2018

Bayesian Semiparametric Methods for Analyzing Panel Count Data

Jianhong Wang
University of South Carolina

Bayesian Semiparametric Methods for Analyzing Panel Count Data

by

Jianhong Wang

Bachelor of Mathematics
Henan University, China 1997

Master of Applied Mathematics
PLA University of Science and Technology, China 2000

Master of Statistics
University of South Carolina 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

2018

Accepted by:

Xiaoyan Lin, Major Professor

David Hitchcock, Committee Member

Lianming Wang, Committee Member

Linyuan Lu, Committee Member

Cheryl L. Addy, Vice Provost and Dean of the Graduate School

Acknowledgments

I would like to acknowledge all those who have helped me during my doctoral study. First of all, I am especially indebted to my esteemed advisor, Dr. Xiaoyan (Iris) Lin, for her critical inspiration and continuous support. Her infectious enthusiasm and continuous encouragement have always been the major driving force through my study and research. It is impossible to complete this dissertation without her generous guidance. I also extend my gratitude to the members of my exceptional doctoral committee: Drs. Lianming Wang, David Hitchcock, Linyuan Lu for their continuous support and insightful comments on my work.

Many professors in the Department of Statistics have assisted and encouraged me in various ways during my graduate study. I would like to express my deep gratitude to Dr. Hanson and Dr. Tebbs as graduate chairs for their valuable advice about taking courses as well as their encouragement. I would also like to express my appreciation to Drs. John Grego, Xianzheng Huang, David Hitchcock, Edsel Pena, Liangming Wang and Tim Hanson, for all the courses they have offered from which I have learned so much. I am very appreciative and thankful for the great support of Maureen Petkewich who has always kindly allocated all my teaching assignments by providing me with more chance and more experience to be a much better teaching assistant.

Finally, I am particularly grateful to my family for their love, support and sacrifices during my study. Without them, this dissertation would have never been written.

Abstract

Panel count data commonly arise in epidemiological, social science, medical studies, in which subjects have repeated measurements on the recurrent events of interest at different observation times. Since the subjects are not under continuous monitoring, the exact times of those recurrent events are not observed but the counts of such events within the adjacent observation times are known. Panel count data can be considered as a special type of longitudinal data with a count response variable in the literature. Compared to the frequentist literature, very limited Bayesian approaches have been developed to analyze panel count data. In this dissertation, several Bayesian estimation approaches are proposed for analyzing panel count data under different semiparametric regression models.

Chapter 1 of this dissertation provides some description of panel count data, literature review on existing methods, and background knowledge of related tools used in the proposed methods. Chapter 2 proposes a Bayesian estimation approach under the Poisson proportional mean model, in which we model the baseline mean function with the monotone splines of Ramsay (1988) [1]. An efficient Gibbs sampler is proposed, all parameters can be either sampled directly from their full conditional distributions in standard forms or updated through automatic adaptive rejection sampling. Our proposed method is evaluated through extensive simulations and compared with two exiting methods. Our method is applied to a bladder cancer data set for illustration. Chapter 3 proposes a new Bayesian estimation approach for analyzing panel count data when there is heterogeneity in the population (that cannot be described by the available covariates). A frailty Poisson proportional mean model is proposed with

the unobserved gamma frailties representing the heterogeneity among the subjects. Simulation studies suggest that our method not only has a good performance when such frailty exists but also provides robust estimation when there is no frailty. The bladder cancer tumor data is analyzed for illustration. Chapter 4 investigates the robustness of our proposed Bayesian approaches in Chapter 2 and Chapter 3 through simulations. We draw the conclusion that our proposed Bayesian methods still have a good performance in most cases when the assumptions are invalid.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Recurrent Data

1.1.1 Introduction

Recurrent data arise in a lot of areas such as epidemiology, econometrics, criminology, social sciences, reliability and clinic trial study. Recurrent data consist of times to repeated events for each sample subject. In other words, all these subjects in the study could experience recurrences of the same event at different times. For example, in a clinic study, recurrent episodes of a disease in patients appear multiple times, such as repeated heart attacks, cancer tumors recurrence even after previous tumors are removed etc.. In a reliability study, when a unit breaks down, it can be put back into service after being repaired. This situation could occur multiple times with warranty claims for some particular unit. More examples were given by Kalbfleisch and Lawless (1981, 1985)[2] [3], Thall and Lachin (1988) [4], and Sun and Kalbfleisch (1995) [5].

One typical type of recurrent study is the history study of the event of interest. Since the subjects experience event of interest multiple times, the resulting data are usually classified as event history data. Generally there are two types of event history data. One type is the data that can be collected at a continuous time period. We name this type of data recurrent event data (Byar 1980 [6], Gail 1981 [7], Pepe and Cai 1993 [8], Lin et al. 2000 [9]), which record the times of all occurrences of events. In the other type of data, the subjects are observed only at discrete time points, and

only the number of occurrences of the event between observation times is known. The exact time of the occurring event of interest is never known. They are often referred to as panel count data (Kalbfleisch and Lawless 1985 [2], Thall and Lachin 1988 [4], Sun and Kalbfleisch 1995 [5]).

Recurrent event data are frequently encountered in longitudinal follow-up studies such as epidemiological and medicinal studies. The observation of the recurrent events can be terminated at or before the end of the study. For instance, the recurrent events can have multiple occurrences of hospitalizations, and will be terminated by patients who loss to follow-up, a fatal event such as death, or the end of the study.

Panel count data are often encountered in periodic follow-up studies. It may be due to being impractical, too expensive, or not realistic to keep subjects under continuous observation over the entire study period. The number of observation times usually varies from one subject to another subject, and the observation times also vary from subject to subject. In practice, when there is only one observation time for the event of interest, such as death or onset of disease, the panel count data are often referred to as current status data (Dinse and Lagakos 1983 [10], Diamond and McDonald 1992 [11], Jewell and van 1996 [12], Sun and Kalbfleisch 1995 [5]). Interval censored data can also be viewed as a special case of panel count data.

We are going to mainly focus on panel count data study for this dissertation, and two examples are presented to further illustrate the basic concepts and general structures of panel count data.

1.1.2 Examples

(1) Bladder Cancer Tumor Study

Table 1.1 gives a set of panel count data about the bladder cancer study. This study was conducted by the Veterans Administration Cooperative Urological Research Group (Byar et al. 1977 [13], Byar 1980 [6]). There were 118 patients who had

superficial bladder tumors when they entered the study. For the first time visit, the number of the tumors were counted, and the initially largest tumor size was measured and all the initial tumors were removed transurethrally. Then each patient was randomly assigned to one of the three treatment groups: placebo, pyridoxine pill and thiotepa instillation. At each follow-up visit for each patient, the same procedure was followed: the number of the new tumors was counted, then all of them transurethrally removed, and then each patient was given the same treatment as before. Usually a lot of patients had multiple follow-up visits, and had recurrence tumors through the entire study. In all three treatment groups, the average follow-up time was about 31 months, but some patients may have been followed as long as five years. The objective of the study was to determine if the treatments, especially thiotepa instillation, reduced the recurrence of bladder tumors. The treatment effects have been analyzed by many authors such as Wei, Lin and Weissfeld (1989) [14], Wellner and Zhang (2007) [15], Lu, Zhang and Huang (2007, 2009) [16] [17].

(2) National Cooperative Gallstone Study (NCGS)

One of the major interests was to study the safety of the drug chenodiol for the treatment of cholesterol gallstones. This is a 10-year, multicenter, doubleblinded, placebo-controlled clinical trial on the use of the natural bile acid chenodeoxycholic acid, cheno, for the dissolution of cholesterol gallstones (Sun and Zhao 2013 [18]). There were a total of 916 patients who were assigned one of three treatments: high-dose (750 mg per day), low-dose (375 mg per day), or placebo. One of the primary objectives of the study was to assess the impact of the treatments on the incidence of digestive symptoms commonly associated with gallstone disease. The symptoms could be nausea, vomiting, dyspepsia etc. For example, episodes of nausea vomiting was viewed as a very common symptom associated with gallstone diseases and very important for the investigators to determine whether there exists a significant difference between the incidence of nausea for the patients in the three treatment groups.

Table 1.1: Numbers of new tumors at each visit for selected subjects with placebo treatment and Thiotepa treatment

Patient ID	Start Size	Numbers of new tumors at following months 0-10	11 -20	21-30
1	3	1 0	. . . . .	. . . . .
2	1	2 0 . . 0 . . 0 . . .	. . . . .	. . . . .
12	3	3 . . 1 . . 0 . . 1 0	. . 0 . 0 . 0 . . 0	8 . 0 . . . . .
17	4	1 . 4 . . . 0	. . . . .	. . . . . 8 . . . .
23	5	1 . 4 0	. . . . . 2 . . .	. . 4 0 . . . . 0 .
		Thiotepa group		
48	3	1 0	. . . . .	. . . . .
50	1	8 . . . . 8	. . . . .	. . . . .
79	4	1 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0
81	1	4 . . . 1 . . . 0 . .	. . 0	0 . . 1 . . . . .
85	3	1 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 . 0 . 0	0 0 . 0 0 0 0 0 .

The symptoms were observed shortly after they achieved maximal dose, since it was believed that any treatment had an effect right after taking the doses, and the effect might later begin to dissipate. During the study, the patients were scheduled to return for clinic observations for each follow-up visit. For the first year, they were asked to return at 1, 2, 3, 6, 9, and 12 months, and to report the total number of each type of symptom that had occurred between successive visits such as the number of the incidences of nausea. However, the actual visit times varied from patient to patient. For example, some patients may visit more often at the first half a year than the later half a year, some patients may drop off the study. Table 1.2 is a selected data subset which has visit times in weeks and observed counts of episodes of nausea for patients with floating gallstones. The whole data set has been studied in many references (Wei and Lachin 1984 [19], Thall and Lachin 1988 [4], and Park et al. 2007 [20]).

Table 1.2: Visit times in weeks and observed counts of episodes of nausea for some patients with floating gallstones in the National Cooperation Gallstone Study

Patient	Visit times and episodes of nausea																	
ID	t_1	N_1	t_2	N_2	t_3	N_3	t_4	N_4	t_5	N_5	t_6	N_6	t_7	N_7	t_8	N_8	t_9	N_9
							High-dose chemo group											
1	4	0	8	0	13	0	26	0	38	0	51	0	69	0	.	.	.	.
10	4	0	9	0	13	0	17	0	22	0	26	0	38	0	43	0	62	0
32	3	1	9	3	13	0	26	0	38	0	52	0	70	0	.	.	.	.
50	3	0	8	0	13	0	25	8	37	20	53	0	73	0	.	.		
							Placebo group											
64	4	0	8	0	12	0	25	0	38	0	52	0	68	0	.	.	.	.
70	4	0	8	0	13	0	24	0	37	0	50	0	67	0	.	.	.	.
103	4	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
111	4	0	9	0	14	0	25	0	.	.	.	.	.	.	.	.	.	.

1.2 Analysis of Panel Count Data

As we know the study subjects in panel count data are monitored at discrete time points, and the observable information is the event counts for each subject corresponding to the time interval between two adjacent time points, but the exact times of the event is not observable. Moreover, the observation times for each subject are not often the same, which makes the counts among subjects not directly comparable.

Based on the incomplete and unbalanced nature of observed information, to analyze recurrent event data, it is common and convenient to characterize the occurrences of recurrent events by counting process (Andersen et al. 1982 [21]) and to model the intensity process of the counting process. On the other hand, for the analysis of panel count data, it is usually more convenient to work directly on the mean function of the counting processes conditional on covariate processes (details in Sun and Zhao 2013 [18]). A natural and simple approach is to fit the panel count data through

parametric Poisson processes or mixed parametric Poisson processes. For example, Lawless (1987) [22] provided a thorough parametric analysis of the proportional intensity Poisson process model, where subject-specific gamma random effects and a Weibull baseline intensity function were assumed. Thall (1988) [4] developed some regression approaches for mixed Poisson. There are quite a few methods for panel count data with the assumption of nonhomegeneous Poisson process. For example, Wellner and Zhang (2000) [23] studied nonparametric maximum pseudolikelihood and nonparametric maximum likelihood estimators based on a nonhomogeneous Poisson process. They showed that the maximum pseudolikelihood estimator was exactly the one proposed by Sun and Kalbfleisch (1995) [5]. They also discussed the asymptotic properties of the maximum pseudolikelihood and maximum likelihood estimators. Compared with maximum pseudolikelihood estimator, the maximum likelihood estimator is much more efficient, but its computation is more difficult.

Many investigators have studied spline estimation of a hazard function for panel count data. Whittemore and Keller (1986) [24] used step functions and linear splines to obtain non-parametric estimates of the baseline hazard function. Etezadi-Amoli and Ciampi (1987) [25] applied quadratic splines to obtain smoother estimates. Rosenberg (1995) [26] modeled the hazard function as a linear combination of cubic-splines and obtained maximum likelihood estimates. Lu, Zhang and Huang (2007, 2009) [16] [17] proposed nonparametric spline likelihood-based estimators using monotone polynomial splines and semiparametric monotone B-splines, respectively.

In some cases, panel count data may show a higher incidence of zero counts than expected if the data were Poisson distributed. Zero-inflated Poisson regression models are a useful class of models for modeling such data, but parameter estimates may be seriously biased if the nonzero counts are overdispersed in relation to the Poisson distribution. Ridout et al. (2001) provided zero-inflated negative binomial alternatives to zero-inflated Poisson regression models, and conducted a score test for

testing zero-inflated Poisson regression models against zero-inflated negative binomial alternatives.

In the rest of these sections, we provide a review about counting process, the mean cumulative function and regression models for panel count data, respectively.

1.3 Counting Process

We know that panel count data study is a special type of event history study. The observation counts are inevitably involved in the study. The counting processes have been playing an essential and important role in the development of statistical models and inferential procedures for this kind of event history analysis. Especially, it is widely applied in analyzing panel count data. Some of the early and significant contributions to this were provided by Aalen (1975, 1978) [27] [28], Andersen and Borgan (1985) [29], and Liaw (1995) [30]. They established the connection between the counting process and the event history analysis, and provided the theory of counting processes and a general framework for event history analysis. In particular, Andersen and Gill (1982) [21] proposed the Cox type intensity model for counting processes, developed the partial likelihood estimation procedure for regression parameters, and established the large sample theory for the resulting estimators. Andersen et al. (1993) [31] gave more details about counting processes.

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and $T = [0, \tau)$ a continuous time interval, where τ is a given terminal time, $0 < \tau \leq \infty$. A stochastic process X is a family of random variables $\{X(t) : t \in T\}$. A filtration or history $\{\mathcal{F}_t : t \in T\}$ is an increasing right-continuous family of sub σ-algebras of $\mathcal{F}$, it contains all the information generated by the stochastic process X on $[0, t]$. If $X(t)$ is $\mathcal{F}_t$-measurable for every $t \in T$, then the stochastic process X is said to be adapted to the filtration. If $X(t)$ is known given the history $\mathcal{F}_{t-}$ generated by $\{X(s) : 0 \leq s < t\}$, then the stochastic process X is predictable with respect to $\mathcal{F}_t$. When $N(0) = 0$ and $N(t) < \infty$, a count-

ing process is a stochastic process $\{N(t) : t \geq 0\}$ with almost surely such that the path is right-continuous with probability one, piecewise constant, and has only jump discontinuities with jumps of size plus 1.

Usually we use the intensity to model a counting process, it could be defined by

$$P\{N(t+dt) - N(t) = 1|\mathcal{F}_t\} = \lambda(t)dt + o(dt)$$

and

$$P\{N(t+dt) - N(t) \geq 2|\mathcal{F}_t\} = o(dt)$$

with $\lambda(t)$ being a left-continuous function. If the counting process is Poisson process $N(t)$, based on the definition above, we can see that the Poisson process $N(t)$ only has at most one jump over a small time interval and does not depend on its history. A non-homogeneous Poisson process (Karlin and Taylor 1981 [32]) is a very common and widely applied special counting process. If $\lambda(t)$ is a constant, the process is usually called a homogeneous Poisson process. For a Poisson process $\{N(t) : t \geq 0\}$ at each t, we have that $N(t)$ follows the Poisson distribution.

Among all counting processes, the Poisson process $\{N(t) : t \geq 0\}$ as the most commonly used one has been shown a large contribution to queuing theory, risk theory and a lot of types of the application in model building area for our real life. There are a lot of examples of phenomena that can be modeled by Poisson processes with many kinds of ways. In traffic processes in networks such as computer systems, machine repair shops etc, traffic flows within the network are postulated as Poisson processes in Melamed (1979) [33]. A compound Poisson process had been used to model cumulative energy release of main shocks in the Balkan region for Balkan earthquake sequences by Gospodinov and Rotondi (2001) [34].

Landfalling hurricanes recorded along the U.S. Gulf and Atlantic coasts between 1900 and 2010, and their associated maximum wind speed and damages were studied by modeling the sequence of nonhomogeneous density functions in Xiao et al. (2015) [35].

We will discuss and study our models about panel count data based on the non-homogeneous Poisson process in the following chapters.

1.4 Mean Cumulative Function

Recurrence data consist of the times to any number of repeated events for each sample unit, for example, times of new bladder tumor observed in patients or times of repair of warrantied product. Sometimes the data are censored due to different ends of histories or study. Panel count data contains the counts of repeated events till particular observation times. After Cox (1972) [36] provided the proportional hazards model, Lawless and Nadeau (1995) [30] and Lin et al. (2000) [9] described proportional rates and proportional means regression models for recurrence data. The mean cumulative function for the number of the events usually contains the information of interest in the analysis of panel count data. Actually in many cases, the mean function is of more interest and modeling it directly is desirable. And we know that the aim of fitting a Cox model to time-to-event data is to estimate the effect of covariates on baseline hazard function, the aim of fitting proportional mean models is very smaller to it, i.e, the aim is to estimate the effect of covariates on the baseline mean cumulative function. Since the baseline hazard function is not itself estimated, there is no big risk of misspecifying the baseline distribution. how to construct or define the baseline function depends on one's perspective on what else effects will be considered in the models, and depends on how to combine it with the covariates in practice in order to get the ideal results in study.

The proportional mean model is as follows:

$$U(t|X) = E\{N(t)|X\} = U_0(t)\exp(X'\boldsymbol{\beta}), \tag{1.1}$$

where $\{N(t), t > 0\}$ is a counting process, X is a vector of invariant covariates. U_0 is baseline mean cumulative function, which is is unspecified non-negative and

nondecreasing function, β is a p x 1 vector of the regression coefficients.

In practice, if we obtain the mean cumulative function, it is easy to get information not only from the quantity of it, but also from the plot visually. For example, by some time t, whether the rate of occurrence of events is increasing, decreasing, or constant, and whether several groups differ significantly in expected number of events can be explained by the mean cumulative function. In Chapters 2, 3 and 4, we will adopt the proportional mean models to analyze the panel count data.

1.4.1 I-Splines

A spline is a real function which was introduced by Schoenberg (1964) [37] in connection with smooth, piecewise polynomial approximation. It has been developed from different purposes in terms of B-splines, G-splines, smoothing splines, I-splines, etc.. In the following, we will briefly introduce how to construct splines, and then mainly focus on introduce monotone I-splines, which will be used as an approximation for the baseline mean cumulative function in our models.

Let's start with M-splines. First, two crucial components are needed in order to construct M-splines. One is the degree of the polynomial, and the other is the knots which partition an interval into a number of subintervals. Let's choose the degree d and set up m interior knots $\xi_1 < \cdots < \xi_m$ within the interval $[a, b]$.

M-spline can be written as the form $f = \sum_{i=1}^{n} c_i M_i$, where M_i is the basis functions constructed by the following recursive formulas, and c_i is coefficient of basis function. Set $s_1 = a, s_2 = \xi_1, \cdots, s_{m+1} = \xi_m$, and $s_{m+2} = b$, then for $i = 1, 2, \ldots, m+1$, and $d = 1$,

$$M_l(t|1) = \begin{cases} \frac{1}{s_{l+1} - s_l}, & s_l \leq t < s_{l+1}, \\ 0, & \text{otherwise;} \end{cases}$$

for $d \geq 2$, let $s_1 = \cdots = s_d = a, s_{d+1} = \xi_1, \cdots, s_{d+m} = \xi_m$, and $s_{m+d+1} = \cdots = s_{m+2d} = b$, then for $d > 1$ and $l = 1, \ldots, m+d$,

$$M_l(t|d) = \begin{cases} \frac{[(t-s_l)M_l(t|d-1)+(s_{l+d}-t)M_{l+1}(t|d-1)]}{(k-1)(s_{l+d}-s_l)}, & s_l \leq t < s_{l+d}, \\ 0, & \text{otherwise.} \end{cases}$$

Note that each $M_l(\cdot|d)$ is a piecewise polynomial with nonzero only within $[s_l, s_{l+d})$ for $l = 1, \ldots, m+d$, and $\int M_l(x)dx = 1$. The reason is that this partition points only affect the change of the coefficient a_i within this interval. Also we can see that the individual polynomials always have the same degree and connect smoothly at join points (knots).

Based on M-splines, one can construct two specific splines: B-spline and I-splines. The B-splines can be constructed in the form: $f = \sum_{i=0}^{q} c_i B_i$, where B_i is the basis functions constructed by $B_i = (t_{i+k} - t_i)M_i/d$. q needs to be specified. The I-splines are the integrated functions of M-splines defined as follows:

$$I_l(t|d) = \int_L^t M_l(u|d)du. \tag{1.2}$$

Specifically, for $t \in [s_j, s_{j+1})$. $I_l(t|d)$ takes the following form,

$$I_l(t|d) = \begin{cases} 1, & l < j-d+1, \\ \sum_{h=1}^{j}(s_{h+d+1-s_h})\frac{M_h(t|d+1)}{d+1} & j-d+1 \leq l \leq j, \\ 0, & l > j. \end{cases} \tag{1.3}$$

The I-splines can be constructed in the form:

$$g(t) = \sum_{i=0}^{k} c_i I_i(t),$$

where $I_i(t)$ is the basis functions as (1.3), $I_0(t) = 1$ for any t. c_i is coefficient of basis function, it has to be nonnegative eo ensure that g is nondecreasing. Similar to B-splines, two crucial components are needed to specify the I spline basis functions: the knots and the degree. The placement of the knots determines the shape, and the degree determines the smoothness of the I splines. The I spline basis functions

are totally determined after the knots and the degree are specified. The number of spline basis functions k equals the number m of interior knots plus the degree d of the splines. We refer to Ramsay (1988) [1] where more details about I splines were provided.

In general, the shape of a spline function is robust to knot placement. We will discuss more about how to choose the number of the interior knots and the degree of the splines in the following chapter.

1.5 Deviance Information Criterion

In Bayesian method study, The deviance information criterion, Akaike information criterion and Bayesian information criterion are widely used to get deviance information in model selections. There are many literitures that can be found about their application. Comparative performance of Bayesian and AIC-based measures of phylogenetic model certainty was studied in Alfaro and Huelsenbeck (2006) [38]. Wang (2009) [39] developed a generalized Bayesian information criterion for regression model selection. Ando (2007) [40] discussed about Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. Erven and Grunwald (2012) [41] proposed a predictive approach to adaptive estimation with an application to the AIC-BIC dilemma.

The deviance information criterion (DIC) is a hierarchical modeling generalization of the Akaike information criterion (AIC) and Bayesian information criterion (BIC). It is very similar to AIC and BIC, since it is also an asymptotic approximation as the sample size becomes large. Claeskens and Hjort (2008) [42] showed that the DIC is large-sample equivalent to the natural model-robust version of the AIC. It is particularly useful in Bayesian model selection problems, especially, for some model selections, and the posterior distributions can be easily obtained by Markov chain Monte Carlo (MCMC) simulation.

Define the deviance as $D(\theta) = -2log(p(x|\theta))+C$, where x is the data, θ is unknown parameter of the model and $p(x|\theta)$ is the likelihood function. C is a constant, it will not be necessary to know, because it will be cancelled out in all the calculations when comparing different models.The expectation $\bar{D} = \mathrm{E}_\theta[D(\theta)]$ is a measure of how well the model fits the data; the smaller it is, the better the model fits.

In order to get DIC, some calculations have to be made. There are two ways that can be considered for getting the effective number of parameters included in the model. The first way, as described in Spiegelhalter et al. (2002) [43], is to calculate

$$p_D = \bar{D} - D(\bar{\theta}),$$

where $\bar{\theta}$ is the expectation of θ.

The second one, as described in Gelman et al. (2004) [44] is to get p_D by calculating the varivances as follows,

$$p_D = p_V = \frac{1}{2}\widehat{\mathrm{var}}\left(D(\theta)\right).$$

The larger the effective number of parameters is, the easier it is for the model to fit the data, and so the deviance needs to be penalized. The formula for calculating the deviance information criterion is

$$DIC = p_D + \bar{D},$$

or equivalent to

$$DIC = D(\bar{\theta}) + 2p_D.$$

or

$$DIC = 2\bar{D} - D(\bar{\theta}). \tag{1.4}$$

The advantage of DIC over other criteria in the case of Bayesian model selections is that the DIC can easily be calculated from the samples generated by a Markov

chain Monte Carlo simulation. AIC and BIC require calculating the likelihood at its maximum over θ, which is not readily available from the MCMC simulation. But to calculate DIC, as long as one can find the likelihood function, then simply compute $\bar{D}$ as the average of $D(\theta)$ over the samples of θ, and $D(\bar{\theta})$ as the value of D evaluated at the average of the samples of θ, then just take some calculations by the formula of getting the DIC. In the following chapters, DIC is used as one of criteria for the model selections about the panel count study.

Chapter 2

Bayesian Semiparametric Proportional Mean Model for Panel Count Data

2.1 Introduction

Panel count data often arise in epidemiological studies, industry reliability, long-term clinic trials, and animal experiments. Kalbfleisch and Lawless, Gaver and OMuircheataigh, Thall and Lachin, Thall, Sun and Kalbfleisch, and Wellner and Zhang etc. have been studying in the area a lot. By panel count data, it means that it is not feasible or not practical to continuously observe each subject, but the event of interests has a property of recurrence, and the study subjects are monitored periodically. For such studies, since the subjects are not under continuous monitoring, the exact time of each recurrent event is not observed but the count of such events between any two adjacent observation times is known. For example, in the bladder cancer study (Byar 1980 [6]), the famous example in the literature of panel count data, the number of new tumors is counted for each follow-up visit after the old tumors are removed at each visit. The observation times, i.e., the times at clinic visits, are different for different subjects, which makes the counts among subjects not directly comparable. Panel count data often can be considered as a special type of longitudinal data with a count response variable (sun and Zhao 2016 [45]).

Poisson models are commonly used for analyzing count data. However, the regular Poisson models ignore the longitudinal property of panel count data. To incorporate the incomplete data information and unbalanced data structure of panel count data,

a counting process is usually assumed and the observed counts are assumed to be collected at discrete time points from the counting process. Earlier research on panel count data focused on the estimation of the mean function of the counting process when there are no covarites. For example, Sun and Kalbfleisch (1995) [5] applied the isotopic regression method and proposed a nonparametric estimate of the mean function. Wellner and Zhang (2000) [23] proposed both a nonparametric pseudo-maximum likelihood estimator and the full maximum likelihood estimator of the mean function based on a non-homogeneous Poisson process. Hua and Zhang [46] discussed a semiparametric analysis of panel count data within the framework of GEE, assuming the proportional mean model as in Wellner and Zhang (2007) [15] and Lu et al. (2009) [17].

Spline techniques have been widely used to approximate unknown functions in statistics literature. For modeling panel count data, Lu et al. (2007) [16] adopted monotone B-splines to model the mean function in the nonparametric setting and showed that their spline likelihood estimators outperformed the estimators proposed by Wellner and Zhang (2000) [23] in terms of convergence rate and finite sample performance. Lu et al. (2009) [17] then extended the work of Lu et al. (2007) [16] to the regression setting and used the monotone cubic B-splines to approximate the logarithm baseline mean function. Hua and Zhang (2012) [46] adopted the same spline specification and obtained the estimates of the regression parameters and spline coefficients by projecting the GEE estimates into a feasible domain under the proportional mean model. In this chapter, we assume the proportional mean model for the panel count data and use monotone I-splines (Ramsay 1988 [1]) to model the baseline mean function. Our spline coefficients only require being nonnegative to ensure the monotonicity of the baseline mean cumulative function. Furthermore, the monotone I-spline specification and the nonnegative restriction naturally fit in the Bayesian prior specification and the Markov Chain Monte Carlo (MCMC) update.

Compared to the flourish work in the frequentest literature, Bayesian approaches for panel count data seem to be very limited. Chib and Greenberg (1998) [47] conducted Poisson regression on correlated count data with the correlation among the counts modeled by multivariate normal latent effects. Sinha and Maiti (2004) [48] proposed a Bayesian approach for the analysis of panel count data with a dependent termination following a discreet Cox model. However, Sinha and Maiti (2004) [48] made restrictive assumptions that the observation times are the same for three all subjects and that the termination times only take values of the observations times.

In many applications, the effects of covariates on the panel count are of interest. Wellner and Zhang (2007) [15] extended their two likelihood-based methods (Wellner and Zhang 2000 [23]) by incorporating covariates under a proportional mean regression model (2.1) associated with the assumption of a non-homogeneous Poisson process.Their methods are shown to be robust against misspecification of the underlying counting process, as long as the proportional mean model 2.1 holds. The likelihood estimator is more efficient than the pseudolikelihood estimator, but its computation is very intensive. And also it can be very inefficient when the distribution of the number of observations times is heavily tailed, which is shown in an ex-ample given by Wellner, Zhang and Liu (2004) [49]. In Wellner and Zhang (2007) [15], a substantial amount of computation was required in the bootstrap semiparametric inference procedure. Therefore it is desirable to develop some methods that not only maintain good statistical properties, but also are less computationally intensive.

Generally, various methods of panel count data and recurrent timed data were developed with mainly putting high emphasis on partially specified models. In suck work, the models and associated inferential procedures depend on the particular underlying inferential objective. These can be of the following four types as mentioned in Sinha and Maiti (2004) [48]. (1) Inference about the event count under the assumption that given covariates, the event process is independent of the observation

time; (2) Marginal inference about event count unconditional on past history of event ; (3) Inference about observation time given the covariate information or the history of event count; and (4) Inference about the termination modulation process (Ghosh and Lin 2003 [50]). For example, Sinha and Maiti (2004) [48] developed and analyzed a fully specified stochastic model of the joint distribution of the nonfatal events and the termination time through the likelihood of their model and associated Bayesian methods for analysis under the assumption of a compound nonhomogeneous Poisson process.

In this chapter, we provide a very flexible Bayesian approach for analyzing panel count data, where the dependent termination is not considered, but the number and the sequence of observation times are allowed to be different across subjects. The proposed Bayesian approach provides an accurate estimation of both the regression parameters and the baseline mean function. Beside, our approach is computationally efficient and easy to implement.

The remainder of this chapter is organized as follows. First the details of data structure are given in Section 2.2. The specification of monotone splines is presented in Section 2.3. Section 2.4 augments the likelihood function under the Poisson process assumption about the panel count of interest. Prior specification and an efficient algorithm of Gibbs sampler are presented in Section 2.5. Section 2.6 presents the simulation results of the estimates of regression parameters and the estimates of baseline function for two scenarios, respectively. Model selections in terms of the number of knots and degrees of the splines are discussed in Section 2.6 as well. The comparison about the estimates of regression parameters and baseline mean function is provided with two other existing methods in Section 2.6.4. A real data set is illustrated by our proposed method in Section 2.7. Finally, the conclusion and some comments are presented in Section 2.8.

2.2 Notation and Model

Panel count data is to record the counts of recurrent events between every two visit times for each subject. In practice, it can be treated as cluster data, and all the observations from the same subject can be thought of as a cluster. Suppose we have n independent subjects (clusters) for a set of panel count data, for each subject i, the data are collected at K_i random time points $0 < t_{i1} < t_{i2} < \ldots < t_{iK_i}$, where K_i is a random variable that takes positive integer values, and t_{iK_i} is the last observation time of subject i. Denote $N_i(t)$ as the number of events observed before and at time t for subject i, then $N_i(t)$ is a counting process for cluster i. So the panel counts for the counting process $N_i(t)$ are $N_i = (N_i(t_{i1}), N_i(t_{i2}), \ldots, N_i(t_{iK_i})$.

Without loss of generality, we assume initial counts $N_i(t_{i0}) = 0$ for $i = 1, \ldots, n$. We have the whole set of observed panel count data denoted by $\mathscr{D}_i = \{t_{ij}, N_i(t_{ij}), X_i,$ for $i = 1, \ldots, n, j = 1, \ldots, K_i\}$, where X_i is a p-dimensional vector of covariates. Denote $\mathscr{D} = (\mathscr{D}_1, \ldots, \mathscr{D}_n)$. Assume that the counting process $N_i(\cdot)$ is independent of the number K_i of the observation times and the sequence of the observation times $t_{i1},\ t_{i2}, \ldots, t_{iK_i}$. Then we have the following proportional mean model:

$$E(N_i(t)|X_i) = U_0(t)\exp(X_i'\boldsymbol{\beta}), \tag{2.1}$$

for $i = 1, \ldots, n$, where $U_0(t)$ is the baseline mean cumulative function and $\boldsymbol{\beta}$ is a vector of covariate effects. Model (2.1) has been widely studied in semiparametric or parametric analysis, for instance, in Balshaw and Dean (2002) [51], Huang, Wang and Zhang (2006) [52], Lu, Zhang and Huang (2007) [16], and Wellner and Zhang (2007) [15]. Let Z_{ij} be the count of recurrences within the jth time interval $[t_{i,j-1}, t_{ij})$, then Z_{ij} will be written as $Z_{ij} = N_i(t_{ij}) - N_i(t_{i,j-1})$, for $j = 1, \ldots, K_i$. In this chapter, we assume that $N_i(t)$ follows a nonhomogeneous Poisson process, then Z_{ij} independently follows the Poisson distribution as follows:

$$Z_{ij}|X_i \sim \mathcal{P}\Big(\big\{U_0(t_{ij}) - U_0(t_{ij-1})\big\}\exp(X_i'\boldsymbol{\beta})\Big). \tag{2.2}$$

The whole set of observed data can also be denoted by $\mathscr{D} = \{t_{ij}, Z_{ij}, X_i, \text{ for } i = 1, \ldots, n, \ j = 1, \ldots, K_i\}$.

2.3 Modeling $U_0(t)$ with Monotone I- splines

Monotone I-splines (Ramsay 1988 [1]) have recently been widely applied in semi-parametric survival models such as in Wang and Dunson (2011) [53], Wu et al. (2012) [54], and Lin et al. (2015) [55]. Monotone I-splines are often adopted to approximate unknown nonnegative and nondecreasing functions. In this dissertation, monotone I-splines (Ramsay 1988 [1]) are also adopted to approximate the unknown nonnegative and nondecreasing function $U_0(t)$. The formula is as follows:

$$U_0(t) = \sum_{l=1}^{L} r_l I_l(t|d), \tag{2.3}$$

where $I_l(\cdot|d)$ i monotone I-spline basis function with degree d and each is nondecreasing from 0 to 1, and r_ls are nonnegative spline coefficients to ensure the monotonicity of the nonnegative function $U_0(t)$. The detailed formulation of the basis function $I_l(\cdot|d)$ is provided in Chapter 1, and more details can be found in Ramsay (1988) [1] and Lin et al. (2015) [55]. The degree d and the knot placement are two crucial components to determine the basis functions. The number L of the basis functions is equal to the degree d plus the number m of interior knots and plus 1. In general, the degree d controls the smoothness of the splines and taking degree as 2 or 3 is usually enough to ensure the smoothness. The placement of knots controls the shapes of the basis splines, and as a consequence, they effect the shape of the final fitted function. Usually, the more knots in a region, the greater the flexibility of the function in that region; the more data points between a pair of knots, the better fitted the curve in that interval (Lawless 1987 [22]). According to Cai et al. (2011) [56] and Wang and Dunson (2011) [53], in general, using 10 to 30 knots can provide adequate modeling flexibility even for thousands of observations in data sets. One may choose

equally-spaced knots or quantile-based knots, which depends on data itself. Moreover, Bayesian model comparison criteria such as the deviance information criterion (DIC) (Spiegelhalter et al. 2002 [43]) and log pseudo marginal likelihood (LPML) (Ibrahim et al. 2001 [57]) would also be useful to help with selection of the best setup of the degree and the number of knots.

In this chapter, we use the comparison criterion of deviance information criterion (DIC) (Spiegelhalter et al. 2002 [43]) to determine the best setup of I-splines in the simulation part in Section 2.6 and the real data analysis in Section 2.7.

2.4 Likelihood Augmentation with Poisson Latent Variables

Since $\{N_i(t), t > 0\}$ follows a non-homogeneous Poisson process, and Z_{ij} is defined as the count difference between time points $t_{i,j-1}$ and $t_{i,j}$, i.e., $Z_{ij} = N_i(t_{ij}) - N_i(t_{i,j-1})$, for $i = 1, \ldots, n$, and $j = 1, \ldots, K_i$. Based on the properties of a nonhomogenous Poisson process, Z_{ij} independently follows Poisson distribution conditionally on the covariates X_i,

$$Z_{ij}|X_i \sim \mathcal{P}\bigg(\Big\{U_0(t_{ij}) - U_0(t_{ij-1})\Big\}\exp(X_i'\boldsymbol{\beta})\bigg).$$

Under the nonhomogenous Poisson process assumption (given X), the observed likelihood function is

$$L(\theta|\mathscr{D}) = \prod_{i=1}^{n}\prod_{j=1}^{K_i}\exp\bigg[-\Big\{U_0(t_{ij}) - U_0(t_{ij-1})\Big\}\exp(X_i'\boldsymbol{\beta})\bigg] \bigg[\Big\{U_0(t_{ij}) - U_0(t_{ij-1})\Big\}\exp(X_i'\boldsymbol{\beta})\bigg]^{Z_{ij}} \Big/ Z_{ij}!, \tag{2.4}$$

where θ is a vector of all unknown parameters including $\boldsymbol{\beta}$ and $r = (r_1, \ldots, r_L)$. Note that for the notation convenience, we omit degree d in the basis functions. Under the monotone I-spline approximation to $U_0(t)$, $U_0(t) = \sum_{l=1}^{L} r_l I_l(t)$, then the likelihood can be further written as

$$L(\theta|\mathscr{D}) = \prod_{i=1}^{n}\prod_{j=1}^{K_i}\exp\Big[-\Big\{\sum_{l=1}^{L} r_l I_l(t_{ij}) - \sum_{l=1}^{L} r_l I_l(t_{ij-1})\Big\}\exp(X_i'\boldsymbol{\beta})\Big]$$

$$\times\Big[\Big\{\sum_{l=1}^{L} r_l I_l(t_{ij}) - \sum_{l=1}^{L} r_l I_l(t_{ij-1})\Big\}\exp(X_i'\boldsymbol{\beta})\Big]^{Z_{ij}}\Big/ Z_{ij}!. \tag{2.5}$$

This likelihood function is very complex, especially for the term

$$\Big[\Big\{\sum_{l=1}^{L} r_l I_l(t_{ij}) - \sum_{l=1}^{L} r_l I_l(t_{ij-1})\Big\}\exp(X_i'\boldsymbol{\beta})\Big]^{Z_{ij}}. \tag{2.6}$$

This term makes sampling each r_l very difficult. To facilitate the computation, independent Poisson latent variables $Z_{ij1}, \ldots, Z_{ijL}$ are introduced to decompose Z_{ij} such that $Z_{ij} = \sum_{l=1}^{L} Z_{ijl}$. By the additivity property of Poisson random variable, each single Z_{ijl} independently follows a Poisson distribution as follows:

$$Z_{ijl}|X_i \sim \mathcal{P}\Big(\Big\{r_l I_l(t_{ij}) - r_l I_l(t_{ij-1})\Big\}\exp(X_i'\boldsymbol{\beta})\Big),$$

for $i = 1, \ldots, n,\ j = 1, \ldots, K_i$, and $l = 1, \ldots, L$.

Then the augmented likelihood function with Z_{ijl}'s is

$$L_{aug} = \prod_{i=1}^{n}\prod_{j=1}^{K_i}\prod_{l=1}^{L}\exp\Big[-\Big\{r_l I_l(t_{ij}) - r_l I_l(t_{ij-1})\Big\}\exp(X_i'\boldsymbol{\beta})\Big]$$

$$\times\Big[\Big\{r_l I_l(t_{ij}) - r_l I_l(t_{ij-1})\Big\}\exp(X_i'\boldsymbol{\beta})\Big]^{Z_{ijl}}\Big/ Z_{ijl}!. \tag{2.7}$$

The augmented likelihood function (2.7) gets rid of all the summation signs in the term of (2.6), and adds another of multiplication in front of the whole expression of the likelihood function. As a result, each parameter r_l can be easily factored out, which is much helpful for obtaining their full conditional distributions and ease the MCMC sampling in the following section.

2.5 Prior Specification and Posterior Computation

For Bayesian posterior computation, we need to assign priors for the unknown parameters $\boldsymbol{\beta}$ and r_l. For $\beta_m, m = 1, \ldots, p$, we assign conventional independent vague normal priors with mean zero and large variance σ_m^2 such as 100. This leads to a log-concave full conditional posterior distribution for each β_m, and can be sampled easily using the adaptive rejection sampling (ARS) (Gilks and Wild 1992 [58]). For $r_l,\ l = 1, \ldots, L$, we assign independent exponential $\exp(\lambda)$ priors with a Gamma hyper prior $\mathcal{G}(a_\lambda, b_\lambda)$ for λ with mean a_λ/b_λ and variance a_λ/b_λ^2. This prior specification is appealing from the computational perspective because it leads to conjugate forms for each of the conditional posterior distributions of $r_l\prime s$ and λ. Theoretically, such a prior specification is closely related to Bayesian Lasso (Park and Casella 2008 [59]) and is equivalent to the penalized likelihood approach with L_1 penalty on those spline coefficients, in which λ serves as a tuning parameter. This prior specification penalizes large values of the coefficients $r_l s$ and functions to shrink the coefficients of those unnecessary spline bases towards zero. This property allows us to use many knots to provide adequate modeling flexibility but does not cause over-fitting problems. In the penalized likelihood approach, selecting a proper λ value requires much additional work using cross-validation method. In contrast, our approach treats λ as random and assigns the Gamma hyper prior for λ to allow for automatic tuning with much less computational efforts. Our simulations show that our approach is very robust to the choice of the hyperparameters in the Gamma hyper prior of λ. Here we choose $a_\lambda = 1$ and $b_\lambda = 1$.

Gibbs sampling (Geman and Geman 1984 [60], Gelfand and Smith 1990 [61]) is one of the most popular MCMC algorithm (Casella and robert 2004 [62]) for sampling the joint posterior distribution. We adopt Gibbs sampling for sampling the joint posterior distribution. Typically, it requires sampling all the unknown parameters and latent variables from their full conditional distributions iteratively. When

the MCMC chains converge, the samples can be considered from the target joint posterior distribution. Based on the augmented likelihood function (2.7) and the above specified prior distributions, the Gibbs sampler is derived with the four sets of the full conditionals as below:

1. Sample r_l from a Gamma distributions $\mathcal{G}(A_l, B_l)$, for $l = 1, \ldots, L$, where

$$\begin{aligned} A_l &= \sum_{i=1}^{n}\sum_{j=1}^{K_i} Z_{ijl} + 1, \\ \\ B_l &= \sum_{i=1}^{n}\Big\{I_l(t_{iK_i}) - I_l(t_{i0})\Big\}\exp(X_i'\boldsymbol{\beta}) + \lambda. \end{aligned}$$

2. Sample β_m by using the adaptive rejection sampling (ARS) (Gilks and Wild 1992 [58]) method, for $m = 1, \ldots, p$, because the full conditional distribution of each β_m is proportional to

$$\exp\Bigg[-\sum_{i=1}^{n}\sum_{l=1}^{L} r_l\Big\{I_l(t_{iK_i}) - I_l(t_{i0})\Big\}\exp(X_i'\boldsymbol{\beta}) + \sum_{i=1}^{n}\sum_{j=1}^{K_i}(X_i'\boldsymbol{\beta})Z_{ij} - \beta_m^2/(2\sigma^2)\Bigg],$$

which is a log-concave function of β_m. This readily available software program for the adaptive rejection sampling method (Gilks and Wild 1992 [58]) has been widely used to sample from these log-concave densities such as Sinha and Maiti (2004) [48], Yao and Wang (2016) [63] etc.

3. Sample Z_{ijl}, for $i = 1, \ldots, n,\ j = 1, \ldots, K_i$ and $l = 1, \ldots, L$. It is from a multinomial distribution with parameters Z_{ij} and $\mathbf{P}_{ij}$, denoted by $\mathcal{M}(Z_{ij}, \mathbf{P}_{ij})$, where $\mathbf{P}_{ij} = (p_{ij1}, \cdots, p_{ijL})$ with $\sum_{l=1}^{L} p_{ijl} = 1$, and

$$p_{ijl} = \frac{r_l\{I_l(t_{ij}) - I_l(t_{i,j-1})\}}{\sum_{l=1}^{L} r_l\{I_l(t_{ij}) - I_l(t_{i,j-1})\}},$$

$$(Z_{ij1}, \cdots, Z_{ijL}|Z_{ij}) \sim \mathcal{M}(Z_{ij}, \mathbf{P}_{ij}).$$

4. Sample λ from a Gamma distribution $\mathcal{G}(a_\lambda + L, b_\lambda + \sum_{l=1}^{L} r_l)$.

The above developed Gibbs sampler is very efficient and easy to implement, because all the full conditional distributions either have closed forms or they are log-concave.

2.6 Simulation Studies

2.6.1 Data Generation

An extensive simulation study is conducted to evaluate the proposed approach. Data sets are generated as follows. The event count of each subject i is assumed to follow a nonhomogeneous Poisson process with the mean function modeled as (2.1),

$$E(N_i(t)|X_i) = U_0(t)\exp(X_i'\boldsymbol{\beta}),$$

where covariates $X_i = (X_{i1}, X_{i2})$ with X_{i1} generated from B$ernoulli(0.5)$ and X_{i2} from $N(0, 0.5)$. Each subject is repeatedly observed at a random number P$oisson(6) + 1$ of times and the counts of event are collected, and the gap times between any two adjacent observation times are drawn from an exponential distribution with mean 0.5. We set up the true regression coefficients $\beta_1 = 0, 1,$ or -1 and $\beta_2 = 0$ or 1, resulting in 6 combinations of regression coefficients. For the true baseline mean function U_0, we choose two scenarios:

Scenario 1, $U_0(t) = t + log(1 + t)$, which is approximately linear.

Scenario 2, $U_0(t) = 0.5\Phi(t, 1, 1) + 0.5\Phi(t, 5, 0.5) + t^{0.5}$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. It is clear that for this scenario, $U_0(t)$ is curvilinear and has multiple reflection points.

We totally perform 12 simulation setups. For each setup, 500 data sets are simulated and each data set has 150 subjects. The proposed Gibbs sampler in Section 5 is implemented for each simulated data set. All the summarized results are based on 5000 iterations of the Gibbs samples after discarding the first 1000 iterations as a burn-in.

2.6.2 Specification of Splines

In order to get the specification of I-spline basis functions, we choose the degree and the number of knots as necessary. Model selection criterion DIC (deviance information criterion) can help work it out. As we described in Chapter 1, we can use the formula (1.4) to get DIC. For our particular Model 2.1, we have

$$\begin{aligned}\log(p(y|\theta)) = &-\sum_{i=1}^{n}\sum_{l=1}^{L} r_l b_l(t_{iK_i})\exp(X_i'\boldsymbol{\beta}) + \sum_{i=1}^{n}\sum_{j=1}^{K_i} Z_{ij}\log\Big(\sum_{l=1}^{L} r_l df_{ijl}\Big) \\ &+\sum_{i=1}^{n}\sum_{j=1}^{K_i} Z_{ij}\exp(X_i'\boldsymbol{\beta}), \end{aligned} \tag{2.8}$$

where $df_{ijl} = b_l(t_{ij}) - b_l(t_{ij-1})$.

Based on (2.8), we can easily calculate $\bar{D}$ and $D(\bar{\theta})$. So DICs can be obtained for different knots and degrees. The model with smallest DIC value could be preferred.

In both scenarios, we obtain DICs for equally-spaced knots and quantile-based knots for each simulated data set, respectively. We choose equally-spaced knots between 0 and maximum of observation times from each data set, and choose quantile-based knots by setting quantiles between 0 to 1, and each quantile is increased by $1/(q-1)$, where q is the number of the knots. The results of DICs suggest that there is almost no big difference between using equally-space knots and quantile-based knots regardless of degrees or scenarios (Pan et al. 2014 [64]). Therefore, in this dissertation, we present the results based on 20 equally-spaced knots regardless of the degree or scenario. For the degree choice, the DIC difference between using degree 2 and using degree 3 is obtained for each different $\boldsymbol{\beta}$ combination based on 500 simulated data sets, and the boxplots of DIC differences for six different $\boldsymbol{\beta}$ combinations for both scenarios are shown in Figure 2.1. For scenario 1, it indicates that DICs for each $\boldsymbol{\beta}$ combination with degree 2 is very close to those with degree 3. However, for scenario 2, the overall DICs for each $\boldsymbol{\beta}$ combination with degree 2 is greater than those with degree 3. Hence, from now on, all the following results are based on

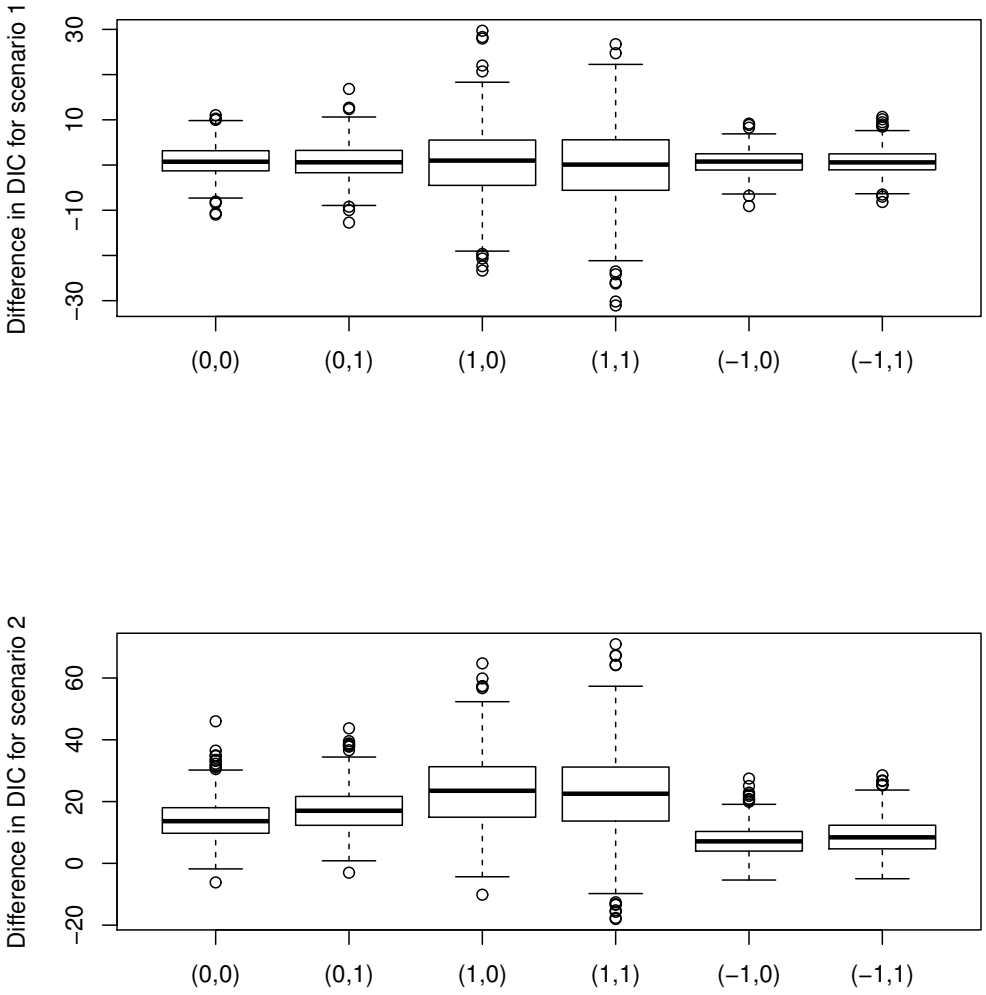

Figure 2.1: Boxplots of DIC difference between degree 2 and degree 3 based on 500 simulated data sets for the six $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ configurations for scenario 1 and scenario 2, respectively.

equally-spaced knots with degree 2 for scenario 1, and with degree 3 for scenario 2. Usually, if there is no big difference between degree 2 and degree 3, lower degree is preferred, because it decreases the dimension of the estimation which contributes to the computing efficiency.

2.6.3 Simulation Results

Based on the selection of knots and degree for each scenario, we implement the proposed Gibbs sampler in Section 2.5. The results suggest that there is a good mixing and fast convergence in MCMC chains of the regression parameters and the spline coefficients regardless of scenarios. Tables 2.1 and 2.2 present the frequentist operating characteristics of estimates of the regression parameters for both scenarios, with Bias denoting the difference between the average of 500 point estimates and the true value, SSE as the sample standard deviation of the 500 point estimates, ESD as the average of the 500 posterior standard deviations, and CP95 as the 95% coverage probability. The point estimates are taken to be posterior means, and the coverage probabilities are based on the 95% credible intervals constructed by 2.5th and 97.5th percentiles. In Tables 2.1 and 2.2, the results from our proposed method indicate that our method performs very well regardless of scenarios, because the estimates show no bias, and SSE and ESD are small and very close to each other, and the coverage probabilities are all close to 0.95. One advantage of the proposed method is to provide a smooth estimate of the baseline mean function. Figures 2.2 and 2.3 present the estimation of the baseline mean function and their corresponding pointwise credible intervals for all the configurations. It is clear that the estimation of the baseline mean function is very good as the average of the estimates overlaps the true curve. The departure of the estimates from the true curve at the right tail area is reasonable because wherein there are rare observations at the end of time period.

Table 2.1: Estimation of regression parameters for scenario 1 based on 500 simulated data sets from the proposed Bayesian method, parametric method and Rosen algorithm method. Bias refers to the difference between the average of the 500 point estimates and the true value, ESD refers to the average of the 500 posterior standard deviations, SSE refers to the sample standard deviation of the 500 point estimates, and CP95 is the 95% coverage probability.

Scenario 1: $U_0(t) = t + \log(1+t)$										
			Results on β_1				Results on β_2			
β_1	β_2	Methods	Bias	SSE	ESD	CP95	Bias	SSE	ESD	CP95
0	0	Proposed	-0.0002	0.0800	0.0740	0.932	-0.0020	0.0800	0.0750	0.926
		Parametric	-0.0006	0.0787	0.0522	0.776	-0.0024	0.0796	0.0743	0.918
		Rosen	0.0009	0.0892	0.0376	0.578	-0.0053	0.0926	0.0378	0.594
0	1	Proposed	0.0009	0.0742	0.0704	0.930	0.0027	0.0773	0.0715	0.942
		Parametric	0.0027	0.0741	0.0492	0.796	0.0030	0.0764	0.0633	0.912
		Rosen	0.0033	0.0869	0.0360	0.584	0.0045	0.0871	0.0371	0.604
1	0	Proposed	0.0012	0.0602	0.0616	0.948	0.0015	0.0563	0.0551	0.956
		Parametric	0.0016	0.0602	0.0315	0.706	0.0017	0.0564	0.0544	0.948
		Rosen	0.0010	0.0679	0.0294	0.578	0.0019	0.0592	0.0279	0.632
1	1	Proposed	0.0013	0.0592	0.0580	0.950	0.0001	0.0523	0.0525	0.962
		Parametric	0.0019	0.0593	0.0297	0.638	0.0012	0.0522	0.0465	0.934
		Rosen	0.0010	0.0679	0.0294	0.578	0.0019	0.0592	0.0279	0.632
-1	0	Proposed	-0.0026	0.0990	0.1019	0.950	-0.0030	0.0949	0.0913	0.944
		Parametric	-0.0027	0.0816	0.0718	0.920	0.0040	0.0704	0.0748	0.952
		Rosen	-0.0007	0.1166	0.0509	0.618	-0.0031	0.1075	0.0470	0.614
-1	1	Proposed	-0.0007	0.0907	0.0961	0.962	-0.0040	0.0862	0.0870	0.956
		Parametric	0.0033	0.0908	0.0812	0.918	-0.0031	0.0864	0.0774	0.924
		Rosen	0.0003	0.1053	0.0486	0.646	-0.0019	0.1040	0.0457	0.618

Table 2.2: Estimation of regression parameters for scenario 2 based on 500 simulated data sets from the proposed Bayesian method, parametric method and Rosen algorithm method. Bias refers to the difference between the average of the 500 point estimates and the true value, ESD refers to the average of the 500 posterior standard deviations, SSE refers to the sample standard deviation of the 500 point estimates, and CP95 is the 95% coverage probability.

Scenario 2: $U_0(t) = 3(0.5\ \text{pnorm}(t,1,1) + 0.5\text{pnorm}(t,5,0.5) + t^{0.5})$										
			Results on β_1				Results on β_2			
β_1	β_2	Methods	Bias	SSE	ESD	CP95	Bias	SSE	ESD	CP95
0	0	proposed	-0.0025	0.0600	0.0621	0.968	0.0029	0.0594	0.0624	0.956
		parametric	-0.0010	0.0601	0.04352	0.850	0.0028	0.0592	0.0617	0.956
		Rosen	-0.0033	0.0670	0.0293	0.624	0.0033	0.0661	0.0296	0.622
0	1	proposed	-0.0007	0.0595	0.0583	0.946	-0.0020	0.0585	0.0586	0.950
		parametric	0.0003	0.0596	0.0409	0.812	-0.0017	0.0586	0.0517	0.091
		Rosen	0.0013	0.0657	0.0280	0.620	-0.0007	0.0652	0.0286	0.620
1	0	proposed	0.0015	0.0512	0.0511	0.948	-0.0042	0.0447	0.0456	0.946
		parametric	0.0014	0.0509	0.0264	0.692	-0.0042	0.0449	0.0451	0.948
		Rosen	0.0012	0.0588	0.0239	0.568	-0.0052	0.0539	0.0219	0.556
1	1	proposed	-0.0012	0.0658	0.0482	0.946	-0.0045	0.0610	0.0431	0.966
		parametric	-0.0018	0.0662	0.0249	0.672	-0.0050	0.0617	0.0383	0.930
		Rosen	0.0029	0.0550	0.0230	0.598	-0.0004	0.0491	0.0215	0.600
-1	0	proposed	-0.0068	0.0818	0.0845	0.958	0.0039	0.0708	0.0755	0.956
		parametric	-0.0027	0.0816	0.0718	0.920	0.0039	0.0704	0.0747	0.952
		Rosen	-0.0053	0.0912	0.0395	0.608	0.0039	0.0792	0.0364	0.614
-1	1	proposed	-0.0034	0.0921	0.0798	0.940	-0.0031	0.0844	0.0719	0.944
		parametric	-0.0021	0.0800	0.0679	0.914	-0.0020	0.0714	0.0640	0.922
		Rosen	-0.0002	0.1015	0.0378	0.560	-0.0038	0.0957	0.0355	0.582

2.6.4 Comparison With Two Bench Methods

In this section, we compare our proposed method with a parametric method and the method performed by Lu et al. (2009) [17]. In order to make the result comparable and meaningful, we use the same simulated data sets for all three methods.

Let's start with a parametric method. The baseline mean function U_0 in (2.1) assumes the Weibull form $(t/\lambda)^p$. This parametric specification results in a very finite number of parameters $(\lambda, p, \boldsymbol{\beta})$, and the maximum likelihood method is used to estimate the parameters through the following log-likelihood function:

$$\begin{aligned} log(L_p) = &- \sum_{i=0}^{n} \left[\frac{t_{iK_i}}{\lambda}\right]^p \exp(X_i'\boldsymbol{\beta}) \\ &+ \sum_{i=1}^{n}\sum_{j=1}^{K_i} Z_{ij} \log\left([\frac{t_{ij}}{\lambda}]^p - [\frac{t_{ij-1}}{\lambda}]^p \right) \exp(X_i'\boldsymbol{\beta}) + \sum_{i=1}^{n}\sum_{j=1}^{K_i} Z_{ij}!. \end{aligned}$$

Tables 2.1 and 2.2 also show the estimation results of regression coefficients $\boldsymbol{\beta}$ using the parametric method for the same simulated data sets in Section 2.6. The estimates suggest that there is no bias of the estimates of regression coefficients for any $\boldsymbol{\beta}$ combinations. These estimates of $\boldsymbol{\beta}$ are similar to the results in our proposed method. However, the standard deviations are clearly underestimated, since their ESDs are consistently smaller than SSEs, which leads to bad coverage probabilities for most cases. The underestimation of standard errors may be caused by the misspecification of the parametric model. Figures 2.4 and 2.5 show the comparison about the estimation of baseline function with the parametric method. The green dash-dotted curves are the estimates of the unknown baseline function from the Weibull parametric method in both scenarios. It fits well for scenario 1, but it doesn't fit well for scenario 2 at turning points. Another drawback of using the MLE for the parametric model is the non-convergence problem. We noticed that 26% of data sets have no convergent results if random initial values are given for the MLE. Here we take the Bayesian estimates as the initial values and it converges for all data sets.

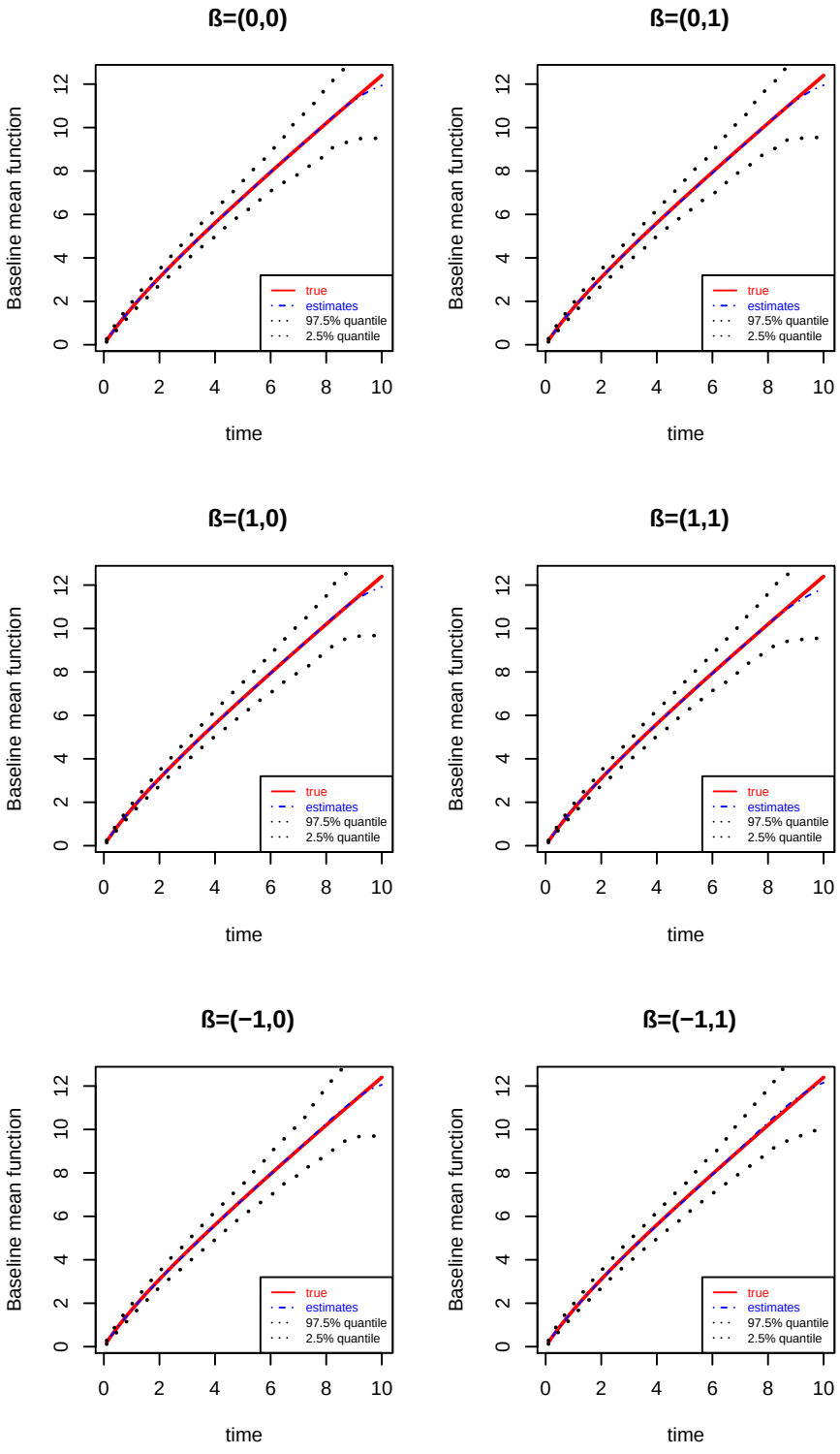

Figure 2.2: Estimates of baseline mean functions for different $\boldsymbol{\beta}$ in scenario 1

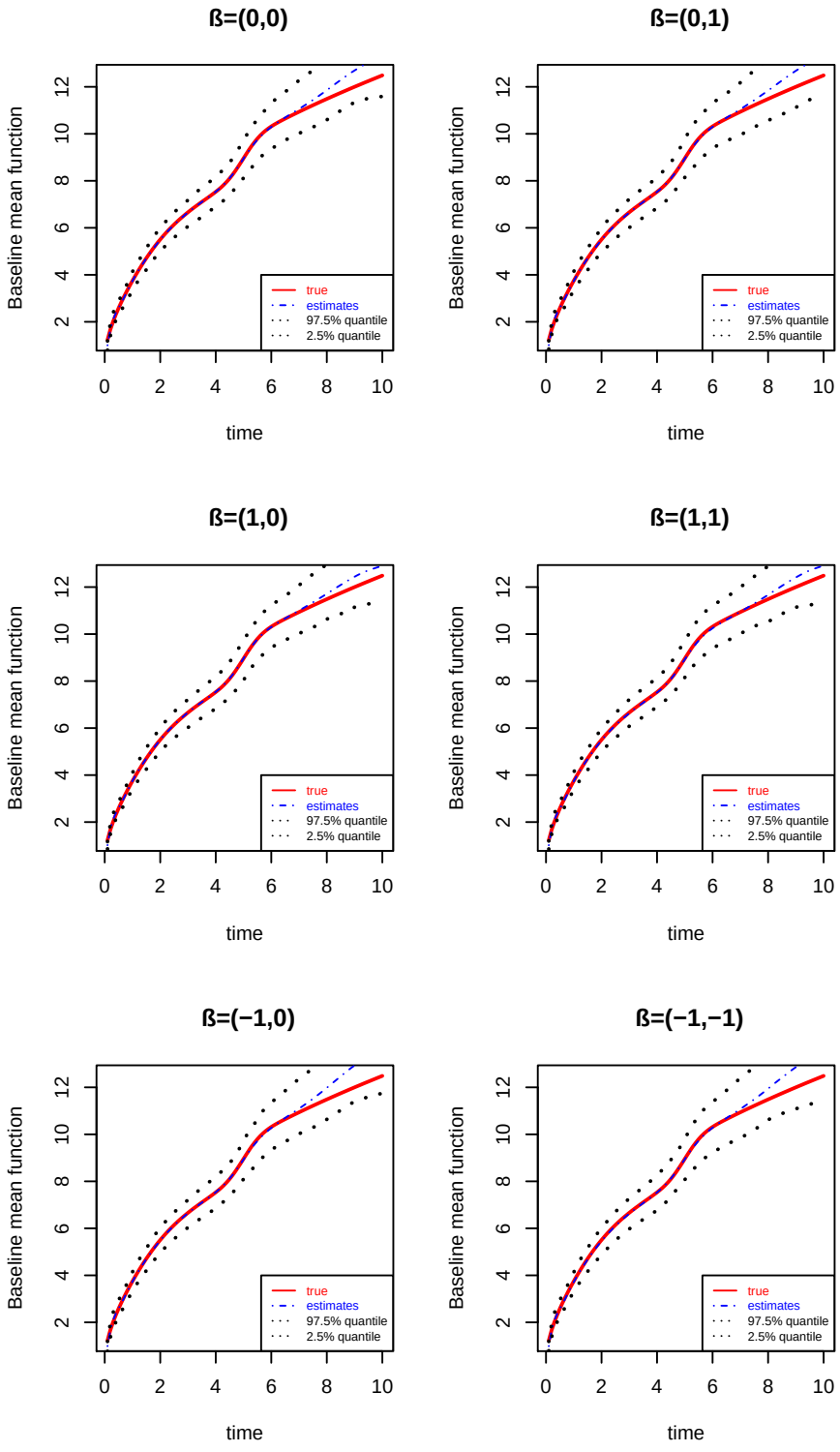

Figure 2.3: Estimates of baseline mean functions for different $\boldsymbol{\beta}$ in scenario 2

In this chapter, we also compare our proposed method with the method used by Lu, Zhang and Huang (2009) [17] for the same simulated data sets. This method was originally proposed by Rosen (1960) [65] and it was generalized by Jamshidian (2004) [66] and Zhang and Jamshidian (2004) [67]. Lu et al. (2009) [17] modified the generalized Rosen algorithm for computing the estimates of the coefficients of B-splines and regression parameters jointly through maximizing likelihood subject to nondecreasing constraints.

In order to compare with our proposed method, we choose quadratic B-splines for scenario 1, and cubic B-splines for scenario 2 to approximate the logarithm of true baseline mean function

$$\log(U_0(t)) \approx \sum_{j=1}^{q} \alpha_j \boldsymbol{\beta}_j(t),$$

subject to $\alpha_1 \leq \cdots \leq \alpha_q$. The results about the estimates of regression parameters are presented in Tables 2.1 and 2.2 for scenario 1 and scenario 2, respectively. The results are very similar to those from the parametric method. The estimates about the regression coefficients suggest that there is no bias as our proposed method and the parametric method. The average of the 500 posterior standard deviations ESDs are consistently smaller than SSEs, which also leads to bad coverage probabilities for most cases.

For the estimates of baseline function, visually from Figure 2.4 and Figure 2.5, Rosen method estimates the true baseline function better than the parametric method for scenario 2, but it still doesn't fit as well at the period with multiple reflection points as our proposed method.

MSEs of the baseline mean function from each method are presented in Table 2.3 in order to further compare the estimation of the baseline mean functions for the three different methods. MSEs are calculated as follows. We set a time sequence between 0.1 and 10 with 100 equally-spaced points, and obtain its estimates of baseline mean function at each time point for each data set, then get the mean estimates of the

baseline function based on 500 simulated data sets. MSEs are obtained by summation of all the squares of the differences between the mean estimates and the true baseline values divided by the number of estimate points. Here we just present MSEs based on the first 75 estimated points. Since most of data lie in before 7.5 units time and there are only few observations after 7.5 units time. Table 2.3 indicates that our proposed method has relatively smaller MSE than parametric method and the adapted Rosen method (Lu et al. 2009 [17]).

Another drawback of using the MLE method for the parametric model is the nonconvergence problem. We noticed that around 26% of data sets have no convergent results if random initial values are given for the MLE method. When we take the Bayesian estimates as the initial values, the MLE method converges for all data sets.

Table 2.3: Comparison MSE of baseline mean functions for different $\boldsymbol{\beta}$ among the proposed, parametric and Rosen methods for scenarios 1 and 2

	$\boldsymbol{\beta}$	(0,0)	(0,1)	(1,0)	(1,1)	(-1,0)	(-1,1)
Scenario 1	Proposed method	0.0001	0.0005	0.0001	0.0001	0.0004	0.0008
	Parametric method	0.0006	0.0012	0.0008	0.0008	0.0008	0.0008
	Rosen method	0.0019	0.0024	0.0008	0.0013	0.0022	0.0020
Scenario 2	Proposed method	0.0041	0.0036	0.0018	0.0021	0.0067	0.0059
	Parametric method	0.1024	0.1040	0.1024	0.1153	0.1073	0.1042
	Rosen method	0.0090	0.0081	0.2932	0.0084	0.0121	0.0086

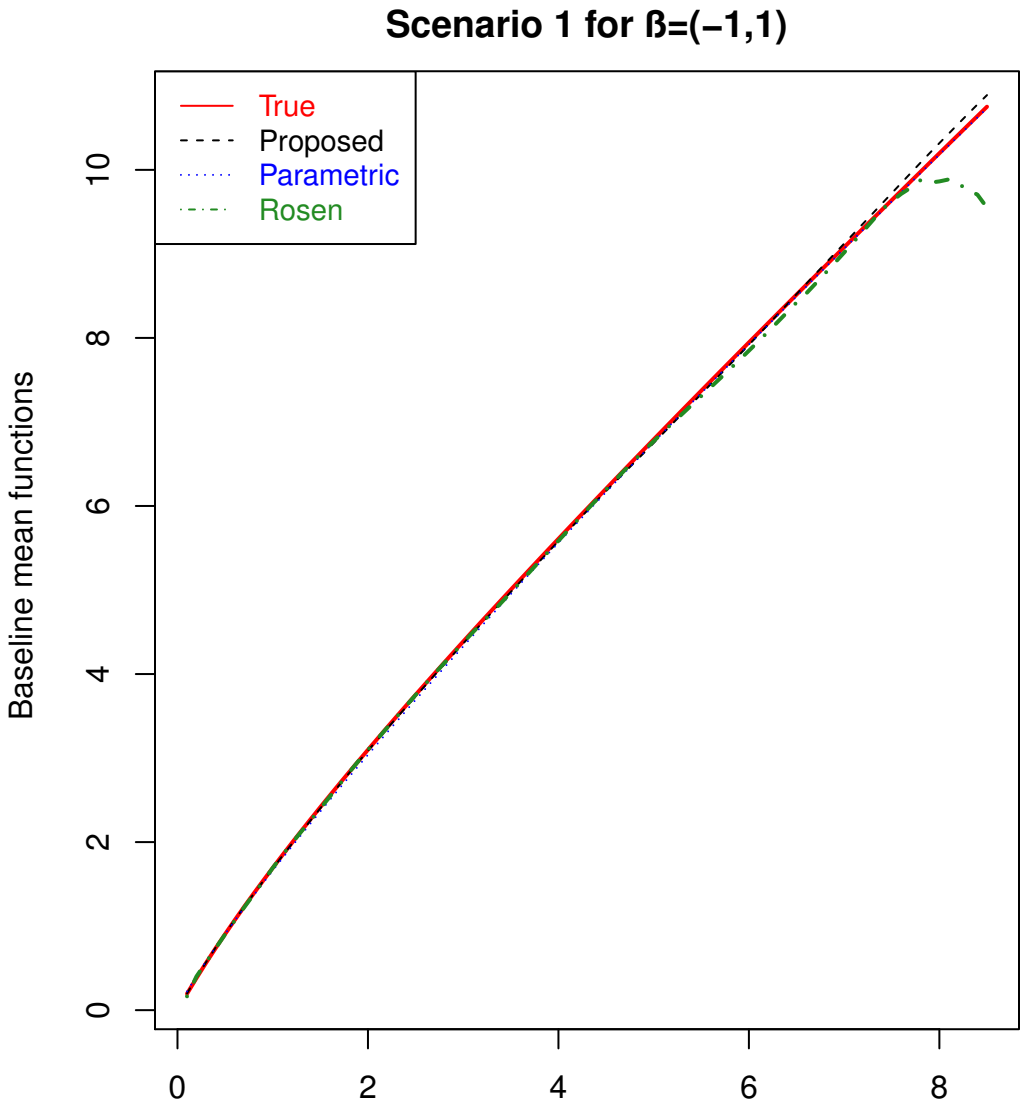

Figure 2.4: Comparison among the true and estimates of cumulative baseline mean functions among the proposed method, parametric method and Rosen method for $\boldsymbol{\beta} = (-1, 1)$ for scenario 1

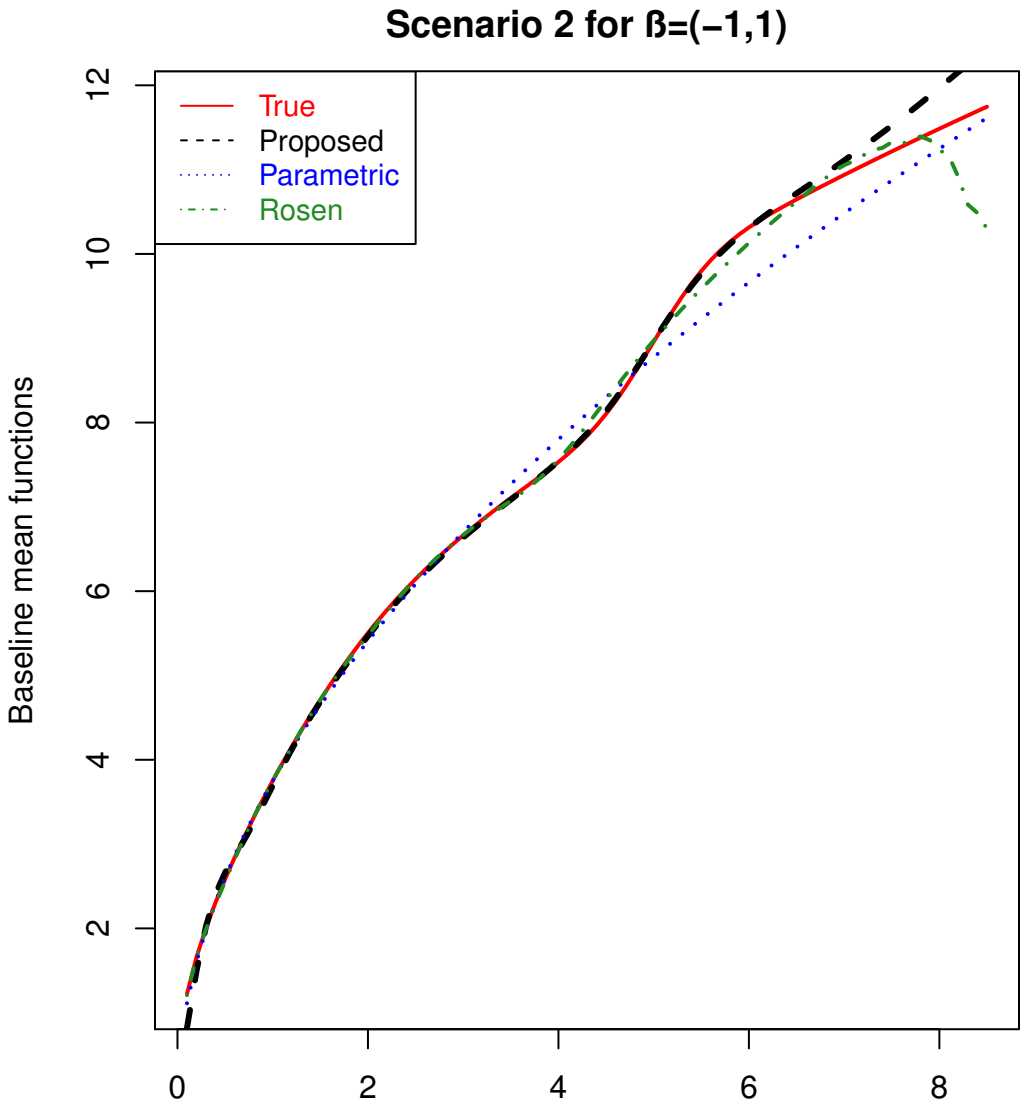

Figure 2.5: Comparison among the true and estimates of cumulative baseline mean functions among the proposed method, parametric method and Rosen method for $\boldsymbol{\beta} = (-1, 1)$ for scenario 2

2.7 A Real Data Example

In this section, we apply the proposed method to analyze the bladder tumor data as described in Chapter 1. Although a total of 118 subjects were observed, we just use 116 subjects because two of them had missing values due to no follow-up visit. y. For the first time visit, the number of the tumors was counted, the initially largest tumor size was measured, and all the initial tumors were removed transurethrally. Then these patients were randomly assigned to one of the three treatment groups: 47 of them to the placebo group, 31 of them to the pyridoxine pill group, and 38 of them to the thiotepa instillation group. At each follow-up visit for each patient, the same procedure was followed: the number of the new tumors was counted, then all of them were transurethrally removed, and then each patient was given the same treatment as before. A lot of patients had multiple follow-up visits, and had tumors recurred through the entire study. In all three treatment groups, the average follow-up time was about 31 months, but some patients may have been followed as long as five year. The covariates for this data set are denoted as $\mathbf{X} = (X_1, X_2, X_3, X_4)$, for patient $i = 1, \ldots, 116$, where X_1 and X_2 represent the number and the size of the tumors at the entry of the study, X_3 and X_4 are the indicators of the treatments of pyridoxine pills and thiotepa instillation, respectively.

In order to get the best spline specification, we consider different combinations in terms of degrees and knots, we choose degrees as 2 or 3, and choose the number of knots from 10 to 30. Table 2.5 presents the estimates of covariate coefficients and standard deviation and DICs from some combinations (not all combinations are listed here) for each degree and knot combination. The estimated posterior means and standard deviations of covariate coefficients are very similar along with the similar DCIs across the different combinations of degree and knots, which implies that this real data analysis is robust to the spline specification. The DIC results suggest that the best spline specification be with degree 3 and 25 knots, since it has the smallest

Table 2.4: The results from three different methods: proposed method, Rosen method and paramatric method for bladder tumor cancer data. Variables: X1 and X2 represent the number and the largest size of bladder tumors at beginning of the trial, respectively. X3 and X4 are indicators for pyridoxine pill and thiotepa installation, respectively.

Method	Proposed method		Rosen method		Parametric method	
Variable	$\hat{\beta}$	$\hat{sd}(\hat{\beta})$	$\hat{\beta}$	$\hat{sd}(\hat{\beta})$	$\hat{\beta}$	$\hat{sd}(\hat{\beta})$
X1	0.205	0.020	0.208	0.083	0.206	0.012
X2	-0.040	0.026	-0.035	0.085	-0.038	0.017
X3	0.062	0.094	0.063	0.414	0.081	0.073
X4	-0.801	0.112	-0.798	0.342	-0.781	0.095

DIC. In the following, we summarize the result based on this spline specification.

The number of the tumors at the entrance (X_1) of the study is positively related to the recurrence of the bladder tumor. The individuals with one more tumor when entering the study tend to have the number of recurrent tumors increased by around 22.7%. Thiotepa instillation (X_4) effectively decreases the number of recurrent tumors. On average, the number of recurrent tumors of patients in the thiotepa instillation treatment group is about 44.9% of that in the placebo treatment group for the patients with the same tumor number and size. In other words, the thiotepa instillation treatment significantly suppresses the recurrence of the tumors. The size of tumors (X_2) and pyridoxin pills (X3) are not significantly related to the number of recurrent tumors at follow-up visits. All of the results based on our proposed method are consistent with the results in Wellner and Zhang (2007) [15], Sun and Wei (2000) [68], and Lu, Zhang and Huang (2009) [17]. We also compare the results of our proposed method with those using the parametric method and Lu' s method (called Rosen method) in Table 2.5. Similarly, as observed in the simulation study, the point

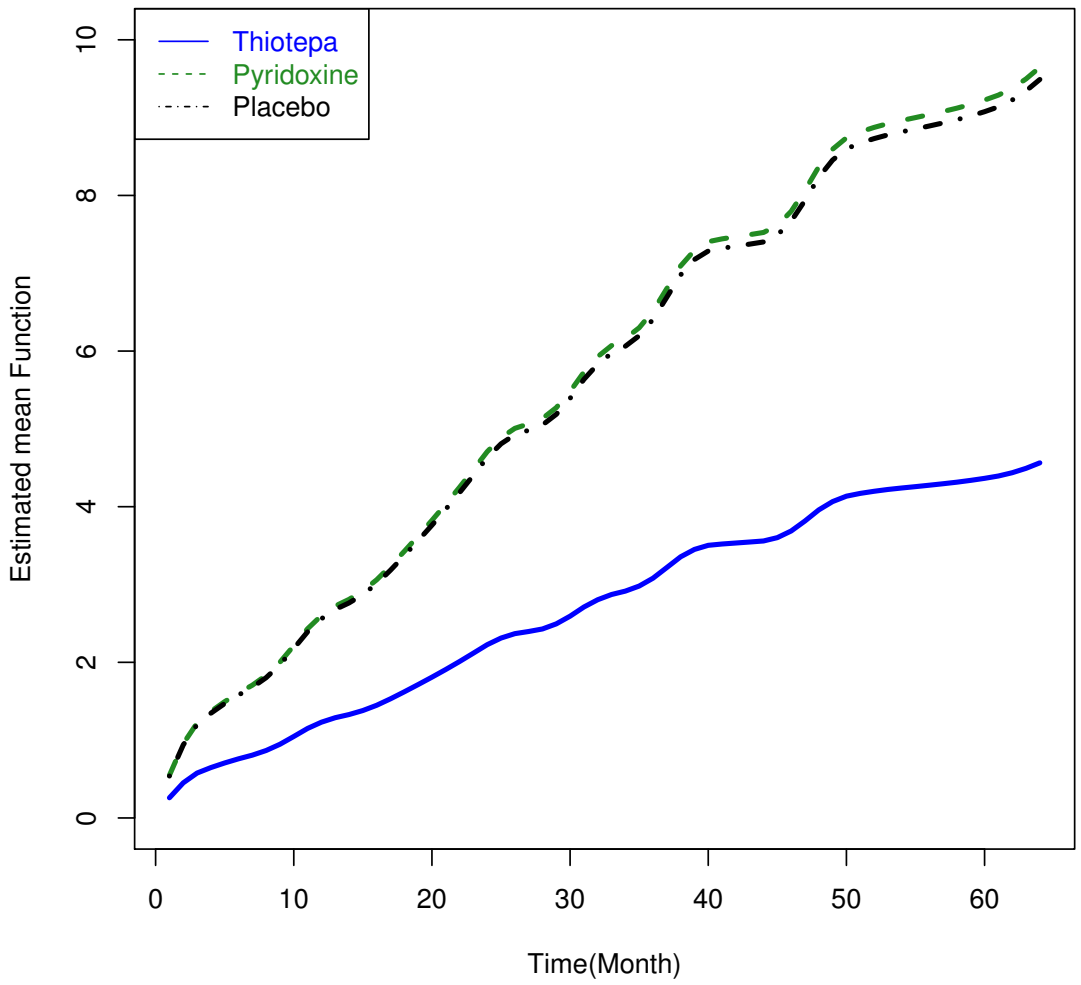

Figure 2.6: Plot for estimated mean function for different treatments

estimates of covariate coefficients are close to those from our proposed methods but with smaller standard errors. Figure 2.6 shows the estimation of the mean cumulative function for the three treatments. The estimated mean functions for the placebo group and pyridoxine pill group are very close to each other, while the estimated mean function of the thiotepa group is apparently below them, which indicates that the treatment of thiotepa significantly lowered the number of recurrent tumors.

Table 2.5: Results about regression parameters for different knots with degrees 2 and 3

d	Kn	β_1		β_2		β_3		β_4		DIC
		est.	sd	est.	sd	est.	sd	est.	sd	
2	15	0.2055	0.0198	-0.0356	0.0249	0.0724	0.0911	-0.7969	0.1110	1190.288
	19	0.2041	0.0193	-0.0368	0.0285	0.0606	0.0882	-0.8028	0.1155	1196.649
	20	0.2053	0.0204	-0.0388	0.0261	0.0635	0.0953	-0.7976	0.1149	1177.684
	25	0.2049	0.0199	-0.039100	0.0264	0.0597	0.0942	-0.8009	0.1129	1182.369
3	15	0.2086	0.0195	-0.0382	0.0246	0.0551	0.0906	-0.8092	0.1236	1192.616
	19	0.2039	0.0205	-0.042	0.0268	0.0615	0.0940	-0.8005	0.112	1181.535
	20	0.2047	0.0197	-0.0404	0.0260	0.0656	0.0939	-0.7967	0.1134	1192.453
	25	0.2049	0.0200	-0.0400	0.0261	0.0616	0.0947	-0.8009	0.1126	1176.094

2.8 Discussion

In this chapter, we propose a Bayesian estimation approach for semiparametric regression analysis of panel count data under the proportional mean model with an assumption of a nonhomogeneous Poisson process for the penal count response. Monotone splines are adopted to approximate the baseline mean function to reduce the number of unknown parameters while maintaining adequate modeling flexibility. Based on a Poisson data augmentation, we develop an efficient and easy-to-implement Gibbs sampler, in which the regression parameters are updated automatically using adaptive-rejection sampling and all other parameters and latent variables are directly sampled from standard distributions. Numerical evidences have shown that the proposed approach provides accurate estimates of the regression parameters and the baseline mean function and outperforms the parametric method and the adapted Rosen method (Lu et al. 2009 [17]). On the other hand, In terms of 95% coverage

probability, it is very close to 95% for our proposed method, but not close to 95% for the other two methods. Moreover, unlike parametric maximum likelihood method, our proposed method has no convergent problem. Furthermore, the proposed method is comparable to the other two methods through the real data analysis, and we have the consistent results with other published results.

Although only time-independent covariates are considered in this chapter, our approach can be adapted to accommodate time-dependent covariates. Another appealing extension of our approach is to allow subject specific baseline mean function to account for the heterogeneity among the subjects. Specifically, a gamma frailty term can be added multiplicatively to model 2.1 for each subject. A slight modification of the proposed Gibbs sampler will suffice the Bayesian posterior computation in this case while maintaining all the computation advantages.

Chapter 3

Frailty Effect on Panel Count Data

3.1 Introduction

3.1.1 Heterogeneity

Heterogeneity of subjects is very common in clinical and observational studies. For example, there is such population consisting of two subgroups with distinct failure rates, and the subjects in one subgroup are much more prone to experience events or failures than those in the other subgroup. Then the failure rate in the surviving population will fall over time if we ignore the unobserved heterogeneity and just simply use the proportional hazards model (Cox 1972 [69]). In many cases, if the heterogeneity is not appropriately taken into account, then the estimation may be inconsistent.

Heterogeneity can be caused from different sources. It may be due to differences in life situations, experiences, environment, genetic variation or heritability. For instance, some subjects may experience changes in gene expression that increase the chance to develop new tumors. However, this kind of factors are never observable or measurable. In the other cases, factors are not observed due to the lack of time effort or great cost. Vaupel et al. (1979) [70] introduced the concept of frailty to account for unobserved and unobservable heterogeneity, and Lawless (1987) [22] introduced it as a random effect in models. To summarize, the motivation for the frailty term is to explicitly account for the extra variation associated with unobserved or unmeasured risk factors in survival data. For example, Sun, Tong and He (2007) [71] studied the

bladder data. In the study, the patients visited the clinical centers periodically and at each visit, the number of bladder tumors that occurred since the last visit was recorded, measured and removed, some patients in the study had more visits than others, and the occurrence of bladder tumors and the visit may be related. The bigger number of the tumors was detected at the first time of visit, the more frequently the patients would visit the clinical centers. The correlation between the first visit times and the numbers of recurrent bladder tumors and can be explained by a frailty term in the data analysis study.

3.1.2 PH Model with Frailty Effects

Cox (1972) proposed the proportional hazard model (PH):

$$h(t|X) = h_0(t)\exp(X'\boldsymbol{\beta}), \tag{3.1}$$

where the hazard rate is a function of covariates and strictly proportional to the covariate effect. To account for unobserved or unmeasured frailties, the PH model can be extended as follows:

$$h_i(t|X) = h_0(t)\exp(X_i'\boldsymbol{\beta} + v_i\phi), \tag{3.2}$$

for $i = 1, \ldots, n$, where v_j is the frailty for subject i and assumed to be an independent sample from a distribution with mean 0 and variance 1 in Keiding et al. (1997) [72]. Notice that we have the regular PH model if $\phi = 0$. Moreover, if we can measure and include v_i in the model, then ϕ will go to zero as well. The model can be written as:

$$h_i(t|X_i, w_i) = h_0(t)w_i\exp(X_i'\boldsymbol{\beta}), \tag{3.3}$$

where $w_i = e^{v_i\phi}$. Denote $W = (w_1, \ldots, w_n)'$. From this expression, we can see that the frailty term w_i acts multiplicative on the hazard function, and the hazard rate for different subjects varies besides the covariate effect.

We notice that a scale factor common to all subjects in a population can be absorbed into the baseline hazard function. Therefore, for the identification purpose, it is conventionally assumed that the mean of W is 1, and the variance is unknown and equal to some parameter. If the variance is 0, it indicates the independence among subjects. A high value of the variance indicates a high correlation among the subjects. For standard non-frailty models, we can assume that $w_i's$ are all one with probability 1. Therefore, standard non-frailty models are naturally nested in the frailty model (3.5).

There is another frailty model which is called shared frailty model and defined as follows. Suppose that there are n groups (clusters). For the ith group, the n_i individuals are associated with a common unobserved frailty w_i. Conditional on the frailty w_i, the survival times T_{ij}'s, for $j = 1, \ldots, n_i$, are assumed to be independent and their hazard functions has the following form:

$$h_{ij}(t|X_i, w_i) = h_{0j}(t)w_i \exp(X_{ij}'\boldsymbol{\beta}). \tag{3.4}$$

Generally, these are two categories of frailty models that are discussed in the literature. The first one is the class of univariate frailty models (3.3) that consider univariate data. It is used to model heterogeneity among individuals. It was discussed with details by Vaupel et al (1979) [70], Lancaster (1979) [73], and Hougaard (1984) [74]. The second one is the class of multivariate frailty models (3.4) that take into account of multivariate (clustered) data, where the unobserved frailty within the same group is shared by the individuals in the group. The shared frailty model may be considered as a random effect model for survival data. Typical groups that share common risk factors may be the members of a family, twins, mice born in the same litter, or simply a single subject for which multiple episodes are observed (repeated measurements for one subject). For the last one, panel count data is a very good example. Clayton (1978) [75] first studied the multivariate situation on chronic disease incidence within families. More shared frailty models have been discussed in

detail by Hougaard (2000) [76], Therneau and Grambsch (2000) [77], Duchateau and Janssen (2008) [78], and Hanagal and Dabade (2013) [79].

3.1.3 Analysis of Frailty Models

For analyzing frailty models, usually it is assumed that the frailty term follows a parametric distribution. The distributions can be gamma, lognormal (McGilchrist and Aisbett 1991 [80]), inverse Gaussian family, a power variance function exponential family, or some formulas remarkably similar for the gamma and the inverse Gaussian families (Hougaad 1995 [81]). The common assumption is gamma frailty. The effect of different frailty distributions was investigated by Congdon (1995) [82]. Under the parametric proportional hazards model, the estimation of the regression coefficients, baseline function and variance components can be obtained together through maximizing log-likelihood method (Andersen et al 1993 [31]). There are a lot of literature analyzing gamma frailty models parametrically such as Sahu et al. (1997) [83], Yu (2006) [84], Ibrahim et al. (2013) [85].

Estimation of frailty models can also be semi-parametric, which has been much widely discussed to date. In this case, the baseline function is usually left unspecified, and advanced techniques are applied such as splines to analyze this situation (Abrahantes et al. 2007 [86]). The baseline hazard is often approximated by splines such as B-splines (Lu. et al. 2009 [17], Sharef et al. 2010 [87]), S-splines, polynomial splines, or smoothing spline ANOVA (Du et al. 2011 [88]). Since frailties are usually regarded as unobserved covariates, this has led to the use of the expectation-maximization (EM) algorithm as a very common estimation tool. There are a lot of literature where the EM algorithm has been applied in frailty models (Klein 1992 [89]), Nielsen et al. 1992 [90], Wang et al. 2015 [91], and Yao et al. 2016 [63]). The frailties as random effects were also intensively studied through penalized likelihood methods and hierarchical likelihood combined with splines tools. For example, see

Do Ha (2011) [92], Therneau et al. (2003) [93], and Du and Ma (2010) [94].

We investigate the frailty effects on the proportional mean model for panel count in the rest of the chapter. A gamma-frailty is used to account for the heterogeneity across subjects. The rest of the chapter is organized as follows. In Section 3.2, we present the data structure, and the proportional mean model with frailty effects. In Section 3.3, the likelihood and the augmented likelihood are presented. Priors for the unknown parameters and hyper parameters are assigned, and the full conditional posterior distributions are derived in Section 3.4. In Section 3.5, we evaluate the proposed approach first through simulation studies, and investigate how the frailty affects the frailty model and no frailty model (2.1) for frailty data, respectively. In Section 3.6, the utility of the proposed approach is illustrated with the bladder tumor data. The results are compared with the results in Chapter 2 where no frailty model was applied. Finally, in Section 3.7 we give a brief summary and discuss potential directions for further research.

3.2 Proportional Mean Model with Frailty Effect

As we discussed, panel count data can be treated as cluster data. All the observations from the same subject can be thought of from a cluster, and the same cluster shares the same frailty random variable in frailty models. Suppose we have n independent clusters (subjects) for a panel count data set. For each cluster i, the total number of observation times is denoted by K_i, and $\{t_{ij}\}_{j=1}^{K_i}$ is a random time sequence. Let $N_i(t)$ denote the number of events observed before and at time t for cluster i. Then the whole set of data can be described as $\mathscr{D} = \{t_{ij}, N_i(t_{ij}), X_i, \text{ for } i = 1, \ldots, n,\ j = 1, \ldots, K_i\}$, X_i is a vector of covariates.

In order to explicitly account for the extra variation associated with unobserved or unmeasured factors and the correlation among the cumulative count observations for the same subject in panel count data, frailty effects $W = (w_1, \ldots, w_n)'$ are introduced

into the proportional mean model (2.1). Following the same flavor as the PH frailty model, the mean function of the counting process $N_i(t)$ is

$$E\Big[N_i(t)|X_i, w_i\Big] = U_i(t)|X_i, w_i = U_0(t)w_i\exp(X_{ij}'\boldsymbol{\beta}), \tag{3.5}$$

where $U_0(t)$ is an unknown, nonegative, and nondecreasing baseline mean cumulative function, X_i is a p-dimensional covariate vector, $\boldsymbol{\beta}$ is a vector of covariate effects, and w_i is frailty variable for ith cluster. We can see that model (2.1) is the special situation of model (3.5) when w_i is identical to be 1.

Frailty models for panel count data has been studied from frequentist and Bayesian perspectives. Sun and Kalbfleisch (1995) [5], Zhang (2002) [95], and Wellner et al. (2004) [49] studied nonparametric maximum pseudo-likelihood estimator, nonparametric maximum likelihood estimator, and made comparison of the estimators between the maximum pseudo-likelihood estimator and the maximum likelihood estimator through large-sample properties. Sun and Wei (2000) [68] formulated estimation equations for regression parameters in the semiparametric proportional rate models. Huang et al. (2006) [52] had the estimates of the baseline rate function and the regression parameters by maximizing a conditional likelihood function of observed event counts and solving estimation equations. Wang et al. (2013) [96] proposed an augmented estimating equation (AEE) approach for a semiparametric mean regression model with panel count data under possibly informative observation schemes and censoring, where the regression coefficients and the unspecified baseline mean function were estimated with an Expectation-Solving (ES) algorithm. Hua et al. (2014) [97] studied a gamma-frailty non-homogeneous Poisson process model for analyzing over-dispersed panel count data, in which a cubic B-spline function was used to approximate the logarithm of the baseline mean function in the semiparametric proportional mean model, and regression parameters and spline coefficients were jointly estimated by maximizing a spline-based sieve pseudo-likelihood and by replacing the nuisance over-dispersion parameter with its moment estimate. Sinha

and Maiti (2004) [48] proposed a Bayesian approach for the analysis of panel count data with dependent termination due to an adverse response such as detection of a relapse to malignancy or due to a positive response such as cure. They assigned frailty with an independent gamma prior, and the unknown baseline mean cumulative function with a joint increment prior (Kalbfleisch, 1978 [98]), then derived the full conditional distributions.

There are a lot of ways to study model (3.5), one may focus on the prediction of counts between time intervals, or focus on estimation of gap time hazard, or other estimations of interest such as the baseline mean function and regression coefficients. For example, Du et al. (2011) [88] studied the gap time hazard estimation in recurrent event data. They used smoothing spline analysis of variance decomposition to model log gap time hazard, and introduced general frailty to account for between-subject heterogeneity and within-subject correlation. Here we focus on the counts that occur between two adjacent time points, and aim to estimate the baseline mean function and the regression coefficients jointly.

For the baseline function, it is usually left unspecified. Approximation of the baseline function can be achieved by splines such as B-splines (Lu et al. 2009 [17], Sharef et al. 2010 [87]), S-splines, polynomial splines, or smoothing spline ANOVA (Du et al. 2011 [88]). Here we still model the unknown baseline mean cumulative function by monotone I-splines as in Chapter 2:

$$U_0(t) = \sum_{l=1}^{L} r_l I_l(t|d), \tag{3.6}$$

where $I_l(\cdot|d)$ is a basis spline function with degree d, which is nondecreasing from 0 to 1, and r_ls are nonnegative spline coefficients to ensure the monotonicity of the nonnegative function $U_0(t)$. The detailed formulation of the basis function $I_l(\cdot|d)$ can be found in chapter 1. The degree d and the knot placement are two crucial components to determine the basis functions. The number L of the basis functions is equal to the degree d plus the number m of interior knots and plus 1.

For the frailty random variable, various distribution assumptions have been proposed in the literature of Duchateau and Janssen (2008) [78]. The most common assumption for frailty distribution is to use a gamma distribution. The popularity of gamma frailty in survival model is contributed by the following reasons. First, a gamma distribution has a positive support, so it is suitable to explain the positive correlation of the panel count data. Second, a gamma distribution is a very flexible distribution. If the shape parameter is known, especially when it is equal to one, it is identical to the well-known exponential distribution. If the shape parameter is known but not equal to one, and the rate parameter is unknown, then a gamma distribution for the rate forms a conjugate prior for some models in Bayesian method. It can be chi-squared, inverse-gamma, generalized gamma distributions when different shape and sale parameters are used or some transformations are made. Sometimes a negative binomial distribution is also considered as the discrete analogue of the gamma distribution. Third, Abbring and van den Berg (2007) [99] rationalized the use of gamma distributions for frailties in time-to-event data analysis. They showed that in a large class of univariate and multivariate frailty models, the distribution of the frailty among survivors converges to a gamma distribution under mild regularity assumptions. Fourth, based on a computational and analytic point of view, it is easy to derive the closed form for unconditional survival, cumulative density and hazard function due to the simplicity of the Laplace transform. These calculations were demonstrated through the Laplace transform by Hougaard (1984) [74].

In this chapter, we assume that the frailty terms w_i independently and identically follow a gamma distribution with both shape and rate parameters equal to ϕ.

3.3 Likelihood and Augmentation

Denote Z_{ij} the count of recurrences within the jth time interval $[t_{i,j-1}, t_{ij})$. That is, Z_{ij} is the count difference between time points $t_{i,j-1}$ and $t_{i,j}$. By the definition of

$N_i(t)$, we have $Z_{ij} = N_i(t_{ij}) - N_i(t_{i,j-1})$, for $i = 1, \ldots, n$, and $j = 1, \ldots, K_i$. Under the non-homogeneous Poisson process assumption of $N_i(t)$, Z_{ij} independently follows Poisson distribution conditional on the covariates X_i and the frailty w_i,

$$Z_{ij}|X_i, w_i \sim \mathcal{P}\bigg(\Big\{U_0(t_{ij}) - U_0(t_{ij-1})\Big\}\exp(X_i'\boldsymbol{\beta})w_i\bigg). \tag{3.7}$$

Then the observed likelihood function is

$$\begin{aligned} L(\theta|\mathscr{D}) = \prod_{i=1}^{n}\prod_{j=1}^{K_i} \exp\bigg[&- \Big\{U_0(t_{ij}) - U_0(t_{ij-1})\Big\}\exp(X_i'\boldsymbol{\beta})w_i\bigg] \\ &\times \bigg[\Big\{U_0(t_{ij}) - U_0(t_{ij-1})\Big\}\exp(X_i'\boldsymbol{\beta})w_i\bigg]^{Z_{ij}} \Big/ Z_{ij}!, \end{aligned} \tag{3.8}$$

where θ is the vector of all parameters and latent frailties w_i. With the monotone I-spline specification $U_0(t) = \sum_{l=1}^{L} r_l I_l(t)$, we have

$$\begin{aligned} L(\theta|\mathscr{D}) = \prod_{i=1}^{n}\prod_{j=1}^{K_i} \exp\bigg[&- \Big\{ \sum_{l=1}^{L} r_l I_l(t_{ij}) - \sum_{l=1}^{L} r_l I_l(t_{ij-1})\Big\}\exp(X_i'\boldsymbol{\beta})w_i\bigg] \\ &\times \bigg[\Big\{ \sum_{l=1}^{L} r_l I_l(t_{ij}) - \sum_{l=1}^{L} r_l I_l(t_{ij-1})\Big\}\exp(X_i'\boldsymbol{\beta})w_i\bigg]^{Z_{ij}} \Big/ Z_{ij}!. \end{aligned} \tag{3.9}$$

This likelihood function is very complex, especially for the term

$$\bigg[\Big\{ \sum_{l=1}^{L} r_l I_l(t_{ij}) - \sum_{l=1}^{L} r_l I_l(t_{ij-1})\Big\}\exp(X_i'\boldsymbol{\beta})w_i\bigg]^{Z_{ij}},$$

which makes hard to sample r_ls. To facilitate the computation, independent Poisson latent variables $Z_{ij1}, \ldots, Z_{ijL}$ are introduced to decompose Z_{ij} such that $Z_{ij} = \sum_{l=1}^{L} Z_{ijl}$. By the additivity property of Poisson random variable, each single Z_{ijl} independently follows a Poisson distribution as follows:

$$Z_{ijl}|X_i, w_i \sim \mathcal{P}\bigg(\Big\{r_l I_l(t_{ij}) - r_l I_l(t_{ij-1})\Big\}\exp(X_i'\boldsymbol{\beta})w_i\bigg),$$

for $i = 1, \ldots, n,\ j = 1, \ldots, K_i$ and $l = 1, \ldots, L$.

Then the augmented likelihood function with Z_{ijl}s is

$$L_{aug} = \prod_{i=1}^{n}\prod_{j=1}^{K_i}\prod_{l=1}^{L} \exp\bigg[- \Big\{r_l I_l(t_{ij}) - r_l I_l(t_{ij-1})\Big\}\exp(X_i'\boldsymbol{\beta})w_i\bigg]$$

$$\times \left[\Big\{ r_l I_l(t_{ij}) - r_l I_l(t_{ij-1}) \Big\} \exp(X_i' \boldsymbol{\beta}) w_i \right]^{Z_{ijl}} \Big/ Z_{ijl}!. \tag{3.10}$$

From the likelihood function (3.10), we can easily factor out all the parameters r_ls.

3.4 Prior Specification and Posterior Computation

Now we assign priors for unknown parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ and $r = (r_1, \ldots, r_L)'$. For $\beta_m,\ m = 1, \ldots, p$, we assign independent normal prior with mean zero and large variance $N(0, \sigma_m^2)$ with $\sigma_m^2 = 100$. For $r_l,\ l = 1, \ldots, L$, we assign independent exponential priors Exp(λ) with a gamma hyper prior $\mathcal{G}(a_\lambda, b_\lambda)$ on λ. Here we choose $a_\lambda = 1$ and $b_\lambda = 1$ and find that it is robust to choose values of a_λ and b_λ. For the parameter ϕ from the gamma frailties, we assign it a hyper prior as gamma distribution $\mathcal{G}(c, c)$ with parameter c. We assign c a small value to make the variance of ϕ large. Here we choose $c = 0.1$.

We adopt Gibbs sampling for sampling the joint posterior distribution. Typically, it requires sampling all the unknown parameters and latent variables from their full conditional distributions. Based on the augmented likelihood function (3.10) and the above specified priors, the Gibbs sampler is derived as below. After specifying the initial values of the unknown parameters, our Gibbs sampler iterates through the following steps:

1. Sample r_l from a gamma distributions $\mathcal{G}(A_l, B_l)$, for $l = 1, \ldots, L$, where

$$A_l = \sum_{i=1}^{n} \sum_{j=1}^{K_i} Z_{ijl} + 1,$$

$$B_l = \sum_{i=1}^{n} \Big\{ I_l(t_{iK_i}) - I_l(t_{i0}) \Big\} \exp(X_i' \boldsymbol{\beta}) w_i + \lambda.$$

2. Sample β_m by using the adaptive rejection sampling (ARS) (Gilks and Wild 1992 [58]) method, for $m = 1, \ldots, p$, because the full conditional distribution of each β_m is proportional to

$$\exp\Big[-\sum_{i=1}^{n}\sum_{l=1}^{L} r_l\Big\{I_l(t_{iK_i}) - I_l(t_{i0})\Big\}\exp(X_i'\boldsymbol{\beta})w_i + \sum_{i=1}^{n}\sum_{j=1}^{K_i}(X_i'\boldsymbol{\beta})Z_{ij} - \sum_{m=1}^{p}\beta_m^2/(2\sigma^2)\Big],$$

which is a log-concave function of β_m. Note that $I_l(t_{i0}) = 0$.

3. Sample Z_{ijl}, for $i = 1, \ldots, n,\ j = 1, \ldots, K_i$ and $l = 1, \ldots, L$, from a multinomial distribution with parameters Z_{ij} and $\mathbf{P}_{ij}$, denoted by $\mathcal{M}(Z_{ij}, \mathbf{P}_{ij})$, where $\mathbf{P}_{ij} = (p_{ij1}, \cdots, p_{ijL})$ with $\sum_{l=1}^{L} p_{ijl} = 1$ and

$$p_{ijl} = \frac{r_l\{I_l(t_{ij}) - I_l(t_{i,j-1})\}}{\sum_{l=1}^{L} r_l\{I_l(t_{ij}) - I_l(t_{i,j-1})\}},$$
$$(Z_{ij1}, \cdots, Z_{ijL}|Z_{ij}) \sim \mathcal{M}(Z_{ij}, \mathbf{P}_{ij}).$$

4. Sample λ from a gamma distribution $\mathcal{G}(a_\lambda + L, b_\lambda + \sum_{l=1}^{L} r_l)$.

5. Sample w_i from a gamma distribution $\mathcal{G}(c_i, d_i)$ for $i = 1, \ldots, n$, where

$$c_i = \phi + \sum_{j=1}^{K_i} Z_{ij},$$

$$d_i = \phi + \sum_{l=1}^{L} r_l b_l(t_{iK_i})\exp(X_i'\boldsymbol{\beta}).$$

6. Sample ϕ from

$$p(\phi|\cdot) \propto \prod_{i=1}^{n}\left(\frac{\phi^\phi w_i^{\phi-1}\exp(-\phi w_i)}{\Gamma(\phi)}\right)\frac{c^c\phi^{c-1}\exp(-\phi c)}{\Gamma(c)}$$

$$\propto \frac{\phi^{n\phi+c-1}(\prod_{i=1}^{n} w_i)^{\phi-1}\exp\Big(-(c+\sum_{i=1}^{n} w_i)\phi\Big)}{\Big(\Gamma(\phi)\Big)^n},$$

which is log-concave. We then sample it through the adaptive rejection sampling (ARS) method.

The above developed Gibbs sampler is efficient and easy to implement because all of the full conditional distributions either have closed forms or they are log-concave.

3.5 Simulation

We have two main purposes for simulation studies. One purpose is to evaluate the effectiveness of the proposed approach for dealing with the frailty effects. The second purpose is to compare frailty model (3.5) with no frailty model (2.1) and to investigate how the model misspecification affects the final inferences.

3.5.1 Data Generation

There are two sets of data we generated. The first set is the same as the data generated in simulation part of Chapter 2. The second set of data will be generated by considering the frailty effects. How to generate the first set of data is described in Chapter 2. In order to make the results comparable, the second set of data is generated by the same setting as the first set of data except for adding the frailty item. The event count of each subject i is assumed to follow a non-homogeneous Poisson process with the mean function modeled as (3.5),

$$E(N_i(t)|X_i, w_i) = U_i(t|X_i, w_i) = U_0(t)\exp(X_{ij}'\boldsymbol{\beta})w_i.$$

The frailty effects w_is are independent, and identically follow $\mathcal{G}(\phi, \phi)$ with $\phi \sim \mathcal{G}(c, c)$ and $c = 0.1$. We still study the two scenarios for the baseline function.
Scenario 1: $U_0(t) = t + log(1 + t)$, which is approximately linear;
scenario 2: $U_0(t) = 0.5\Phi(t, 1, 1) + 0.5\Phi(t, 5, 0.5) + t^{0.5}$, where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard normal distribution. It is curvilinear and has multiple reflection points.

3.5.2 Evaluating the Proposed Approach

In this subsection, we evaluate the effectiveness of the proposed approach for dealing with the frailty effects. We check the estimation of the regression coefficients and the baseline functions. For fitting the model, we still use DIC to determine the

Table 3.1: Estimates of regression parameters for scenario 1, when fitting model with frailty effect for frailty data

		Results on β_1				Results on β_2				
β_1	β_2	Bias	SSE	ESD	CP95	Bias	SSE	ESD	CP95	$1/\phi$
0	0	-0.0128	0.1878	0.1840	0.942	-0.0118	0.1773	0.1871	0.960	0.96
0	1	-0.0226	0.1856	0.1856	0.938	0.0045	0.1941	0.1935	0.962	0.97
1	0	-0.0071	0.1744	0.1776	0.966	0.017	0.1738	0.1804	0.956	0.97
1	1	-0.0144	0.1651	0.1778	0.970	0.0058	0.1989	0.1832	0.936	0.97
-1	0	-0.0278	0.1988	0.2014	0.942	-0.0221	0.2005	0.1967	0.960	0.95
-1	1	-0.0072	0.2070	0.2014	0.946	-0.0045	0.2011	0.2060	0.962	0.96

degree and the number of knots for the I-spline specification. The details on how to calculate DIC's are presented in Chapter 1. We choose 20 equally-spaced knots between the minimum and the maximum of observation times from each data set. For scenario 1, we choose degree 2. For scenario 2, we choose degree 3. Based on the selection of knots and degree for each scenario, we implement the proposed Gibbs sampler in Section 3.4. The results suggest that there is a good mixing and fast convergence in MCMC chains of the regression parameters and the spline coefficients regardless of scenarios. All the summarized results are based on 5000 iterations of the Gibbs samples after discarding the first 1000 iterations as a burn-in.

Tables 3.1 and 3.2 present the characteristics of estimates of the regression parameters for both scenarios. Bias denotes the difference between the average of 500 point estimates and the true value of $\boldsymbol{\beta}$, SSE as the sample standard deviation of the 500 point estimates, ESD as the average of the 500 posterior standard deviations, and CP95 as the 95% coverage probability. The point estimates are taken to be posterior means, and the coverage probabilities are based on the 95% credible intervals

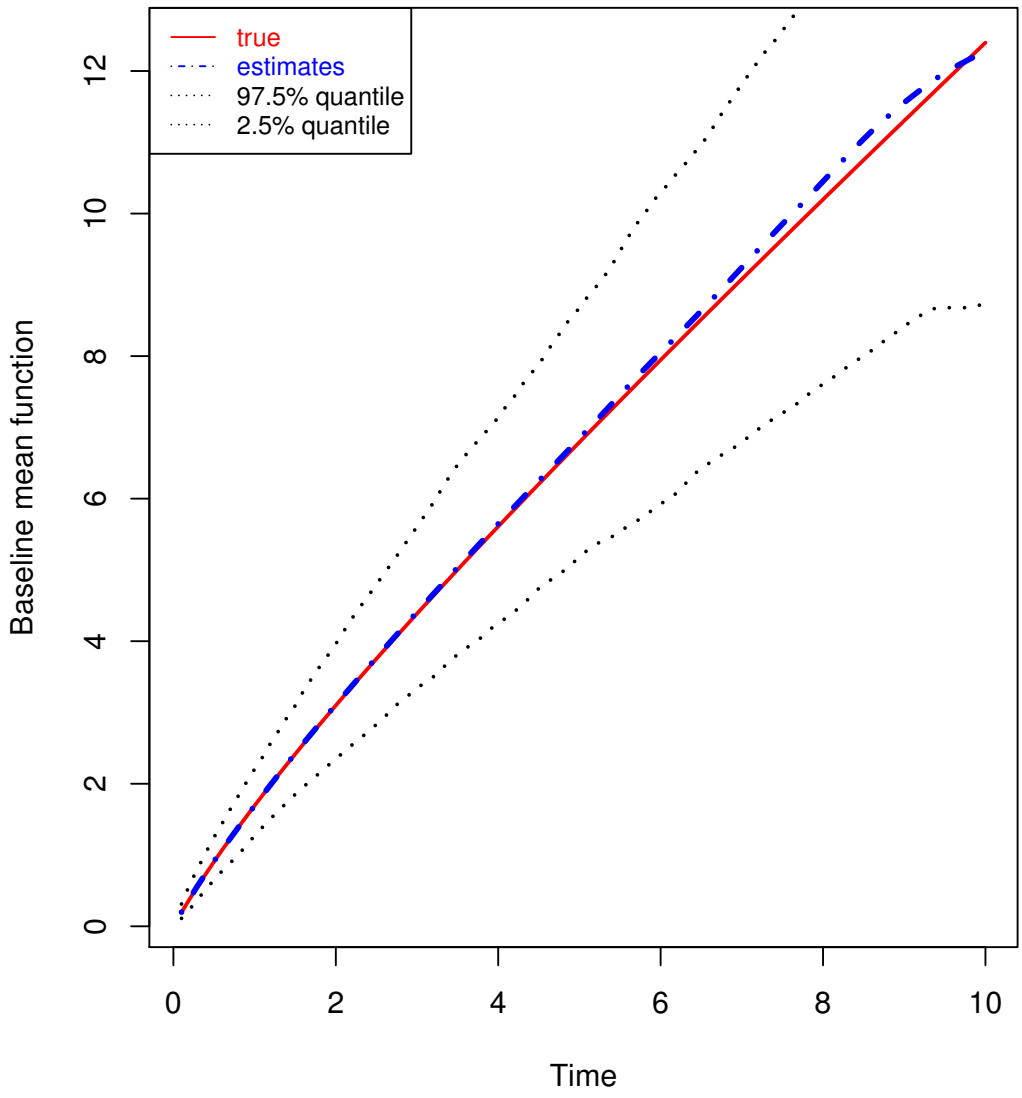

Figure 3.1: Estimates of baseline function with β=(-1,1) for scenario 1

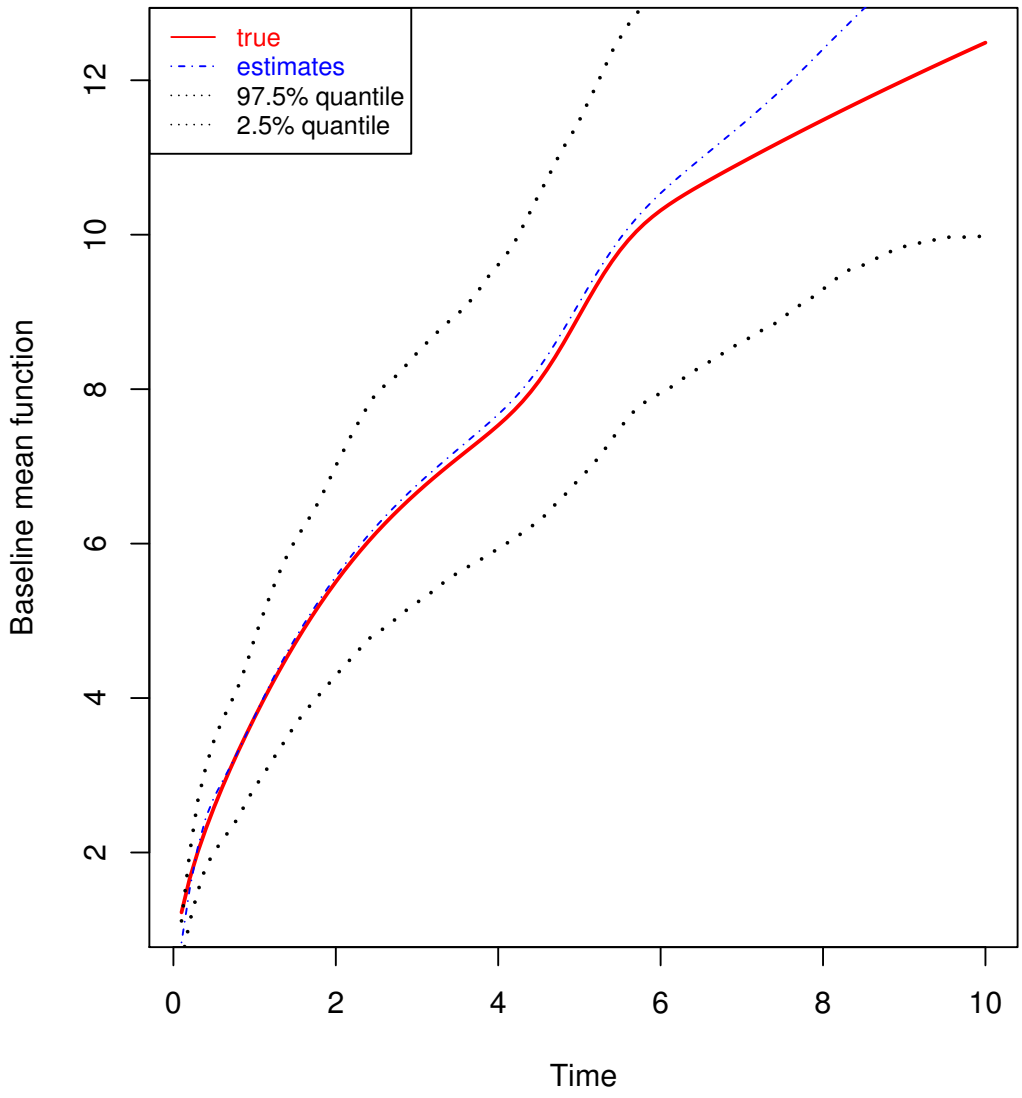

Figure 3.2: Estimates of baseline function with β=c(-1,1) for scenario 2

Table 3.2: Estimates of regression parameters for scenario 2, when fitting the model with frailty effect for frailty data

		Results on β_1				Results on β_2				
β_1	β_2	Bias	SSE	ESD	CP95	Bias	SSE	ESD	CP95	$1/\phi$
0	0	-0.0078	0.1799	0.1763	0.928	-0.011	0.1715	0.1802	0.974	0.99
0	1	-0.0057	0.1735	0.1781	0.952	0.0072	0.1925	0.1842	0.936	0.98
1	0	-0.0019	0.1795	0.1724	0.926	-0.0077	0.1736	0.1739	0.940	0.98
1	1	-0.0092	0.1816	0.1721	0.944	-0.0101	0.1862	0.1782	0.942	1.61
-1	0	-0.0186	0.2016	0.1876	0.936	0.0107	0.1943	0.1913	0.952	0.98
-1	1	-0.0102	0.1901	0.1886	0.948	0.006	0.2081	0.1931	0.932	0.97

constructed by 2.5th and 97.5th percentiles.

Based on the results from Tables 3.1 and 3.2, we can see that the proposed method performs very well. All the estimated $\boldsymbol{\beta}$s are very close to the true $\boldsymbol{\beta}$s regardless of scenarios. The sample standard deviation of the 500 point estimates is close to the average of the 500 posterior standard deviations. The 95% coverage probability is close to the nominal value 0.95.

One of the advantages of the proposed method is to provide a smooth estimate of the baseline mean function. Figures 3.1 and 3.2 present the estimates of the baseline functions with the pointwise 95% credible intervals when the true $\boldsymbol{\beta}$ is equal to $(-1, 1)$ for scenario 1 and scenario 2, respectively. It is clear that the estimates of the baseline mean functions fit the true baseline functions very well. They almost overlap through the whole observation time period except for the end of the time period. This is reasonable, because there are rare observations at the end of time period.

3.5.3 Model Misspecification

In this subsection, we compare frailty model (3.5) with no frailty model (2.1), and investigate how the model misspecification affects the final estimation results. Table (3.3) shows two types of misspecifications. Misspesification 1 stands for using model with frailty (3.5) to analyze the no frailty data. Misspecification 2 stands for using model without frailty (2.1) to analyze the frailty data.

Table 3.3: Types of Misspecification

Data / **Model**	**Frailty data** (Simulated in Chapter 3)	**No frailty data** (Simulated in Chapter 2)
with frailty (Model 3.5)	corrcet	misspecification 1
without frailty (Model 2.1)	misspecification 2	correct

For misspecification 1, Table 3.4 and Table 3.5 present the estimation results of the regression parameters for scenario 1 and scenario 2, respectively. Compared with the results by fitting the true model, the point estimates are still unbiased. SSEs almost do not change, while ESDs increase a little bit, which results in high coverage probabilities. ESDs' slight increase is reasonable as frailty effects introduce additional variation in the model fitting.

We present the DIC differences between fitting no frailty model (2.1) and fitting frailty model (3.5) for no frailty data for scenarios 1 and 2 in Figures 3.3 and 3.4. For the scenario 1, based on the top panel boxplots of Figure 3.3, we can see that the ranges of all the DIC differences for 6 different $\boldsymbol{\beta}$ combinations is (-100, 20). For scenario 2, based on the top panel boxplots of Figure 3.4, the range of the DIC

difference is (-130, 50), which is a little bit wider than for scenario 1. However, these ranges are not big while comparing with the DICs themselves (more than 1000) for each data set. Therefore, we conclude that introducing frailty term into the model for no frailty effect still provides robust estimation results.

Table 3.4: Estimates of regression parameters for scenario 1, when fitting the model with frailty effect for no frailty data

		Results on β_1				Results on β_2				
β_1	β_2	Bias	SSE	ESD	CP95	Bias	SSE	ESD	CP95	$1/\phi$
0	0	-0.0042	0.0801	0.1088	0.998	-0.0012	0.0819	0.1102	0.988	0.21
0	1	-0.0001	0.0786	0.1087	0.994	0.0083	0.0830	0.1129	0.990	0.21
1	0	-0.0006	0.0621	0.0994	1.000	0.0022	0.0606	0.0982	1.000	0.21
1	1	0.0013	0.0648	0.0993	0.998	0.0076	0.0631	0.1012	1.000	0.21
-1	0	-0.0062	0.1009	0.1302	0.988	-0.0037	0.0983	0.1263	0.990	0.22
-1	1	-0.0046	0.0947	0.1291	0.984	0.002	0.0932	0.1283	0.992	0.22

For misspecification 2, Table 3.6 and Table 3.7 present the estimation results of the regression parameters for scenario 1 and scenario 2, respectively. Compared with the results by fitting the true model, the point estimates are still unbiased. SSEs almost do not change. However, ESDs are significantly underestimated, which results in low 95% coverage probabilities.

The DIC differences between fitting frailty model (3.5) and fitting no frailty model (2.1) for frailty data for scenarios 1 and 2 are presented at the bottom panels of Figures 3.3 and 3.4. The big negative differences clearly show the true model with

Table 3.5: Estimates of regression parameters for scenario 2, when fitting the model with frailty effect for no frailty data

		Results on β_1				Results on β_2				
β_1	β_2	Bias	SSE	ESD	CP95	Bias	SSE	ESD	CP95	$1/\phi$
0	0	-0.0094	0.0707	0.1146	0.984	0.0038	0.0721	0.1116	0.998	0.21
0	1	-0.0018	0.0624	0.0981	0.998	-0.0054	0.0893	0.1007	0.994	0.21
1	0	0.001	0.052	0.0907	0.998	-0.0050	0.0486	0.0904	1.000	0.21
1	1	-0.0016	0.0687	0.0916	0.998	-0.0004	0.0677	0.0929	0.998	0.21
-1	0	-0.0099	0.0827	0.1148	0.990	0.0036	0.0731	0.1116	1.000	0.21
-1	1	-0.0084	0.0811	0.1142	0.984	0.0055	0.083	0.1149	0.986	0.21

frailty effects performs significantly better than the misspecificated model. Therefore, we conclude that the misspecification 2 may cause a severe problem. Hence, to summarize, for real data analysis practice, we recommend the model with frailty effects. After all, the model (2.1) with no frailty effect is naturally nested in the frailty model (3.5).

3.6 Real Data Analysis

In this section, the same bladder tumor data described in Chapter 2 are analyzed by using the proportional mean model with frailty effect (3.5). The 116 patients were randomly assigned into three treatment groups. 31 of them were given pyridoxine pills, 38 of them were instilled with thiotepa, and 47 were in placebo group. The total follow-up times varied from one week to sixty-four weeks. The covariates are

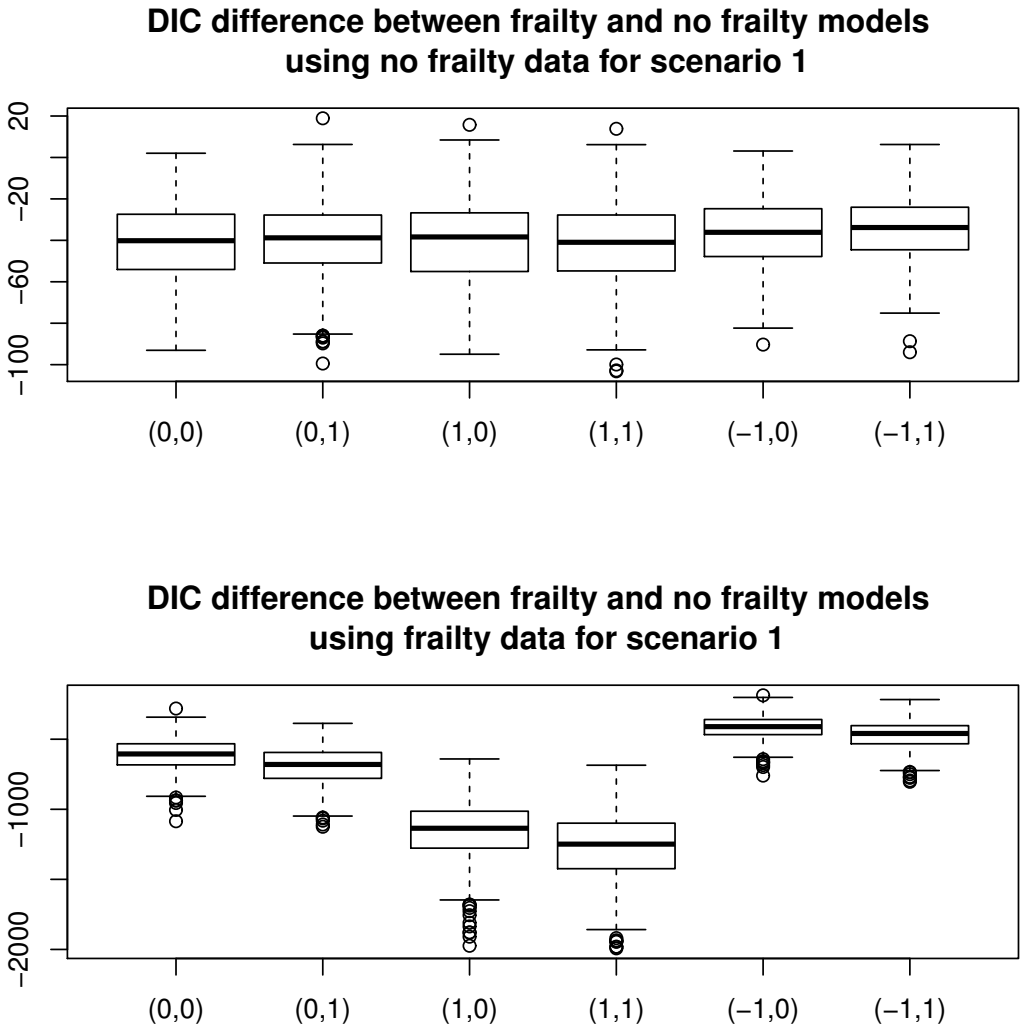

Figure 3.3: Boxplots of DIC differences between frailty and no frailty models for scenario 1

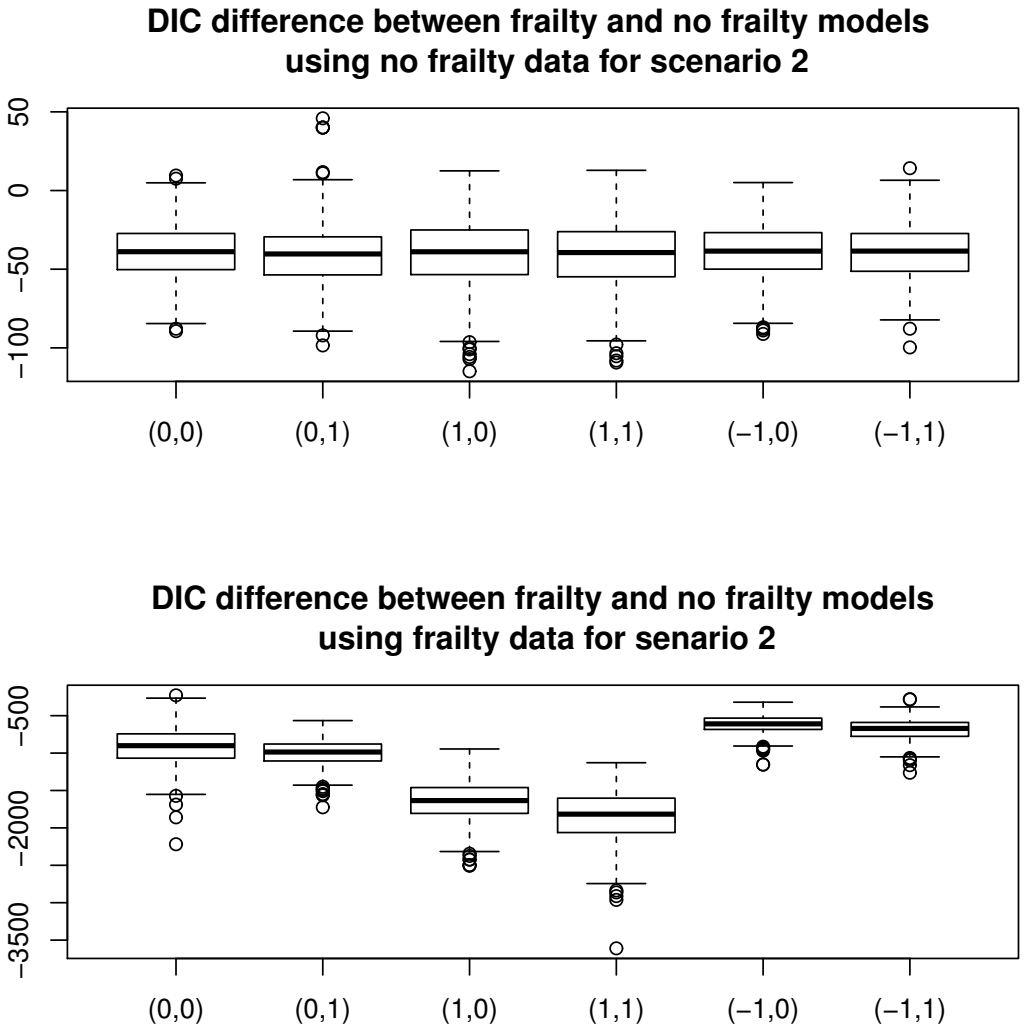

Figure 3.4: Boxplots of DIC differences between frailty and no frailty models for scenario 2

Table 3.6: Estimates of regression parameters for scenario 1, when fitting the model with no frailty effect for frailty data

		Results on β_1				Results on β_2			
β_1	β_2	Bias	SSE	ESD	CP95	Bias	SSE	ESD	CP95
0	0	0.0012	0.2021	0.075	0.524	-0.0092	0.1823	0.0749	0.596
0	1	-0.0123	0.2142	0.0712	0.510	-0.0201	0.2238	0.0718	0.464
1	0	0.0047	0.1825	0.0617	0.494	0.0171	0.1998	0.0549	0.406
1	1	-0.0014	0.1993	0.0584	0.430	0.0001	0.2452	0.0526	0.330
-1	0	-0.0185	0.2082	0.1024	0.676	-0.0201	0.2125	0.0910	0.586
-1	1	0.0039	0.2289	0.0972	0.566	-0.0301	0.2485	0.0867	0.484

denoted as $\mathbf{X}_i = (X_{i1}, X_{i2}, X_{i3}, X_{i4})$, where X_{i1} and X_{i2} represent the number and the size of the tumors at the entry of the study, X_{i3} and X_{i4} are the indicators of the treatments of pyridoxine pills and thiotepa instillation, respectively. Two purposes are for the application. The first one is to determine which covariates have significant effects on the recurrence of the bladder tumors, and which treatment can reduce the recurrence of bladder tumors. The second one is to compare the results with those no frailty model (2.1) in Chapter 2.

In order to obtain the best model, we consider different I-spline specifications with different combinations of degree and the number of knots. We choose the combination with the number of knots from 10 to 30 and degree 2 or 3. The equally spaced knots are taken. Table 3.8 presents the estimated posterior mean and standard deviation for the regression coefficients with DICs. The estimates of $\boldsymbol{\beta}$s and standard deviation and

Table 3.7: Estimates of regression parameters for scenario 2, when fitting the model with no frailty effect for frailty data

		Results on β_1				Results on β_2			
β_1	β_2	Bias	SSE	ESD	CP95	Bias	SSE	ESD	CP95
0	0	-0.0154	0.1951	0.0857	0.532	0.004	0.2071	0.0762	0.510
0	1	0.003	0.1958	0.0590	0.464	-0.0034	0.2212	0.0597	0.392
1	0	0.0053	0.1845	0.0515	0.416	-0.006	0.1882	0.0459	0.364
1	1	0.0098	0.2003	0.0489	0.366	-0.0205	0.2326	0.0437	0.278
-1	0	-0.0156	0.1950	0.0857	0.628	0.0037	0.2069	0.0762	0.518
-1	1	-0.0061	0.2106	0.0806	0.556	0.0017	0.2398	0.0726	0.398

DICs from some representative combinations. The estimates of covariate coefficients and their standard deviations are very similar for different selections of degree and knots, which demonstrates that the point estimates are robust to the choice of knots and degree in the application of real data. DICs for the different knots and degree suggest that there exists difference for these combinations. The best combination is the one with degree 3 with 25 knots, since it has the smallest DIC 97.2177. Therefore, our summarized results are based on degree 3 and 25 knots.

The estimate of baseline mean function is shown in Figure 3.5. It is a smooth and nondecreasing curve. Figure 3.6 shows the estimated mean function for each treatment. It is clear that the treatment of thiotepa is a much better treatment on reducing the recurrence of bladder tumors comparing with the treatment of pyridoxine pills and placebo. Based on the estimates with standard deviations in Table 3.8, we

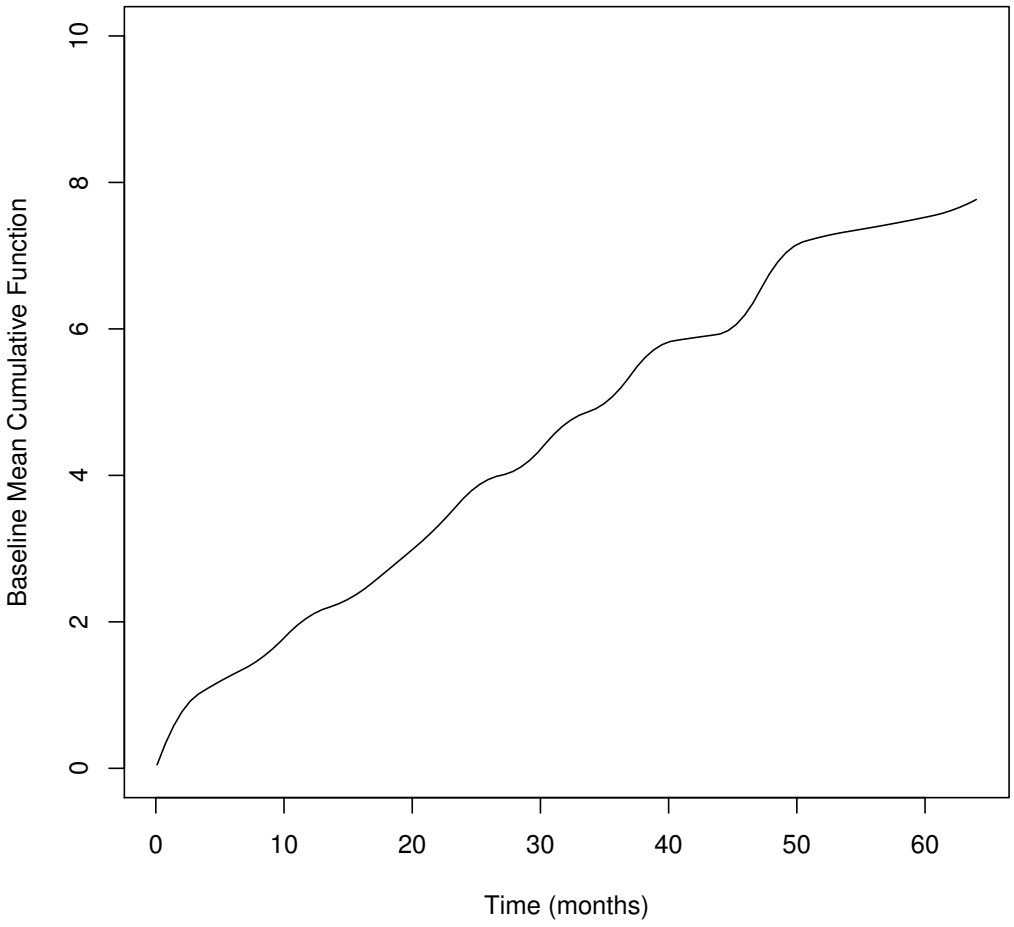

Figure 3.5: The estimator of baseline mean function for bladder tumor example under the proportional mean model with frailty

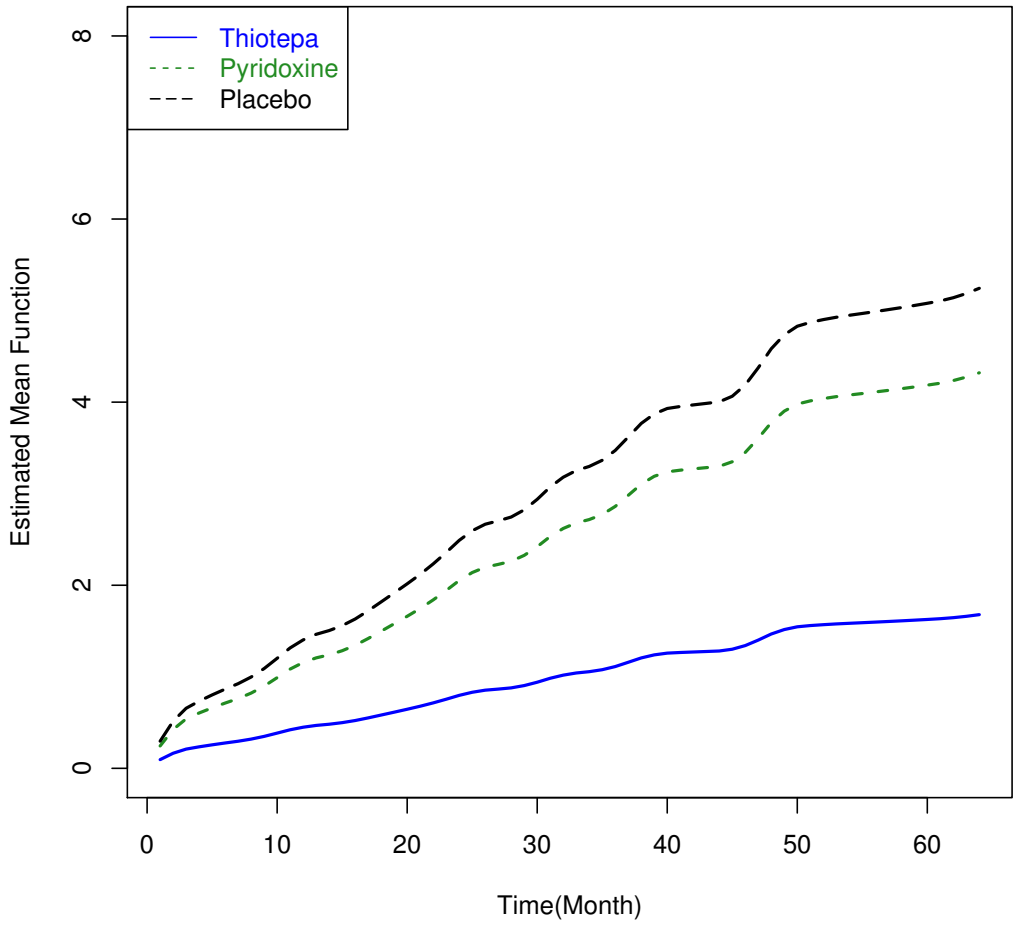

Figure 3.6: The estimated mean functions for different treatments for bladder tumor example under the proportional mean model with frailty

Table 3.8: Results about regression parameters for different knots with degree=2 and 3

d	Kn	β_1		β_2		β_3		β_4		DIC	$1/\phi$
		est.	sd	est.	sd	est.	sd	est.	sd		
2	15	0.3303	0.1163	-0.0067	0 .1252	-0.1384	0.4203	-1.2176	0.4513	199.857	3.03
	20	0.3545	0.1075	-0.0443	0.1061	-0.1766	0.4215	-1.2340	0.4482	97.218	3.03
	25	0.2978	0.1106	-0.0607	0.1250	-0.1326	0.5167	-1.2295	0.4369	256.255	3.03
3	15	0.3174	0.1024	-0.0508	0.01162	-0.1498	0.4766	-1.2917	0.4738	292.093	3.03
	20	0.3321	0.1023	-0.0547	0.1135	-0.09832	0.5146	-1.2709	0.4702	205.388	3.12
	25	0.3096	0.1086	0.0855	0.1116	-0.1310	0.4389	-1.2258	0.4482	273.359	3.03

can draw conclusion that the number of tumors at the beginning of the study and with the treatment of thiotepa are significant. The number of the tumors at the entrance of the study is positively related to the recurrence of the bladder tumor. Thiotepa instillation much more effectively reduces the number of recurrent tumors than pyridoxine pill. The individuals with one more tumor at entering the study tend to have more recurrent tumors by increasing around 42%. On average, the number of recurrent tumors in patients with thiotepa instillation treatment group is about 29% of that with placebo treatment group for the similar patients with the same tumor number and size at the entry of the study. The size of tumors and pyridoxine pills are not significantly related to the number of recurrent tumors at follow-up visits.

Compared with the results in real data analysis of Chapter 2, the effect of the number of the tumors at the study entrance and the treatment of thiotepa are more significant when accounting for the correlation between cumulative counts using the frailty variable. There does exist frailty in the bladder tumor data since the variance

$1/\phi$ as large as round 3. We conclude that frailty model should be better than the no frailty model on analyzing bladder tumor data.

Table 3.9: The estimates of covariate coefficients for Bladder tumor data from different methods. WZ method: Wellner and Zhang (2007). YWH method: Yao, Wang and He (2016)

Method	Proposed method		YWH method		WZ method	
Variable	$\hat{\beta}$	$\hat{sd}(\hat{\beta})$	$\hat{\beta}$	$\hat{sd}(\hat{\beta})$	$\hat{\beta}$	$\hat{sd}(\hat{\beta})$
X1	0.3545	0.1075	0.336	0.106	0.2069	0.078
X2	-0.0443	0.1061	0.012	0.120	-0.0355	0.6801
X3	-0.1766	0.4215	-0.033	0.409	0.664	0.4310
X4	-1.2340	0.4482	-1.140	0.435	-0.7972	0.3603

Table 3.9 lists the estimates of covariate coefficients under the frailty model along with results from other two published papers for the same data set. All the results based on our proposed method are consistent with the results in Wellner and Zhang (2007) [15], Lu, Zhang and Huang (2009) [17], and Yao, Wang and He (2016) [63].

3.7 Summary

In this chapter, a proportional mean model with gamma frailty effect is discussed. Following Chapter 2, a semiparametric Bayesian approach is developed to analyze panel count data under the proportional mean model with frailty effect. Monotone I-splines are used to approximate the baseline mean function. Poisson latent variables are introduced to decompose the counts at each time interval, so that it simplifies the complicated sampling. On one hand, the simulation study shows that the proposed

approach performs well for analyzing the panel count data with the frailty effect. The estimates of the regression coefficients are unbiased with the standard error close to the posterior standard deviation. The estimate of the baseline mean cumulative function is smooth and close to the true baseline function regardless of scenarios. Moreover, the estimation procedure doesn't rely on the frailty distribution, and hence is more robust against departure from the true frailty distribution. On the other hand, the simulation study shows that fitting model with no frailty effect for frailty data may cause significant worse results, while introducing frailty term into the model for no frailty data still provides robust estimation results. Therefore, in practice, we recommend the model with frailty effect for real data analysis.

For the bladder tumor data, the frailty model overperforms the no frailty model with much smaller DICs. The analyzed results are consistent with the published results in Wellner and Zhang (2007) [15], Lu, Zhang and Huang (2009) [17], and Yao et al. (2016) [63].

Chapter 4

Robustness Study of our Proposed Bayesian Approaches

4.1 Introduction

In this chapter, we investigate the robustness of our proposed Bayesian approaches in Chapter 2, and specifically, the following situations are investigated: 1. The sample size is small; 2. The underlying event count follows a nonhomogeneous negative binomial process instead of a nonhomegeneous Poisson process; 3. The observation times depend on covariate variables; 4. The event count process is an informative observation time process.

4.2 Small Size Situation

We first investigate whether our approaches provided accurate estimates when the sample size is small. We generate data as the way in Chapter 2 for two scenarios except that the sample size is 50 instead of 150. For the fitting in both scenarios, 20 equally-spaced knots are chosen between 0 and the maximum of observation times from each data set, and for the choice of the degree, degree 3 is chosen in order to get adequate smoothness although there is no big difference between choosing degrees 2 and 3 according to Chapter 2 and Cai et al. (2011) [56] and Wang and Dunson (2011) [53].

The simulation results are shown in tables 4.1 and 4.2 for scenario 1 and scenario 2, respectively. We can see that all the results are very close to the ones in Chapter

Table 4.1: Estimation of regression parameters for **scenario 1** based on 500 simulated data sets with **sample size 50** from the proposed Bayesian method in Chapter 2. Bias refers to the difference between the average of the 500 point estimates and the true value, ESD refers to the average of the 500 posterior standard deviations, SSE refers to the sample standard deviation of the 500 point estimates, and CP95 is the 95% coverage probability

Scenario 1: $U_0(t) = t + \log(1+t)$										
			Results on β_1				Results on β_2			
β_1	β_2	Methods	Bias	SSE	ESD	CP95	Bias	SSE	ESD	CP95
0	0	Proposed	-0.0022	0.1402	0.1321	0.924	-0.008	0.1367	0.133	0.936
0	1	Proposed	-0.0034	0.1336	0.1262	0.930	0.0012	0.136	0.129	0.934
1	0	Proposed	0.0019	0.107	0.1078	0.952	0.0013	0.0995	0.0991	0.948
1	1	Proposed	-0.0038	0.1003	0.1029	0.948	-0.0078	0.1002	0.097	0.960
-1	0	Proposed	-0.0142	0.1999	0.1809	0.920	-0.0049	0.1621	0.1616	0.950
-1	1	Proposed	0	0.1872	0.1713	0.942	-0.0015	0.1579	0.1566	0.944

2 regardless of scenarios. The only difference is that all the estimates are slightly bigger than the ones with sample size 150. For instance, for "$\beta_1 = 0$" in scenario 1, the bias changes from -0.0002 to -0.0022, SSE from 0.0800 to 0.1402, ESD from 0.0740 to 0.1321; for "$\beta_2 = 0$", the bias changes from 0.0020 to 0.008, SSE from 0.008 to 0.1367, ESD from 0.0750 to 0.133. ESDs, the average of the 500 posterior standard deviations, are still close to SSEs, the sample standard deviation of the 500 point estimates. Therefore, the 95% coverage probabilities are still close to 0.95. Figures 4.1 and 4.2 present the estimates of the cumulative baseline functions with the pointwise 95% credible intervals and the true cumulative baseline function for scenario 1 and

scenario 2, respectively. We can see that the true and estimate values are almost overlapped, and that the estimates of the cumulative mean baseline function well characterize the true baseline mean function. To summarize, our proposed method in Chapter 2 still performs very well when the sample size is relatively small.

Table 4.2: Estimation of regression parameters for **scenario 2** based on 500 simulated data sets with **sample size 50** from the proposed Bayesian method in Chapter 2. Bias refers to the difference between the average of the 500 point estimates and the true value, ESD refers to the average of the 500 posterior standard deviations, SSE refers to the sample standard deviation of the 500 point estimates, and CP95 is the 95% coverage probability

Scenario 2: $U_0(t) = 3(0.5 \text{ pnorm}(t, 1, 1) + 0.5\text{pnorm}(t, 5, 0.5) + t^{0.5})$

			Results on β_1				Results on β_2			
β_1	β_2	Methods	Bias	SSE	ESD	CP95	Bias	SSE	ESD	CP95
0	0	proposed	-0.0032	0.1128	0.1097	0.938	-0.0023	0.1153	0.1115	0.940
0	1	proposed	-0.0011	0.1066	0.1044	0.948	0.0025	0.1073	0.1080	0.960
1	0	proposed	-0.0033	0.0917	0.0899	0.954	0.0008	0.0848	0.0811	0.944
1	1	proposed	0.0029	0.0811	0.0861	0.958	0.0004	0.0872	0.0001	0.936
-1	0	proposed	-0.0142	0.1496	0.1498	0.952	-0.0039	0.1475	0.1361	0.944
-1	1	proposed	-0.0146	0.1373	0.1424	0.958	0.0129	0.1317	0.1315	0.954

4.3 Relaxing Nonhomegenous Poisson Process Assumption

4.3.1 Introduction

In Chapter 2 we consider the proportional mean model, where the number of events are assumed to follow a nonhomogeneous Poisson process, and the observation

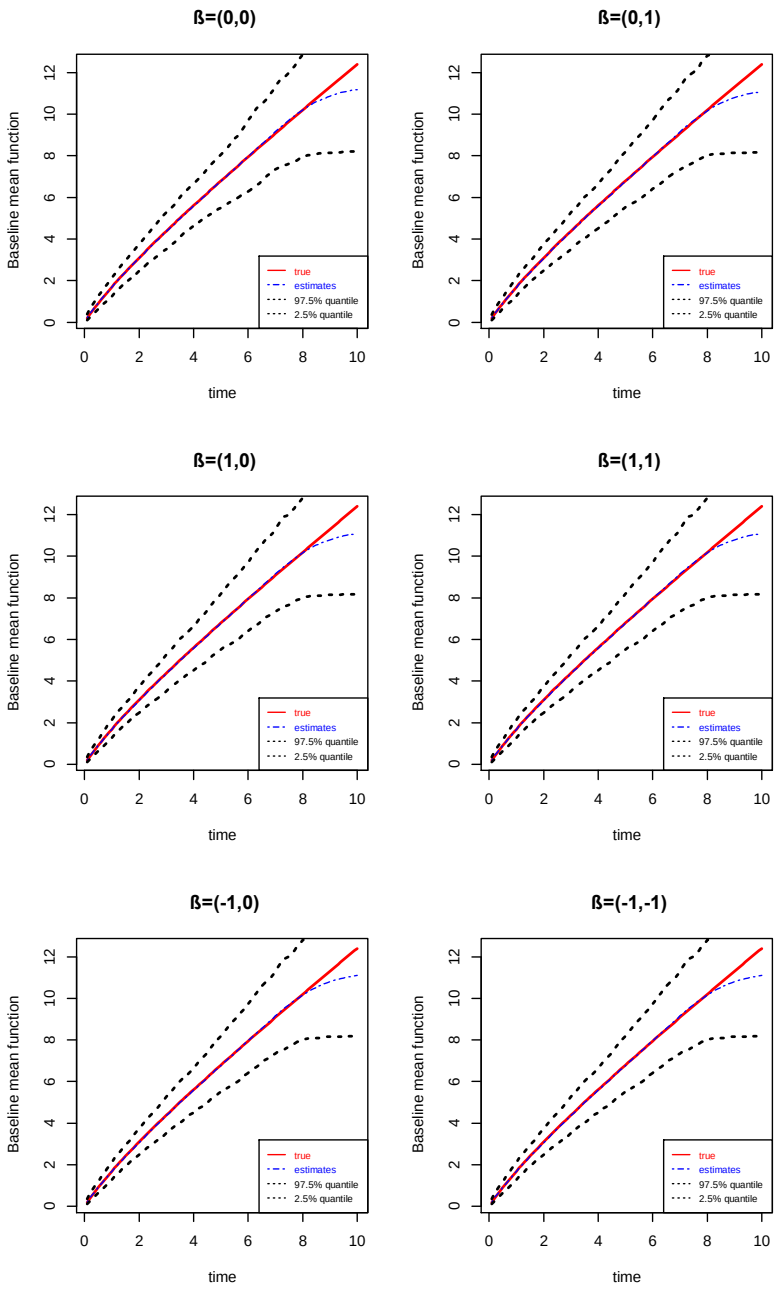

Figure 4.1: Estimates of baseline mean functions for different $\boldsymbol{\beta}$ in **scenario 1** for sample size 50

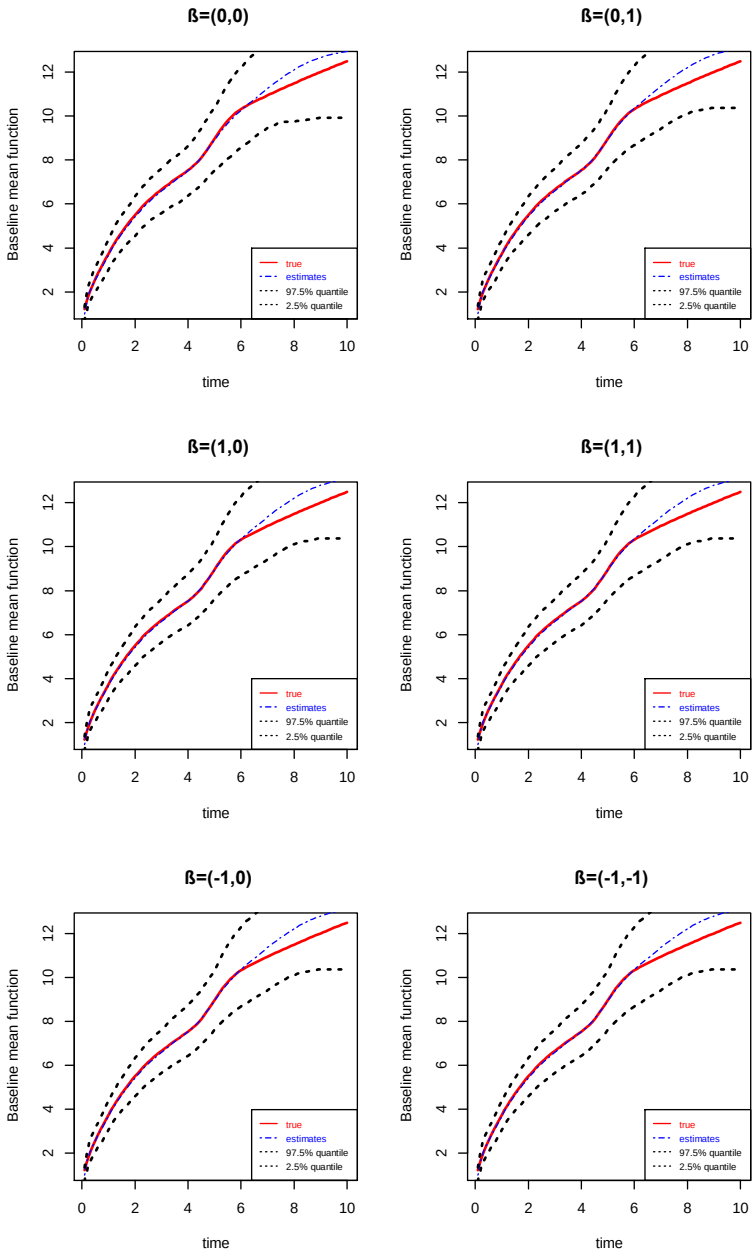

Figure 4.2: Estimates of baseline mean functions for different $\boldsymbol{\beta}$ in **scenario 2** for sample size 50

times are allowed to be different for different subjects. This model with various extensions are reviewed in Chapter 1. One important topic is to relax the Poisson process assumption. Wellner and Zhang (2007) [15] explored both semiparametric maximum pseudolikelihood and likelihood estimators under the proportional mean model where the underlying counting process was assumed as a nonhomogeneous Poisson process conditional on covariates. They showed that the estimation procedures were robust against possible misspecification of the underlying counting process, as long as the proportional mean model held. It is not hard to get the pseudolikelihood estimator, but it is fairly inefficient when there exists a heavy tail on the distribution of the number K of observation times, which was illustrated in an example given by Wellner et al. (2004) [49]. Ramaswamy (1994) [100] presented a disaggregate negative binomial regression procedure for analysis of count data observed from a heterogeneous sample of cross-sections, where over some fixed time periods the variance was much greater than the mean. In this situation, overdispersion occured. Overdispersion is a very common feature in applied data analysis because populations are frequently heterogeneous in practice. Cameron and Trivedi (1990) [101] discussed some regression-based tests for overdispersion in the Poisson regression model. These tests could be used to test the appropriateness of the Poisson regression model. Cox (1983) [102] showed that overdispersion had little effect on parameter estimates but led to the underestimation of standard errors in the context of standard Poisson regression. This underestimation may be corrected by the use of quasi-likelihood methods as Breslow (1984) [103] suggested.

If the Poisson regression model isn't deemed appropriate for the count data, alternative stochastic models may be explored to accommodate overdispersion (Cameron and Trivedi 1986 [104], Cameron et al. 1990 [101]). The negative binomial regression model is perhaps the most popular model to be used, in part due to its simplicity (Gardner et al. 1995 [105]).

Our goal in this subsection is to investigate if our approach is feasible for the data generated from the negative binomial process through simulations.

4.3.2 Data Generation

We generate data from a negative binomial counting process with the same mean function as in Chapter 2. For each subject i, the data are collected at K_i random time points $0 < t_{i1} < t_{i2} < \ldots < t_{iK_i}$, where K_i is a random variable that takes positive integer values, and t_{iK_i} is the last observation time of subject i. Denote $N_i(t)$ as the number of events observed before or at time t for subject i, then $N_i(t)$ is a counting process for subject i. So, the panel counts from the counting process $N_i(t)$ are $N_i = (N_i(t_{i1}), N_i(t_{i2}), \ldots, N_i(t_{iK_i}))$.

In general, a negative binomial random variable Y defined as the number of failures before r (can be a real number) successes, has the density

$$\frac{\Gamma(y+r)}{\Gamma(r)y!}p^r(1-p)^y,$$

where p is the probability of success. The density can be represented as a Poison mixture. That is, it arises as a mixture of Poisson distributions with mean distributed as a gamma distribution with scale parameter $(1-p)/p$ and shape parameter size r.

For the data generation, we follow all settings as described in Chapter 2 in order to make the results comparable except that the event count between two adjacent observation times now follows a negative binomial distribution with parameter $r = \Big\{U_0(t_{ij}) - U_0(t_{ij-1})\Big\}\exp(X_i'\boldsymbol{\beta})$ and $p = 0.5$ instead of following a Poisson distribution. That is,

$$Z_{ij}|X_i = N_i(t_{ij}) - N_i(t_{i,j-1})|X_i \sim \mathcal{NB}\bigg(\Big\{U_0(t_{ij}) - U_0(t_{ij-1})\Big\}\exp(X_i'\boldsymbol{\beta}), 0.5)\bigg). \quad (4.1)$$

In this way, the mean of the negative binomial counting process is still the same as $U_0(t)\exp(X_i'\boldsymbol{\beta})$.

Table 4.3: Estimation of regression parameters for **scenario 1** based on 500 simulated data sets **for the negative binomial event process.** Bias refers to the difference between the average of the 500 point estimates and the true value, ESD refers to the average of the 500 posterior standard deviations, SSE refers to the sample standard deviation of the 500 point estimates, and CP95 is the 95% coverage probability

Scenario 1: $U_0(t) = t + \log(1+t)$										
			Results on β_1				Results on β_2			
β_1	β_2	Methods	Bias	SSE	ESD	CP95	Bias	SSE	ESD	CP95
0	0	Proposed	0.0034	0.1035	0.0746	0.846	-0.0062	0.1120	0.0747	0.794
0	1	Proposed	-0.0066	0.1009	0.0706	0.826	0.0042	0.1011	0.0715	0.838
1	0	Proposed	0.0060	0.0942	0.0616	0.802	0.0007	0.0786	0.0552	0.812
1	1	Proposed	-0.0035	0.083	0.0580	0.826	-0.0039	0.0753	0.0529	0.840
-1	0	Proposed	-0.0013	0.1431	0.1018	0.842	0.0058	0.1239	0.0910	0.836
-1	1	Proposed	-0.0230	0.1420	0.0964	0.780	-0.0026	0.1209	0.0864	0.834

We also have the same two scenarios:

Scenario 1, $U_0(t) = t + log(1+t)$, which is approximately linear; and

Scenario 2, $U_0(t) = 0.5\Phi(t,1,1) + 0.5\Phi(t,5,0.5) + t^{0.5}$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. It is clear that for this scenario, $U_0(t)$ is curvilinear and has multiple reflection points.

There are totally 12 simulation setups. For each setup, 500 data sets are simulated and each data set has 150 subjects. The proposed Gibbs sampler in Section 5 of Chapter 2 is implemented for each simulated data set. All the summarized results are based on 5000 iterations of the Gibbs samples after discarding the first 1000

Table 4.4: Estimation of regression parameters for **scenario 2** based on 500 simulated data sets **for the negative binomial event process.** Bias refers to the difference between the average of the 500 point estimates and the true value, ESD refers to the average of the 500 posterior standard deviations, SSE refers to the sample standard deviation of the 500 point estimates, and CP95 is the 95% coverage probability

Scenario 2: $U_0(t) = 3(0.5\ \text{pnorm}(t,1,1) + 0.5\text{pnorm}(t,5,0.5) + t^{0.5})$										
			Results on β_1				Results on β_2			
β_1	β_2	Methods	Bias	SSE	ESD	CP95	Bias	SSE	ESD	CP95
0	0	proposed	-0.0060	0.0863	0.0622	0.846	0.0040	0.0835	0.0626	0.862
0	1	proposed	0.0021	0.0816	0.0584	0.834	0.0027	0.0883	0.0589	0.806
1	0	proposed	0.0022	0.0739	0.0512	0.838	-0.0015	0.0637	0.0458	0.828
0	1	proposed	0.0022	0.0739	0.0512	0.838	-0.0015	0.0637	0.0458	0.828
-1	0	proposed	-0.0210	0.1222	0.0849	0.826	-0.0078	0.1142	0.0752	0.796
-1	1	proposed	-0.0051	0.1090	0.0799	0.848	0.0021	0.1006	0.0720	0.844

iterations as a burn-in. Additionally, in both scenarios, 20 equally-spaced knots are chosen between 0 and the maximum of observation times from each data set, for the degree choice, degree 3 is chosen in order to get adequate smoothness although there is no big difference between choosing degrees 2 and 3 according to Chapter 2.

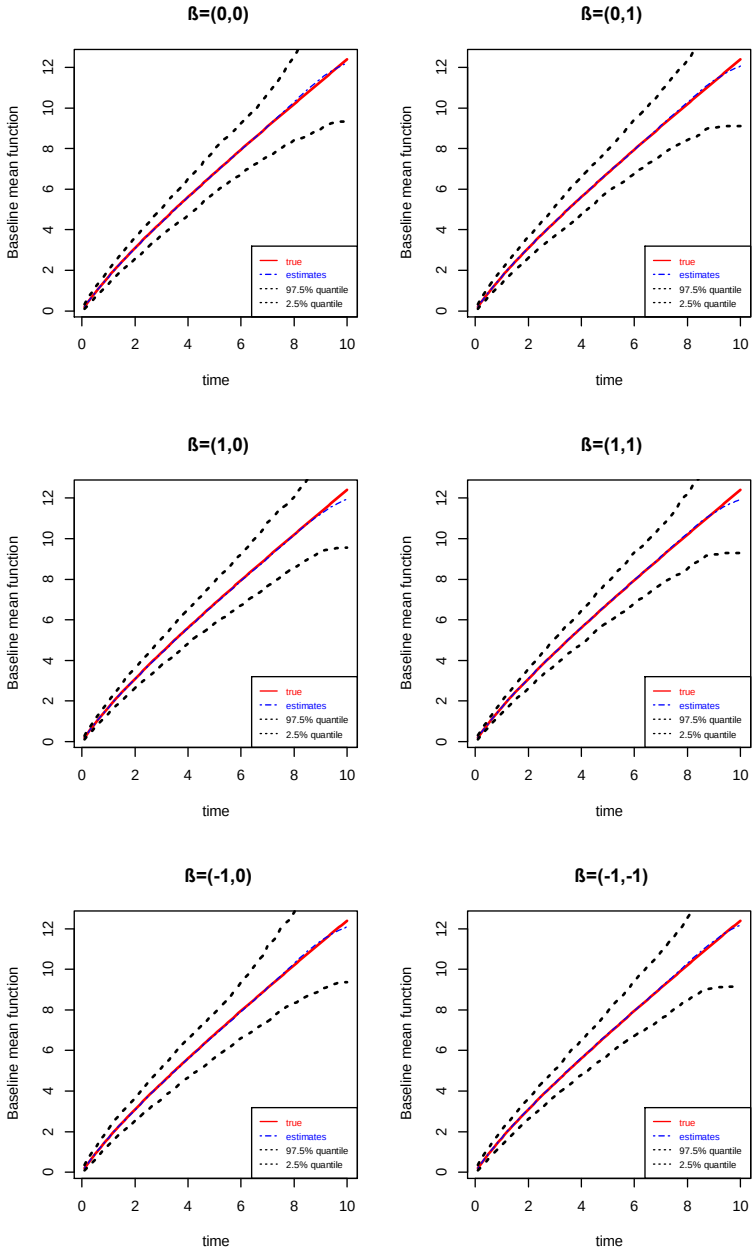

Figure 4.3: Estimates of baseline mean functions for different $\boldsymbol{\beta}$ in **scenario 1 for the negative binomial event process**

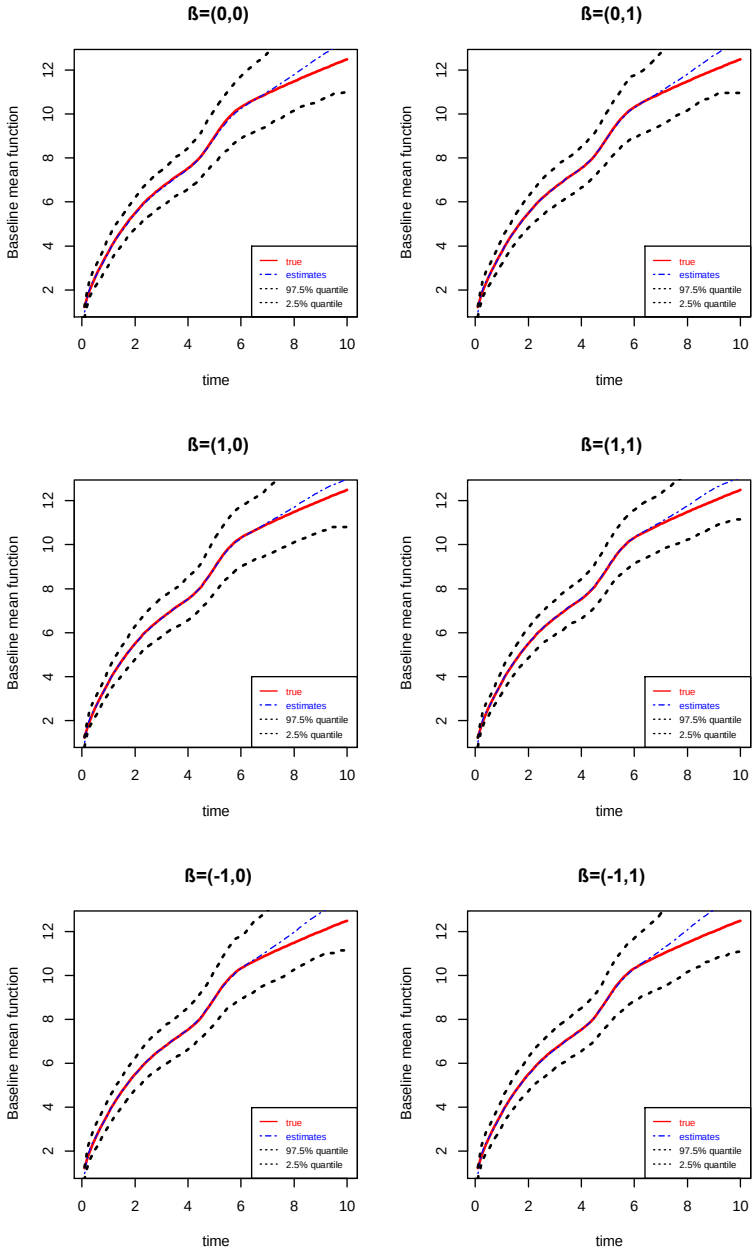

Figure 4.4: Estimates of baseline mean functions for different $\boldsymbol{\beta}$ in **scenario 2 for the negative binomial event process**

4.3.3 Simulation Results

Based on the selection of knots and degree for each scenario, we implement the proposed Gibbs sampler in Section 2.5. The trace plots of Gibbs samples suggest that there is a good mixing and fast convergence in MCMC chains of the regression parameters and the spline coefficients regardless of scenarios. Tables 4.3 and 4.4 present the estimates of the regression parameters for both scenarios, with Bias denoting the difference between the average of 500 point estimates and the true value, SSE as the sample standard deviation of the 500 point estimates, ESD as the average of the 500 posterior standard deviations, and CP95 as the 95% coverage probability. The point estimates are taken to be posterior means, and the coverage probabilities are based on the 95% credible intervals constructed by 2.5th and 97.5th percentiles.

Most of the estimates perform very well regardless of scenarios. There is almost no bias shown, and SSEs and ESDs are fairly small. However, coverage probabilities are constantly smaller than the nomial level 0.95, with most of them close to 0.85. This happens due to ESDs being smaller than SSEs. The estimation of the baseline mean function and their corresponding pointwise credible intervals for all the configurations are presented in Figures 4.3 and 4.4. It is clear that the estimation of the baseline mean function is very good as the average of the estimates overlaps the true curve except for the right tail. The departure of the estimates from the true curve at the right tail area is reasonable because wherein there are rare observations at the end of time period. Overall, the estimating approach that assumes the Poisson event process is quite robust when the event process follows the negative binomial process. The only problem we find is that the posterior standard deviations of the regression parameters are slightly underestimated. To solve this problem, we try the frailty model approach to fit the data.

4.3.4 Correcting Underestimated Posterior Standard Deviations

In Chapter 3, we observed that the frailty model approach in general inflated the posterior standard deviations when fitting no frailty data. In this subsection, we try the frailty model approach to remedy the underestimation of the posterior standard deviations.

Table 4.5: Estimation of regression parameters by **using the frailty model approach for scenario 1 based on 500 simulated data sets under the negative binomial event process.** Bias refers to the difference between the average of the 500 point estimates and the true value, ESD refers to the average of the 500 posterior standard deviations, SSE refers to the sample standard deviation of the 500 point estimates, and CP95 is the 95% coverage probability

Scenario 1: $U_0(t) = t + \log(1 + t)$

			Results on β_1				Results on β_2			
β_1	β_2	Methods	Bias	SSE	ESD	CP95	Bias	SSE	ESD	CP95
0	0	proposed	0.0019	0.1063	0.1136	0.956	-0.0054	0.1137	0.1144	0.946
0	1	proposed	-0.0093	0.1066	0.1129	0.956	0.0089	0.1118	0.1170	0.958
1	0	proposed	0.0037	0.0965	0.1027	0.954	0.0011	0.0838	0.0997	0.984
1	1	proposed	-0.0080	0.0897	0.1018	0.964	-0.0022	0.0878	0.1028	0.976
-1	0	proposed	-0.0063	0.1455	0.1400	0.938	0.0050	0.1322	0.1366	0.972
- 1	1	proposed	-0.0264	0.1499	0.1376	0.928	0.0050	0.1338	0.1375	0.964

We apply the algorithm in Section 3.4 to fit the negative binomial process data. We have the results for scenarios 1 and 2 in tables 4.5 and 4.6 and figures 4.5 and 4.6. Comparing SSE with ESD of each covariate coefficients, we can see that this

Table 4.6: Estimation of regression parameters by **using the frailty model approach for scenario 2 based on 500 simulated data sets under the negative binomial event process.** Bias refers to the difference between the average of the 500 point estimates and the true value, ESD refers to the average of the 500 posterior standard deviations, SSE refers to the sample standard deviation of the 500 point estimates, and CP95 is the 95% coverage probability

Scenario 2: $U_0(t) = 3(0.5 \text{ pnorm}(t, 1, 1) + 0.5\text{pnorm}(t, 5, 0.5) + t^{0.5})$

			Results on β_1				Results on β_2			
β_1	β_2	Methods	Bias	SSE	ESD	CP95	Bias	SSE	ESD	CP95
0	0	proposed	-0.0067	0.0877	0.1003	0.979	0.003	0.086	0.1013	0.965
0	1	proposed	-0.0004	0.0867	0.1004	0.978	0.0064	0.0958	0.1036	0.972
1	0	proposed	0.0018	0.0767	0.0915	0.974	-0.001	0.0697	0.0911	0.992
1	1	proposed	-0.0015	0.0720	0.0920	0.986	-0.0015	0.0699	0.0932	0.990
-1	0	proposed	-0.0244	0.1232	0.1199	0.938	-0.0104	0.1167	0.1168	0.955
-1	1	proposed	-0.0072	0.1128	0.1185	0.960	0.0088	0.1110	0.1195	0.969

time the differences are much smaller and close to 0.006 in most of cases. The 95% coverage probabilities are all fairly close to 0.95, although they are slightly lower or higher than 0.95 in some cases.

Figures 4.5 and 4.6 show that there are no problems for fitting the true baseline mean functions except the location at the end of the time period due to the lack of data.

Overall, the frailty model is a better choice for fitting the data simulated from a negative binomial process.

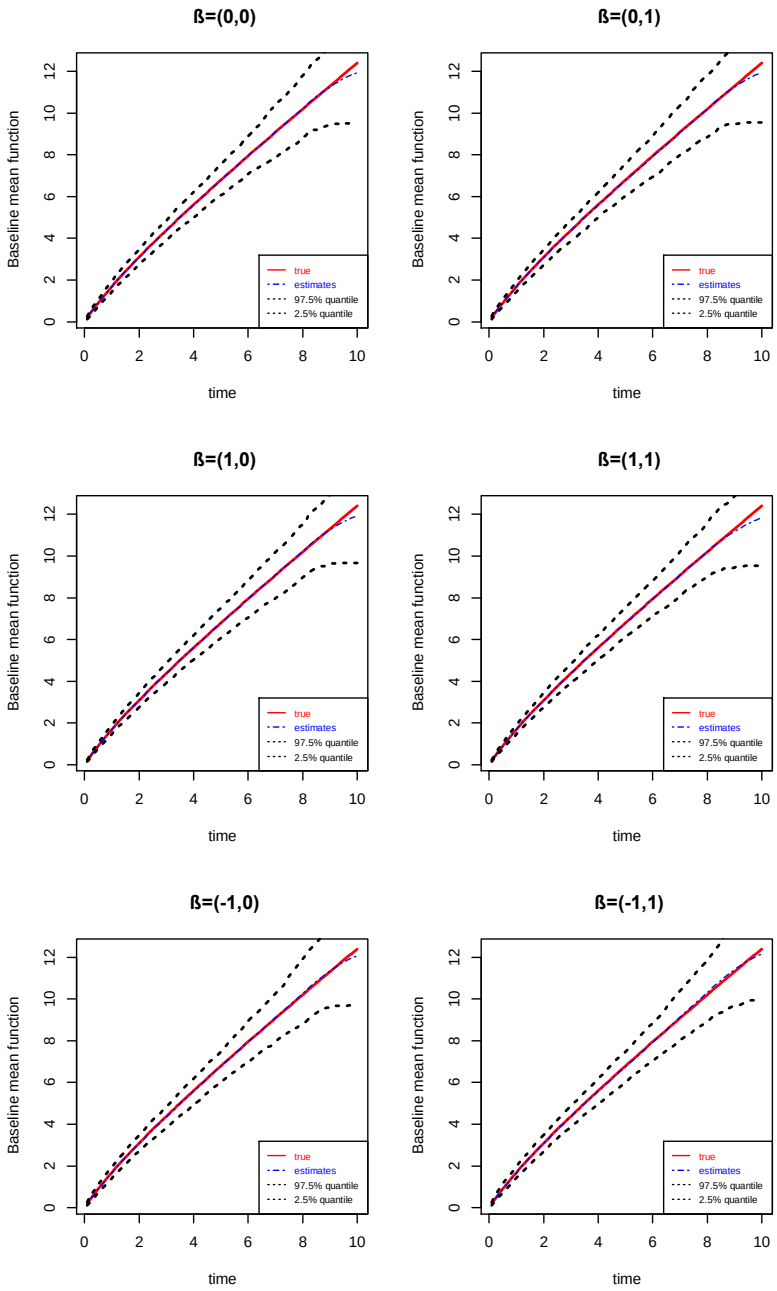

Figure 4.5: Estimates of baseline mean functions for different β in **scenario 1 using the frailty model fitting the data from negative binomial event process**

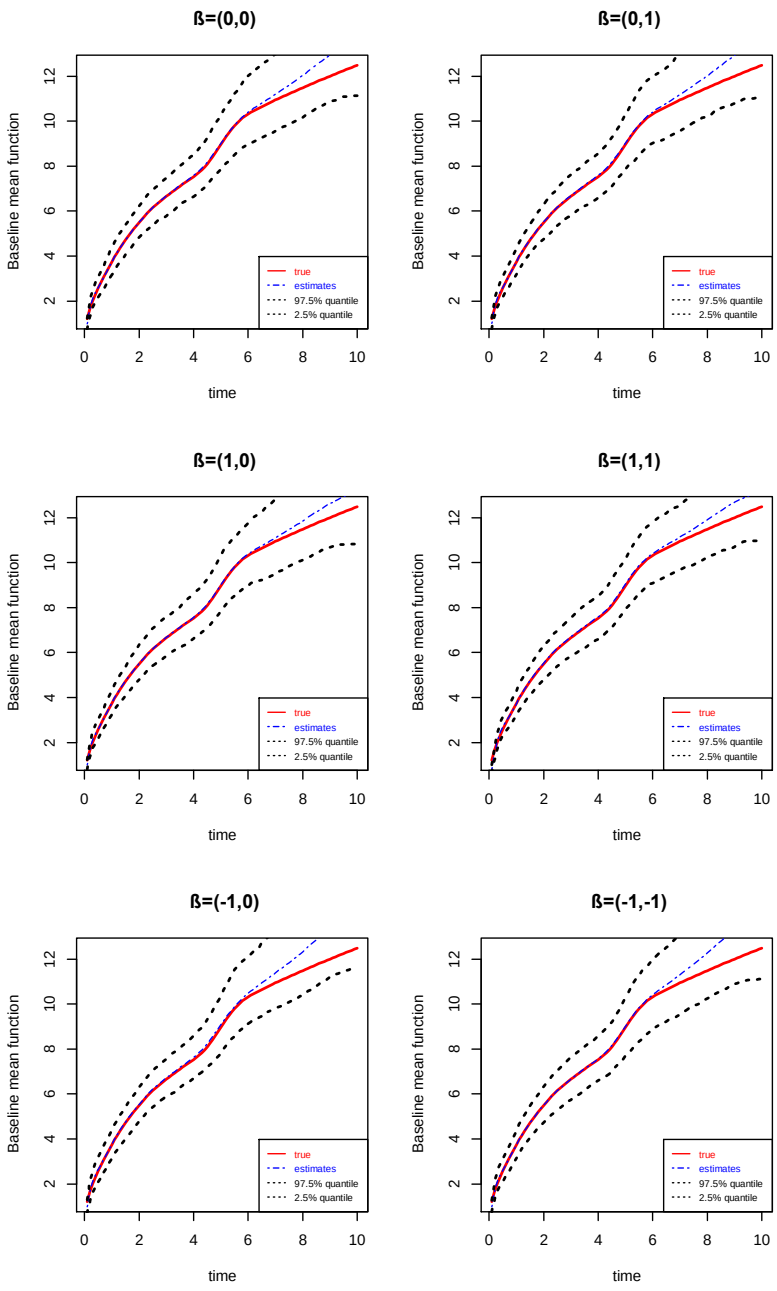

Figure 4.6: Estimates of baseline mean functions for different β in **scenario 2 using the frailty model fitting the data from negative binomial event process**

4.4 Observation Times Dependent on Covariates

In this section, we investigate the validation of the proposed approach in Chapter 2 when the observation times are dependent on the covariates. For example, the length of time between two adjacent observation times t_i follows an exponential distribution with rate $\exp(X_i'\gamma)$. For the simulation, we take $\gamma = c(1, 1)$, simulate data in the same way as described in Chapter 2.

Table 4.7: Estimation of regression parameters for **scenario 1** based on 500 simulated data sets **when the observation times depend on covariates**. Bias refers to the difference between the average of the 500 point estimates and the true value, ESD refers to the average of the 500 posterior standard deviations, SSE refers to the sample standard deviation of the 500 point estimates, and CP95 is the 95% coverage probability

Scenario 1: $U_0(t) = t + \log(1 + t)$

			Results on β_1				Results on β_2			
β_1	β_2	Methods	Bias	SSE	ESD	CP95	Bias	SSE	ESD	CP95
0	0	proposed	0.0026	0.0625	0.0633	0.950	0.0041	0.0623	0.0638	0.966
0	1	proposed	-0.0032	0.0618	0.0595	0.948	-0.0021	0.0630	0.0604	0.952
1	0	proposed	-0.0016	0.0551	0.0523	0.934	0.0017	0.0466	0.0466	0.950
1	1	proposed	0.0003	0.0512	0.0488	0.936	0.0005	0.0427	0.0446	0.968
-1	0	proposed	-0.0054	0.0861	0.0857	0.948	0.0064	0.0801	0.0765	0.932
-1	1	proposed	-0.0069	0.0818	0.0812	0.960	0.0037	0.0758	0.0740	0.942

Table 4.8: Estimation of regression parameters for **scenario 2** based on 500 simulated data sets **when the observation times depend on covariates**. Bias refers to the difference between the average of the 500 point estimates and the true value, ESD refers to the average of the 500 posterior standard deviations, SSE refers to the sample standard deviation of the 500 point estimates, and CP95 is the 95% coverage probability

Scenario 2: $U_0(t) = 3(0.5\ \text{pnorm}(t, 1, 1) + 0.5\text{pnorm}(t, 5, 0.5) + t^{0.5})$										
			Results on β_1				Results on β_2			
β_1	β_2	Methods	Bias	SSE	ESD	CP95	Bias	SSE	ESD	CP95
0	0	proposed	0.0011	0.0575	0.0561	0.938	-0.0010	0.0554	0.0564	0.950
0	1	proposed	0.0020	0.0531	0.0531	0.946	0.0017	0.0559	0.0537	0.952
1	0	proposed	0.0038	0.0458	0.0463	0.954	-0.0018	0.0419	0.0413	0.942
0	1	proposed	-0.0012	0.0440	0.0438	0.952	0.0001	0.0389	0.0399	0.962
-1	0	proposed	-0.0009	0.0760	0.0768	0.952	-0.0012	0.0724	0.0687	0.930
-1	1	proposed	-0.0009	0.0717	0.0718	0.942	0.0017	0.0648	0.0650	0.962

Tables 4.7 and 4.8 show the estimation results of the regression parameter for scenario 1 and scenario 2, respectively. Figures 4.7 and 4.8 present the estimates of the cumulative baseline functions with the pointwise 95% credible intervals and the true cumulative baseline function for scenario 1 and scenario 2, respectively. We can see that all the parameter estimates are very close to the true values. The estimates of the cumulative mean baseline function well characterize the true baseline mean function, having a very close overlapping each other. To summarize, our proposed method in

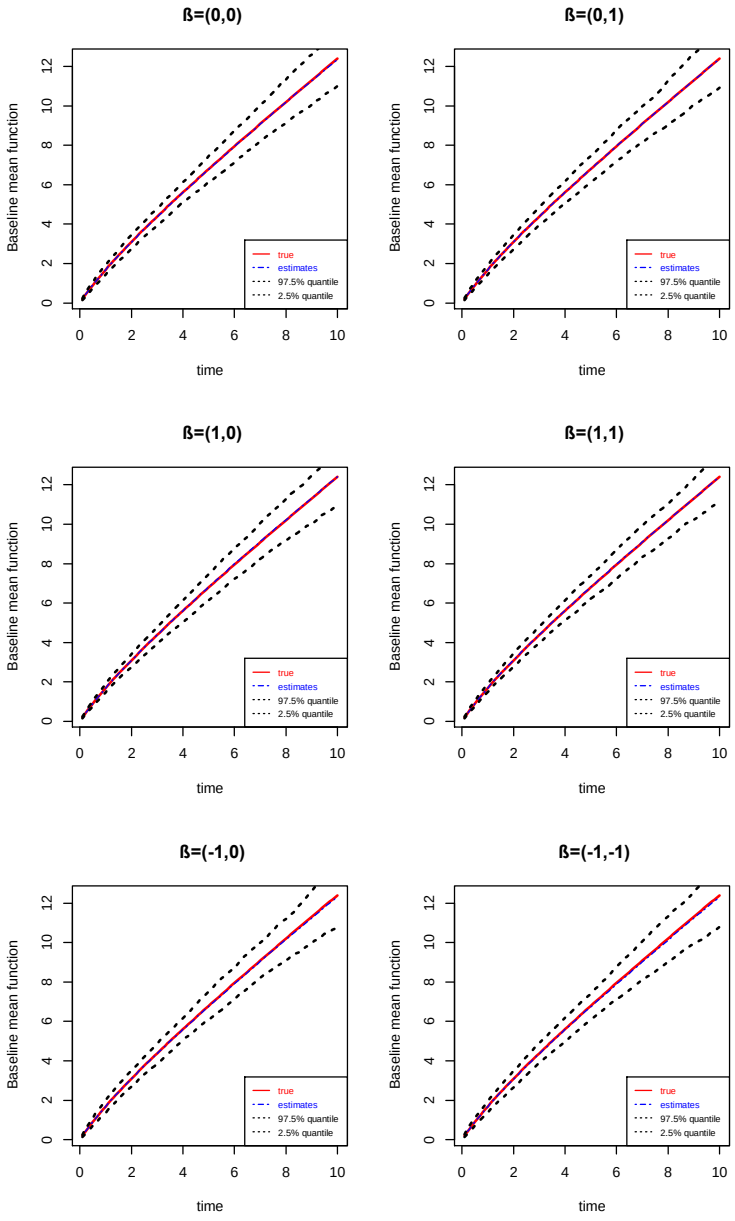

Figure 4.7: Estimates of baseline mean functions for different β in **scenario 1 when observation times depend on covariates**

Chapter 2 still performs very well when the observation times are dependent on the covariates.

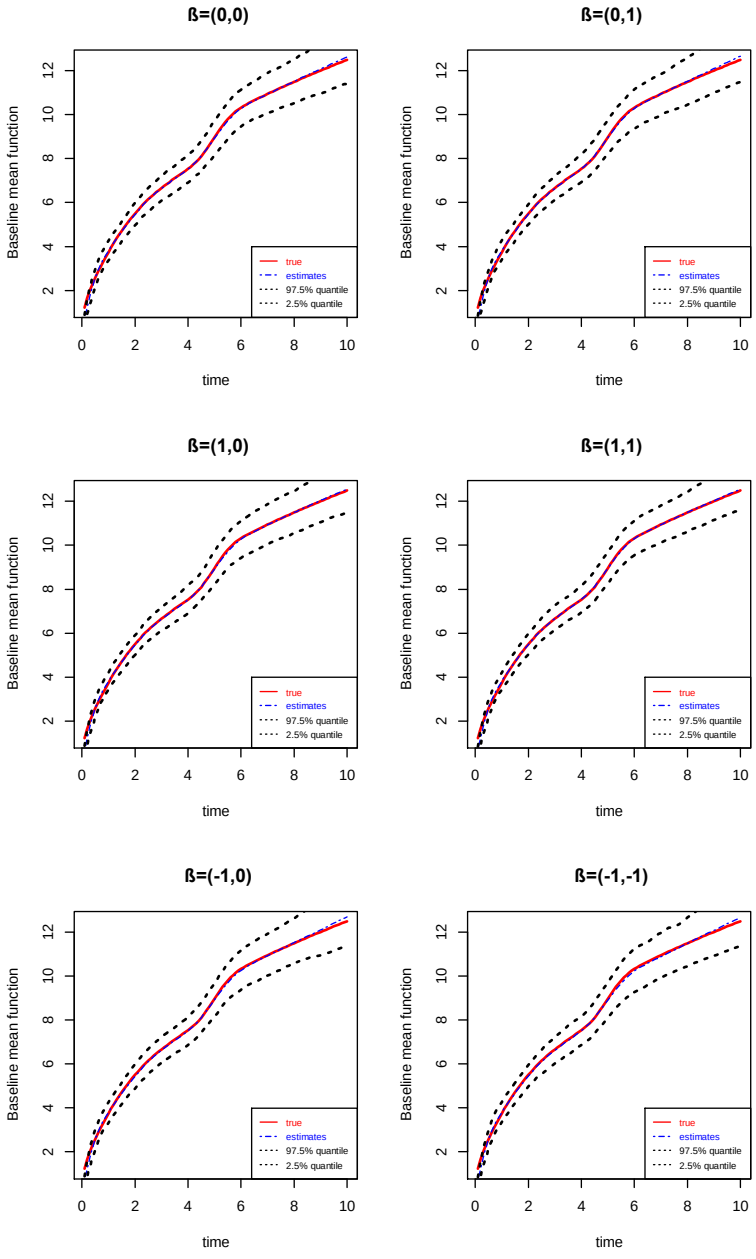

Figure 4.8: Estimates of baseline mean functions for different $\boldsymbol{\beta}$ in **scenario 2 when observation times depend on covariates**

4.5 Informative Observation Process

4.5.1 Introduction

In panel count data study, a general assumption behind most methods is that the observation counting process that characterizes the observation times is independent on the occurrences of recurrent events. In other words, the underlying event process of interest that controls the occurrence of the event and the observation counting process are independent. In practice, however, this may not be true. For instance, in a bladder cancer data study in Sun and Wei (2000) [68], the patients visited the clinical centers periodically, and the number of bladder tumors that occurred since the last visit was recorded at each clinical visit. Some patients in the study had more visits than others and the occurrence of bladder tumors and the number of visits may be related. In Sun, Tong and He (2007) [71], they calculated the sample correlation between the first visit times and the numbers of bladder tumors observed at these times and obtained the correlation -0.1180 for the patients in the placebo group. Although the evidence is weak, it indicates that there exists a negative correlationship between the tumor occurrence process and the observation time process. As they commented, the more visits a patient had, the less time the patient had for tumor recurrence, thus a lower tumor recurrence rate. Therefore, the correlation should be taken into account for analysis if possible.

Huang (2006) [52] proposed estimation procedures for nonparametric and semiparametric models that required no parametric assumption about the distribution of the frailty variable and the observation process. The models allowed observation times to be correlated with the event process through a frailty variable. The results suggested that the proposed method gave valid results, while the nonparametric maximum likelihood estimator and the nonparametric pseudo-likelihood estimator were substantially biased in their simulation scenarios. Sun, Tong and He (2007) [71] built

estimating equations for getting consistent estimates of the covariate covariance matrices by applying a bootstrap approach and discussed asymptotic properties of the parameter estimates.

In Chapters 2 and 3, we assume noninformative observation process. That is, the event process is independent on the observation process give coviarates. In the following, we are going to study the situation where the event process is dependent on the observation process through a common frailty term.

Consider a study involving n subjects, for each subject i, define $N_i(t)$ to be the cumulative number of events that have occurred prior to time t, $\quad 0 \leq t \leq \tau$, where τ is a known constant time point that can be viewed as the end of time point of the study or the event. For each subject i, $N_i(\cdot)$ is observed only at finite time points $t_{i1} < t_{i2} < \ldots < t_{iK_i}$, where K_i is a random variable that takes positive integer values. That is, only values of the $N_i(\cdot)'$s at these observation times are known. In general, not every subject can be followed until τ and for each subject there exists a censoring time C_i. Without loss of generality, we assume initial counts $N_i(t_{i0}) = 0$ for $i = 1, \ldots, n$. We have the whole set of observed panel count data denoted by $\mathscr{D}_i = \{t_{ij}, N_i(t_{ij}), X_i, \text{ for } i = 1, \cdots, n, \quad j = 1, \ldots, K_i, \quad C_i, \quad \tau\}$, where X_i is a p-dimensional vector of covariates.

We assume that given $X_i = (x_{i1}, \cdots, x_{ip})'$ and a subject-specific unobservable positive frailty w_i, the mean function of $N_i(t)$ has the form

$$E(N_i(t)|x_i, w_i] = w_i U_0(t) \exp(x_i'\beta),$$

where $U_0(t)$ is the baseline mean function.

Define $\bar{N}_i(t) = H_i\{\min(t_i, C_i)\}$, where $H_i(t) = \sum_{l=1}^{K_i} I(t_{il} \leq t)$ and $I(\cdot)$ is the indicator function for $i = 1, \ldots, n$. Therefore, $\bar{N}_i(t)$ is a point process, which characterizes the ith subject's observation process and only jumps at the observation times t_{ij}.

In literature, the observation process $H_i(t)$ is assumed a nonhomogeneous Poisson

process with the intensity function

$$\lambda_i(t) = w_i \lambda_0(t) \exp(x_i' \gamma),$$

where $\lambda_0(t)$ is an unknown and continuous baseline intensity function and γ denotes the vector of regression parameters. Therefore, the cumulative baseline intensity function is

$$\Lambda_0(t) = \int_0^t \lambda_0(s)\, d\, s.$$

Table 4.9: Estimation of regression parameters for **scenario 1 based on 500 simulated data sets for the informative observation times.** Bias refers to the difference between the average of the 500 point estimates and the true value, ESD refers to the average of the 500 posterior standard deviations, SSE refers to the sample standard deviation of the 500 point estimates, and CP95 is the 95% coverage probability

Scenario 1: $U_0(t) = t + \log(1+t)$										
			Results on β_1				Results on β_2			
β_1	β_2	Methods	Bias	SSE	ESD	CP95	Bias	SSE	ESD	CP95
0	0	Proposed	0.0032	0.0834	0.0581	0.830	0.0042	0.0820	0.0582	0.834
0	1	Proposed	-0.0018	0.0774	0.0548	0.820	0.0002	0.0888	0.0551	0.798
1	0	Proposed	0.0023	0.0695	0.0479	0.824	0.0044	0.0682	0.0430	0.770
1	1	Proposed	-0.0066	0.0746	0.045	0.764	-0.001	0.0779	0.0409	0.700
-1	0	Proposed	-0.0014	0.0971	0.0792	0.890	0.0039	0.0922	0.0704	0.864
-1	1	Proposed	0.0015	0.0959	0.0744	0.864	0.0019	0.0980	0.0667	0.816

Table 4.10: Estimation of regression parameters for **scenario 2 based on 500 simulated data sets for the informative observation times.** Bias refers to the difference between the average of the 500 point estimates and the true value, ESD refers to the average of the 500 posterior standard deviations, SSE refers to the sample standard deviation of the 500 point estimates, and CP95 is the 95% coverage probability

Scenario 2: $U_0(t) = 3(0.5\ \text{pnorm}(t,1,1) + 0.5\text{pnorm}(t,5,0.5) + t^{0.5})$										
			Results on β_1				Results on β_2			
β_1	β_2	Methods	Bias	SSE	ESD	CP95	Bias	SSE	ESD	CP95
0	0	proposed	-0.0014	0.0723	0.0523	0.828	-0.0022	0.0738	0.0521	0.820
0	1	proposed	0.0070	0.0755	0.0494	0.816	0.0031	0.0820	0.0500	0.754
1	0	proposed	-0.0019	0.0692	0.0431	0.776	0.0017	0.0679	0.0386	0.750
0	1	proposed	-0.0005	0.0724	0.0408	0.724	0.0008	0.0846	0.0366	0.626
-1	0	proposed	0.0020	0.1113	0.0711	0.888	-0.0030	0.0838	0.0631	0.874
-1	1	proposed	-0.0096	0.0900	0.0674	0.850	0.0006	0.0908	0.0609	0.806

For the identifiability issue, it assumes that $E(w|x) = E(w) = 1$ (Wang et al. 2001). It also assumes that $N_i(\cdot)$ and $H_i(\cdot)$ are mutually independent conditional on (x_i, w_i).

For the simulation, we adopt the similar setup as in Sun, Tong and He [71]. First let censoring time C_i be sampled from a uniform $(4, 10)$ distribution for each subject i. Suppose $\lambda_i(t) = w_i\lambda_0$, which results in a homogeneous Poisson process without depending on x_i. We let $\lambda_0 = 1$ and sample w_i from a gamma distribution with shape and rate parameters both as 10. Secondly, given C_i, sample the number of events

K_i as 1 plus a Poisson with mean parameter equal to $\int_0^{C_i} \lambda_i(s)\, d\, s = w_i C_i \lambda_0$. Then the observation times $T_{i1}, \ldots, T_{iK_i}$ are order statistics of a random sample of size K_i from a uniform $(0, C_i)$ distribution. Finally, generate counts between two adjacent observations as described at Section 2.6 in Chapter 2.

Tables 4.9 and 4.10 show the results of the parameter estimates for scenario 1 and scenario 2, respectively. Figures 4.9 and 4.10 present the estimates of the cumulative baseline functions with the pointwise 95% credible intervals and the true cumulative baseline function for scenario 1 and scenario 2, respectively. We can see that the estimates of coefficient parameters are very close to the true values. However, ESD, the average of the 500 posterior standard deviations is a little bit greater than SSE, the sample standard deviation of the 500 point estimates. In a result, the coverage probability is smaller than 95% for most cases.

However, the cumulative baseline mean function still have good estimates, which can be seen from figures 4.9 and 4.10 for the two scenarios by overlapping between the estimates and true values at most points. To summarize, our proposed method in Chapter 2 still performs well but underestimates the 95% coverage probability. In the following we investigate if the frailty model (3.5) will better fit the situation where the event process and the observation process are dependent.

4.5.2 Using frailty Model Approach to fit the Data from the Informative Observation Process

We apply the frailty model approach to fit the data with the event process dependent on the informative observation times. The results are shown in tables 4.11 and 4.12, and figures 4.11 and 4.12. We find that the estimates of the baseline mean function and covariate coefficients are still good. But the posterior standard deviations are overall overestimated, resulting in higher coverage probabilities.

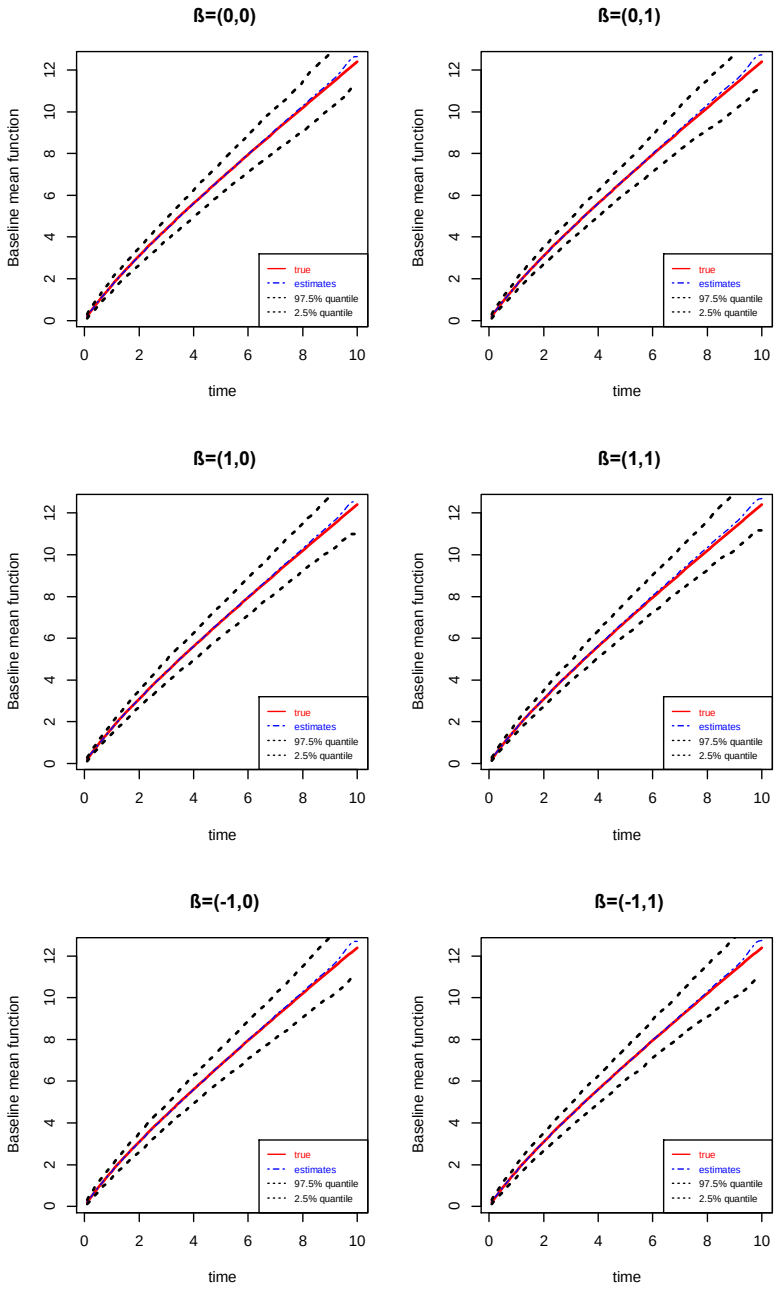

Figure 4.9: Estimates of baseline mean functions for different β in **scenario 1 for the informative observation process**

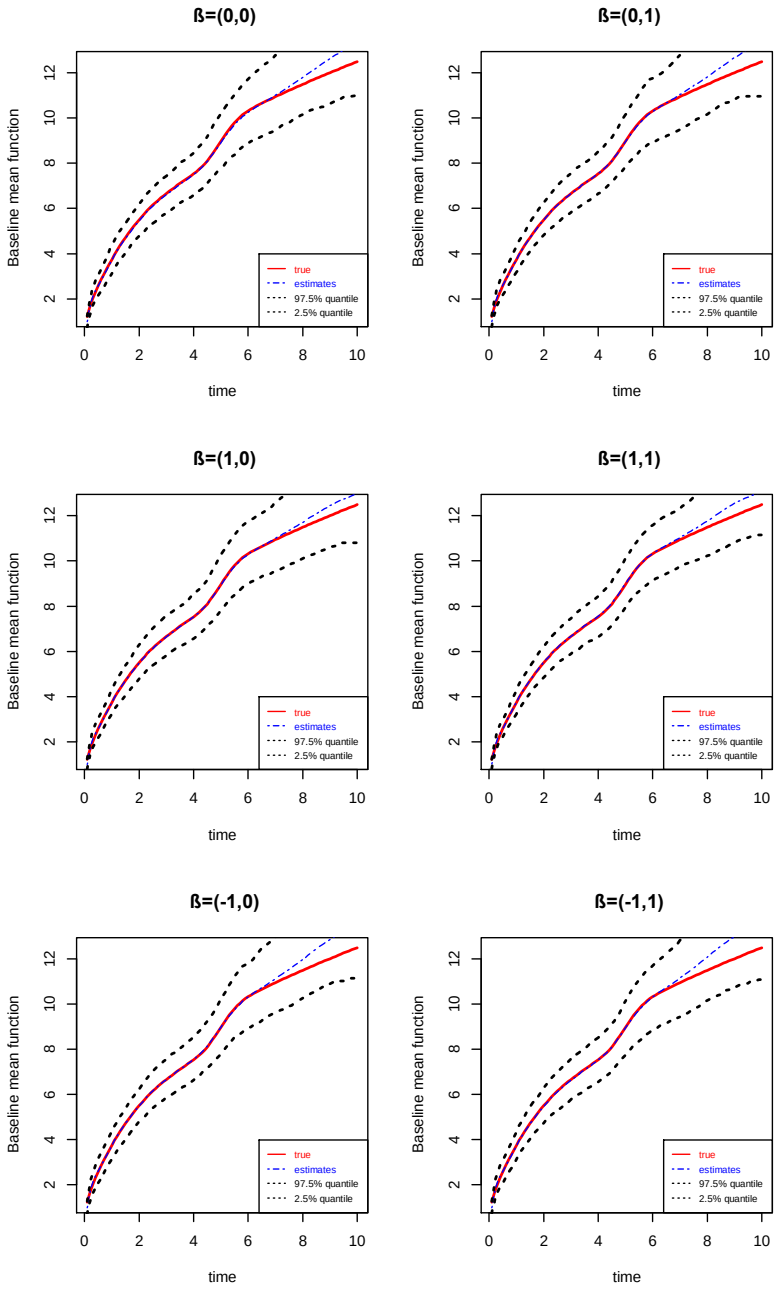

Figure 4.10: Estimates of baseline mean functions for different β in **scenario 2 for the informative observation process**

Table 4.11: Estimation of regression parameters for **scenario 2 using the frailty model approach based on 500 simulated data sets under the informative observation process.** Bias refers to the difference between the average of the 500 point estimates and the true value, ESD refers to the average of the 500 posterior standard deviations, SSE refers to the sample standard deviation of the 500 point estimates, and CP95 stands for the 95% coverage probability

Scenario 1: $U_0(t) = t + \log(1+t)$										
			Results on β_1				Results on β_2			
β_1	β_2	Methods	Bias	SSE	ESD	CP95	Bias	SSE	ESD	CP95
0	0	proposed	0.0017	0.0841	0.0966	0.969	-0.0048	0.0819	0.0970	0.982
0	1	proposed	-0.0016	0.0745	0.0971	0.994	0.0038	0.0818	0.1001	0.988
1	0	proposed	0.0037	0.0965	0.1027	0.990	0.0011	0.0838	0.0997	0.992
1	1	proposed	-0.0103	0.0705	0.0902	0.984	-0.0010	0.0706	0.0923	0.982
-1	0	proposed	-0.0036	0.0964	0.1115	0.980	0.0026	0.0929	0.1087	0.976
-1	1	proposed	-0.0010	0.0941	0.1112	0.990	0.0046	0.0931	0.1116	0.976

4.6 Conclusion

In this chapter, we investigate the robustness of the proposed methods in Chapter 2 through simulation when the assumptions fail to hold. The proposed Bayesian methods still have a good performance at most cases when the assumptions are invalid. When the observation process depends on the covariate variables, the proposed method with no frailty model has a very good performance. When the underlying event count is from a negative binomial process, the method for frailty model (3.5) has a better performance than the method with no frailty model (2.1). However, in

Table 4.12: Estimation of regression parameters for **scenario 2 using the frailty model approach based on 500 simulated data sets under the informative observation process.** Bias refers to the difference between the average of the 500 point estimates and the true value, ESD refers to the average of the 500 posterior standard deviations, SSE refers to the sample standard deviation of the 500 point estimates, and CP95 stands for the 95% coverage probability

Scenario 2: $U_0(t) = 3(0.5\ \text{pnorm}(t,1,1) + 0.5\text{pnorm}(t,5,0.5) + t^{0.5})$										
			Results on β_1				Results on β_2			
β_1	β_2	Methods	Bias	SSE	ESD	CP95	Bias	SSE	ESD	CP95
0	0	proposed	-0.0032	0.0721	0.0921	0.982	-0.0020	0.0741	0.0922	0.990
0	1	proposed	0.0042	0.0744	0.0926	0.988	0.0037	0.0759	0.0958	0.982
1	0	proposed	-0.0047	0.0686	0.0864	0.978	0.0024	0.0657	0.0866	0.988
1	1	proposed	-0.0022	0.0673	0.0874	0.986	-0.0007	0.0742	0.0887	0.978
-1	0	proposed	-0.0023	0.0910	0.1052	0.975	-0.0031	0.0848	0.1029	0.979
-1	1	proposed	0.0099	0.0878	0.1054	0.978	0.0039	0.0860	0.1068	0.988

the case that the event process and the observation process are dependent, the no frailty model method (2.1) underestimates the posterior standard deviations while the frailty model (3.5) overestimates the posterior standard deviations. Therefore, the next step, we may revise the model (3.5) as

$$E(N(t)|x,w) = w^{\alpha} U_0(t) \exp(x'\beta), \tag{4.2}$$

where α is a scale between 0 and 1. The similar strategy as in Chapters 2 and 3 can be applied in this model.

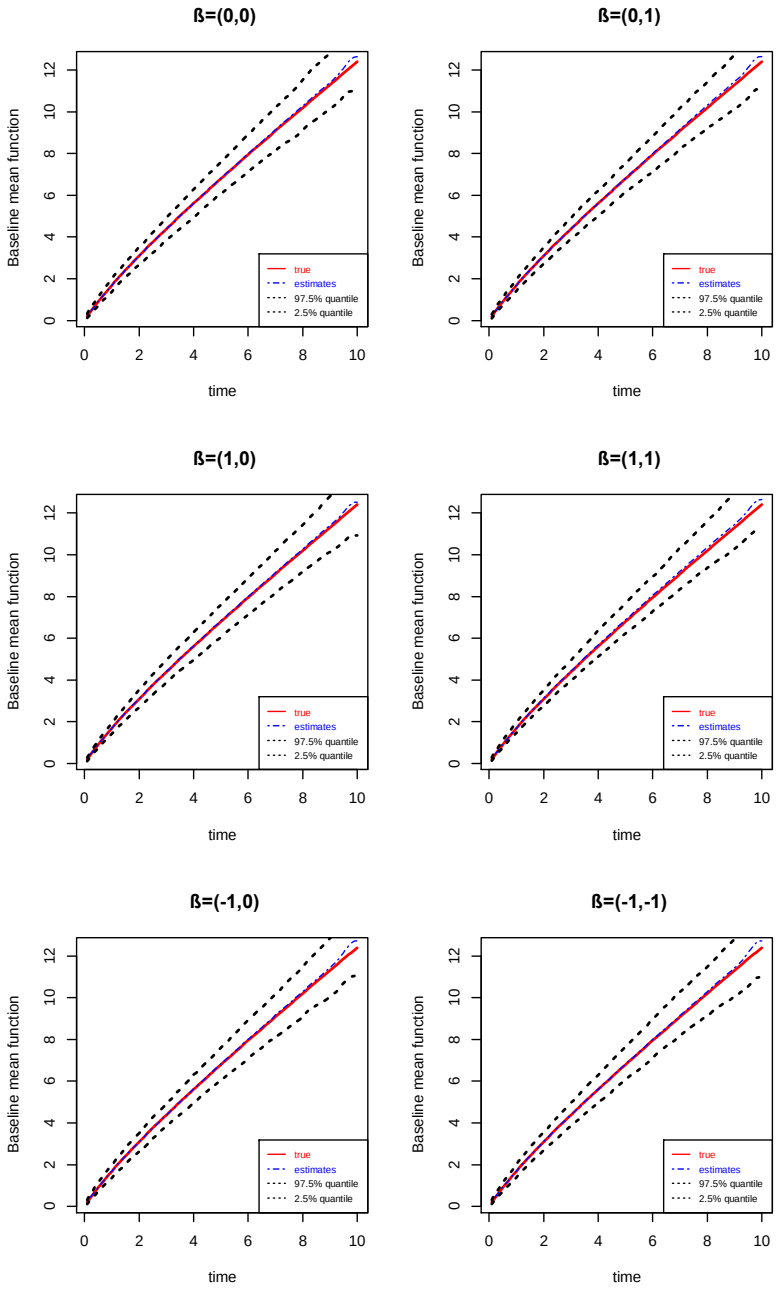

Figure 4.11: Estimates of baseline mean functions for different β in **scenario 1 using the frailty model fitting the data from the informative observation process**

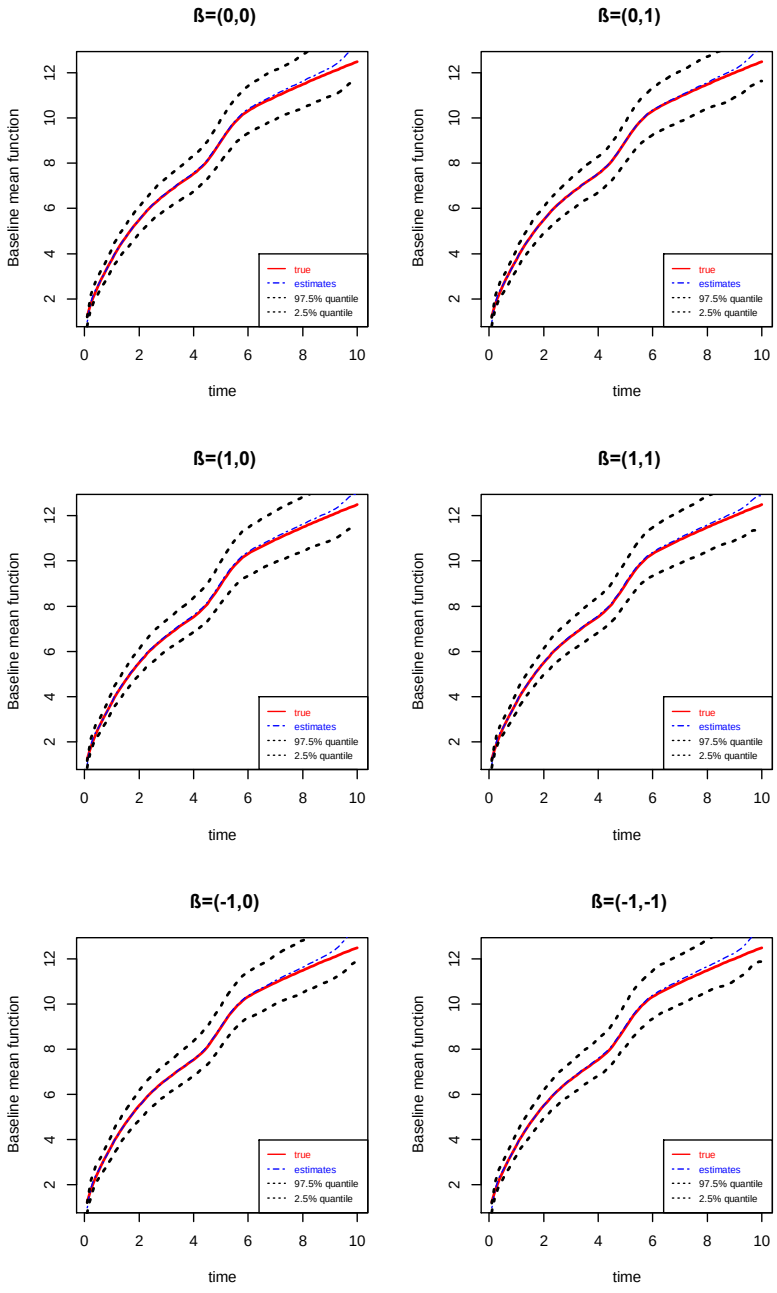

Figure 4.12: Estimates of baseline mean functions for different β in **scenario 2 using the frailty model fitting the data from the informative observation process**

Chapter 5

Future Plan

In Chapter 3, we studied the proportional mean model with gamma frailty effects for explaining the extra variation arising from the heterogeneity of subjects and the correlation among the cumulative counts within the same subject. We found out that fitting no frailty effect model for frailty data caused significantly worse results than fitting frailty effect model. Even for no frailty effect data, the frailty model provides robust estimation results. Therefore, including frailty into the model is a safe practice. Of course, when it is not necessary, including the frailty effect, will increase the computation burden and decrease the efficiency of the Gibbs sampler. Therefore, what type of models is applied really depends on the data themselves. We can use DIC to choose a better model.

For the future research, we will continue to study the frailty effect. We may first study the model (4.2) and see whether or how a scale parameter affects the estimates of posterior standard deviations.

In practice, for convenience, frailties are often assumed to follow a parametric distribution. However, it may come from different clusters, and each cluster has different distributions. From this perspective, we may study the frailty effects nonparametrically. Compared with the parametric and semiparametric methods, nonparametric frailty is more robust against the violation of parametric assumptions commonly made in parametric and semiparametric models. To ensure the continuity of the distribution, we may use a Dirichlet process (DP) mixture prior (Antoniak, 1974 [106]). Specifically, we let frailties independently and identically follow a Dirichlet process

(DP) mixture of gammas. We refer to West et al. (1994), Bush and MacEachern (1996) [107], Muller and Rosner (1997) [108], and Neal (2000) [109] etc.

Another potential research direction is to extend our proposed Bayesian approaches in Chapters 2 and 3 to analyze multivariate panel count data. Multivariate panel count data usually contain counts of several types of events of interest for the same subjects. The random effect (frailty) of subjects can play an important role to model the correlation among different types of counts.

Bibliography

1. Ramsay, J. O. Monotone regression splines in action. *Statistical science,* 425–441 (1988).

2. Kalbfleisch, J. & Lawless, J. F. The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association* **80,** 863–871 (1985).

3. Kalbfleisch, J. & Lawless, J. *Statistical inference for observational plans arising in the study of life history processes* in *Symposium on Statistical Inference and Applications In Honour of George Barnard's 65th Birthday* (1981), 5–8.

4. Thall, P. F. & Lachin, J. M. Analysis of recurrent events: Nonparametric methods for random-interval count data. *Journal of the American Statistical Association* **83,** 339–347 (1988).

5. Sun, J. & Kalbfleisch, J. Estimation of the mean function of point processes based on panel count data. *Statistica Sinica,* 279–289 (1995).

6. Byar, D. in *Bladder tumors and other topics in urological oncology* 363–370 (Springer, 1980).

7. Gail, M. H. Evaluating serial cancer marker studies in patients at risk of recurrent disease. *Biometrics,* 67–78 (1981).

8. Pepe, M. S. & Cai, J. Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *Journal of the American statistical Association* **88,** 811–820 (1993).

9. Lin, D., Wei, L., Yang, I. & Ying, Z. Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62,** 711–730 (2000).

10. Dinse, G. E. & Lagakos, S. Regression analysis of tumour prevalence data. *Applied statistics,* 236–248 (1983).

11. Diamond, I. D. & McDonald, J. W. Analysis of current-status data. (1992).

12. Jewell, N. P. & van der Laan, M. in *Lifetime Data: models in Reliability and Survival Analysis* 141–148 (Springer, 1996).

13. Byar, D., Blackard, C., Group, V. A. C. U. R., *et al.* Comparisons of placebo, pyridoxine, and topical thiotepa in preventing recurrence of stage I bladder cancer. *Urology* **10,** 556–561 (1977).

14. Wei, L.-J., Lin, D. Y. & Weissfeld, L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American statistical association* **84,** 1065–1073 (1989).

15. Wellner, J. A., Zhang, Y., *et al.* Two likelihood-based semiparametric estimation methods for panel count data with covariates. *The Annals of Statistics* **35,** 2106–2142 (2007).

16. Lu, M., Zhang, Y. & Huang, J. Estimation of the mean function with panel count data using monotone polynomial splines. *Biometrika* **94,** 705–718 (2007).

17. Lu, M., Zhang, Y. & Huang, J. Semiparametric estimation methods for panel count data using monotone B-splines. *Journal of the American Statistical Association* **104,** 1060–1070 (2009).

18. Sun, J. & Zhao, X. in *Statistical Analysis of Panel Count Data* 91–120 (Springer, 2013).

19. Wei, L. & Lachin, J. Two-sample asymptotically distribution-free tests for incomplete multivariate observations. *Journal of the American Statistical Association* **79,** 653–661 (1984).

20. Park, D.-H., Sun, J. & Zhao, X. A class of two-sample nonparametric tests for panel count data. *Communications in Statistics—Theory and Methods* **36,** 1611–1625 (2007).

21. Andersen, P. K. & Gill, R. D. Cox's regression model for counting processes: a large sample study. *The annals of statistics,* 1100–1120 (1982).

22. Lawless, J. F. Regression methods for Poisson process data. *Journal of the American Statistical Association* **82,** 808–815 (1987).

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

24. Whittemore, A. S. & Keller, J. B. Survival estimation using splines. *Biometrics,* 495–506 (1986).

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

26. Rosenberg, P. S. Hazard function estimation using B-splines. *Biometrics,* 874–887 (1995).

27. Aalen, O. *Statistical inference for a family of counting process* PhD thesis (Ph. D. Thesis, University of California, Berkeley, 1975).

28. Aalen, O. Nonparametric inference for a family of counting processes. *The Annals of Statistics,* 701–726 (1978).

29. Andersen, P. K. *et al.* Counting process models for life history data: A review [with discussion and reply]. *Scandinavian Journal of Statistics,* 97–158 (1985).

30. Liaw, L. *et al.* The adhesive and migratory effects of osteopontin are mediated via distinct cell surface integrins. Role of alpha v beta 3 in smooth muscle cell migration to osteopontin in vitro. *The Journal of clinical investigation* **95,** 713–724 (1995).

31. Andersen, E. & Cimasoni, G. A rapid and simple method for counting crevicular polymorphonuclear leucocytes. *Journal of clinical periodontology* **20,** 651–655 (1993).

32. Karlin, S. & Taylor, H. E. *A second course in stochastic processes* (Elsevier, 1981).

33. Melamed, B. Characterizations of Poisson traffic streams in Jackson queueing networks. *Advances in Applied probability* **11,** 422–438 (1979).

34. Gospodinov, D. & Rotondi, R. Exploratory analysis of marked Poisson processes applied to Balkan earthquake sequences. *Journal of the Balkan Geophysical Society* **4,** 61–68 (2001).

35. Xiao, S., Kottas, A., Sansó, B., *et al.* Modeling for seasonal marked point processes: An analysis of evolving hurricane occurrences. *The Annals of Applied Statistics* **9,** 353–382 (2015).

36. Cox, D. R. in *Breakthroughs in statistics* 527–541 (Springer, 1992).

37. Schoenberg, I. J. Spline functions and the problem of graduation. *Proceedings of the National Academy of Sciences* **52,** 947–950 (1964).

38. Alfaro, M. E. & Huelsenbeck, J. P. Comparative performance of Bayesian and AIC-based measures of phylogenetic model uncertainty. *Systematic Biology* **55,** 89–96 (2006).

39. Wang, L. Wilcoxon-type generalized Bayesian information criterion. *Biometrika* **96,** 163–173 (2009).

40. Ando, T. Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. *Biometrika* **94,** 443–458 (2007).

41. Erven, T. v., Grünwald, P. & De Rooij, S. Catching up faster by switching sooner: a predictive approach to adaptive estimation with an application to the AIC–BIC dilemma. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **74,** 361–417 (2012).

42. Claeskens, G., Hjort, N. L., *et al.* Model selection and model averaging. *Cambridge Books* (2008).

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

44. Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. *Bayesian data analysis. Texts in statistical science series* 2004.

45. Sun, J. & Zhao, X. *Statistical analysis of panel count data* (Springer, 2016).

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

47. Chib, S. & Greenberg, E. Analysis of multivariate probit models. *Biometrika* **85,** 347–361 (1998).

48. Sinha, D. & Maiti, T. A Bayesian Approach for the Analysis of Panel-Count Data with Dependent Termination. *Biometrics* **60,** 34–40 (2004).

49. Wellner, J. A., Zhang, Y. & Liu, R. *A semiparametric regression model for panel count data: when do pseudo-likelihood estimators become badly inefficient?* in *Proceedings of the Second Seattle Symposium in Biostatistics* (2004), 143–174.

50. Ghosh, D. & Lin, D. Semiparametric analysis of recurrent events data in the presence of dependent censoring. *Biometrics* **59,** 877–885 (2003).

51. Balshaw, R. F. & Dean, C. A Semiparametric Model for the Analysis of Recurrent-Event Panel Data. *Biometrics* **58,** 324–331 (2002).

52. Huang, C.-Y., Wang, M.-C. & Zhang, Y. Analysing panel count data with informative observation times. *Biometrika* **93,** 763–775 (2006).

53. Wang, L. & Dunson, D. B. Semiparametric Bayes' Proportional Odds Models for Current Status Data with Underreporting. *Biometrics* **67,** 1111–1118 (2011).

54. Wu, Y., Jiang, X., Kim, J. & Ohno-Machado, L. I-spline Smoothing for Calibrating Predictive Models. *AMIA Summits on Translational Science Proceedings* **2012,** 39 (2012).

55. Lin, X., Cai, B., Wang, L. & Zhang, Z. A Bayesian proportional hazards model for general interval-censored data. *Lifetime data analysis* **21,** 470–490 (2015).

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

57. Ibrahim, J. G., Chen, M.-H. & Sinha, D. in *Bayesian Survival Analysis* 1–29 (Springer, 2001).

58. Gilks, W. R. & Wild, P. Adaptive rejection sampling for Gibbs sampling. *Applied Statistics,* 337–348 (1992).

59. Park, T. & Casella, G. The bayesian lasso. *Journal of the American Statistical Association* **103,** 681–686 (2008).

60. Geman, S. & Geman, D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence,* 721–741 (1984).

61. Gelfand, A. E. & Smith, A. F. Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association* **85,** 398–409 (1990).

62. Casella, G., Robert, C. P. & Wells, M. T. Mixture models, latent variables and partitioned importance sampling. *Statistical Methodology* **1,** 1–18 (2004).

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

64. Pan, C., Cai, B., Wang, L. & Lin, X. Bayesian semiparametric model for spatially correlated interval-censored survival data. *Computational Statistics & Data Analysis* **74,** 198–208 (2014).

65. Rosen, J. B. The gradient projection method for nonlinear programming. Part I. Linear constraints. *Journal of the society for industrial and applied mathematics* **8,** 181–217 (1960).

66. Jamshidian, M. On algorithms for restricted maximum likelihood estimation. *Computational statistics & data analysis* **45,** 137–157 (2004).

67. Zhang, Y. & Jamshidian, M. On algorithms for the nonparametric maximum likelihood estimator of the failure function with censored data. *Journal of Computational and Graphical Statistics* **13,** 123–140 (2004).

68. Sun, J. & Wei, L. Regression analysis of panel count data with covariate-dependent observation and censoring times. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62,** 293–302 (2000).

69. Cox, D. Regression models and life tables. JR Statist. Soc. B 34, 187-202. Cox18734J. R. *Statist. Soc B* **1972** (1972).

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

71. Sun, J., Tong, X. & He, X. Regression analysis of panel count data with dependent observation times. *Biometrics* **63,** 1053–1059 (2007).

72. Keiding, N., Andersen, P. K. & Klein, J. P. The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in medicine* **16,** 215–224 (1997).

73. Lancaster, T. Econometric methods for the duration of unemployment. *Econometrica: Journal of the Econometric Society,* 939–956 (1979).

74. Hougaard, P. Life table methods for heterogeneous populations: distributions describing the heterogeneity. *Biometrika* **71,** 75–83 (1984).

75. Clayton, D. G. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65,** 141–151 (1978).

76. Hougaard, P. Analysis of Multivariate Survival Data. Springer-Verlag, New York. (2000).

77. Therneau, T. M. & Grambsch, P. M. in *Modeling Survival Data: Extending the Cox Model* 261–287 (Springer, 2000).

78. Duchateau, L. & Janssen, P. the frailty model (Statistics for Biology and Health). *Springer: New York* **10,** 978– (2008).

79. Hanagal, D. D. & Dabade, A. D. Bayesian estimation of parameters and comparison of shared gamma frailty models. *Communications in Statistics-Simulation and Computation* **42,** 910–931 (2013).

80. McGilchrist, C. & Aisbett, C. Regression with frailty in survival analysis. *Biometrics,* 461–466 (1991).

81. Hougaard, P. Frailty models for survival data. *Lifetime data analysis* **1,** 255–273 (1995).

82. Congdon, P. Modelling frailty in area mortality. *Statistics in medicine* **14,** 1859–1874 (1995).

83. Sahu, S. K., Dey, D. K., Aslanidou, H. & Sinha, D. A Weibull regression model with gamma frailties for multivariate survival data. *Lifetime data analysis* **3,** 123–137 (1997).

84. Yu, B. Estimation of shared gamma frailty models by a modified EM algorithm. *Computational statistics & data analysis* **50,** 463–474 (2006).

85. Ibrahim, J. G., Chen, M.-H. & Sinha, D. *Bayesian survival analysis* (Springer Science & Business Media, 2013).

86. Abrahantes, J. C. *et al.* Comparison of different estimation procedures for proportional hazards model with random effects. *Computational statistics & data analysis* **51,** 3913–3930 (2007).

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

88. Du, P., Jiang, Y. & Wang, Y. Smoothing spline ANOVA frailty model for recurrent event data. *Biometrics* **67,** 1330–1339 (2011).

89. Klein, J. P. Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics,* 795–806 (1992).

90. Nielsen, G. G., Gill, R. D., Andersen, P. K. & Sørensen, T. I. A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian journal of Statistics,* 25–43 (1992).

91. Wang, N., Wang, L. & McMahan, C. S. Regression analysis of bivariate current status data under the Gamma-frailty proportional hazards model using the EM algorithm. *Computational Statistics & Data Analysis* **83,** 140–150 (2015).

92. Doha, E. H., Bhrawy, A. & Saker, M. Integrals of Bernstein polynomials: an application for the solution of high even-order differential equations. *Applied Mathematics Letters* **24,** 559–565 (2011).

93. Therneau, T. M., Grambsch, P. M. & Pankratz, V. S. Penalized survival models and frailty. *Journal of computational and graphical statistics* **12,** 156–175 (2003).

94. Du, P., Ma, S. & Liang, H. Penalized variable selection procedure for Cox models with semiparametric relative risk. *Annals of statistics* **38,** 2092 (2010).

95. Zhang, J. Powerful goodness-of-fit tests based on the likelihood ratio. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64,** 281–294 (2002).

96. Wang, X., Ma, S. & Yan, J. Augmented estimating equations for semiparametric panel count regression with informative observation times and censoring time. *Statistica Sinica,* 359–381 (2013).

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

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

99. Abbring, J. H. & Van Den Berg, G. J. The unobserved heterogeneity distribution in duration analysis. *Biometrika* **94,** 87–99 (2007).

100. Ramaswamy, V., Anderson, E. W. & DeSarbo, W. S. A disaggregate negative binomial regression procedure for count data analysis. *Management Science* **40,** 405–417 (1994).

101. Cameron, A. C. & Trivedi, P. K. Regression-based tests for overdispersion in the Poisson model. *Journal of econometrics* **46,** 347–364 (1990).

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

103. Breslow, N. E. Extra-Poisson variation in log-linear models. *Applied statistics,* 38–44 (1984).

104. Cameron, A. C. & Trivedi, P. K. Econometric models based on count data. Comparisons and applications of some estimators and tests. *Journal of applied econometrics* **1,** 29–53 (1986).

105. Gardner, W., Mulvey, E. P. & Shaw, E. C. Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological bulletin* **118,** 392 (1995).

106. Antoniak, C. E. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The annals of statistics,* 1152–1174 (1974).

107. Bush, C. A. & MacEachern, S. N. A semiparametric Bayesian model for randomised block designs. *Biometrika* **83,** 275–285 (1996).

108. Müller, P. & Rosner, G. L. A Bayesian population model with hierarchical mixture priors applied to blood count data. *Journal of the American Statistical Association* **92,** 1279–1292 (1997).

109. Neal, R. M. Markov chain sampling methods for Dirichlet process mixture models. *Journal of computational and graphical statistics* **9,** 249–265 (2000).

110. Breslow, N. Tests of hypotheses in overdispersed Poisson regression and other quasi-likelihood models. *Journal of the American Statistical Association* **85,** 565–571 (1990).

111. Lawless, J. F. & Nadeau, C. Some simple robust methods for the analysis of recurrent events. *Technometrics* **37,** 158–168 (1995).

112. Zhao, X., Tong, X. & Sun, J. Robust estimation for panel count data with informative observation times. *Computational Statistics & Data Analysis* **57,** 33–40 (2013).

113. Hu, X. J., Sun, J. & WEI, L.-J. Regression parameter estimation from panel counts. *Scandinavian Journal of Statistics* **30,** 25–43 (2003).