Summer 2020

Semiparametric Regression Analysis of Survival Data and Panel Count Data

Lu Wang

Semiparametric Regression Analysis of Survival Data and Panel Count Data

by

Lu Wang

Bachelor of Medicine
Sichuan University 2013

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

2020

Accepted by:

Lianming Wang, Major Professor

Edsel Peña, Committee Member

Xiaoyan Lin, Committee Member

Bo Cai, Committee Member

Cherly Addy, Vice Provost and Dean of the Graduate School

Acknowledgments

Foremost, I would like to express my sincere gratitude to my advisor Dr. Lianming Wang for the continuous support of my Ph.D study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study.

Besides my advisor, I would like to thank the rest of my thesis committee: Dr. Edsel Peña, Dr. Xiaoyan Lin, and Dr. Bo Cai, for their encouragement, insightful comments, and inspirational questions. I am grateful for their patience and support in overcoming numerous obstacles I have been facing through my research.

My sincere thanks also goes to Chunling Wang for the stimulating discussions and all the nice time that we spent on the project together.

I am also grateful to my friends in Dr.Peña's research group: Shiwen Shen, Piaomu Liu, Taeho Kim, Beidi Qiang, Jeff Thompson, Tahmidul Islam and Lili Tong. I thank them for their feedback, cooperation and of course friendship.

Last but not the least, I would like to thank my family: my husband and my parents for supporting me spiritually throughout writing this thesis and my life in general.

Abstract

Both censored survival data and panel count data arise commonly in real-life studies in many fields such as epidemiology, social science, and medical research. In these studies, subjects are usually examined multiple times at periodical or irregular follow-up examinations. Censored data are studied when the exact failure times of the events are of interest but not all of these exact times are directly observed. Some of the failure times of event of interest are only known to fall within some intervals formed by the observation times. Panel count data are under investigation when the exact times of the recurrent events are not of interest but the counts of the recurrent events of interest occurring within the time intervals are available and of interest. This dissertation devotes to discussing three semiparametric regression models that can be used to analyze censored survival data and panel count data.

Chapter 1 of this dissertation proposes an estimation approach for regression analysis of arbitrarily censored survival data under the proportional odds model. Arbitrarily censored data contains a mixture of exactly observed, left-censored, interval-censored, and right-censored observations. Existing research work on regression analysis on arbitrarily censored data is sparse and limited to the proportional hazards model only. In this chapter, a novel estimation approach based on an EM algorithm is proposed for analyzing arbitrarily censored data under the proportional odds model. The proposed EM algorithm is robust to initial values, easy to implement, converging fast, and providing the variance estimate of the regression parameter estimate in closed form. This method has shown excellent performance in estimating the regres-

sion parameters as well as the baseline survival function in an extensive simulation study. Several real-life data applications are provided for illustration purpose.

In Chapter 2, a novel Bayesian approach is proposed to analyze panel count data. The widely used gamma frailty Poisson process model has been shown to have good estimation performance and some robustness against misspecification of the frailty distribution but may still produce biased estimation in some cases when the gamma frailty assumption is violated. In this chapter, we tackle the problem by modeling the frailty distribution nonparametrically by adopting a Dirichlet Process Gamma Mixture (DPGM) prior for the frailty distribution. An easy-to-implement Gibbs sampler is developed to facilitate the Bayesian computation. The proposed Bayesian approach has an excellent performance in estimating the regression parameters and the baseline mean function in our simulation. It outperforms the gamma frailty Poisson model when the gamma frailty distribution is misspecified. The proposed method is applied to the famous bladder cancer data for illustration and comparison with existing methods.

In Chapter 3, a novel unified Bayesian approach is developed for analyzing panel count data under the Gamma frailty Poisson process mode and interval-censored data under Cox's proportional hazards model and the proportional odds model. The baseline functions in these models share the same property of being nondecreasing positive functions and are modeled nonparametrically by assigning a Gamma process prior. Efficient and easy-to-implement Gibbs samplers are developed for the posterior computation under these three models for the two types of data. The proposed methods are evaluated in extensive simulation studies and illustrated by real-life data applications.

Table of Contents

List of Tables

List of Figures

Chapter 1

Regression analysis of arbitrarily censored data under the proportional odds models

1.1 Introduction

The analysis of survival data plays an indispensable role in a lot of areas, such as epidemiology, biomedical science, engineering and sociology. Nowadays, there are more and more instances where complex observation schemes have to be treated. Interval-censored data occurs naturally in clinical trials and epidemiology studies in which patients visit the clinic periodically and the event of interest is assessed on repeated visits. It also appears in retrospective cohort study when some of the event time are exactly observed but the others are only known to lie in certain time intervals. Current status data can be encountered when study the onset of tumor on mice. Since researchers can only examine whether a tumor has developed after a rat is sacrificed so the time is either right- or left-censored. Also, prevalent cases of a disease can be viewed as left-censored observations, etc. The diverse situation leads to a demand for versatile approaches which can accommodate arbitrarily censored data produced by complex observation schemes.

A series of nonparametric maximum likelihood estimators (NPMLE) with no covariates have been developed for general censoring data. Peto [1973] proposed a nonparametric maximum likelihood estimator (NPMLE) of the survival function for interval-censored data. Turnbull [1974] extended it to arbitrarily censored data and further extended it to fit arbitrarily censored and truncated data in 1976. Pan and

Chappell [1998] suggested an iterative Nelson estimator (INE) to estimate the survival function nonparametrically. Other methods includes Wang et al. [1986] and Uzunog-Ullari and Wang [1992],among others. Meanwhile, since regression analysis of censored time-to-event data is of central interest in health sciences research, some widely used approaches have been developed based on semiparametric models. For the most widely used proportional hazard regression model, Finkelstein [1986] developed a maximum likelihood estimator for arbitrarily censored data. Tu et al. [1993] described a general discrete-time proportional hazards model and fitted it with EM algorithm. Alioum and Commenges [1996] extended Turnbull [1976]'s method to the proportional hazards model for arbitrarily censored and truncated data in continuous time.

Compared to the large amount of methods established for the PH model, the methodology development with respect to the proportional odds (PO) model is very limited. PO model performs as an alternative to PH model on analyzing time to event data. It specifies the log ratio of odds of survival given covariates to the baseline odds as a parametric regression function of covariates. The associated baseline odds function is left unspecified. Regression parameters in PO model is more interpretable than PH model in terms of odds ratio. Different from proportional hazard model, PO model constrains the ratio of the hazards converges to unity as time increases, so Bennett [1983a] and Murphy et al. [1997] suggest that PO model is more appropriate for demonstrating an effective cure or the case that the morbidity rates converge with time. Despite its pleasing interpretation, the PO model is rarely used, likely due to difficulty of implementation. Efforts have been done on addressing simpler instances, say right censored data. For example, reasonable estimations for the regression coefficients have been proposed by Bennett [1983a], Murphy et al. [1997], Yang and Prentice [1999] and Royston and Parmar [2002], among others. Because the complexity of data structure adds more complexity to the PO model there are only a handful of studies

on fitting PO model with interval-censored data. Rossini and Tsiatis [1996] adapted the semi-parametric framework for modeling current status data by approximating the infinite-dimensional nuisance parameter, the baseline log-odds of failure, with a step function, and carried out a maximum likelihood procedure. Huang and Rossini [1997] proposed a sieve maximum likelihood estimator for proportional odds model with interval censored data. Shen [1998] use monotone splines of variable orders and knots for approximating the odds of failure time and proposed a sieve maximum likelihood estimator for right-censored and case 2 interval-censored data. Lin and Wang [2011] proposed a Bayesian approach for analyzing case 2 interval-censored data under the semiparametric proportional odds model. This situation requires the analyst to seek specialized, distinct techniques according to different censoring patterns.

The aim of this paper is to propose an easy to implemented approach that can fit the proportional odds model for arbitrarily censored data in continuous time. The expectation-maximization (EM) algorithm we propose is so flexible that it can handle any combination of incomplete data. Specifically, it can successfully fit randomly right-censored data, left-censored data, current status data, case 2 interval-censored data, or a mixture of them. To our knowledge, icenReg[Anderson-Bergman, 2017] is the only approach available for analyzing arbitrarily censored data under the PO model by now. This method is a combination of conditional Newton-Raphson, ICM algorithm [Pan, 1999] and constrained gradient ascent algorithm. It is an efficient algorithm but it does not offer closed-form of standard errors. Inference on the regression parameters needs to be done using bootstrap standard errors.
Usually composing 100$\sim$ 1000 bootstrap samples will significantly prolong the analytic time span. Our method gets rid of this problem by providing a closed-form expressions of the asymptotic variance estimates. We will compare our approach with **icenReg** later in both simulation study and real data analysis.

In Section 2, we provide the methodological details of the proposed method. These details include the use of monotone splines for approximating the baseline odds function in the PO model, a four-stage data augmentation process that leads to the development of an EM algorithm that can be used to find the maximum likelihood estimates of all unknown parameters, and closed-form expressions of the asymptotic variance estimates. In Section 3, the performance of the proposed approach is evaluated in simulated data against competing package *Icenreg*. In Section 4 the proposed approach is applied to three real datasets. Section 5 provides a summary discussion and future plans.

1.2 The Method

1.2.1 Proportional Odds Model

Let T_i denote the survival time of interest and $\mathbf{x}_i$ a $p \times 1$ vector of potential covariates for subject i, for $i = 1, \cdots, n$. In this article, we take a general notation $[L_i, R_i]$ to denote the observed interval for the failure time T_i, with $0 \leq L_i \leq R_i \leq \infty$. This general interval yields an exactly observed failure time when $0 < L_i = R_i < \infty$, a left-censored observation when $0 = L_i < R_i < \infty$, a strictly interval-censored observation when $0 < L_i < R_i < \infty$, and a right-censored observation when $0 < L_i < R_i = \infty$. It is assumed that the failure time is conditionally independent with the observational process (i.e., the set of examination times) given covariates. Under this non-informative censoring assumption, the observed likelihood takes the following form

$$L_{obs} = \prod_{i=1}^{n} f(R_i|\mathbf{x}_i)^{\delta_{i0}} \{1 - S(R_i|\mathbf{x}_i)\}^{\delta_{i1}} \{S(L_i|\mathbf{x}_i) - S(R_i|\mathbf{x}_i)\}^{\delta_{i2}} \{S(L_i|\mathbf{x}_i)\}^{\delta_{i3}}, \quad (1.1)$$

where $f(t|\mathbf{x})$ and $S(t|\mathbf{x})$ are the density and survival functions respectively given covariate $\mathbf{x}$, and δ_{i0}, δ_{i1}, δ_{i2}, and δ_{i3} are all binary censoring indicators for exactly ob-

served, left-censored, interval-censored, and right-censored observations, respectively, with the constraint $\delta_{i0} + \delta_{i1} + \delta_{i2} + \delta_{i3} = 1$ for subject i.

Under the PO model, the survival function is specified as $S(t|\mathbf{x}) = \{1 + \Lambda_0(t)\exp(\mathbf{x}'\boldsymbol{\beta})\}^{-1}$ and the density function takes the form

$$f(t|\mathbf{x}) = \frac{\Lambda_0'(t)\exp(\mathbf{x}_i'\boldsymbol{\beta})}{\{\Lambda_0(t)\exp(\mathbf{x}_i'\boldsymbol{\beta}) + 1\}^2},$$

where $\Lambda_0(t) = F_0(t)/\{1 - F_0(t)\}$ is the baseline odds function, $F_0(t)$ is the baseline cumulative distribution function when $\mathbf{x} = 0$, and $\Lambda_0'(t)$ is the first derivative of $\Lambda_0(t)$. Note that the baseline odds function $\Lambda_0(t)$ is an usepecified non-negative and non-decreasing function, and $\Lambda_0'(t)$ is a non-negative function under the PO model. Thus, the unknown parameters in the observed likelihood (1.1) include the regression parameters $\boldsymbol{\beta}$, the baseline odds function $\Lambda_0(t)$, and its derivative $\Lambda_0'(t)$.

1.2.2 Monotone Splines

The infinite dimension in the baseline odds function $\Lambda_0(\cdot)$ and its derivative $\Lambda_0'(\cdot)$ causes great trouble from both theoretic and computational perspectives. To reduce the number of unknown parameters while still allowing adequate modeling flexibility, we adopt the monotone splines of Ramsay [1988] for modeling $\Lambda(\cdot)$ as follows,

$$\Lambda_0(t) = \sum_{l=1}^{K} \gamma_l b_l(t), \tag{1.2}$$

where $b_l(\cdot)$'s are integrated spline (or I-spline) basis functions and γ_l's are non-negative spline coefficients to ensure the monotonicity of the Λ_0. Each of the I-spline basis function is a piecewise polynomial of specified degree $d - 1$ or order d, taking 0 in an initial flat region, increasing in a mid region, and remaining at at 1 in the third region[Wang and Dunson, 2011]. The same or similar strategy has been effectively used to model unknown non-decreasing functions such as the transformed baseline cumulative distribution function in the probit model [Lin and Wang, 2010], the log-

arithm of the baseline odds in the PO model [Wang and Dunson, 2011], and the cumulative hazard function in the PH model [McMahan et al., 2013] among others.

Another benefit of using the I-splines is that it allows us to model $\Lambda_0'(\cdot)$ directly without the need of introducing any additional parameters. That is,

$$\Lambda_0'(t) = \sum_{l=1}^{K} \gamma_l M_l(t),$$

where $M_l(\cdot)$'s are the so-called M-splines, the derivatives of the I-splines. All these basis functions are determined once the degree and knots are specified and can be obtained using iterative algorithms in our R functions. Note that these functions are calculated just once and do not need to be recalculated during the estimation process.

In general the degree determines the smoothness of the monotone splines, and together with the degree the placement of the knots determines the shape of the splines. Setting degree as 2 or 3 typically provides adequate smoothness. As for the placement of knots, it is reported that using $10 \sim 30$ equally-spaced knots provides adequate modeling flexibility for data sets containing up to thousands of observations (Cai et al. [2011], Wang and Dunson [2011]). Lin and Wang [2011] and Lin et al. [2015] showed that for general interval-censored data, adopting equally-spaced knots in their methods outperforms the strategy of using quantile-based knots in terms of two commonly used Bayesian model selection criteria: the deviance information criteria (DIC) and the logarithm of psedu-marginal likelihood (LPML). It is worthy noting that those Bayesian methods employ shrinkage priors for the spline coefficients and thus allow to use a large number of knots without causing over-fitting problems (Cai et al. [2011], Wang and Dunson [2011], Lin and Wang [2011]). From a frequentist perspective, we recommend to follow the idea of Rosenberg [1995], McMahan et al. [2013], and Wang et al. [2016] to determine the number of knots. That is, we will fit the PO model using our method with different values for the number of knots and then choose the number that leads to the smallest value of Akaike's information criterion (AIC) or Bayesian information criteria (BIC).

1.2.3 A four-stage data augmentation

Direct optimization of the observed likelihood (1) encountered many numerical problems such as non-convergence from our experiences even though the number of unknown parameters is finite with the use of monotone splines. The main reason is that the optimization is very sensitive to initial values of the spline coefficients in addition to the complexity of the observed likelihood. To overcome such difficulties, we seek to explore an EM algorithm to obtain the MLE. To this end, we first introduce a four-stage data augmentation that leads to a complete data likelihood that has a nice form for our EM algorithm. The details of the four-stage data augmentation are given below.

The first-stage augmentation takes advantage of the relationship between the proportional odds model and the frailty proportional hazards model (Shen [1998], Murphy et al. [1997], McMahan et al. [2013]). Specifically, one can write the survival function of the proportional odds model as the marginal survival function in the frailty proportional hazards model with the frailty following an exponential distribution with mean 1 in the following manner,

$$S(t|\mathbf{x}) = \{\Lambda_0(t)\exp(\mathbf{x}'\boldsymbol{\beta}) + 1\}^{-1} = \int_0^\infty \exp\{-\Lambda_0(t)\exp(\mathbf{x}'\boldsymbol{\beta})\phi\}\exp(-\phi)d\phi. \quad (1.3)$$

Based on this, we introduce independent latent variables $\phi_i \sim \mathcal{E}xp(1)$ for all subjects and $\psi_i \sim \mathcal{E}xp(1)$ only for those exactly observed subjects (i.e, $\delta_{i0} = 1$).
Conditioning on these latent variables, the augmented data likelihood takes the fol-

lowing form

$$\begin{aligned}\mathcal{L}_1 = \prod_{i=1}^{n} &\{\Lambda_0'(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})\}^{\delta_{i0}}\exp\{-\Lambda_0(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})(\phi_i+\psi_i)\}^{\delta_{i0}} \\ &\times [1-\exp\{-\Lambda_0(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i\}]^{\delta_{i1}} \\ &\times [\exp\{-\Lambda_0(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i\}-\exp\{-\Lambda_0(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i\}]^{\delta_{i2}} \\ &\times [\exp\{-\Lambda_0(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i\}]^{\delta_{i3}} \\ &\times \exp(-\phi_i)\exp(-\psi_i\delta_{i0}).\end{aligned} \tag{1.4}$$

In the augmented likelihood $\mathcal{L}_1$, the middle three multiplicative terms essentially form the likelihood for interval-censored data under the PH model (Lin et al. [2015], Wang et al. [2016]). Our second-stage data augmentation generalizes the ideas in Lin et al. [2015] and Wang et al. [2016] with additional frailties. In order to make this part self-complete, we provide the following motivations and justifications.

Let $N_i(t)$ denote a latent non-homogeneous Poisson process with cumulative intensity function $\Lambda_0(t)\exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i$ conditioning on unobserved frailty ϕ_i for subject i with $\delta_{i0}=0$ (i.e., not exactly observed). Define two time points t_{i1} and t_{i2} based on the observed data for subject i as follows: $t_{i1}=R_iI(\delta_{i1}=1)+L_iI(\delta_{i2}=1)$ and $t_{i2}=R_iI(\delta_{i2}=1)+L_iI(\delta_{i3}=1)$. Then define $Z_i=N(t_{i1})$ and $W_i=N(t_{i2})-N(t_{i1})$ for subject i with $\delta_{i0}=0$ (i.e., not exactly observed). Based on the independent increment property of the Poisson process, Z_i and W_i are conditionally independent Poisson random variables given frailty ϕ_i, $Z_i\sim\text{Poisson}(\Lambda_0(t_{i1})\exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i)$, and $W_i\sim\text{Poisson}(\{\Lambda_0(t_{i2})-\Lambda_0(t_{i1})\}\exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i)$ for subject i with $\delta_{i0}=0$. With the introduced latent variables Z_i's and W_i's, the augmented data likelihood in this stage is

$$\begin{aligned}\mathcal{L}_2 = \prod_{i=1}^{n} &\{\Lambda_0'(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})\}^{\delta_{i0}}\exp\{-\Lambda_0(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})(\phi_i+\psi_i)\delta_{i0}\} \\ &\times \mathcal{P}(Z_i)^{1-\delta_0}\mathcal{P}(W_i)^{\delta_{i2}+\delta_{i3}}\exp(-\psi_i\delta_{i0})\exp(-\phi_i),\end{aligned} \tag{1.5}$$

where $\mathcal{P}(\cdot)$ denote the Poisson probability mass function. In this augmented likelihood, the latent variables Z_i's and W_i's are subject to the following constraints:

$Z_i > 0$ if $\delta_{i1} = 1$; $W_i > 0$ and $Z_i = 0$ if $\delta_{i2} = 1$; and $W_i = Z_i = 0$ if $\delta_{i3} = 1$. Integrating out the augmented likelihood with respect to Z_i's and W_i's leads to the second augmented likelihood function in equation (1.4).

To fully take advantage of the additive form of the monotone spline representation (1.2), we decompose both Z_i and W_i as a sum of K conditionally independent Poisson random variables given ϕ_i as $Z_i = \sum_{l=1}^{K} Z_{il}$ and $W_i = \sum_{l=1}^{K} W_{il}$, where

$$Z_{il} \sim \text{Poisson}(\gamma_l b_l(t_{i1}) \exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i),$$

$$W_{il} \sim \text{Poisson}(\{b_l(t_{i2}) - b_l(t_{i1})\}\gamma_l \exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i),$$

for each i with $\delta_{i0} = 0$. Treating all Z_{il}'s and W_{il}'s as missing data, the augmented likelihood in this stage takes the following form

$$\mathcal{L}_3 = \prod_{i=1}^{n} \{\Lambda_0'(L_i) \exp(\mathbf{x}_i'\boldsymbol{\beta})\}^{\delta_{i0}} \exp\{-\Lambda_0(L_i) \exp(\mathbf{x}_i'\boldsymbol{\beta})(\phi_i + \psi_i)\delta_{i0}\} \exp(-\psi_i\delta_{i0}) \exp(-\phi_i) \times \prod_{l=1}^{K} \mathcal{P}(Z_{il})^{1-\delta_{i0}} \mathcal{P}(W_{il})^{\delta_{i2}+\delta_{i3}}, \tag{1.6}$$

where the latent variables are subject to the following constraints: $\sum_{l=1}^{K} Z_{il} > 0$ when $\delta_{i1} = 1$; $\sum_{l=1}^{K} W_{il} > 0$ and $Z_{il} = 0$ for $l = 1, \cdots, K$ when $\delta_{i2} = 1$; $Z_{il} = 0$ and $W_{il} = 0$ for $l = 1, \cdots, K$ when $\delta_{i3} = 1$.

Notice that the first multiplicative term in the augmented likelihood (1.6) involves the summation of M-splines for exactly observed observations. In order to facilitate the computation and get rid of the summation, we introduce a multinomial latent vector $(U_{i1}, ..., U_{iK}) \sim \text{Multinomial}\{1, (1/K, ..., 1/K)\}$ in the forth stage of our augmentation for subject i with $\delta_{i0} = 1$. It is clear that summing $\prod_{l=1}^{K}[\gamma_l M_l(t)]^{u_l}$ over all possible combinations of $(u_1, ..., u_k)$ leads to $\sum_{l=1}^{K} \gamma_l M_l(t)$. With these new latent variables, the augmented data likelihood function now has the following multiplicative

form:

$$\mathcal{L}_c = \prod_{i=1}^{n} \exp(\mathbf{x}_i'\boldsymbol{\beta}\delta_{i0}) \exp\{-\sum_{l=1}^{K} \gamma_l b_l(t_{i1}) \exp(\mathbf{x}_i'\boldsymbol{\beta})(\phi_i + \psi_i)\delta_{i0}\} \exp(-\psi_i\delta_{i0}) \exp(-\phi_i)$$
$$\times \prod_{l=1}^{K} \{\gamma_l M_l(R_i)\}^{U_{il}\delta_{i0}} P(Z_{il})^{1-\delta_{i0}} P(W_{il})^{\delta_{i2}+\delta_{i3}},$$

subject to the same constrains for the augmented likelihood $\mathcal{L}_3$. The augmented data likelihood (1.7) is extremely appealing because it only contains multiplicative terms of simple functions and will be viewed as the complete data likelihood for the derivation of our EM algorithm below.

1.2.4 EM Algorithm

Viewing the observed data as incomplete data, EM algorithm computes the maximum-likelihood estimates iteratively through an expectation step followed by a maximization step. In E-step, the expectation of logarithm of the complete data likelihood with respect to the latent variables condition on the observed data $\mathcal{D}$ and the current parameter estimate $\boldsymbol{\theta}^{(d)} = (\boldsymbol{\beta}^{(d)'}, \boldsymbol{\gamma}^{(d)'})'$ is obtained and denoted as $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)}) = E[\log\{\mathcal{L}_c(\boldsymbol{\theta})\}|\mathcal{D}, \boldsymbol{\theta}^{(d)}]$. To provide a concise expression, we omit $\mathcal{D}$ and $\boldsymbol{\theta}^{(d)}$ in all the conditional expectation in the rest of the paper. Benefit to the well-designed data augmentation, $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ has the following simple addition terms,

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)}) = \sum_{i=1}^{n} \Bigg\{ & \mathbf{x}_i'\boldsymbol{\beta}\delta_{i0} - E(\psi_i)\delta_{i0} - E(\phi_i) \\ & + \sum_{l=1}^{K} \Bigg[\{E(U_{il})\delta_{i0} + E(Z_{il})\delta_{i1} + E(W_{il})\delta_{i2}\} \log(\gamma_l) \\ & + \{E(\psi_i)\delta_{i0} b_l(L_i) + E(\phi_i)(b_l(L_i)(\delta_{i0} + \delta_{i3}) + b_l(R_i)(\delta_{i1} + \delta_{i2}))\} e^{\mathbf{x}_i'\boldsymbol{\beta}} \gamma_l \\ & + \ \{E(Z_{il})\delta_{i1} + E(W_{il})\delta_{i2}\}\mathbf{x}_i'\boldsymbol{\beta} \Bigg] \Bigg\} + g(\boldsymbol{\theta}^{(d)}). \end{aligned}$$

In addition, these conditional expectations in $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ all have explicit forms as shown in appendix A. The M-step in the EM algorithm devotes to find $\boldsymbol{\theta}^{(d+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$. Taking the partial derivatives of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ with respect to $\boldsymbol{\theta}$

leads to

$$\frac{\partial Q}{\partial \gamma_l} = \sum_{i=1}^{n} \Big(\{E(U_{il})\delta_{i0} + E(Z_{il})\delta_{i1} + E(W_{il})\delta_{i2}\} \frac{1}{\gamma_l} - \Big[E(\psi_i)\delta_{i0} b_l(L_i) + E(\phi_i)\{b_l(L_i)(\delta_{i0} + \delta_{i3}) + b_l(R_i)(\delta_{i1} + \delta_{i2})\} \Big] \exp(\mathbf{x}_i'\boldsymbol{\beta}) \Big)$$

$$\frac{\partial Q}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \Big(\delta_{i0} + E(Z_i)\delta_{i1} + E(W_i)\delta_{i2} - \sum_{l=1}^{K} \gamma_l \exp(\mathbf{x}_i'\boldsymbol{\beta}) \Big[E(\psi_i)\delta_{i0} b_l(L_i) + E(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)})\{b_l(L_i)(\delta_{i0} + \delta_{i3}) + b_l(R_i)(\delta_{i1} + \delta_{i2})\} \Big] \Big) \mathbf{x}_i.$$

Solving the system of equations: $\partial Q/\partial \boldsymbol{\beta} = 0$ and $\partial Q/\partial \gamma_l = 0$ for $l = 1, ..., K$ returns the value of $\boldsymbol{\theta}^{(d+1)}$. Specifically, solving $\partial Q/\partial \gamma_l = 0$ for γ_l offers a closed-form expression for $\gamma_l^{(d+1)}$ in terms of $\boldsymbol{\beta}^{(d+1)}$ for each l. Thus, by plug in this expression of γ_l in $\partial Q/\partial \boldsymbol{\beta} = 0$ one can directly obtain $\boldsymbol{\beta}^{(d+1)}$, which then allows for the direct calculation of $\gamma_l^{(d+1)}$. Moreover, it can be shown that $\boldsymbol{\theta}_l^{(d+1)}$ is the unique global maximizer of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ and $\hat{\boldsymbol{\theta}}$ solves the score equations based on the observed likelihood. The proof is shown in web Appendix A.

Although the derivation seems tricky, the EM algorithm that is actually used to fit the model turns out to be quite succinct and easy to operate. The whole process can be summarized as follows. First set $d = 0$ and initialize $\boldsymbol{\theta}^{(d)} = \boldsymbol{\theta}(\boldsymbol{\beta}^{(d)}, \gamma^{(d)})$. Then repeat the following two steps until convergence:

1. Obtain $\boldsymbol{\beta}^{(d+1)}$ by solving the following system of p equations

$$\sum_{i=1}^{n} \{\delta_{i0} + E(Z_i)\delta_{i1} + E(W_i)\delta_{i2}\}\mathbf{x}_i = \sum_{i=1}^{n} \sum_{l=1}^{K} \gamma_l^* \Big[E(\psi_i)\delta_{i0} b_l(L_i) + E(\phi_i)\{b_l(L_i)(\delta_{i0} + \delta_{i3}) + b_l(R_i)(\delta_{i1} + \delta_{i2})\} \Big] \exp(\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i \tag{1.7}$$

where

$$\gamma_l^*(\boldsymbol{\beta}) = \frac{\sum_{i=1}^{n} \{E(U_{il})\delta_{i0} + E(Z_{il})\delta_{i1} + E(W_{il})\delta_{i2}\}}{\Big[E(\psi_i)\delta_{i0} b_l(L_i) + E(\phi_i)\{b_l(L_i)(\delta_{i0} + \delta_{i3}) + b_l(R_i)(\delta_{i1} + \delta_{i2})\} \Big] \exp(\mathbf{x}_i'\boldsymbol{\beta})}. \tag{1.8}$$

2. Let $\gamma_l^{(d+1)} = \gamma_l^*(\boldsymbol{\beta}^{(d+1)})$ and increase d by 1.

Solving the system of equations in the first step of the iteration part can be accomplished using standard root finding routines, available in practically all existing statistical software packages. The second step of the iteration part is a simple updating of $\gamma_l^{(d)}$ in closed form. Thus, the implementation of the EM algorithm is straightforward and computationally inexpensive.

1.2.5 Asymptotic Properties and Variance Estimation

Obtained by EM algorithm, our estimator is actually an MLE which has all the good properties of MLE in general. Suppose the number and position of the knots are pre-specified and do not depend on the sample size, under the standard regularity conditions, as $n \to \infty$, $\hat{\boldsymbol{\theta}} \sim \mathcal{AN}(\boldsymbol{\theta}, \{I(\boldsymbol{\theta})\}^{-1})$. Depending on the missing information principle, Louis's method (Louis [1982]) gives a closed-form expression for the the Fisher information matrix as shown in below:

$$I(\boldsymbol{\theta}) = -\frac{\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \text{var}\{\frac{\partial \log L_c(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} | \mathcal{D}, \hat{\boldsymbol{\theta}}\} \tag{1.9}$$

In addition, all the entries in $\text{var}\{\partial \log L_c(\boldsymbol{\theta})/\partial \boldsymbol{\theta} | \mathcal{D}, \hat{\boldsymbol{\theta}}\}$ and $\partial^2 Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'$ have explicit forms. Thus Wald inference and the according confidence intervals for the regression coefficients $\boldsymbol{\beta}$ along with the baseline odds function can be obtained easily. The details pertaining to the calculation of the two matrix on the right hand side are provided, in closed-form, in Appendix A.

1.3 Simulation study

An extensive simulation study was conducted to evaluate the proposed method and compare it with the existing methods. We considered the following PO model for failure time T,

$$F(t|\mathbf{x}) = \frac{\Lambda_0(t) \exp(x_1\beta_1 + x_2\beta_2)}{1 + \Lambda_0(t) \exp(x_1\beta_1 + x_2\beta_2)},$$

where $\Lambda_0(t)$ is the baseline odds function, x_1 is a $N(0,1)$ random variable and x_2 is a $Bernoulli(0.5)$ random variable. The true values of β_1 and β_2 are taken to be $\{-1,0,1\}$ and the true baseline odds function $\Lambda_0(t)$ was taken to be $\Lambda_0(t) = \log(1+t) + t^3 + sin(t)$ and $\Lambda_0(t) = \log(1+t) + t^{1.5}$.

For each data set, we first generated failure time t_i for subject i by solving $F(t_i|x_i) = u_i$, where u_i was a random number from uniform distribution $U_{(0,1)}$. Then an indicator w_i was generated from $Bernoulli(0.2)$ to determine if the the failure time is exactly observed or censored. If $w_i = 1$, t_i was treated as exactly observed; and otherwise a series of examination times were then generated to determine the observed interval for subject i. For this part, the number of examination times was determined by 1 plus a Poisson random variable with mean 6, and the gap times between adjacent examination times were generated from independent exponential distributions with mean 0.2. Subsequently, the examination times were taken to be the cumulative sums of the gap times. The observed interval $[L_i, R_i]$ was determined by the two adjacent examination times (including 0 and ∞) that bracket the generated failure time t_i. For each parameter configuration, 500 independent data sets were generated each with sample size $n = 200$. On average, the simulated data contain 19.7% to 20.2% of exactly observed failure times, 11.1% to 23.1% of left-censored observations, 31.0% to 38.9% of interval-censored observations, and 19.9% to 36.1% of right-censored observations across all the setups.

We applied the proposed method to each simulated data set under PO model by fixing the degree of I-spline to be 3 and taking 9 equally spaced knots between 0 and maximum of the finite endpoints of all the observed intervals. For comparison purpose, we also implemented *ic_np* function in R package *Icenreg* on all the simulated data because *Icenreg* is the only publically available package to analyze arbitrarily censored data. Since a closed-form of the standard errors are not available in *Incereg*,

bootstrap method was applied to make statistical inference based on 100 of bootstrap samples for each data set.

Table 1.1 and Table 1.2 show the summarized estimation results of the regression parameters from both methods in terms of the average empirical bias (BIAS), the average of the 500 estimated standard errors (ESE), the sample standard deviation of the 500 point estimates (SSD), and the empirical coverage probability associated with 95% Wald confidence intervals for each parameter configuration.

As shown in Table 1.1 and Table 1.2, the small values of BIAS suggest that the point estimate of regression parameters obtained by the proposed method are all close to their corresponding true values. Additionally, ESE and SSD are in close agreement for all configurations, which indicates that the variance estimates based on Louis's method are accurate. The empirical coverage probabilities CP95s are close to the nominal level 0.95, indicating that the asymptotic normality of our estimators holds. As for comparison, it is clear from Tables 1 and 2 that the proposed methodology performed as good, if not better than, as *Icenreg* in estimating the regression parameters.

Table 1.1: Simulation results of estimated regression parameters from the proposed method and *Icenreg* when baseline odds is $\Lambda_0(t) = \log(1+t) + t^3 + sin(t)$.

		$\hat{\beta}_1$				$\hat{\beta}_2$			
Method	(β_1,β_2)	BIAS	ESE	SSD	CP95	BIAS	ESE	SSD	CP95
EM	(-1,-1)	0.0218	0.1308	0.1347	0.942	0.0143	0.2611	0.2679	0.946
Icenreg		0.0211	0.1366	0.1343	0.950	0.0152	0.2673	0.2668	0.948
EM	(-1,0)	0.0060	0.1306	0.1335	0.938	-0.0084	0.2590	0.2669	0.942
Icenreg		0.0058	0.1352	0.1333	0.948	-0.0060	0.2665	0.2672	0.954
EM	(-1,1)	0.0009	0.1298	0.1328	0.934	-0.0111	0.2597	0.2786	0.928
Icenreg		0.0004	0.1355	0.1330	0.942	-0.0101	0.2679	0.2784	0.932
EM	(0,-1)	0.0065	0.1304	0.1327	0.948	-0.0085	0.2597	0.2912	0.914
Icenreg		0.0066	0.1348	0.1331	0.960	-0.0077	0.2679	0.2907	0.914
EM	(0,0)	0.0045	0.1307	0.1275	0.954	-0.0019	0.2605	0.2575	0.950
Icenreg		0.0045	0.1359	0.1276	0.960	-0.0007	0.2686	0.2559	0.954
EM	(0,1)	-0.0054	0.1307	0.1366	0.948	-0.0012	0.2588	0.2614	0.952
Icenreg		-0.0054	0.1354	0.1362	0.946	-0.0014	0.2674	0.2614	0.960
EM	(1,-1)	-0.0023	0.1308	0.1333	0.946	-0.0120	0.2601	0.2682	0.938
Icenreg		-0.0021	0.1352	0.1331	0.950	-0.0117	0.2673	0.2670	0.946
EM	(1,0)	0.0008	0.1311	0.1307	0.958	0.0178	0.2593	0.2596	0.944
Icenreg		0.0005	0.1360	0.1305	0.962	0.0179	0.2677	0.2583	0.952
EM	(1,1)	-0.0071	0.1310	0.1294	0.948	0.0195	0.2603	0.2566	0.960
Icenreg		-0.0069	0.1355	0.1292	0.960	0.0182	0.2675	0.2575	0.960

In order to evaluate the performance of our method in estimating survival functions, we calculated the mean squared errors (MSE) of the estimates of the baseline survival function $S_0(t)$ at a set of pre-specified time points (taking 119 evenly-spaced grid points between 0 and 6). Table 1.3 summarizes the mean and maximum of those local MSEs for the two comparative methods. As seen in Table 1.3, both methods have an excellent performance in estimating the baseline odds function for all the simulation setups; However, the proposed method seems to perform slightly better than *Icenreg*.

Table 1.2: Simulation results of estimated regression parameters from the proposed method and *Icenreg* when baseline odds is $\Lambda_0(t) = \log(1+t) + t^{1.5}$.

		$\hat{\beta}_1$				$\hat{\beta}_2$			
Method	(β_1, β_2)	BIAS	ESE	SSD	CP95	BIAS	ESE	SSD	CP95
EM	(-1,-1)	-0.0068	0.1571	0.1531	0.958	-0.0127	0.2813	0.2898	0.950
Icenreg		-0.0190	0.1666	0.1546	0.966	-0.0244	0.2925	0.2940	0.950
EM	(-1,0)	-0.0104	0.1540	0.1593	0.952	0.0021	0.2672	0.2567	0.968
Icenreg		-0.0207	0.1635	0.1617	0.960	0.0032	0.2786	0.2613	0.980
EM	(-1,1)	-0.0053	0.1549	0.1520	0.958	-0.0148	0.2759	0.2679	0.960
Icenreg		-0.0194	0.1650	0.1561	0.966	-0.0022	0.2861	0.2726	0.950
EM	(0,-1)	-0.0009	0.1331	0.1320	0.956	-0.0151	0.2713	0.2788	0.938
Icenreg		-0.0014	0.1389	0.1327	0.958	-0.0209	0.2824	0.2793	0.936
EM	(0,0)	0.0017	0.1301	0.1328	0.946	-0.0024	0.2597	0.2713	0.946
Icenreg		0.0012	0.1352	0.1338	0.940	-0.0014	0.2670	0.2738	0.944
EM	(0,1)	0.0027	0.1311	0.1339	0.960	0.0109	0.2674	0.2666	0.954
Icenreg		0.0030	0.1365	0.1350	0.960	0.0163	0.2784	0.2690	0.964
EM	(1,-1)	0.0126	0.1580	0.1612	0.946	-0.0075	0.2805	0.2923	0.940
Icenreg		0.0224	0.1653	0.1630	0.950	-0.0173	0.2936	0.2959	0.956
EM	(1,0)	0.0231	0.1534	0.1607	0.944	-0.0132	0.2657	0.2654	0.948
Icenreg		0.0339	0.1634	0.1623	0.952	-0.0156	0.2634	0.2685	0.960
EM	(1,1)	0.0072	0.1537	0.1565	0.952	0.0134	0.2751	0.2724	0.952
Icenreg		0.0199	0.1628	0.1579	0.954	0.0262	0.2862	0.2754	0.954

Table 1.3: Mean and maximum of local MSEs for the estimated survival function $S_0(t)$.

		$\Lambda_0(t) = \log(1+t) + t^{1.5}$		$\Lambda_0(t) = \log(1+t) + t^3 + sin(t)$	
(β_1, β_2)	M	meanMSE	maxMSE	meanMSE	maxMSE
(-1,-1)	EM	0.0013	0.0022	0.0073	0.0300
	Icenreg	0.0020	0.0038	0.0075	0.0338
(-1,0)	EM	0.0013	0.0023	0.0072	0.0298
	Icenreg	0.0017	0.0029	0.0075	0.0332
(-1,1)	EM	0.0016	0.0024	0.0073	0.0299
	Icenreg	0.0020	0.0032	0.0075	0.0336
(0,-1)	EM	0.0012	0.0022	0.0075	0.0302
	Icenreg	0.0016	0.0078	0.0077	0.0340
(0,0)	EM	0.0016	0.0022	0.0075	0.0304
	Icenreg	0.0019	0.0028	0.0077	0.0339
(0,1)	EM	0.0018	0.0022	0.0073	0.0298
	Icenreg	0.0023	0.0029	0.0075	0.0337
(1,-1)	EM	0.0013	0.0024	0.0073	0.0298
	Icenreg	0.0018	0.0029	0.0075	0.0330
(1,0)	EM	0.0015	0.0024	0.0077	0.0310
	Icenreg	0.0019	0.0042	0.0079	0.0352
(1,1)	EM	0.0016	0.0024	0.0075	0.0310
	Icenreg	0.0021	0.0030	0.0078	0.0340

1.4 Real Data Application

To demonstrate that the algorithm finds the correct solution, results obtained from EM algorithm were compared with *Icenreg*'s *ic_sp* function on three real datasets. Since *Icenreg* needs bootstrap for estimating the standard errors, we fixed the number of bootstrap samples to 100 in all three analysis. The first dataset is a mixture of exact observed, left-, interval- and right censored data, which dose not have any covariate; the second dataset is an example of type-2 interval censored data with one covariate; the third data set contains exactly observed and interval-censored data with one covariate.

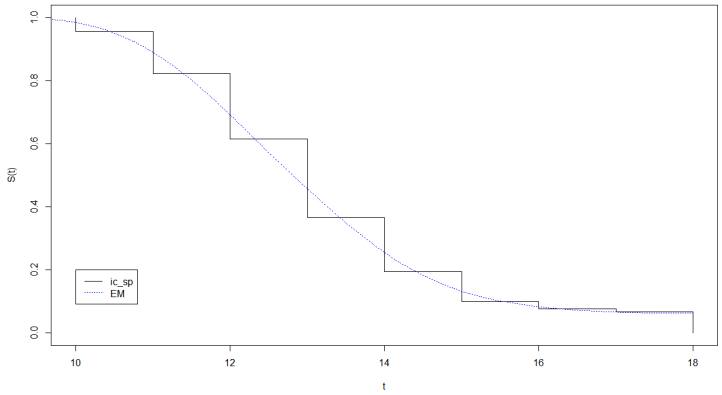

Figure 1.1: Time to the first use of marijuana

1.4.1 First Use of Marijuana

In 1975, on the Stanford-Palo Alto Peer Counseling Program, Hamburg et al. studied drug use in a representative sample of suburban junior and senior high school students. In this study, they found a distinctive age-related pattern of drug use among students. Later in 1987, Turnbull and Weiss extracted a set of failure time data of marijuana use from Hamburg et al.'s study. This data set summarizes the answer of 191 California high school students to the question "When did you first use marijuana?". A direct answer give rise to an exact observation. If the student answered, "I have never used it," then this gives rise to an observation which is censored on the right at his/her present age. The final possibility was someone who answered, "I have used it but cannot recall just when the first time was." This gives rise to a left censored observation where age of first use is known only to be prior to the student's current age.

Since there is no covariate in this dataset, the goal of the data analysis is to estimate the baseline survival function. In Figure 1.1 the smooth dotted line and the solid step line are survival functions obtained by our EM algorithm and *Icenreg*, respectively. As can be seen from this figure, the two methods match each other very well.

1.4.2 HIV

The dataset *Hemophilia* from r package **ICsurv** was collected in 1980's as part of a multi-center prospective study. Patients with hemophilia need blood products made from donors' plasma, so they are at risk of HIV-1 infection. This study was conducted to quantify the dose effect of blood products. Specifically, it aimed at assessing the HIV-1 infection rate in hemophilia patients with different average annual dose of blood products. In this study, 544 patients were classified into high, medium, low, or no dose group based on their average annual dose of blood products. The exact HIV-1 infection times were never observed and only observed intervals are available. Among all the patients, 63 of them are left-censored, 204 are interval-censored, and 277 are right-censored. Please refer to Goedert et al. [1989] and Kroner et al. [1994] for more detail about this study. This typical interval-censored data set has also been analyzed by Sun [2006] and Lin and Wang [2010], among many others. We specified monotone splines by fixing degree to 3 and taking 10 equally spaced interior knots within (0, 57.01).

The estimated regression coefficients for low, medium and high average annual doses obtained by the proposed EM algorithm and *ic_sp* are summarized in Table 1.4. From Table 1.4, those two methods give comparable results. These results suggest that there is a significant dose effect between each dose group and the non-dose group. In particular, under the proportional odds assumption the odds of HIV infection for patients using low-, medium- and high- dose of blood products is estimated to be approximately 9.88, 68.92 and 174.69 times that for patients who do not use blood products.

In Figure 1.2, we superimposed the estimated survival functions obtained from the proposed approach and *ic_np* for all the dose groups. The difference among these survival functions is clearly seen by comparing the three plots. Meanwhile, the esti-

mated survival function obtained by our method is consistent with the nonparametric estimate offered by *ic_np*.

Table 1.4: HIV data analysis: estimated regression coefficients for the average annual blood products does level.

Dose-Level	Method	Estimate	Exp(Est)	Std. Error	z-value	p-value
Low	EM	2.291	9.8848	0.2620	8.744	0.000
	ic_sp	2.278	9.7571	0.2682	8.494	0.000
Medium	EM	4.233	68.9237	0.3149	13.442	0.000
	ic_sp	4.226	68.4429	0.2772	15.250	0.000
High	EM	5.163	174.6877	0.3683	14.019	0.000
	ic_sp	5.163	174.6877	0.3775	13.680	0.000

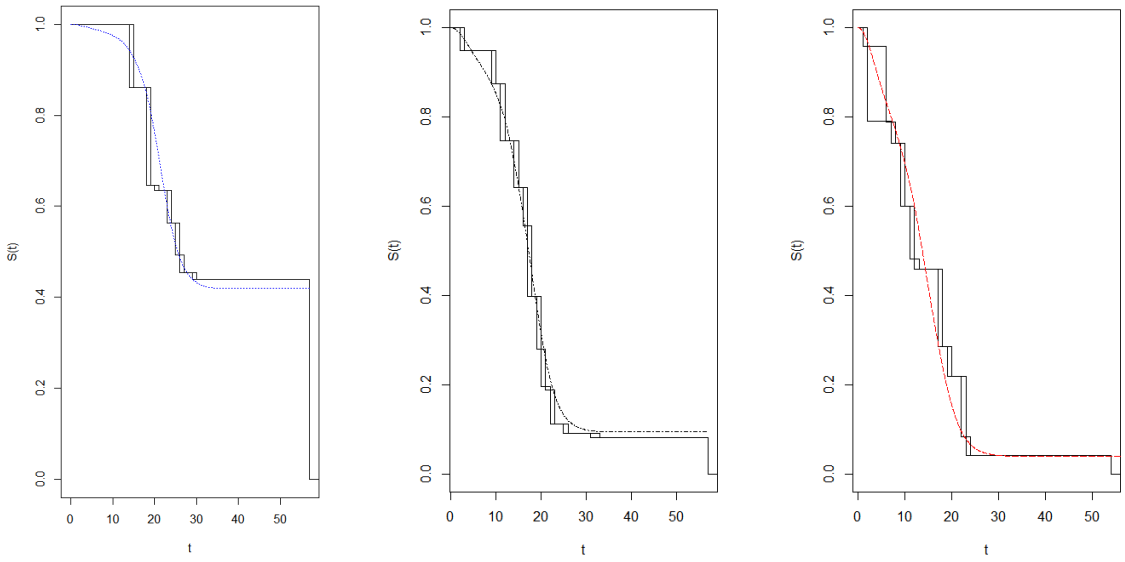

Figure 1.2: The estimated survival functions obtained from the proposed EM algorithm and *ic_np* for low, medium and high dose groups.

1.4.3 Diabetes

The *IR_diabetes* dataset from **icenReg** was analyzed by Anderson-Bergman [2017] with function *ic_sp*. We reanalyzed it for the purpose of comparison. This dataset is initially introduced by Zhao and Sun [2015] in their package **glrt**. The

data frame is based on a study conducted at the Steno Memorial Hospital in Denmark from 1933-1984. In this study, the time from onset of diabetes to onset of diabetic nephropathy is the response time of interest. It contains data from 731 patients (454 males and 277 females), for many of the patients (595), the event time was known exactly but for others (136) the event time was known only up to an interval due to limited follow up. The dataset contains three variables: left, right and gender. The variables left and right represent the observational interval and the effect of gender will be examined by the proposed EM algorithm under PO model. In this part, we also fixed the degree of I-spine to 3 for adequate flexibility and took 10 equally spaced interior knots within (0, 44.01).

The estimated regression coefficient for gender obtained by the proposed EM algorithm and *ic_sp* is summarized in Table 1.5. The proposed procedure resulted in practically identical estimates of the regression coefficients and inferential conclusions with *ic_sp*. From the result, it can be seen that there is a statistically significant difference in the odds of having experienced diabetic nephropathy at a given time after diabetes between men and women in the study. As the female is treated as baseline group, it is estimated that the odds of onset of diabetic nephropathy for men at any given time will be almost 0.681 times lower than for women under the assumption of proportional odds model.

Figure 1.3 plots the estimated survival functions for male and female obtained from the proposed method and *ic_sp*. The smooth dotted lines represent the estimated survival functions obtained by the proposed EM algorithm and the solid step lines are the estimated survival functions obtained by *ic_sp*. Figure 1.3 indicates the odds of survival for males is estimated to be higher than females at all times.

Table 1.5: IR_diabetes data analysis: estimated regression coefficients for gender.

	Estimate	Exp(Est)	Std. Error	z-value	p-value	Time
EM	-0.3833	0.682	0.1387	-2.763	0.0057	1.65
ic_sp	-0.4013	0.669	0.1407	-2.851	0.0044	24.14

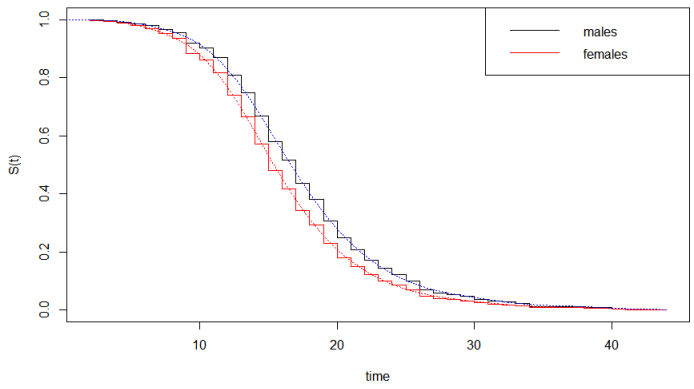

Figure 1.3: The estimated survival function for the diabetes data set.

1.5 Discussion

In this article a new method for analyzing arbitrary censored data is proposed under the proportional odds model. After a reparameterization basing on the semi-parametric framework, an approximation of the nondecreasing baseline odds function has been achieved by using monotone splines, therefore leading to a finite number of parameters to estimate. The EM algorithm developed in this paper can be used to find the maximum likelihood estimates of the baseline odds function and regression parameters simultaneously and to provide a closed-form of the asymptotic variance-covariance matrix. The key step in the derivation of the proposed EM algorithm is a four-step data augmentation which expands the observed data likelihood to a complete data likelihood. The expanding process involves the relationship between PO model and frailty PH model along with the first failure time under the PH model with a latent non-homogeneous Poisson process. Simulation study and real dataset

applications shown that the proposed method can provide accurate estimation results. The proposed method also has a full advantage in terms of computational speed. This is because the proposed approach provides a closed-form expression of the asymptotic variance estimates while *ic_sp* in *Icenreg* relies on bootstrap method to obtain variance estimates. In the simulation, it took 0.3s for the proposed approach to complete model fitting on average, and *ic_sp* took 3 times longer. This advantage could render the proposed approach preferable when analyzing larger data sets. As shown in real data application, for the *IR_diabetes* dataset with $n = 731$, it takes 1.65s and 24.14s for the proposed approach and *ic_sp* to fit the PO model, respectively. The reason is that the model fitting times for both methods increase as the sample size increases, but the bootstrap in *ic_sp* significantly magnified the prolonged fitting time thus makes it much slower than the proposed method. (All the simulations was conducted in R on a computer with a 3.60 GHz processor and 32.0 GB of memory.) In addition to taking less time and effort, the proposed approach does not require model assumptions about the observational process and thus is widely applicable. In summary, the proposed approach can be easily applied to various situations, it is accurate, reliable, and computationally efficient. We expect our approach to be widely used for analyzing any arbitrarily censored data under proportional odds model.

Chapter 2

Regression Analysis of Panel Count Data Accounting for Within-Subject Correlation with Nonparametric Frailty Distribution

2.1 Introduction

Panel count data is longitudinal count data that have the following characteristics: subjects are observed at several discrete time points during the study period; the number of observations varies from subject to subject; the observation times are treated as continuous random variables [Zhang and Jamshidian, 2003]. Data like this often occur in a long-term clinical, industrial or animal study where the primary end point is the time to a specific event and each subject may experience several such events over time [Sun and Wei, 2000]. Usually an estimation of the mean function for the recurrent event over time is researcher's main interest.

Various methods have been established to analyze the panel count data. Among all the methods, two types are most studied. One type is the likelihood-based methods and the other is generalized estimating equation methods. When Sun and Kalbfleisch [1995] first estimated the mean function of panel count data, they constructed a non-parametric estimator based on isotonic regression technique. Depending on their study, Wellner and Zhang [2000] studied a pseudo-likelihood estimator and the full maximum likelihood estimator of the mean function based on the nonhomogeneous Poisson process model. The pseudo-likelihood estimator ignores the dependence be-

tween counts in the counting process while the full nonparametric maximum likelihood Estimator (NPMLE) takes account of the dependence of the successive counts. Their corresponding method for panel count data with covariate [Wellner and Zhang, 2007] considered the proportional mean regression model under nonhomogeneous Poisson process assumption. Poisson process then become a commonly used tool for analyzing panel count data [Hua et al., 2014]. For example, Lu et al.[2007,2009] studied the spline-based sieve version of MPLE and MLE by approximating the baseline mean function using monotone B-spline functions.

Although Poisson likelihood-based estimation methods are consistent and robust against the underlying Poisson process assumption [Wellner and Zhang, 2000], the Poisson process-based likelihood does not address the overdispersion problem that often occurs in various applications of longitudinal count data [Hua and Zhang, 2012]. According to Cox [1983], overdispersion in general has two effects: underestimation of standard errors of the estimated regression parameters and loss of estimation efficiency. Both effects have been observed in the analysis of panel count data when overdispersion is neglected. Similar to introduce latent variable in GLM (Nielsen et al. [1992], Murphy [1995], Pan [1999], etc), adding a multiplicative or additive frailty term in Poisson likelihood has been widely used in the analysis of panel count data. Zhang and Jamshidian [2003] proposed an EM algorithm based on the gamma frailty Poisson model without incorporating covariates. Huang et al. [2006] introduced a latent frailty to account for informative observation times and avoided specifying its distribution through a conditional maximum likelihood approach. Yao et al. [2016] studied semiparametric regression analysis of panel count data under the gamma frailty Poisson model and derived an estimator of the within-subject correlations. Besides likelihood based methods, Hu et al. [2009] discussed an alternative method based on quasi-score equations with additional quadratic estimation equations to account for the overdispersion. Hua and Zhang [2012] developed a spline-based semi-

parametric projected generalized estimating equation method and showed that the semiparametric GEE method is actually equivalent to a semiparametric likelihood method based on a gamma-frailty Poisson process model. Later, Hua et al. [2014] established the asymptotic properties of this spline based estimators and claimed that the gamma-frailty Poisson process model is robust to frailty distribution misspecification. On the other hand, Yao et al. [2016] showed in their simulation studies that the estimation on the regression parameters may be biased when the gamma frailty assumption does not hold. In this paper, the primary objective is to model the distribution of frailty nonparametrically in the Poisson process model. Hence, frailty can properly account for the within-subject correlation to solve the overdispersion problem in all circumstances and provide a more accurate estimation. Specifically, Dirichlet process [Ferguson, 1973] mixture technique has been used to model the distribution of the frailties and we call this proposed approach nonparametric frailty Possion model (NPFPM).

The rest of the paper is organized as follows. Section 2.2 introduces the proportional mean model with frailty, the modeling of the baseline mean function with monotone splines and the distribution of the frailty with Dirichlet process. Section 3.5 provides the details of an easy-to-implement blocked Gibbs sampler for the posterior computation. Section 2.4 provides extensive simulation studies to evaluate the performance of our approach. Section 2.5 provides a real-life data application which involves the analysis of the bladder tumor data. Finally, we give some concluding remarks and discuss some further issues in Section 2.6.

2.2 The proposed model

2.2.1 Notation, model, and the likelihood

Consider a study that consists of n independent subjects. We assume that the observational process and the recurrent event process are conditionally independent

given covariates. For subject i, let $N_i(t)$ denote the counting process that is observed only at discrete examimation times $\{t_{ij}, j = 1, ..., J_i\}$, where J_i is the total number of observations and t_{iJ_i} is the last observation time. In order to account for the within-subject correlation, we introduce frailty term ϕ_i whose distribution is unspecified for each subject. Specifically, conditional on ϕ_i, $N_i(t)$ is a non-homogeneous Poisson process with mean function $\mu_0(t)\exp(\mathbf{x}'\boldsymbol{\beta})\phi_i$, where x_i is a vector of $p \times 1$ time-independent covariates and $\mu_0(t)$ is an unspecified nondecreasing baseline mean function with $\mu_0(0) = 0$.

By the properties of non-homogeneous Poisson process, define $Z_{ij} = N_i(t_{ij}) - N_i(t_{ij-1})$ as the count of recurrent events within time interval $(t_{ij-1}, t_{ij}]$, we have

$$Z_{ij}|\phi_i \sim Poi\Big[\{\mu_0(t_{ij}) - \mu_0(t_{ij-1})\}\exp(\mathbf{x}'\boldsymbol{\beta})\phi_i\Big],$$

and all Z_{ij}'s are conditionally independent given ϕ_i. Thus the observed data likelihood is

$$L_{obs} = \prod_{i=1}^{n} p(\phi_i) \prod_{j=1}^{J_i} \mathcal{P}(z_{ij}|\phi_i),$$

where $p(\phi_i)$ is the pdf of the frailty, whose form is unspecified; $\mathcal{P}(\centerdot|\phi_i)$ is the pdf of Poisson distribution with mean $\{\mu_0(t_{ij}) - \mu_0(t_{ij-1})\}\exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i$.

2.2.2 Monotone splines

Estimating the baseline mean function $\mu_0(\centerdot)$ is important as it is an indispensable part of the mean function but is also challenging because it is infinitely dimensional. In a nonparametric estimation, the number of parameters involved in $\mu_0(\centerdot)$ is on the order of sample size when the observation times differ from subject to subject.
To handle this situation, we use Yao et al. [2016]'s tactic and approximate the baseline mean function $\mu_0(\centerdot)$ with monotone spline of Ramsay [1988] in the following manner,

$$\mu_0(t) = \sum_{l=1}^{L} \gamma_l b_l(t), \tag{2.1}$$

where $b_l(\centerdot)$'s are integrated spline basis functions and γ_l's are spline coefficients, for $l = 1...L$. Each of the I-spline basis function is a piecewise polynomials of specified degree $d-1$. Each starts from 0 in an initial flat region, increases in a mid region, and then plateaus at 1 at higher values [Wang and Dunson, 2011]. In such a way, by constraining the basis coefficients to be nonnegative, the monotonicity of $\Lambda_0(\centerdot)$ can be guaranteed.

One can easily determine the form of monotone spline basis functions by specifying knots and degree. The placement of the knots determines the shape and the degree determines the smoothness of the monotone splines. According to previous studies, 2 or 3 degree can provide adequate smoothness. As for the placement of knots, Cai et al. [2011] have shown that using $10-30$ knots (equally-spaced or based on quantiles) provide adequate modeling flexibility for data sets containing up to thousands of observations [Cai et al., 2011, Wang and Dunson, 2011]. Additionally, we adopt a shrinkage prior for the spline coefficients in the Gibbs sampler to prevent over-fitting problems.

2.2.3 Dirichlet process mixture

Modeling the frailty distribution nonparametrically with DP is widely used in the analysis of clustered survival data. Naskar and Das [2006] propose a semiparametric bivariate binary model in which the subject-specific effects involved in the bivariate log odds ratio and the univariate logit components are assumed to follow a nonparametric Dirichlet proces. Naskar [2008] and Manda [2011] demonstrate the use of Dirichlet process prior for the frailty under Cox proportional hazard model, where the cluster-specific shared frailty is modeled nonparametrically with DP. Additionally, Pennell and David B. Dunson [2006] use Dirichlet process priors for modeling subject-specific shared frailty and for modeling multiplicative innovations on this frailty over time intervals.

However, because of the almost sure discreteness of the random measure generated by the Dirichlet process [Blackwell, 1973], modeling the distribution of ϕ_i nonparametrically by using Dirichlet process directly incurs the problem of unidentifiability and the discreteness of the posterior distribution. To circumvent the discrete constrain of DP and guarantee the identifiability of the parameters (i.e. make sure the mean of ϕ_i equal to 1), the Dirichlet process mixing of frailty distribution has been utilized instead of simple DP. More specifically, we assume the frailty to follow an unspecified distribution $p(\phi_i) = \int ga(\phi_i|v_i)d\pi(v_i)$. This is equivalent to a Dirichlet process mixture model, in which the frailty follows a conditional Gamma distribution with mean 1 and variance $1/v_i$, denoted by $\phi_i|v_i \sim \mathcal{G}a(v_i, v_i)$. The distribution of v_i is generated from Dirichlet process, i.e. $v_i \sim \pi(\centerdot)$ and $\pi(\centerdot) \sim DP(\alpha G_0)$. In this expression, $DP(\alpha G_0)$ refers to $\pi(\centerdot)$ being a random distribution generated by a Dirichlet process with base measure G_0 and total mass parameter α.

2.3 The proposed Gibbs sampler

Since the posterior distribution is intractable for exact inference, we built a Gibbs sampler [Geman and Geman, 1984] for our posterior computation. Gibbs sampler is one of the most popular Markov chain Monte Carlo (MCMC) [Robert and Casella, 2004] algorithms for Bayesian computation. It repeatedly and sequentially generates all unknown parameters and latent variables from their full conditional distributions. The MCMC theory guarantees that the limiting distribution of the samples from a Gibbs sampler is the same as the joint posterior distribution under certain regularity conditions. The Gibbs sampler we developed is a combination of non-homogeneous Poisson process and Dirichlet process. The former part is just a standard derivation of full conditional distribution while the latter part is an application of blocked Gibbs sampler.

2.3.1 Data augmentation

In order to exploit the monotone spline representation of $\mu_0(\cdot)$ in (2.1) and estimate the spline coefficients, a data augmentation is considered. By taking advantage of the Poisson likelihood and additive form of spline expression, we decompose Z_{ij} into the sum of L conditionally independent Poisson latent variables $\{Z_{ijl}\}_{l=1}^{L}$ given ϕ_i, for subject i and time interval $(t_{i,j-1}, t_{ij}]$, such that

$$Z_{ij} = \sum_{l=1}^{L} Z_{ijl}$$

$$Z_{ijl}|\phi_i \sim Poi[\gamma_l\{b_l(t_{ij}) - b_l(t_{ij-1})\}\exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i],$$

for $l = 1, ..., L$. Under this setting, the augmented data likelihood obtained a simple multiplication form:

$$L_c = \prod_{i=1}^{n} p(\phi_i) \prod_{j=1}^{J_i} \prod_{l=1}^{L} \mathcal{P}(z_{ijl}|\phi_i), \tag{2.2}$$

where $\mathcal{P}(\cdot|\phi_i)$ is the pdf of Poisson distribution with rate $\gamma_l\{b_l(t_{ij}) - b_l(t_{ij-1})\}\exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i$.

2.3.2 Approximated DP

Fitting the Dirichlet process gamma mixture model for the frailty distribution can be tricky and challenging. Here we adopted the blocked Gibbs sampler introduced by Ishwaran and James [2001]. The blocked Gibbs sampler is based on an approximated DP. In this approximation, the prior is assumed to be a finite dimensional measure whose random weight is generated by the stick-breaking construction:

$$p_1 = V_1 \quad and \quad p_k = (1 - V_1)(1 - V_2)\ldots(1 - V_{k-1})V_k, \quad k = 2, ..., N - 1,$$

where $V_1, V_2, ..., V_{N-1}$ are independent and identically distributed Beta(1, α) random variables. V_N is fixed to 1 so that $\sum_{k=1}^{N} p_k = 1$. Sethuraman [1994] showed that this truncated DP converges almost surely to DP(αG_0) as $N \to \infty$.

In particular, under the proposed model, define $\boldsymbol{\theta} = \{\theta_1, ..., \theta_N\}$ as the set of distinct v_i's, where $N \leq n$ is the number of distinct elements in $\mathbf{v} = \{v_1, ..., v_n\}$. Let $\mathcal{K} = \{K_1, ..., K_i\}$ denote the vector of configuration indicators such that $v_i = \theta_{K_i}$. Then denote the size of the h-th cluster as n_h, that is $n_h = \sum_{i=1}^{n} I(K_i = h)$. The random distribution which generated by a truncated DP has the from $\pi_N(\centerdot) = \sum_{h=1}^{N} p_h \delta_{\theta_h}$.

The finite dimensionality of such priors allows us to construsct our model in terms of a finite number of random variables. This then allows the blocked Gibbs sampler to update blocks of parameters, which, because of the nature of the prior, are drawn from simple multivariate distributions. Then the frailty part in the Bayesian semiparametric model can be written hierarchically as a truncated DP mixture:

$$\begin{aligned} (\phi_i|\boldsymbol{\theta}, \mathcal{K}) &\overset{iid}{\sim} \mathcal{G}a(v_i, v_i), \quad i = 1, ..., n \\ (K_i|\mathbf{p}) &\overset{iid}{\sim} \sum_{h=1}^{N} p_h \delta_{\theta_h} \quad h = 1, ..., N \\ (\mathbf{p}, \boldsymbol{\theta}) &\sim \pi(\mathbf{p}) \times G_0^N(\boldsymbol{\theta}) \end{aligned} \tag{2.3}$$

where θ_h are iid G_0. Note that usually Gibbs sampling schemes in mixture of Dirichlet process models are restricted to using conjugate base measures which allow analytic evaluation of the transition probabilities or alternatively need to rely on approximate numeric evaluations of some transition probabilities [Maceachern and Muller, 1998]. For convinience of computation, we assume G_0 is a gamma distribution.

2.3.3 Prior distributions

We need to specify the prior distributions for the unknown parameters $\boldsymbol{\beta}$ and γ_l's, then combining with the complete likelihood (Equation 2.2) we can obtain full conditional distribution of the parameters. We simply adopt conventional vague priors and assign independent exponential priors $\mathcal{E}xp(\lambda)$ to γ_l's, $\mathcal{G}a(a_\alpha, b_\alpha)$ prior to α and prior distribution $\mathcal{N}(\mu_0, \Sigma_0)$ to $\boldsymbol{\beta}$, with mean vector zero and large independent

variance such as 10. As mentioned before, we also assign a $\mathcal{G}a(a_\lambda, b_\lambda)$ hyper prior for λ. Theoretically, such a prior specification is closely related to Bayesian Lasso [Park and Casella, 2008] and is equivalent to the penalized likelihood approach with L1 penalty imposed on those spline coefficients, where λ serves as a tuning parameter. Lin et al. [2015] showed this shrinkage priors for the spline coefficients naturally prevents over-fitting and allow for automatic tuning with much less computational efforts.

2.3.4 Gibbs sampler

The initial values of θ_h's are sampled from $\mathcal{G}a(1,1)$ independently. Their corresponding p_h's are generated from Dirichlet$(\alpha/N, ..., \alpha/N)$, where α is generated from $\mathcal{G}a(a_\alpha, b_\alpha)$. Then the initial value of K_i's are generated independently from $Multinomial(1, \mathbf{p})$. With $\mathcal{K}$ we identify v_i for each observation and generate frailty ϕ_i independently. The Gibbs sampler iterates through the following steps:

1. Sample $Z_{ij1}, ..., Z_{ijL} | z_{ij} \sim Multinomial(z_{ij}, (q_{ij1}, ..., q_{ijL}))$, where

$$q_{ijl} = \frac{\gamma_l\{b_l(t_{ij}) - b_l(t_{ij-1})\}}{\sum_{j=1}^{L}\gamma_j\{b_j(t_{ij}) - b_j(t_{ij-1})\}} \qquad l = 1, ..., L$$

2. Sample γ_l from $\mathcal{G}a(\sum_{i=1}^{n}\sum_{j=1}^{J_i} Z_{ijl} + 1, \sum_{i=1}^{n}\{b_l(t_{iJ_i}) - b_l(t_{i0})\}\exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i + \lambda)$

3. Sample λ from $\lambda \sim \mathcal{G}a(\alpha_\lambda + L, \beta_\lambda + \sum_{l=1}^{L}\gamma_l)$

4. Sample $\boldsymbol{\beta}$ by using adaptive rejection metropolis sampling (ARMS)

$$L(\boldsymbol{\beta}|\mathbf{.}) \propto \exp\left\{\sum_{i=1}^{n}(\sum_{j=1}^{J_i} Z_{ij})\mathbf{x}_i'\boldsymbol{\beta} - \sum_{i=1}^{n}\{\mu_0(t_{iJ_i}) - \mu_0(t_{i0})\}\phi_i\exp(\mathbf{x}_i'\boldsymbol{\beta}) - (\boldsymbol{\beta} - \boldsymbol{\mu}_0)'\Sigma_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_0)/2\right\}$$

5. Sample ϕ_i from $\mathcal{G}a(Z_i + v_i, \{\mu_0(t_{iJ_i}) - \mu_0(t_{i0})\}\exp(\mathbf{x}_i'\boldsymbol{\beta}) + v_i)$

6. Sample θ_h, for $h = 1, ..., N$, by using ARMS

$$\theta_h \propto \exp(-\theta_h)\prod_{\{i:K_i=h\}}\frac{\theta_h^{\theta_h}}{\Gamma(\theta_h)}\phi_i^{\theta_h - 1}\exp(-\theta_h\phi_i)$$

7. Sample $K_i \sim Multinomial(1, \mathbf{p_i})$, where

$$(p_{1i}, ..., p_{Ni}) \propto (p_1 ga(\phi_i|\theta_1), ..., p_N ga(\phi_i|\theta_N)$$

8. Sample $V_h \overset{ind}{\sim} Beta(1 + n_h, \alpha + \sum_{s=h+1}^{N} n_s)$, for $h = 1, ..., N-1$ and let $V_N = 1$. Then let $p_1 = V_1$ and $p_h = V_h(1 - V_{h-1})...(1 - V_1)$, for $h = 2, ..., N$.

9. Sample α from $\mathcal{G}a(a_\alpha + N - 1, b_\alpha - \sum_{h=1}^{N-1} log(1 - V_h))$

2.4 Simulation study

Extensive simulation studies are conducted to assess the performance of the proposed approach thoroughly. To demonstrate the robustness and the other advantages of the proposed nonparametric frailty Poisson model (NPFPM), we compared it with gamma frailty Poisson model (GFPM) under six different frailty distributions: (1) the simplest cases where data were generated from gamma frailty Poisson model where the frailty ϕ_i's were generated from $\mathcal{G}a(0.5, 0.5)$; (2) the data were generated from mixture gamma frailty Poisson model, in which the frailty follows mixture gamma distribution $0.5ga(1, 1) + 0.5ga(50, 50)$; (3) To further explore the performance of our approach with more complicated frailty form, we generate data with the frailty follows a mixture of four gamma distributions, i.e., $\phi_i \sim 0.25ga(0.5, 0.5) + 0.25ga(1, 1) + 0.25ga(10, 10) + 0.25ga(50, 50)$; some other distributions are also considered, such as (4) ϕ_i's follow a lognormal distribution with mean 1 and variance 2; (5) ϕ_i's follow a log-logistic distribution with shape π and scale sin(1), so that it has mean 1 and variance around 0.56; and (6) ϕ_i's follow mixture lognormal distribution $0.5\mathcal{LN}(1, 2) + 0.5\mathcal{LN}(1, 0.02)$, which has mean 1 and variance 1.01. Note the distributions of the frailty become more complicated in a progressive manner. For the purpose of illustration, we fix the order of monotone spline to 3 and use 18 equally-spaced interior knots in all the situations. For the part

involves Dirichlet process in the NPFPM, we fixed N, the number of distinct values of v_i's to 20.

For each setting, we simulated 500 data sets and there are $n = 100$ subjects in each data set. To generate the observational process for subject i, we first generate the total number of observation times from 1 plus a Poisson random variable with mean 6, then generate gap times independently from an exponential distribution with a rate parameter 2. The counting process associated with subject i was generated from the following model,

$$Z_{ij}|\phi_i = N_i(t_{ij}) - N_i(t_{ij-1}) \sim Poi\Big[\{\mu_0(t_{ij}) - \mu_0(t_{ij-1})\}\exp(x_{i1}\beta_1 + x_{i2}\beta_2)\phi_i\Big],$$

where $\mu_0(t) = log(1+t)+t$, $x_{i1} \sim Bernoulli(0.5)$, $x_{i2} \sim N(0, 0.5^2)$, the true values of regression coefficients (β_1, β_2) have values $(1, -1)$ or $(-1, 1)$, and the frailty ϕ_i comes from different distributions shown previously.

Table 2.1 summarizes the simulation results on the estimation of (β_1, β_2) from NPFPM and GFPM in terms of bias, the difference between the average of 500 posterior means and the true parameter value; ESE, the average of the estimated posterior standard errors; SSD, the sample standard deviation of the 500 posterior means; and CP95, the coverage rate based on the 500 95% credible intervals. As shown in Table 2.1, NPFPM performs well and steady with the frailty following different distributions. The small values of BIAS suggest that the point estimate of regression parameters obtained by NPFPM are close to their corresponding true values hence the estimation is unbiased. Additionally, ESE is close to SSD in each situation, indicating the proposed approach effectively solves the problem of overdispersion. This advantage is more obvious when compared to GFPM, whose ESE and SSD have larger differences when frailty does not follow gamma distribution. Meanwhile, the magnitudes of ESE and SSD are smaller in NPFPM than that in GFPM, so that NPFPM can offer a shorter confidence interval and provide more precise inference. The empirical coverage probabilities (CP95) for the confidence intervals for the re-

gression parameters are close to the nominal level 0.95, which means the Bayesian inference based on NPFPM is reliable.

Table 2.1: Simulation results from NPFPM and GFNPM when the data were generated from Poisson model in which the frailty generated 6 different distributions.

Dist	(β_1, β_2)	Est	NPFPM				GFNPM			
			Bias	ESE	SSD	CP95	Bias	ESE	SSD	CP95
GM	$(1,-1)$	$\hat{\beta}_1$	-0.054	0.300	0.285	0.966	-0.054	0.300	0.291	0.950
		$\hat{\beta}_2$	0.016	0.317	0.328	0.944	0.018	0.317	0.321	0.942
	$(-1,1)$	$\hat{\beta}_1$	-0.048	0.328	0.332	0.924	-0.049	0.326	0.324	0.928
		$\hat{\beta}_2$	-0.003	0.347	0.352	0.952	-0.004	0.344	0.344	0.952
TGM	$(1,-1)$	$\hat{\beta}_1$	-0.001	0.142	0.137	0.957	-0.004	0.161	0.163	0.945
		$\hat{\beta}_2$	-0.010	0.144	0.143	0.955	-0.012	0.166	0.166	0.945
	$(-1,1)$	$\hat{\beta}_1$	0.004	0.186	0.179	0.954	0.009	0.189	0.189	0.948
		$\hat{\beta}_2$	0.005	0.188	0.188	0.958	0.007	0.196	0.193	0.952
FGM	$(1,-1)$	$\hat{\beta}_1$	-0.007	0.159	0.160	0.958	-0.014	0.186	0.200	0.912
		$\hat{\beta}_2$	-0.000	0.162	0.162	0.948	-0.001	0.192	0.201	0.946
	$(-1,1)$	$\hat{\beta}_1$	-0.010	0.204	0.205	0.954	-0.010	0.212	0.222	0.938
		$\hat{\beta}_2$	0.007	0.207	0.214	0.948	0.016	0.218	0.231	0.934
$\mathcal{LN}$	$(1,-1)$	$\hat{\beta}_1$	-0.032	0.227	0.255	0.918	-0.016	0.227	0.288	0.884
		$\hat{\beta}_2$	0.008	0.238	0.255	0.926	-0.014	0.237	0.272	0.904
	$(-1,1)$	$\hat{\beta}_1$	-0.017	0.264	0.290	0.916	-0.034	0.262	0.318	0.884
		$\hat{\beta}_2$	0.026	0.270	0.290	0.932	0.034	0.272	0.306	0.914
$\mathcal{LL}$	$(1,-1)$	$\hat{\beta}_1$	-0.017	0.141	0.149	0.930	-0.012	0.144	0.169	0.916
		$\hat{\beta}_2$	0.003	0.146	0.160	0.922	-0.003	0.150	0.176	0.900
	$(-1,1)$	$\hat{\beta}_1$	-0.001	0.178	0.190	0.922	-0.012	0.181	0.2010	0.914
		$\hat{\beta}_2$	0.007	0.178	0.178	0.954	0.023	0.183	0.194	0.938
M$\mathcal{LN}$	$(1,-1)$	$\hat{\beta}_1$	-0.013	0.134	0.130	0.956	-0.005	0.145	0.156	0.942
		$\hat{\beta}_2$	-0.000	0.136	0.130	0.962	-0.009	0.148	0.159	0.934
	$(-1,1)$	$\hat{\beta}_1$	-0.002	0.173	0.173	0.946	-0.010	0.177	0.196	0.926
		$\hat{\beta}_2$	-0.008	0.172	0.176	0.946	0.002	0.180	0.186	0.934

Summarized results include the bias (Bias), the average of the estimated posterior standard errors(ESE), the sample standard deviation of the 500 posterior means (SSD), and the 95% coverage rate (CP95). The true frailty distributions are: GM: gamma distribution; TGM: mixture of two gamma distributions; FGM: mixture of four gamma distributions; $\mathcal{LN}$, lognormal distribution; $\mathcal{LL}$, log-logistic distribution; and mix-$\mathcal{LN}$, mixture lognormal distribution.

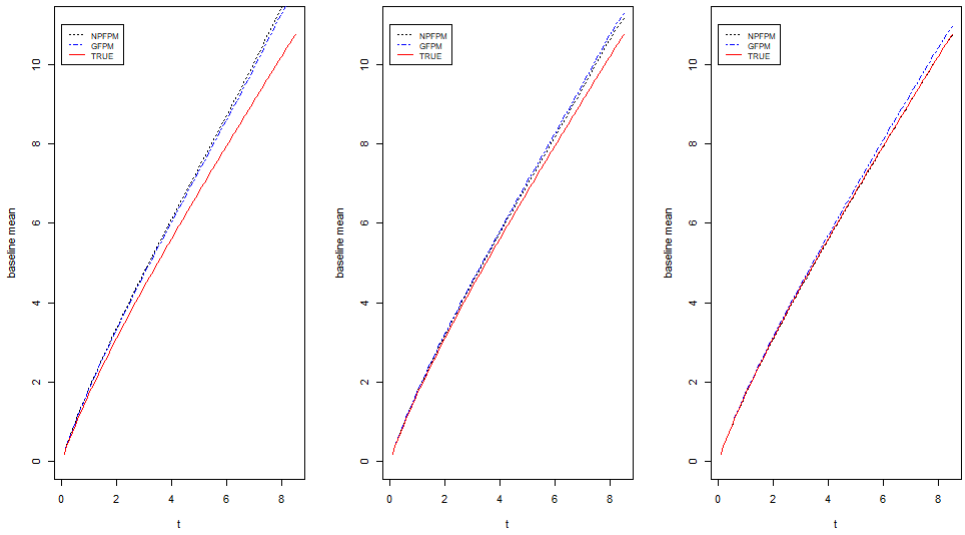

Figure 2.1: The true baseline mean function and the average of the estimated baseline mean curves under NPFPM and GFPM, when frailty follows gamma distribution, mixture of four gamma distribution and log-logistic distribution.

Regarding baseline mean function, Figure 2.1 shows the true baseline mean function and the average of the baseline mean function estimates from NPFPM and GFPM when $(\beta_1, \beta_2) = (1, -1)$ and frailty follows $\mathcal{G}a(0.5, 0.5)$, mixture of four gamma distribution specified above, and log-logistic$(\pi, \sin(1))$. As seen in Figure 2.1, NPFPM and GFPM have similar performance in general because the averaged baseline mean estimates from NPFPM and GFPM essentially overlaps with the true curve. However, after close inspection, we can see GFPM gives better estimation when frailty follows gamma distribution while NPFPM gives better estimation in the other two settings.

In summary, these two methods have comparable performance when frailty follows gamma distribution. However, NPFPM performs better in terms of parameter estimation, inferential characteristics, and baseline mean function estimation when frailty distribution is different from gamma. Thus, we conclude that estimating the frailty nonparametrically with Dirichlet process mixture is robust and it solves the problem of underestimating variances and increases the coverage probabilities.

This makes sense because there are multiple counts from the same subjects which can provide adequate information to estimate the frailty distribution accurately [Yao et al., 2016].

2.5 Real-life data application

In this section, we apply the proposed method to the most widely used panel count data example in the literature, which arose from a bladder cancer study conducted by the Veterans Administration Cooperative Urological Research Group [Byar and Blackard, 1977]. In this randomized clinical trial study, all the 118 patients had experienced superficial bladder tumors when they entered the trial. They were randomized into one of three treatment groups: placebo, thiotepa, and pyridoxine. During the study at each follow-up visit, new tumors since the last visit were counted, measured and then removed transurethrally. The number of follow-up clinical visits and follow-up times vary noticeably from patient to patient. The primary objective of the study is to determine if any treatment could significantly reduce the recurrence of bladder tumor.

This data set has been analyzed extensively using many different approaches in the literature. Following Wellner and Zhang [2007] and Lu et al. [2009], we focused on 116 patients in the study, who had at least one follow-up observation after the study enrollment. Let $\mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4})'$ denote the covariate vector for patient i, where x_{i1} and x_{i2} represent the number of bladder tumors and the size of the largest bladder tumors for patient i at the beginning of the trial, and x_{i3} and x_{i4} are the binary variables indicate whether patient i was assigned to the treatment of pyridoxine pills or thiotepa installation, respectively. When applying the proposed method, we use 20 equally-spaced knots within the data range $0 - 64$ months for the monotone spline specification. In addition, we fix the number of distinct values of v_i's to 20 as in the simulation.

Table 2.2: Bladder tumor data analysis from the proposed approach, the GFNPM approach and the WZ approach in Wellner and Zhang (2007).

	NPFPM			GFNPM			WZ		
	Point	SE	CI95	Point	SE	CI95	Point	SE	CI95
$\hat{\beta}_1$	0.317	0.101	(0.109,0.502)	0.336	0.106	(0.128,0.544)	0.207	0.078	(0.054,0.360)
$\hat{\beta}_2$	-0.026	0.129	(-0.260,0.255)	0.012	0.120	(-0.223,0.247)	-0.036	0.086	(-0.133,0.205)
$\hat{\beta}_3$	-0.107	0.391	(-0.838,0.716)	-0.033	0.409	(-0.835,0.769)	0.066	0.431	(-0.779,0.911)
$\hat{\beta}_4$	-1.219	0.442	(-2.043,-0.330)	-1.140	0.435	(-1.993,-0.287)	-0.797	0.360	(0.091,1.503)

Summarized results are the point estimates (Point), the standard errors (SE), and the p-values for all the regression parameters and the frailty variance parameter v.

Table 2.2 shows the results from the proposed approach and two other competitive approaches, i.e. Yao et al. [2016] and Wellner and Zhang [2007]. The results from these two competitors are directly drawn from their papers. Both of these two competitive approaches are likelihood-based approaches under the non-homogeneous Poisson model. Yao et al. [2016]'s method considered the within-subject correlation while Wellner and Zhang [2007]'s method did not consider the within-subject correlation.

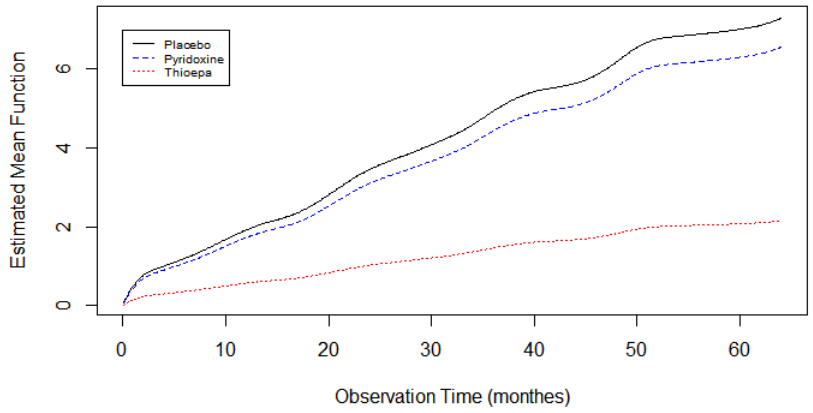

Figure 2.2: The estimated mean functions for different groups for the bladder tumor data.

As seen in Table 2.2, the results from our method indicates that the number of initial bladder tumors was positively related to the recurrence of the tumor while the size of the largest tumor at the enrollment did not have a significant effect. It also reveals that the thiotepa instillation treatment significantly reduced the recurrence rate of bladder tumors, while the treatment of pyridoxine pills did not have a significant effect. Figure 2.2 plots the estimated mean functions of bladder tumor counts for the control and the other two treatment groups. It is clear that the estimated mean functions for the control and the pyridoxine treatment groups are close to each other and they are higher than the one for the thiotepa treatment group. These conclusions

are consistent with those made in Wellner and Zhang [2007] and more close to Yao et al. [2016] in terms of regression coefficients estimation and baseline mean function estimation. This is because Yao et al.'s method also accounts for the within-subject correlation. In addition, the estimated density function of the frailty from the proposed method and Yao et al.'s method highly matches with each other as shown in Figure 2.3. For this data, the within-subject correlation is not ignorable because the tumor number at baseline is positively related to the recurrence of bladder tumor [Hua and Zhang, 2012].

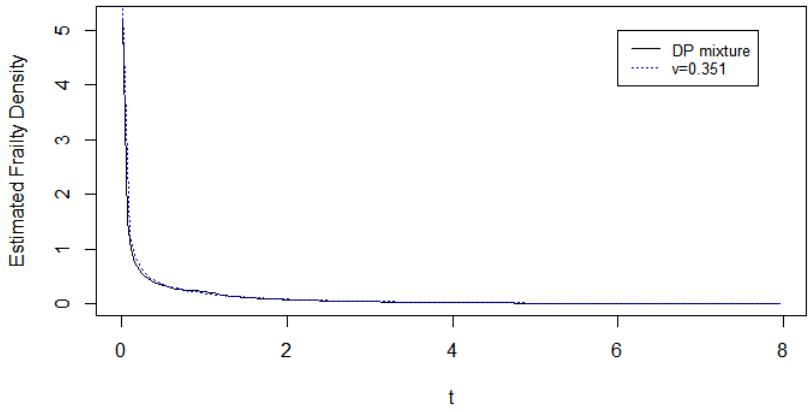

Figure 2.3: The estimated probability density function of the frailty for the bladder tumor data.

2.6 Discussion

In this paper, we proposed a Bayesian estimation approach to analyze panel count data in a non-homogeneous Poisson process framework. By introducing multiplicative frailty term for the proportional mean model, this approach is able to account for within-subject correlation. The distribution of the frailty is estimated nonparametrically with a Dirichlet process mixture which solves the problem of overdispersion in likelihood based methods. The baseline mean function is approximated by monotone splines which leads to a finite number of parameters to estimate and thus save the

computation effort. An easy-to-implement Gibbs sampler is established upon Poisson data augmentation and the blocked Gibbs sampler of Ishwaran and James [2001]. The proposed method shows an excellent performance of estimating the regression parameters and the baseline mean function when frailty follows different distributions as shown in our simulation studies and the real data application. Our future effort will be devoted to developing more robust methods and extend this strategy to multivariate penal count data.

Chapter 3

Bayesian inferences for panel count data and interval-censored data with nonparametric modeling of the baseline functions

3.1 Introduction

Panel count data frequently occur in epidemiological and social-behavioral studies. In such studies, subjects experience multiple recurrences of an event of interest such as smoking or infections but they are monitored or observed only at finite discrete time points instead of continuously. A consequence of such design is that the exact occurrence times of the events are not recorded and only the numbers of the occurrences of the events within the time windows are available, which leads to the panel count data structure. The primary interests for panel count data are to estimate the mean function of the counting process and/or the covariate effects on the counts. Sun and Kalbfleisch [1995] first studied nonparametric estimation of the mean function of panel count data using the monotonicity of the mean function of a counting process. They applied the Isotonic regression technique to estimate the mean function nonparametrically. Wellner and Zhang [2000] modeled the counting process with a nonhomogeneous Poisson process and established a pseudo-likelihood by ignoring the within-subject correlations. Subsequent researches [Wellner and Zhang[2000, 2007]] showed that methods based on the underlying conditional Poisson process assumption are robust to the actual distribution of the underlying counting process. To account

for the within-subject correlation, Zhang and Jamshidian [2003] and Yao et al. [2016] introduced gamma-frailty variable and established EM algorithms to fit the model. Researches [Hua et al., 2014, Xu et al., 2018] have shown that gamma frailty models exhibit a certain level of robustness to some misspecified frailty distributions.

In some situations, the event of interest is not recurrent for each subject, such as death or the onset of HIV infection. In this case, the resulting counting process becomes a 0-1 process and the panel count data reduce to interval-censored survival data [Wellner and Zhang, 2000]. Among all the models, Cox's proportional hazards (PH) model [Cox, 1972] is unquestionably the most popular and widely used semiparametric regression model in survival literature. It assumes that covariates have a multiplicative effect on the hazard function of the failure time of interest. Many approaches have been developed for the regression analysis of interval censored data under the PH model. For example, Finkelstein [1986] proposed a Newton-Raphson algorithm to fit the model, Satten [1996] proposed a marginal likelihood approach, Goggins et al. [1998] developed a Monte Carlo EM algorithm, Satten et al. [1998] proposed estimating equations; Pan proposed a generalized gradient projection method Pan [1999] and a multiple imputation method Pan [2000], Cai and Betensky [2003] developed a penalized likelihood approach, Zhang et al. [2010] proposed a seive maximum likelihood method, Sinha et al.[1999] used piecewise constant functions to model the baseline hazard function. Besides Cox's model, the semiparametric proportional odds (PO) model also has plausible properties such as interpretable, closely related to logistic regression and ratio of the hazards converges to unity as time increases. It assumes the covariates have a multiplicative effect on the unspecified baseline odds function instead. Compare to PH model, there are only a handful of studies on fitting PO model with interval-censored data due to the complexity of data structure. Huang and Rossini [1997] proposed a sieve maximum likelihood estimator for proportional odds model with interval censored data. Shen [1998] use monotone splines

of variable orders and knots for approximating the odds of failure time and proposed a sieve maximum likelihood estimator for right-censored and interval-censored data. Lin and Wang [2010] proposed a Bayesian approach for analyzing interval-censored data under the semiparametric proportional odds model.

For all three semiparametric models, the baseline function, say, the baseline mean function of panel count data along with the baseline cumulative hazards function and baseline odds function of the interval censored data, are usually assumed to be unspecified fixed nondecreasing functions that equal to zero at the initial time point 0. In most existing studies, they are approximated by linear combinations of spline functions [Hua and Zhang, 2012, Hua et al., 2014, Yao et al., 2016] or by a step-function with fixed jump sizes at every observation time points [Zhu et al., 2018]. Different from previous researches, we propose to approximate the baseline function nonparametrically with a nondecreasing stochastic process with independent nonnegative increments. One such is the gamma process, which can be thought of as arising from a compound Poisson process of gamma-distributed increments in which the Poisson rate tends to infinity while the sizes of the increments tend to zero in proportion [Lawless and Crowder, 2004]. Upon this gamma process assumption, Bayesian estimations of the baseline functions can be established and Gibbs samplers can be developed to fit the model.

It is not rare to see the gamma process be used to estimate nondecreasing function. Kalbfleisch [1978] used gamma process to model the cumulative hazard function and built up a Bayesian analysis of the semi-parametric regression and life model of Cox[1972]. Ferguson and Phadia [1979] used the processes neutral to the right as prior distributions for the unknown distribution function F and derived general theory that can be used in the estimation of F given some right censored data. However, as mentioned in Wellner and Zhang [1998] most of those results are limited to the special case in which each subject has the same number of observation times. For

example, Groeneboom and Wellner [1992] and Huang [1996] developed method for cases when each subject is only observed once or twice. The proposed method solved this concern by creating a find partition of the time space and define a gamma process on these intervals. The proposed method allows each individual has totally different number of observation times and even different observation schemes. It can also easily handle missing observation points.

The organization of the rest of this paper is as follows. In Section 2, we includes the setting of the model; an introduction of the notations; a brief description of the successive intervals based on the observed values; a detail introduction of the application of gamma process; and a step of data augmentation which is important for the computation of posterior distributions. In section 3, the proposed Gibbs sampler is exhibited. In Section 4, the performance of proposed Bayesian approach is assessed by a group of simulation study. In Section 5, the proposed method is applied to real data. Section 6 is summary discussion.

3.2 Models, Notations, and the observed data likelihoods

3.2.1 Gamma frailty Poisson process for panel count data

Suppose that there are n independent subjects in a study. For subject i, the counting process of the recurrent event of interest $N_i(t)$ is only observed at examination time points $\{t_{ij}, j = 1, ..., J_i\}$, where J_i is the total number of examination time points for subject i. Conditioning on an unobserved frailty ϕ_i, we assume a non-homogeneous Poisson process for $N_i(t)$ with the following conditional mean function

$$E(N_i(t)|\phi_i) = \mu_{it} = \mu_0(t)\exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i,$$

where $\mu_0(t)$ is an unspecified nondecreasing baseline mean function with $\mu_0(0) = 0$, $\mathbf{x}_i$ is a $p \times 1$ vector of time-independent covariates, and $\phi_i \sim \mathcal{G}a(v, v)$ is a frailty with

mean 1 and variance v^{-1}. The purpose of introducing the unobserved frailty ϕ_i is to account for the within-subject correlation between the counts for subject i.

Define $Z_{ij} = N_i(t_{ij}) - N_i(t_{ij-1})$, the count of recurrent events within time interval $(t_{ij-1}, t_{ij}]$, the properties of the non-homogeneous Poisson process guarantee that Z_{ij}'s are conditionally independent Poisson random variables given ϕ_i and

$$Z_{ij}|\phi_i \sim Poi\Big[\{\mu_0(t_{ij}) - \mu_0(t_{ij-1})\}\exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i\Big]$$

for $j = 1, \cdots, J_i$ and $i = 1, \cdots, n$. By assuming that the observational process and the recurrent event process are conditionally independent given the time-independent covariates, the observed data likelihood takes the form

$$L_{obs} = \prod_{i=1}^{n} \int g(\phi_i|v) \prod_{j=1}^{J_i} \mathcal{P}(z_{ij}|\mu_{ij}) d\phi_i,$$

where $g(.|v)$ is the probability density function (PDF) of $\mathcal{G}a(v, v)$, $\mathcal{P}(.|\mu_{ij})$ is the PDF of Poisson distribution with mean equal to $\mu_{ij} = \{\mu_0(t_{ij}) - \mu_0(t_{ij-1})\}\exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i$. A similar stochastic model was proposed by Sinha [2004], which assumes that all the subjects are examined at the same set of time points.

3.2.2 The PH and PO models for general interval-censored data

We consider the PH and PO models for general interval-censored data in this subsection. Let T_i denote the survival time of interest and $\mathbf{x}_i$ a $p \times 1$ vector of potential covariates for subject i, for $i = 1, \cdots, n$. Due to the study design of examining subjects periodically, T_i is not exactly observed but is known to fall within some observed interval $(L_i, R_i]$, with $0 \leq L_i < R_i \leq \infty$ for $i = 1, \cdots, n$. This general interval $(L_i, R_i]$ yields a left-censored observation when $0 = L_i < R_i < \infty$, a strictly interval-censored observation when $0 < L_i < R_i < \infty$, and a right-censored observation when $0 < L_i < R_i = \infty$. Under the assumption that the failure time and the observation process are independent conditional on the covariates, the observed data likelihood

takes the form

$$L_{obs} = \prod_{i=1}^{n} \{1 - S(R_i|\mathbf{x}_i)\}^{\delta_{i1}} \{S(L_i|\mathbf{x}_i) - S(R_i|\mathbf{x}_i)\}^{\delta_{i2}} S(L_i|\mathbf{x}_i)^{\delta_{i3}}, \tag{3.1}$$

where $S(\cdot|\mathbf{x})$ is the survival function of the failure time given covariate $\mathbf{x}$, and δ_{i1}, δ_{i2}, and δ_{i3} are binary censoring indicators for left-censored, interval-censored, and right-censored observations, respectively, with the constraint $\delta_{i1} + \delta_{i2} + \delta_{i3} = 1$ for each i. The survival time $S(t|\mathbf{x})$ takes the form $S(t|\mathbf{x}) = \exp\{-\mu_0(t)\exp(\mathbf{x}'\boldsymbol{\beta})\}$ under the PH model and $S(t|\mathbf{x}) = \{1 + \mu_0(t)\exp(\mathbf{x}'\boldsymbol{\beta})\}^{-1}$ under the PO model, where $\mu_0(\cdot)$ can be interpreted as the baseline cumulative hazard function and the baseline odds function under the PH and PO models, respectively.

It is known that the survival function in the PO model can be rewritten as the marginal survival function of the frailty PH model with the frailty following a $\mathcal{G}a(1,1)$ distribution, i.e.,

$$S(t|\mathbf{x}) = \{1 + \mu_0(t)\exp(\mathbf{x}'\boldsymbol{\beta})\}^{-1} = \int_0^{\infty} \exp\{-\mu_0(t)\exp(\mathbf{x}'\boldsymbol{\beta})\phi\}\exp(-\phi)d\phi. \tag{3.2}$$

This fact suggests that we can rewrite the observed likelihood (3.1) under the PH and PO models as the following unified form

$$L_{obs} = \prod_{i=1}^{n} \int_0^{\infty} [1 - S(R_i|\mathbf{x}_i,\phi_i)]^{\delta_{i1}} [S(L_i|\mathbf{x}_i,\phi_i) - S(R_i|\mathbf{x}_i,\phi_i)]^{\delta_{i2}} S(L_i|\mathbf{x}_i,\phi_i)^{\delta_{i3}} g(\phi_i)d\phi_i, \tag{3.3}$$

where $S(t|\mathbf{x},\phi) = \exp\{-\mu_0(t)\exp(\mathbf{x}'\boldsymbol{\beta})\phi\}$ is the conditional survival function of the failure time given covariate $\mathbf{x}$ and frailty ϕ, ϕ_i's are i.i.d. frailties with a density function g, and g takes degenerated Point mass distribution at 1 for the PH model and $\mathcal{G}a(1,1)$ for the PO model.

3.3 Modeling the baseline functions nonparametrically

3.3.1 Gamma process

The baseline mean function in the Poisson process, the baseline cumulative hazard function in the PH model, and the baseline odds function in the PO model are all

unspecified nondecreasing functions taking 0 at time 0. We model these unknown functions μ_0 nonparametrically by assigning them a Gamma process prior. Gamma process is a Lévy process with gamma distributed increments. It has been extensively studied and applied in various studies since its introduction by Doksum [1974] as one of the processes neutral to the right. For example, Nozer [1995] use gamma process to model the hazard rate process in a dynamic environment; Lawless and Crowder [2004] extend gamma process model by incorporating random effect to explore its use as a degradation model; Wang [2009] introduced a nonparametric estimation of the shape function in gamma process for degradation data; Sinha et al. [2015] assumes the wear process is a gamma process and developed a Bayesian analysis of a stochastic wear process model to fit survival data that might have a large number of ties, etc.

Specifically, we assign a Gamma process prior $\mathcal{GP}(H_0, \eta)$ for μ_0, where H_0 is the expected function of μ_0 and η quantifies the uncertainty level of this guess. The larger value of η, the closer μ_0 is to H_0. This prior implies the following two facts. First, for any $t > 0$, $\mu_0(t)$ has a Gamma distribution $\mathcal{G}a(\eta H_0(t), \eta)$. Second, the increments in non-overlapping time intervals are independent of each other. That is, $\mu_0(t+s) - \mu_0(t)$ is independent of $\mu_0(t)$ and

$$\mu_0(t+s) - \mu_0(t) \sim \mathcal{G}a(\eta\{H_0(t+s) - H_0(t)\}, \eta)$$

for any $t > 0$ and $s > 0$.

3.3.2 A fine partition of the time space

To further adjust the proposed approach so that it allows subjects to have different examination intervals, we build a find partition of the time space based on the observed values (t_{ij}'s for panel count data and all L_i's and R_i's for interval-censored data). Basically, they are intervals over the time line upon which any observed interval is a union of some of the partitioned intervals.

For panel count data, we define interval $\{s_m, s_{m-1}\}$ to be the non-empty intersection of observed intervals $\{t_{ij}, t_{ij-1}\}$ such that $\{s_m, s_{m-1}\} \cap \{t_{ij}, t_{ij-1}\}$ is either an empty set or $\{s_m, s_{m-1}\}$ [Li et al., 1997]. In this way, the time between two successive examination time points for subject i can be expressed as a union of certain intervals, $\{t_{ij-1}, t_{ij}\} = \bigcup_{l=0}^{k_{i1}-k_{ij-1}-1}\{s_{k_{ij-1}+l}, s_{k_{ij-1}+l+1}\}$, where k_{ij} is an index of the position of t_{ij} on the scale of $\mathbf{s} = \{s_0, ..., s_M\}$ and $s_{k_{ij}} = t_{ij}$ for each i and j. Similarly, for interval censored data $\{L_i, R_i\} = \bigcup_{l=0}^{k_{i2}-k_{i1}-1}\{s_{k_{i1}+l}, s_{k_{i2}+l+1}\}$, where k_{i1} and k_{i2} are indexes of positions of L_i and R_i on the scale of $\mathbf{s} = \{s_0, ..., s_M\}$ so that $s_{k_{i1}} = L_i$ and $s_{k_{i2}} = R_i$.

Define $\lambda_m = \mu_0(s_m) - \mu_0(s_{m-1})$ as the increment of the baseline mean function on interval $\{s_m, s_{m-1}\}$, for $m = 1, \cdots, M$. It is known that λ_m follows $\mathcal{G}a(\{H_0(s_m) - H_0(s_{m-1}\}\eta, \eta)$. Since gamma process is robust to the choice of H_0, we simply choose $H_0(t) = at$. For panel count data, the increment of baseline mean function for subject i in the jth time interval is $\mu_0(t_{ij}) - \mu_0(t_{ij-1}) = \sum_{l=1}^{k_{ij}-k_{ij-1}} \lambda_{k_{ij-1}+l}$. So the conditional likelihood after incorporating the Gamma process for μ_0 for panel count data can be written as:

$$L_{pc1} \propto \Big(\prod_{i=1}^{n} \Big[\prod_{j=1}^{J_i} \Big(\sum_{l=1}^{k_{ij}-k_{ij-1}} \lambda_{k_{ij-1}+l} \Big)^{z_{ij}} e^{\mathbf{x}_i'\boldsymbol{\beta} z_{ij}} \phi_i^{z_{ij}} \exp\Big\{ - \Big(\sum_{l=1}^{k_{ij}-k_{ij-1}} \lambda_{k_{ij-1}+l} \Big) e^{\mathbf{x}_i'\boldsymbol{\beta}} \phi_i \Big\} \Big] \times g(\phi_i|v) \Big) \prod_{m=1}^{M} g\{\lambda_m | a(s_m - s_{m-1})\eta, \eta\}. \tag{3.4}$$

For interval censored data, each individual only have one observed interval, the increment of the cumulative baseline hazard or the baseline odds function for subject i is $\mu_0(t_{i2}) - \mu_0(t_{i1}) = \sum_{m=k_{i1}+1}^{k_{i2}} \lambda_m$. So the conditional likelihood functions given the frailties ϕ_i's under PH and PO models after incorporating the Gamma process prior

for μ_0 take the form

$$L_{IC1} \propto \Big(\prod_{i=1}^{n} [1 - \exp\{-(\sum_{m=1}^{k_{i2}} \lambda_m) \exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i\}]^{\delta_{i1}} [\exp\{-(\sum_{m=1}^{k_{i1}} \lambda_m) \exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i\}]^{\delta_{i3}}$$
$$\times [\exp\{-(\sum_{m=1}^{k_{i1}} \lambda_m) \exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i\} - \exp\{-(\sum_{m=1}^{k_{i2}} \lambda_m) \exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i\}]^{\delta_{i2}} \exp(-\phi_i) \Big)$$
$$\times \prod_{m=1}^{M} g\{\lambda_m | a(s_m - s_{m-1})\eta, \eta\}. \tag{3.5}$$

3.4 Data Augmentation

3.4.1 For Panel count data

Based on the fact that for $(u_1, ..., u_K) \sim Multi(1, (\frac{1}{K}, ..., \frac{1}{K}))$, integrate $\prod_{l=1}^{K} [\lambda_l]^{u_l}$ with respect to $(u_1, ..., u_K)$ equals to $\sum_{l=1}^{K} \lambda_l / K$. We introduce $\mathbf{u}_{ij} \sim Multi(1, (\frac{1}{n_{ij}}, ..., \frac{1}{n_{ij}}))$, where $n_{ij} = k_{ij} - k_{ij-1}$, for $i = 1, ..., n$ and $j = 1, ..., J_i$. The augmented likelihood function has a simple multiplication form:

$$L_{aug} \propto \Big(\prod_{i=1}^{n} \Big[\prod_{j=1}^{J_i} (\prod_{l=1}^{k_{ij}-k_{ij-1}} \lambda_{k_{ij-1}+l}^{u_{ijl}})^{z_{ij}} e^{\mathbf{x}_i'\beta z_{ij}} \phi_i^{z_{ij}} \exp\Big\{ - (\sum_{l=1}^{k_{ij}-k_{ij-1}} \lambda_{k_{ij-1}+l}) e^{\mathbf{x}_i'\beta} \phi_i \Big\} \Big]$$
$$\times g a(\phi_i | v) \Big) \prod_{m=1}^{M} g\{\lambda_m | a(s_m - s_{m-1})\eta, \eta\}. \tag{3.6}$$

This complete data likelihood is the product of Poisson probability mass functions multiplied by the gamma densities for the increments of the baseline mean function μ_0. A Gibbs sampler is to be developed based on this complete data likelihood.

3.4.2 For interval-censored data

The data augmentation below is based on the connection between a failure time T_i following a PH and a latent non-homogeneous Poisson process as in Lin et al. (2015) and Wang et al. (2016). Let $N_i(t)$ denote a latent non-homogeneous Poisson process with cumulative intensity function: $\mu_0(t) \exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i$ for subject i, and let $\mathcal{T}_i$ denote

the time of the first jump of the counting process, that is: $\mathcal{T}_i = \inf\{t : N_i(t) > 0\}$. Then it is clear that the probability of the first jump hasn't shown up yet at time t is

$$P(\mathcal{T}_i > t) = P(N_i(t) = 0) = \exp\{-\mu_0(t)\exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i\},$$

which suggests that $\mathcal{T}_i$ has a frailty PH model with frailty ϕ.

Define $t_{i1} = R_i I(\delta_{i1} = 1) + L_i I(\delta_{i2} = 1)$ and $t_{i2} = R_i I(\delta_{i2} = 1) + L_i I(\delta_{i3} = 1)$ for each i. Let $Z_i = N(t_{i1})$ and $W_i = N(t_{i2}) - N(t_{i1})$ depending on the availability of t_{i1} and t_{i2} from the observed interval $(L_i, R_i]$. Based on the properties of Poisson process, one has

$$Z_i \sim Poisson(\mu_0(t_{i1})\exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i),$$

and

$$W_i \sim Poisson(\{\mu_0(t_{i2}) - \mu_0(t_{i1})\}\exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i).$$

Conditional on the latent variables Z_i's, W_i's, and ϕ_i's, the augmented data likelihood is

$$\begin{aligned} L_{aug_2} &= \prod_{i=1}^{n} \mathcal{P}(Z_i|\mu_0(t_{i1})\exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i)[\mathcal{P}(W_i|\{\mu_0(t_{i2}) - \mu_0(t_{i1})\}\exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i)]^{\delta_{i2}+\delta_{i3}} e^{-\phi_i} \\ &\propto \Big(\prod_{i=1}^{n}\{(\sum_{m=1}^{k_{i1}}\lambda_m)e^{\mathbf{x}_i'\boldsymbol{\beta}}\phi_i\}^{z_i}\exp\Big\{-(\sum_{l=1}^{k_{i1}}\lambda_m)e^{\mathbf{x}_i'\boldsymbol{\beta}}\phi_i\Big\}\{(\sum_{m=k_{i1}+1}^{k_{i2}}\lambda_m)e^{\mathbf{x}_i'\boldsymbol{\beta}}\phi_i\}^{W_i(\delta_{i2}+\delta_{i3})} \\ &\times \exp\Big\{-(\sum_{m=k_{i1}+1}^{k_{i2}}\lambda_m)e^{\mathbf{x}_i'\boldsymbol{\beta}}\phi_i(\delta_{i2}+\delta_{i3})\Big\}\Big)\prod_{m=1}^{M} g\{\lambda_m|a(s_m - s_{m-1})\eta, \eta\} \end{aligned}$$

subject to : when $\delta_{i1} = 1$, $Z_i > 0$; when $\delta_{i2} = 1$, $Z_i = 0$ and $W_i > 0$; when $\delta_{i3} = 1$, $Z_i = 0$ and $W_i = 0$. Here $\mathcal{P}(\cdot|\gamma)$ denote the Poisson probability mass function with the rate parameter γ.

Taking advantage of the fact that the sum of independent Poisson random variables is still a Poisson random variable, we decompose both Z_i and W_i as a sum of K_{i1} and $K_{i2} - K_{i1}$ conditionally independent Poisson random variables given ϕ_i:

$$Z_{im} \sim Poisson(\lambda_m\exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i)$$

$$W_{im} \sim Poisson(\lambda_m \exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i)$$

where $\sum_{m=1}^{K_{i1}} Z_{im} = Z_i$ and $\sum_{m=K_{i1}-1}^{K_{i2}} W_{im} = W_i$. Conditional on the latent variables Z_{im}'s, W_{im}'s, and ϕ_i's, the agumented data likelihood takes the following form

$$\begin{aligned} L_C \propto \Big(& \prod_{i=1}^{n} [\prod_{m=1}^{K_{i1}} \{\lambda_m \exp(\mathbf{x}_i'\beta)\phi_i\}^{z_{im}} \exp\{-\lambda_m \exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i\}] \exp(-\phi_i) \\ & \times [\prod_{m=K_{i1+1}}^{K_{i2}} \{\lambda_m \exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i\}^{w_{im}\delta_{i2}} \exp\{-\lambda_m \exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i(\delta_{i2}+\delta_{i3})\}] \Big) \\ & \times \prod_{m=1}^{M} g\{\lambda_m | a(s_m - s_{m-1})\eta, \eta\} \end{aligned} \tag{3.7}$$

Again, this augmented likelihood is the product of Poisson probability mass functions and Gamma densities, and this will be used as the complete data likelihood for our Bayesian computation.

3.5 Gibbs sampler

3.5.1 Panel count data

For the purpose of providing flexible modeling while also allowing for efficient posterior computation, we assign conventional vague priors for all of the parameters in the Bayesian approach. Specifically, we assign a multivariate normal $\mathcal{N}(\mu_0, \Sigma_0)$ prior for the regression coefficients $\boldsymbol{\beta}$, with mean vector zero and large independent variances such as 10 or 10^6. In practical, the very noninformative prior can balance both the skeptical and the enthusiastic views about the effects of covariates such as assigned treatments [Sinha, 2004]. We adopt independent $\mathcal{G}a(1,1)$ priors for v, a and η. By giving all the parameters fixed initial values 1, our Gibbs sampler iterates through the following steps:

1. Sample λ_m for $m = 1, ..., M$, from

$$\lambda_m \sim \mathcal{G}a(\sum_i \sum_j \sum_l u_{ijl} Z_{ij} + a(s_m - s_{m-1})\eta, \sum_i \exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i + \eta),$$

for all i, j, l such that $k_{ij-1} + l = m$.

2. Sample ϕ_i from

$$\mathcal{G}a(\sum_{j=1}^{J_i} Z_{ij} + v, \{\sum_{j=1}^{J_i} \sum_{l=1}^{k_{ij}-k_{ij-1}} \lambda_{k_{ij-1}+l} \exp(\mathbf{x}_i'\boldsymbol{\beta})\} + v)$$

3. Sample $U_{ij1}, ..., U_{ijn_{ij}} \sim Multi(1, (p_{ij1}, ..., p_{ijn_{ij}}))$, where

$$p_{ijl} = \frac{\lambda_{k_{ij-1}+l}}{\sum_{l=1}^{k_{ij}-k_{ij-1}} \lambda_{k_{ij-1}+l}} \qquad l = 1, ..., L$$

4. Sample $\boldsymbol{\beta}$ by using adaptive rejection metropolis sampling (ARMS)

$$L(\boldsymbol{\beta}|\boldsymbol{.}) \propto \exp\left\{\sum_{i=1}^{n} \left\{ Z_i \mathbf{x}_i'\boldsymbol{\beta} - (\sum_{j=1}^{J_i} \sum_{l=1}^{k_{ij}-k_{ij-1}} \lambda_{k_{ij-1}+l})\phi_i \exp(\mathbf{x}_i'\boldsymbol{\beta}) \right\} - (\boldsymbol{\beta}-\boldsymbol{\mu}_0)'\Sigma_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_0)/2 \right\}$$

5. Sample v by using ARMS

$$L(v|\boldsymbol{.}) \propto \exp(-v)\{\frac{v^v}{\Gamma(v)}\}^n (\prod_{i=1}^{n} \phi_i)^{v-1} \exp(-v \sum_{i=1}^{n} \phi_i)$$

6. Sample a by using ARMS

$$L(a|\boldsymbol{.}) \propto \frac{\eta^{a\eta s_M}}{\prod_{m=1}^{M} \Gamma\{a(s_m - s_{m-1})\eta\}} \prod_{m=1}^{M} \lambda_m^{a(s_m - s_{m-1})\eta - 1} \exp(-a)$$

7. Sample η by using ARMS

$$L(\eta|\boldsymbol{.}) \propto \frac{\eta^{a\eta s_M}}{\prod_{m=1}^{M} \Gamma\{a(s_m - s_{m-1})\eta\}} \prod_{m=1}^{M} \lambda_m^{a(s_m - s_{m-1})\eta - 1} \exp(-\eta \sum_{m=1}^{M} \lambda_m) \exp(-\eta)$$

3.5.2 Interval censored data under PO model

We adopt independent $\mathcal{G}a(1,1)$ priors for a and η. By giving all the parameters fixed initial values 1, our Gibbs sampler iterates through the following steps:

1. Sample λ_m for $m = 1, ..., M$, from $\mathcal{G}a(\alpha_m, \beta_m)$ with

$$\alpha_m = \sum_{i}^{n} \{Z_{im} I(m \leq K_{i1}) + W_{im}\delta_{i1} I(K_{i1} < m \leq K_{i2})\} + a(s_m - s_{m-1})\eta$$

$$\beta_m = \sum_{i}^{n} \exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i \{I(m \leq K_{i1}) + I(K_{i1} < m \leq K_{i2})(\delta_{i2} + \delta_{i3})\} + \eta$$

2. Sample ϕ_i from

$$\mathcal{G}a(Z_i + W_i\delta_{i2} + 1, \{1 + (\sum_{m=1}^{k_{i1}} \lambda_m)\delta_{i1} + (\sum_{m=1}^{k_{i2}} \lambda_m)(\delta_{i2} + \delta_{i3})\} \exp(\mathbf{x}_i'\boldsymbol{\beta}))$$

3. Sample $\boldsymbol{\beta}$ by using ARMS

$$L(\boldsymbol{\beta}|\boldsymbol{.}) \propto \exp\Big(\sum_{i=1}^{n} \Big[\{Z_i + W_i(\delta_{i2}+\delta_{i3})\}\mathbf{x}_i'\boldsymbol{\beta} - \{(\sum_{m=1}^{k_{i1}} \lambda_m) + (\sum_{m=k_{i1}+1}^{k_{i2}} \lambda_m)(\delta_{i2}+\delta_{i3})\}\phi_i \exp(\mathbf{x}_i'\boldsymbol{\beta})\Big]\Big)$$

4. Sample $Z_i's, Z_{im}'s, W_i's$ and $W_{im}'s$ by setting all of them to be 0. Then for each i:

a. If $\delta_{i1} = 1$ (i.e. left-censored) sample:

$$Z_i \sim Poisson((\sum_{m=1}^{K_{i1}} \lambda_m) \exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i) I(Z_i > 0)$$

$$Z_{i1}, ..., Z_{iK_{i1}} | z_i \sim Multi(z_i, (p_{i1}, ..., p_{iK_{i1}}))$$

$$p_m = \frac{\lambda_m}{\sum_{m=1}^{K_{i1}} \lambda_m} \qquad m = 1, ..., K_{i1}$$

b. If $\delta_{i2} = 1$ (i.e. interval-censored) sample:

$$W_i \sim Poisson((\sum_{m=K_{i1}+1}^{K_{i2}} \lambda_m) \exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i) I(W_i > 0)$$

$$W_{i1}, ..., W_{i(K_{i2}-K_{i1})} | w_i \sim Multi(z_i, (q_{i1}, ..., q_{i(K_{i2}-K_{i1})}))$$

$$q_{im} = \frac{\lambda_m}{\sum_{m=K_{i1}+1}^{K_{i2}} \lambda_m} \qquad m = 1, ..., K_{i2} - K_{i1}$$

5. Sample a by using ARMS

$$L(a|\boldsymbol{.}) \propto \frac{\eta^{a\eta s_M}}{\prod_{m=1}^{M} \Gamma\{a(s_m - s_{m-1})\eta\}} \prod_{m=1}^{M} \lambda_m^{a(s_m - s_{m-1})\eta - 1} \exp(-a)$$

6. Sample η by using ARMS

$$L(\eta|\boldsymbol{.}) \propto \frac{\eta^{a\eta s_M}}{\prod_{m=1}^{M} \Gamma\{a(s_m - s_{m-1})\eta\}} \prod_{m=1}^{M} \lambda_m^{a(s_m - s_{m-1})\eta - 1} \exp(-\eta \sum_{m=1}^{M} \lambda_m) \exp(-\eta)$$

As for PH model, the whole sampling process keeps the same in general except for fixing the value of ϕ_i to 1.

3.6 Simulation study

Comprehensive simulations are conducted to evaluate the proposed approach. To generate simulated data for each subject, we set 50 evenly allocated examination time points on time interval $(0, 10]$ to imitate the pre-decided research time span and observation scheme. Then we randomly remove 20% of examination time points for each of the subject. In this way, all the subjects have different total number of examination times and gap times.

3.6.1 Panel count data

To generate panel count data, the counting process associated with subject i was generated from the following model,

$$Z_{ij}|\phi_i = N_i(t_{ij}) - N_i(t_{ij-1}) \sim Poi\Big[\{\mu_0(t_{ij}) - \mu_0(t_{ij-1})\}\exp(x_{i1}\beta_1 + x_{i2}\beta_2)\phi_i\Big],$$

where x_{i1} is continuous variable that follows a normal distribution, $N(0, 0.5^2)$ and x_{i2} is a binary variable that follows the Bernoulli distribution, $Bernoulli(0.5)$. The true regression coefficients are $\beta_1 = \{-1, 1\}$, $\beta_2 = \{-1, 1\}$. The distribution of ϕ_i is $\mathcal{G}a(1, 1)$. We assessed the proposed approach with two different baseline mean function: $\mu_0(t) = \log(1 + t) + t^{1.5}$ and $\mu_0(t) = t + \sin(t)$. For each setting, 100 data sets with sample size $n = 100$ are generated.

The proposed Gibbs sampler in Section 3.5 is implemented to fit the Gamma frailty proportional mean model for each of the simulated data set. Table 3.2 present the frequentist operating characteristics of the estimates of the regression parameters. Bias is the difference between the average of 100 posterior means and the true parameter value; ESE is the average of the estimated posterior standard errors; SSD is the sample standard deviation of the 100 posterior means; and CP95 is the empirical coverage probability based on the 100 95% credible intervals.The results from our proposed method indicate that the proposed method performs well in terms of the

estimation of the regression parameters, as the estimates show no bias, ESD and SSD are close to each other, and the coverage probabilities are all close to 0.95.

Table 3.1: Estimation of regression parameters for panel count data based on 100 simulated data sets from the proposed Bayesian method.

		$\mu_0(t) = log(1+t) + t^{1.5}$				$\mu_0(t) = t + sin(t)$			
Method	(β_1,β_2)	BIAS	ESD	SSD	CP95	BIAS	ESD	SSD	CP95
GP	(-1,-1)	0.0217	0.2072	0.1715	0.98	0.0014	0.2228	0.2526	0.94
		0.0066	0.2128	0.2186	0.97	0.0145	0.2343	0.2567	0.94
SP		0.0427	0.2056	0.1697	0.97	-0.0185	0.2246	0.2501	0.95
		0.0030	0.2133	0.2218	0.93	0.0142	0.2341	0.2568	0.94
GP	(-1,1)	0.0250	0.2092	0.2119	0.93	-0.0366	0.2249	0.2403	0.93
		-0.0216	0.2126	0.2283	0.95	-0.0099	0.2323	0.2567	0.96
SP		0.0370	0.2111	0.2167	0.92	-0.0607	0.2277	0.2394	0.93
		-0.0168	0.2135	0.2301	0.92	-0.0089	0.2347	0.2545	0.96
GP	(1,-1)	-0.0038	0.1934	0.2096	0.93	0.0074	0.2101	0.2329	0.91
		0.0017	0.2032	0.2076	0.92	-0.0164	0.2194	0.2196	0.95
SP		0.0039	0.1933	0.2026	0.93	-0.0148	0.2106	0.2250	0.90
		0.0111	0.1994	0.2116	0.94	-0.0158	0.2211	0.2272	0.95
GP	(1,1)	0.0160	0.1997	0.2009	0.95	0.0248	0.2112	0.2284	0.93
		0.0355	0.2091	0.2005	0.94	-0.0016	0.2150	0.1986	0.97
SP		0.0170	0.1947	0.2013	0.91	-0.0034	0.2103	0.2245	0.94
		0.0295	0.2107	0.2075	0.91	-0.0050	0.2175	0.1981	0.97

Empirical bias (BIAS), the average of the estimated standard errors (ESD) and standard deviation (ESD) of β, and the empirical coverage probabilities associated with 95% confidence probability (CP95).

3.6.2 Interval censored data under PH and PO model

To generate interval censored data under PH and PO model, we first generate the failure time T from

$$F(t|\mathbf{x}) = 1 - \exp\{-\Lambda_0(t)\exp(x_1\beta_1 + x_2\beta_2)\}$$

under PH model and from

$$F(t|\mathbf{x}) = \frac{\Lambda_0(t)\exp(x_1\beta_1 + x_2\beta_2)}{1 + \Lambda_0(t)\exp(x_1\beta_1 + x_2\beta_2)},$$

under PO model, where $\Lambda_0(t) = \log(1+t) + t^{1.5}$, x_1 is a $N(0,1)$ random variable and x_2 is a $Bernoulli(0.5)$ random variable. The true values of β_1 and β_2 are taken

to be $\{0, 1\}$ and $\Lambda_0(t)$ is the baseline cumulative hazards function and baseline odds function for PH and PO model, respectively.

The observed interval $(L_i, R_i]$ was then determined by the two adjacent examination times (including 0 and ∞) that bracket the generated failure time t_i. For each parameter configuration, 100 independent data sets were generated each with sample size $n = 200$. On average, the simulated data contain 11.1% to 23.1% of left-censored observations, 31.0% to 38.9% of interval-censored observations, and 19.9% to 36.1% of right-censored observations across all the setups.

Table 3.2: Estimation of regression parameters for interval censored data based on 100 simulated data sets from the proposed Bayesian method.

		$\hat{\beta}_1$				$\hat{\beta}_2$			
Model	(β_1,β_2)	BIAS	ESD	SSD	CP95	BIAS	ESD	SSD	CP95
PH	(1,1)	-0.0539	0.0905	0.0961	0.88	-0.0495	0.1490	0.1416	0.97
PO		-0.0827	0.1301	0.1226	0.92	-0.0517	0.2221	0.2380	0.94
PH	(1,0)	-0.0438	0.0897	0.0731	0.95	-0.0066	0.1398	0.1259	0.97
PO		-0.0547	0.1309	0.1151	0.96	-0.0368	0.2172	0.2297	0.95
PH	(0,1)	0.0041	0.0764	0.0751	0.94	-0.0096	0.1492	0.1685	0.92
PO		0.0053	0.1287	0.1226	0.96	-0.0205	0.2268	0.2287	0.93
PH	(0,0)	0.0017	0.0737	0.0696	0.97	0.0015	0.1359	0.1495	0.93
PO		0.0006	0.1251	0.1267	0.95	-0.0162	0.2182	0.2298	0.93

BIAS: the empirical bias; ESD: the average of the estimated standard errors; SSD: standard deviation of β; CP95: the empirical coverage probabilities associated with 95% confidence probability.

3.7 Real-life data application

3.7.1 The patent study

We applied the proposed method to analysis of an industrial economics data set from the r package 'pglm'. The current data set is an extract from a larger data set that is collected by Hall et al. for their study of the relationship between patenting and research and development activity at the firm level by the U.S. manufacturing sector

during the 1970's. This dataset contains 346 firms in the United States. Among them, 147 firms are in the scientific sector. During 1970 to 1979, the number of patents applied for in each year that were eventually granted is recorded for every firm. The data set also includes all the firms' book value of capital in 1972 and their annual research and development (R & D) spending.

In this section, the primary objective of our analysis is to assess the relationship between the mean number of patents and the characteristics of the firm. x_{i1} is binary variable that indicate if the firm i is in the scientific sector. x_{i2} and x_{i3} are the book value of capital in 1972 and the average annual research and development (R & D) spending for the firm i, respectively. To mitigate the problem of collinearity, we standardized x_2 and x_3 before fitting the model. For the purpose of comparison, we also analyzed this data set with GFNPM [Yao et al., 2016].

Table 3.3: Patent data analysis from the proposed approach (GFGP) and GFNPM.

	GFGP			GFNPM		
	Point	SE	CI95	Point	SE	CI95
$\hat{\beta}_1$	0.537	0.146	(0.240,0.834)	0.561	0.127	(0.311, 0.798)
$\hat{\beta}_2$	1.032	0.148	(0.812,1.427)	1.130	0.122	(0.818, 1.303)
$\hat{\beta}_3$	0.617	0.144	(0.335,0.801)	0.795	0.144	(0.483, 1.005)
$\hat{v}$	0.549	0.037	(0.480,0.624)	0.555	0.037	(0.485, 0.630)

Summarized results are the point estimates (Point), the standard errors (SE), and the 95% credible interval for all the regression parameters and the frailty variance parameter v.

As shown in Table 3.3, the estimation of regression coefficients from both methods are accordance with each other. The result indicates that the mean number of patent applied by firms in the scientific sector is 0.7 times higher than firms that not in the scientific sector. At the same time, a firm's book value and its R & D spending have significant positive effect to the patenting development. In Figure 3.1, we superimposed the estimated baseline mean functions of patent counts between 1970 and 1979

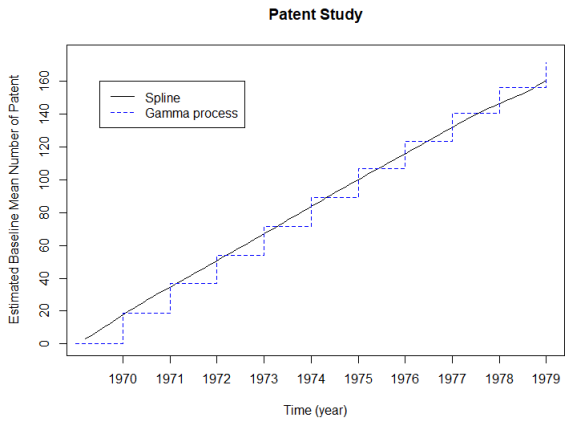

Figure 3.1: Estimate of the baseline mean function for the patent study.

obtained by both methods. The two lines are very close to each other, which implies the proposed method provides a similar estimation of the baseline mean function to GFNPM.

3.7.2 The bladder tumor study

We also apply the proposed method to the most widely used panel count data example in the literature, which arose from a bladder cancer study conducted by the Veterans Administration Cooperative Urological Research Group [Byar and Blackard, 1977]. In this randomized clinical trial study, all the 118 patients had experienced superficial bladder tumors when they entered the trial. They were randomized into one of three treatment groups: placebo, thiotepa, and pyridoxine. During the study at each follow-up visit, new tumors since the last visit were counted, measured and then removed transurethrally. The number of follow-up clinical visits and follow-up times vary noticeably from patient to patient. The primary objective of the study is to determine if any treatment could significantly reduce the recurrence of bladder tumor.

This data set has been analyzed extensively using many different approaches in the literature. Following Wellner and Zhang [2007], we focused on 116 patients in the

study, who had at least one follow-up observation after the study enrollment. Let $\mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4})'$ denote the covariate vector for patient i, where x_{i1} and x_{i2} represent the number of bladder tumors and the size of the largest bladder tumors for patient i at the beginning of the trial, and x_{i3} and x_{i4} are the binary variables indicate whether patient i was assigned to the treatment of pyridoxine pills or thiotepa installation, respectively. When applying the proposed method, we use 20 equally-spaced knots within the data range $0 - 64$ months for the monotone spline specification.

Table 3.4: Bladder tumor data analysis from the proposed approach (GFGP), the GFNPM approach and the WZ approach in Wellner and Zhang [2007].

	GFGP			GFNPM			WZ		
	Point	SE	CI95	Point	SE	CI95	Point	SE	CI95
$\hat{\beta}_1$	0.333	0.107	(0.131,0.550)	0.336	0.106	(0.128,0.544)	0.207	0.078	(0.054,0.360)
$\hat{\beta}_2$	0.001	0.122	(-0.224,0.244)	0.012	0.120	(-0.223,0.247)	-0.036	0.086	(-0.133,0.205)
$\hat{\beta}_3$	-0.021	0.427	(-0.851,0.833)	-0.033	0.409	(-0.835,0.769)	0.066	0.431	(-0.779,0.911)
$\hat{\beta}_4$	-1.152	0.427	(-2.051,-0.261)	-1.140	0.435	(-1.993,-0.287)	-0.797	0.360	(0.091,1.503)
$\hat{v}$	0.326	0.058	(0.225, 0.453)	0.351	0.062	(0.229,0.473)	-	-	-

Table 3.4 shows the results from the proposed approach and two other competitive approaches, i.e. Yao et al. [2016] and Wellner and Zhang [2007]. The results from these two competitors are directly drawn from their papers. Both of these two competitive approaches are likelihood-based approaches under the non-homogeneous Poisson model. Yao et al. [2016]'s method considered the within-subject correlation while Wellner and Zhang [2007]'s method did not consider the within-subject correlation.

As seen in Table 3.4, the results from our method indicates that the number of initial bladder tumors was positively related to the recurrence of the tumor while the size of the largest tumor at the enrollment did not have a significant effect. It also reveals that the thiotepa instillation treatment significantly reduced the recurrence rate of bladder tumors, while the treatment of pyridoxine pills did not have a significant effect. Figure 3.2 plots the estimated mean functions of bladder tumor counts for the control and the other two treatment groups. It is clear that the estimated mean functions for the control and the pyridoxine treatment groups are close to each other and they are higher than the one for the thiotepa treatment group. These conclusions are consistent with those made in Wellner and Zhang [2007] and more close to Yao et al. [2016] in terms of regression coefficients estimation and baseline mean function estimation. This is because Yao et al.'s method also accounts for the within-subject correlation.For this data, the within-subject correlation is not ignorable because the tumor number at baseline is positively related to the recurrence of bladder tumor [Hua and Zhang, 2012].

3.7.3 Breast cosmesis data

To illustrate the proposed method on interval censored data, we applied it to analysis the commonly used interval-censored breast cosmesis data set of Finkelstein and Wolfe [1985]. The data come from a study of 94 early breast cancer patients who

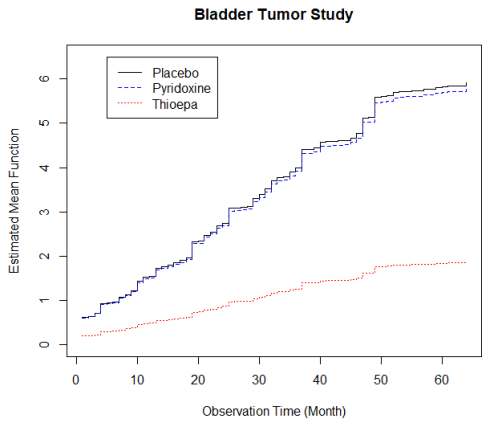

Figure 3.2: Plot for the bladder tumor study.

were treated with adjuvant therapy following tumorectomy. The primary goal of this study is to identify if treating patients with primary radiation therapy and adjuvant chemotherapy has better long-term cosmetic results than treating with radiotherapy alone. Patients in this study were divided into two groups, 48 patients in one group were treated with radiation therapy and chemotherapy, 46 patients in the other group were treated with radiation therapy alone. The response is time (in months) until the appearance of breast retraction. The data are interval-censored between the last clinic visit before the event and the first visit when the event was observed (or Inf if the event was not observed).

We applied the proposed Bayesian approach (GFGP) and compared it with the results obtained by fitting the semi-parametric Turnbull model with a modified ICM algorithm. As shown in Table 3.5, under PH model both methods suggests combining radiation therapy with chemotherapy has positive affect on patient's survival. Under PH and PO models, the estimate of survival functions for two treatments are illustrated in Figure 3.3.

Table 3.5: Breast cosmesis data analysis from the proposed approach (GFGP) and *ic_sp*. Summarized results are the point estimates (Point), the standard errors (SE), and the 95% credible interval for all the regression parameters and the frailty variance parameter v.

	GFGP			*ic_sp*		
Model	Estimate	SE	CI95	Estimate	SE	CI95
PH	0.814	0.232	(0.340,1.257)	0.797	0.345	(0.121,1.473)
PO	0.798	0.337	(0.142,1.469)	0.902	0.406	(0.106,1.698)

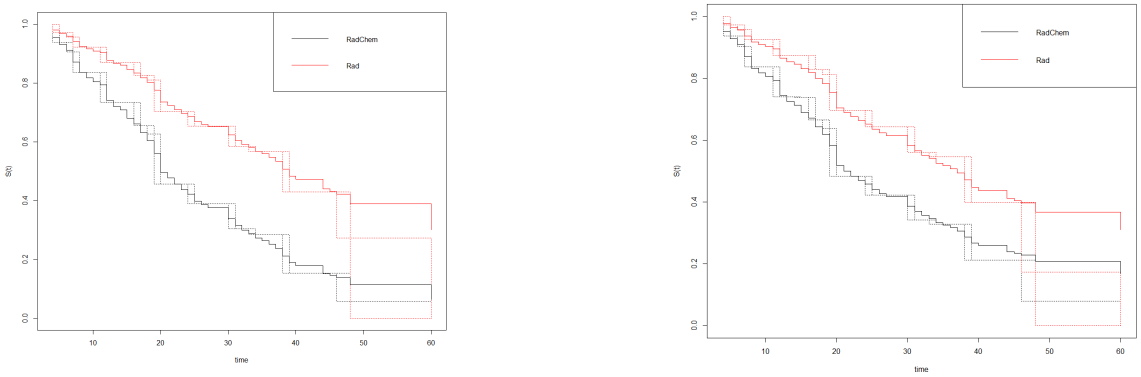

Figure 3.3: The estimated survival functions obtained from the proposed approach (GFGP) and *ic_sp* under PH model (left) and PO model (right).

The results from the two methods are shown in Table 3.5. All the confidence intervals under PH and PO model indicate that chemotherapy increase the hazards and odds of breast retraction for patients who have been previously treated with radiotherapy. This conclusion is in accordance with Finkelstein and Wolfe [1985]'s analysis. Figure 3.3 exhibits the estimated survival functions of appearance of retraction under PH and PO models. Under both models, treating patients with radiotherapy alone has higher survival rate than treatment group.

3.8 Discussion

This article introduces a Bayesian approach which can fit the gamma frailty non-homogeneous Poisson process model for panel count data and survival data. The

approach decomposed each observation time interval using the idea of innermost interval, which results in the ability of processing continuous observation and missing data. The method estimates the baseline mean function nonparametrically by adopting gamma process prior, which allows us to develop a straightforward and easy to implement Gibbs sampler. This approach have appealing numerical performance in terms of providing efficient, accurate and reliable estimation of regression coefficients and baseline mean function. Additionally, because of the intrinsic connection between panel count data and interval censored data this framework is further extended to fit interval censored data under PO and PH model without introducing any complexity of the optimization problem. Because of using gamma process as prior, one limitation of our approach is that the realization $\Lambda(t)$ is discrete with probability one. We hope to tackle this potential issue in the future research.

Bibliography

A. Alioum and D. Commenges. A proportional hazards model for arbitrarily censored and truncated data. *Biometrics*, 52:512–524, 1996.

C. Anderson-Bergman. icenreg: Regression models for interval censored data in r. *Journal of Statistical Software*, 81, 2017.

S. Bennett. Analysis of survival data by the proportional odds model. *Statistics in Medicine*, 2:273–277, 1983a.

D. Blackwell. The discreteness of ferguson selections. *Annals of Statistics*, 1:356–358, 1973.

D.P. Byar and C. Blackard. Comparisons of placebo, pyridoxine, and topical thiotepa in preventing recurrence of stage i bladder cancer. *The Veterans Administration Cooperative Urological Research Group*, 10:556–561, 1977.

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

D. R. Cox. Regression models and life tables. *Journal of the Royal Statistical Society*, 34:187–220, 1972.

D.R. Cox. Some remarks on overdispersion. *Biometrika*, 70(1):269–274, 1983.

K. Doksum. Tailfree and neutral random probabilities and their posterior distributions. *The Annals of Probability*, 2:183–201, 1974.

S. T. Ferguson and G. E. Phadia. Bayesian nonparametric estimation based on censored data. *The Annals of Statistics*, 1:163–186, 1979.

T. S. Ferguson. A bayesian analysis of some nonparametric problems. *Annals of Statistics*, 1:209–230, 1973.

D.M. Finkelstein and R.A. Wolfe. A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics*, 41:731–740., 1985.

M. D. Finkelstein. A proportional hazards model for interval-censored failure time data. *Biometrics*, 42:845–854, 1986.

S. Geman and D. Geman. Stochastic relaxiation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

J. J. Goedert, C.M. Kessler, L.M. Aledort, R. J. Biggar, W. A. Andes, G. C. White 2nd, J.E. Drummond, K. Vaidya, D. L. Mann, M. E. Eyster, and et al. A prospective study of human immunodeficiency virus type 1 infection and the development of aids in subjects with hemophilia. *N Engl J Med*, 321(17):1141–1148, 1989.

P. Groeneboom and J. A. Wellner. Information bounds and nonparametric maximum likelihood estimation. *Birkhauser*, page Basel, 1992.

B. Hall, G. Zvi, and H. Jerry. Patents and r and d: Is there a lag? *International Economic Review*, 27:265–283, 1986.

A. B. Hamburg, C.H. Kraemer, and W. Jahnke. A hierarchy of drug use in adolescence: Behavioral and attitudinal correlates of substantial drug use. *Am J Psychiatry*, 132:1155–1163, 1975.

X. J. Hu, S. W. Lagakos, and R. A. Lockhart. Marginal analysis of panel counts through estimating functions. *Biometrika*, 96:445–456, 2009.

L. Hua and Y. Zhang. Spline-based semiparametric projected generalized estimating equation method for panel count data. *Biostatistics*, 13:440–454, 2012.

L. Hua, Y. Zhang, and W. Tu. A spline-based semiparametric sieve likelihood method for over-dispersed panel count data. *The Canadian Journal of Statistics*, 42:217–245, 2014.

C. Huang, M.C. Wang, and Y. Zhang. Analysing panel count data with informative observation times. *Biometrika*, 93:763–775, 2006.

J. Huang. Effcient estimation for the cox model with interval censoring. *International Economic Review*, 24:540–568, 1996.

J. Huang and J.A. Rossini. Sieve estimation for the proportional odds model with interval-censoring. *Journal of American Statistical Association*, 92:960–967, 1997.

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

D. J. Kalbfleisch. Non-parametric bayesian analysis of survival time data. *Journal of the Royal Statistical Society*, 40:214–221, 1978.

B. L. Kroner, P.S. Rosenberg, L. M. Aledort, W. G. Alvord, and J. J. Geodert. Hiv-1 infection incidence among persons with hemophilia in the united states and western europe, 1978-1990. *Journal of Acquired Immune Deficiency Syndromes*, 7:279–286, 1994.

J. Lawless and M. Crowder. Covariates and random effects in a gamma process model with application to degradation and failure. *Lifetime Data Analysis*, 10:213–227, 2004.

L. Li, T. Watkins, and Q. Yu. An em algorithm for smoothing the selfconsistent estimator of survival functions with interval-censored data. *Scandinavian Journal of Statistics*, 24:531–542, 1997.

X. Lin and L. Wang. A semiparametric probit model for case 2 interval-censored failure time data. *Statistics in Medicine*, 29:972–981, 2010.

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

X. Lin, B. Cai, L. Wang, and Z. Zhang. A bayesian proportional hazards model for general interval-censored data. *Lifetime Data Analysis*, 21:470–490, 2015.

T. Louis. Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society*, 44:226–233, 1982.

M. Lu, Y. Zhang, and J. Huang. Estimation of the mean function with panel count data using monotone polynomial splines. *Biometrika*, 94:1060–1070, 2007.

M. Lu, Y. Zhang, and J. Huang. Semiparametric estimation methods for panel count data using monotone b-splines. *Journal of the American Statistical Association*, 104:1060–1070, 2009.

N.S. Maceachern and P. Muller. Estimating mixture of dirichlet process models. *Journal of Computational and Graphical Statistics*, 7(2):223–238, 1998.

O. M. S. Manda. A nonparametric frailty model for clustered survival data. *Communications in Statistics-Theory and Methods*, 40:863–875, 2011.

C.S. McMahan, L. Wang, and J.M. Tebbs. Regression analysis for current status data using the em algorithm. *Statistics in Medicine*, 32:4452–4466, 2013.

S. A. Murphy. Asymptotic theory for the frailty model. *Annals of Statistics*, 23: 182–198, 1995.

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

M. Naskar. Semiparametric analysis of clustered survival data under nonparametric frailty. *Statistica Neerlandica*, 62:155–172, 2008.

M. Naskar and K. Das. Semiparametric analysis of two-level bivariate binary data. *Biometrics*, 62:1004–1013, 2006.

G. G. Nielsen, R. D. Gill, P. K. Andersen, and T. I. A. Sorensen. A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics*, 19:25–44, 1992.

D. S. Nozer. Survival in dynamic envirnoments. *Statistical Science*, 10(1):86–103, 1995.

W. Pan. Extending the iterative convex minorant algorithm to the cox model for interval-censored data. *Journal of Computational and Graphical Statistics*, 8(1): 109–120, 1999.

W. Pan and Rick Chappell. A nonparametric estimator of survival functions for arbitrarily truncated and censored data. *Lifetime Data Analysis*, 4:187–202, 1998.

T. Park and G Casella. The bayesian lasso. *Royal Statistical Society*, 103:681–686, 2008.

L. M. Pennell and B. D. David B. Dunson. Bayesian semiparametric dynamic frailty models for multiple event time data. *Biometrics*, 62:1044–1052, 2006.

R. Peto. Experimental survival curves for interval-censored data. *Applied Statistics*, 22:86–91, 1973.

J. O. Ramsay. Monotone regression splines in action. *Statistical Science*, 3:425–461, 1988.

C. P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer, New York, 2004.

P. S. Rosenberg. Hazard function estimation using b-splines. *Biometrics*, 51:874–887, 1995.

A. J. Rossini and A. A. Tsiatis. A semiparametric proportional odds regression model for the analysis of current status data. *Journal of the American Statistical Association*, 91:713–721, 1996.

P. Royston and M. K. B. Parmar. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine*, 21: 2175–2197, 2002.

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

X. T. Shen. Proportional odds regression and sieve maximum likelihood estimation. *Biometrika*, 85:165–177, 1998.

A. Sinha, Z. Chi, and M. Chen. Bayesian inference of hidden gamma wear process model for survival data with ties. *Statistical Sinica*, 25:1613–1635, 2015.

D. Sinha. A bayesian approach for the analysis of panel-count data with dependent termination. *Biometrics*, 60:34–40, 2004.

J. Sun. The statistical analysis of interval-censored data. *Springer*, 2006.

J. Sun and J. D Kalbfleisch. Estimation of the mean function of point processes based on panel count data. *Statistica Sinica*, 5:279–290, 1995.

J. Sun and L. J. Wei. Regression analysis of panel count data with covariate-dependent observation and censoring times. *Royal Statistical Society*, 62:293–302, 2000.

X. M. Tu, X.L. Meng, and M. Pagano. The aids epidemic: Estimating survival after aids diagnosis from surveillance data. *Journal of the American Statistical Association*, 88:26–36, 1993.

B. Turnbull. Nonparametric estimation of a survivorship function with doubly censored data. *Journal of the American Statistical Association*, 69:169–173, 1974.

B. W. Turnbull. The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society, Series B*, 38: 290–295, 1976.

B. W. Turnbull and L. Weiss. A likelihood ratio statistic for testing goodness of fit with randomly censored data. *Biometrics*, 34:367–375, 1978.

Ulku Uzunog-Ullari and Jane-Ling Wang. A comparison of hazard rate estimators for left truncated and right censored data. *Biometrika*, 79(2):297–310, 06 1992.

C.S. Wang, L.and McMahan, M.G. Hudgens, and Z.P. Qureshi. A flexible, computationally efficient method for fitting the proportional hazards model to interval-censored data. *Biometrics*, 72:222–231, 2016.

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

Mei-Cheng Wang, Nicholas P. Jewell, and Wei-Yann Tsai. Asymptotic properties of the product limit estimate under random truncation. *The Annals of Statistics*, 14(4):1597–1605, 1986.

X. Wang. Nonparametric estimation of the shape function in a gamma process for degradation data. *The Canadian Journal of Statistics*, 37(1):102–118, 2009.

A.J Wellner and Y. Zhang. Two estimators of the mean of a counting process with panel count data. *The Annals of Statistics*, 28(3):779–814, 2000.

A.J Wellner and Y. Zhang. Two likelihood-based semiparametric estimation methods for panel count data with covariates. *The Annals of Statistics*, 35(5):2105–2142, 2007.

Jon A. Wellner and Ying Zhang. Large sample theory for an estimator of the mean of a counting process with panel count data, 1998.

D. Xu, H. Zhao, and J. Sun. Joint analysis of interval-censored failure time data and panel count data. *Lifetime Data Analysis*, 24:94–109, 2018.

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

B. Yao, L. Wang, and X. He. Semiparametric regression analysis of panel count data allowing for within-subject correlation. *Computational Statistics and Data Analysis*, 97:47–59, 2016.

Y. Zhang and M. Jamshidian. The gamma-frailty poisson model for the nonparametric estimation of panel count data. *Biometrics*, 59:1099–1106, 2003.

Q. Zhao and J. Sun. glrt: Generalized logrank tests for interval-censored failure time data. *R package version 2.0*, pages URL https://CRAN.R–project.org/package=glrt, 2015.

L. Zhu, Y. Zhang, Y. Li, J Sun, and L. L. Robison. A semiparametric likelihood-based method for regression analysis of mixed panel-count data. *Biometrics*, 74: 488–497, 2018.

Appendix A

Asymptotic Covariance Matrix Estimation

Once applied the four-step data augmentation as shown in the paper, the complete likelihood has the form:

$$\begin{aligned} L_c(\theta) \propto \prod_{i=1}^{n} \Big(& \exp(\mathbf{x}_i'\boldsymbol{\beta}\delta_{i0}) \exp(-\psi_i\delta_{i0}) \exp(\phi_i) \\ & \times \Big[\prod_{l=1}^{K} \{\gamma_l M_l(t_{i1})\}^{u_{il}} \exp\{-\gamma_l b_l(t_{i1}) \exp(\mathbf{x}_i'\boldsymbol{\beta})\psi_i\delta_{i0}\} \\ & \times \{\gamma_l b_l(t_{i1}) \exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i\}^{z_{il}\delta_{i1}} \exp\{-\gamma_l b_l(t_{i1}) \exp(x'\boldsymbol{\beta})\phi_i\} \\ & \times \{\gamma_l (b_l(t_{i2}) - b_l(t_{i1})) \exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i\}^{w_{il}\delta_{i2}} \\ & \times \exp\{-\gamma_l \{b_l(t_{i2}) - b_l(t_{i1})\} \exp(\mathbf{x}_i'\boldsymbol{\beta})\phi_i(\delta_{i2} + \delta_{i3})\} \Big] \Big), \end{aligned} \tag{A.1}$$

where

$$\begin{aligned} t_{i1} &= L_i I(\delta_{i0} = 1) + R_i I(\delta_{i1} = 1) + L_i I(\delta_{i2} = 1) + L_i^- I(\delta_{i3} = 1), \\ t_{i2} &= L_i^+ I(\delta_{i0} = 1) + R_i^+ I(\delta_{i1} = 1) + R_i I(\delta_{i2} = 1) + L_i I(\delta_{i3} = 1). \end{aligned} \tag{A.2}$$

Subject to the constrain that

$$\sum_{l=1}^{K} Z_{il} = Z_i = \begin{cases} 0 & \text{when} \quad \delta_{i0} = 1 \quad \text{or} \quad \delta_{i2} = 1 \quad \text{or} \quad \delta_{i3} = 1 \\ > 0 & \text{when} \quad \delta_{i2} = 1 \end{cases} \tag{A.3}$$

$$\sum_{l=1}^{K} W_{il} = W_i = \begin{cases} 0 & \text{when} \quad \delta_{i3} = 1 \\ > 0 & \text{when} \quad \delta_{i2} = 1 \end{cases} \tag{A.4}$$

Under regularity conditions, as $n \to \infty$, $\hat{\theta} \overset{a}{\sim} N(\theta, \{I(\theta)\}^{-1})$. Basing on Louis's method:

$$I(\theta) = -\frac{\partial^2 Q(\theta, \hat{\theta})}{\partial\theta\partial\theta'} - \text{var}\{\frac{\partial log L_c(\theta)}{\partial\theta} | \mathcal{D}, \hat{\theta}\} \tag{A.5}$$

Items on the right-hand side:

$$\frac{\partial^2 Q(\theta,\hat{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} = -\sum_{i=1}^{n}\sum_{l=1}^{K}\Big\{\Big[E(\psi_i|\mathcal{D},\hat{\theta})b_l(L_i) + E(\phi_i|\mathcal{D},\hat{\theta})\{b_l(L_i)(\delta_{i0}+\delta_{i3}) + b_l(R_i)(\delta_{i1}+\delta_{i2})\}\Big]\exp(\mathbf{x}_i'\boldsymbol{\beta})\gamma_l\mathbf{x}_i\mathbf{x}_i'\Big\} \tag{A.6}$$

$$\frac{\partial^2 Q(\theta,\hat{\theta})}{\partial\boldsymbol{\beta}\partial\gamma_l} = -\sum_{i=1}^{n}\Big(\Big[E(\psi_i|\mathcal{D},\hat{\theta})b_l(L_i) + E(\phi_i|\mathcal{D},\hat{\theta})\{b_l(L_i)(\delta_{i0}+\delta_{i3}) + b_l(R_i)(\delta_{i1}+\delta_{i2})\}\Big]\exp(\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i\Big) \tag{A.7}$$

$$\frac{\partial^2 Q(\theta,\hat{\theta})}{\partial\gamma_l\partial\gamma_l} = -\sum_{i=1}^{n}\frac{E(U_{il}|\mathcal{D},\hat{\theta}) + E(Z_{il}|\mathcal{D},\hat{\theta}) + E(W_{il}|\mathcal{D},\hat{\theta})}{\gamma_l^{(d)2}} \tag{A.8}$$

$$\frac{\partial^2 Q(\theta,\hat{\theta})}{\partial\gamma_l\partial\gamma_k} = 0, k \neq l$$

$$\begin{aligned}\mathrm{var}\Big(\frac{\partial log L_c(\theta)}{\partial\boldsymbol{\beta}}|\mathcal{D},\hat{\theta}\Big) = \sum_{i=1}^{n}\Big[&\{\Lambda_0(L_i)\delta_{i0}\exp(\mathbf{x}_i'\boldsymbol{\beta})\}^2\mathrm{var}(\psi_i)\\ &+\{\Lambda_0(L_i)(\delta_{i0}+\delta_{i3})+\Lambda_0(R_i)(\delta_{i1}+\delta_{i2})\}^2 e^{2\mathbf{x}_i'\beta}\mathrm{var}(\phi_i)\\ &+var(Z_i)\delta_{i1}+var(W_i)\delta_{i2}-2\mathrm{cov}(\phi_i,Z_i)\delta_{i1}\Lambda_0(R_i)e^{\mathbf{x}_i'\beta}\\ &-2\mathrm{cov}(\phi_i,W_i)\delta_{i2}\Lambda_0(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})\Big]\mathbf{x}_i\mathbf{x}_i'\end{aligned} \tag{A.9}$$

$$\begin{aligned}\mathrm{var}\Big(\frac{\partial log L_c(\theta)}{\partial\gamma_l}|\mathcal{D},\hat{\theta}\Big) = \sum_{i=1}^{n}\Big(&\frac{\mathrm{var}(U_{il})\delta_{i0}+\mathrm{var}(W_{il})\delta_{i2}+\mathrm{var}(Z_{il})\delta_{i1}}{\gamma_l^2}\\ &+\Big[\delta_{i0}b_l(L_i)^2\mathrm{var}(\psi_i)+\{b_l(L_i)(\delta_{i0}+\delta_{i3})+b_l(R_i)(\delta_{i1}+\delta_{i2})\}^2\mathrm{var}(\phi_i)\Big]e^{2\mathbf{x}_i'\beta}\\ &-2\frac{\delta_{i2}\exp(\mathbf{x}'\boldsymbol{\beta})b_l(R_i)}{\gamma_l}\mathrm{cov}(W_{il},\phi_i)\\ &-2\frac{\delta_{i1}\exp(\mathbf{x}'\boldsymbol{\beta})b_l(R_i)}{\gamma_l}\mathrm{cov}(Z_{il},\phi_i)\Big)\end{aligned} \tag{A.10}$$

$$\begin{aligned}\operatorname{cov}(\frac{\partial logL_c(\theta)}{\partial \boldsymbol{\beta}}, \frac{\partial logL_c(\theta)}{\partial \gamma_l}|\mathcal{D}, \hat{\theta}) = \sum_{i=1}^{n} \Big[& \Lambda_0(L_i)\exp(2\mathbf{x}_i\boldsymbol{\beta})b_l(L_i)\delta_{i0}\operatorname{var}(\psi_i) \\ & + \{\Lambda_0(L_i)(\delta_{i0}+\delta_{i3}) + \Lambda_0(R_i)(\delta_{i1}+\delta_{i2})\}\{b_l(L_i)(\delta_{i0}+\delta_{i3}) + b_l(R_i)(\delta_{i1}+\delta_{i2})\} \\ & \times \exp(2\mathbf{x}_i'\boldsymbol{\beta})\operatorname{var}(\phi_i) + \frac{\delta_{i1}\operatorname{cov}(Z_i, Z_{il})}{\gamma_l} + \frac{\delta_{i2}\operatorname{cov}(W_i, W_{il})}{\gamma_l} \\ & - \frac{\delta_{i2}\Lambda_0(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})\operatorname{cov}(\phi_i, W_{il})}{\gamma_l} - \frac{\delta_{i1}\Lambda_0(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})\operatorname{cov}(\phi_i, Z_{il})}{\gamma_l} \\ & - \exp(\mathbf{x}_i'\boldsymbol{\beta})b_l(R_i)\{\operatorname{cov}(Z_i, \phi_i)\delta_{i1} + \operatorname{cov}(W_i, \phi_i)\delta_{i2}\}\Big]\mathbf{x}_i \end{aligned} \tag{A.11}$$

$$\begin{aligned}\operatorname{cov}(\frac{\partial logL_c(\boldsymbol{\theta})}{\partial \gamma_l}, \frac{\partial logL_c(\boldsymbol{\theta})}{\partial \gamma_k}|\mathcal{D}, \hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \Big(& \frac{\operatorname{cov}(U_{il}, U_{ik})\delta_{i0} + \operatorname{cov}(W_{il}, W_{ik})\delta_{i2} + \operatorname{cov}(Z_{il}, Z_{ik})\delta_{i1}}{\gamma_l\gamma_k} \\ & + \Big[b_l(L_i)b_k(L_i)\operatorname{var}(\psi_i)\delta_{i0} + \{b_l(L_i)(\delta_{i0}+\delta_{i3}) + b_l(R_i)(\delta_{i1}+\delta_{i2})\} \\ & \times \{b_k(L_i)(\delta_{i0}+\delta_{i3}) + b_k(R_i)(\delta_{i1}+\delta_{i2})\}\operatorname{var}(\phi_i)\Big]\exp(2\mathbf{x}_i'\boldsymbol{\beta}) \\ & - \frac{b_k(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})}{\gamma_l}\{\operatorname{cov}(Z_{il}, \phi_i)\delta_{i1} + \operatorname{cov}(W_{il}, \phi_i)\delta_{i2}\} \\ & - \frac{b_l(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta})}{\gamma_k}\{\operatorname{cov}(Z_{ik}, \phi_i)\delta_{i1} + \operatorname{cov}(W_{ik}, \phi_i)\delta_{i2}\}\Big) \end{aligned} \tag{A.12}$$

All the conditional expectations, variance and covariance:

When $\delta_{i0} = 1$

$$E(\psi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) = \frac{1}{\Lambda_0^{(d)}(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)}) + 1}$$

$$\operatorname{var}(\psi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) = E(\psi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)})^2$$

$$\operatorname{var}(U_{il}|\mathcal{D}, \boldsymbol{\theta}^{(d)}) = \frac{\gamma_l^{(d)}M_l(L_i)}{\Lambda_0'^{(d)}(L_i)}\{1 - \frac{\gamma_l^{(d)}M_l(L_i)}{\Lambda_0'^{(d)}(L_i)}\}$$

When $\delta_{i1} = 1$

$$E(Z_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) = \Lambda_0^{(d)}(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)}) + 1$$

$$\operatorname{var}(Z_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) = \{\Lambda_0^{(d)}(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)}) + 1\}\Lambda_0^{(d)}(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)})$$

$$\text{var}(Z_{il}|\mathcal{D},\boldsymbol{\theta}^{(d)}) = E(Z_i|\mathcal{D},\boldsymbol{\theta}^{(d)})\frac{\gamma_l^{(d)}b_l(R_i)}{\Lambda_0^{(d)}(R_i)}\{1-\frac{\gamma_l^{(d)}b_l(R_i)}{\Lambda_0^{(d)}(R_i)}\}+\text{var}(Z_i|\mathcal{D},\boldsymbol{\theta}^{(d)})\{\frac{\gamma_l^{(d)}b_l(R_i)}{\Lambda_0^{(d)}(R_i)}\}^2$$

$$\text{cov}(Z_i, Z_{il}|\mathcal{D},\boldsymbol{\theta}^{(d)}) = \frac{b_l(R_i)\mathbf{x}_i'\boldsymbol{\beta}^{(d)}}{\Lambda_0^{(d)}(R_i)}\text{var}(Z_i|\mathcal{D},\boldsymbol{\theta}^{(d)})$$

$$\text{cov}(Z_{il}, Z_{ik}|\mathcal{D},\boldsymbol{\theta}^{(d)}) = \frac{\gamma_l^{(d)}\gamma_k^{(d)}b_l(R_i)b_k(R_i)}{\{\Lambda_0^{(d)}(R_i)\}^2}\{\text{var}(Z_i) - E(Z_i)\}$$

$$\text{cov}(\phi_i, Z_i|\mathcal{D},\boldsymbol{\theta}^{(d)}) = \Lambda_0^{(d)}(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)})$$

$$\text{cov}(\phi_i, Z_{il}|\mathcal{D},\boldsymbol{\theta}^{(d)}) = \gamma_l^{(d)}b_l(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)})$$

When $\delta_{i2} = 1$

$$E(W_i|\mathcal{D},\boldsymbol{\theta}^{(d)}) = \frac{\Lambda_0^{(d)}(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)}) + 1}{\Lambda_0^{(d)}(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)}) + 1}$$

$$\text{var}(W_i|\mathcal{D},\boldsymbol{\theta}^{(d)}) = E(W_i|\mathcal{D},\boldsymbol{\theta}^{(d)})\{E(W_i|\mathcal{D},\boldsymbol{\theta}^{(d)}) - 1\}$$

$$\begin{aligned}\text{var}(W_{il}|\mathcal{D},\boldsymbol{\theta}^{(d)}) = {} & E(W_i|\mathcal{D},\boldsymbol{\theta}^{(d)})\{1 - \frac{\gamma_l^{(d)}[b_l(R_i) - b_l(L_i)]}{\Lambda_0^{(d)}(R_i) - \Lambda_0^{(d)}(L_i)}\} + \text{var}(W_i|\mathcal{D},\boldsymbol{\theta}^{(d)}) \\ & \times [\frac{\gamma_l^{(d)}\{b_l(R_i) - b_l(L_i)\}}{\Lambda_0^{(d)}(R_i) - \Lambda_0^{(d)}(L_i)}]^2\end{aligned}$$

$$\text{cov}(W_i, W_{il}|\mathcal{D},\boldsymbol{\theta}^{(d)}) = \text{var}(W_i|\mathcal{D},\boldsymbol{\theta}^{(d)})\frac{\gamma_l^{(d)}\{b_l(R_i) - b_l(L_i)\}}{\Lambda_0^{(d)}(R_i) - \Lambda_0^{(d)}(L_i)}$$

$$\begin{aligned}\text{cov}(W_{il}, W_{ik}|\mathcal{D},\boldsymbol{\theta}^{(d)}) = {} & \frac{\gamma_l^{(d)}\gamma_k^{(d)}\{b_l(R_i) - b_l(L_i)\}\{b_k(R_i) - b_k(L_i)\}}{\{\Lambda_0^{(d)}(R_i) - \Lambda_0^{(d)}(L_i)\}^2} \\ & \times \{\text{var}(W_i|\mathcal{D},\boldsymbol{\theta}^{(d)}) - E(W_i|\mathcal{D},\boldsymbol{\theta}^{(d)})\}\end{aligned} \tag{A.13}$$

$$\text{cov}(W_i, \phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) = \frac{\{\Lambda_0^{(d)}(R_i) - \Lambda_0^{(d)}(L_i)\}\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)})}{\{\Lambda_0^{(d)}(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)}) + 1\}^2}$$

$$\text{cov}(W_{il}, \phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) = \frac{\gamma_l^{(d)}\{b_l(R_i) - b_l(L_i)\}\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)})}{\{\Lambda_0^{(d)}(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)}) + 1\}^2}$$

For ϕ_i:

a. When $\delta_{i0} = 1$ or $\delta_{i3} = 1$,

$$\text{var}(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) = \{\frac{1}{\Lambda_0^{(d)}(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)}) + 1}\}^2$$

b. When $\delta_{i1} = 1$,

$$\text{var}(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) = 1 + \{\frac{1}{\Lambda_0^{(d)}(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)}) + 1}\}^2$$

c. When $\delta_{i2} = 1$,

$$\text{var}(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) = \{\frac{1}{\Lambda_0^{(d)}(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)}) + 1}\}^2 + \{\frac{1}{\Lambda_0^{(d)}(R_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)}) + 1}\}^2$$

$$\text{var}(\psi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) = \{\frac{1}{\Lambda_0^{(d)}(L_i)\exp(\mathbf{x}_i'\boldsymbol{\beta}^{(d)}) + 1}\}^2$$

Appendix B

Proof that $\boldsymbol{\theta}^{d+1}$ is a unique global maximizer

In each iteration of the EM algorithm, the M-step seeks the maximizer of $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$, i.e., $\boldsymbol{\theta}^{(d+1)} = (\boldsymbol{\beta}^{(d+1)'}, \boldsymbol{\gamma}^{(d+1)'})' = \arg\max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$. This maximization can be accomplished through a two step procedure. Firstly, get $\boldsymbol{\gamma}^{(d)} = \arg\max_{\gamma} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$. The maximizer $\boldsymbol{\gamma}^{(d)}(\boldsymbol{\beta})$ can be obtained as the solution to the system of equations $\partial\mathcal{Q}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}^{(d)})/\partial\boldsymbol{\gamma} = 0$ and can be expressed in the form of equation (15). Because $\mathcal{Q}^2(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}^{(d)})/\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'$, the Hessian matrix of $\mathcal{Q}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}^{(d)})$, is a diagonal matrix with the lth diagonal element takes the form in equation (24) in Appendix A. We can easily verify it is negative definite for all $\boldsymbol{\gamma}$ by noting $\gamma_1 > 0$, for $l = 1, ..., K$, so that $\mathcal{Q}^2(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}^{(d)})/\partial\gamma_l^2$ is strictly less than 0. In this way, we can see for each value of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}^{(d)}(\boldsymbol{\beta})$ is the unique maximizer. Secondly, we need to get $\boldsymbol{\beta}^{(d+1)} = \arg\max_{\boldsymbol{\beta}} \mathcal{Q}(\boldsymbol{\beta}, \boldsymbol{\gamma}^{(d)}(\boldsymbol{\beta}), \boldsymbol{\theta}^{(d)})$. Again, we can show $\boldsymbol{\beta}^{(d+1)}$ is the unique maximizer of $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\gamma}^{(d)}(\boldsymbol{\beta}), \boldsymbol{\theta}^{(d)})$ by showing the Hessian matrix of $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\gamma}^{(d)}(\boldsymbol{\beta}), \boldsymbol{\theta}^{(d)})$ is negative definite for all $\boldsymbol{\beta}$.

For notational convenience, let's first define

$$A_{il} = E(U_{il}|\mathcal{D}, \boldsymbol{\theta}^{(d)})\delta_{i0} + E(Z_{il}|\mathcal{D}, \boldsymbol{\theta}^{d})\delta_{i1} + E(W_{il}|\mathcal{D}, \boldsymbol{\theta}^{(d)})\delta_{i2},$$

$$B_{il} = E(\psi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)})\delta_{i0}b_l(L_i) + E(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)})\{b_l(L_i)(\delta_{i0} + \delta_{i3}) + b_l(R_i)(\delta_{i1} + \delta_{i2})\},$$

$$C_{il} = E(Z_{il}|\mathcal{D}, \boldsymbol{\theta}^{(d)})\delta_{i1} + E(W_{il}|\mathcal{D}, \boldsymbol{\theta}^{(d)})\delta_{i2}.$$

Based on the previous result, we have

$$\begin{aligned}\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\gamma}^{(d)}(\boldsymbol{\beta}), \boldsymbol{\theta}^{(d)}) = \sum_{i=1}^{n} \Big\{ & \mathbf{x}_i'\boldsymbol{\beta}\delta_{i0} - E(\psi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)})\delta_{i0} - E(\phi_i|\mathcal{D}, \boldsymbol{\theta}^{(d)}) \\ & + \sum_{l=1}^{k} \Big[A_{il} log(\sum_{i=1}^{n} A_{il}) - A_{il} log\{\sum_{i=1}^{n} B_{il} \exp(\mathbf{x}_i'\boldsymbol{\beta})\} \\ & - B_{il} \exp(\mathbf{x}_i'\boldsymbol{\beta}) \frac{\sum_{i=1}^{n} A_{il}}{\sum_{i=1}^{n} B_{il} \exp(\mathbf{x}_i'\boldsymbol{\beta})} + C_{il}\mathbf{x}_i'\boldsymbol{\beta} \Big] \Big\} + g(\boldsymbol{\theta}^d)\end{aligned} \tag{B.1}$$

Then

$$\frac{\partial \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\gamma}^{(d)}(\boldsymbol{\beta}), \boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \delta_{i0}\mathbf{x}_i - \sum_{i=1}^{n}\sum_{l=1}^{k} \frac{\sum_{i=1}^{n} A_{il}\{\sum_{i=1}^{n} B_{il} \exp(\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i\}}{\sum_{i=1}^{n} B_{il} \exp(\mathbf{x}_i'\boldsymbol{\beta})} + \sum_{i=1}^{n}\sum_{l=1}^{k} C_{il}\mathbf{x}_i \tag{B.2}$$

Then if we denote $\sum_{i=1}^{n} A_{il} = A_l$ and $B_{il} \exp(\mathbf{x}_i'\boldsymbol{\beta}) = D_{il}$. The second order derivative of $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\gamma}^{(d)}(\boldsymbol{\beta}), \boldsymbol{\theta}^{(d)})$ with respect to $\boldsymbol{\beta}$ is

$$\frac{\partial^2 \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\gamma}^{(d)}(\boldsymbol{\beta}), \boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = -\sum_{l=1}^{k} A_l \frac{(\sum_{i=1}^{n} D_{il}\mathbf{x}_i\mathbf{x}_i')(\sum_{i=1}^{n} D_{il}) - (D_{il}\mathbf{x}_i)(\sum_{i=1}^{n} D_{il}\mathbf{x}_i')}{(\sum_{i=1}^{n} D_{il})^2}. \tag{B.3}$$

For any $\mathbf{z} \in \mathbb{R}^P$, where p is the dimension of $\boldsymbol{\beta}$. Let $H_l = \frac{A_l}{\{\sum_{i=1}^{n} D_{il}\}^2}$, then

$$\begin{aligned}\mathbf{z}' \frac{\partial^2 \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\gamma}^{(d)}(\boldsymbol{\beta}), \boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \mathbf{z} &= -\sum_{l=1}^{k} H_l \Big[\{\sum_{i=1}^{n} D_{il}\mathbf{z}'\mathbf{x}_i\mathbf{x}_i'\mathbf{z}\}\{\sum_{i=1}^{n} D_{il}\} \\ &\quad - \{\sum_{i=1}^{n} D_{il}\mathbf{z}'\mathbf{x}_i\}\{\sum_{i=1}^{n} D_{il}\mathbf{x}_i'\mathbf{z}\} \Big] \\ &= -\sum_{l=1}^{k} H_l \Big[\{\sum_{i=1}^{n} D_{il}(\sum_{j=1}^{p} z_j x_{ij})^2\}\{\sum_{i=1}^{n} D_{il}\} \\ &\quad - \{\sum_{i=1}^{n} D_{il}(\sum_{j=1}^{p} z_j x_{ij})\}^2 \Big] \\ &= -\sum_{l=1}^{k} H_l \Big[\sum_{i<h} D_{il}D_{hl}\{(\sum_{j=1}^{p} z_j x_{ij})^2 + (\sum_{j=1}^{p} z_j x_{hj})^2\} \\ &\quad - 2\sum_{i<h} D_{il}D_{hl}(\sum_{j=1}^{p} z_j x_{ij})(\sum_{j=1}^{p} z_j x_{hj}) \Big] \\ &= -\sum_{l=1}^{k} H_l \sum_{i<h} D_{il}D_{hl}\{(\sum_{j=1}^{p} z_j x_{ij}) - (\sum_{j=1}^{p} z_j x_{hj})\}^2\end{aligned} \tag{B.4}$$

For nonzero $\mathbf{z}$, $\mathbf{z}' \frac{\partial^2 \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\gamma}^{(d)}(\boldsymbol{\beta}), \boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \mathbf{z} = 0$ only when $\mathbf{z}'(\mathbf{x}_j - \mathbf{x}_h) = 0$ for all $i \neq h$. This only happens when all subjects have the same value for a particular covariate. In

this situation the corresponding regression parameter and Λ_0 are not identifiable. In other words, as long as the model is identifiable, we have $\mathbf{z}' \frac{\partial^2 \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\gamma}^{(d)}(\boldsymbol{\beta}), \boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \mathbf{z} < 0$ for all $\mathbf{z} \in \mathbb{R}^P \{0\}$. This shows that $\frac{\partial^2 \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\gamma}^{(d)}(\boldsymbol{\beta}), \boldsymbol{\theta}^{(d)})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}$ is negative definite. Thus, $\boldsymbol{\beta}^{(d+1)} = \arg\max_{\boldsymbol{\beta}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\gamma}^{(d)}(\boldsymbol{\beta}), \boldsymbol{\theta}^{(d)})$ is the unique maximizer. Consequently, $\boldsymbol{\theta}^{(d+1)}$ as determined by the two step process described in the EM algorithm is the unique maximizer of $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$.

Appendix C

Derivation of the within-subject correlation

Consider two non-overlapping intervals $(t_1, t_2]$ and $(t_3, t_4]$. Let Z_1 and Z_2 denote the counts of recurrent events within these two intervals, respectively, from the same subject with covariates $\mathbf{x}$. The derivation of Pearson's correlation coefficient between Z_1 and Z_2 is shown below under the proposed nonparametric frailty Poisson model. Define $\lambda_1 = \{\mu_0(t_2) - \mu_0(t_1)\}\exp(\mathbf{x}'\boldsymbol{\beta})$ and $\lambda_2 = \{\mu_0(t_4) - \mu_0(t_3)\}\exp(\mathbf{x}'\boldsymbol{\beta})$, it is clear that $Z_1|\phi \sim Poi(\lambda_1\phi)$ and $Z_2|\phi \sim Poi(\lambda_2\phi)$ under the proposed model. Meanwhile, the frailty ϕ follows the distribution $h(\centerdot)$ which is generated by the DP mixture. First, using the law of iterated conditional expectations, one obtains

$$\begin{aligned} \text{cov}(Z_1, Z_2) &= \text{cov}\{E(Z_1|\phi), E(Z_1|\phi)\} + E\{\text{cov}(Z_1, Z_2|\phi)\} \\ &= \text{cov}\{\lambda_1\phi, \lambda_2\phi\} + 0 \\ &= \lambda_1\lambda_2\text{var}(\phi) \end{aligned} \tag{C.1}$$

$$\begin{aligned} \text{var}(Z_1) &= E\{\text{var}(Z_1|\phi)\} + \text{var}\{E(Z_1|\phi)\} \\ &= E\{\lambda_1\phi\} + \text{var}\{\lambda_1\phi\} \\ &= \lambda_1 + \lambda_1^2\text{var}(\phi) \end{aligned} \tag{C.2}$$

Similarly, $\text{var}(Z_2) = \lambda_2 + \lambda_2^2\text{var}(\phi)$. Then the correlation between Z_1 and Z_2 can be written as:

$$\rho(Z_1, Z_2) = \frac{1}{\sqrt{\{1 + \lambda_1^{-1}\text{var}(\phi)^{-1}\}\{1 + \lambda_2^{-1}\text{var}(\phi)^{-1}\}}}.$$

As for var(ϕ) we can use iterated rule again based on the fact that $\phi|v \sim Ga(v, v)$. So we can get:

$$\text{var}(\phi) = E\{\text{var}(\phi|v)\} + \text{var}\{E(\phi|v)\} = E(v^{-1}) + \text{var}(1) = \int v^{-1} d\pi(v). \qquad \text{(C.3)}$$

Using the notation in Chapter 2 it can be further written as $\text{var}(\phi) = \sum_{h=1} \theta_h^{-1} p_h$.