. In this work, Kijima type I, Kijima type II, and NHPP models ($D_{ij} = 1$) are fitted for the data using the proposed method. The baseline distribution F is given a tailfree prior: $J = 4$ or 5 ; $c \sim \Gamma(5, 1)$ or $\Gamma(10, 1)$; $\boldsymbol{\theta} \sim N_2(\boldsymbol{\mu}_\theta, \mathbf{V}_\theta)$ where $\boldsymbol{\mu}_\theta$ and \mathbf{V}_θ are obtained from a fit of the Poisson process assuming Weibull for the baseline distribution. For the Kijima models, we choose exponential link for the age reduction factor, i.e. $D_{ij} = \exp(\beta_0)$. Two sets of priors are considered for β_0 : $N(0, 2^2)$ and $N(0, 3^2)$.

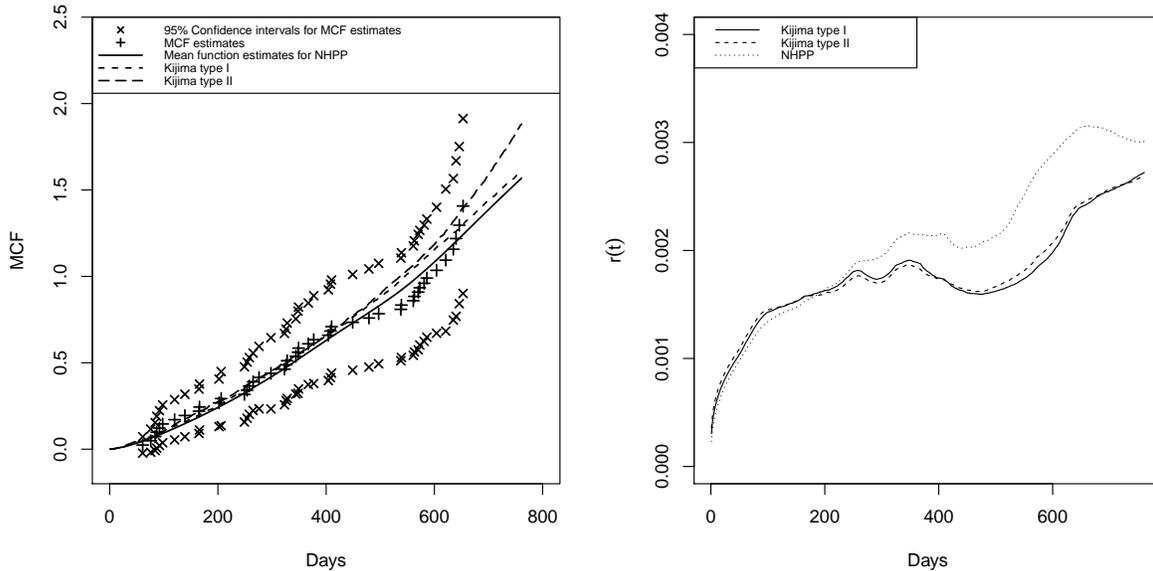


Figure 3.3: Plots for valve seats maintenance data; left panel is the MCF plot for data where ‘+’ is the empirical point estimate and ‘×’ is its 95% confidence interval, overlaid with estimates of the mean function for NHPP (solid line), Kijima type I (short-dashed line) and Kijima type II (long-dashed line); right panel is the estimated baseline hazard function for Kijima type I (solid line), Kijima type II (dashed line) and NHPP (dotted line).

Based on the results in Table 3.5, both Kijima type I and type II models show high probabilities for D_{ij} being greater than 1, i.e. posterior $P(\beta_0 > 0)$. There is little difference in both estimation of β_0 and goodness-of-fit measures (LPML, DIC) when increasing the tailfree level J . The prior favoring lower c shows slightly better values for LPML and DIC and has some effects on estimation of β_0 due to a less weight of the centering Weibull family. The prior on β_0 with larger variance results in wider 95% credible intervals but the point estimate of β_0 remains stable. Under $J = 5, c \sim \Gamma(5, 1)$ and $\pi(\beta_0) \sim N(0, 2^2)$, the estimated baseline hazards $\widehat{r(t)}$ (right panel of Figure 3.3) are nondecreasing in general and with slight decreases around 500 days. Interpretation of the age reduction is then related to effectiveness of the repairs, i.e. the repairs have high probability of being worse than “bad as old”, explaining to some extent the rapid increase of failures around 600 days. To compare data fits,

estimates of $E(N(t))$ for NHPP, Kijima type I and type II models are also plotted on the left panel in Figure 3.3 where estimates for Kijima types I and II are based on simulated failure times using the posterior means of β_0 and the baseline survival function (Krivtsov 2000, Veber et al. 2008). The plot shows that MCF estimates based on all three of NHPP, Kijima type I, and type II interpolate the nonparametric MCF estimates well, except for slight differences in the tail. Under $J = 5$, $c \sim \Gamma(5, 1)$, the nonparametric fit of the NHPP model has LPML and DIC -336.0 and 666.6 respectively. Little difference can be found in the goodness-of-fit measures, compared to those in Table 3.5. Nevertheless, our models provide information on the age reduction factor which is helpful for understanding the repairs and future modeling. We also fit parametric Kijima type I and II models assuming Weibull baseline distribution and obtain LPML as -334.6 and -334.7 and DIC as 669.4 and 669.6 respectively. Therefore, the simpler Weibull family would be an adequate choice for the baseline distribution.

Table 3.5: Summaries of β_0 for Kijima type I and type II models for the valve seats maintenance data.

Kijima Model	J	$\pi(c)$	$\pi(\beta_0)$	$\widehat{\beta}_0$ (95% CI.)	$P(\beta_0 > 0)$	LPML	DIC
Type I	4	$\Gamma(5, 1)$	$N(0, 3^2)$	1.11(-1.73,3.06)	0.92	-334.0	665.4
	5	$\Gamma(5, 1)$	$N(0, 3^2)$	1.07(-1.79,2.98)	0.93	-334.9	665.2
	4	$\Gamma(5, 1)$	$N(0, 2^2)$	1.08(-1.47,2.76)	0.93	-334.0	665.0
	5	$\Gamma(5, 1)$	$N(0, 2^2)$	1.04(-1.48,2.61)	0.93	-334.1	664.0
	5	$\Gamma(10, 1)$	$N(0, 2^2)$	0.96(-1.78,2.85)	0.88	-334.9	668.0
Type II	4	$\Gamma(5, 1)$	$N(0, 3^2)$	0.80(-2.69,2.89)	0.89	-334.7	667.2
	5	$\Gamma(5, 1)$	$N(0, 3^2)$	0.79(-2.46,2.66)	0.89	-334.5	666.2
	4	$\Gamma(5, 1)$	$N(0, 2^2)$	0.78(-1.36,2.23)	0.90	-334.6	666.3
	5	$\Gamma(5, 1)$	$N(0, 2^2)$	0.84(-1.43,2.39)	0.91	-334.5	665.7
	5	$\Gamma(10, 1)$	$N(0, 2^2)$	0.64(-1.77,2.27)	0.85	-335.1	669.5

The second dataset includes failures and repairs of 12 syringe-driver pumps (Baker 1991, Singh 2011). Most systems are maintained up until 106 months and we assume that this censoring time is independent of the failure processes. The pumps receive

preventive maintenances (*pm* mode) and corrective maintenances (*cm* mode) with 48 *pms* and 94 *cms* in total. The cost for each *cm* is also available and we consider it as a covariate interacting with corrective maintenance mode only. Denote \mathbf{w} as the covariate vector including: $w_0 = 1$ if the maintenance is *pm* and 0 otherwise, $w_1 = 1$ if the maintenance is *cm* and 0 otherwise and w_2 is the cost of the *cm* repair. Kijima type I and type II models are fitted to the data. The baseline distribution F is given a tailfree prior: $J = 5$; $c \sim \Gamma(10, 1)$ or $\Gamma(5, 1)$; $\boldsymbol{\theta} \sim N_2(\boldsymbol{\mu}_\theta, \mathbf{V}_\theta)$ where $\boldsymbol{\mu}_\theta$ and \mathbf{V}_θ are obtained from a parametric fit of the first events of the 12 pumps. The coefficient vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$ is considered with g-prior and other Gaussian priors. We assume the exponential link for both models.

Table 3.6 shows the point estimates (posterior means) and 95% credible intervals for $\boldsymbol{\beta}$. For the Kijima type I model, the g-prior results in narrower credible intervals for the intercept β_0 and shrinks β_1 toward zero. For Kijima type II model, there is little difference in the estimates and credible intervals by using different priors on $\boldsymbol{\beta}$. The effective age reduction factor due to preventive maintenance in type II models is significantly less than 1, by exponentiating the estimate of the intercept β_0 , indicating that preventive maintenances are better than “bad as old” repairs. Baker (1991) also observes that the preventive maintenances are very effective in maintaining the systems. Type I models does not show strong evidence for β_0 less than zero. Across all fitting specifications, the corrective repairs perform significantly worse than the preventive repairs and the cost of the corrective repairs shows no significant effect. Table 3.7 presents estimates of Goodness-of-fit measures LPML and DIC. Since larger LPML or lower DIC indicates a better fit of the data, the results show that type I models fit the data slightly better than type II models; $\Gamma(5, 1)$ yields a better fit than $\Gamma(10, 1)$.

We also fit the Kijima type I and II models with the underlying Weibull baseline ($c \rightarrow \infty$) and $\boldsymbol{\beta} \sim N(\mathbf{0}, 3^2\mathbf{I}_3)$. The type I model has LPML -314.6 and DIC 628.7.

The type II model has LPML -319.1 and DIC 636.6. The nonparametric Bayes method gives slightly better fits to the data than the parametric Weibull models. The differences are not significant, suggesting adequacy of the Weibull assumption for the data. Figure 3.4 also shows that the parametric estimates for the baseline density and hazard functions (smooth lines) stay in the credible intervals (dashed lines) of the nonparametric estimates.

Table 3.6: Summary of the coefficients for Kijima type I and type II models for Syringe-driver maintenance data; $J = 5$, link function is exponential, estimates are posterior means, and 95% CIs are credible intervals.

Coefficient	$\pi(c)$	Type I	Type II
		$\hat{\boldsymbol{\beta}}$ (95% CI)	$\hat{\boldsymbol{\beta}}$ (95% CI)
	$\Gamma(10, 1)$	$\boldsymbol{\beta} \sim N(\mathbf{0}, gn(\mathbf{W}'\mathbf{W})^{-1});$	$g \sim \text{Exp}(1)$
$\beta_0(pm)$		-0.13(-1.96,1.13)	-0.36 (-0.65,-0.06)
$\beta_1(cm)$		1.29 (0.46,2.31)	0.05 (-0.04,0.19)
$\beta_2(cost)$		-0.20 (-0.78,0.25)	-0.02 (-0.16,0.11)
	$\Gamma(10, 1)$	$\boldsymbol{\beta} \sim N(\mathbf{0}, 3^2\mathbf{I}_3)$	
$\beta_0(pm)$		0.04 (-3.47,2.18)	-0.36 (-0.66,-0.07)
$\beta_1(cm)$		2.19 (0.67,4.44)	0.05 (-0.05,0.19)
$\beta_2(cost)$		-0.28 (-1.03,0.31)	-0.02 (-0.14,0.11)
	$\Gamma(5, 1)$	$\boldsymbol{\beta} \sim N(\mathbf{0}, 3^2\mathbf{I}_3)$	
$\beta_0(pm)$		-0.06 (-4.4,2.55)	-0.34 (-0.63,-0.06)
$\beta_1(cm)$		2.25 (0.65,5.17)	0.04 (-0.05,0.16)
$\beta_2(cost)$		-0.29 (-1.16,0.29)	0.01 (-0.12,0.13)

Table 3.7: Goodness-of-fit measures for Kijima type I and type II model for Syringe-driver maintenance data; $J = 5$, and link function is exponential.

	$\pi(c)$	$\pi(\boldsymbol{\beta})$	Type I	Type II
LPML	$\Gamma(10, 1)$	g-prior	-313.2	-319.2
LPML	$\Gamma(10, 1)$	$N(\mathbf{0}, 3^2\mathbf{I}_3)$	-312.2	-319.1
LPML	$\Gamma(5, 1)$	$N(\mathbf{0}, 3^2\mathbf{I}_3)$	-311.4	-316.8
DIC	$\Gamma(10, 1)$	g-prior	624.2	631.3
DIC	$\Gamma(10, 1)$	$N(\mathbf{0}, 3^2\mathbf{I}_3)$	621.8	631.1
DIC	$\Gamma(5, 1)$	$N(\mathbf{0}, 3^2\mathbf{I}_3)$	616.4	627.2

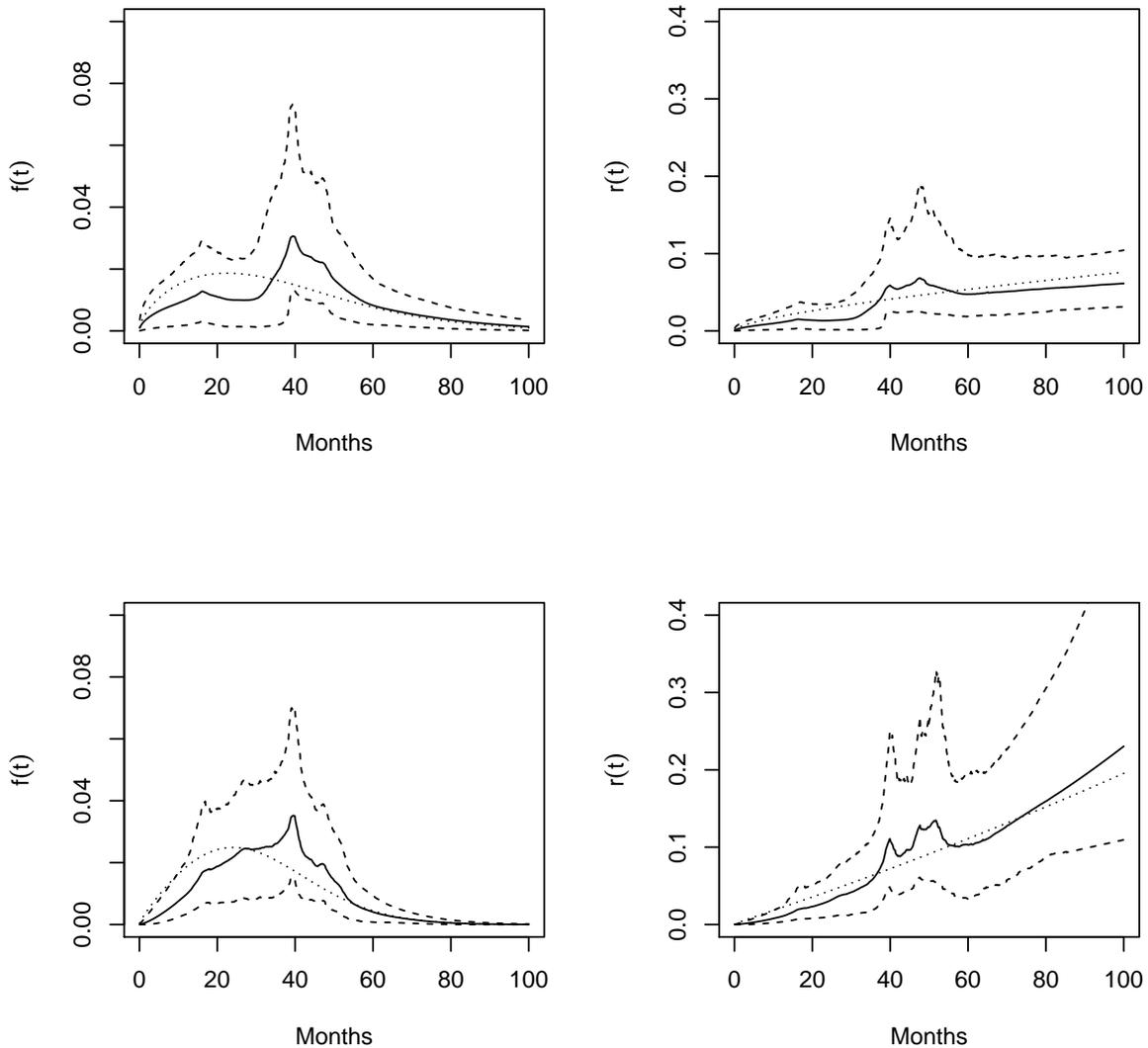


Figure 3.4: Plots for Syringe-driver maintenance data for Kijima type I (left) and type II (right) model; solid lines are baseline density (hazard) estimates and dashed lines are 95% credible intervals. Smooth estimates (dotted lines) are fitted from the parametric Weibull fit.

Finally, we fit the Kijima type I model with $h(w_2)$ modeled by a B-spline as outlined in Section 3.3. Let $M = 6$ and set one B-spline coefficient as zero for model identifiability. Now $\boldsymbol{\beta} = (\beta_0, \beta_1, b_1, \dots, b_5)$ and let $\boldsymbol{\beta} \sim N(0, ng(\mathbf{X}'\mathbf{X})^{-1})$ where \mathbf{X} is the new design matrix and g is fixed at 0.5 or 1. For the tailfree prior, let $J = 5$

and $c \sim \Gamma(10, 1)$. Based on Table 3.8, point estimates for β_0 and β_1 are close to those fitted with linearity assumption for $h(w_2)$ but credible intervals are much wider due to an increased number of parameters. Figure 3.5 plots the point-wise estimates (solid lines) and 95% credible intervals (dashed lines) for $h(w_2)$ showing no significant nonlinear trend. Bayes factors for the test of linearity of $h(w_2)$ are less than one and hence the linear assumption is preferred.

Table 3.8: Summary of the coefficients for Kijima type I model for Syringe-driver maintenance data; $h(w_2)$ is approximated by a B-spline, estimates are posterior means, and 95% CI are credible intervals.

Coefficient	$g = 0.5$	$g = 1$
	$\hat{\beta}$ (95% CI)	$\hat{\beta}$ (95% CI)
$\beta_0(pm)$	-0.12 (-1.5, 1.17)	-0.04 (-2.12, 4.63)
$\beta_1(cm)$	2.02 (-2.02, 8.06)	2.53 (-5.37, 9.43)
$\beta_2(cost)$	–	–

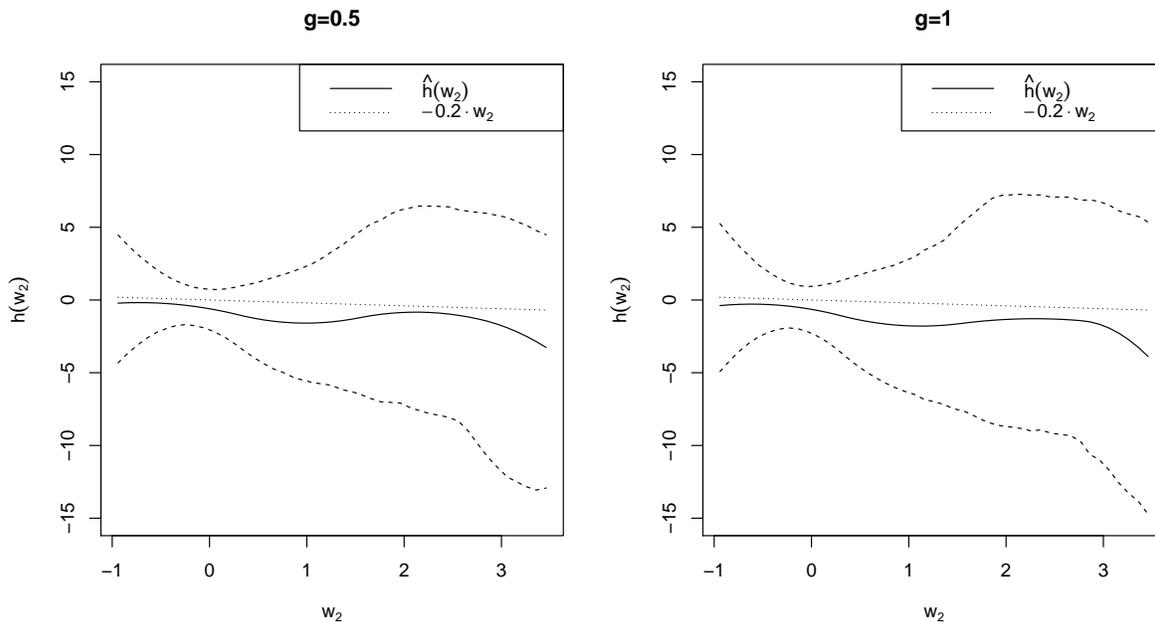


Figure 3.5: Estimate (solid) of $h(w_2)$ and its 95% credible intervals (dashed) for type I model with logistic (exponential) link on the left (right) panel; dotted lines are linear functions, i.e. $h(w_2) = -0.2w_2$ on the left panel.

3.6 DISCUSSION

We proposed a new semiparametric regression model for recurrent events arising from maintenances of repairable systems where effectiveness of repairs characterized by covariates are taken into account in the joint modeling. We generalized the Kijima effective age models (Kijima 1989) by regressing the age reduction factors on covariates. The baseline distribution is flexibly modeled using a tailfree prior, which generalizes the commonly-used Weibull family allowing for data-driven flexibility. Logistic and exponential links are proposed for the regression and efficient, adaptive, and easy-to-implement MCMC is described. The proposed method was illustrated using simulations and two data analyses. We found useful and interesting interpretations of regression coefficients when examining the effect of covariates on the effectiveness of repairs. When the link is the exponential function, the proposed semiparametric regression model provides an easy test for the common assumption of minimal repair (“bad as old repair”) which is also appealing to practitioners.

The regression parameters are interpretable since an increase or decrease of effective age is closely related to intensity of the system. However when the hazard of the system is not monotone, the interpretation becomes more difficult. Finally, we note that it is also straightforward to generalize our model to include heterogeneous systems by including random system effects in the linear predictor or times the intensity function.

CHAPTER 4

SPATIAL EXTENDED HAZARD MODEL WITH APPLICATION TO PROSTATE CANCER SURVIVAL

This paper 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 copula model of Li and Lin (2006), using a correlation structure implied by an intrinsic conditionally autoregressive prior (ICAR). Efficient MCMC 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.

Keywords: Censored data; Copula; Normal transformation model; intrinsic autoregressive prior.

4.1 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; Komarek 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

$$\lambda(x|\mathbf{z}) = \lambda_0(xe^{\beta'\mathbf{z}})e^{\gamma'\mathbf{z}}. \quad (4.1)$$

The more easily interpretable PH, AFT, and AH models occur as special cases when $\beta = \mathbf{0}$, $\beta = \gamma$, and $\gamma = 0$ 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 $\lambda(x|\mathbf{z}) = \lambda_0(xe^{\beta z_1})e^{\beta z_1 + \gamma z_2}$. Holding z_1 constant, z_2 has PH interpretation; e^γ is how the instantaneous risk of death changes when z_2 is increased by one unit. Holding z_2 constant, z_1 has AFT interpretation; mean or any percentile of lifetime (e.g. median) changes by a factor of e^β when increasing z_1 by a unit. Such reduced semiparametric models have enhanced interpretability, separating inference into easily interpretable parametric (regression coefficients) and nonparametric (baseline hazard) portions.

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. Multivariate survival data can be handled conditionally through the use of frailties, marginally through modifications of the hazard (e.g. Cai et al., 2007), or marginally through a transformation model, also called a copula model. Both marginal methods produce the "population-averaged" covariate effects, however the normal transformation model 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

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

where, for our data, b_1, \dots, b_{46} and g_1, \dots, g_{46} are county-level frailties for South Carolina and c_i is the county subject i belongs to. To test that PH is adequate, the

hypotheses $H_0 : \boldsymbol{\beta} = \mathbf{0}, b_j = 0, j = 1, \dots, 46$ need be considered; the per-variable tests are even more complex. In contrast, the copula approach is both more easily implemented and allows ready interpretation. Li and Lin (2006) consider estimation in the PH model for spatially correlated georeferenced data. Note that the georeferenced approach of Li and Lin (2006) 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 MCMC schemes for posterior updating. Since all of these models are formally nested within the EH model, Bayes factors are quickly computed using the Savage-Dickey ratio (Verdinelli and Wasserman 1995).

Define $Y_i = \Phi^{-1} \{1 - e^{\Lambda_i(X_i)}\}$ where $\Phi(\cdot)$ is the standard normal cumulative distribution function, X_i the random failure time and $\Lambda_i(\cdot)$ the cumulative hazard function for the i th subject. Let $\mathbf{Y} = (Y_1, \dots, Y_n)'$. Under the marginal 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,

$$\mathbf{Y} \sim N(\mathbf{0}, \boldsymbol{\Gamma}). \quad (4.2)$$

The normal transformation model incorporates covariate effects in $\lambda_i(\cdot)$ through (4.1) and spatial dependence by $\boldsymbol{\Gamma}$. We introduce the spatial transformation model (4.9) in Section 3.3 for county-level data where transformed survival times follow the correlation structure $\boldsymbol{\Gamma}$ implied by the intrinsic conditionally autoregressive (ICAR) model (Besag, York, and Mollie, 1991; Banerjee, Carlin, and Gelfand, 2004). The framework developed here allows immediate extension to proper CAR and exchangeable correlation structures as well.

There has been renewed, recent interest in stably estimating the EH model. Both Tseng and Shu (2011) and Tong et al. (2013) 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 , 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 amount of hazard jumps needed to be used in that it is not known *a priori* what the effective support of λ_0 is and the factor $e^{\beta'z}$ by which time accelerates or decelerates in the argument of the hazard, survival, and density functions. 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_{θ} , have such scale parameters, e.g. 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 $\lambda_{\theta}(t)$ 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. More investigation of the baseline hazard modeling is included in Section 2.1 and Section 5.

We analyze prostate cancer data 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 incidence registry covering the entire state of South Carolina that consistently receives the highest gold rating from the North American Association of Central Cancer Registries, indicating 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 extended hazard model.

This paper is organized as follows. Section 2 presents the proposed method for

fitting the extended hazards 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 without the introduction of frailties. Section 4 proposes Bayes factors for testing whether PH, AFT, or AH holds globally, as well as per-variable tests. Simulation results are summarized in Section 5. An analysis of the SCCCR prostate cancer data using the proposed model is presented in Section 6. We conclude the paper in Section 7.

4.2 EXTENDED HAZARD MODEL

In the EH model (4.1), $\boldsymbol{\beta}$ 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 X and C be random failure and censoring time respectively. Conditioning on a p -dimensional covariate vector \mathbf{z} , we assume X and C are independent. Consider n subjects in the study; each subject is observed with an event time $t_i = \min\{c_i, x_i\}$ and a censoring indicator $\delta_i = I(x_i < c_i)$. The likelihood based on data $\mathcal{D} = \{(t_i, \delta_i, \mathbf{z}_i)\}_{i=1}^n$ under model (4.1) is

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda_0(\cdot)) = \prod_{i=1}^n \left\{ e^{\boldsymbol{\gamma}'\mathbf{z}_i} \lambda_0(e^{\boldsymbol{\beta}'\mathbf{z}_i} t_i) \right\}^{\delta_i} \exp \left\{ -e^{\boldsymbol{\gamma}'\mathbf{z}_i} \int_0^{t_i} \lambda_0(t e^{\boldsymbol{\beta}'\mathbf{z}_i}) dt \right\} \quad (4.3)$$

Proper priors are required to compute Bayes factors; the most common choice are g-priors. For $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, i.e.

$$\pi(\boldsymbol{\beta}) \sim N_d(\mathbf{0}, g_1(\mathbf{Z}'\mathbf{Z})^{-1}); \quad \pi(\boldsymbol{\gamma}) \sim N_d(\mathbf{0}, g_2(\mathbf{Z}'\mathbf{Z})^{-1}); \quad g_1, g_2 \sim \text{Gamma}(a_g, b_g), \quad (4.4)$$

where $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_n)'$. Recently, the g-prior has been advocated for nonlinear regression models (e.g., Rathbun and Fei, 2006; Bové and Held, 2011). We have found inference to be robust to the prior specification on (g_1, g_2) .

Baseline hazard

B-splines are now a standard tool in the modeling of hazard functions, and in fact generalize the piecewise constant (first order B-spline) hazards extensively used in Bayesian survival analysis in the 1990's (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 accounting for lattice spatial correlation marginally through a copula, 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

$$\lambda_0(t) = \sum_{j=1}^J b_j B_{kj}(t), \quad (4.5)$$

where $B_{kj}(\cdot), \dots, B_{kJ}(\cdot)$ are k th order B-spline basis functions expanded over a knot sequence $\mathbf{s} = (s_{1-k}, \dots, s_J)$ (De Boor 2001) and b_1, \dots, b_J are the positive B-spline coefficients. Set the boundary knots $\mathbf{s}_{(1-k):0} = 0$, and $\mathbf{s}_{(J-k+1):J} = s_{max}$. Let $F_{\boldsymbol{\theta}}(\cdot)$ be the cumulative distribution function for $\lambda_{\boldsymbol{\theta}}(\cdot)$. Let p_1, \dots, p_{J-k+1} be probabilities between 0 and 1 in an increasing order. Our default choice is $p_j = jp_{max}/(J-k+1)$ where p_{max} is a constant set to close to one. Set $s_j = F_{\boldsymbol{\theta}}^{-1}(p_j)$, $j = 1, \dots, J-k+1$. The proposed method automatically allocates more knots in regions of higher mass under the parametric family, and works very well in simulations and in our data analysis. 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), $\lambda_0(t) = \lambda_{\boldsymbol{\theta}}(\tilde{s}_j) B_{kj}(t)$ approximates $\lambda_{\boldsymbol{\theta}}(t)$ over $t \in (0, s_{max})$ with uniformly bounded error, i.e. $\max_{0 \leq t \leq s_{max}} \|\lambda_0(t) - \lambda_{\boldsymbol{\theta}}(t)\| \leq 2 \max\{|\lambda_{\boldsymbol{\theta}}(x) - \lambda_{\boldsymbol{\theta}}(y)| : |x - y| \leq \min\{s_{max}/\sqrt{2k-2}, \max_j\{s_{j+1} - s_j\}\sqrt{k/12}\}\}$.

To center $\lambda_0(t)$ at λ_θ in our Bayesian framework, we take the prior mean $E(b_j) = \lambda_\theta(\tilde{s}_j)$, specifically $b_j \sim \text{Gamma}(c\lambda_\theta(\tilde{s}_j), c)$ where the prior variance is $\lambda_\theta(\tilde{s}_j)/c$. The scalar c controls the how stochastically ‘close’ λ_0 is to λ_θ under the prior and is assigned a prior $\Gamma(a_c, b_c)$. The distribution family F_θ anchors the prior shape of λ_0 . A data-driven prior for θ is assumed by first obtaining maximum likelihood estimates under the underlying parametric EH model $\hat{\theta}$, and its associated inverse information matrix \mathbf{V}_θ ; θ is then assigned Gaussian prior $N(\hat{\theta}, a_\theta \mathbf{V}_\theta)$ where $a_\theta > 1$ is a scalar. The number of B-spline basis function J is a typically chosen between 20 to 40; see Ruppert (2002) for a detailed discussion on selection of J .

In summary,

$$b_j \stackrel{\text{ind.}}{\sim} \text{Gamma}(c\lambda_\theta(\tilde{s}_j), c), j = 1, \dots, J; c \sim \text{Gamma}(a_c, b_c); \theta \sim N_2(\hat{\theta}, a_\theta \mathbf{V}_\theta). \quad (4.6)$$

Sharef et al. (2010) state that “*Unfortunately, it is not easy to define an informative prior on the spline parameters that induces shrinkage towards a specified parametric target*” and thus posit a hazard model that is a weighted sum of a parametric hazard and a penalized B-spline. 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.

MCMC sampling

The joint posterior of $\{\beta, \gamma, \mathbf{b}, \theta, c\}$ following likelihood (4.3) and priors (4.6) and (4.4) is

$$p(\beta, \gamma, \mathbf{b}, \theta, c | \mathcal{D}) \propto L(\beta, \gamma, \mathbf{b}, \theta) \pi(\beta) \pi(\gamma) \pi(\theta) p(c) \prod_{j=1}^J p(b_j), \quad (4.7)$$

where $\mathbf{b} = (b_1, \dots, b_J)$. 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 λ_0 , we follow Lin and Wang (2011), and introduce latent variables $u_{ij}, j = 1, \dots, J$ constrained such that $\sum_{j=1}^J u_{ij} = 1$ for $i \in S^c$. The MCMC sampling steps follow.

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

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

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

$$\mathbf{u}_i | \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{b}, \boldsymbol{\theta}, \mathbf{u}_{-i} \sim \text{Multinomial} \left(\frac{b_1 B_1(e^{\boldsymbol{\beta}'\mathbf{z}_i t_i})}{\sum_{j=1}^n b_j B_j(e^{\boldsymbol{\beta}'\mathbf{z}_i t_i})}, \dots, \frac{b_J B_J(e^{\boldsymbol{\beta}'\mathbf{z}_i t_i})}{\sum_{j=1}^n b_j B_j^*(e^{\boldsymbol{\beta}'\mathbf{z}_i t_i})} \right).$$

Step 4: B-spline coefficients are updated by

$$b_j | \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{b}_{-j}, \boldsymbol{\theta}, \{\mathbf{u}_i\}_{i \in S^c} \sim \text{Gamma} \left(\sum_{i \in S} u_{ij} + c \lambda_{\boldsymbol{\theta}}(\tilde{s}_j), c + \sum_{i=1}^n e^{\boldsymbol{\gamma}'\mathbf{z}_i} \int_0^{t_i} B_j(te^{\boldsymbol{\beta}'\mathbf{z}_i}) dt \right).$$

Updating the baseline hazard using the augmented $\{u_{ij}\}$ and adaptive Metropolis, coupled with a data-driven prior on $\boldsymbol{\theta}$ has given very efficient MCMC chains. More details on computation time are presented in Section 5.

4.3 SPATIAL CORRELATION

The normal transformation model of Li and Lin (2006) is extended to areal data. Relevant papers 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.

Likelihood of the normal transformation model

Under the the normal transformation model (4.2), Li and Lin's (2006) likelihood simplifies to

$$L_s(\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{b}, \boldsymbol{\theta}, \boldsymbol{\Gamma}) = \int \left[\prod_{i \in S} \frac{f_i(t_i)}{\phi(y_i)} \right] \left[\prod_{i \in S^c} \frac{f_i(x_i)}{\phi(y_i)} I(x_i > t_i) \right] \phi(\mathbf{y}; \mathbf{0}, \boldsymbol{\Gamma}) \prod_{i \in S^c} dx_i \quad (4.8)$$

where

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

and $\mathbf{\Gamma}$ is a positive definite matrix with diagonal elements being one. It is difficult to integrate out $\{x_i, i \in S^c\}$ both theoretically and numerically.

In the next section, we design $\mathbf{\Gamma}$ for areal data based on the commonly used CAR and ICAR model.

CAR and ICAR correlation structures

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 $i = 1, \dots, m$. Let d_i be the total number of adjacent units for unit i , i.e. $d_i = \sum_{j=1}^m w_{ij}$. Denote $\mathbf{D} = \text{diag}(d_1, \dots, d_m)$ and assume \mathbf{Y} is the vector of transformed failure times $\{Y_i = \Phi^{-1}(F_i(X_i)), i = 1, \dots, n\}$ arranged by areal units. Let \tilde{Y}_{ij} be a normal random variable for the j th individual in unit i . Denote $\tilde{\mathbf{Y}} = (\tilde{Y}_{11}, \dots, \tilde{Y}_{1n_1}, \dots, \tilde{Y}_{m1}, \dots, \tilde{Y}_{mn_m}) = (\tilde{\mathbf{Y}}'_1, \dots, \tilde{\mathbf{Y}}'_m)$ where n_i is the number of observations in unit i . Assume $\mathbf{\Gamma} = \text{cov}(\mathbf{Y}) = \text{corr}(\tilde{\mathbf{Y}})$.

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

$$\tilde{Y}_{ij} = \alpha_i + \epsilon_{ij}; \quad \boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_m)' \sim N(\mathbf{0}, \boldsymbol{\Omega}); \quad \epsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2). \quad (4.9)$$

where $\alpha_i, i = 1, \dots, m$ are the random effects with a covariance matrix $\boldsymbol{\Omega} = (\omega_{ij})$ and ϵ_{ij} is the error term with variance σ^2 for the j th subject in the unit i . The covariance structure of $\boldsymbol{\alpha}$ characterizes the spatial dependence among areal units. Popular models for $\boldsymbol{\alpha}$ include independent normals (i.e. exchangeable), conditional autoregressive (CAR) models, ICAR, simultaneous autoregressive models (SAR), and many others. Under the proper CAR model, $\boldsymbol{\alpha} | \varphi, r \sim N_m(\mathbf{0}, \varphi^{-1}(\mathbf{D} - r\mathbf{W})^{-1})$

where r and φ are scalars. That is, $\boldsymbol{\Omega} = \varphi^{-1}(\mathbf{D} - r\mathbf{W})^{-1}$ implying $\alpha_j | \boldsymbol{\alpha}_{-j}, \varphi \sim N(r \sum_{j=1}^n w_{ij} \alpha_j / w_{j+}, 1/(\varphi w_{j+}))$. When $r = 0$, $\tilde{\mathbf{Y}}_i$ is independent of $\tilde{\mathbf{Y}}_j$ if $j \neq i$. It is difficult to estimate r and φ simultaneously, as noticed by many authors. One way is to fix r . Allowing $r \rightarrow 1$ leads to the ICAR model. However, the implied covariance matrix of $\boldsymbol{\alpha}$ under ICAR model is improper as $p(\boldsymbol{\alpha}) \propto \exp(-\varphi \boldsymbol{\alpha}'(\mathbf{D} - \mathbf{W})\boldsymbol{\alpha}/2)$ and $\mathbf{D} - \mathbf{W}$ is singular. A common strategy to restore the propriety is to impose the constraint $\sum_{j=1}^m \alpha_j = 0$ during Gibbs sampling. In Appendix A, we derive the implied covariance matrix under this constraint and yield $\text{cov}(\boldsymbol{\alpha}) = \varphi^{-1}\boldsymbol{\Omega}^*$ where $\boldsymbol{\Omega}^* = (\omega_{ij}^*)$ is a matrix only depending on the neighboring matrix \mathbf{W} . The resulting correlation matrix $\boldsymbol{\Gamma}$ under the ICAR model only involves one unknown quantity $\varphi\sigma^2$. Denote $\varphi^* = \varphi\sigma^2$. A smaller value of φ^* corresponds to stronger spatial dependence within county and across counties.

Latent survival times approach

As mentioned in Section 3.1, it is difficult to evaluate the likelihood function (4.8). In this section, we introduce a latent failure time X_i for each $i \in S^c$. The augmented likelihood for the spatial model is

$$L_s^A(\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{b}, \boldsymbol{\theta}, \boldsymbol{\Gamma}, \{x_i, i \in S^c\}) = \left[\prod_{i \in S} \frac{f_i(t_i)}{\phi(y_i)} \right] \left[\prod_{i \in S^c} \frac{f_i(x_i)}{\phi(y_i)} I(x_i > t_i) \right] \phi(\mathbf{y}; \mathbf{0}, \boldsymbol{\Gamma}) \quad (4.10)$$

The conditional posterior of X_i given the other parameters is

$$p(x_i | \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{b}, \boldsymbol{\theta}, \boldsymbol{\Gamma}, \{x_j, j \neq i\}) \propto \frac{f_i(x_i)}{\phi(\Phi^{-1}(F_i(x_i)))} \phi(\Phi^{-1}(F_i(x_i)) | \mathbf{y}_{-i}, \boldsymbol{\Gamma}) I(x_i > t_i), \quad i \in S^c.$$

Although this looks horrendous, the x_i are easily sampled. Firstly y_i is sampled from a truncated normal distribution $N(y_i | \mathbf{y}_{-i}, \boldsymbol{\Gamma}) I(y_i > \Phi^{-1}(F_i(t_i)))$ (e.g. Geweke, 1991) then $x_i = F_i^{-1}(\Phi(y_i))$ can be obtained through bisection or the Newton-Raphson algorithm. In addition to the imputation of censored failure times, we need to evaluate Γ^{-1} in an efficient way. For high-dimensional lattice data, we obtain a closed-form of

Γ^{-1} for the Γ induced from the random effects model (4.9), with details presented in Appendix B.

Based on the augmented likelihood (4.10) and the priors (4.6) and (4.4), MCMC sampling steps 1 to 3 are similar as those in Section 2.2. Sampling $\{\mathbf{b}, \varphi^*\}$ are accomplished as follows.

Step 4: Propose b_j^{new} from

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

and accept it with probability

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

where \mathbf{y}^{new} is new transformed failure time vector corresponding to b_j^{new} . Evaluation of $\mathbf{y}' \Gamma^{-1} \mathbf{y}$ is efficiently carried out in Appendix B.

Step 5: Sample $Y_i \sim N(y_i | \mathbf{y}_{-i}, \mathbf{\Gamma}) I(y_i > \Phi^{-1}(F_i(t_i)))$ then set $x_i = F_i^{-1}(\Phi(y_i))$.

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 next section, we propose an alternative approach.

Random-effect approach for lattice data

Based on model (4.9), consider the following random effects model

$$\tilde{Y}_{ij} = \alpha_i^* + \epsilon_{ij}^*; \quad \alpha_i^* = \alpha_i / \sqrt{\omega_{ii} + \sigma^2}; \quad \epsilon_{ij}^* = \epsilon_{ij} / \sqrt{\omega_{ii} + \sigma^2}. \quad (4.11)$$

Note that the correlation structure under model (4.9) and (4.11) are equal. We refer α_i^* as random effect.

Let \tilde{X}_{ij} be the failure time for the j th observation in county i and \tilde{t}_{ij} be the observed failure time associated with \tilde{X}_{ij} . Conditional on the county effect, the

likelihood contribution for a censored observation is

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

and for an observed observation 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} increases as α_i^* increases, holding φ^* constant. With the addition of $\boldsymbol{\alpha}^*$, the augmented joint likelihood (4.8) is written as

$$L = \int \prod_{(i,j) \in S} f_{ij}(\tilde{t}_{ij}|\alpha_i^*) \prod_{(i,j) \in S^c} S_{ij}(\tilde{t}_{ij}|\alpha_i^*) P(\alpha_1^*, \dots, \alpha_m^*) d\alpha_1^* \dots d\alpha_m^*. \quad (4.12)$$

The dimension of integration in (4.12) is typically much lower than that in (4.9) for highly censored data. The sampling steps for \mathbf{b} and $\alpha_1^*, \dots, \alpha_m^*$ are carried out through adaptive Metropolis-Hasting steps.

4.4 HYPOTHESIS TESTS USING BAYES FACTORS

Tests on null hypotheses $H_0 : \boldsymbol{\beta} = \mathbf{0}$, $H_0 : \boldsymbol{\gamma} = \mathbf{0}$, and $H_0 : \boldsymbol{\beta} = \boldsymbol{\gamma}$ lead to global comparisons of EH to PH, AH and AFT model respectively. Let M_1 denotes the model under the null hypothesis, i.e. PH, AH, and AFT and M_2 denotes EH under the alternative. The Bayes factor for comparing models M_1 and M_2 is

$$BF_{12} = \frac{\int f(\mathcal{D}|\Upsilon, M_1) \pi_0(\Upsilon) d\Upsilon}{\int f(\mathcal{D}|\tilde{\Upsilon}, M_2) \pi(\tilde{\Upsilon}) d\tilde{\Upsilon}},$$

where $f(\mathcal{D}|\cdot)$ is the likelihood of the data \mathcal{D} , Υ and $\tilde{\Upsilon}$ are the parameters of model M_1 and M_2 respectively, and $\pi_0(\Upsilon)$ and $\pi(\tilde{\Upsilon})$ are the prior densities. Our M_1 models are restricted versions of M_2 , e.g. when M_1 is PH, $\tilde{\Upsilon} = (\boldsymbol{\beta}, \Upsilon)$ and $f(\mathcal{D}|\Upsilon, M_1) = f(\mathcal{D}|\boldsymbol{\beta} = \mathbf{0}, \Upsilon, M_2)$. If $\pi(\Upsilon|\boldsymbol{\beta} = \mathbf{0}) = \pi_0(\Upsilon)$, the Bayes factor BF_{12} for comparing PH to EH is reduced to a Savage-Dicky ratio (Verdinelli and Wasserman 1995),

$$BF_{12} = \frac{\pi(\boldsymbol{\beta} = \mathbf{0}|\mathcal{D})}{\pi(\boldsymbol{\beta} = \mathbf{0})},$$

where $\pi(\boldsymbol{\beta})$ and $\pi(\boldsymbol{\beta}|\mathcal{D})$ are the prior and posterior distributions of $\boldsymbol{\beta}$ under M_2 . When the assumption does not hold, Verdinelli and Wasserman(1995) provided the correct version: $BF_{12} = \pi(\boldsymbol{\beta} = \mathbf{0}|\mathcal{D})/\pi(\boldsymbol{\beta} = \mathbf{0})E[\pi_0(\boldsymbol{\gamma})/\pi(\boldsymbol{\gamma}|\boldsymbol{\beta} = \mathbf{0})]$ assuming that the expectation is finite, where the expectation is with respect to $\pi(\boldsymbol{\gamma}|\boldsymbol{\beta} = \mathbf{0}, \mathcal{D})$.

Raftery (1994) noted that often only a crude approximation to the Bayes factor is needed. Suggested by Verdinelli and Wasserman(1995), the posterior distribution can be approximated by a normal distribution using MCMC iterates. Bayes factors per variable can also be computed, which provide further guidance on choosing sub-models, as illustrated in our data analysis.

4.5 SIMULATIONS

Simulation I examines the proposed method described in Section 2 for fitting the EH model. Simulation II illustrates the MCMC procedure presented in Section 3 for fitting the spatial model.

Simulation I: Data of size $n = 300$ are repeatedly simulated from the (1) AFT, (2) AH, (3) PH, and (4) EH models; 300 replicates are simulated for each scenario. The baseline distribution is $0.5\text{lognormal}(1, 0.2) + 0.5\text{lognormal}(2, 0.2)$. The first covariate Z_1 is generated from Bernoulli(0.5) and the second Z_2 is generated from $N(0, 0.5^2)$. A uniform distribution $U(0, a)$ is used to induce 30% and 0% right-censored observations. We set the number of B-spline basis functions $J = 20$ and knots to be $F_{\boldsymbol{\theta}}^{-1}(p_j)$ where p_j s are equally spaced over $[0, p_{max} = 0.995]$ and $F_{\boldsymbol{\theta}}$ is a log-normal distribution. The centering hazard family $\boldsymbol{\theta}$ follows $N(\hat{\boldsymbol{\theta}}, a_{\boldsymbol{\theta}}\mathbf{V}_{\boldsymbol{\theta}})$ where $\hat{\boldsymbol{\theta}}$ and $\mathbf{V}_{\boldsymbol{\theta}}$ are obtained through a fit of parametric AFT, AH, PH, or EH. Set hyper-parameters $a_g = b_g = 0.1$ and $a_{\boldsymbol{\theta}} = 10$. For each simulated dataset, it takes approximately ten seconds to run a chain of 30,000 iterates using Fortran. We diagnosed the iterates using tests and trace plots provided in the CODA package in R and found that the chain converged quickly after a few hundreds steps.

Summaries of regression coefficients from fitting EH model are presented in Table 4.1. Bias, SSD, ESE, and 95% are the averaged bias of the posterior means, standard error of the posterior means, average of sample standard errors and coverage probability of the 95% credible intervals respectively. The coefficients are estimated with low bias. SSD and ESE are close across all simulation setups. Coverage probabilities get higher when the censoring rate decreases.

Based on the posterior mean estimates of survival, density, and hazard functions for the 300 data sets with 0% censoring in scenario (1), Figure 4.1 plots the means, 2.5%, and 97.5% quantiles of the estimates on the left panels. The true functions are captured well. To investigate the effect of baseline choice, we refit the data sets choosing F_θ to be log-logistic distribution. The estimates are plotted on the right panels of Figure 4.1; there is essentially no difference over the observed range of the data.

Based on the posterior samples for datasets with 30% censoring, we compute Bayes factors for comparing the simpler models to EH. When the data are generated according to AFT, $BF > 1$ for AFT vs. EH 100% of the time; $BF < 1$ for PH and AH 100%. When the data are generated according to PH, $BF > 1$ 100% of the time for PH vs. EH; $BF < 1$ 100% for AH vs. EH and 39% and for AFT vs. EH. When data are truly AH, $BF > 1$ 97% of the time for AH vs. EH; $BF < 1$ 100% of the time for AFT vs. EH and PH vs. EH. Finally, when data are truly EH, $BF < 1$ 100% of the time for AH and PH vs. EH, but $BF < 1$ 65% of the time for AFT vs. EH.

Simulation II: We sample data from EH model with ICAR spatial dependence (4.9) across SC counties. Consider a sample size $n = 500$. The subjects are assigned to the counties with equal probabilities; each county has at least one subject. The spatial dependence is characterized by φ^* . We consider two values for φ^* : 5 and 10. Assume the true regression parameters β be (0.5, 0.5) and γ be (0.5, 0.5). Also let

Table 4.1: Summaries for Simulation I: $n = 300$ and censoring rates are 30% and 0%; baseline distribution is $0.5\text{lognormal}(1, 0.2) + 0.5\text{lognormal}(2, 0.2)$; SSD: standard error of the posterior means; ESE: average of sample standard errors.

		30%					0%				
(1) AFT	True	Bias	SSD	ESE	95%	Bias	SSD	ESE	95%		
	β_1	0	0.004	0.042	0.036	0.887	0.000	0.036	0.030	0.913	
	β_2	1	0.003	0.046	0.038	0.903	0.004	0.036	0.030	0.946	
	γ_1	0	-0.017	0.144	0.146	0.970	-0.015	0.118	0.124	0.960	
	γ_2	1	0.012	0.160	0.160	0.927	0.033	0.131	0.129	0.947	
(2) AH		Bias	SSD	ESE	95%	Bias	SSD	ESE	95%		
	β_1	0.5	0.005	0.046	0.037	0.870	0.002	0.035	0.030	0.940	
	β_2	0.5	0.005	0.044	0.039	0.896	0.002	0.038	0.032	0.903	
	γ_1	0	-0.015	0.142	0.147	0.953	0.029	0.140	0.138	0.946	
	γ_2	0	-0.024	0.161	0.157	0.953	-0.002	0.152	0.146	0.940	
(3) PH		Bias	SSD	ESE	95%	Bias	SSD	ESE	95%		
	β_1	0	0.001	0.047	0.034	0.860	0.004	0.037	0.032	0.918	
	β_2	0	0.000	0.046	0.037	0.900	0.003	0.036	0.033	0.918	
	γ_1	1	0.004	0.164	0.160	0.940	-0.013	0.125	0.126	0.948	
	γ_2	1	0.008	0.181	0.171	0.930	-0.009	0.125	0.130	0.960	
(4) EH		Bias	SSD	ESE	95%	Bias	SSD	ESE	95%		
	β_1	0	0.002	0.038	0.032	0.903	-0.001	0.036	0.029	0.900	
	β_2	0.5	-0.004	0.039	0.034	0.933	0.004	0.037	0.030	0.923	
	γ_1	0.5	0.032	0.145	0.145	0.953	0.023	0.126	0.127	0.937	
	γ_2	1	0.027	0.158	0.152	0.960	0.046	0.131	0.130	0.923	

the baseline distribution be $0.5\text{lognormal}(1, 0.2) + 0.5\text{lognormal}(2, 0.2)$; 300 replicates are simulated for each scenario, each dataset has 30% censored observations. Prior specifications are the same as in Simulation I. The imputation method of Section 3.3 is used. For each simulated dataset, it takes approximately a few minutes to run a chain of 30,000 iterates using Fortran. Results are displayed in Table 4.2.

Table 4.2: Summaries for Simulation II: sample size $n = 500$ with 30% censored observations; baseline distribution is $0.5\text{lognormal}(1, 0.2) + 0.5\text{lognormal}(2, 0.2)$; SSD: standard error of the posterior means; ESE: average of sample standard errors.

	True value	Bias	SSD	ESE	95%CP	True value	Bias	SSD	ESE	95%CP
β_1	0.5	-0.004	0.056	0.048	0.905	0.5	-0.001	0.058	0.049	0.900
β_2	0.5	0.005	0.032	0.027	0.915	0.5	0.000	0.032	0.028	0.890
γ_1	0.5	0.019	0.204	0.198	0.940	0.5	0.001	0.201	0.198	0.940
γ_2	0.5	-0.020	0.123	0.112	0.910	0.5	0.000	0.123	0.113	0.915
φ^*	5	-0.093	4.470	6.060	0.930	10	-1.080	7.450	12.85	0.915

4.6 DATA ANALYSIS

We analyze SCCCR prostate cancer data for the period 1996–2004. Covariates include county of residence at diagnosis, standardized age at diagnosis, race, marital status at diagnosis, grade of tumor differentiation, and SEER summary stage. Table 4.3 provides summaries for the categorical covariates. There are $N = 20599$ patients in the dataset after excluding subjects with missing information; 72.3% of the survival times are right-censored.

Table 4.3: Summary characteristics of prostate cancer patients in SC from 1996-2004.

Covariate		n	Sample percentage
Race	Black	6483	0.32
	White	14116	0.68
Marital status	Non-married	4525	0.22
	Married	16074	0.78
Grade	well or moderately differentiated	15309	0.74
	poorly differentiated or undifferentiated	5290	0.26
SEER summary stage	Localized or regional	19792	0.96
	Distant	807	0.04

The purpose of the study is to quantify racial disparity in prostate cancer survival, adjusting for the remaining risk factors and accounting for the county the subject lives in. We expect patients residing in the same county to be positively 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 4.2 which suggests strong spatial patterns in the northwestern and eastern parts of South Carolina. Test on the PH assumption using methods proposed in Grambsch and Therneau (1994) yield a global p-value less than 0.01. We fit the EH model with the following specifications: F_{θ} 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, $\theta \sim N(\hat{\theta}, a_{\theta}\hat{\Sigma}_{\theta})$ where $\hat{\theta}$ and $\hat{\Sigma}_{\theta}$ are obtained by assuming parametric EH model, $\beta \sim N(0, g_1n(\mathbf{X}'\mathbf{X})^{-1})$, $\gamma \sim N(0, g_2n(\mathbf{X}'\mathbf{X})^{-1})$, $a_g = b_g = 0.1$, $a_{\theta} = 1000$, $a_c = b_c =$

0.1, and $J = 30$. After a burn-in of 2000, we obtain 6000 iterates after thinning every other five.

Table 4.4: Summary of fitting the extended hazard model EH, the reduced model, AFT, and PH; * indicates $LPML - 21000$ and $DIC - 42000$.

Covariates		EH	Reduced	AFT $\beta = \gamma$	PH $\beta = 0$	PH+additive age $\beta = 0$
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

Table 4.5: Bayes factors for comparing EH to PH, AFT, and AH with and without spatial correlation.

Covariate	EH			Spatial+EH		
	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

The column under EH in Table 4.4 gives the fitted results. The overall Bayes factors for EH versus PH, AH, and AFT are much greater than 1000, indicating evidence against those commonly assumed models. Variable-specific Bayes factors in Table 4.5 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. To compare to the general EH model, the reduced model just described, AFT, and PH models are fitted using the same prior specifications, except that for PH, knots are fixed and equally spaced—a commonly used way for fitting PH model. The results are displayed in Table 4.4. 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 differences in parameter estimations, LPML, and DIC for the alternative priors $g_1^{-1} \sim \Gamma(0.001, 0.001)$, $g_2^{-1} \sim \Gamma(0.001, 0.001)$. There is also very little difference when increasing J to 50.

We further fit the extended hazards model with spatial dependence (4.2) via the 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 4.6. The Bayes factors after taking into account spatial correlation are presented Table 4.5 under the column Spatial+EH, implying the same reduced model as that in Table 4.4 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, i.e. a partially-linear Cox model with spatial frailties (e.g. Kneib and Fahrmeir, 2007). The nonlinear transformation 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-copula 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 introduced in Section 3.4 and the frailties in the PH model are mapped in Figure (4.2). Note that the random-effects have opposite interpretation from the PH frailties that smaller random effect

Table 4.6: Summary of spatial models; * indicates $LPML - 21000$ and $DIC - 42000$.

Covariates		Marginal EH	Marginal reduced	PH+ICAR+additive age $\beta = 0$
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

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.

Based on the fitted results of the reduced models with and without spatial dependence, white South Carolina subjects diagnosed with prostate cancer in live 22% longer ($e^{0.20} \approx 1.22$) than black patients (95% CI is 18% to 25%), fixing age, stage, and SEER stage. Note that this interpretation pertains to individuals randomly selected from any South Carolina county, i.e. 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 4.3. To compare the survival probabilities for white and black patients, in Figure 4.4 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 one 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.

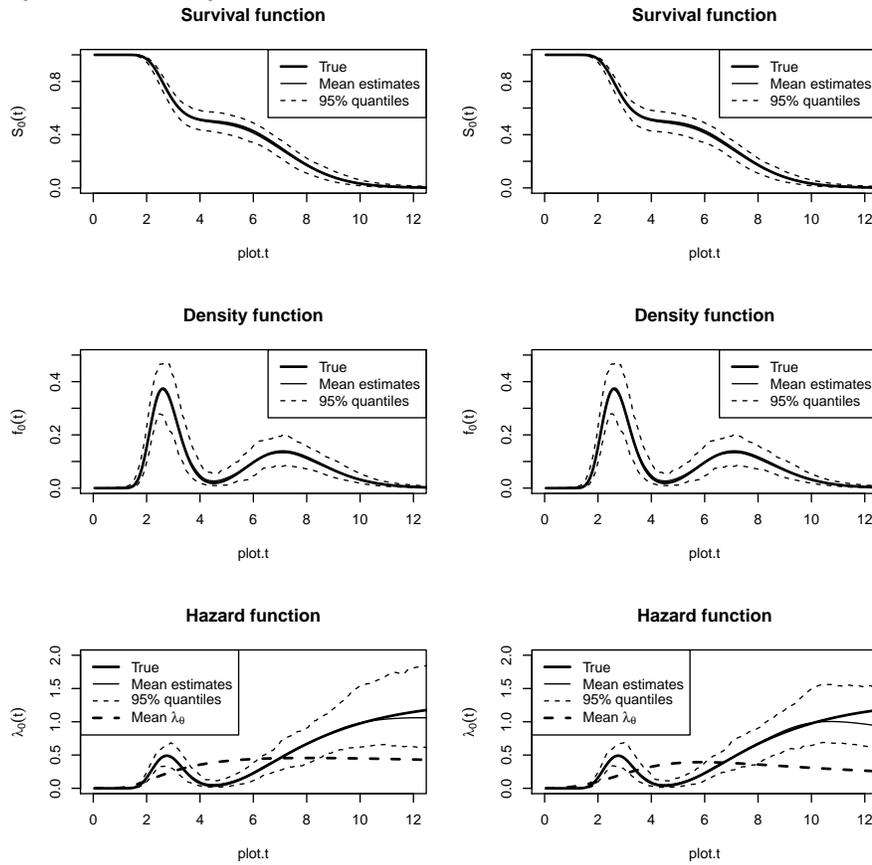


Figure 4.1: Estimates and 95% credible intervals of the baseline survival, density, and hazard functions based on scenario (1) in simulation I with 0% censored subjects. F_{θ} is log-normal distribution for the left panels and log-logistic distribution for the right panels. Bold solid lines are the true functions; solid lines are the mean estimates; dashed lines are the 95% credible intervals; Bold dashed lines on the bottom panels are the mean functions of λ_{θ} .

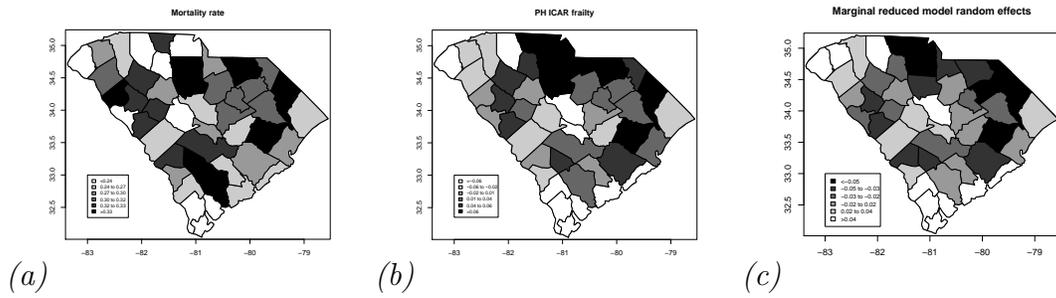


Figure 4.2: Map of (a) Mortality rate, (b) ICAR frailties in the PH model and (c) random effects in the marginal reduced model for SC counties. Larger values of frailties in (b) corresponds to higher risk of hazard function; larger values of random effects in (c) are related to higher survival probabilities.

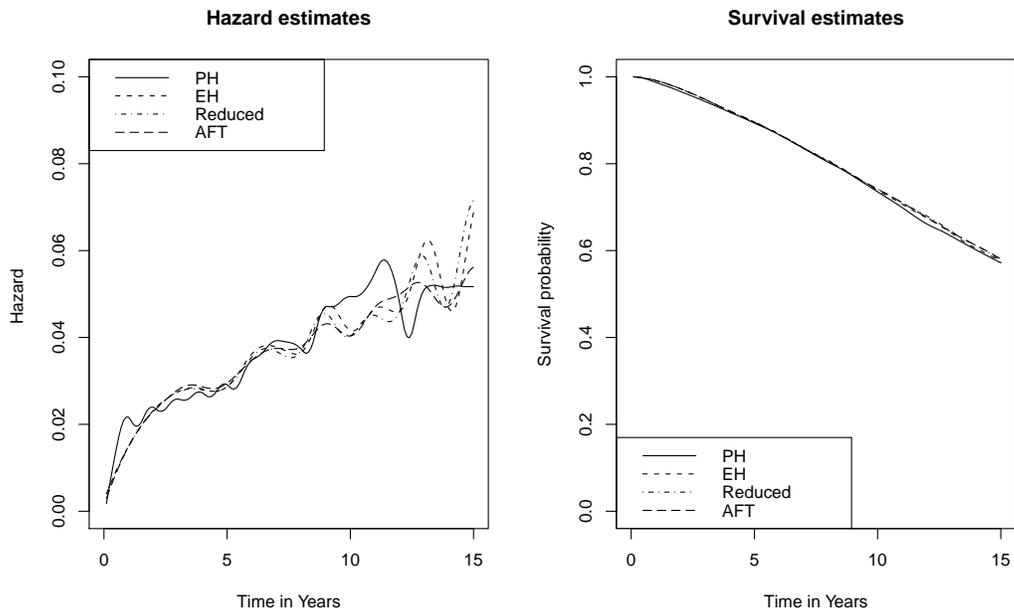


Figure 4.3: Baseline hazard (left) and survival probabilities (right) estimates.

4.7 CONCLUSION

A Bayesian semiparametric method for fitting the extended hazard model to data on South Carolina subjects diagnosed with prostate cancer is developed, and further generalized to include spatially correlated data through a Gaussian copula. 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

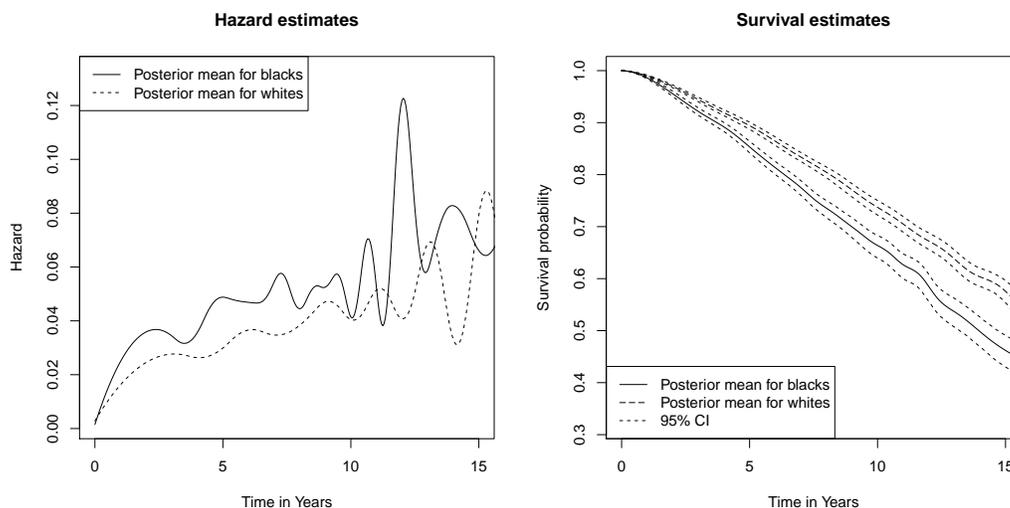


Figure 4.4: Baseline hazard (left) and survival probabilities (right) estimates for black patients (solid line) and white patients (long-dashed line). Short-dashed lines are 95% credible intervals.

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 adjusted mortality for further etiologic research, adjusted for other risk factors.

Cox (1972) is the second most cited statistical paper of all time (Ryan and Woodall, 2005). However, as seen in the the SCCCR analysis presented here, PH can fail to fit actual survival data. Cox himself suggested that PH models are overused, stating “...*the physical or substantive basis for...proportional hazards models...is one of its weaknesses...*” and goes on to suggest that “...*accelerated failure time models are in many ways more appealing because of their quite direct physical interpretation*” (Reid, 1994). Echoing this sentiment, the SCCCR analysis showed that the main covariate of interest, race, is best modeled as an AFT effect.

We propose fitting a large ‘super’ model that encompasses simpler, interpretable models, and deciding on a reduced model where covariates can have one of many plau-

sible, interpretable affects on survival. Other super models have also been proposed in the literature. Scharfstein et al. (1998) propose a special case of transformation models termed the generalized odds-rate model:

$$q_\rho\{S_{\mathbf{x}}(t)\} = -\mathbf{x}'\boldsymbol{\beta} + q_\rho\{S_0(t)\}$$

where $q_\rho(s) = \log\{\rho s^\rho/(1 - s^\rho)\}$. Here, $\rho = 1$ gives PO and $\rho \rightarrow 0^+$ gives PH. Yin and Ibrahim (2005b) propose a Box-Cox transformation of the baseline hazard:

$$\frac{h_{\mathbf{x}}(t)^\rho - 1}{\rho} = \frac{h_0(t)^\rho - 1}{\rho} + \boldsymbol{\beta}'\mathbf{x}(t).$$

Here, $\rho = 1$ gives the AH model and $\rho \rightarrow 0$ gives PH. These authors treat ρ as known when fitting; for these models $\boldsymbol{\beta}$ loses simple interpretability when $\rho \neq 1$ and the estimation of ρ is problematic.

APPENDIX

Covariance matrix of ICAR model under the constraint

In the following, we derive the covariance matrix of $\boldsymbol{\alpha}$ under the constraint $\sum_{j=1}^m \alpha_j = 0$. Under the ICAR prior for $\boldsymbol{\alpha}$, $p(\boldsymbol{\alpha}) \propto \exp(-\varphi\boldsymbol{\alpha}'(\mathbf{D} - \mathbf{W})\boldsymbol{\alpha}/2)$. Note that

$$\boldsymbol{\alpha}'(\mathbf{D} - \mathbf{W})\boldsymbol{\alpha} = \sum_{j=1}^m w_{j+}\alpha_j^2 - \sum_{j=1}^m \sum_{i=1}^m w_{ji}\alpha_i\alpha_j. \quad (13)$$

Under the constraint $\sum_{j=1}^m \alpha_j = 0$, let $\alpha_m = -\alpha_1 - \alpha_2 - \cdots - \alpha_{m-1}$ and plug it into (13), then

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

Let $\mathbf{D}^* = \text{diag}(w_{1+}, \dots, w_{(m-1)+})$, $\mathbf{W}^* = (w_{ij}^*)$ with $w_{ij}^* = w_{ji} - w_{jm} - w_{mi} - w_{m+}$.

Let $\boldsymbol{\alpha}^* = (\alpha_1, \dots, \alpha_{m-1})$. Then under the constraint

$$\boldsymbol{\alpha}'(\mathbf{D} - \mathbf{W})\boldsymbol{\alpha} = \boldsymbol{\alpha}'^*(\mathbf{D}^* - \mathbf{W}^*)\boldsymbol{\alpha}^*. \quad (14)$$

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

$$\Omega^* = \begin{pmatrix} \xi_{11} & \cdots & \xi_{1,m-1} & -\sum_{j=1}^{m-1} \xi_{1j} \\ \vdots & \vdots & \ddots & \vdots \\ \xi_{m-1,1} & \cdots & \xi_{m-1,m-1} & -\sum_{j=1}^{m-1} \xi_{m-1,j} \\ \sum_{j=1}^{m-1} \xi_{j1} & \cdots & \sum_{j=1}^{m-1} \xi_{j,m-1} & -\sum_{j=1}^{m-1} \sum_{i=1}^{m-1} \xi_{ij} \end{pmatrix} \quad (15)$$

The covariance matrix of $\boldsymbol{\alpha}$ under the constraint as $\phi^{-1}\Omega^*$,

Matrix inversion

In the following, we find the inverse of $\boldsymbol{\Sigma} = \text{cov}(\tilde{\mathbf{Y}})$. Based on the random effects model (4.9),

$$\text{cov}(\tilde{\mathbf{Y}}) = \boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{J}_{n_1}\omega_{11} + \mathbf{I}_{n_1}\sigma^2 & \mathbf{J}_{n_1n_2}\omega_{12} & \cdots & \mathbf{J}_{n_1n_m}\omega_{1m} \\ \mathbf{J}_{n_2n_1}\omega_{21} & \mathbf{J}_{n_2}\omega_{22} + \mathbf{I}_{n_2}\sigma^2 & \cdots & \mathbf{J}_{n_2n_m}\omega_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{J}_{n_mn_1}\omega_{m1} & \mathbf{J}_{n_mn_2}\omega_{m2} & \cdots & \mathbf{J}_{n_m}\omega_{mm} + \mathbf{I}_{n_m}\sigma^2 \end{pmatrix} \quad (16)$$

where $\omega_{ij} = \text{cov}(\alpha_i, \alpha_j)$, \mathbf{J}_{n_i} and $\mathbf{J}_{n_in_j}$ are matrix of ones with dimension $n_i \times n_i$ and $n_i \times n_j$ respectively, and \mathbf{I}_{n_i} is an identity matrix with dimension $n_i \times n_i$. Note that

$$\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Omega}\mathbf{P}' + \sigma^2\mathbf{I}_{n^*} \quad (17)$$

where $\mathbf{P} = \text{blockdiag}(\mathbf{1}_{n_1}, \dots, \mathbf{1}_{n_m})$ 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} and $\mathbf{1}_{n_i}$ is a vector of ones with length n_i . Define $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_{n^*})$ where $\mathbf{u}_i = \tilde{\mathbf{u}}_i$ for $i = 1, \dots, m$ and $\mathbf{u}_{m+1}, \dots, \mathbf{u}_{n^*}$ are the orthonormal expansion of $\mathbf{u}_1, \dots, \mathbf{u}_m$. Define $\mathbf{U}_1 = (\mathbf{u}_1, \dots, \mathbf{u}_m)$, $\mathbf{V} = \mathbf{I}_m$, $\mathbf{S}_0 = \text{diag}(\sqrt{n_1}, \dots, \sqrt{n_m})$, 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, $\mathbf{P} = \mathbf{U}\mathbf{S}\mathbf{V}'$.

Therefore, $\Sigma = \mathbf{U}\mathbf{S}\mathbf{U}' + \sigma^2\mathbf{I}_{n^*} = \mathbf{U}\mathbf{K}\mathbf{U}' + \sigma^2\mathbf{I}_{n^*}$ where

$$\mathbf{K} = \begin{pmatrix} \mathbf{K}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \quad \text{and} \quad \mathbf{K}^* = \begin{pmatrix} n_1\omega_{11} & \sqrt{n_1n_2}\omega_{12} & \cdots & \sqrt{n_1n_m}\omega_{1m} \\ \sqrt{n_2n_1}\omega_{21} & n_2\omega_{22} & \cdots & \sqrt{n_2n_m}\omega_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{n_mn_1}\omega_{m1} & \sqrt{n_mn_2}\omega_{m2} & \cdots & n_m\omega_{mm} \end{pmatrix}$$

Since \mathbf{U} is orthonormal matrix, $\Sigma = \mathbf{U}(\mathbf{K} + \sigma^2\mathbf{I}_{n^*})\mathbf{U}'$. Therefore,

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

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

$$\mathbf{A} = \text{blockdiag} \left(\sqrt{1/(\omega_{11} + \sigma^2)}\mathbf{I}_{n_1}, \dots, \sqrt{1/(\omega_{mm} + \sigma^2)}\mathbf{I}_{n_m} \right),$$

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

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

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

BIBLIOGRAPHY

- [1] Andersen, P. K., and Gill, R. D. (1982). Cox's Regression Model for Counting Processes: A Large-Sample Study. *The Annals of Statistics*, 4, 1100–1120.
- [2] Aitkin, M. (1991). Posterior Bayes factors. *Journal of the Royal Statistical Society, Series B*, 53, 111–142.
- [3] Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical modeling and analysis for spatial Data*, Boca Raton, FL: Chapman & Hall/CRC Press.
- [4] 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.
- [5] Baker, R.D. (1991). Testing the efficiency of preventive maintenance. *Journal of the Operational Research Society*, 42, 493–503.
- [6] Berger, J. O. and Guglielmi, A. (2001). Bayesian and conditional frequentist testing of a parametric model versus nonparametric alternatives. *Journal of the American Statistical Association*, 96, 174–184.
- [7] 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.
- [8] Block, H. W., Borges, W. S., and Savits, T. H. (1985). Age-dependent minimal repair. *Journal of Applied Probability*, 22, 370–385.
- [9] Bové, D.S. and Held, L. (2011). Hyper- g priors for generalized linear models. *Bayesian Analysis*, 6, 1–24.
- [10] Buckley, J., and James, I. (1979). Linear regression with censored data. *Biometrika*, 66, 429–436.

- [11] 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.
- [12] Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B*, 187–220.
- [13] Cox, D. R. (1972b). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*, 34, 187–220.
- [14] Cox D.R, Snell E.J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society, B*, 248–275.
- [15] Cook, R. J., and Lawless, J. F. (2007). *The statistical analysis of recurrent events*. Springer.
- [16] Chen, Y. Q. and Wang, M. C. (2000). Analysis of accelerated hazards models. *Journal of the American Statistical Association*, 95, 608–618.
- [17] Chen, Y. Q. and Jewell, N. P. (2001). On a general class of semiparametric hazards regression models. *Biometrika*, 88, 687–702.
- [18] Chen, Y., Hanson, T., and Zhang, J. (2014). Accelerated hazards model based on parametric families generalized with Bernstein polynomials. *Biometrics*, in press.
- [19] Cooper, W. L., de Mello, and T.H., Kleywegt, A.J. (2006). Models for the spiral-down effect in revenue management. *Operations Research*, 54, 968–987.
- [20] Christensen, R. and Johnson, W. (1988). Modelling accelerated failure time with a Dirichlet process. *Biometrika* 75, 693–704.
- [21] Cressie, N. (1993). *Statistics for Spatial Data*, Wiley, New York.
- [22] De Boor, C. (2001). *A Practical guide to splines*. Applied Mathematical Sciences 27. Springer.
- [23] Dorado, C., Hollander, and M., Sethuraman, J. (1997). Nonparametric estimation for a general repair Model. *Annals of Statistics*, 25, 1140–1160.

- [24] Doyen, L. and Gaudoin, O. (2004). Classes of imperfect repair models based on reduction of failure intensity or virtual age. *Reliability Engineering and System Safety*, 84, 45–56.
- [25] 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.
- [26] Escobar, M. D., and West, M. (1995), Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90, 1301–1312.
- [27] 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.
- [28] Eilers, P. H., and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical science*, 89–102.
- [29] Fabius J. (1964), Asymptotic behavior of Bayes \check{S} estimates. *Annals of Mathematical Statistics*, 35, 846–856.
- [30] Ferguson, T.(1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, 1, 209–230.
- [31] Ferguson, T. (1974). Prior distribution on the spaces of probability measures. *Annals of Statistics*, 2, 615–629.
- [32] Ferguson, T. and Phadia. E. G (1979). Prior distribution on spaces of probability measures. *Annals of Statistics*, 2, 615–629.
- [33] Fouskakis, D., Ntzoufras, I. and Draper, D. (2009). Bayesian variable selection using cost-adjusted BIC, with application to cost-effective measurement of quality of health care. *Annals of Applied Statistics*, 3, 663–690.
- [34] Freedman, D. (1963). On the asymptotic distribution of Bayes' estimates in the discrete case. *Annals of Mathematical Statistics*, 34, 1386–1403.
- [35] Gelfand, A. E., and Dey, D. K. (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society, Series B*, 56, 501–514.

- [36] Geisser, S. and Eddy, W. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, 74, 153–160.
- [37] Geweke, J. (1991). Efficient simulation from the multivariate normal and Student-t distributions subject to linear constraints and the evaluation of constraint probabilities. In Keramidas, E. M., editor, *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, pp. 571–578, Fairfax Station, VA. Interface Foundation of North America, Inc.
- [38] 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.
- [39] Grambsch, P. and Therneau, T. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81, 515–526
- [40] Hanson, T. (2006). Inference for mixtures of finite Polya tree models. *Journal of the American Statistical Association*, 101, 1548–1565.
- [41] Hanson, T. and Johnson, W. (2002). Modeling regression error with a mixture of Polya trees. *Journal of the American Statistical Association*, 97, 1020–1033.
- [42] Hanson, T., and Yang, M. (2007). Bayesian semiparametric proportional odds models. *Biometrics*, 63, 88–95.
- [43] Hanson, T., Branscum, A., and Gardner, I. (2008). Multivariate mixtures of Polya trees for modelling ROC data. *Statistical Modelling*, 8, 81–96.
- [44] Haario, H., Saksman, E., and Tamminen, J. (2005). Componentwise adaptation for high dimensional MCMC. *Computational Statistics*, 20, 265–273.
- [45] Hanson, T., Kottas, A., and Branscum, A. (2008). Modelling stochastic order in the analysis of receiver operating characteristic data: Bayesian nonparametric approaches. *Journal of the Royal Statistical Society, Series C*, 57, 207–225.
- [46] Hennerfeind, A., Brezger, A., Fahrmeir, L. (2006). Geoaddivitive survival model. *Journal of the American Statistical Association*, 101, 1065–1075.
- [47] 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, April 2009.

- [48] Ibrahim, J. G., Chen, M. H., and Sinha, D. (2001). Bayesian survival analysis. John Wiley & Sons, Ltd.
- [49] Ishwaran, H., and James. L. F. (2004), Computational methods for multiplicative intensity models using weighted gamma processes. *Journal of the American Statistical Association*, 99, 175–190.
- [50] Jara, A. and Hanson, T. (2011). A class of mixtures of dependent tailfree processes. *Biometrika*, 98, 553–566.
- [51] Jin, Z., Lin, D. Y., Wei, L. J., and Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika*, 90(2), 341–353.
- [52] Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 773–795.
- [53] Kammann, E. E., and Wand, M. P. (2003). Geoadditive Models. *Journal of the Royal Statistical Society, Ser. C*, 52, 1–18.
- [54] Kijima, M. (1989). Some results for repairable systems with general repair. *Journal of Applied Probability*, 26, 89–102.
- [55] Kalbfleisch, J. D. and Prentice, R. L. (2002). The statistical analysis of failure time data. Hoboken, NJ: Wiley.
- [56] Kalbfleisch, J. D. (1978). Non-parametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society. Series B*, 214–221.
- [57] Kuo, L., and Ghosh, S. K. (1997). Bayesian nonparametric inference for nonhomogeneous Poisson processes. Technical Report 9718, University of Connecticut, Department of Statistics.
- [58] Kneib, T., and Fahrmeir, L. (2006). Structured additive regression for categorical space-time Data: a mixed model approach. *Biometrics*, 62, 109–118.
- [59] Kneib, T. and Fahrmeir, L. (2007). A mixed model approach for geoadditive hazard regression. *Scandinavian Journal of Statistics*, 34, 207–228.

- [60] Krivtsov V. (2000). A Monte Carlo approach to modeling and estimation of the generalized renewal process in repairable system reliability analysis. Dissertation for the degree of Doctor of Philosophy, University of Maryland.
- [61] 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.
- [62] Lang, S., and Brezger, A. (2004). Bayesian P-splines. *Journal of computational and graphical statistics*, 13, 183–212.
- [63] Lavine, M. (1992). Some aspects of Polya tree distributions for statistical modeling. *Annals of Statistics*, 20, 1222–1235.
- [64] Lawless, J. F. and Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics*, 37, 158–168.
- [65] Lawless, J. F. and Thiagarajah, K. (1996). A point process model incorporating renewals and time trends, with application to repairable systems. *Technometrics*, 38, 131–138.
- [66] Lawless, J. F. (2003a). *Statistical models and methods for lifetime data*, 2nd edition. Wiley.
- [67] Li, Y., and Lin, X. (2006). Semiparametric normal transformation models for spatially correlated survival data. *Journal of the American Statistical Association*, 101, 591–603.
- [68] Li, Y. and Ryan, L. (2002). Modeling spatial survival data using semi-parametric frailty models. *Biometrics* 58, 287–297.
- [69] Li, P. and Rahman, M.A. (2011). Bayesian analysis of multivariate sample selection models using Gaussian copulas. In *Missing Data Methods: Cross-sectional Methods and Applications (Advances in Econometrics, Volume 27)*, pp. 269–288. D.M. Drukker, ed. Emerald Group Publishing Limited.
- [70] Lindqvist, B. H. (2006). On the statistical modeling and analysis of repairable systems. *Statistical Science*, 21, 532–551.
- [71] 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.

- [72] Lo, A. Y. (1992). Bayesian inference for Poisson process models with censored data. *Journal of Nonparametric Statistics*, 2, 71–80.
- [73] Marin, J-M. and Robert, C.P. (2007). *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*. Springer-Verlag, New York.
- [74] Mauldin, R. D., Sudderth, W. D. and Williams, S. C. (1992). Polya trees and random distributions. *Annals of Statistics*, 20, 1203–1221.
- [75] Martino, S., Akerkar, R., and Rue, H. (2011). Approximate Bayesian inference for survival models. *Scandinavian Journal of Statistics*, 38, 514–528.
- [76] Marsden, M. J. (1972). On uniform spline approximation. *Journal of Approximation Theory*, 6, 249–253.
- [77] Nelson, W. (1995). Confidence limits for recurrence data applied to cost or number of product repairs. *Technometrics*, 37, 147–157.
- [78] Peña, E., Slate, E. and González, J. (2007). Semiparametric inference for a general class of models for recurrent events. *Journal of Statistical Planning and Inference*, 137, 1727–1747.
- [79] Presnell, B., Hollander, M., Sethuraman, J. (1994). Testing the minimal repair assumption in an imperfect repair model. *Journal of the American Statistical Association*, 89, 289–297.
- [80] Raftery, A. E. (1996). Hypothesis testing and model selection via posterior simulation. *Markov chain Monte Carlo in practice*, 163–188.
- [81] Rathbun, S.L. and Fei, S. (2006). A spatial zero-inflated poisson regression model for oak regeneration. *Environmental and Ecological Statistics*, 13,409–426.
- [82] Reid, N. (1994). A conversation with Sir David Cox. *Statistical Science*, 9, 439–455.
- [83] Ryan, T. and Woodall, W. (2005). The most-cited statistical papers. *Journal of Applied Statistics*, 32, 461–474.
- [84] Ruppert D (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics*, 11, 735–757.

- [85] Scott, D.W. (1985). Averaged shifted histograms: effective nonparametric density estimators in several dimensions. *Annals of Statistics*, 13, 1024–1040.
- [86] Singh, I. (2011). Risk-averse periodic preventive maintenance optimization. Doctoral dissertation, The University of Texas at Austin.
- [87] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, 64, 583–639.
- [88] 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.
- [89] Smith, M.S. (2013). Bayesian approaches to copula modeling. In *Bayesian Theory and Applications*. P. Damien, P. Dellaportas, N. Polson, and D. Stephens, eds. Oxford University Press: Oxford.
- [90] Scharfstein, D.O., Tsiatis, A.A., and Gilbert, P.B. (1998). Semiparametric efficient estimation in the generalized odds-rate class of regression models for right-censored time-to-event data. *Lifetime Data Analysis*, 4, 355–391.
- [91] Taddy, M. and Kottas, A. (2012). Mixture modeling for marked Poisson processes. *Bayesian Analysis*, 7, 335–362.
- [92] Tsiatis, A. A. (1990). Estimating Regression Parameters Using Linear Rank Tests for Censored Data. *The Annals of Statistics*, 18, 354–372.
- [93] Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics*, 22, 1701–1728, with discussion 1728–1762.
- [94] Tseng, Y. K., Hsieh, F., and Wang, J. L. (2005). Joint modelling of accelerated failure time and longitudinal data. *Biometrika*, 92, 587–603.
- [95] 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.
- [96] Tong, X., Zhu, L., Leng, C., Leisenring, W., and Robison, L.L. (2013). A general semiparametric hazards regression model: efficient estimation and structure selection. *Statistics in Medicine*, in press.

- [97] Veber B., Nagode M., Fajdiga M. (2008). Generalized renewal process for repairable systems based on finite Weibull mixture. *Reliability Engineering and System Safety*, 93, 1461–1472.
- [98] 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.
- [99] Fuqing, Y., Kumar, U. (2012). A general imperfect repair model considering time-dependent repair effectiveness. *Reliability, IEEE Transactions*, 61, 95–100.
- [100] Yin, G. and Ibrahim, J.G. (2005). Bayesian frailty models based on Box-Cox transformed hazards. *Statistica Sinica* 15, 781–794.
- [101] Walker, S., and Mallick, B. K. (1999). A Bayesian semiparametric accelerated failure time model. *Biometrics*, 55, 477–483.
- [102] Walker, S. G., Damien, P., Laud, P. W., and Smith, A. F. (1999). Bayesian nonparametric inference for random distributions and related functions. *Journal of the Royal Statistical Society: Series B*, 61, 485–527.
- [103] Zeng, D., and Lin, D. Y. (2007). Efficient estimation for the accelerated failure time model. *Journal of the American Statistical Association*, 102, 1387–1396.
- [104] Zellner, A. and Siow, A. (1980). Posterior odds ratios for selected regression hypotheses. In Bernardo, J. M., DeGroot, M. H., Lindley, D. V., and Smith, A. F. M. (eds.). *Bayesian Statistics: Proceedings of the First International Meeting Held in Valencia*, 585–603.
- [105] Zellner, A. (1983). Applications of Bayesian analysis in econometrics. *The Statistician*, 32, 23–34.
- [106] Zhao, L., Hanson, T., and Carlin, B. (2009). Flexible spatial frailty modeling via mixtures of Polya trees. *Biometrika*, 96, 263–276.
- [107] Whitaker, L. R. and Samaniego, F. J. (1989). Estimating the reliability of systems subject to imperfect repair. *Journal of the American Statistical Association*, 84, 301–309.

APPENDIX A

COPYRIGHT PERMISSION TO REPRINT CHAPTER 2 & 3

A.1 CHAPTER 2

Dear Ms. Li:

We are in receipt of your request to reproduce your article “Li Li, Timothy Hanson, Paul Damien & Elmira Popova (2013) A Bayesian Nonparametric Test for Minimal Repair Technometrics Accepted Manuscript” for use in your dissertation. T & F reference number is P040114 – 01.

This permission is all for editions, both print and electronic. We will be pleased to grant you permission free of charge on the condition that:

This permission is limited to non-exclusive English rights for this usage only.

This permission does not cover any third party copyrighted work which may appear in the material requested.

Full acknowledgment must be included showing article title, author, and full Journal title, reprinted by permission of Taylor & Francis LLC (<http://www.tandfonline.com>)

Thank you very much for your interest in Taylor & Francis publications. Should you have any questions or require further assistance, please feel free to contact me directly.

Sincerely,

Mary Ann Muller

Permission Coordinator

Telephone: 215.606.4334

Email: maryann.muller@taylorandfrancis.com

A.2 CHAPTER 3

Elsevier has partnered with RightsLink to license its content. This notice is a confirmation that your order was successful. Your order details and publisher terms and conditions are available by clicking the link: <http://s100.copyright.com/CustomerAdmin/PLF.jsp?ref=981701d9-a0dd-4176-aa0e-91044dd823bb>.

Order Details:

Licensee: LI LI

License Date: Mar 27, 2014

License Number: 3357190107702

Publication: Computational Statistics & Data Analysis

Title: A Bayesian semiparametric regression model for reliability data using effective age

Type Of Use: reuse in a thesis/dissertation