Time-Varying Coefficient Models for Recurrent Events

Yi Liu

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Statistics

Feng Guo, Chair
Xinwei Deng
Yili Hong
Inyoung Kim

November 8, 2018
Blacksburg, Virginia

Time-Varying Coefficient Models for Recurrent Events

Yi Liu

Abstract

I have developed time-varying coefficient models for recurrent event data to evaluate the temporal profiles for recurrence rate and covariate effects. There are three major parts in this dissertation. The first two parts propose a mixed Poisson process model with gamma frailties for single type recurrent events. The third part proposes a Bayesian joint model based on multivariate log-normal frailties for multi-type recurrent events. In the first part, I propose an approach based on penalized B-splines to obtain smooth estimation for both time-varying coefficients and the log baseline intensity. An EM algorithm is developed for parameter estimation. One issue with this approach is that the estimating procedure is conditional on smoothing parameters, which have to be selected by cross-validation or optimizing certain performance criterion. The procedure can be computationally demanding with a large number of time-varying coefficients. To achieve objective estimation of smoothing parameters, I propose a mixed-model representation approach for penalized splines. Spline coefficients are treated as random effects and smoothing parameters are to be estimated as variance components. An EM algorithm embedded with penalized quasi-likelihood approximation is developed to estimate the model parameters. The third part proposes a Bayesian joint model with time-varying coefficients for multi-type recurrent events. Bayesian penalized splines are used to estimate time-varying coefficients and the log baseline intensity. One challenge in Bayesian penalized splines is that the smoothness of a spline fit is considerably sensitive to the subjective choice of hyperparameters. I establish a procedure to objectively determine the hyperparameters through a robust prior specification. A Markov chain Monte Carlo procedure based on Metropolis-adjusted Langevin algorithms is developed to sample from the high-dimensional distribution of spline coefficients. The procedure includes a joint sampling scheme to achieve better convergence and mixing properties. Simulation studies

in the second and third part have confirmed satisfactory model performance in estimating time-varying coefficients under different curvature and event rate conditions. The models in the second and third part were applied to data from a commercial truck driver naturalistic driving study. The application results reveal that drivers with 7-hours-or-less sleep prior to a shift have a significantly higher intensity after 8 hours of on-duty driving and that their intensity remains higher after taking a break. In addition, the results also show drivers' self-selection on sleep time, total driving hours in a shift, and breaks. These applications provide crucial insight into the impact of sleep time on driving performance for commercial truck drivers and highlights the on-road safety implications of insufficient sleep and breaks while driving. This dissertation provides flexible and robust tools to evaluate the temporal profile of intensity for recurrent events.

Keywords: Multivariate Frailty, Penalized Likelihood, Variance Component, Laplace Approximation, Markov Chain Monte Carlo, Metropolis-Adjusted Langevin Algorithm, Truck Driving Safety.

Time-Varying Coefficient Models for Recurrent Events

Yi Liu

Abstract

(General Audience)

The overall objective of this dissertation is to develop models to evaluate the time-varying profiles for event occurrences and the time-varying effects of risk factors upon event occurrences. There are three major parts in this dissertation. The first two parts are designed for single event type. They are based on approaches such that the whole model is conditional on a certain kind of tuning parameter. The value of this tuning parameter has to be pre-specified by users and is influential to the model results. Instead of pre-specifying the value, I develop an approach to achieve an objective estimate for the optimal value of tuning parameter and obtain model results simultaneously. The third part proposes a model for multi-type events. One challenge is that the model results are considerably sensitive to the subjective choice of hyperparameters. I establish a procedure to objectively determine the hyperparameters. Simulation studies have confirmed satisfactory model performance in estimating the temporal profiles for both event occurrences and effects of risk factors. The models were applied to data from a commercial truck driver naturalistic driving study. The results reveal that drivers with 7-hours-or-less sleep prior to a shift have a significantly higher intensity after 8 hours of on-duty driving and that their driving risk remains higher after taking a break. In addition, the results also show drivers' self-selection on sleep time, total driving hours in a shift, and breaks. These applications provide crucial insight into the impact of sleep time on driving performance for commercial truck drivers and highlights the on-road safety implications of insufficient sleep and breaks while driving. This dissertation provides flexible and robust tools to evaluate the temporal profile of both event occurrences and effects of risk factors.

Acknowledgments

I would like to thank my advisor Dr. Feng Guo, for his illuminating guidance, support, and encouragement. It has been an honor to be his student. I am fortunate enough to have the opportunity to work with him. I sincerely appreciate my committee members, Dr. Xinwei Deng, Dr. Yili Hong and Dr. Inyoung Kim for their assistance and advice. I would like to thank Dr. Eric A. Vance for providing me an opportunity to work as a collaborative researcher for Dr. Denise R. Simmons. I want to thank Dr. Simmons for always being kind and encouraging to me. Thank you to the faculty and staff from the Department of Statistics and Virginia Tech for all the support. My thanks go to all the talented teammates in Dr. Guo's research group, Dr. Youjia Fang, Dr. Qing Li, Dr. Chen Chen, Jon Atwood, Danni Lu and Huiying Mao for their generous help. Finally, my most special thanks go to my mother, Yuxian Geng, and my father, Zhiliang Liu, for their endless love and support.

Contents

List of Figures

List of Tables

Chapter 1 Introduction

This chapter introduces the background of the dissertation in Section 1.1. Sections 1.2 to 1.4 review the recurrent events models, varying coefficient models, and spline smoothing methodology. The Commercial Truck Driver Naturalistic Driving Study (NDS) is used as case studies and the study is described in Section 1.5. The structure of the dissertation is introduced in Section 1.6.

1.1 Background

Recurrent events are commonly observed in clinical trials and epidemiological studies in which individuals can experience repeated events of interest over the observation period. Examples include multiple occurrences of bladder tumors for patients in bladder cancer trials (Byar et al., 1977), multiple mammary tumors for animals in carcinogenicity experiments (Gail et al., 1980), and recurrent pulmonary exacerbations for patients with cystic fibrosis (Therneau, Hamilton, 1997). One typical feature of recurrent event data is that event times within the same subject are correlated. Approaches for recurrent event data analysis include marginal methods based on estimating functions (e.g., Pepe, Cai, 1993; Lawless, Nadeau, 1995; Lin et al., 2000) and conditional methods based on frailty distributions (e.g., Lawless, 1987; Nielsen et al., 1992; Vaida, Xu, 2000). Methods using estimating functions directly model marginal means and treat the correlation of event times as a nuisance parameter. On the contrary, conditional models use frailty distributions to characterize the heterogeneity and correlation among events, allowing both marginal and conditional inferences. One challenge in recurrent event modeling is to quantify how recurrences vary over time and to compare the temporal profile of recurrences across different covariate values while accounting

for the correlation structure between events. Yet limited research has been conducted on developing approaches that can be directly implemented to address these issues. General and flexible methods are in need for characterizing the time-varying patterns of recurrent event data.

The effect of a covariate can vary over time for recurrent events and it could be of primary interest. For example, total sleep time the night before driving is a critical risk factor affecting driver fatigue, and its impact on driving performance is likely to vary over a long course of driving due to fatigue (Hanowski et al., 2007). In the Commercial Truck Driver NDS (see Section 1.5 for details), interests lie in evaluating how sleep time impacts on the temporal profile of driving performance. Varying coefficient models for longitudinal data include kernel-based methods (e.g., Carroll et al., 1998; Cai et al., 2000) and spline-based methods (e.g., Hastie, Tibshirani, 1993; Huang et al., 2002). Kernel-based estimators are constructed through a weighted mean of the nearby observations and therefore obtain great intuitive appeal. However, the kernel weights are asymmetric at boundaries of the predictor space and could cause substantial bias for the estimate (Hastie, Loader, 1993). Spline-based estimators are based on linear combinations of piecewise polynomial functions and have a straightforward extension of linear regression models. In addition, the computation is relatively fast and can be easily implemented (Eilers, Marx, 1996). Varying coefficient models have also been studied in the recurrent event data setting (Amorim et al., 2008; Sun et al., 2011; Yu et al., 2013). However, these approaches are based on the partial likelihood approach (Cox, 1975), in which the baseline intensity is considered as a nuisance parameter and not explicitly estimated. In the Commercial Truck Driver NDS, driving performance could change over a long shift due to fatigue and other factors, and its temporal profile is also of interest. Therefore, smooth estimation of the baseline intensity needs to be incorporated for characterizing the time-varying pattern of driving performance.

The methodologies of this dissertation is motivated by interests in evaluating driving safety and the related risk factors for commercial truck driving. Traffic crashes are a major safety hazard for commercial truck drivers and are also a general public health concern. In the

United States, the operational hours for commercial truck drivers are regulated by the Federal Motor Carrier Safety Administration (FMCSA). Referred to as the hours-of-service (HOS) regulations, truck drivers can work 14 hours, 11 hours of which can be driving, before taking a 10-hour off-duty break (Federal Register, 2011). Though not all drivers work the fully 14 hours, or drive the 11 hours as HOS allowed, for drivers that do push the work/driving limits (e.g., long-haul truck drivers), fatigue is a major risk factor (National Academies of Sciences, Engineering, and Medicine, 2016). Fatigue issues associated with time-on-task and deterioration of driving performance over the course of a long work day or drive are of particular concern for this cohort. Though the importance of being well-rested for optimal performance has been documented (e.g., Banks, Dinges, 2007), there has been limited research to evaluate the change in driving performance and safety risk over time as a function of pre-shift sleep time.

Many factors can affect driver fatigue, including total sleep time, sleep debt, circadian rhythm, breaks during long shifts, and caffeine consumption. Among these factors, sleep time the night before driving has been noted as a critical risk factor (Hanowski et al., 2007). Being well rested prior to a long shift has a direct impact on driver fatigue and driving performance (Banks, Dinges, 2007; Lim, Dinges, 2008, 2010). Insufficient sleep is typically defined as sleep less than 7 hours per day (Van Dongen et al., 2003a,b; Ford et al., 2015; Watson et al., 2015a,b). Sleep debt, the cumulative lack of sleep for a period of time, also negatively affects driver fatigue (Abe et al., 2012; Scott et al., 2007). Furthermore, it has also been shown that breaks during a long shift reduce the number of on-road safety-critical events for truck drivers (Soccolich et al., 2013).

Chapter 2 develops a mixed Poisson process model with time-varying coefficients for recurrent event data. The correlation among event times is incorporated via gamma frailties. The approach of penalized B-splines (Eilers, Marx, 1996) is adopted for modeling both the log baseline intensity and time-varying coefficients. The cumulative intensity in the penalized likelihood is approximated by numerical integration based on the trapezoid method (Kauermann, 2005). An EM algorithm is used for parameter estimation. One issue with

the penalized likelihood approach is that the spline fit is conditional on the smoothing parameter, whose value has to be selected by cross-validation or optimizing some performance criterion. As the number of time-varying coefficients increases, these procedures could be computationally demanding.

Chapter 3 proposes the same mixed Poisson process model with time-varying coefficients as in Chapter 2, but with a mixed-model representation approach based on the connection between penalized likelihood and random effect models (Brumback, Rice, 1998; Wang, 1998a,b; Brumback et al., 1999). By the mixed-model representation approach, spline coefficients are treated as random effects and smoothing parameters are estimated as variance components. The EM algorithm developed in Chapter 2 is embedded with the penalized quasi-likelihood (PQL) approach (Breslow, Clayton, 1993) for likelihood approximation.

Chapter 4 develops a Bayesian joint model with time-varying coefficients for multi-type recurrent event data. The joint model is based on multivariate log-normal frailties to characterize the heterogeneity and correlation among different event types. I use Bayesian penalized splines (Lang, Brezger, 2004) to achieve smooth estimates for both time-varying coefficients and the baseline intensity. In Bayesian penalized splines, the smoothness of a spline fit depends substantially on the subjective choice of hyperparameters (Brezger, Lang, 2006). However, re-fitting the model with a number of choices can be computationally demanding. I adopt a robust prior specification (Jullion, Lambert, 2007) to achieve an objective fitting procedure. One challenge in using Markov chain Monte Carlo (MCMC) methods for posterior inference is to sample from the high-dimensional distribution of spline parameters. I use Metropolis-adjusted Langevin (MALA, Roberts, Tweedie, 1996) algorithms, a class of Metropolis-Hastings algorithms whose proposals are based on the gradient of the target density. Another concern in sampling spline parameters is that their MALA proposal strongly depends on a roughness penalty parameter introduced by Bayesian penalized splines. Separately updating these two parameters could lead to convergence issues. I integrate MALA algorithms into a joint sampling scheme (Knorr-Held, Rue, 2002) for better convergence and mixing properties.

This dissertation uses data from the Commercial Truck Driver NDS (Blanco et al., 2011) as an example for application. The study was sponsored by the FMCSA to collect large-scale naturalistic driving data and to evaluate issues on commercial motor vehicles. Repeated on-road safety-critical events were identified for the recruited drivers, including crashes, near-crashes, crash-relevant conflicts and unintentional lane deviations. The application is to evaluate how driving performance changes over time and to compare this temporal profile across sleep time before driving.

1.2 Recurrent Events

This section introduces notations and frameworks for modeling recurrent event data. For a single recurrent event process starting for at $t = 0$, let $0 \leq T_1 < T_2 < \cdots$ denote the event times, where T_i is the time of the ith event. The associated counting process $\{N(t), t \geq 0\}$ records the cumulative number of events generated by the process. $N(t) = \sum_{i=1}^{\infty} I(T_i \leq t)$ indicates the number of occurrences over the time interval $[0, t]$. For $s < t$, $N(t) - N(s)$ represents the number of events that have occurred in the interval $(s, t]$. In addition, t^- and t^+ are used to indicate times infinitesimally smaller or larger than t. Consider a small interval $[t, t + h)$, where $h > 0$. Let $H(t) = \{N(s), 0 \leq s < t\}$ denote the history of the process at time t. The intensity function for the process $\{N(t), t \geq 0\}$ associated with $H(t)$ is defined as

$$\lambda(t|H(t)) = \lim_{h \downarrow 0} \frac{P\left\{N(t + h^-) - N(t^-) = 1 \,\middle|\, H(t)\right\}}{h}, \tag{1.1}$$

where $N(t + h^-) - N(t^-)$ equals the increment of events over the interval $[t, t + h)$. The intensity function $\lambda(t|H(t))$ indicates the instantaneous probability of an event occurring at time t, conditional on the process history. The intensity is a fundamental concept in that it mathematically defines an event process, and all process characteristics can be derived from it.

Consider an event process with the intensity function (1.1) observed over a time interval $[c_0, c]$. Conditional on $H(c_0)$, the probability density function of observing n events at times $t_1 < t_2 < \cdots < t_n$, where $n \geq 0$, is

$$\prod_{i=1}^{n} \lambda(t_i|H(t_i)) \exp\left\{-\int_{c_0}^{c} \lambda(s|H(s))\,\mathrm{d}s\right\}. \tag{1.2}$$

The Poisson process is one commonly used framework for modeling recurrent event data. It has been discussed with extensive details in books on stochastic processes (e.g., Karlin, Taylor, 1975; Ross, 1996). One way to define a Poisson process is via the intensity function: a Poisson process is one for which the intensity function follows the form

$$\lambda(t|H(t)) = \rho(t),\ t > 0, \tag{1.3}$$

where $\rho(t)$ is a non-negative integrable function. If $\rho(t)$ is time constant, the process is called homogeneous; otherwise, it is called non-homogeneous.

1.3 Varying Coefficient Models

The varying coefficient model is an important extension of the linear regression model with constant coefficients. It was orignally introduced by Cleveland et al. (1991). Later Hastie, Tibshirani (1993) gave a detailed description on the estimating procedures for the class of models.

Consider a random variable Y whose distribution depends on a parameter η, and the associated predictors $X_1, X_2, \ldots, X_p$ and $R_1, R_2, \ldots, R_p$. The varying coefficient model is of the form

$$\eta = \beta_0 + \beta_1(R_1)X_1 + \cdots + \beta_p(R_p)X_p\,. \tag{1.4}$$

R_j, often known as the "effect modifier", changes the coefficient of X_j through the unspecified function $\beta_j(\cdot)$. The dependence of $\beta_j(\cdot)$ on R_j indicates a special kind of interaction between each pair of X_j and R_j. One special case of the varying coefficient models is to let R_j be the time variable. The time-varying coefficient model naturally arises when one is interested in exploring how the regression coefficient affects the response over time.

1.4 Spline Smoothing

Let the data be n pairs $(x_i,\, y_i)$, where for simplicity x_i is univariate. Consider the model of the form

$$y_i = f(x_i) + \epsilon_i,$$

where f is an unknown function representing the nonlinear relationship between y and x, and ϵ_i's are independent and identically distributed random errors, with mean zero and a finite variance. The objective is to obtain a smooth estimate of the unknown function f using the methods of spline smoothing.

The theory of spline functions is briefly described as follows. A spline function is defined as a piecewise polynomial of degree p. Let $[a, b]$ be a finite closed interval and let $a = \kappa_0 < \kappa_1 < \cdots < \kappa_k < \kappa_{k+1} = b$. The set $\{\kappa_i\}_{i=0}^{k+1}$ partitions the interval $[a,\, b]$ into $k+1$ subintervals: $I_i = [\kappa_i, \kappa_{i+1})$, $i = 0, 1, \ldots, k-1$, and $I_k = [\kappa_k, \kappa_{k+1}]$. Let

$$\mathcal{S}_{p,\boldsymbol{\kappa}} = \mathcal{S}_p(\kappa_1, \ldots, \kappa_k) = \Big\{ f \in \mathcal{C}^{p-1}[a, b] : f(x \in I_i) \in \mathcal{P}_p, \text{ for } i = 0, 1, \ldots, k \Big\},$$

where $\mathcal{C}^{p-1}[a, b]$ is the space of functions having continuous $(p-1)$th derivative in $[a,\, b]$, and $\mathcal{P}_p$ is the space of polynomials with degree p. Then $\mathcal{S}_{p,\boldsymbol{\kappa}}$ is called the space of polynomial splines of degree p with knots $\kappa_1, \ldots, \kappa_k$. The spline space $\mathcal{S}_{p,\boldsymbol{\kappa}}$ is a finite dimensional linear space. Particularly, $p+k+1$ basis functions are needed to span the space. Every continuous

function on the interval $[a, b]$ can be approximated arbitrarily well by polynomial splines of degree p, given a sufficient number of knots.

Assume $f \in \mathcal{S}_{p,\boldsymbol{\kappa}}$ and let $\{B_1(x), \ldots, B_{p+k+1}(x)\}$ be the set of basis function for $\mathcal{S}_{p,\boldsymbol{\kappa}}$. Then f can be approximated in terms of a linear combination of the basis functions:

$$f(x) = \sum_{j=1}^{p+k+1} \beta_j B_j(x). \tag{1.5}$$

The smoothing problem is simply reduced to the regression of n pairs of data points (x_i, y_i) on a set of basis functions $B_j(\cdot)$. The objective function to minimize is

$$\sum_{i=1}^{n} \left\{ y_i - \sum_{j=1}^{p+k+1} \beta_j B_j(x_i) \right\}^2. \tag{1.6}$$

Unfortunately the spline smoothing problem is complicated by choosing the knots and the degree of splines. Especially knot selection has a strong influence upon the spline fit.

Two main approaches to knot selection have been proposed in the literature. The first approach, known as the free-knot approach, is to treat the knots as free parameters that need to estimate. Computational methods, such as adaptive regression splines, treat the free knot problem as a predictor subset selection problem. The knots are selected from a set of pre-specified candidate points by a technique similar to stepwise regression. Given the selected knots, the coefficients are estimated by ordinary least squares. Related literature includes Friedman, Silverman (1989), Friedman (1991), Breiman (1993), and Stone et al. (1997).

The second approach is called the roughness penalty approach which uses a large number of knots and a penalty to control the smoothness of the fit. The method of smoothing splines (Wahba, 1990; Eubank, 1999; Green, Silverman, 1994) has a knot placed at each unique data point of x with a penalty on the integral of the square of a specified derivative. Compared to smoothing splines, the method of penalized splines is similar but with fewer knots and

a somewhat more general penalty term. In particular, Eilers, Marx (1996) proposed to regression on a basis of B-splines and use a difference penalty on the regression coefficients; Ruppert, Carroll (2000) proposed to smooth based on truncated power functions and use a ridge penalty.

1.5 The Commercial Truck Driver Naturalistic Driving Study

The application examples in the dissertation use data collected from the Commercial Truck Driver NDS (Blanco et al., 2011), a large-scale data collection study conducted from November 2005 to March 2007 at four for-hire trucking companies in the United States. The study was directed at investigating issues on the commercial motor vehicle (CMV) hours-of-service regulations.

The study recruited a total of 100 CMV drivers. However, 3 drivers were removed from the analyses due to missing data. 96 of the rest drivers provided demographic information. Among these 96 drivers, 91 were male and 5 were female; 75 were long-haul (on the road for an extended period of time) and 21 were line-haul (out for one day/night); the average age was 44 years old (range: 21 to 73 years old); the average years of experience in driving CMVs was 9.13 (range: 4 weeks to 54 years).

The study used a naturalistic-data-collection approach to collect the driving data. Each participant was to drive an instrumented company truck during the normal revenue-producing runs for approximately four weeks. The trucks were installed with unobtrusive data-collection equipment, including a data acquisition system, sensors to measure driver performance, and video cameras to record the driver's face, the steering wheel, and three views outside of the truck. Figure 1.1 shows the views from the installed video camera. The time-stamped video and sensor data were collected continuously when the vehicle was on and in motion. The final project data set had approximately 735,000 miles of driving data.

To collect data on drivers' activities, the study asked each participant to report on- and

Figure 1.1: The Commercial Truck Driver NDS: Photo for five views from the installed video cameras (Blanco et al., 2011).

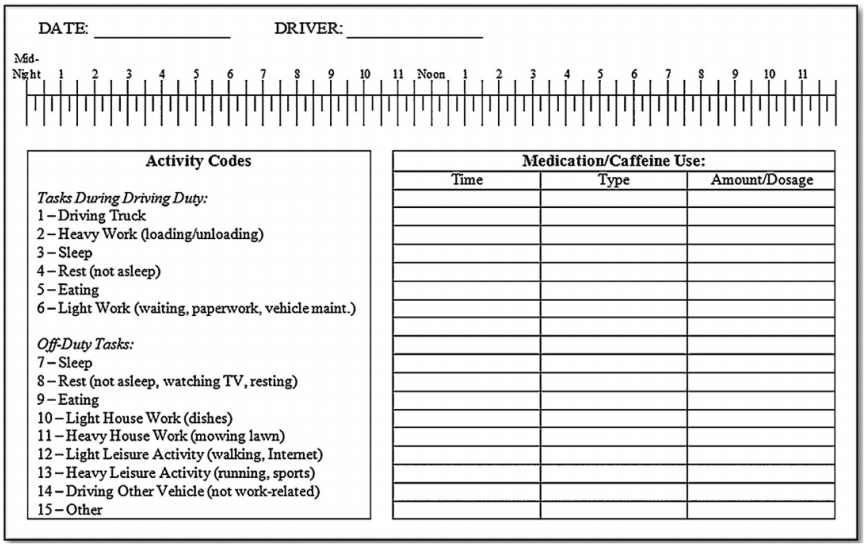


Figure 1.2: The Commercial Truck Driver NDS: A sample page of the daily activity register (Blanco et al., 2011).

off-duty activities using an activity register shown in Figure 1.2. A daily activity register included a 24-hour time line scaled with 15-minute increments. As the participants went about their days, they were to record their activities using the 15 activity codes (listed in Figure 1.2) and to mark the duration for each activity on the time line.

To establish a comprehensive time line of the drivers' activities, the driving data and the activity data were combined into a single data set. This hybrid activity data included both driving and non-driving activities for both on-duty and off-duty periods, and was used for all analyses of the 8-Truck study.

The study identified safety critical events (SCEs) to assess driver performance and risk. Computer softwares were used to filter the collected driving data by identifying driving episodes that potentially contained SCEs. These video episodes were then inspected by data reductionists to determine whether any valid events occurred. The study identified four types of SCEs: crashes, near-crashes, crash-relevant conflicts and unintentional lane deviations.

1.6 Overview

The remainder of this dissertation is described as follows. Chapter 2 proposes a mixed Poisson process model with time-varying coefficients for single-type recurrent event data and develops the penalized likelihood approach conditional the pre-specified smoothing parameters. Chapter 3 proposes the same model as in Chapter 2 but develops a mixed-model representation approach to estimate the smoothing parameters as variance components. Chapter 4 develops a Bayesian joint model with time-varying coefficients for multi-type recurrent event data and designs MCMC algorithms for posterior inference. Chapter 5 provides summary and discussion.

Chapter 2 A Time-Varying Coefficient Model Based on the Penalized Likelihood Approach

2.1 Introduction

Recurrent event data are commonly observed in clinical trials and epidemiological studies in which individuals can experience repeated events of interest over the study period. Examples include multiple occurrences of bladder tumors for patients in bladder cancer trials (Byar et al., 1977), multiple mammary tumors for animals in carcinogenicity experiments (Gail et al., 1980), and recurrent pulmonary exacerbations for patients with cystic fibrosis (Therneau, Hamilton, 1997). One typical feature of recurrent events is that event times within the same subject are correlated. Common approaches for recurrent event data analysis include marginal models based on estimating functions (e.g., Pepe, Cai, 1993; Lawless, Nadeau, 1995; Lin et al., 2000) and conditional models based on frailty distributions (e.g., Lawless, 1987; Nielsen et al., 1992; Vaida, Xu, 2000). Methods using estimating functions are to directly model marginal means and treat the dependence structure of events as nuisance parameters. In contrast, conditional models use frailty distributions to characterize the heterogeneity and correlation among events, and allow both marginal and conditional inferences.

The effect of a covariate in recurrent event modeling may not be constant over time and could be of interest. For the Commercial Truck Driver Naturalistic Driving Study (see Section 1.5 for details), the temporal profile of pre-shift sleep time affecting subsequent driving performance is of key interest for fatigue management and hours-of-service regulations. Varying coefficient models for longitudinal data include kernel-based approaches (e.g., Carroll et al., 1998; Cai et al., 2000) and spline-based approaches (e.g., Hastie, Tibshirani, 1993). Kernel

estimators are constructed through a weighted mean of the nearby observations and have great intuitive appeal. However, the kernel weights are asymmetric at boundaries of the predictor space and could cause substantial bias for the estimate (Hastie, Loader, 1993). Spline estimators are based on linear combinations of piecewise polynomial functions and have a straightforward extension of linear regression models. The computations are relatively fast and can be easily implemented (Eilers, Marx, 1996). Varying coefficient models have also been studied in the recurrent event data setting (Amorim et al., 2008; Sun et al., 2011; Yu et al., 2013). However, they are based on the partial likelihood approach (Cox, 1975), in which the baseline intensity is considered as a nuisance parameter and not explicitly estimated. For the Commercial Truck Driver Naturalistic Driving Study, driving performance is likely to change over a long shift due to fatigue and other factors. Therefore, smooth estimation of the baseline intensity needs to be incorporated to characterize the time-changing pattern of driving performance.

This chapter develops a mixed Poisson process model with time-varying coefficients for recurrent event data. The correlation among event times is characterized by using gamma frailties. The approach of penalized B-splines (Eilers, Marx, 1996) is used for modeling both the log baseline intensity and time-varying coefficients. The trapezoid method (Kauermann, 2005) is adopted for numerical integration in the likelihood function. An EM algorithm is developed for parameter estimation.

The rest of this chapter is arranged as follows. In Section 2.2, a mixed Poisson process model with time-varying coefficients is introduced and the approach of penalized B-splines is described. The EM algorithm is developed in Section 2.3. The asymptotic distribution for the estimator is derived in Section 2.4. Summary and discussion are presented in Section 2.5.

2.2 Model

Suppose there are m event processes observed in the data. For the ith process $\{N_i(t), t \geq 0\}$, n_i events ($n_i \geq 0$) are observed at times $0 < t_{i1} < \cdots < t_{i,n_i}$ before being censored at time c_i. The independent censoring scheme is assumed here. Let $\boldsymbol{x_i'} = (x_{i1}, \ldots, x_{ip})$ denote a p-dimensional covariate vector with constant coefficients, and $\boldsymbol{z_i'} = (z_{i0}, z_{i1}, \ldots, z_{iq})$ denote a $(q+1)$-dimensional covariate vector with time-varying coefficients, where $z_{i0} = 1$ for all i. The intensity function of the ith event process is formulated as

$$\lambda_i(t \,|\, \boldsymbol{x_i}, \boldsymbol{z_i}, u_i) = u_i \, \exp\left\{\boldsymbol{x_i'}\,\boldsymbol{\alpha} + \boldsymbol{z_i'}\,\boldsymbol{\beta}(t)\right\}. \tag{2.1}$$

$\boldsymbol{\beta}(t) = (\beta_0(t), \beta_1(t), \ldots, \beta_q(t))'$ is a vector of unknown smooth functions of time t. Particularly, the baseline intensity is expressed as $\exp \beta_0(t)$. The u_i's are independent and identically distributed gamma variables with density

$$p(u; \phi) = \frac{u^{1/\phi - 1} \exp\{-u/\phi\}}{\Gamma(1/\phi)\phi^{1/\phi}}, \quad \phi > 0. \tag{2.2}$$

With such distribution for the frailty, the mean of u_i is 1 and the variance is ϕ. Large values of ϕ indicate great variation between processes and strong association of event times within a process. Conditional on u_i and covariates, (2.1) characterizes a non-homogeneous Poisson process. This model extends the Cox proportional hazards model (Cox, 1972) in two ways: covariate effects are allowed to change with time by assuming time-varying coefficients; intra-process correlation and inter-process variation that cannot be accounted for by covariates are included in the frailty.

Let $\boldsymbol{t_i} = [t_{ij}]_{1 \leq j \leq n_i+1}$ denote a vector of the time to event observed in the ith process, where $t_{i,n_i+1} = c_i$, and let $\boldsymbol{\delta_i} = [\delta_{ij}]_{1 \leq j \leq n_i+1}$ denote the censoring indicator for $\boldsymbol{t_i}$. Conditional on

u_i, the log density function of the observed data $\boldsymbol{D}_{i,\text{obs}} = (\boldsymbol{t_i}, \boldsymbol{\delta_i})$ would be

$$\log f(\boldsymbol{D}_{i,\text{obs}}|u_i) = \sum_{j=1}^{n_i} \Big[\log u_i + \boldsymbol{x_i'\alpha} + \boldsymbol{z_i'\beta}(t_{ij}) \Big] - u_i \int_0^{c_i} \exp\big\{ \boldsymbol{x_i'\alpha} + \boldsymbol{z_i'\beta}(t) \big\} \mathrm{d}t. \tag{2.3}$$

Therefore the log likelihood given the complete data $\{(\boldsymbol{D}_{i,\text{obs}}, u_i), i = 1, \ldots, m\}$ is written as

$$l_\text{c} = \sum_{i=1}^{m} \log f(\boldsymbol{D}_{i,\text{obs}}|u_i) + \sum_{i=1}^{m} \log p(u_i). \tag{2.4}$$

By integrating the u_i's out, the log marginal likelihood is obtained as

$$\begin{aligned} l_\text{m} = & \sum_{i=1}^{m} \sum_{j=1}^{n_i} \Big[\boldsymbol{x_i'\alpha} + \boldsymbol{z_i'\beta}(t_{ij}) \Big] \\ & - \sum_{i=1}^{m} \left(n_i + \frac{1}{\phi} \right) \log \left[1 + \phi \int_0^{c_i} \exp\big\{ \boldsymbol{x_i'\alpha} + \boldsymbol{z_i'\beta}(t) \big\} \, \mathrm{d}t \right] \\ & + \sum_{i=1}^{m} \sum_{j=0}^{n_i^*} \log(1 + j\phi), \end{aligned} \tag{2.5}$$

where $n_i^* = \max(0, n_i - 1)$.

Smooth estimation of time-varying coefficients is achieved by using penalized B-splines (Eilers, Marx, 1996). The approach assumes that for $l = 0, 1, \ldots, q$, $\beta_l(t)$ can be approximated by a polynomial spline of degree v with equally spaced knots

$$t_{\min} = \zeta_{l0} < \zeta_{l1} < \cdots < \zeta_{l,k_l-1} < \zeta_{lk_l} = t_{\max} \tag{2.6}$$

chosen within the domain of time t. The spline can be written in terms of a linear combination of $K_l = k_l + v$ B-spline basis functions $\boldsymbol{B_l'}(t) = (B_{l1}(t), \ldots, B_{lK_l}(t))$, that is,

$$\beta_l(t) = \sum_{k=1}^{K_l} \beta_{lk} B_{lk}(t) = \boldsymbol{B_l'}(t)\boldsymbol{\beta_l}, \tag{2.7}$$

with $\boldsymbol{\beta_l} = (\beta_{l1}, \ldots, \beta_{lK_l})'$ as a vector of spline coefficients for modeling $\beta_l(t)$. The grid of

knots should be relatively dense to allow for detailed structure in $\beta_l(t)$ to be estimated. Let $\boldsymbol{\beta}' = (\boldsymbol{\beta}_{\mathbf{0}}', \ldots, \boldsymbol{\beta}_{\boldsymbol{q}}')$ denote all the spline coefficients for modeling unknown functions in $\boldsymbol{\beta}(t)$, and $\boldsymbol{w}_{\boldsymbol{i}}'(t) = (z_{i0}\, \boldsymbol{B}_{\mathbf{0}}'(t), \ldots, z_{iq}\, \boldsymbol{B}_{\boldsymbol{q}}'(t))$, (2.1) can be rewritten as

$$\lambda_i(t \,|\, \boldsymbol{x_i}, \boldsymbol{z_i}, u_i) = u_i \, \exp\left\{\boldsymbol{x}_{\boldsymbol{i}}' \, \boldsymbol{\alpha} + \boldsymbol{w}_{\boldsymbol{i}}'(t) \, \boldsymbol{\beta}\right\},$$

and the set of unknown parameters in the model becomes $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \phi)$.

If $\boldsymbol{\beta_l}$ is treated as ordinary parameter and estimated via maximizing the likelihood function, the resulting estimate of $\beta_l(t)$ would be a somewhat wiggly piecewise polynomial function. The approach of penalized splines uses a penalty to restrict the flexibility of the fit. O'Sullivan (1986) introduced a penalty on the second derivative of the fitted curve. Eilers, Marx (1996) proposed a more generalized and simpler penalty based on finite differences of adjacent spline coefficients. Let Δ^r denote the rth difference operator. For instance, $\Delta\beta_{lj} = \beta_{lj} - \beta_{l,j-1}$, and $\Delta^2\beta_{lj} = \Delta(\Delta\beta_{lj}) = \beta_{lj} - 2\beta_{l,j-1} + \beta_{l,j-2}$. The rth difference penalty for modeling $\beta_l(t)$ is

$$\frac{1}{2}\,\lambda_l \sum_{j=r+1}^{K} \left(\Delta^r \beta_{lj}\right)^2 = \frac{1}{2}\,\lambda_l\, \boldsymbol{\beta_l'}(\boldsymbol{D_l'D_l})\boldsymbol{\beta_l}, \tag{2.8}$$

where λ_l, often known as the smoothing parameter, is a non-negative constant which can be adjusted to control the smoothness of the fit. For $\lambda_l = 0$, it reduces to ordinary likelihood maximization, resulting in a overfitted $\beta_l(t)$. By increasing λ_l, the smoothness can be tuned. In the limit of a large λ_l, a polynomial fit of degree $r-1$ will be obtained. $\boldsymbol{D_l}$ is a difference matrix of $(K_l - r) \times K_l$, such that $\boldsymbol{D_l\beta_l} = \Delta^r\boldsymbol{\beta_l}$ constructs the vector of the rth differences of $\boldsymbol{\beta_l}$. For instance, when $r = 2$, $\boldsymbol{D_l}$ is a second-order difference matrix, that is,

$$\boldsymbol{D_l} = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & -2 & 1 \end{bmatrix}.$$

The penalties are to be subtracted from the log marginal likelihood (2.5) to form the penalized likelihood function

$$pl_{\mathrm{m}} = l_{\mathrm{m}} - \frac{1}{2}\sum_{l=0}^{q} \lambda_l \, \boldsymbol{\beta}_{\boldsymbol{l}}'(\boldsymbol{D}_{\boldsymbol{l}}'\boldsymbol{D}_{\boldsymbol{l}})\boldsymbol{\beta}_{\boldsymbol{l}}. \tag{2.9}$$

Conditional on the λ_l's, the parameter can be estimated by maximizing (2.9).

2.2.1 Numerical Integration

The likelihood functions (2.4) and (2.5) involve integrals based on the intensity function. Since there's no analytic solution, numerical integration is implemented. A computationally efficient technique is to approximate the integrals by the trapezoid rule.

Let $\rho_i(t) = \exp\left\{\boldsymbol{x}_{\boldsymbol{i}}'\boldsymbol{\alpha} + \boldsymbol{z}_{\boldsymbol{i}}'\boldsymbol{\beta}(t)\right\}$ and $\mu_i(c_i) = \int_0^{c_i} \rho_i(t)\,\mathrm{d}t$. To approximate $\mu_i(c_i)$ for $i = 1, \ldots, m$, consider a grid of equally spaced points spanning the range of the c_i's:

$$0 = s_0 < s_1 < \cdots < s_R = \max\{c_i, i = 1, \ldots, m\}.$$

Define index R_i such that $s_{R_i-1} < c_i \leq s_{R_i}$. The integral $\mu_i(c_i)$ is approximated by a polygon going through the points $(s_r,\ \rho_i(s_r))$ for $r = 0, 1, \ldots, R_i - 1$, that is,

$$\begin{aligned}\mu_i(c_i) &\approx (c_i - s_{R_i-1})\,\frac{\rho_i(s_{R_i}) + \rho_i(s_{R_i-1})}{2} + I(R_i > 1)\sum_{r=1}^{R_i-1}(s_r - s_{r-1})\,\frac{\rho_i(s_r) + \rho_i(s_{r-1})}{2} \\ &= \min(c_i,\ s_1)\,\rho_i(s_0)/2 + \sum_{r=1}^{R_i}[\min(c_i,\ s_{r+1}) - \min(c_i,\ s_{r-1})]\,\rho_i(s_r)/2, \end{aligned} \tag{2.10}$$

with $I(\cdot)$ denoting the indicator function.

Consider the following notations

$$\boldsymbol{s_a} = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ \vdots \\ s_R \\ s_R \end{bmatrix}, \qquad \boldsymbol{s_b} = \begin{bmatrix} s_0 \\ s_0 \\ s_1 \\ \vdots \\ s_{R-2} \\ s_{R-1} \end{bmatrix}, \qquad \boldsymbol{\rho_i} = \begin{bmatrix} \rho_i(s_0) \\ \rho_i(s_1) \\ \rho_i(s_2) \\ \vdots \\ \rho_i(s_{R-1}) \\ \rho_i(s_R) \end{bmatrix}.$$

Then the approximation in (2.10) can be rewritten as

$$\mu_i(c_i) \approx [\min(c_i, \boldsymbol{s_a}) - \min(c_i, \boldsymbol{s_b})]' \boldsymbol{\rho_i}/2, \tag{2.11}$$

where $\min(c, \boldsymbol{s})$, with c as a number and $\boldsymbol{s}$ as a vector, is a vector where each element is the minimum between c and the corresponding element in $\boldsymbol{s}$.

2.3 EM Algorithm

Since direct maximization of the penalized marginal likelihood is difficult, the estimation procedure proceeds by using a EM algorithm (Dempster et al., 1977; Klein, 1992).

Consider the following notations:

$$\begin{aligned} \boldsymbol{\delta} &= [\boldsymbol{\delta_i}]_{1\le i\le m}, \\ \boldsymbol{X_i} &= [\boldsymbol{x_i'}]_{1\le j\le n_i+1}, & \boldsymbol{X} &= \left[\boldsymbol{X_i}\right]_{1\le i\le m}, \\ \boldsymbol{W_i} &= [\boldsymbol{w_i'}(t_{ij})]_{1\le j\le n_i+1}, & \boldsymbol{W} &= \left[\boldsymbol{W_i}\right]_{1\le i\le m}. \end{aligned}$$

The penalized log likelihood given the complete data $\{(\boldsymbol{D}_{i,\text{obs}}, u_i), i = 1, \ldots, m\}$ is, up to a

term free of the parameter values,

$$pl_{\mathrm{c}} = l_{\mathrm{c}} - \frac{1}{2}\sum_{l=0}^{q} \lambda_l \, \boldsymbol{\beta}_{\boldsymbol{l}}'(\boldsymbol{D}_{\boldsymbol{l}}'\boldsymbol{D}_{\boldsymbol{l}})\boldsymbol{\beta}_{\boldsymbol{l}} = l_1(\boldsymbol{\alpha}, \boldsymbol{\beta}) + l_2(\phi),$$

where

$$l_1(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \boldsymbol{\delta}'(\boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{W}\boldsymbol{\beta}) - \sum_{i=1}^{m} u_i \mu_i(c_i) - \frac{1}{2}\boldsymbol{\beta}'\boldsymbol{P}\boldsymbol{\beta}, \tag{2.12}$$

and

$$l_2(\phi) = \frac{1}{\phi}\sum_{i=1}^{m}\left(\log u_i - u_i\right) + m\left[\frac{1}{\phi}\log\left(\frac{1}{\phi}\right) - \log\Gamma\left(\frac{1}{\phi}\right)\right]. \tag{2.13}$$

By inheriting the notations in Section 2.2.1, $\sum_{i=1}^{m} u_i \mu_i(c_i)$ is the sum of the cumulative intensities evaluated at the c_i's. $\boldsymbol{P}$ is a block diagonal matrix with blocks $\lambda_l(\boldsymbol{D}_{\boldsymbol{l}}'\boldsymbol{D}_{\boldsymbol{l}})$ in the terms corresponding to the $\boldsymbol{\beta}_{\boldsymbol{l}}$ and 0's elsewhere.

The estimating algorithm proceeds by first making an initial guess at the parameter values. To apply the E-step of the algorithm, notice that, conditional on the observed data $\boldsymbol{D}_{i,\mathrm{obs}}$, the u_i's are independent gamma variables with shape parameters $A_i = n_i + 1/\phi$ and rate parameters $B_i = \mu_i(c_i) + 1/\phi$. The resulting expectation of pl_{c} given the observed data and the current estimates of A_i and B_i can be written as

$$E(pl_{\mathrm{c}}) = Q_1(\boldsymbol{\alpha}, \boldsymbol{\beta}) + Q_2(\phi),$$

where

$$Q_1(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \boldsymbol{\delta}'(\boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{W}\boldsymbol{\beta}) - \sum_{i=1}^{m}\left(\frac{A_i}{B_i}\right)\mu_i(c_i) - \frac{1}{2}\boldsymbol{\beta}'\boldsymbol{P}\boldsymbol{\beta}, \tag{2.14}$$

and

$$Q_2(\phi) = \frac{1}{\phi}\sum_{i=1}^{m}\left(\psi(A_i) - \log(B_i) - A_i/B_i\right) + m\left[\frac{1}{\phi}\log\left(\frac{1}{\phi}\right) - \log\Gamma\left(\frac{1}{\phi}\right)\right], \tag{2.15}$$

with $\psi(\cdot)$ as the digamma function.

The M-step of the algorithm requires separate maximization of (2.14) and (2.15). The updated estimate of ϕ involves maximizing (2.15) numerically. To update the estimate of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, the score equation for Q_1 is

$$\begin{bmatrix} \boldsymbol{X}' \\ \boldsymbol{W}' \end{bmatrix} \boldsymbol{\delta} - \boldsymbol{S}\Lambda - \begin{bmatrix} \mathbf{0} \\ \boldsymbol{P}\boldsymbol{\beta} \end{bmatrix} = \mathbf{0}, \tag{2.16}$$

where $\Lambda(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{i=1}^{m}(A_i/B_i)\mu_i(c_i)$ and $\boldsymbol{S}\Lambda$ is the first derivative of $\Lambda(\boldsymbol{\alpha}, \boldsymbol{\beta})$, that is,

$$\boldsymbol{S}\Lambda = \sum_{i=1}^{m}\left(\frac{A_i}{B_i}\right)\int_0^{c_i}\begin{bmatrix} \boldsymbol{x_i} \\ \boldsymbol{w_i}(t) \end{bmatrix}\exp\left\{\boldsymbol{x_i}'\boldsymbol{\alpha} + \boldsymbol{w_i}'(t)\boldsymbol{\beta}\right\}\mathrm{d}t. \tag{2.17}$$

The Newton-Raphson method is implemented to find the solution to (2.16). The Hessian matrix of Q_1 is

$$-\Big(\boldsymbol{H}\Lambda + \begin{bmatrix} \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \boldsymbol{P} \end{bmatrix}\Big), \tag{2.18}$$

where $\boldsymbol{H}\Lambda$ is the Hessian matrix of $\Lambda(\boldsymbol{\alpha}, \boldsymbol{\beta})$, that is,

$$\boldsymbol{H}\Lambda = \sum_{i=1}^{m}\left(\frac{A_i}{B_i}\right)\int_0^{c_i}\begin{bmatrix} \boldsymbol{x_i} \\ \boldsymbol{w_i}(t) \end{bmatrix}\begin{bmatrix} \boldsymbol{x_i}' & \boldsymbol{w_i}'(t) \end{bmatrix}\exp\left\{\boldsymbol{x_i}'\boldsymbol{\alpha} + \boldsymbol{w_i}'(t)\boldsymbol{\beta}\right\}\mathrm{d}t. \tag{2.19}$$

The iteration between the E-step and M-step continues until the difference in estimates at successive iterations drops below a desired tolerance.

2.4 Asymptotic Distribution

To approximate the distribution of maximum penalized likelihood estimator $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\phi})$, the knot locations and the number of parameters are assumed to be fixed as the sample size m increases, and that the family of models incorporates the true distribution. Writing $\boldsymbol{\theta}$ for the full set of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and ϕ parameters, the penalized log marginal likelihood in (2.9) can be rewritten as

$$pl_{\mathrm{m}}(\boldsymbol{\theta}) = l_{\mathrm{m}}(\boldsymbol{\theta}) - \frac{1}{2}\boldsymbol{\theta}'\tilde{\boldsymbol{P}}\boldsymbol{\theta}, \tag{2.20}$$

where $l_{\mathrm{m}}(\boldsymbol{\theta})$ is the log marginal likelihood defined in (2.5), and $\tilde{\boldsymbol{P}}$ is a block diagonal matrix with blocks $\lambda_l(\boldsymbol{D}_l'\boldsymbol{D}_l)$ in the terms corresponding to the $\boldsymbol{\beta}_l$ parameters and 0's elsewhere. With the sample size m increasing while $\boldsymbol{D}_l'\boldsymbol{D}_l$ fixed, the magnitude of l_{m}'s contribution to (2.20) also increases. To maintain the same amount of smoothing in the fit of $\beta_l(t)$, λ_l needs to be increased at a rate of $O(m)$ as well (see Gray, 1992).

Consider the following notations

$$\begin{aligned} \boldsymbol{S}_p(\boldsymbol{\theta}) &= \partial pl_{\mathrm{m}}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}, & \boldsymbol{S}(\boldsymbol{\theta}) &= \partial l_{\mathrm{m}}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}, \\ \boldsymbol{I}_p(\boldsymbol{\theta}) &= -\partial^2 pl_{\mathrm{m}}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^2, & \boldsymbol{I}(\boldsymbol{\theta}) &= -\partial^2 l_{\mathrm{m}}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^2. \end{aligned}$$

Notice that $\boldsymbol{S}_p(\boldsymbol{\theta}) = \boldsymbol{S}(\boldsymbol{\theta}) - \tilde{\boldsymbol{P}}\boldsymbol{\theta}$, and $\boldsymbol{I}_p(\boldsymbol{\theta}) = \boldsymbol{I}(\boldsymbol{\theta}) + \tilde{\boldsymbol{P}}$. Let $\boldsymbol{\theta}_0$ denote the true parameter values. Using standard arguments based on first-order Taylor expansion and the fact that $\hat{\boldsymbol{\theta}}$ is defined through $\boldsymbol{S}_p(\hat{\boldsymbol{\theta}}) = \boldsymbol{0}$,

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = \boldsymbol{I}_p^{-1}(\boldsymbol{\theta}_0)\,\boldsymbol{S}_p(\boldsymbol{\theta}_0) = \boldsymbol{I}_p^{-1}(\boldsymbol{\theta}_0)\left[\boldsymbol{S}(\boldsymbol{\theta}_0) - \tilde{\boldsymbol{P}}\boldsymbol{\theta}_0\right].$$

Therefore the bias and variance of $\hat{\boldsymbol{\theta}}$ are:

$$\begin{aligned}\text{bias}(\hat{\boldsymbol{\theta}}) &= -\boldsymbol{I}_p^{-1}(\boldsymbol{\theta}_0)\,\tilde{\boldsymbol{P}}\,\boldsymbol{\theta}_0,\\ \text{Var}(\hat{\boldsymbol{\theta}}) &= \boldsymbol{I}_p^{-1}(\boldsymbol{\theta}_0)\,\text{Var}[\boldsymbol{S}(\boldsymbol{\theta}_0)]\,\boldsymbol{I}_p^{-1}(\boldsymbol{\theta}_0).\end{aligned}$$

Gray (1992) discussed two special cases where the bias is negligible. The first case is when $\tilde{\boldsymbol{P}}\boldsymbol{\theta}_0 = \mathbf{0}$, that is, the penalty does not induce bias in the estimates. The second is when $\tilde{\boldsymbol{P}}\boldsymbol{\theta}_0$ is not necessarily $\mathbf{0}$ but the amount of smoothing decreases as $m \to \infty$. Since $\lambda_l^{(m)}$ must be $O(m)$ to maintain the same degree of smoothing, any smaller power will result in a decrease. Let $\lambda_l^{(m)} = O(m^a)$ with $0 < a < 1$, then $\text{bias}(\hat{\boldsymbol{\theta}})$ goes to $\mathbf{0}$ at a rate of $O(m^{a-1})$. Under either case, it can be shown that if $\lambda_l^{(m)}$ is replaced by the estimate $\hat{\lambda}_l$ such that $\hat{\lambda}_l/\lambda_l^{(m)} \to 1$ in probability, $\sqrt{m}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ follows asymptotically a normal distribution with mean $\mathbf{0}$ and covariance matrix $\lim m\text{Var}(\hat{\boldsymbol{\theta}})$, where $\text{Var}(\hat{\boldsymbol{\theta}})$ can be estimated by

$$\boldsymbol{V}(\hat{\boldsymbol{\theta}}) = \left[\boldsymbol{I}(\hat{\boldsymbol{\theta}}) + \tilde{\boldsymbol{P}}\right]^{-1} \boldsymbol{I}(\hat{\boldsymbol{\theta}}) \left[\boldsymbol{I}(\hat{\boldsymbol{\theta}}) + \tilde{\boldsymbol{P}}\right]^{-1}. \tag{2.21}$$

To obtain confidence intervals for the estimated intensity function with covariates $\boldsymbol{x_i}$ and $\boldsymbol{z_i}$,

$$\hat{\lambda}(t|\boldsymbol{x_i}, \boldsymbol{z_i}) = \exp\left\{\boldsymbol{x_i'}\,\hat{\boldsymbol{\alpha}} + \boldsymbol{w_i'}(t)\,\hat{\boldsymbol{\beta}}\right\},$$

the variance of $\log\hat{\lambda}(t|\boldsymbol{x_i}, \boldsymbol{z_i})$ is estimated based on (2.21):

$$\widehat{\text{Var}}\left[\log\hat{\lambda}(t|\boldsymbol{x_i}, \boldsymbol{z_i})\right] = \begin{bmatrix}\boldsymbol{x_i'} & \boldsymbol{w_i'}(t)\end{bmatrix} \boldsymbol{V}\left(\begin{bmatrix}\hat{\boldsymbol{\alpha}}\\ \hat{\boldsymbol{\beta}}\end{bmatrix}\right) \begin{bmatrix}\boldsymbol{x_i}\\ \boldsymbol{w_i}(t)\end{bmatrix}.$$

It is important to note that the asymptotic distribution relies on the assumption that the time-varying coefficients are approximated well with keeping $\boldsymbol{\theta}$ at a fixed dimension. Since the B-spline basis is taken to be high-dimensional, one can argue heuristically that the approximation bias should be negligible. Therefore the asymptotic results will be used here.

2.4.1 Test Statistics

This section considers a general linear null hypothesis of the form: $\boldsymbol{C\eta} = \mathbf{0}$, where $\boldsymbol{\eta}' = (\boldsymbol{\alpha}', \boldsymbol{\beta}')$, and $\boldsymbol{C}$ is a matrix of full row rank. Gray (1992) suggested a Wald version test with the test statistic

$$(\boldsymbol{C}\,\hat{\boldsymbol{\eta}})' \left(\boldsymbol{C}\left[\boldsymbol{I}_p^{-1}(\hat{\boldsymbol{\theta}})\right]_\eta \boldsymbol{C}'\right)^{-1} \boldsymbol{C}\,\hat{\boldsymbol{\eta}}, \tag{2.22}$$

where $[\boldsymbol{I}_p^{-1}(\hat{\boldsymbol{\theta}})]_\eta$ is the block in $\boldsymbol{I}_p^{-1}(\hat{\boldsymbol{\theta}})$ corresponding to the $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ parameter. Under the null hypothesis, (2.22) asymptotically follows the distribution of a linear combination of independent Chi-squared random variables with one degree of freedom, that is,

$$\sum e_i\, \chi^2(1),$$

where the e_i's are the eigenvalues of

$$\left(\boldsymbol{C}\left[\boldsymbol{I}_p^{-1}(\hat{\boldsymbol{\theta}})\right]_\eta \boldsymbol{C}'\right)^{-1} \boldsymbol{C}\left[\boldsymbol{V}(\hat{\boldsymbol{\theta}})\right]_\eta \boldsymbol{C}',$$

with $\left[\boldsymbol{V}(\hat{\boldsymbol{\theta}})\right]_\eta$ being the block in $\boldsymbol{V}(\hat{\boldsymbol{\theta}})$ corresponding to the $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ parameter.

Three types of hypothesis are of interest. The first one is the test of no effect in the lth covariate of $\boldsymbol{z}$, that is, $H_0 : \beta_l(t) = 0$. Since $\beta_l(t)$ is modeled by B-splines as $\beta_l(t) = \boldsymbol{B_l'}(t)\boldsymbol{\beta_l}$, the null and the alternative hypothesis are equivalently

$$H_0 : \boldsymbol{\beta_l} = \mathbf{0} \quad \text{vs} \quad H_1 : \boldsymbol{\beta_l} \neq \mathbf{0}.$$

The second one is the test of no time-varying effect in the lth covariate of $\boldsymbol{z}$, that is, $H_0 : \mathrm{d}\beta_l(t)/\mathrm{d}t = 0$. For any B-spline basis of degree v, $\{B_{lk}(t; v), k = 1, \ldots, K_l\}$, De Boor (1978)

gives a simple formula for derivatives of B-splines:

$$h\,\frac{\mathrm{d}}{\mathrm{d}t}\sum_{k=1}^{K_l}\beta_{lk}B_{lk}(t;v) = -\sum_{k=1}^{K_l-1}\Delta\beta_{l,k+1}B_{lk}(t;v-1), \tag{2.23}$$

where h denotes the distance between knots. Therefore in the test of no time-varying effect, the null and the alternative hypothesis are equivalently

$$H_0 : \boldsymbol{C_l\beta_l} = \mathbf{0} \quad \text{vs} \quad H_1 : \boldsymbol{C_l\beta_l} \neq \mathbf{0},$$

where $\boldsymbol{C_l}$ is a first-order difference matrix of $(K_l - 1) \times K_l$. The third one is the test of a linear time-varying effect in the lth covariate of $\boldsymbol{z}$, that is, $H_0 : \mathrm{d}^2\beta_l(t)/\mathrm{d}t^2 = 0$. According to (2.23), the null and the alternative hypothesis can be written as

$$H_0 : \boldsymbol{C_l\beta_l} = \mathbf{0} \quad \text{vs} \quad H_1 : \boldsymbol{C_l\beta_l} \neq \mathbf{0},$$

where $\boldsymbol{C_l}$ is a second-order difference matrix of $(K_l - 2) \times K_l$.

2.5 Summary and Discussion

This chapter develops a mixed Poisson process model with time-varying coefficients for recurrent event data. I use gamma frailties to account for the correlation among event times. The approach of penalized B-splines is adopted for modeling both the log baseline intensity and time-varying coefficients. The trapezoid method is adopted for numerical integration in the likelihood function. An EM algorithm is developed for parameter estimation.

It is important to note that the estimating procedure is conditional on the values of smoothing parameters. Typically for estimating the smoothing parameter , one first needs to determine a grid of potential values. After obtaining parameter estimates at each grid value, the "optimal" smoothing parameter is found by optimizing a performance criterion, such

as cross-validation or an information criterion. The whole process becomes computationally demanding when the number of time-varying coefficients grows, as it yields a high dimensional grid from which the combination of smoothing parameters are to be selected. In addition, these procedures often fail in practice because no optimal solution can be found (Lang, Brezger, 2004). To address the issue, the next chapter develops a mixed-model representation approach based on the connection between penalized likelihood and random effect models (Brumback, Rice, 1998; Wang, 1998a,b; Brumback et al., 1999). Spline coefficients could be treated as random effects and smoothing parameters are estimated as variance components.

Chapter 3 A Time-Varying Coefficient Model Based on the Mixed-Model Representation Approach

3.1 Introduction

Traffic crashes are a major safety hazard for commercial truck drivers and are also a general public health concern. In 2015, heavy trucks were involved in 3,598 fatal crashes in the United States (U.S.), which led to 667 truck driver fatalities and 4,067 total fatalities (Federal Motor Carrier Safety Administration, 2016). In addition, 116,000 persons were injured due to heavy vehicle crashes. In the U.S., the operational hours for commercial truck drivers are regulated by the Federal Motor Carrier Safety Administration (FMCSA). Referred to as the hours-of-service (HOS) regulations, truck drivers can work 14 hours, 11 hours of which can be driving, before taking a 10-hour off-duty break (Federal Register, 2011). Though not all drivers may work the fully allocated 14 hours, or drive the allocated 11 hours, for drivers that do push the work/driving limits (e.g., long-haul truck drivers), fatigue is a major risk factor (National Academies of Sciences, Engineering, and Medicine, 2016). Fatigue issues associated with time-on-task and deterioration of driving performance over the course of a long work day or drive are of particular concern for this cohort. Though the importance of being well-rested for optimal performance has been documented (e.g., Banks, Dinges, 2007), there has been limited research to evaluate the change in driving performance and safety risk over time as a function of pre-shift sleep time.

Many factors can affect driver fatigue, including total sleep time, sleep debt, circadian rhythm, breaks during long shifts, and caffeine consumption. Among these factors, sleep time the night before driving has been noted as a critical risk factor (Hanowski et al., 2007). Being well rested prior to a long shift has a direct impact on driver fatigue and driving

performance (Banks, Dinges, 2007; Lim, Dinges, 2008, 2010). Insufficient sleep is typically defined as sleep less than 7 hours per day (Van Dongen et al., 2003a,b; Ford et al., 2015; Watson et al., 2015a,b). Sleep debt, the cumulative lack of sleep for a period of time, also negatively affects driver fatigue (Abe et al., 2012; Scott et al., 2007). Furthermore, it has also been shown that breaks during a long shift reduce the number of on-road safety-critical events for truck drivers (Soccolich et al., 2013).

Recurrent event data are commonly observed in clinical trials and epidemiological studies in which individuals can experience repeated events of interest over the study period. Examples include multiple occurrences of bladder tumors for patients in bladder cancer trials (Byar et al., 1977), multiple mammary tumors for animals in carcinogenicity experiments (Gail et al., 1980), and recurrent pulmonary exacerbations for patients with cystic fibrosis (Therneau, Hamilton, 1997). One typical feature of recurrent events is that event times within the same subject are correlated. Common approaches for recurrent event data analysis include marginal models based on estimating functions (e.g., Pepe, Cai, 1993; Lawless, Nadeau, 1995; Lin et al., 2000) and conditional models based on frailty distributions (e.g., Lawless, 1987; Nielsen et al., 1992; Vaida, Xu, 2000). Methods using estimating functions are to directly model marginal means and treat the dependence structure of events as nuisance parameters. In contrast, conditional models use frailty distributions to characterize the heterogeneity and correlation among events, and allow both marginal and conditional inferences.

The effect of a covariate in recurrent event data analysis may not be constant over time. For this study, the temporal profile of pre-shift sleep time affecting subsequent driving performance is of key interest for fatigue management and HOS regulations. Varying coefficient models for longitudinal data include kernel-based approaches (e.g., Carroll et al., 1998; Cai et al., 2000) and spline-based approaches (e.g., Hastie, Tibshirani, 1993). Kernel estimators are constructed through a weighted mean of the nearby observations and have great intuitive appeal. However, the kernel weights are asymmetric at boundaries of the predictor space and could cause substantial bias for the estimate (Hastie, Loader, 1993). Spline estimators

are based on linear combinations of piecewise polynomial functions and have a straightforward extension of linear regression models. The computations are relatively fast and can be easily implemented (Eilers, Marx, 1996). Varying coefficient models have also been studied in the recurrent event data setting (Amorim et al., 2008; Sun et al., 2011; Yu et al., 2013). However, they are based on the partial likelihood approach (Cox, 1975), in which the baseline intensity is considered as a nuisance parameter and not explicitly estimated. For this study, driving performance is likely to change over a long shift due to fatigue and other factors. Therefore, smooth estimation of the baseline intensity needs to be incorporated to characterize the time-changing pattern of driving performance.

This chapter develops a mixed Poisson process model with time-varying coefficients for recurrent event data. The correlation among event times is characterized by using gamma frailties. The approach of penalized B-splines (Eilers, Marx, 1996) is used for modeling both the log baseline intensity and time-varying coefficients. For penalized splines, the fit is conditional on the smoothing parameter, which has to be selected by cross-validation or optimizing some performance criterion. These procedures could be computationally demanding as the number of time-varying coefficients increases. To address the issue, I adopt a mixed-model representation approach based on the connection between penalized likelihood and random effect models (Brumback, Rice, 1998; Wang, 1998a,b; Brumback et al., 1999). Spline coefficients are treated as random effects and smoothing parameters are estimated as variance components. An EM algorithm is used to fit the model, in which the penalized quasi-likelihood (PQL) approach (Breslow, Clayton, 1993) is adopted for likelihood approximation.

The rest of this chapter is arranged as follows. In Section 3.2, I describe the Commercial Truck Driver Naturalistic Driving Study (NDS) that is used for evaluating the temporal profile of pre-shift sleep time affecting subsequent driving performance. In Section 3.3, a mixed Poisson process model with time-varying coefficients is firstly introduced. Then the mixed-model representation of penalized splines is described. The model performance is evaluated by a simulation study presented in Section 3.4. The results from the model's application to the Commercial Truck Driver NDS data are shown in Section 3.5. Summary

and discussion are presented in Section 3.6.

3.2 Commercial Truck Driver Naturalistic Driving Study

The chapter uses data from the Commercial Truck Driver NDS (Blanco et al., 2011), a large-scale data collection study sponsored by the FMCSA to evaluate issues related to commercial motor vehicles. The study recruited 100 drivers from four for-hire trucking companies. A naturalistic-data-collection approach was used to collect the driving data. Each driver drove an instrumented truck for approximately four weeks. The trucks were fitted with unobtrusive data-collection equipment, including four video cameras, radar, a global positioning system device, and kinematic sensors. Driving data were continuously recorded from ignition-on to ignition-off at high frequency, for example, 10 Hz for video, radar, and three-dimensional accelerometers. The study also collected drivers' on-duty and off-duty activities through a daily activity register. The activity register requested data such as the beginning and ending times of sleep, break, driving and other activities. The NDS driving data and the activity register data were combined to provide a complete time-line of participants' driving and non-driving activities. The detailed driving data allow various driving behavior to be examined. For this study, unintentional lane deviation (ULD) is selected to measure driving performance. ULD is defined as "any circumstance where the subject vehicle crosses over a solid lane line (e.g., on to the shoulder) where there is not a hazard (guardrail, ditch, vehicle, etc.) present. In these cases, the driver does not signal or intentionally make a lane change maneuver" (reg, 2017, pp. 145). ULD has been shown to be a measure of driving performance decrement and sensitive to driving fatigue (Hanowski et al., 2008; Van Dongen et al., 2010).

To evaluate the impact of off-duty sleep time on drivers' subsequent on-duty driving performance, the data need to be organized into pairs of an off-duty period followed by an on-duty one. The daily schedules of commercial truck drivers are highly irregular. For example, some drivers opt to drive late at night to avoid traffic congestion. For the purpose of this

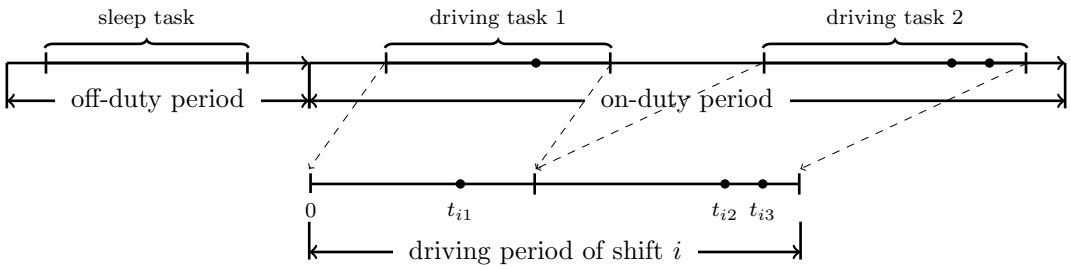


Figure 3.1: Illustration of data processing. Black dots represent ULD events, and t_{ij}, where $j = 1, 2, 3$, represents the driving time to event j in shift i.

study, the off-duty duration has to be reasonably long so that the beginning of a succeeding on-duty period is a "fresh start" for the driver. Another complication from data processing is that some participants failed to distinguish between on-duty and off-duty activity codes. For example, on-duty tasks were coded as off-duty and vice versa. To address these issues, the data were processed based on the following three criteria.

- Driving the truck, light work, and heavy work were considered on-duty tasks. Other tasks were considered off-duty.
- An off-duty period consisted of off-duty tasks only and met either of the two following criteria: (1) its total duration was 10 hours or longer (to be consistent with the HOS regulations that require at least 10 consecutive hours off duty for truck drivers to start a new on-duty shift); (2) it included at least 7 hours of sleep.
- An on-duty period, which followed immediately after an off-duty one, consisted of on-duty tasks and potentially short off-duty tasks. An on-duty period longer than 14 hours was truncated (to be consistent with the 14-hour on-duty limit in the HOS regulations).

The procedure extracted 2,056 on-duty periods, of which 2,004 involved driving tasks. In each of these 2,004 periods, episodes of driving tasks were connected in time to form the driving period of a shift (see Figure 3.1). The episodes between successive driving tasks were

either breaks (e.g., eat, rest, sleep) or on-duty work (e.g., loading, unloading). Due to the HOS regulations that require drivers to drive no more than 11 hours on-duty, driving longer than 11 hours in a shift was truncated. For each ULD event, the driving time to event from the beginning of the shift was calculated. Total sleep time from the preceding off-duty period was calculated. Since some off-duty periods involved multiple days or even weeks, sleep time within the 12-hour window right before an on-duty period was used. A total of 1,880 shifts from 96 drivers contain valid sleep data and are included in the analysis.

3.3 Methodology

I propose a mixed Poisson process model with time-varying coefficients in a recurrent event data setting. Both the log baseline intensity and time-varying coefficients are modeled by penalized B-splines. A mixed-model representation of penalized splines is adopted to estimate the amount of smoothness for the spline fit.

3.3.1 Model

Consider a sample of I shifts. Let $\{N_i(t), t \geq 0\}$ be a right-continuous counting process that records the number of events over driving time $[0, t]$ in shift i. Let $\boldsymbol{x_i} = (x_{i1}, \ldots, x_{ip})'$ denote a covariate vector with constant coefficients, and $\boldsymbol{z_i} = (z_{i1}, \ldots, z_{iq})'$ denote a covariate vector with time-varying coefficients. Given a random effect u_i, $\{N_i(t), t \geq 0\}|\boldsymbol{x_i}, \boldsymbol{z_i}, u_i$ is assumed to be an independent Poisson process with conditional intensity function as

$$\lambda_i(t|\boldsymbol{x_i}, \boldsymbol{z_i}, u_i) = u_i \exp\{\boldsymbol{x_i'\alpha} + \beta_0(t) + z_{i1}\beta_1(t) + \cdots + z_{iq}\beta_q(t)\}. \tag{3.1}$$

$\boldsymbol{\alpha}$ is a vector of unknown regression coefficients and $\boldsymbol{x_i'\alpha}$ represents the linear component of the log intensity. $\beta_0(t), \beta_1(t), \ldots, \beta_q(t)$ are unknown smooth functions: $\beta_0(t)$ is the log baseline intensity and the rest are time-varying coefficients of the covariates in $\boldsymbol{z_i}$. McCulloch, Neuhaus (2011) showed that statistical inferences are robust to the distribution assumptions

for random effects. The chapter assumes that u_i follows an independent and identical gamma distribution with density $p(u_i;\phi)$, where $\mathrm{E}(u_i) = 1$ and $\mathrm{var}(u_i) = \phi$. Large values of ϕ indicate greater heterogeneity between shifts and stronger association among event times.

Let t_{ij} denote the driving time to the jth event that occurred in shift i. Suppose $n_i(\geq 0)$ events are observed in shift i at driving times $0 < t_{i1} < \ldots < t_{i,n_i}$ before the censoring time c_i. Let $\boldsymbol{D_i^{\mathrm{obs}}}$ denote the observed data for shift i, including event times, censoring time, and covariate values. The log likelihood given the complete data $\{\boldsymbol{D_i^{\mathrm{obs}}}, u_i, 1 \leq i \leq I\}$ is

$$l_{\mathrm{c}} = \sum_{i=1}^{I} \left\{ \sum_{j=1}^{n_i} \Big[\log u_i + \log \rho_i(t_{ij}) \Big] - u_i \mu_i(c_i) \right\} + \sum_{i=1}^{I} \log p(u_i;\phi), \tag{3.2}$$

with $\rho_i(t) = \exp\{\boldsymbol{x_i'\alpha} + \beta_0(t) + z_{i1}\beta_1(t) + \cdots + z_{iq}\beta_q(t)\}$ and $\mu_i(c_i) = \int_0^{c_i} \rho_i(t)\,\mathrm{d}t$. By integrating u_i's from the complete-data likelihood, I obtain the log marginal likelihood as

$$l_{\mathrm{m}} = \sum_{i=1}^{I} \sum_{j=1}^{n_i} \log \rho_i(t_{ij}) - \sum_{i=1}^{I} \Big(n_i + \frac{1}{\phi} \Big) \log \Big(1 + \phi \mu_i(c_i) \Big) + \sum_{i=1}^{I} \sum_{j=0}^{n_i^*} \log(1 + j\phi), \tag{3.3}$$

with $n_i^* = \max(0, n_i - 1)$. The integral $\mu_i(c_i)$ in likelihood functions (3.2) and (3.3) is based on $\beta_0(t), \beta_1(t), \ldots, \beta_q(t)$. Since there is no analytic solution, I use the trapezoid method to approximate the integral. Consider a grid of points $0 = s_0 < s_1 < \cdots < s_m = t_{\max}$ over the domain of time. For shift i, define index $m_i = \min\{k : s_k \geq c_i\}$. Then $\mu_i(c_i)$ can be approximated by a polygon going through the points $(s_k,\ \rho_i(s_k))$, for $0 \leq k \leq m_i - 1$. The approximation yields

$$\begin{aligned} \mu_i(c_i) &\approx (c_i - s_{m_i-1}) \frac{\rho_i(s_{m_i}) + \rho_i(s_{m_i-1})}{2} + I(m_i \geq 2) \sum_{k=1}^{m_i-1} (s_k - s_{k-1}) \frac{\rho_i(s_k) + \rho_i(s_{k-1})}{2} \\ &= \frac{1}{2} \min(c_i, s_1) \rho_i(s_0) + \frac{1}{2} \sum_{k=1}^{m_i} \big[\min(c_i, s_{k+1}) - \min(c_i, s_{k-1}) \big] \rho_i(s_k), \end{aligned}$$

where $I(\cdot)$ denotes the indicator function.

To model the baseline intensity and time-varying coefficients, I assume that $\beta_l(t)$, $0 \leq l \leq q$,

can be approximated by polynomial splines of degree v with equally spaced knots $0 = \zeta_{l0} < \zeta_{l1} < \cdots < \zeta_{l,k_l} = t_{\max}$ over the domain of time. Let $B_{lk}(t)$ denote the value at t of the kth B-spline, and $\boldsymbol{B_l}(t) = (B_{l1}(t), \ldots, B_{l,K_l}(t))'$ denote $K_l = k_l + v$ B-spline basis functions. The parameterization used for the splines is

$$\beta_l(t) = \sum_{k=1}^{K_l} \beta_{lk} B_{lk}(t) = \boldsymbol{B_l'}(t)\boldsymbol{\beta_l}. \tag{3.4}$$

$\boldsymbol{\beta_l} = (\beta_{l1}, \ldots, \beta_{l,K_l})'$ is a vector of unknown regression coefficients for the splines. To ensure sufficient flexibility for the spline fit, a moderately large number of knots should be chosen. However, this could lead to severe overfitting. Eilers, Marx (1996) proposed a penalized approach to prevent overfitting by incorporating a difference penalty on the coefficients of adjacent splines. Let Δ denote the difference operator so that the first order difference of β_{lk} is $\Delta\beta_{lk} = \beta_{lk} - \beta_{l,k-1}$, the second order difference is $\Delta^2\beta_{lk} = \Delta(\Delta\beta_{lk}) = \beta_{lk} - 2\beta_{l,k-1} + \beta_{l,k-2}$, and so on. In general, a difference penalty based on the rth order difference is subtracted from the log likelihood; parameters are estimated by maximizing the resulting penalized likelihood function, that is,

$$l_{\mathrm{c}} - \sum_{l=0}^{q} \left\{ \frac{\lambda_l}{2} \sum_{k=r+1}^{K_l} (\Delta^r \beta_{lk})^2 \right\}. \tag{3.5}$$

λ_l, often known as the smoothing parameter, is a non-negative constant which can be adjusted to control the smoothness of the fit. For $\lambda_l = 0$, (3.5) reduces to ordinary likelihood function, yielding a overfitted $\beta_l(t)$. By increasing λ_l, the amount of penalty increases and the smoothness can be tuned. In the limit of a large λ_l, a linear ($r = 2$) or quadratic ($r = 3$) fit can be obtained. The value of λ_l has to be selected by cross-validation or optimizing some information criterion. However, these procedures can be computationally demanding in our setting: as the number of time-varying coefficients increases, it yields a high dimensional grid from which the combination of smoothing parameters are to be selected.

To estimate the smoothing parameter, I adopt a mixed-model representation of penalized

splines based on the connection between penalized likelihood and random effects models (Brumback, Rice, 1998; Wang, 1998a,b; Brumback et al., 1999). Cai et al. (2002) and Cai, Betensky (2003) employed such connection in baseline hazard estimation and treated the coefficients of truncated power splines as independent Gaussian random effects. For B-splines used in the chapter, I assume that the rth order difference $\Delta^r \beta_{lk}$, $r+1 \leq k \leq K_l$, is a random effect following an independent and identical normal distribution $N(0, 1/\tau_l)$. Let $\boldsymbol{D_l}$ be a matrix such that $\boldsymbol{D_l}\boldsymbol{\beta_l} = (\Delta^r \beta_{l,r+1}, \ldots, \Delta^r \beta_{l,K_l})'$. The probability density of $\boldsymbol{\beta_l}$ can be written as

$$p(\boldsymbol{\beta_l}; \tau_l) \propto \tau_l^{(K_l - r)/2} \exp\left\{-\frac{\tau_l}{2} \boldsymbol{\beta_l}'(\boldsymbol{D_l}'\boldsymbol{D_l})\boldsymbol{\beta_l}\right\}. \tag{3.6}$$

The smoothness of the fit is now controlled by the inverse variance τ_l, which corresponds to λ_l in the penalized likelihood approach. It should be noted that formulation (3.6) is originally proposed as a prior in the Bayesian approach to penalized splines (Berry et al., 2002; Lang, Brezger, 2004). Yet in the frequentist setting by assuming spline coefficients to be random, smoothing parameter selection turns into inverse variance estimation. In the following section I show that, with the PQL method (Breslow, Clayton, 1993), the estimate is approximately equivalent to a penalized spline fit under a certain amount of difference penalty.

3.3.2 Estimation

The model is fitted by an EM algorithm with the PQL approach embedded for likelihood approximation. Let $\tilde{l}$ denote the likelihood functions in which $\boldsymbol{\beta_l}$ is random with density (3.6). The log likelihood given the complete data $\{\boldsymbol{D_i^{\text{obs}}}, u_i, 1 \leq i \leq I, \boldsymbol{\beta_l}, 0 \leq l \leq q\}$ is

$$\begin{aligned}\tilde{l}_{\text{c}} &= \sum_{i=1}^{I} \left\{ \sum_{j=1}^{n_i} \Big[\log u_i + \log \rho_i(t_{ij}) \Big] - u_i \mu_i(c_i) \right\} + \sum_{i=1}^{I} \log p(u_i; \phi) + \sum_{l=0}^{q} \log p(\boldsymbol{\beta_l}; \tau_l) \\ &= \tilde{l}_1(\boldsymbol{\alpha}, \boldsymbol{\tau}) + \tilde{l}_2(\phi),\end{aligned}$$

where $\boldsymbol{\tau} = (\tau_0, \tau_1, \ldots, \tau_q)'$,

$$\tilde{l}_1(\boldsymbol{\alpha}, \boldsymbol{\tau}) = \frac{1}{2}\sum_{l=0}^{q}(K_l - r)\log\tau_l + \sum_{i=1}^{I}\sum_{j=1}^{n_i}\log\rho_i(t_{ij}) - \sum_{i=1}^{I}u_i\mu_i(c_i) - \frac{1}{2}\boldsymbol{\beta}'\boldsymbol{P}(\boldsymbol{\tau})\boldsymbol{\beta},$$

and

$$\tilde{l}_2(\phi) = \frac{1}{\phi}\sum_{i=1}^{I}(\log u_i - u_i) + I\left[\frac{1}{\phi}\log\left(\frac{1}{\phi}\right) - \log\Gamma\left(\frac{1}{\phi}\right)\right].$$

$\boldsymbol{\beta}' = (\boldsymbol{\beta_0}', \boldsymbol{\beta_1}', \ldots, \boldsymbol{\beta_q}')$, and $\boldsymbol{P}(\boldsymbol{\tau})$ is a block diagonal matrix with block $\tau_l(\boldsymbol{D_l}'\boldsymbol{D_l})$ in the term corresponding to $\boldsymbol{\beta_l}$ and 0's elsewhere.

In the E-step of the algorithm, conditional on the observed data, u_i independently follows a gamma distribution, with shape parameter $A_i = n_i + 1/\phi$ and rate parameter $B_i = \mu_i(c_i) + 1/\phi$. Given the data and the current values of A_i and B_i, the expectation of $\tilde{l}_{\mathrm{c}}$ with respect to u_i is $\widetilde{Q}_1(\boldsymbol{\alpha}, \boldsymbol{\tau}) + \widetilde{Q}_2(\phi)$, where

$$\widetilde{Q}_1(\boldsymbol{\alpha}, \boldsymbol{\tau}) = \frac{1}{2}\sum_{l=0}^{q}(K_l - r)\log\tau_l + \sum_{i=1}^{I}\sum_{j=1}^{n_i}\log\rho_i(t_{ij}) - \sum_{i=1}^{I}\left(\frac{A_i}{B_i}\right)\mu_i(c_i) - \frac{1}{2}\boldsymbol{\beta}'\boldsymbol{P}(\boldsymbol{\tau})\boldsymbol{\beta},$$

and

$$\widetilde{Q}_2(\phi) = \frac{1}{\phi}\sum_{i=1}^{I}\left(\psi(A_i) - \log(B_i) - \frac{A_i}{B_i}\right) + I\left[\frac{1}{\phi}\log\left(\frac{1}{\phi}\right) - \log\Gamma\left(\frac{1}{\phi}\right)\right].$$

$\psi(\cdot)$ denotes the digamma function.

In the M-step of the algorithm, I can update the value of ϕ by maximizing $\widetilde{Q}_2$. To update the estimates of $(\boldsymbol{\alpha}, \boldsymbol{\tau})$, it is important to notice that $\widetilde{Q}_1$ is conditional on the random effects $\boldsymbol{\beta_l}$'s. Therefore $\boldsymbol{\beta_l}$'s need to be integrated out. The integrated $\widetilde{Q}_1$ is

$$\log\int\exp\widetilde{Q}_1\,\mathrm{d}\boldsymbol{\beta} = \frac{1}{2}\sum_{l=0}^{q}(K_l - r)\log\tau_l + \log\int\exp\{-\kappa(\boldsymbol{\beta})\}\,\mathrm{d}\boldsymbol{\beta}, \tag{3.7}$$

where,

$$-\kappa(\boldsymbol{\beta}) = \sum_{i=1}^{I}\sum_{j=1}^{n_i} \log \rho_i(t_{ij}) - \sum_{i=1}^{I} \left(\frac{A_i}{B_i}\right) \mu_i(c_i) - \frac{1}{2}\boldsymbol{\beta}'\boldsymbol{P}(\boldsymbol{\tau})\boldsymbol{\beta}. \tag{3.8}$$

Since there is no closed form for the integral in (3.7), I apply Laplace's method (Tierney, Kadane, 1986) for approximation. Let κ' and κ'' denote the first and second partial derivative of κ with respect to $\boldsymbol{\beta}$. The approximation yields

$$\log \int \exp \widetilde{Q}_1 \, \mathrm{d}\boldsymbol{\beta} \approx \frac{1}{2}\sum_{l=0}^{q}(K_l - r)\log \tau_l - \frac{1}{2}\log\left|\kappa''(\tilde{\boldsymbol{\beta}})\right| - \kappa(\tilde{\boldsymbol{\beta}}), \tag{3.9}$$

where $\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\boldsymbol{\alpha}, \boldsymbol{\tau})$ is the solution to $\kappa'(\boldsymbol{\beta}) = \mathbf{0}$ that maximizes $-\kappa(\boldsymbol{\beta})$. I follow Breslow, Clayton (1993) in their PQL approach for generalized linear mixed models. That is, I ignore the first two terms in (3.9) and choose $\boldsymbol{\alpha}$ to maximize $-\kappa(\tilde{\boldsymbol{\beta}})$. Consequently, $(\hat{\boldsymbol{\alpha}}(\boldsymbol{\tau}), \hat{\boldsymbol{\beta}}(\boldsymbol{\tau}))$, where $\hat{\boldsymbol{\beta}}(\boldsymbol{\tau}) = \tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}(\boldsymbol{\tau}))$, jointly maximizes $-\kappa(\boldsymbol{\beta})$. If $\boldsymbol{\tau}$ were known and $\boldsymbol{\beta_l}$'s were fixed effects, $-\kappa(\boldsymbol{\beta})$ would be the penalized likelihood with $-(1/2)\boldsymbol{\beta}'\boldsymbol{P}(\boldsymbol{\tau})\boldsymbol{\beta}$ being equivalent to the difference penalty term in (3.5). The PQL approximation implies that, given $\boldsymbol{\tau}$, maximizing the integrated $\widetilde{Q}_1$ is approximately equivalent to maximizing the penalized likelihood. The estimation procedure is then reduced to solving the score equation of $-\kappa(\boldsymbol{\beta})$ for $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, and the Newton-Raphson algorithm can be applied.

By assigning the maximized values $(\hat{\boldsymbol{\alpha}}(\boldsymbol{\tau}), \hat{\boldsymbol{\beta}}(\boldsymbol{\tau}))$ back into (3.9), I obtain an approximate profile likelihood of $\boldsymbol{\tau}$ as

$$\frac{1}{2}\sum_{l=0}^{q}(K_l - r)\log \tau_l - \frac{1}{2}\log\left|\kappa''(\hat{\boldsymbol{\beta}})\right| - \frac{1}{2}\hat{\boldsymbol{\beta}}'\boldsymbol{P}(\boldsymbol{\tau})\hat{\boldsymbol{\beta}}. \tag{3.10}$$

Differentiation of (3.10) with respect to τ_l gives the following estimating equation

$$\frac{1}{2}\left[\frac{K_l - r}{\tau_l} - \mathrm{tr}\big(\kappa''(\hat{\boldsymbol{\beta}})^{-l}\boldsymbol{D_l'D_l}\big) - \hat{\boldsymbol{\beta}}_{\boldsymbol{l}}'(\boldsymbol{D_l'D_l})\hat{\boldsymbol{\beta}}_{\boldsymbol{l}}\right] = 0,$$

where $\mathrm{tr}(\cdot)$ is the trace operation of a matrix, and $\kappa''(\hat{\boldsymbol{\beta}})^{-l}$ is the diagonal block corresponding to $\boldsymbol{\beta_l}$ in $\kappa''(\hat{\boldsymbol{\beta}})^{-1}$. The solution is of the simple form:

$$\hat{\tau}_l = \frac{K_l - r}{\mathrm{tr}\big(\kappa''(\hat{\boldsymbol{\beta}})^{-l}\boldsymbol{D_l'D_l}\big) + \hat{\boldsymbol{\beta}}_{\boldsymbol{l}}'(\boldsymbol{D_l'D_l})\hat{\boldsymbol{\beta}}_{\boldsymbol{l}}}. \tag{3.11}$$

The estimation procedure can be conducted in two steps. Firstly, a set of random initial values were assigned to $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau}, \phi)$. Conditional on the current values, I update ϕ by maximizing $\widetilde{Q}_2$ and update $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ by solving the score equation of $-\kappa(\boldsymbol{\beta})$. Secondly, through the updated values of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, I obtain a new value for τ_l using (3.11). These two steps are iterated until convergence.

After obtaining estimates for $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \phi)$, I follow the similar method by Gray (1992) and obtain the covariance matrix of $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ given $\boldsymbol{\tau}$ approximately as

$$\mathrm{cov}\left(\left[\begin{array}{c}\hat{\boldsymbol{\alpha}}\\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\\ \hat{\phi}\end{array}\right]\middle|\,\boldsymbol{\tau}\right) = [\boldsymbol{I}(\boldsymbol{\theta}) + \boldsymbol{P}(\boldsymbol{\tau})]^{-1}\boldsymbol{I}(\boldsymbol{\theta})[\boldsymbol{I}(\boldsymbol{\theta}) + \boldsymbol{P}(\boldsymbol{\tau})]^{-1}, \tag{3.12}$$

where $\boldsymbol{I}(\boldsymbol{\theta}) = -\partial^2 l_{\mathrm{m}}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^2$ and l_{m} is the log marginal likelihood defined in (3.3). To obtain the confidence interval for $\hat{\beta}_l(t)$, I estimate the variance of $\hat{\beta}_l(t)$ as $\widehat{\mathrm{var}}\big[\hat{\beta}_l(t)\big] = \boldsymbol{B_l'}(t)\,\widehat{\mathrm{cov}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}|\boldsymbol{\tau})\,\boldsymbol{B_l}(t)$. It should be noted that, as the uncertainty in estimating $\boldsymbol{\tau}$ is ignored in (3.12), these estimates tend to underestimate the true variances; however, the source of variation due to the estimation of $\boldsymbol{\tau}$ is negligible for large samples (Cai, Betensky, 2003). The potential influence is evaluated in the simulation study in Section 3.4.

3.4 Simulation

I conducted a simulation study to evaluate the model performance. Two major factors could affect the fitting results are the curvature of time-varying coefficients and the magnitude of

event rate. I generated data from low and high event rates with time-varying coefficients of different curvatures.

Simulated data were generated from a Poisson process with conditional intensity function

$$\lambda_i(t|u_i) = u_i \exp\{x_{i1}\alpha_1 + \beta_0(t) + z_{i1}\beta_1(t) + z_{i2}\beta_2(t)\},\ 0 < t \leq 10.$$

I considered $I = 500$ shifts. The censoring time c_i was generated from a gamma distribution independent of the shift. shifts were censored at the maximum follow-up time, 10, if not censored before. u_i's were generated from a gamma distribution with mean of 1 and variance $\phi = 5$. The covariate x_{i1} was generated from a uniform distribution on $[-1, 1]$, and its coefficient was $\alpha_1 = 0.5$. The covariates z_{i1} and z_{i2} were binary with probability distribution $p(z_{i1}, z_{i2}) = (1 - p_1 - p_2)^{1-z_{i1}-z_{i2}} p_1^{z_{i1}} p_2^{z_{i2}}$, where $0 \leq z_{i1} + z_{i2} \leq 1$ and $p_1 = p_2 = 1/3$. There were two settings for the time-varying coefficients. Setting 1 considered functions of low curvature (shown in Figure 3.2):

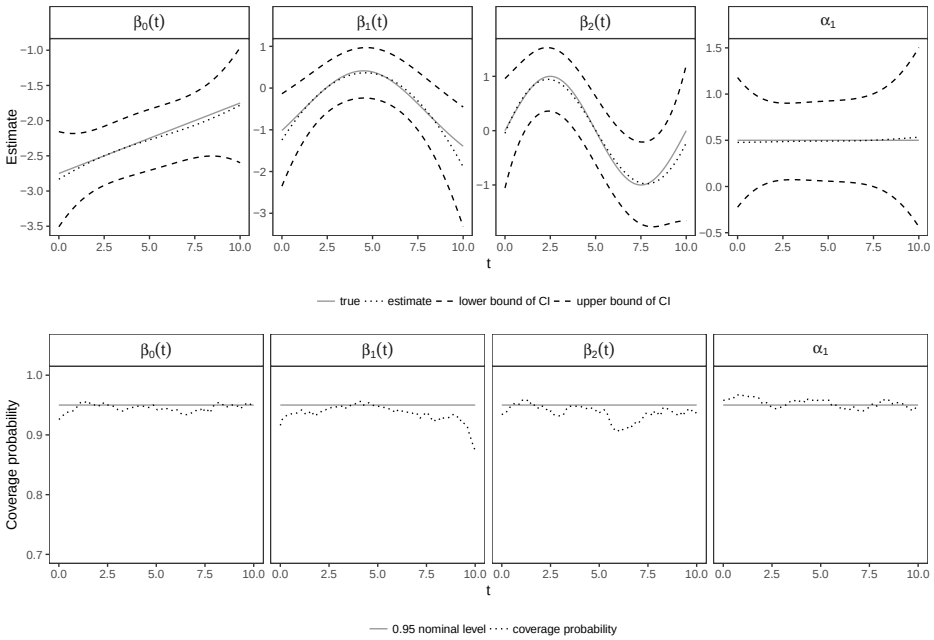
$$\begin{aligned}
\beta_1(t) &= \frac{10}{1 + e^{-0.4(t-3)}} + \frac{10}{1 + e^{0.4(t-6)}} - 12.5, \\
\beta_2(t) &= \sin(0.2\pi t).
\end{aligned}$$

Setting 2 considered functions of high curvature (shown in Figure 3.3):

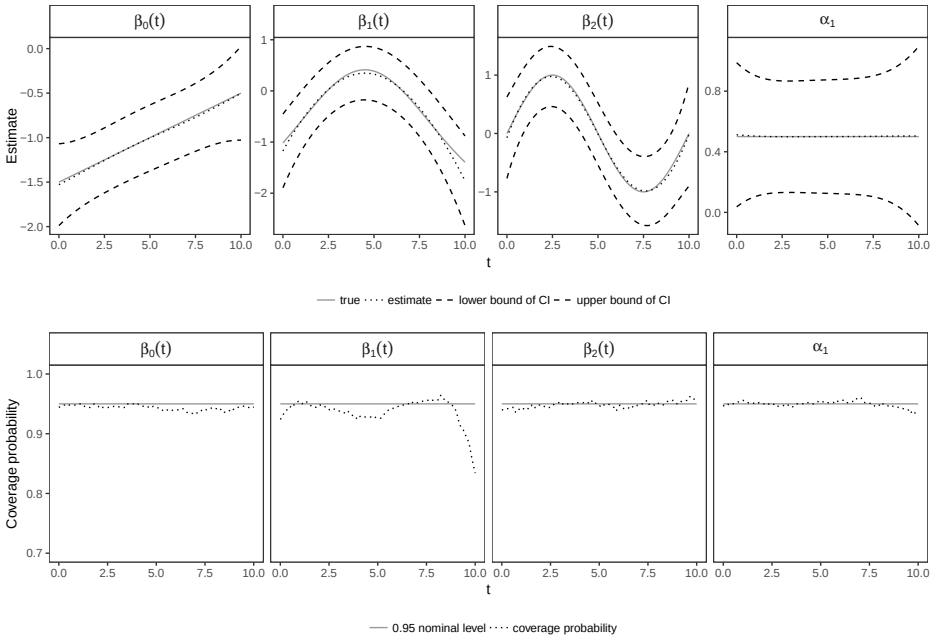
$$\begin{aligned}
\beta_1(t) &= \frac{8}{1 + e^{-(t-4.5)}} + \frac{8}{1 + e^{(t-5.5)}} - 9, \\
\beta_2(t) &= \frac{15}{1 + e^{-(t-4)}} + \frac{15}{1 + e^{(t-4.5)}} + \frac{15}{1 + e^{(t-5.5)}} + \frac{15}{1 + e^{-(t-6)}} - 30.
\end{aligned}$$

There were two settings for the log baseline intensity: a low event rate setting with $\beta_0(t) = 0.1t - 2.75$ and a high event rate setting with $\beta_0(t) = 0.1t - 1.5$.

I generated 500 data sets under each setting. The average number of events per shift (shown in Table 3.1) was 0.8 for the low event rate setting and 3 for the high event rate setting. $\beta_l(t)$'s were approximated using cubic B-splines with the number of knots $k_l = 10$ and the

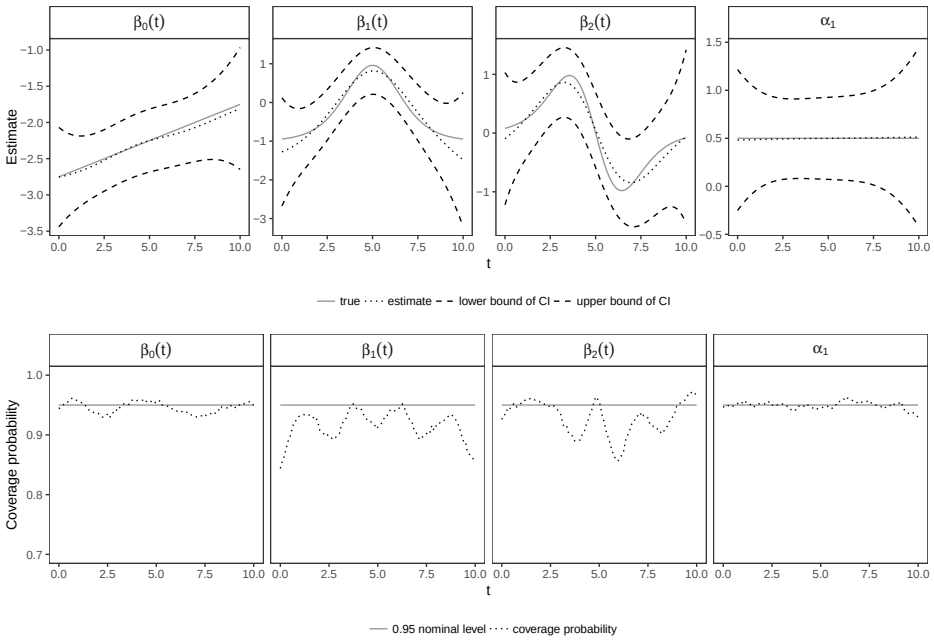


(a) Low event rate.

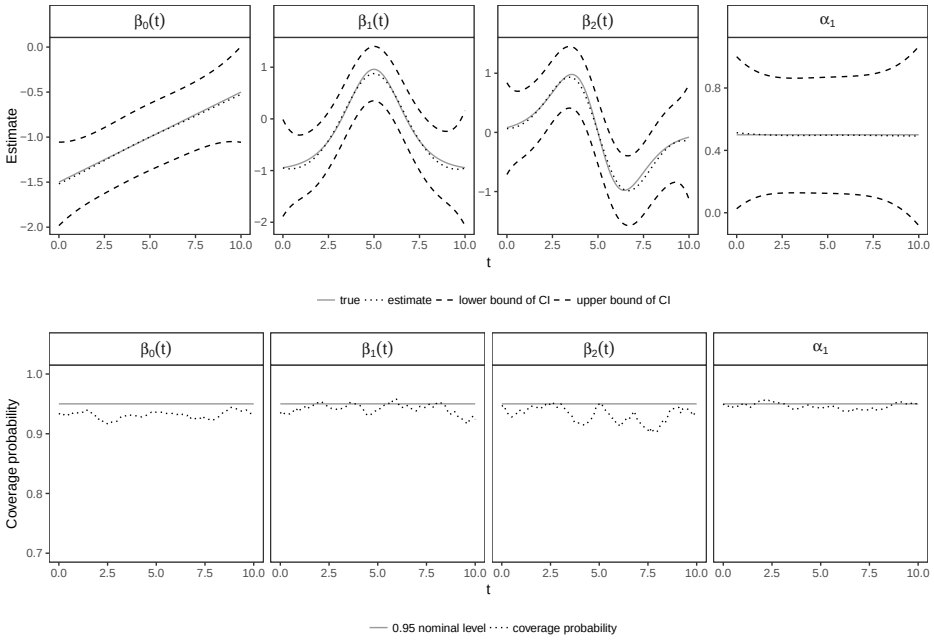


(b) High event rate.

Figure 3.2: Estimates, 95% confidence intervals (CIs) and coverage probability for the low-curvature time-varying coefficients simulation setting.



(a) Low event rate.



(b) High event rate.

Figure 3.3: Estimates, 95% confidence intervals (CIs) and coverage probability for the high-curvature time-varying coefficients simulation setting.

difference penalty order $r = 3$. Each data set was fitted in two approaches: (1) using splines for both $\beta_l(t)$'s and α_1; and (2) using splines for $\beta_l(t)$'s only.

Simulation results for the spline fit are shown in Figure 3.2 and 3.3. For time-varying coefficients of low curvature, the mean of spline estimate shows very little bias, and the pointwise coverage probability is close to 95% over most of the follow-up period. For time-varying coefficients of high curvature, the mean estimates of $\beta_1(t)$ and $\beta_2(t)$ show larger bias under the low event rate setting. The corresponding coverage probabilities also display greater variation. However, in the high event rate setting, the mean estimates show negligible bias with the coverage probabilities more stable and close to the 95% nominal value.

Simulation results for the parametric coefficients are shown in Table 3.1. The $\hat{\phi}$ is approximate 2% smaller than the true value in the low event rate setting and 0.6% smaller than the true value in the high event rate setting. I assess the standard error estimation by comparing the mean of standard error (SEM, the average of estimated standard error) and the empirical standard error of the estimates (SE), which should be equal ideally. The SEM for $\hat{\phi}$ is 4% smaller than the SE under the setting of low-curvature time-varying coefficients and low event rate. For the rest of settings the SEM and SE are similar.

The bias of $\hat{\phi}$ and the underestimation of standard error are consistent with the results of Ripatti, Palmgren (2000), who used the PQL approximation in the inference procedure for multivariate frailty models. Since the uncertainty in the estimation of $\boldsymbol{\tau}$ is ignored when estimating the variances, the estimates tend to underestimate the true variances (see Equation (3.12) and the related discussion).

The mean estimate of α_1 has similar bias (2% smaller) in the low event rate setting, yet such bias is negligible in the high event rate setting. The bias in $\hat{\alpha}_1$ is connected to the bias in $\hat{\phi}$. The SEM for $\hat{\alpha}_1$ is very close to its SE, showing no underestimation. The coverage probability of the 95% confidence intervals for $\hat{\alpha}_1$ is close to or even slightly higher than the nominal level.

In summary, the simulation study shows the model provide satisfactory estimation for the

Table 3.1: Simulation: parametric coefficient estimates.

Curvature	Number of events per shift	Fit		True	Mean	Bias	SE[c]	SEM[d]	CP[e](%)
low	0.83	1[a]	ϕ	5	4.91	−0.09	0.68	0.65	
	(low event rate)	2[b]	α_1	0.5	0.49	−0.01	0.19	0.20	96.2
			ϕ	5	4.91	−0.09	0.68	0.65	
	2.93	1	ϕ	5	4.97	−0.03	0.46	0.46	
	(high event rate)	2	α_1	0.5	0.50	0.00	0.18	0.18	95.6
			ϕ	5	4.97	−0.03	0.46	0.46	
high	0.86	1	ϕ	5	4.90	−0.10	0.65	0.64	
	(low event rate)	2	α_1	0.5	0.49	−0.01	0.20	0.20	94.4
			ϕ	5	4.90	−0.10	0.65	0.64	
	2.99	1	ϕ	5	4.97	−0.03	0.47	0.46	
	(high event rate)	2	α_1	0.5	0.50	−0.00	0.18	0.18	95.6
			ϕ	5	4.97	−0.03	0.46	0.46	

[a] Fit 1, use splines to estimate both $\beta_l(t)$'s and α_1.
[b] Fit 2, use splines to estimate $\beta_l(t)$'s only.
[c] SE, empirical standard error;
[d] SEM, average of estimated standard error;
[e] CP, coverage probability.

true parameters in various settings. The biases of spline estimates and parametric coefficient estimates are small under both low and high event rates. The confidence intervals have close to nominal coverage probabilities. Although the frailty variance tends to be underestimated, the magnitude of the bias is limited.

3.5 Application

I apply the proposed model to the Commercial Truck Driver NDS data. The objective is to evaluate the temporal profile of driving performance over long driving hours in a shift by the total sleep duration prior to the shift. The driving performance was measured by the unintentional lane deviation.

3.5.1 Characteristics of Driving Duration, Unintentional Lane Deviations, and Breaks

The data contain 1,880 off-duty and on-duty pairs from 96 drivers. Each pair includes an off-duty period followed by an on-duty one as introduced in Section 3.2. Studies show that sleep time is generally considered insufficient if the total off-duty sleep duration is less than 7 hours (Van Dongen et al., 2003a,b; Ford et al., 2015; Watson et al., 2015a,b). The off-duty sleep time is considered as normal if it is between 7 and 9 hours, and abundant if the duration is more than 9 hours. Among the 1,880 shifts, 20.6% (388) are in the insufficient sleep time group, 58.2% (1,095) in the normal sleep time group, and 21.1% (397) in the abundant sleep time group.

Table 3.2 compares the rate of ULD events by driving hour since a shift starts among the three sleep time groups. For each driving hour since the beginning of a shift, I calculated the the number of ULDs occurred in the specific hour and the exposure, i.e., total driving time for all shifts that occurred in the specific hour. The event rate for a specific driving hour is a ratio of the ULD count over the driving exposure. As can be seen, the event rate for the insufficient sleep time group increases substantially after 8 hours of driving and reaches a peak of 0.26 in the 10^{th} driving hour. The normal sleep time group, i.e., sleep time between 7 and 9 hours, observes no substantial increase at later hours of driving. The event rate of the abundant group increased considerably in the 10^{th} hour of driving. These results suggest that temporal profile of the ULD event rates vary substantially by sleep time groups. The

Table 3.2: ULDs and driving exposure (time in hours) in the 1st–11th driving hours by sleep time group.

Driving time since a shift starts	Sleep hours ≤ 7			$7 <$ Sleep hours ≤ 9			Sleep hours > 9		
	ULDs	Exposure[a]	Rate[b]	ULDs	Exposure	Rate	ULDs	Exposure	Rate
1^{st} hour	21	384.11	0.055	52	1092.16	0.048	31	395.48	0.078
2^{nd} hour	38	374.18	0.102	52	1078.64	0.048	31	392.78	0.079
3^{rd} hour	13	371.50	0.035	60	1055.97	0.057	31	386.98	0.080
4^{th} hour	14	360.58	0.039	106	1016.10	0.104	24	374.35	0.064
5^{th} hour	33	333.32	0.099	83	970.26	0.086	20	358.31	0.056
6^{th} hour	11	304.90	0.036	94	898.70	0.105	40	334.00	0.120
7^{th} hour	24	251.49	0.095	96	781.78	0.123	36	301.14	0.120
8^{th} hour	26	182.29	0.143	32	665.80	0.048	19	265.40	0.072
9^{th} hour	19	130.99	0.145	40	547.23	0.073	14	228.44	0.061
10^{th} hour	26	99.04	0.263	18	393.08	0.046	22	181.31	0.121
11^{th} hour	10	61.60	0.162	8	267.69	0.030	22	130.49	0.169

[a] Exposure in the i^{th} hour, where $i = 1, \ldots, 11$, is the total driving time for all shifts that occurred in the i^{th} hour.
[b] Rate is the ratio of ULDs to exposure.

group with insufficient sleep time (< 7 hours) tend to have a higher ULD rate in the later driving part of a shift.

Analysis shows that the sleep time prior to a shift is related to the driving length of a shift. The mean driving length for shifts with insufficient sleep time is 7.36 hours, for normal sleep time group is 8.01 hours and for abundant sleep time group is 8.43 hours. A statistically significant correlation can be found between total driving length and sleep time. Figure 3.4 shows the percent of shifts driving into the 1^{st}–11^{th} driving hour for each sleep time group. The shifts with abundant sleep time has the highest percentage of driving in each of the 11 hours, followed by the normal sleep time group, and the insufficient group has the lowest

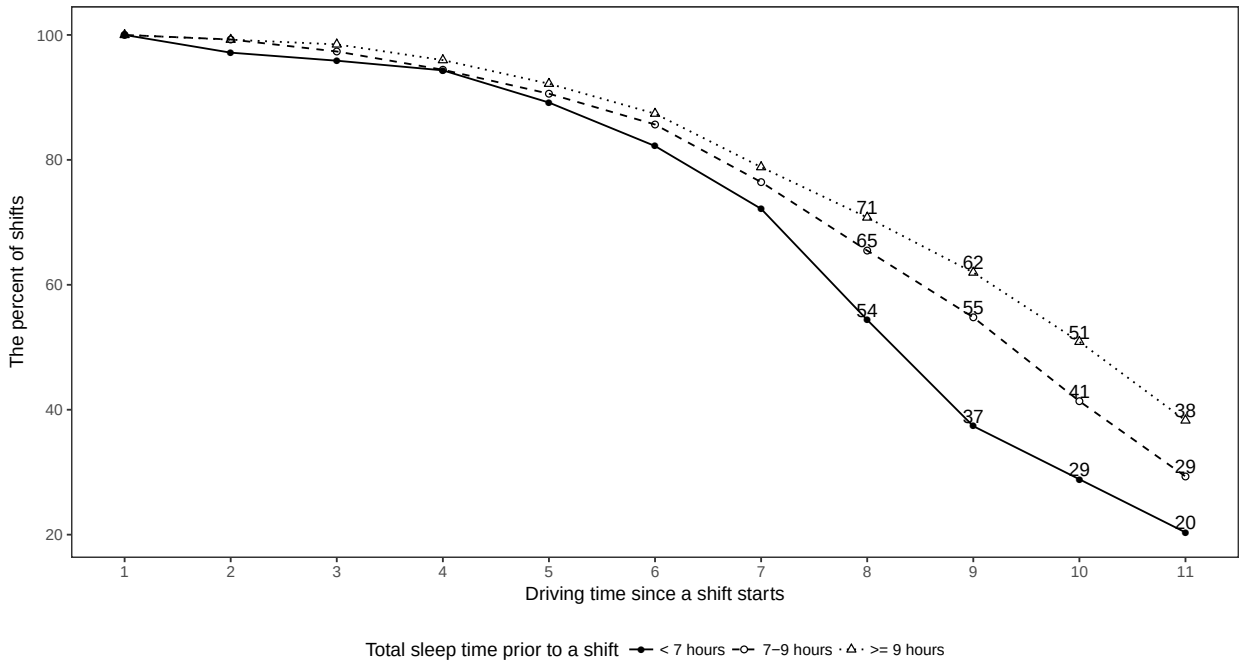


Figure 3.4: The percent of shifts driving into the 1st-11th driving hour by sleep time group.

percentage. The abundant group has 38% of shifts driving into the 11^{th} driving hour, the normal group has 29% while the insufficient group has only 20%. This implies drivers' self-selection in terms of taking sufficient sleep prior to a long shift or shortening driving time with insufficient sleep.

Studies have shown that breaks during long driving periods could also affect the subsequent driving performance (Soccolich et al., 2013). Figure 3.5 compares the ratios of break length to driving exposure among the three sleep time groups. For the i^{th} hour since a shift starts, where $i = 1, \ldots, 11$, the break duration is the sum of break duration from all shifts in the i^{th} hour of driving. The ratio of the break duration over the driving exposure for the i^{th} hour represents the average length of breaks for every hour of driving. As shown in Figure 3.5, in the first 8 driving hours, the ratios for all three groups are stably below 0.2. Starting from the 8^{th} driving hour, the difference becomes substantial: the ratio for insufficient sleep time climbs rapidly and reaches the maximum of 0.8 in the 11^{th} driving hour; the ratio for normal sleep time has a slower increasing trend and reaches its maximum of 0.4 at the end;

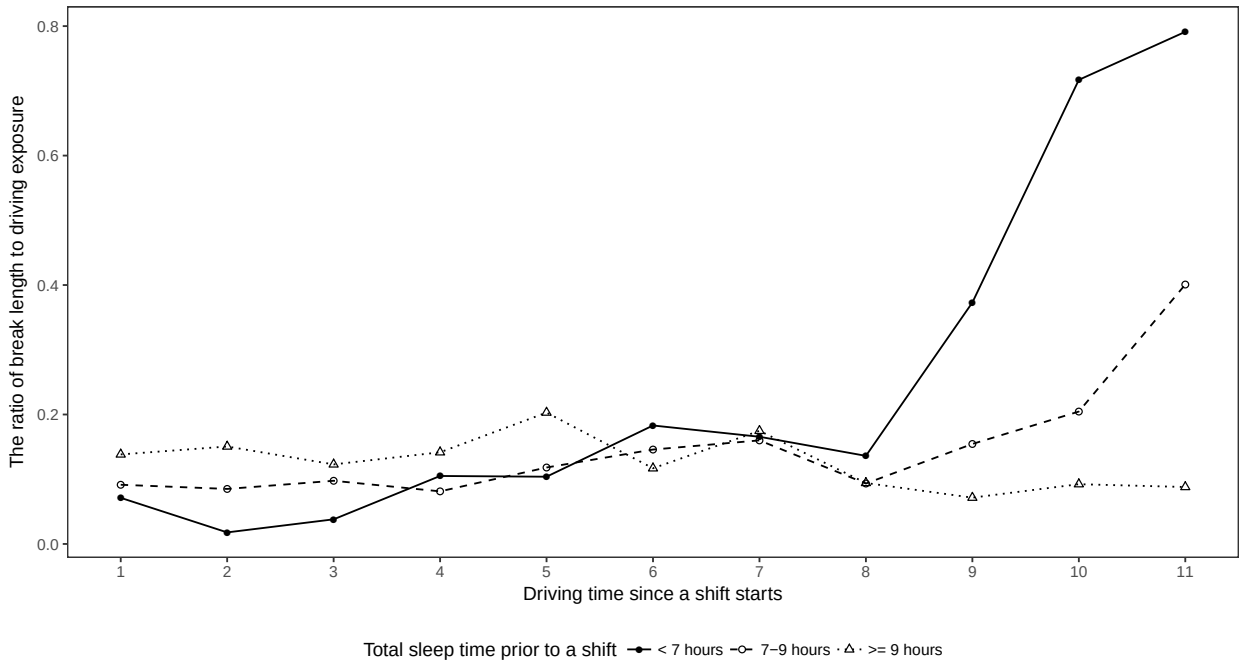


Figure 3.5: The ratio of break length (in hours) to driving exposure (in hours) in the 1st–11th driving hours by sleep time group.

the abundant group stays constantly low around 0.1. This result indicates that drivers who slept less tended to take longer breaks in the later driving part of a shift.

3.5.2 Assessing Driving Performance Profile over Time Using the Time-Varying Coefficient Model

I applied the proposed time-varying coefficient model to the Commercial Truck Driver NDS data. Two dummy variables were created to represent the three sleep time groups. The coefficients for these two dummy variables were assumed to be time-varying. The log baseline (normal sleep time) intensity and the time-varying coefficients were modeled by cubic B-splines with the number of knots $k_l = 10$ and the order of difference penalty $r = 3$. The results show that the estimate of the frailty variance is $\hat{\phi} = 7.68$ and its standard error is 0.59.

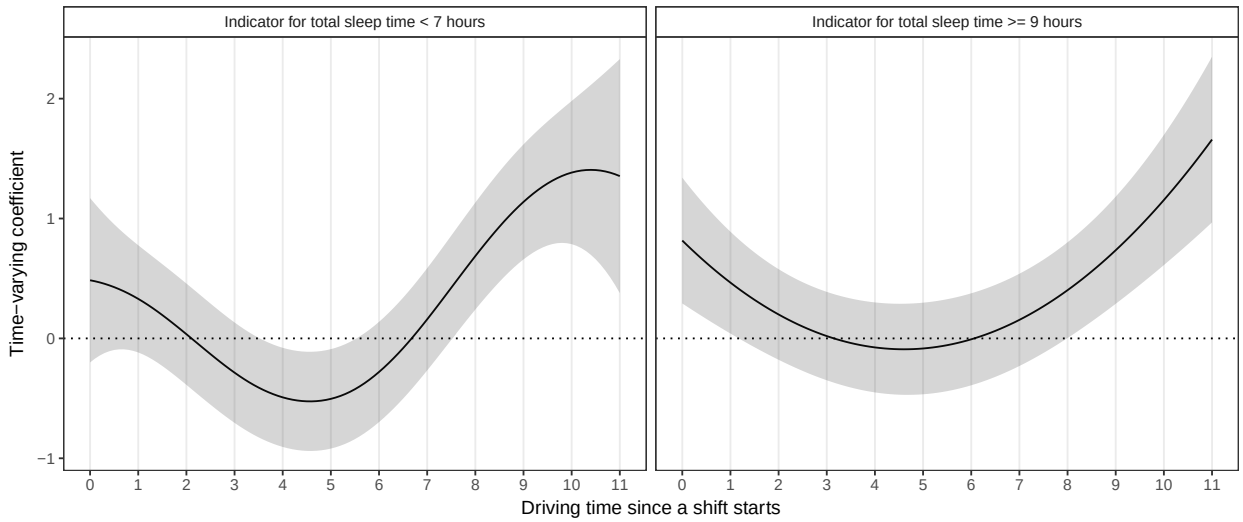


Figure 3.6: The time-varying coefficients: spline estimate (in solid line) and 95% pointwise confidence interval (in shaded area).

Figure 3.6 shows the estimates for time-varying coefficients. The time-varying coefficient of the insufficient indicator variable gives a temporal profile of the difference betIen the insufficient and normal sleep time groups. Starting from the 8^{th} driving hour, the coefficient function is above zero and keeps increasing as driving time increases. This indicates that the intensity of ULD for the insufficient and normal sleep time groups are similar for the first 8 hours of driving. After 8 hours, shifts with insufficient sleep time have a higher intensity than those with normal sleep time. The difference keeps increasing with driving hours. Similar pattern was observed for the difference between the abundant and normal sleep time groups. The shifts with abundant sleep time show a significant higher intensity after 8 hours of driving hour.

Figure 3.7 displays the estimated intensities for the three sleep time groups. The intensity for insufficient sleep time stays low in the first 6 driving hours and then starts to increase. The intensity for normal sleep time slowly climbs up and reaches the peak around the 5th driving hour, after that the function declines as time goes on. The intensity for abundant sleep time remains constant over the whole driving time.

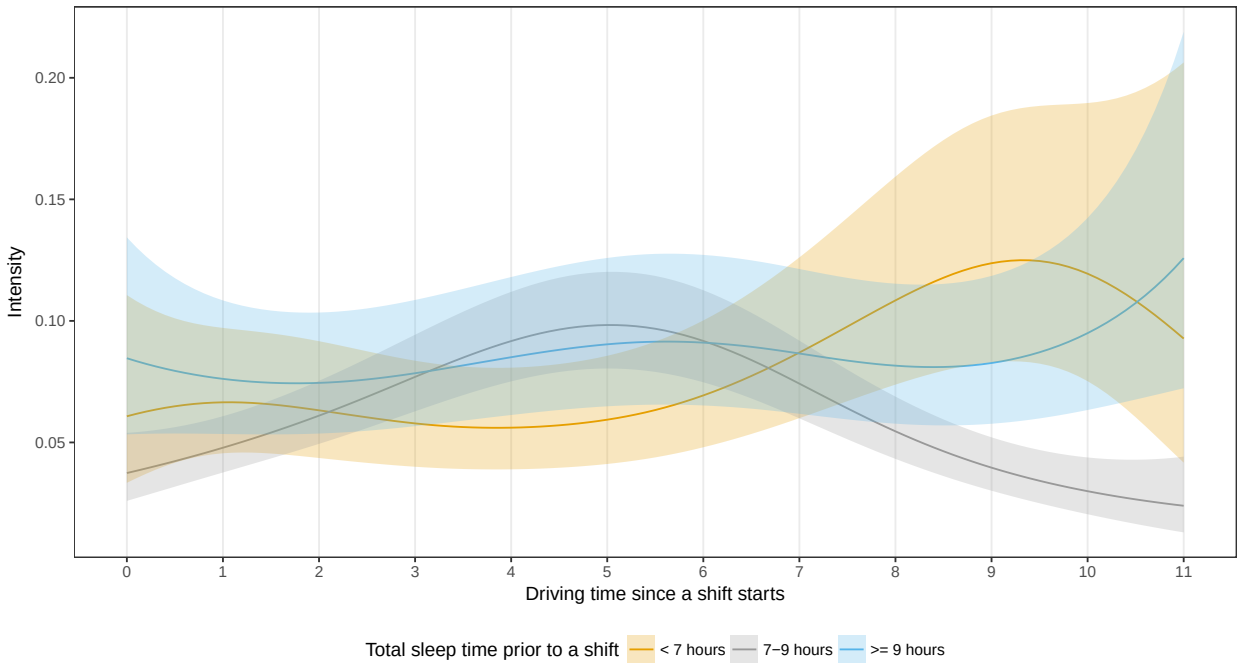


Figure 3.7: The intensity by sleep time group: spline estimate (in solid line) and 95% pointwise confidence interval (in shaded area).

The results indicate that shifts with normal sleep time show a lower intensity compared to both insufficient and abundant sleep time. While the higher intensity for shifts with insufficient is expected, it is counter-intuitive that the abundant sleep group also shows a higher intensity. As many factors could affect driving performance, this result could be results of other factors, for example, the characteristics of breaks described in Section 3.5.1. As shown in Figure 3.5, the ratio of break length to driving exposure for normal sleep time starts to increase from the 8^{th} hour of driving. On the contrary, the abundant group has a low break ratio over the entire driving time. The shorter break length at later part of a shift for the abundant sleep time group could be a potential reason for the high ULD intensity.

The difference between insufficient and normal sleep time is most pronounced. From the 8^{th} driving hour, drivers with insufficient sleep time have the highest break ratio, taking breaks of more than half a hour for every hour of driving. However, in the same time range, the associated coefficient function rises above zero and increases rapidly (see Figure 3.6). This

implies that, compared to normal sleep time, drivers with insufficient sleep time tend to take considerably longer breaks after 8 hour of driving, yet these substantial amount of breaks could not prevent a higher intensity; their gap is even wider as the driving continues. The result implies that longer break in later part of a shift is not sufficient to compensate lack of sleep prior to the shift.

3.6 Discussion

Fatigue is a major issue associated with commercial truck drivers, especially long-haul truck drivers who may work 14 hours, and drive 11 hours in a single shift. A recent report by the National Academies acknowledges the hazard of fatigue and suggests that research needs to be conducted to understand the impact of fatigue on the safety and health of truck drivers (National Academies of Sciences, Engineering, and Medicine, 2016). However, there is limited research on how fatigue affects drivers' performance using high-resolution *in situ* collected driving data. This study advances the research in understanding driving fatigue by evaluating the time-varying effects of total sleep time prior to a shift using the proposed model.

I propose a mixed Poisson process model with time-varying coefficients for recurrent event data. Gamma frailties are incorporated to accommodate the dependence among event times. A mixed-model representation of penalized splines is adopted for modeling both the log baseline intensity and time-varying coefficients. An EM algorithm is used to fit the model, in which the PQL approach is adopted for likelihood approximation. The simulation study has confirmed that the model performs well for time-varying coefficients with different degrees of curvature under both low and high event rate scenarios. There are several possible extensions of the model. This chapter uses one frailty term for shift-level heterogeneity in the data. However, there may be heterogeneity existing at the driver level. A nested frailty model with random effects for both levels is worth pursuing yet requires a larger sample size.

The application results reveal that lack of sleep primarily affects driving performance, as measured by unintentional lane deviations, after 8 hours of driving in a given shift. That is, the negative impacts of insufficient sleep are most apparent in terms of performance decrement well into the driver's shift. The results also implies a complex relationship among total sleep time, the driver's choice to take breaks while on duty, and driving performance. Compared to normal sleep time, drivers with insufficient sleep tend to take markedly more breaks after 8 hours of driving, yet these extra breaks could not prevent a higher risk; their gap is even wider as the driving continues. This highlights the importance of drivers beginning their shift well-rested and having sufficient sleep the night before driving. Furthermore, a driver's fitness-to-drive, in terms of increased safety risk while on the road, can be assessed, in part, by measuring how much sleep the driver received prior to beginning the shift. The finding that a performance decrement appears after 8 hours into the shift also provides important information for technology developers aimed at fatigue detection, and for those interested in determining the role that fatigue may play in a road crash.

Chapter 4 A Bayesian Time-Varying Coefficient Model for Multi-Type Recurrent Events

4.1 Introduction

Recurrent events can be observed in medical and epidemiological studies in which individuals experience events of interest repeatedly over a period of time. It is common for individuals to be at risk for more than one type of event. For example, patients in skin cancer trials are at risk of recurrence of both basal cell epithelioma and squamous cell carcinoma tumors (Abu-Libdeh et al., 1990), and recurrent asthma exacerbations for respiratory studies can be classified by sputum cell counts as eosinophilic or non-eosinophilic (Jayaram et al., 2006). One interest aspect of multi-type recurrent event modeling is to evaluate how the intensity vary over a period of time as well as how the profile is affected by factors. Limited research has been conducted to address this issue and the purpose of this study is to develop a general and flexible method for characterizing the temporal patterns of multi-type recurrent events.

Multi-type recurrent event analysis typically includes marginal models based on estimating functions (e.g., Ng, Cook, 1999; Cai, Schaubel, 2004) and joint models based on multivariate frailties (e.g., Abu-Libdeh et al., 1990; Chen et al., 2005). Methods using estimating functions are constructed to directly describe marginal means and treat the dependence structure of events as nuisance parameters. In contrast, joint models use multivariate frailties to characterize the heterogeneity and correlation among events, and allow both marginal and conditional inferences.

The temporal profile of the baseline intensity could be of interest in many studies. One common approach for baseline modeling is to use a piecewise-constant intensity form (e.g., Chen

et al., 2005; Chen, Cook, 2009). Li et al. (2017) and Li et al. (2018) used change-point models based on piecewise-constant intensity functions to detect the time of change in driving risks for novice teenage drivers. However, baseline modeling that relies on piecewise-constant forms requires accurate specification of the number and locations of pieces (Friedman, 1982). An alternative approach is based on nonparametric baseline assumptions and yields an estimator similar to the Nelson-Aalen estimator (Aalen, 1978) from survival analysis. The estimator is robust with respect to assumptions but might be unstable in regions of sparse data (Cai, Betensky, 2003).

There are situations where a smooth intensity form is desired. For example, it is beneficial to estimate covariate effects for interval-censored data (Cai et al., 2002). Nielsen, Dean (2005) employed regression splines to model recurrent events, though there is no guarantee of monotonicity for the cumulative intensity estimate in small samples. The development of a more general and flexible method is required for smooth estimation of baseline intensity.

The effect of a covariate can vary over time for recurrent events and could be of primary interest. For example, total sleep time the night before driving is a critical risk factor affecting driver fatigue, and its impact on driving performance is likely to change over a long shift due to increasing fatigue over time (Hanowski et al., 2007). In the Commercial Truck Driver Naturalistic Driving Study (see Section 4.5 for details), the investigators were interested in assessing how driving performance varies over time and the impact of sleep time on its temporal profile. Varying coefficient models for longitudinal data include kernel-based methods (e.g., Carroll et al., 1998; Fan, Zhang, 1999; Cai et al., 2000) and spline-based methods (e.g., Hastie, Tibshirani, 1993; Huang et al., 2002). Kernel estimators are constructed through a weighted mean of the nearby observations and have great intuitive appeal. However, the kernel weights are asymmetric at boundaries of the predictor space and could cause substantial bias for the estimate (Hastie, Loader, 1993). Spline estimators are based on linear combinations of piecewise polynomial functions and are a straightforward extension of linear regression models. The computations are relatively fast and can be easily implemented (Eilers, Marx, 1996). Although research has been conducted on varying

coefficients for recurrent event data (Amorim et al., 2008; Sun et al., 2011; Yu et al., 2013; Lin et al., 2015), there is limited work for multi-type recurrent events that simultaneously incorporates baseline modeling, time-varying coefficient estimation while also accounting for the dependence structure between event types.

To assess the time-varying patterns for multi-type recurrent events, this chapter proposes a Bayesian joint model with time-varying coefficients. The joint model is based on multivariate log-normal frailties to characterize the heterogeneity and correlation among different event types. The proposed model uses Bayesian penalized splines (Lang, Brezger, 2004) to achieve smooth estimates for both time-varying coefficients and the baseline intensity. The cumulative intensity in the posterior is then approximated by numerical integration based on the trapezoid method (Kauermann, 2005). In Bayesian penalized splines, the smoothness of a spline fit depends substantially on the subjective choice of hyperparameters (Brezger, Lang, 2006). However, re-fitting the model with a number of choices can be computationally demanding. To address this, I employ a robust prior specification (Jullion, Lambert, 2007) to achieve an objective fitting procedure. One challenge in using Markov chain Monte Carlo methods for posterior inference is sampling from the high-dimensional distribution of spline parameters. This work uses Metropolis-adjusted Langevin algorithms (MALA; Roberts, Tweedie, 1996), a class of Metropolis-Hastings algorithms whose proposals are based on the gradient of the target density. Another concern in sampling spline parameters is that their MALA proposal strongly depends on a roughness penalty parameter introduced by Bayesian penalized splines. Separately updating these two parameters could lead to convergence issues. Thus, I integrate MALA algorithms into a joint sampling scheme (Knorr-Held, Rue, 2002) for better convergence and mixing properties. This chapter provides a general approach that can be directly implemented while still providing smooth estimates for features of multi-type recurrent event data.

The chapter uses data from the Commercial Truck Driver Naturalistic Driving Study (Blanco et al., 2011) as an example for application. The study collected large-scale naturalistic driving data to evaluate issues related to commercial motor vehicles. Repeated on-road safety-critical

events were identified for the recruited drivers, including crashes, near-crashes, crash-relevant conflicts and unintentional lane deviations. The proposed approach was applied to evaluate how driving performance changes over time and to compare this temporal profile across total sleep time before driving.

The remainder of our chapter is organized as follows. Section 4.2 introduces a Bayesian joint model with time-varying coefficients. Section 4.3 describes the joint posterior sampling scheme, in which the MALA algorithm is integrated for sampling from the high-dimensional distribution of spline parameters. A simulation study is presented in Section 4.4. Section 4.5 applies the proposed method to data from the Commercial Truck Driver Naturalistic Driving Study. Section 4.6 provides a summary and discussion.

4.2 Model

The proposed approach is a joint model based on multivariate log-normal frailties with time-varying coefficients. Bayesian penalized splines are used to model both time-varying coefficients and the baseline intensity. To address the issue that subjective choices of hyper-parameters control the smoothness of a spline fit, a robust prior specification approach is adopted to achieve an objective fitting procedure.

Let $1 \leq i \leq I$ be the index of event processes and $1 \leq j \leq J$ be the index of event types. Let $N_{ij}(t)$ denote the number of type j events that occurred over $[0, t]$ for process i. The multivariate event process is $\{\boldsymbol{N_i}(t), t \geq 0\}$, where $\boldsymbol{N_i}(t) = (N_{i1}(t), \ldots, N_{iJ}(t))'$. Let $\boldsymbol{x_i}$ denote a vector of covariates with constant coefficients, and $\boldsymbol{z_i} = (z_{i1}, \ldots, z_{iq})'$ denote covariates with time-varying coefficients. I use $\boldsymbol{H_i}(t) = \{\boldsymbol{N_i}(s), 0 \leq s < t, \boldsymbol{x_i}, \boldsymbol{z_i}\}$ to represent the process history at time t, which includes the times and types of events that occurred over $[0, t)$ as well as their covariate values. Let $\Delta N_{ij}(t)$ denote the number of type j events in $[t, t + \Delta t)$. Conditional on a random effect b_{ij}, the intensity function for type j events is (Cook, Lawless,

2007)

$$\lambda_{ij}(t|\boldsymbol{H_i}(t), b_{ij}) = \lim_{\Delta t \downarrow 0} \frac{P(\Delta N_{ij}(t) = 1|\boldsymbol{H_i}(t), b_{ij})}{\Delta t}. \tag{4.1}$$

I assume that $\{N_{ij}(t), t \geq 0\}|\boldsymbol{x_i}, \boldsymbol{z_i}, b_{ij}$ is an independent Poisson process with the conditional intensity formulated as

$$\lambda_{ij}(t|\boldsymbol{x_i}, \boldsymbol{z_i}, b_{ij}) = \exp\{\boldsymbol{x_i'\alpha_j} + \beta_{0j}(t) + z_{i1}\beta_{1j}(t) + \cdots + z_{iq}\beta_{qj}(t) + b_{ij}\}. \tag{4.2}$$

$\boldsymbol{\alpha_j}$ is a vector of unknown regression coefficients and $\boldsymbol{x_i'\alpha_j}$ represents the linear component of the log intensity. $\beta_{0j}(t), \beta_{1j}(t) \ldots, \beta_{qj}(t)$ are unknown smooth functions: $\beta_{0j}(t)$ is the log baseline intensity and the rest are time-varying coefficients of the covariates in $\boldsymbol{z_i}$. I assume that $\boldsymbol{b_i} = (b_{i1}, \ldots, b_{iJ})'$ is an independent and identical normal variate with density $p(\boldsymbol{b_i}|\boldsymbol{\Sigma}) = N_J(\boldsymbol{0}, \boldsymbol{\Sigma})$. The variances in $\boldsymbol{\Sigma}$ characterize extra-Poisson variation and association of event counts from each process, and the covariances characterize association between different event types. This frailty distribution allows separate parameters for variances and covariances, thus providing flexibility for the dependence structure among event types.

To model the baseline intensity and time-varying coefficients, I assume that $\beta_{lj}(t)$, $0 \leq l \leq q$, can be approximated by polynomial splines of degree v with equally spaced knots $t_{\min} = \zeta_{l0} < \zeta_{l1} < \cdots < \zeta_{l,k_l} = t_{\max}$ over the domain of time t. Let $B_{lk}(t)$ denote the value at t of the kth B-spline, and $\boldsymbol{B_l}(t) = (B_{l1}(t), \ldots, B_{l,K_l}(t))'$ denote $K_l = k_l + v$ B-spline basis functions. The parameterization used for splines is

$$\beta_{lj}(t) = \sum_{k=1}^{K_l} B_{lk}(t)\beta_{klj} = \boldsymbol{B_l'}(t)\boldsymbol{\beta_{lj}}. \tag{4.3}$$

$\boldsymbol{\beta_{lj}} = (\beta_{1lj}, \ldots, \beta_{K_l,lj})'$ is a vector of unknown regression coefficients for splines. To ensure enough flexibility for the fit, a moderately large number of knots should be chosen. However, this commonly leads to severe overfitting. Eilers, Marx (1996) proposed a frequentist solution based on a roughness penalty. Let Δ denote the difference operator such that the first

order difference of β_{klj} is $\Delta\beta_{klj} = \beta_{klj} - \beta_{k-1,lj}$, the second order difference is $\Delta^2\beta_{klj} = \beta_{klj} - 2\beta_{k-1,lj} + \beta_{k-2,lj}$, etc. In general, a roughness penalty based on the rth order difference

$$\text{pen}(\lambda_{ij}) = \frac{\lambda_{lj}}{2} \sum_{k=r+1}^{K_l} (\Delta^r \beta_{klj})^2 \tag{4.4}$$

is subtracted from the log likelihood and parameters are estimated via maximizing the resulting penalized function. The smoothness of a spline fit is controlled by the smoothing parameter λ_{lj}: by increasing λ_{lj}, the amount of roughness penalty increases and the smoothness is tuned accordingly. The smoothing parameter has to be selected by cross-validation or optimizing some information criterion. In our setting, the computational cost grows as the number of time-varying coefficients increases. Sometimes the procedure fails in practice since no optimal solution for λ_{lj} can be found (Lang, Brezger, 2004). Another drawback is that the frequentist approach assumes λ_{lj} to be a known fixed quantity, and therefore underestimates the variances for spline parameters in $\boldsymbol{\beta_{lj}}$, yielding narrow confidence intervals for $\beta_{lj}(t)$.

The model described in this chapter uses a Bayesian approach for penalized splines by replacing the roughness penalty in (4.4) with a random walk of the same order. For example, the first order difference $\Delta\beta_{klj}$ corresponds to a first order random walk $\beta_{klj} = \beta_{k-1,lj} + \epsilon_{klj}$ with $\epsilon_{klj} \sim N(0, 1/\tau_{lj})$. In general, I assume that, for $r+1 \leq k \leq K_l$, $\Delta^r\beta_{klj}$ follows an independent and identical normal distribution $N(0, 1/\tau_{lj})$. Let $\boldsymbol{D_l}$ be a matrix such that $\boldsymbol{D_l}\boldsymbol{\beta_{lj}} = (\Delta^r\beta_{r+1,lj}, \ldots, \Delta^r\beta_{K_l,lj})'$, then the prior for $\boldsymbol{\beta_{lj}}$ can be equivalently written as

$$p(\boldsymbol{\beta_{lj}}|\tau_{lj}) \propto \tau_{lj}^{\rho(\boldsymbol{D_l})/2} \exp\left\{-\frac{\tau_{lj}}{2}\boldsymbol{\beta}_{\boldsymbol{lj}}'(\boldsymbol{D_l}'\boldsymbol{D_l})\boldsymbol{\beta_{lj}}\right\}. \tag{4.5}$$

The prior is a multivariate normal distribution with mean $\mathbf{0}$ and precision matrix $\tau_{lj}(\boldsymbol{D_l}'\boldsymbol{D_l})$, where $\boldsymbol{D_l}'\boldsymbol{D_l}$ is rank deficient with rank $\rho(\boldsymbol{D_l}) = K_l - r$. The unknown τ_{lj}, often called the roughness penalty parameter, plays the role of λ_{lj} as in the frequentist approach. I assign

to τ_{lj} a conjugate Gamma prior $\text{Ga}(a, b)$; i.e.,

$$p(\tau_{lj}) = \frac{b^a}{\Gamma(a)} \tau_{lj}^{a-1} \exp(-b\tau_{lj}). \tag{4.6}$$

The hyperparameters a and b need to be small to yield a non-informative prior: for example, $a = 1$ and $b = 0.005$, or $a = b = 0.001$.

It should be noted that the smoothness of a spline fit depends considerably on the choice of a and b (Brezger, Lang, 2006). Their connection can be demonstrated by integrating τ_{lj} from the joint posterior. Let $\boldsymbol{\Omega}$ denote all the parameters in the model, and $\boldsymbol{\Omega}^{-lj}$ denote all the parameters excluding $(\boldsymbol{\beta_{lj}}, \tau_{lj})$. Let $L(\boldsymbol{\Omega})$ be the likelihood and $p(\boldsymbol{\Omega}^{-lj})$ be the joint prior for parameters in $\boldsymbol{\Omega}^{-lj}$, then the joint posterior is $L(\boldsymbol{\Omega})\, p(\boldsymbol{\Omega}^{-lj})\, p(\boldsymbol{\beta_{lj}}|\tau_{lj})\, p(\tau_{lj})$. By integrating τ_{lj} out, I obtain the log marginal posterior

$$\log \int L(\boldsymbol{\Omega}) p(\boldsymbol{\Omega}^{-lj}) p(\boldsymbol{\beta_{lj}}|\tau_{lj}) p(\tau_{lj})\, \mathrm{d}\tau_{lj} = \log L(\boldsymbol{\Omega}) + \log p(\boldsymbol{\Omega}^{-lj}) - \text{pen}(a, b), \tag{4.7}$$

where

$$\text{pen}(a, b) = [a + \rho(\boldsymbol{D_l})/2] \log \left[1 + \frac{1}{2b} \boldsymbol{\beta}'_{\boldsymbol{lj}} (\boldsymbol{D}'_{\boldsymbol{l}} \boldsymbol{D_l}) \boldsymbol{\beta_{lj}}\right]. \tag{4.8}$$

The marginal posterior corresponds to a frequentist penalized function with a roughness penalty, which differs from that defined in (4.4). The role of b is therefore crucial: as b nears 0, the penalty (4.8) becomes extremely severe, which eventually forces a linear fit (for $r = 2$) or a quadratic fit (for $r = 3$). The influence of a is otherwise limited; Jullion, Lambert (2007) also noted that in practice, no relevant information about a could be obtained from the posterior. To address the control of b over the fit, I adopt a robust specification by assigning hyperpriors in a further stage of the hierarchy. The hyperparameter b is instead

reparameterized as $b = a\delta_{lj}$, while δ_{lj} is assigned to a Gamma hyperprior $\text{Ga}(a_\delta, b_\delta)$, that is,

$$p(\tau_{lj}|\delta_{lj}) \propto \delta_{lj}^{a}\, \tau_{lj}^{a-1}\, \exp(-a\delta_{lj}\tau_{lj}), \tag{4.9}$$

$$p(\delta_{lj}) \propto \delta_{lj}^{a_\delta - 1}\, \exp(-b_\delta \delta_{lj}). \tag{4.10}$$

I choose small values (e.g., 0.001) for a_δ and b_δ to yield a non-informative hyperprior.

4.3 Posterior Inference Using MALA Algorithms

This section first describes the joint posterior. MALA algorithms are next introduced for sampling high-dimensional spline parameters in $\boldsymbol{\beta_{lj}}$. Since $\boldsymbol{\beta_{lj}}$ and τ_{lj} have a strong dependence in the posterior sampling, updating them separately could cause convergence issues. I therefore integrate MALA algorithms in a joint sampling scheme for block updating $(\boldsymbol{\beta_{lj}}, \tau_{lj})$.

4.3.1 The posterior

Let t_{ijk} denote the time to the kth type j event for process i. Suppose $n_{ij}(\geq 0)$ events of type j are observed at times $0 < t_{ij1} < \ldots < t_{ij,n_{ij}}$ before censoring time c_i. Let $\boldsymbol{D_i^{\text{obs}}}$ denote the observed data for process i, which includes the times and types of events, the censoring time, and the covariate values. The conditional likelihood for process i is

$$L(\boldsymbol{\alpha_j}, \boldsymbol{\beta_{lj}}, 1 \leq j \leq J, 0 \leq l \leq q; \boldsymbol{b_i}, \boldsymbol{D_i^{\text{obs}}}) = \prod_{j=1}^{J} \left\{ \prod_{k=1}^{n_{ij}} \lambda_{ij}(t_{ijk}|\boldsymbol{x_i}, \boldsymbol{z_i}, b_{ij}) \Lambda_{ij}(c_i) \right\}, \tag{4.11}$$

where $\Lambda_{ij}(c_i) = \int_0^{c_i} \lambda_{ij}(t|\boldsymbol{x_i}, \boldsymbol{z_i}, b_{ij})\, \mathrm{d}t$. Since there is no analytic solution to $\Lambda_{ij}(c_i)$, I use the trapezoid method for numerical integration (Kauermann, 2005). Consider a grid of points $0 = s_0 < s_1 < \cdots < s_m = t_{\max}$ over the domain of time. For process i, define $m_i = \min\{k : s_k \geq c_i\}$. I approximate $\Lambda_{ij}(c_i)$ by a polygon going through points $(s_k,\ \lambda_{ij}(s_k))$,

where $0 \leq k \leq m_i - 1$, and $\lambda_{ij}(s_k) = \lambda_{ij}(s_k|\boldsymbol{x_i}, \boldsymbol{z_i}, b_{ij})$. The approximation yields

$$\begin{aligned}\Lambda_{ij}(c_i) &\approx (c_i - s_{m_i-1})\frac{\lambda_{ij}(s_{m_i}) + \lambda_{ij}(s_{m_i-1})}{2} + I(m_i \geq 2)\sum_{k=1}^{m_i-1}(s_k - s_{k-1})\frac{\lambda_{ij}(s_k) + \lambda_{ij}(s_{k-1})}{2} \\ &= \frac{1}{2}\min(c_i, s_1)\lambda_{ij}(s_0) + \frac{1}{2}\sum_{k=1}^{m_i}\big[\min(c_i, s_{k+1}) - \min(c_i, s_{k-1})\big]\lambda_{ij}(s_k), \qquad (4.12)\end{aligned}$$

where $I(\cdot)$ is the indicator function.

Let $\boldsymbol{\Omega} = (\boldsymbol{\Sigma}, \boldsymbol{\alpha_j}, \boldsymbol{\beta_{lj}}, \tau_{lj}, \delta_{lj}, 1 \leq j \leq J, 0 \leq l \leq q)$, then the joint posterior is given by

$$\begin{aligned}&p(\boldsymbol{\Omega}|\boldsymbol{D_i^{\text{obs}}}, 1 \leq i \leq I) \\ &\propto \prod_{i=1}^{I} L(\boldsymbol{\alpha_j}, \boldsymbol{\beta_{lj}}, 1 \leq j \leq J, 0 \leq l \leq q; \boldsymbol{b_i}, \boldsymbol{D_i^{\text{obs}}})\, p(\boldsymbol{b_i}|\boldsymbol{\Sigma}) \\ &\quad \prod_{j=1}^{J}\prod_{l=0}^{q} p(\boldsymbol{\beta_{lj}}|\tau_{lj})\, p(\tau_{lj}|\delta_{lj})\, p(\delta_{lj}) \cdot \prod_{j=1}^{J} p(\boldsymbol{\alpha_j}) \cdot p(\boldsymbol{\Sigma}), \qquad (4.13)\end{aligned}$$

where $p(\boldsymbol{b_i}|\boldsymbol{\Sigma}) = N_J(\boldsymbol{0}, \boldsymbol{\Sigma})$, $p(\boldsymbol{\beta_{lj}}|\tau_{lj})$, $p(\tau_{lj}|\delta_{lj})$, and $p(\delta_{lj})$ are given in (4.5), (4.9) and (4.10). Since there is no restriction on $\boldsymbol{\alpha_j}$, I assign normal priors $N(0, 10^3)$ to parameters in $\boldsymbol{\alpha_j}$. A standard choice for prior $p(\boldsymbol{\Sigma})$ is the conjugate inverse-Wishart distribution. However, the prior is restrictive since a single degrees of freedom parameter controls dependences between all the elements in $\boldsymbol{\Sigma}$. The work described in this chapter uses a flexible approach based on Cholesky decompositions. Decomposition has been studied extensively for modeling the covariance matrix of longitudinal data (e.g., Pourahmadi, 1999; Chen, Dunson, 2003), and has been used for recurrent event data as well (Chen et al., 2005; Lin et al., 2015). Using decomposition, I are able to achieve an unconstrained reparameterization of $\boldsymbol{\Sigma}$ while retaining its positive-definiteness property. Given that $\boldsymbol{\Sigma}$ is real-valued, symmetric, and positive definite, the Cholesky decomposition of $\boldsymbol{\Sigma}$ is unique, having the form $\boldsymbol{\Sigma} = \boldsymbol{\Phi}\boldsymbol{\Phi}'$, where $\boldsymbol{\Phi}$ is a real-valued lower triangular matrix with positive diagonal entries (e.g., Golub, Van Loan, 1983, p. 88). For computational convenience, I use a reparameterization: $\boldsymbol{b_i} = \boldsymbol{\Phi}\boldsymbol{u_i}$, with $\boldsymbol{u_i} \sim N_J(\boldsymbol{0}, \boldsymbol{I})$. Normal priors $N(0, 10^3)$ are then assigned to the below-diagonal

entries in $\mathbf{\Phi}$, while truncated priors $N(0, 10^3)$ with a positive truncation range are assigned to the diagonal entries.

4.3.2 Sampling spline parameters

Sampling from the high-dimensional distribution of spline parameters is not trivial. In normal regression models, Lang, Brezger (2004) used a sampling technique based on numerical decompositions for band matrices (Rue, 2001). In generalized additive models, Brezger, Lang (2006) developed sampling schemes based on iteratively weighted least squares originally used for estimating generalized linear models (Fahrmeir, Lang, 2001). In the recurrent event setting discussed here, I sample spline parameters using MALA algorithms, a class of Metropolis-Hastings algorithms whose proposals exploit local properties of the target density. Compared to random-walk Metropolis algorithms, MALA provides faster convergence speed and better mixing chains while remaining simple to implement. Lambert et al. (2005) used the algorithm in a discrete life-table approach for survival analysis. However, as demonstrated later, MALA proposals depend considerably on the value of roughness penalty parameters. Separately updating spline and roughness penalty parameters could cause convergence issues. To remedy the issue, I integrate MALA proposals in a joint sampling scheme for block updating spline and roughness penalty parameters.

Let $\pi(\boldsymbol{\beta_{lj}})$ denote the full conditional for $\boldsymbol{\beta_{lj}}$, that is, $\pi(\boldsymbol{\beta_{lj}}) \propto L(\boldsymbol{\beta_{lj}})p(\boldsymbol{\beta_{lj}}|\tau_{lj})$, where $L(\boldsymbol{\beta_{lj}})$ is a product of likelihoods in (4.11) concerning $\boldsymbol{\beta_{lj}}$ only. Let ∇ denote the gradient operator, and $\boldsymbol{\beta}_{\boldsymbol{lj}}^t$ denote the state of the chain at iteration t. The candidate $\boldsymbol{\beta}_{\boldsymbol{lj}}^*$ for the next state is obtained by a random generation from a multivariate normal proposal

$$q(\boldsymbol{\beta}_{\boldsymbol{lj}}^t, \boldsymbol{\beta}_{\boldsymbol{lj}}^*) \sim N_{K_l}\left(\boldsymbol{\beta}_{\boldsymbol{lj}}^t + \frac{h}{2}\boldsymbol{M}\nabla \log \pi(\boldsymbol{\beta}_{\boldsymbol{lj}}^t),\, h\boldsymbol{M}\right), \tag{4.14}$$

and the candidate is accepted with probability

$$\alpha(\boldsymbol{\beta}_{\boldsymbol{lj}}^{t}, \boldsymbol{\beta}_{\boldsymbol{lj}}^{*}) = \frac{L(\boldsymbol{\beta}_{\boldsymbol{lj}}^{*})\, p(\boldsymbol{\beta}_{\boldsymbol{lj}}^{*}|\tau_{lj}^{t})}{L(\boldsymbol{\beta}_{\boldsymbol{lj}}^{t})\, p(\boldsymbol{\beta}_{\boldsymbol{lj}}^{t}|\tau_{lj}^{t})} \cdot \frac{q(\boldsymbol{\beta}_{\boldsymbol{lj}}^{*}, \boldsymbol{\beta}_{\boldsymbol{lj}}^{t})}{q(\boldsymbol{\beta}_{\boldsymbol{lj}}^{t}, \boldsymbol{\beta}_{\boldsymbol{lj}}^{*})}. \tag{4.15}$$

The proposal contains two unspecified tuning parameters: a positive real h that controls the length of proposed jumps and a symmetric positive definite matrix $\boldsymbol{M}$ that controls the direction of jumps. Roberts, Rosenthal (2001) have shown that the value of h that yields optimal asymptotic efficiency has an acceptance rate equal to 0.574. In practice, an algorithm with an acceptance rate between 0.31 and 0.81 is at least 80% efficient. MALA typically uses the identity matrix for $\boldsymbol{M}$. Yet in our setting, spline parameters associated to close knots are strongly correlated. This is due to the band structure $\boldsymbol{D}_{\boldsymbol{l}}'\boldsymbol{D}_{\boldsymbol{l}}$ in the precision matrix of prior $p(\boldsymbol{\beta}_{\boldsymbol{lj}}|\tau_{lj})$. Using the identity matrix limits the mixing of a chain and slows the convergence. Lambert et al. (2005) also noted that, under the identity matrix, the close-knots spline parameters are forced to take similar values, which causes large cross-correlations in a chain. I therefore suggest $\boldsymbol{M} = \hat{\boldsymbol{H}}_{lj}^{-1}$, where $\boldsymbol{H}_{lj}$ is the Hessian matrix of the negative $\log \pi(\boldsymbol{\beta}_{\boldsymbol{lj}})$ and can be estimated using a "pilot" chain.

Updating spline parameters block-wise is essential to reach convergence for the whole chain. However, separately updating $\boldsymbol{\beta}_{\boldsymbol{lj}}$ and the associated roughness penalty parameter τ_{lj} has yet to provide satisfactory convergence and mixing results (Knorr-Held, Rue, 2002; Brezger, Lang, 2006). The reason is that the prior for $\boldsymbol{\beta}_{\boldsymbol{lj}}$ depends on τ_{lj} through its precision matrix $\tau_{lj}(\boldsymbol{D}_{\boldsymbol{l}}'\boldsymbol{D}_{\boldsymbol{l}})$. The dependence reflects on the MALA proposal (4.14) via the gradient of $\pi(\boldsymbol{\beta}_{\boldsymbol{lj}}^{t})$; i.e.,

$$\nabla \log \pi(\boldsymbol{\beta}_{\boldsymbol{lj}}^{t}) = \nabla \log L(\boldsymbol{\beta}_{\boldsymbol{lj}}^{t}) - \tau_{lj}^{t}(\boldsymbol{D}_{\boldsymbol{l}}'\boldsymbol{D}_{\boldsymbol{l}})\boldsymbol{\beta}_{\boldsymbol{lj}}^{t}. \tag{4.16}$$

It should be also noted that the posterior for τ_{lj} is right-skewed with a long tail towards extremely large values. Once the current state τ_{lj}^{t} gets into the tail, causing too large a jump proposed for $\boldsymbol{\beta}_{\boldsymbol{lj}}^{t+1}$, the chain quickly becomes stuck in the tail.

To achieve better convergence and mixing properties, this chapter uses a joint proposal (Knorr-Held, Rue, 2002) for updating spline and roughness penalty parameters in one block. I first sample the candidate τ_{lj}^* from a proposal which may depend on the current state τ_{lj}^t but not on $\boldsymbol{\beta}_{\boldsymbol{lj}}^t$. A specific proposal $q(\tau_{lj}^t, \tau_{lj}^*)$ can be constructed by setting $\tau_{lj}^* = \tau_{lj}^t \cdot z$, where $z \in [1/f, f]$ is random with density $p(z) \propto 1 + 1/z$, and $f > 1$ is a tuning parameter. One advantage of this proposal is that the ratio $q(\tau_{lj}^*, \tau_{lj}^t)/q(\tau_{lj}^t, \tau_{lj}^*) = 1$. Conditional on τ_{lj}^*, I sample the candidate $\boldsymbol{\beta}_{\boldsymbol{lj}}^*$ from the MALA proposal. This "conditional" MALA proposal is different from that given by (4.14) in terms of the gradient term; i.e.,

$$\nabla \log \pi(\boldsymbol{\beta}_{\boldsymbol{lj}}^t) = \nabla \log L(\boldsymbol{\beta}_{\boldsymbol{lj}}^t) - \tau_{lj}^*(\boldsymbol{D}_{\boldsymbol{l}}'\boldsymbol{D}_{\boldsymbol{l}})\boldsymbol{\beta}_{\boldsymbol{lj}}^t. \tag{4.17}$$

The block of candidates $(\boldsymbol{\beta}_{\boldsymbol{lj}}^*, \tau_{lj}^*)$ is accepted with probability

$$\alpha(\boldsymbol{\beta}_{\boldsymbol{lj}}^t, \tau_{lj}^t, \boldsymbol{\beta}_{\boldsymbol{lj}}^*, \tau_{lj}^*) = \frac{L(\boldsymbol{\beta}_{\boldsymbol{lj}}^*)\, p(\boldsymbol{\beta}_{\boldsymbol{lj}}^*|\tau_{lj}^*)\, p(\tau_{lj}^*|\delta_{lj}^t)}{L(\boldsymbol{\beta}_{\boldsymbol{lj}}^t)\, p(\boldsymbol{\beta}_{\boldsymbol{lj}}^t|\tau_{lj}^t)\, p(\tau_{lj}^t|\delta_{lj}^t)} \cdot \frac{q(\boldsymbol{\beta}_{\boldsymbol{lj}}^*, \boldsymbol{\beta}_{\boldsymbol{lj}}^t)}{q(\boldsymbol{\beta}_{\boldsymbol{lj}}^t, \boldsymbol{\beta}_{\boldsymbol{lj}}^*)}, \tag{4.18}$$

with $p(\tau_{lj}|\delta_{lj})$ give in (4.9), $q(\boldsymbol{\beta}_{\boldsymbol{lj}}^t, \boldsymbol{\beta}_{\boldsymbol{lj}}^*)$ conditional on τ_{lj}^*, and $q(\boldsymbol{\beta}_{\boldsymbol{lj}}^*, \boldsymbol{\beta}_{\boldsymbol{lj}}^t)$ on τ_{lj}^t.

4.4 Simulation

I conducted a simulation study with a focus on evaluating the model performance for time-varying coefficients with different curvatures in both low and high event rate data settings.

Simulated data were generated from a Poisson process with conditional intensity:

$$\lambda_{ij}(t|\boldsymbol{x_i}, \boldsymbol{z_i}, b_{ij}) = \exp\{x_{i1}\alpha_{1j} + \beta_{0j}(t) + z_{i1}\beta_{1j}(t) + z_{i2}\beta_{2j}(t) + b_{ij}\},\ 0 \le t \le 10.$$

I considered $I = 500$ processes and $J = 2$ event types. The censoring time was generated from a gamma distribution independent of the process. Each process was censored at the maximum follow-up time of 10, if not censored before. The covariate x_{i1} was generated from

a uniform distribution on $[-1, 1]$. Its coefficient was $\alpha_{11} = 0.5$ for event type $j = 1$ and $\alpha_{12} = -0.5$ for event type $j = 2$. The covariates z_{i1} and z_{i2} were binary with probability density $p(z_{i1}, z_{i2}) = (1 - p_1 - p_2)^{1-z_{i1}-z_{i2}} p_1^{z_{i1}} p_2^{z_{i2}}$, where $0 \leq z_{i1} + z_{i2} \leq 1$ and $p_1 = p_2 = 1/3$. The time-varying coefficients for event type $j = 1$ were functions of low curvature (shown in Figure 4.1a): a quadratic function $\beta_{11}(t) = -0.1(t - 4.75)^2 + 1.5$, and a sinusoidal function $\beta_{21}(t) = \sin(0.2\pi t)$. The time-varying coefficients for event type $j = 2$ were functions of high curvature (shown in Figure 4.1b):

$$\beta_{12}(t) = \frac{8}{1 + e^{-1.05(t-3.5)}} + \frac{8}{1 + e^{(t-4.5)}} + \frac{8}{1 + e^{(t-5.5)}} + \frac{8}{1 + e^{-1.05(t-6.5)}} - 16,$$
$$\beta_{22}(t) = \cos(1/3\pi(t + 1)) + 0.01t^2.$$

There were two settings for the log baseline intensities $\beta_{0j}(t)$:

(1) a low event rate setting: $\beta_{01}(t) = 0.1t - 3.35$ and $\beta_{02}(t) = 0.01(t - 4)^2 - 2.75$;

(2) a high event rate setting: $\beta_{0j}(t)$ was increased by one unit from the low event rate setting, $j = 1, 2$.

The random effect $\boldsymbol{b_i} = (b_{i1}, b_{i2})'$ was generated from a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$. The diagonal entries of $\boldsymbol{\Sigma}$ were $\sigma_{11} = \sigma_{22} = 1$. There were three settings for the off-diagonal entry, $\sigma_{12} = 0.2$, 0.5, and 0.8, corresponding to a low, moderate and high correlation between event types.

I simulated 500 replications under each parameter setting. In the low event rate setting, the average number of events per process was one for each type; in the high event rate setting, the average number of events per process was three for each type. The unknown functions were approximated by 10 cubic B-splines; the order in the prior for spline parameters was $r = 2$. I fitted each replicate data set in two ways: (1) using splines for $\beta_{lj}(t)$ only; (2) using splines for $\beta_{lj}(t)$ and α_{1j}. For each data set, a final chain of length 8,000 was generated after tuning of parameters to ensure an satisfactory acceptance rate.

Figure 4.1 and 4.2 display the simulation results of using splines to approximate $\beta_{lj}(t)$ and α_{1j} under parameter setting $\sigma_{12} = 0.2$. Figure 4.1 corresponds to low event rate data with one event per process on average, and Figure 4.2 corresponds to high event rate data with three events per process on average. For low-curvature time-varying coefficients, the bias of spline fit is typically negligible for low event rate data, except that the posterior mean (shown in Figure 4.1a) for sinusoidal function $\beta_{21}(t)$ does deviate slightly at the end of the follow-up time. Similar bias for $\beta_{21}(t)$ also occurs for low event rate data with $\sigma_{12} = 0.5$ and 0.8. However, as shown in Figure 4.2a, such bias is no longer noticeable for high event rate data. The simulation results for high-curvature time-varying coefficients are similar: small bias occurs around the bumps of $\beta_{12}(t)$ and $\beta_{22}(t)$ for low event rate data, and is greatly diminished for high event rate data. The spline fits for the log baseline intensities are overall close to the true value, though there is some bias in $\beta_{02}(t)$ at the boundaries. For constant coefficients α_{1j}, event rate does not affect the spline fit—no noticeable bias for α_{1j} exists under either event rate setting.

Tables 4.1 and 4.2 show the simulation results for $\mathbf{\Sigma}$ and α_{1j}. The relative bias (RB; %) was obtained from the bias divided by the true parameter value. Its absolute value is below 2% for low event rate data (except for $\sigma_{12} = 0.2$) and even smaller for high event rate data. The mean of standard error (SEM; the mean of posterior standard deviations) is extremely close to the empirical standard error (SE; the standard error of posterior means). The coverage probability (CP; %) for 95% credible interval is roughly at the nominal level.

In summary, the proposed model performs well under various settings. The bias is small for time-varying coefficients of low and high curvature as well as for parametric coefficients. The credible intervals have appropriate coverage probabilities.

Table 4.1: Simulation results for α_{1j} and $\mathbf{\Sigma}$. Splines were used to approximate $\beta_{lj}(t)$.

		Low Event Rate					High Event Rate				
	True	Mean	RB[a]	SE[b]	SEM[c]	CP[d]	Mean	RB	SE	SEM	CP
α_{11}	0.5	0.504	0.8	0.124	0.124	94.4	0.500	0.1	0.095	0.101	97.2
α_{12}	−0.5	−0.507	1.4	0.125	0.122	94.4	−0.502	0.4	0.103	0.101	94.6
σ_{11}	1.0	1.010	1.0	0.077	0.076	96.0	1.007	0.7	0.053	0.055	94.8
σ_{12}	0.2	0.189	−5.4	0.101	0.103	95.0	0.201	0.3	0.069	0.070	95.6
σ_{22}	1.0	1.016	1.6	0.076	0.076	93.6	1.012	1.2	0.051	0.054	96.8
α_{11}	0.5	0.492	−1.7	0.115	0.124	96.0	0.505	0.9	0.101	0.103	95.0
α_{12}	−0.5	−0.505	1.0	0.123	0.123	94.8	−0.509	1.8	0.107	0.101	92.6
σ_{11}	1.0	1.007	0.7	0.077	0.075	93.0	1.012	1.2	0.056	0.055	92.8
σ_{12}	0.5	0.494	−1.2	0.091	0.090	96.0	0.500	0.0	0.060	0.059	93.6
σ_{22}	1.0	1.019	1.9	0.076	0.076	94.0	1.006	0.6	0.053	0.054	95.2
α_{11}	0.5	0.510	1.9	0.124	0.124	94.2	0.503	0.5	0.101	0.101	93.8
α_{12}	−0.5	−0.510	1.9	0.119	0.122	95.0	−0.501	0.2	0.102	0.101	93.8
σ_{11}	1.0	1.011	1.1	0.076	0.074	93.4	1.000	−0.0	0.084	0.053	94.0
σ_{12}	0.8	0.796	−0.4	0.071	0.068	93.8	0.789	−1.4	0.116	0.039	92.2
σ_{22}	1.0	1.007	0.7	0.074	0.073	95.2	1.007	0.7	0.053	0.053	94.4

[a] Relative bias (%), obtained from the bias divided by the true value.

[b] Empirical standard error; i.e., the standard error of posterior means.

[c] Mean standard error; i.e., the mean of posterior standard deviations.

[d] Coverage probability (%) for 95% credible intervals.

Table 4.2: Simulation results for $\mathbf{\Sigma}$. Splines were used to approximate $\beta_{lj}(t)$ and α_{1j}.

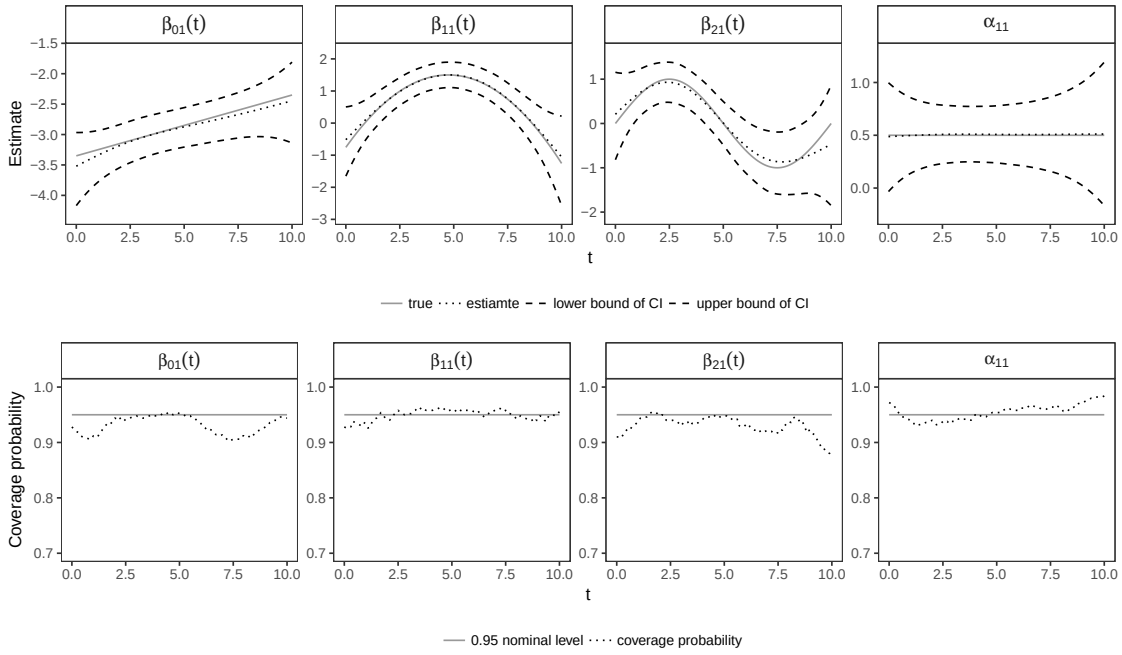
		Low Event Rate					High Event Rate				
	True	Mean	RB[a]	SE[b]	SEM[c]	CP[d]	Mean	RB	SE	SEM	CP
σ_{11}	1.0	1.010	1.0	0.077	0.076	94.2	1.008	0.7	0.054	0.055	94.2
σ_{12}	0.2	0.188	−6.1	0.102	0.103	94.6	0.201	0.4	0.069	0.070	95.6
σ_{22}	1.0	1.016	1.6	0.076	0.076	94.2	1.013	1.3	0.051	0.054	95.8
σ_{11}	1.0	1.008	0.8	0.076	0.076	93.4	1.010	1.0	0.072	0.054	93.4
σ_{12}	0.5	0.492	−1.5	0.090	0.091	95.2	0.497	−0.6	0.088	0.059	93.6
σ_{22}	1.0	1.019	1.9	0.077	0.076	94.6	1.006	0.6	0.052	0.054	95.0
σ_{11}	1.0	1.012	1.2	0.076	0.074	93.6	1.004	0.4	0.056	0.053	93.8
σ_{12}	0.8	0.796	−0.5	0.070	0.067	94.8	0.796	−0.5	0.041	0.039	94.6
σ_{22}	1.0	1.009	0.9	0.074	0.073	94.2	1.007	0.7	0.053	0.053	94.0

[a] Relative bias (%), obtained from the bias divided by the true value.

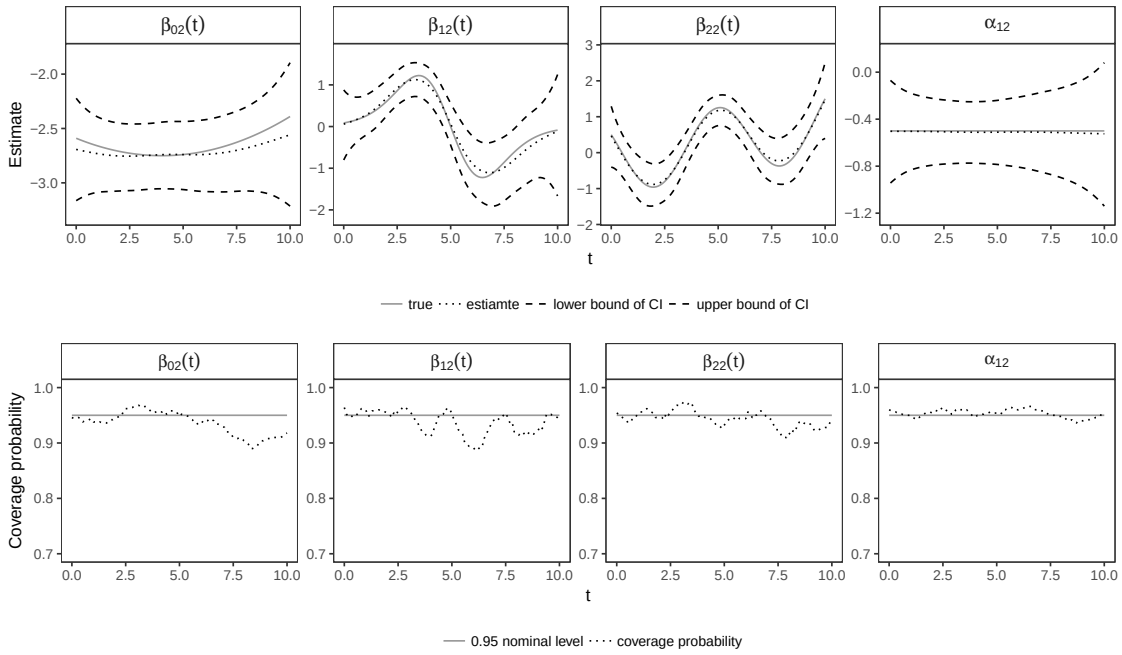
[b] Empirical standard error; i.e., the standard error of posterior means.

[c] Mean standard error; i.e., the mean of posterior standard deviations.

[d] Coverage probability (%) for 95% credible intervals.

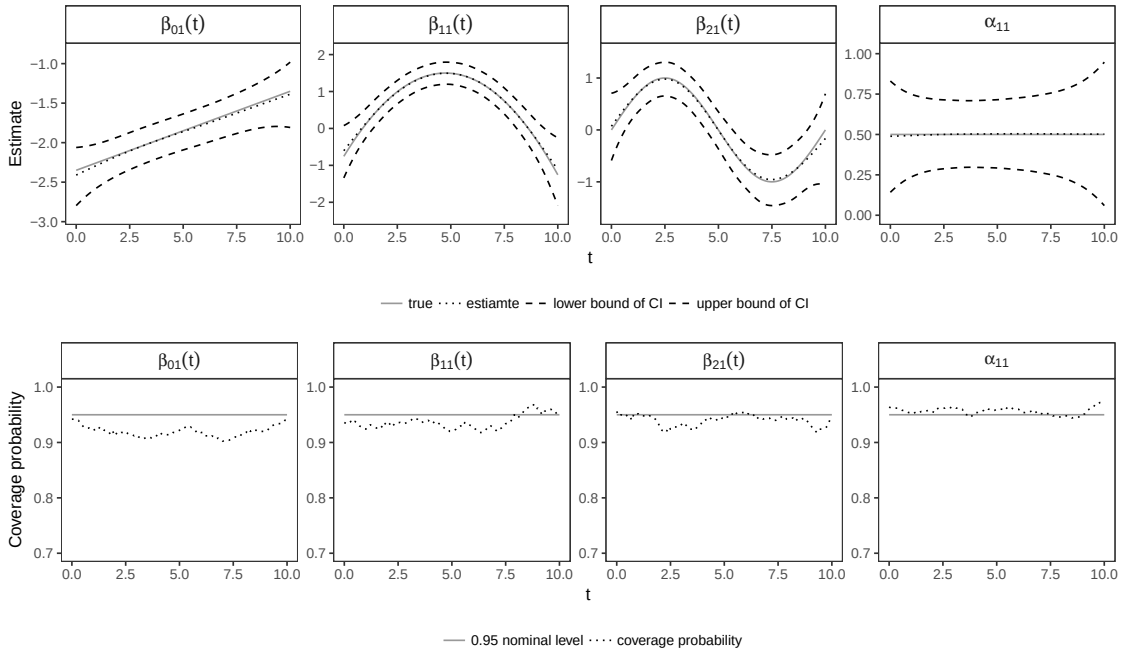


(a) Event type $j = 1$: time-varying coefficients of low curvature.

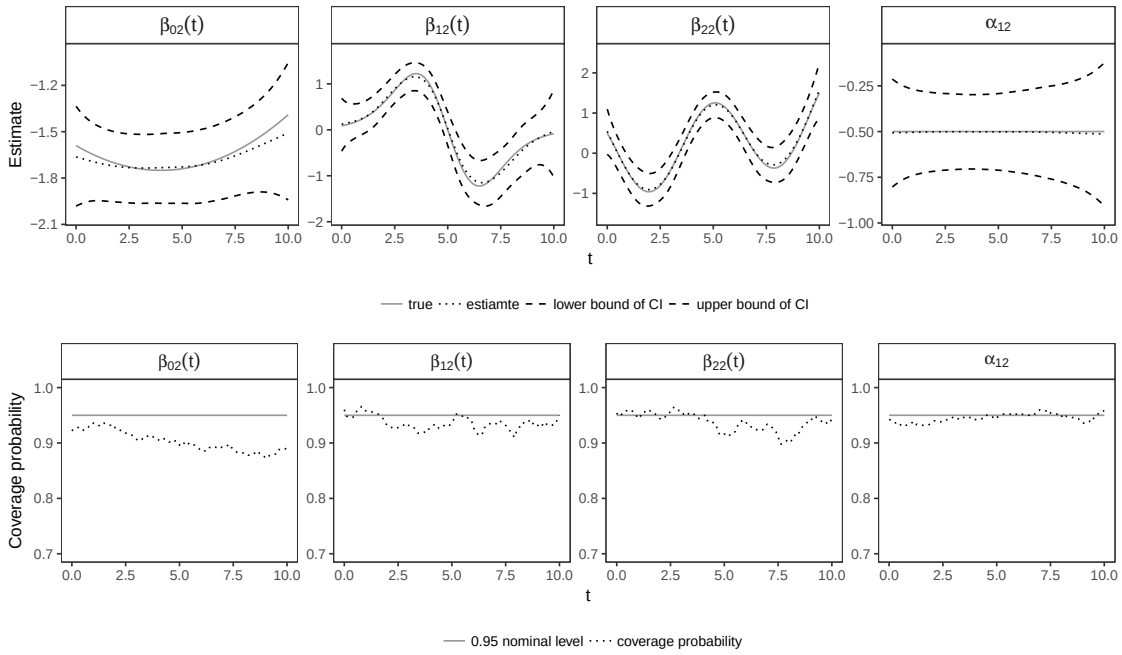


(b) Event type $j = 2$: time-varying coefficients of high curvature.

Figure 4.1: Estimates, 95% credible intervals (CIs) and coverage probability for coefficients $\beta_{lj}(t)$ and α_{1j} under the simulation setting of $\sigma_{12} = 0.2$ and low event rate.



(a) Event type $j = 1$: time-varying coefficients of low curvature.



(b) Event type $j = 2$: time-varying coefficients of high curvature.

Figure 4.2: Estimates, 95% credible intervals (CIs) and coverage probability for coefficients $\beta_{lj}(t)$ and α_{1j} under the simulation setting of $\sigma_{12} = 0.2$ and high event rate.

4.5 Application

Sleep time the night before driving is noted as a critical risk factor affecting driver fatigue (Hanowski et al., 2007). Literature also shows that whether or not a driver is well rested prior to a long on-duty shift directly impacts driver fatigue and driving performance (Banks, Dinges, 2007; Lim, Dinges, 2008, 2010). I used the approach developed in this chapter to assess the time-varying pattern of driving performance and the effect of off-duty sleep time on this temporal profile.

I applied the proposed method to data collected by the Commercial Truck Driver Naturalistic Driving Study (Blanco et al., 2011). The study recruited 100 drivers from four for-hire trucking companies. Each driver drove an instrumented truck for about one month. Driving data were recorded at high frequency by unobtrusive data-collection equipment installed on the trucks. Additionally, each driver was required to report the beginning and ending times for on- and off-duty activities in a daily activity register form. The form included activities such as sleep, break, driving, etc.

Studies have shown that total off-duty sleep time is considered insufficient if the total off-duty sleep duration is less than 7 hours (Van Dongen et al., 2003b,a). In the following analyses, I considered sleep time as normal if the off-duty sleep duration was between 7 and 9 hours, and sleep time as abundant if it was more than 9 hours. Events of interest were safety-critical events identified in the study. Given that a very small number of crashes and near-crashes were observed, the first event type in our analyses combined crashes, near-crashes and crash-relevant conflicts (CRC); the second event type was unintentional lane deviation (ULD). Driving performance was assessed in two time scales: driving time since a shift started (see Section 4.5.1) and driving time from one break until the next break in a shift (see Section 4.5.2).

4.5.1 Within-shifts driving performance

To evaluate the impact of sleep time on the succeeding driving performance in a long shift, I organized the data into pairs of an off-duty period followed by an on-duty period. The off-duty duration was sufficiently long (at least 10 hours or including sleep time for more than 7 hours) such that the beginning of a following on-duty period was a "fresh start" for the driver. Since some off-duty periods spanned several days or even weeks, sleep duration in a 12-hour time window right before an on-duty period was used in analysis. For each on-duty period, episodes of driving tasks were connected in time, yielding a continuous driving period in a shift. Due to the 11-hour driving limit set by the Federal Motor Carrier Safety Administration, driving time in a shift longer than 11 hours was truncated. I refer to Liu et al. (2017) for more details on data preparation.

There were 1,880 shifts in the analysis data: 20.6% (388) of the shifts had insufficient sleep ($<$ 7 hours), 58.2% (1,095) had normal sleep (7–9 hours), and 21.1% (397) had abundant sleep ($\geq$ 9 hours). Figure 4.3 displays the event rate in the 1st–11th driving hours from the beginning of a shift by sleep time group. For the ith driving hour, where $i = 1, \ldots, 11$, the number of events was counted; the exposure (time in hours) was the sum of all shifts' driving duration that occurred in the ith driving hour; the event rate corresponding to the ith driving hour was obtained from the event count divided by the exposure. As shown in Figure 4.3, for crashes, near-crashes and CRC, the difference between insufficient, normal and abundant sleep time is not noticeable until the 10th–11th driving hours. For ULD, the event rate of insufficient sleep is above that of normal sleep in the 8th–11th driving hours, and the rate of abundant sleep also climbs in the last 2 driving hours.

I applied the proposed method to the data, where each continuous driving period in a shift was treated as an event process. Two dummy variables were created for indicating insufficient and abundant sleep time, and were assumed to have time-varying coefficients. The baseline intensities (normal sleep) and time-varying coefficients were estimated by 10 cubic B-splines; the order in the prior for spline parameters was $r = 2$. After tuning of parameters to ensure

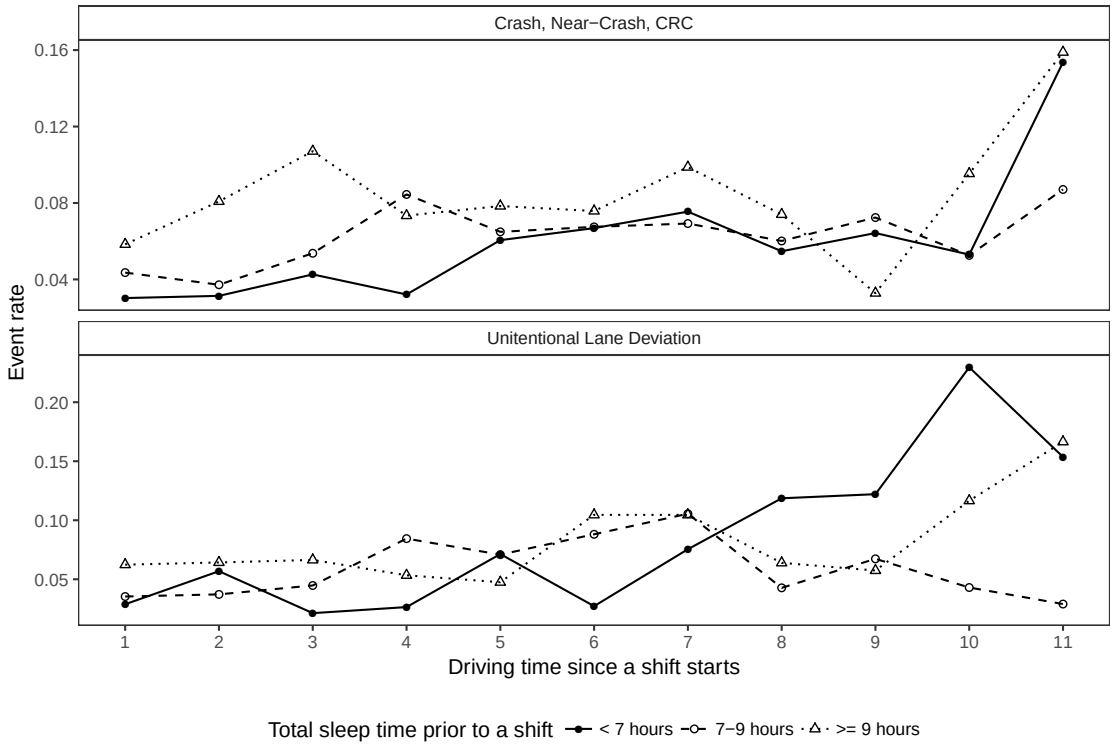


Figure 4.3: Within-shifts driving performance: the event rate in the 1st–11th driving hours by sleep time group.

Table 4.3: The variations and association for event types. σ_{11}: standard deviation of crash, near-crash and CRC; σ_{22}: standard deviation of ULD; ρ_{12}: correlation between the two event types.

	Within-shifts				Between-breaks			
Parameter	Mean	SD[a]	2.5%[b]	97.5%[b]	Mean	SD	2.5%	97.5%
σ_{11}	2.166	0.104	1.974	2.386	2.618	0.220	2.203	3.120
σ_{22}	1.366	0.058	1.256	1.484	1.775	0.146	1.520	2.102
ρ_{12}	0.888	0.022	0.842	0.930	0.810	0.053	0.697	0.903

[a] Standard deviation of posterior samples.
[b] 2.5% and 97.5% quantiles of posterior samples.

a satisfactory acceptance rate, a final chain of length 20,000 was generated. Table 4.3 shows the results for the dependence structure between the two event types. It shows significant evidence that crashes, near-crashes and CRC are positively correlated with ULD.

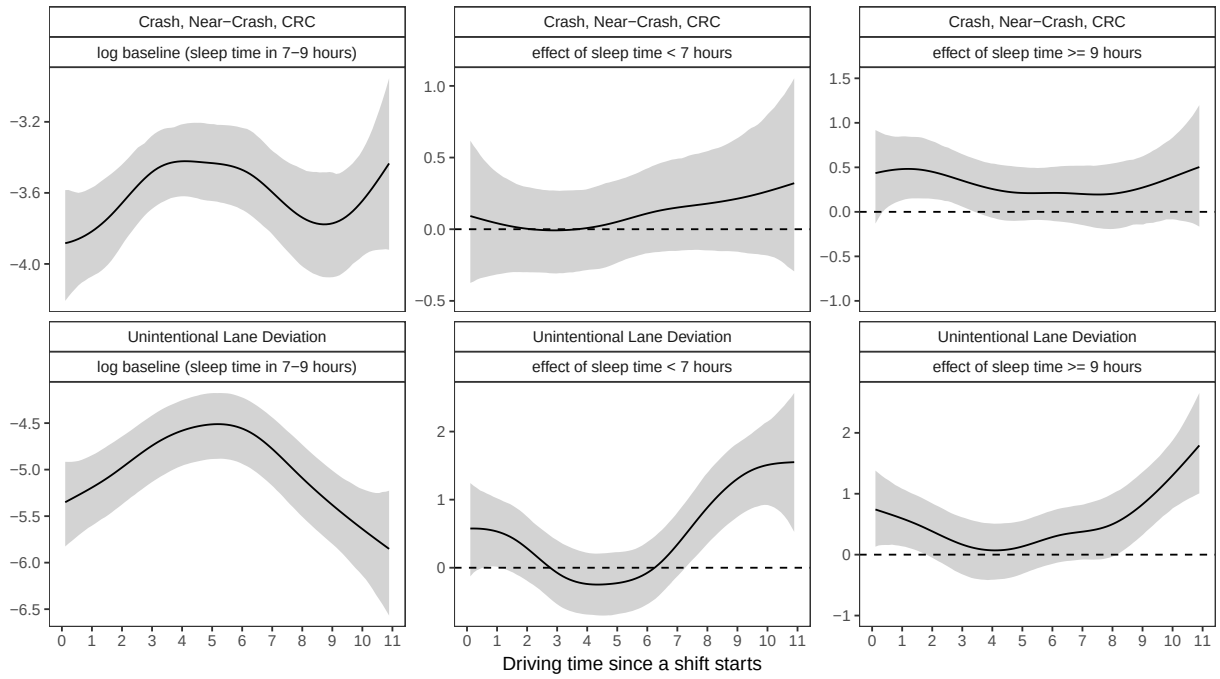


Figure 4.4: Within-shifts driving performance. The log baseline (sleep time in 7–9 hours) intensities and time-varying effects of insufficient (sleep time < 7 hours) and abundant sleep (sleep time ≥ 9 hours): posterior mean (solid line), 95% pointwise credible interval (shaded gray area).

Figure 4.4 shows the log baseline intensities and time-varying coefficients evaluated at the posterior mean along with 95% pointwise credible intervals. The time-varying coefficient of insufficient/abundant sleep gives a temporal profile of the difference between the insufficient/abundant and normal sleep groups. For crashes, near-crashes, and CRC, Figure 4.4 does not show significant difference between the three sleep time groups. However, for ULD events, Figure 4.4 shows that the coefficient functions of both insufficient and abundant sleep start to rise above zero after the 8th driving hour and keep increasing until the end of the shift. The results indicate that, compared to normal sleep time, drivers with insufficient and abundant sleep prior to a shift have a significantly higher ULD risk after 8 hours of driving. Additionally, Figure 4.4 shows that, for ULD events, there is a declining trend in the log baseline (normal sleep) intensity in the later part of a driving shift.

Literature has shown that ULD is a measure of driving performance decrement and that it is sensitive to driving fatigue (Hanowski et al., 2008; Van Dongen et al., 2010). However, it should be noted that many other factors also affect driving performance in a long shift—one important factor being drivers' break behavior. As shown in Liu et al. (2017), drivers with normal sleep time tended to have longer breaks after the 8th driving hour, which could cause the baseline intensity to decline in the same time range. Compared to normal sleep time, drivers with abundant sleep had much shorter breaks over the whole driving time, which could be the reason for the corresponding coefficient function to climb in the later part of a driving shift. After the 8th driving hour, drivers with insufficient sleep had the longest break length among the three sleep time groups. However, within the same time range, their coefficient function is above zero and keeps increasing as driving goes on. This implies that longer breaks in the later part of a driving shift are not sufficient to compensate for lack of sleep prior to the shift.

4.5.2 Between-breaks driving performance

The previous analysis assesses the temporal profile of driving performance in a long shift, which is shown to have a complex relationship between sleep time and drivers' break activities. In this section, I evaluate driving performance from one shift break until the next shift break. I considered a break as a short-time activity (no more than 5 hours) which does not involve driving or on-duty work. Between-breaks driving time longer than 3 hours was truncated due to insufficient exposure. Of the 1,286 between-breaks driving sections extracted from the 1,880 shifts described in Section 4.5.1, 14.3% (184) were associated with insufficient sleep, 60.5% (778) were associated with normal sleep, and 25.2% (324) were associated with abundant sleep.

Figure 4.5 displays the between-breaks event rate in every half an hour for these three sleep time groups. For crashes, near-crashes, and CRC, the event rate of insufficient sleep starts to rise above the other two groups from the second driving hour after a break; the rate of

abundant sleep stays above the normal group in the first 2 driving hours. For ULD, the insufficient sleep group has a substantially higher event rate than the other two groups over the whole between-breaks driving section, while the event rates of both the normal and abundant sleep groups remains low at around 0.1.

I applied the proposed method to the between-breaks data, where each between-breaks driving section was treated as an event process. The fitted dependence structure between the two event types are in Table 4.3. The event types are also positively correlated.

Figure 4.6 shows the log baseline (normal sleep) intensities and time-varying coefficients evaluated at the posterior mean with 95% pointwise credible intervals. Due to the short length of between-breaks driving sections, the intensities for the normal sleep group do not contain significant time-varying patterns. However, for crashes, near-crashes and CRC, the coefficient function of insufficient sleep starts to rise above zero after 1.5 driving hours and

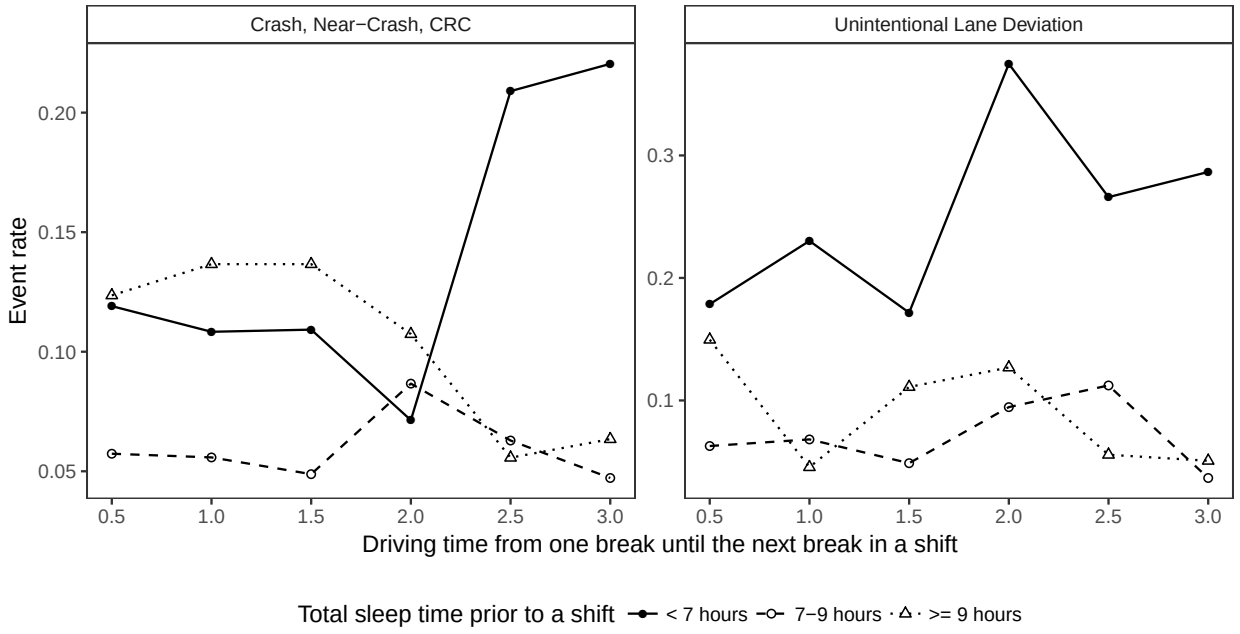


Figure 4.5: Between-breaks driving performance: the event rate in every half an hour by sleep time group.

keeps increasing as driving goes on; the coefficient function of abundant sleep is significantly greater than zero in the first 2 driving hours. These results are consistent with Figure 4.5. This indicates that, after taking a break, drivers with insufficient sleep time have a higher risk of crashes, near-crashes and CRC after 1.5 driving hours as compared to those with normal sleep time. Drivers with abundant sleep also have a significantly higher risk in the first 2 driving hours after a break, though there is no sufficient evidence showing the trend of their risk after the second driving hour. For ULD events, there is no significant result for the coefficient of abundant sleep. However, the coefficient function of insufficient sleep remains significantly above zero over the entire between-breaks driving section. This suggests that, compared to normal sleep time, drivers with insufficient sleep prior to a shift have a higher ULD risk even after taking a break. This result is consistent with that found in Section 4.5.1

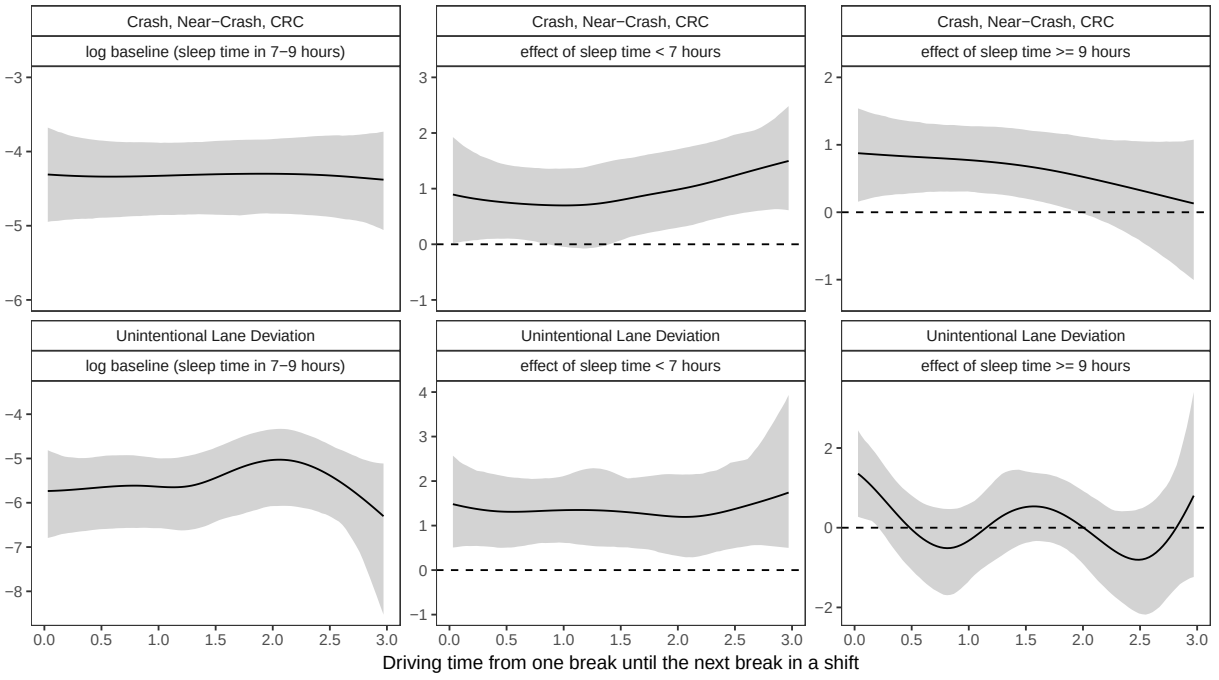


Figure 4.6: Between-breaks driving performance. The log baseline (sleep time in 7–9 hours) intensities and time-varying effects of insufficient (sleep time < 7 hours) and abundant sleep (sleep time ≥ 9 hours): posterior mean (solid line), 95% pointwise credible interval (shaded gray area).

for the difference in ULD risk between insufficient and normal sleep groups.

4.6 Summary and Discussion

Assessing the temporal profiles for recurrence rate and covariate effects is of common interest in studies with multi-type recurrent events. This chapter proposes a Bayesian joint model based on multivariate log-normal frailties for recurrent event data and incorporates time-varying coefficients for analyzing the temporal covariate effects. The Bayesian penalized splines approach is used to achieve smooth estimation for both time-varying coefficients and the baseline intensity. Adopting a robust prior specification approach provides the desired flexibility for the spline fit and achieves an objective fitting procedure without the need for the subjective choice of hyperparameters. Furthermore, MALA algorithms integrated in a joint sampling scheme help achieve better convergence and mixing properties and address the issue of strong dependence between spline and roughness penalty parameters in the posterior sampling. The simulation study demonstrates a good model performance for estimating low- and high-curvature time-varying coefficients; the model performs consistently well for data with low and high event rates.

I applied the proposed model to data from the Commercial Truck Driver Naturalistic Driving Study. The results show that unintentional lane deviation is positively correlated with crashes, near-crashes and crash relevant conflicts. The results from analyzing within-shifts data reveal that lack of sleep has a negative impact on driving performance, measured by unintentional lane deviation, after 8 hours of driving in a shift. The results also imply a complex relationship between sleep time, the driver's choice to take breaks while on duty, and driving performance. The analysis for between-breaks data shows consistent evidence that, after taking a break, insufficient sleep still corresponds to a significantly higher intensity for unintentional lane deviation and crashes, near-crashes and crash relevant conflicts. In addition to the findings for commercial truck drivers, the results provide an illustration and imply a wide applicability of our method to both epidemiological and clinic studies.

Chapter 5 Summary and Discussion

Assessing temporal profiles for recurrence rate and covariate effects is of common interest in many research. This dissertation has developed time-varying coefficient models based on penalized splines under various recurrent event data settings.

Chapter 2 introduces a mixed Poisson process model based on gamma frailties under the single type recurrent event setting, and incorporates time-varying coefficients for evaluating the temporal covariate effects. The penalized B-splines approach is used to obtain smooth estimates for both time-varying coefficients and the log baseline intensity. The trapezoid method has been adopted to approximate the cumulative intensity in the likelihood. An EM algorithm has been developed for parameter estimation. One issue with penalized splines is that the fit is conditional on the smoothing parameter. Procedures such as cross-validation or optimizing some performance criterion are commonly used to select the value for the smoothing parameter. These procedures could be computationally demanding as the number of time-varying coefficients increases.

Built upon the mixed Poisson process model in Chapter 2, Chapter 3 develops a mixed-model representation approach for penalized splines to achieve objective estimation for the smoothing parameter. By the mixed-model representation, spline coefficients are treated as random effects and smoothing parameters are to be estimated as variance components. I develop an EM algorithm embedded with penalized quasi-likelihood approximation for parameter estimation. The simulation study has confirmed that the model performs well for time-varying coefficients with different degrees of curvature under both low and high event rate scenarios. The application to the data from the Commercial Truck Driver Naturalistic Driving Study reveals that lack of sleep (duration < 7 hours) negatively impacts driving performance, as measured by unintentional lane deviations, after 8 hours of driving in a

given shift. The results also implies a complex relationship among total sleep time, the drivers choice to take breaks while on duty, and driving performance. Compared to drivers who have 7 to 9 hours' sleep prior to a shift, drivers with less than 7 hours' sleep tend to take markedly more breaks after 8 hours of driving, yet these extra amount of breaks could not prevent a higher risk; the gap between these two groups is even wider as the driving continues.

Chapter 4 proposes a Bayesian joint model based on multivariate log-normal frailties for multi-type recurrent events and incorporates time-varying coefficients for analyzing the temporal covariate effects. The Bayesian penalized splines approach is used to achieve smooth estimation for both time-varying coefficients and the baseline intensity. Adopting a robust prior specification approach provides the desired flexibility for the spline fit and achieves an objective fitting procedure without the need for the subjective choice of hyperparameters. Furthermore, Metropolis-adjusted Langevin algorithms integrated in a joint sampling scheme help to achieve better convergence and mixing properties and address the issue of strong dependence between spline and roughness penalty parameters in the posterior sampling. The simulation study demonstrates a good model performance for estimating low- and high-curvature time-varying coefficients; the model performs consistently well for data with low and high event rates. The application to the data from the Commercial Truck Driver Naturalistic Driving Study show that unintentional lane deviation is positively correlated with crashes, near-crashes and crash relevant conflicts. The results from analyzing within-shifts data is consistent with those in Chapter 3. The analysis for between-breaks data shows that, after taking a break, less than 7 hours' sleep still corresponds to a significantly higher intensity for unintentional lane deviation and crashes, near-crashes and crash relevant conflicts.

This dissertation provides crucial insight into the impact of sleep time on driving performance for commercial truck drivers and highlights the on-road safety implications of insufficient sleep (duration < 7 hours) and breaks while driving. The finding that performance decrements appear after 8 hours of driving into the shift provides important information for

technology developers aimed at fatigue detection, and for those interested in determining the role that fatigue may play in a road crash.

The methodologies developed in this dissertation provide a tool to reveal temporal profile for recurrent events. I've addressed several key methodology issues including objective selection of smoothing parameters for time-varying coefficients, explicit estimation of the baseline intensity, and characterization of the dependence structure among event types. Possible extensions for this work can be made in the future by incorporating nested frailties, adapting dependent censoring and employing machine learning techniques.

Bibliography

Driver Distraction and Inattention: Advances in Research and Countermeasures, Volume 1. 2017.

Aalen Odd. Nonparametric Inference for a Family of Counting Processes // The Annals of Statistics. 1978. 6, 4. 701–726.

Abe Takashi, Komada Yoko, Inoue Yuichi. Short Sleep Duration, Snoring and Subjective Sleep Insufficiency Are Independent Factors Associated with both Falling Asleep and Feeling Sleepiness while Driving // Internal Medicine. 2012. 51, 23. 3253–3260.

Abu-Libdeh Hasan, Turnbull Bruce W, Clark Larry C. Analysis of multi-type recurrent events in longitudinal studies; application to a skin cancer prevention trial // Biometrics. 1990. 1017–1034.

Amorim Leila D., Cai Jianwen, Zeng Donglin, Barreto Maurcio L. Regression splines in the time-dependent coefficient rates model for recurrent event data // Statistics in Medicine. 2008. 27, 28. 5890–5906.

Banks Siobhan, Dinges David F. Behavioral and Physiological Consequences of Sleep Restriction // Journal of Clinical Sleep Medicine. 2007. 3, 5. 519–528.

Berry Scott M, Carroll Raymond J, Ruppert David. Bayesian smoothing and regression splines for measurement error problems // Journal of the American Statistical Association. 2002. 97, 457. 160–169.

Blanco Myra, Hanowski Richard J., Olson Rebecca L., Morgan Justin F., Soccolich Susan A., Wu Shih-Ching, Guo Feng. The Impact of Driving, Non-driving Work, and Rest Breaks on Driving Performance in Commercial Vehicle Operations. Washington, DC, 2011.

Breiman Leo. Fitting additive models to regression data: Diagnostics and alternative views // Computational Statistics & Data Analysis. 1993. 15, 1. 13–46.

Breslow Norman E., Clayton David .G. Approximate inference in generalized linear mixed models // Journal of the American Statistical Association. 1993. 88, 421. 9–25.

Brezger Andreas, Lang Stefan. Generalized structured additive regression based on Bayesian P-splines // Computational Statistics & Data Analysis. 2006. 50, 4. 967–991.

Brumback Babette A, Rice John A. Smoothing spline models for the analysis of nested and crossed samples of curves // Journal of the American Statistical Association. 1998. 93, 443. 961–976.

Brumback Babette A., Ruppert David, Wand M. P. Variable Selection and Function Estimation in Additive Nonparametric Regression Using a Data-Based Prior: Comment // Journal of the American Statistical Association. 1999. 94, 447. 794–797.

Byar David, Blackard Clyde, Veterans Administration Cooperative Urological Research Group the. Comparisons of placebo, pyridoxine, and topical thiotepa in preventing recurrence of stage I bladder cancer // Urology. 1977. 10, 6. 556–561.

Cai Jianwen, Schaubel Douglas E. Marginal means/rates models for multiple type recurrent event data // Lifetime data analysis. 2004. 10, 2. 121–138.

Cai T, Hyndman Rob J., Wand M. P. Mixed model-based hazard estimation // Journal of Computational and Graphical Statistics. 2002. 11, 4. 784–798.

Cai Tianxi, Betensky Rebecca A. Hazard Regression for Interval-Censored Data with Penalized Spline // Biometrics. 2003. 59, 3. 570–579.

Cai Zongwu, Fan Jianqing, Li Runze. Efficient estimation and inferences for varying-coefficient models // Journal of the American Statistical Association. 2000. 95, 451. 888–902.

Carroll Raymond J, Ruppert David, Welsh Alan H. Local estimating equations // Journal of the American Statistical Association. 1998. 93, 441. 214–227.

Chen Bingshu E, Cook Richard J. The analysis of multivariate recurrent events with partially missing event types // Lifetime data analysis. 2009. 15, 1. 41–58.

Chen Bingshu E, Cook Richard J, Lawless Jerald F, Zhan Min. Statistical methods for multivariate interval-censored recurrent events // Statistics in medicine. 2005. 24, 5. 671–691.

Chen Zhen, Dunson David B. Random Effects Selection in Linear Mixed Models // Biometrics. 2003. 59, 4. 762–769.

Cleveland William S, Grosse Eric, Shyu William M. Local regression models // Statistical Models in S. 1991. 309–376.

Cook Richard J, Lawless Jerald. The statistical analysis of recurrent events. 2007.

Cox D. R. Regression Models and Life-Tables // Journal of the Royal Statistical Society: Series B. 1972. 34, 2. 187–220.

Cox David R. Partial likelihood // Biometrika. 1975. 62, 2. 269–276.

De Boor Carl. A practical guide to splines. 1978.

Dempster Arthur P., Laird Nan M., Rubin Donald B. Maximum likelihood from incomplete data via the EM algorithm // Journal of the royal statistical society: Series B. 1977. 29, 1. 1–38.

Eilers Paul H. C., Marx Brian D. Flexible Smoothing with B-splines and Penalties // Statistical Science. 1996. 11, 2. 89–102.

Eubank Randall L. Nonparametric regression and spline smoothing. 1999. 2.

Fahrmeir Ludwig, Lang Stefan. Bayesian inference for generalized additive mixed models based on Markov random field priors // Journal of the Royal Statistical Society: Series C (Applied Statistics). 2001. 50, 2. 201–220.

Fan Jianqing, Zhang Wenyang. Statistical Estimation in Varying Coefficient Models // The Annals of Statistics. 1999. 27, 5. 1491–1518.

Federal Motor Carrier Safety Administration . Large Truck and Bus Crash Facts 2015. Washington, DC, November 2016.

Federal Register . Hours of Service of Drivers. 12 2011. 76 Fed. Reg. 81134 (to be codified at 49 C.F.R. pts. 385, 386, 390, & 395).

Ford Earl S., Cunningham Timothy J., Croft Janet B. Trends in Self-Reported Sleep Duration among US Adults from 1985 to 2012 // Sleep. 2015. 38, 5. 829–832.

Friedman Jerome H. Multivariate adaptive regression splines // The Annals of Statistics. 1991. 19, 1. 1–67.

Friedman Jerome H, Silverman Bernard W. Flexible parsimonious smoothing and additive modeling // Technometrics. 1989. 31, 1. 3–21.

Friedman Michael. Piecewise Exponential Models for Survival Data with Covariates // The Annals of Statistics. 1982. 10, 1. 101–113.

Gail M. H., Santner T. J., Brown C. C. An Analysis of Comparative Carcinogenesis Experiments Based on Multiple Times to Tumor // Biometrics. 1980. 36, 2. 255–266.

Golub Gene H., Van Loan Charles F. Matrix Computations. Baltimore, MD: The Johns Hopkins University Press, 1983.

Gray Robert J. Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis // Journal of the American Statistical Association. 1992. 87, 420. 942–951.

Green Peter J, Silverman Bernard W. Nonparametric regression and generalized linear models: a roughness penalty approach. 1994.

Hanowski Richard J., Bowman Darrell, Alden Andrew, Wierwille Walter W., Carroll Robert. PERCLOS+: Moving beyond single-metric drowsiness monitors. 2008.

Hanowski Richard J, Hickman Jeffery, Fumero Maria C, Olson Rebecca L, Dingus Thomas A. The sleep of commercial vehicle drivers under the 2003 hours-of-service regulations // Accident Analysis & Prevention. 2007. 39, 6. 1140–1145.

Hastie Trevor, Loader Clive. Local Regression: Automatic Kernel Carpentry // Statistical Science. 1993. 8, 2. 120–129.

Hastie Trevor, Tibshirani Robert. Varying-Coefficient Models // Journal of the Royal Statistical Society. Series B (Methodological). 1993. 55, 4. 757–796.

Huang Jianhua Z., Wu Colin O., Zhou Lan. Varying-coefficient models and basis function approximations for the analysis of repeated measurements // Biometrika. 2002. 89, 1. 111–128.

Jayaram L., Pizzichini M. M., Cook R. J., Boulet L-P., Lemière C., Pizzichini E., Cartier A., Hussack P., Goldsmith C. H., Laviolette M., Parameswaran K., Hargreave F. E. Determining asthma treatment by monitoring sputum cell counts: effect on exacerbations // European Respiratory Journal. 2006. 27, 3. 483–494.

Jullion Astrid, Lambert Philippe. Robust specification of the roughness penalty prior distribution in spatially adaptive Bayesian P-splines models // Computational statistics & data analysis. 2007. 51, 5. 2542–2558.

Karlin Samuel, Taylor Howard M. A first course in stochastic processes. 1975. 2.

Kauermann Göran. Penalized spline smoothing in multivariable survival models with varying coefficients // Computational Statistics & Data Analysis. 2005. 49, 1. 169–186.

Klein John P. Semiparametric estimation of random effects using the Cox model based on the EM algorithm // Biometrics. 1992. 48, 3. 795–806.

Knorr-Held Leonhard, Rue Håvard. On Block Updating in Markov Random Field Models for Disease Mapping // Scandinavian Journal of Statistics. 2002. 29, 4. 597–614.

Lambert Philippe, Eilers Paul HC, others . Bayesian proportional hazards model with time-varying regression coefficients: a penalized Poisson regression approach // Statistics in Medicine. 2005. 24, 24. 3977.

Lang Stefan, Brezger Andreas. Bayesian P-splines // Journal of computational and graphical statistics. 2004. 13, 1. 183–212.

Lawless Jerald F. Regression methods for Poisson process data // Journal of the American Statistical Association. 1987. 82, 399. 808–815.

Lawless Jerald F, Nadeau Claude. Some Simple Robust Methods for the Analysis of Recurrent Events // Technometrics. 1995. 37, 2. 158–168.

Li Qing, Guo Feng, Kim Inyoung, Klauer Sheila G., Simons-Morton Bruce G. A Bayesian finite mixture change-point model for assessing the risk of novice teenage drivers // Journal of Applied Statistics. 2018. 45, 4. 604–625.

Li Qing, Guo Feng, Klauer Sheila G, Simons-Morton Bruce G. Evaluation of risk change-point for novice teenage drivers // Accident Analysis & Prevention. 2017. 108. 139–146.

Lim Julian, Dinges David F. Sleep Deprivation and Vigilant Attention // Annals of the New York Academy of Sciences. 2008. 1129, 1. 305–322.

Lim Julian, Dinges David F. A Meta-Analysis of the Impact of Short-Term Sleep Deprivation on Cognitive Variables // Psychological Bulletin. 2010. 136, 3. 375–389.

Lin D. Y., Wei L. J., 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). 2000. 62, 4. 711–730.

Lin Li-An, Luo Sheng, Chen Bingshu E, Davis Barry R. Bayesian analysis of multi-type recurrent events and dependent termination with nonparametric covariate functions // Statistical methods in medical research. 2015. 0962280215613378.

Liu Yi, Guo Feng, Hanowski Richard. A Time-Varying Coefficient Model for Evaluating Commercial Truck Driver Performance. Manuscript submitted for publication. 12 2017.

McCulloch Charles E., Neuhaus John M. Misspecifying the shape of a random effects distribution: why getting it wrong may not matter // Statistical Science. 2011. 26, 3. 388–402.

National Academies of Sciences, Engineering, and Medicine . Commercial Motor Vehicle Driver Fatigue, Long-Term Health, and Highway Safety: Research Needs. Washington, DC: The National Academies Press, 2016.

Ng Edmund, Cook Richard J. Robust inference for bivariate point processes // Canadian Journal of Statistics. 1999. 27, 3. 509–524.

Nielsen Gert G, Gill Richard D, Andersen Per Kragh, Sørensen Thorkild IA. A counting process approach to maximum likelihood estimation in frailty models // Scandinavian journal of Statistics. 1992. 25–43.

Nielsen JD, Dean CB. Regression splines in the quasi-likelihood analysis of recurrent event data // Journal of Statistical Planning and Inference. 2005. 134, 2. 521–535.

O'Sullivan Finbarr. A statistical perspective on ill-posed inverse problems // Statistical science. 1986. 1, 4. 502–518.

Pepe Margaret Sullivan, Cai Jianwen. Some Graphical Displays and Marginal Regression Analyses for Recurrent Failure Times and Time Dependent Covariates // Journal of the American Statistical Association. 1993. 88, 423. 811–820.

Pourahmadi Mohsen. Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation // Biometrika. 1999. 86, 3. 677–690.

Ripatti Samuli, Palmgren Juni. Estimation of multivariate frailty models using penalized partial likelihood // Biometrics. 2000. 56, 4. 1016–1022.

Roberts Gareth O., Rosenthal Jeffrey S. Optimal scaling for various Metropolis-Hastings algorithms // Statistical Science. 2001. 16, 4. 351–367.

Roberts Gareth O, Tweedie Richard L. Exponential convergence of Langevin distributions and their discrete approximations // Bernoulli. 1996. 341–363.

Ross Sheldon M. Stochastic processes. 1996. 2.

Rue Håvard. Fast sampling of Gaussian Markov random fields // Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2001. 63, 2. 325–338.

Ruppert David, Carroll Raymond J. Theory & Methods: Spatially-adaptive Penalties for Spline Fitting // Australian & New Zealand Journal of Statistics. 2000. 42, 2. 205–223.

Scott Linda D., Hwang Wei-Ting, Rogers Ann E., Nysse Tami, Dean Grace E., Dinges David F. The Relationship between Nurse Work Schedules, Sleep Duration, and Drowsy Driving // Sleep. 2007. 30, 12. 1801–1807.

Soccolich Susan A, Blanco Myra, Hanowski Richard J, Olson Rebecca L, Morgan Justin F, Guo Feng, Wu Shih-Ching. An analysis of driving and working hour on commercial motor vehicle driver safety using naturalistic data collection // Accident Analysis & Prevention. 2013. 58. 249–258.

Stone Charles J, Hansen Mark H, Kooperberg Charles, Truong Young K, others . Polynomial splines and their tensor products in extended linear modeling: 1994 Wald memorial lecture // The Annals of Statistics. 1997. 25, 4. 1371–1470.

Sun Liuquan, Zhou Xian, Guo Shaojun. Marginal regression models with time-varying coefficients for recurrent event data // Statistics in Medicine. 2011. 30, 18. 2265–2277.

Therneau Terry M, Hamilton Scott A. rhDNase as an example of recurrent event analysis // Statistics in Medicine. 1997. 16, 18. 2029–2047.

Tierney Luke, Kadane Joseph B. Accurate approximations for posterior moments and marginal densities // Journal of the american statistical association. 1986. 81, 393. 82–86.

Vaida Florin, Xu Ronghui. Proportional hazards model with random effects // Statistics in medicine. 2000. 19, 24. 3309–3324.

Van Dongen Hans P A, Jackson Melinda Lea, Belenky Gregory. Duration Restart Period Needed to Recycle with Optimal Performance: Phase II. Washington, DC, 2010.

Van Dongen Hans P A, Maislin Greg, Mullington Janet M., Dinges David F. The Cumulative Cost of Additional Wakefulness: Dose-Response Effects on Neurobehavioral Functions and Sleep Physiology From Chronic Sleep Restriction and Total Sleep Deprivation // Sleep. 2003a. 26, 2. 117–126.

Van Dongen Hans P A, Rogers Naomi L, Dinges David F. Sleep debt: Theoretical and empirical issues // Sleep and Biological Rhythms. 2003b. 1, 1. 5–13.

Wahba Grace. Spline models for observational data. 1990.

Wang Yuedong. Mixed effects smoothing spline analysis of variance // Journal of the Royal Statistical Society: Series B (Statistical Methodology). 1998a. 60, 1. 159–174.

Wang Yuedong. Smoothing Spline Models with Correlated Random Errors // Journal of the American Statistical Association. 1998b. 93, 441. 341–348.

Watson Nathaniel F., Badr M. Safwan, Belenky Gregory, Bliwise Donald L., Buxton Orfeu M., Buysse Daniel, Dinges David F., Gangwisch James, Grandner Michael A., Kushida Clete, Malhotra Raman K., Martin Jennifer L., Patel Sanjay R., Quan Stuart F., Tasali Esra. Joint Consensus Statement of the American Academy of Sleep Medicine and Sleep Research Society on the Recommended Amount of Sleep for a Healthy Adult: Methodology and Discussion // Sleep. 2015a. 38, 8. 1161–1183.

Watson Nathaniel F, Badr M Safwan, Belenky Gregory, Bliwise Donald L, Buxton Orfeu M, Buysse Daniel, Dinges David F, Gangwisch James, Grandner Michael A, Kushida Clete, Malhotra Raman K, Martin Jennifer L, Patel Sanjay R, Quan Stuart F, Tasali Esra. Recommended Amount of Sleep for a Healthy Adult: A Joint Consensus Statement of the American Academy of Sleep Medicine and Sleep Research Society // Sleep. 2015b. 38, 6. 843–844.

Yu Zhangsheng, Liu Lei, Bravata Dawn M, Williams Linda S, Tepper Robert S. A semi-parametric recurrent events model with time-varying coefficients // Statistics in Medicine. 2013. 32, 6. 1016–1026.