SEMIPARAMETRIC REGRESSION ANALYSIS OF LONGITUDINAL DATA WITH INFORMATIVE OBSERVATION TIMES

DENG SHIRONG

Ph.D

The Hong Kong Polytechnic University

2012

The Hong Kong Polytechnic University

Department of Applied Mathematics

Semiparametric Regression Analysis of Longitudinal Data with Informative Observation Times

DENG Shirong

A thesis submitted in partial fulfillment of

the requirements for the degree of Doctor of Philosophy

May 2012

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

____________________ (Signed)

DENG Shirong (Name of student)

Abstract

Longitudinal data often occur in a long-term study where each individual is measured repeatedly at distinct time points rather than continuous times and also the observation times and censoring times may vary from subject to subject. Many researchers have considered the analysis of such longitudinal data under the assumption that observation process is independent of response process completely or conditional on covariates, which may not be true in practice. This thesis investigates semiparametric analysis of longitudinal data when the response process is correlated with the observation times.

We develop a new class of semiparametric mean models for longitudinal data which allows for the interaction between the observation history and covariates, leaving patterns of the observation process to be arbitrary. Although panel count data is a special case of longitudinal data, it has particular features which can not be described by general longitudinal models. Thus, to analyze the panel count data, we propose a new class of flexible semiparametric regression models by incorporating the interaction between the observation history and some covariates to the mean model of the recurrent event process, without any formation restriction on the informative observation process. For inference on the regression parameters and the unknown baseline functions involved in both longtidunial data and panel count data models, spline-based least square estimation approachs are proposed, respectively, and asymptotic properties including the consistency, rate of convergence and asymptotic normality of the proposed estimators are established for both models. Simulation studies demonstrate that the proposed inference procedures perform well for both models. The analyses of a bladder tumor data are presented to illustrate the proposed methods.

Furthermore, it would be desirable to develop estimation procedures for panel count data with informative observation times, and also with time-dependent covariates and informative censoring times. Thus we extend the joint frailty models proposed by Zhao and Tong (2011) to panel count data with the time-dependent covariates and informative observation and censoring times. A novel estimating equation approach that does not depend on the distribution of frailty variables and the link function is proposed for estimation of parameters, and the asymptotic properties of the proposed estimators are established. The performance of proposed inference procedure are demonstrated by some simulation studies and illustrated by the analysis of a bladder tumor data.

Key Words Asymptotic normality; B-splines; Empirical process; Estimating equation; Informative observation process; Longitudinal data; Panel count data; Time-dependent covariates.

Acknowledgements

First and foremost, I would like to express my deepest and sincerest gratitude to my supervisor, Dr. Xingqiu Zhao for her support, encouragement, patience and guidance during my PhD graduate studies. Her abiding energy and enthusiasm in life and research has always motivated me to make continuous progress. I appreciate her rigorous scholarship, vast knowledge, active and open mind and many other excellent characters, which has influenced me so much and will be the examples for me to follow all the life. She was so kind that always programmed the academic research road from the position of me, which guided me to do the research so quickly. Without her, this thesis would't be completed.

I also owe my most sincere gratitude to my co-supervisor, Professor Wong Heung, for his guidance and encouragement and also for his teaching of the course of Statistical Inference for me.

I also would like to express my deeply gratitude to my Master's advisor, Professor Luqin Liu in Wuhan University, whose strict requirement for me made me grasp the fundamental theoretical knowledge necessary for my academic research. His guidance, encouragement, consideration had always around me even when I was in Hong Kong for my PhD study. I am also very indebted to Associate Professor Li Liu in Wuhan University, who helped me so much in simulation when I was puzzled in it and treated me as not only a student, but also a friend and influenced me through her personal statements and actions. Many thanks to Professor Yanyan Liu, who is a highly infectious person for her pertinent comments to my statistical study. I also would like to thank Professor Fuqing Gao who taught me the important course - Stochastic Process and gave me many helpful suggestions. My sincere gratitude also belongs to all the members in the statistics study group in Wuhan University, especially to Professor Yiping Zhang, Dr. Jieli Ding for their insightful comments, to Drs. Yuanshan Wu, Qinglong Yang, Yurong Chen for their friendship and help.

Furthermore, I would like to thank all the research students in the Applied Mathematics Department, for the interesting academic discussions and useful assistance over the past three years. And the staff and computer technicians in our department deserve my gratitude for providing assistance and administrative support.

Last but not least, I am forever grateful to my families for their unconditional love and support, especially, to my husband, Tao Zhao, whose love and support accompanied me through my sorrow and happiness all along the way.

Contents

Abstract i

Acknowledgements iii

List of Figures viii

List of Tables ix

Chapter 1 Introduction and Literature Review 1

1.1 Motivation 1

1.1.1 Bladder Cancer Study 1

1.2 Literature Review 4

1.2.1 Longitudinal Data 4

1.2.2 Panel Count data 8

1.2.3 B-Splines in Survival Analysis 16

1.3 Outline of Thesis 21

Chapter 2 Longitudinal Data Analysis Using B-Splines 23

2.1 Introduction 23

2.2 Statistical model 24

2.3 Inference procedure 27

2.4 Asymptotic theory 28

2.5 Simulation study 48

2.6 Application 56

Chapter 3 Panel Count Data Analysis Using Monotone B-Splines 60

3.1 Introduction 60

3.2 Statistical Models 61

3.3 Estimation Procedure 64

3.4 Asymptotic Theory 68

3.5 Simulation study 85

3.6 Application 92

Chapter 4 Panel Count Data Analysis with Time-Dependent Covariates 96

4.1 Introduction 96

4.2 Statistical Models 97

4.3 Estimation Procedure 99

4.4 Asymptotic Theory 101

4.5 Simulation Study 108

4.6 An Application 114

Chapter 5 Conclusions and Future Work 116

5.1 Conclusions 116

5.2 Further Research 119

5.2.1 Proportional partial linear intensity model for recurrent event data 119

5.2.2 New nonparametric tests for panel count data 123

Appendices 127

References **142**

List of Figures

2.1 Estimates of $\mu_0(t) = \log(t+1)$ for simulated longitudinal data with non-Poisson observation process, $h(\mathcal{F}_t, W) = H(t-)X_1$ and $\alpha = 1$ . . 57

3.1 Estimates of $\mu_0(t) = \sqrt{t}$ for simulated panel count data with Poisson and non-Poisson observation processes, $h(\mathcal{F}_t, W) = H(t-)X_1$ and $\alpha = 0.5$ 93

3.2 Estimates of $\mu_0(t) = \log(t+1)$ for simulated panel count data with Poisson and non-Poisson observation processes, $h(\mathcal{F}_t, W) = H(t-)X_1$ and $\alpha = 0.5$ 94

List of Tables

2.1 Estimation results of $(\beta_0', \alpha_0')'$ for simulated longitudinal data with non-Poisson observation process, $h(\mathcal{F}_t, W) = H(t-)X_1$, $\alpha = 1$, $\mu_0(t) = \log(t+1)$ and $n = 50$ by chosing different numbers and placements of knots.. 51

2.2 Estimation results of $(\beta_0', \alpha_0')'$ for simulated longitudinal data with non-Poisson observation process, $h(\mathcal{F}_t, W) = H(t-)X_1$ and $\mu_0(t) = \sin(t/2)$ 52

2.3 Estimation results of $(\beta_0', \alpha_0')'$ for simulated longitudinal data with non-Poisson observation process, $h(\mathcal{F}_t, W) = H(t-)X_1$ and $\mu_0(t) = \log(t+1)$ 53

2.4 Estimation results of $(\beta_0', \alpha_0')'$ for simulated longitudinal data with non-Poisson observation process, $h(\mathcal{F}_t, W) = H(t-)X_2$ and $\mu_0(t) = \sin(t/2)$ 54

2.5 Estimation results of $(\beta_0', \alpha_0')'$ for simulated longitudinal data with non-Poisson observation process, $h(\mathcal{F}_t, W) = H(t-)X_2$ and $\mu_0(t) = \log(t+1)$ 55

3.1 Estimation results of $(\beta_0', \alpha_0')'$ for simulated panel count data with non-Poisson observation processes, $h(\mathcal{F}_t, W) = H(t-)X_1$, $\alpha = 0.3$, $\mu_0(t) = \log(t+1)$ and $n = 100$ by chosing different numbers and placements of knots. 89

3.2 Estimation results of $(\beta_0', \alpha_0')'$ for simulated panel count data with Poisson and non-Poisson observation processes, $h(\mathcal{F}_t, W) = H(t-)X_1$ and $\mu_0(t) = \sqrt{t}$ 90

3.3 Estimation results of $(\beta_0', \alpha_0')'$ for simulated panel count data with Poisson and non-Poisson observation processes, $h(\mathcal{F}_t, W) = H(t-)X_1$ and $\mu_0(t) = \log(t+1)$ 91

4.1 Estimation of β with time-independent and time-dependent covariates respectively and $n = 100$ under the homogeneous Poisson observation process 111

4.2 Estimation of β with time-independent and time-dependent covariates respectively and $n = 100$ under the nonhomogeneous Poisson observation process 112

4.3 Estimation of β under the homogeneous and nonhomogeneous Poisson observation process respectively with n=200 and time-independent covariates 113

Chapter 1

Introduction and Literature Review

In this chapter, we focus on introducing the motivations of our research and reviewing the related literature. In our research, we mainly discuss three different semiparametric analysis procedures for longitudinal data with informative observation times.

1.1 Motivation

The bladder cancer study conducted by the Veterans Administration Cooperative Urological Research Group (VACURG), which will be considered in this dissertation, is described in this section to illustrate that what applications have motivated us to do analysis about longitudinal data with informative observation times.

1.1.1 Bladder Cancer Study

A bladder cancer follow-up study conducted by the VACURG of USA extracted from Andrews and Herzberg (1985, pp.253-260) was first studied by Byar (1980). In the study, 116 subjects had superficial bladder tumors when they entered the study and these tumors were removed transurethrally and then patients were randomly allocated to one of the three treatments, placebo (47), thiotepa (38) and pyridoxine (31). Many patients had multiple recurrences of new tumors during the study. The follow-up times vary from 1 week to 64 weeks. In the study, the patients periodically visited the clinical centers and at each visit, the numbers of bladder new tumors since the last visit were recorded and the new tumors were removed transurethrally and then the treatment was continued. For many patients, more than one bladder new tumors were recorded between two visits, however, times for each tumor occurred were not exactly known. Furthermore, different patients had different visiting times since they had not visited the clinical centers at the scheduled times because of some personal reasons, that is, the observation times and the censoring times vary from patient to patient. Thus only panel count data are available. For each patient, the observed information includes clinical visit times, the numbers of recurrent tumors between clinical visits, two baseline covariates that are the number of initial tumors and the size of the largest initial tumor, and also the type of treatment for the patient. The full data set can be located at `http://www.blackwellpublishers.co.uk/rss/`, and the data for the placebo and thiotepa groups can be found in Hu et al. (2003) and Appendix of Sun (2006).

As indicated in Byar (1980) and Andrews and Herzberg (1985), one of the main objectives of the study is to evaluate the effect of the treatment on the rate of tumor recurrence. The data have been analyzed by Sun and Wei (2000); Wellner and Zhang (2000); Zhang (2002); Wellner and Zhang (2007) among others, where the observation times were assumed to be noninformative.

However, the appearance that some patients in the study had significantly more clinical visits than others indicates that the number of clinical visits may contain some information about the tumor occurrence rate. Thus an important question is how to take into account or make use of this information for inference about the tumor recurrence rate. For the analysis of rate of tumor recurrence with informative observation times, Huang et al. (2006), Li et al. (2010), Zhao and Tong (2011), and Deng (2012) among others have developed different analysis procedures.

Furthermore, by comparing the data in the placebo group with that in the thiotepa group, it is noticeable that the subjects in the thiotepa group tended to visit the clinics more often than those in the placebo group, which may be explained by the reason that the patients in the thiotepa needed to visit the clinics more in order to have their thiotepa installed. The different patterns of observation times should be taken into account in the analysis. Hu et al. (2003), Sun et al. (2005), Li et al. (2010),and Zhao and Tong (2011) have gave some analysis about this data based on the nonhomogeneous Poisson process assumption about the observation times. A nonstationary Poisson process with frailty (Sun et al., 2007; Zhao et al., 2012), a conditional intensity model (Liang et al., 2009) and a marginal rate model (Song et al., 2012) for the observation times have also been proposed. One problem behind all these model assumptions about observation times is how to assess the adequacy of these models.

In addition, the natural logarithm of the total number of observed tumors within the last 3 months plus 1 taken as a time-dependent covariate was used to assess the effect of the dependence among tumor recurrence on the tumor recurrence rate by Sun et al. (2011).

Motivated by the bladder cancer data, we develop marginal conditional models for longitudinal data and panel count data with informative observation times, without any restriction on the pattern of the observation times in Chapter 2 and 3, respectively. Also a class of joint frailty models of panel count data with time-dependent covariates and informative information times is proposed in Chapter 4.

The detailed application discussions on the bladder cancer data are presented in the application parts of Chapters 2 - 4.

1.2 Literature Review

1.2.1 Longitudinal Data

In many longitudinal studies, each individual may experience the same event repeatedly at distinct time points during a relatively long follow-up time. These data may occur frequently in a wide variety of settings, including epidemiology, clinical trials, and economic applications and so on. Examples of longitudinal data include the bladder cancer data (Byar, 1980), a cost-accrual process of chronic heart failure patients from the clinical data repository (CDR) at the University of Virginia Health System (Liu et al., 2008), available online at `http://cdr.virginia.edu/cdr`, seizures counts for epileptic patients (Thall and Vail, 1990; Albert, 1991), a chemotherapy cardiotoxic outcome data for the Acute Lymphoblastic Leukemia (ALL) (Lipshultz et al., 1995; Lipsitz et al., 2002), a medical cost data for the childhood Acute Myeloid Leukemia (AML) trial (Rubnitz et al., 2010; Zhu et al., 2011), the air pollution data (Leitenstorfer and Tutz, 2007), whose original database can be found at `http://www.ime.usp.br/~jmsinger/Polatm9497.zip` and so on. The main characteristic of such data is that the observations are independent between different subjects and may be correlated within each subject. However, in Chapters 2 - 4, we just considered "population-averaged" (PA) models (Zeger et al., 1988), which modelled the population-averaged response as a function of covariates without explicitly accounting for within subject heterogeneity, and thus the effects of the covariates have interpretation for the population rather than for any subject.

For the analysis of longitudinal data, parametric regression analysis has been studied by Laird and Ware (1982) and Liang and Zeger (1986) among others. Diggle et al. (1994) provided an excellent review of frequently used methods including both estimating equation and random effect model approaches and Verbeke and Molenberghs (2000) given a comprehensive review of linear mixed model procedures. In order to avoid the possible modeling biases in parametric analysis, various more flexible nonparametric models have been proposed by several authors including Hoover et al. (1998), Wang (1998), Zhang et al. (1998), and Huang et al. (2004) among others. By composing the parametric and nonparametric models, a number of semiparametric models with nice features have been considered for longitudinal data. Zeger and Diggle (1994) proposed a semiparametric mixed model for longitudinal data and suggested a backfitting procedure for inference. Lin and Ying (2001) developed a novel and simple semiparametric and nonparametric method for the regression analysis of irregularly spaced longitudinal data by formulating the observation times within the framework of counting processes. He et al. (2002) considered an extended M-estimators for analyzing longitudinal data with unspecified dependence structure.

In the longitudinal data analysis, there are two important processes – the response process and the observation process to be considered. A basic assumption behind all the methods mentioned above is that observation times are independent of response variable, completely or given covariates. However, this assumption may be violated in many applications, such as the longitudinal data arising from the bladder cancer follow-up study conducted by the VACURG (Byar, 1980) as mentioned in Section 1.1.1. In the Acute Lymphoblastic Leukemia (ALL) data (Lipsitz et al., 2002), a patient with an abnormally low wall-thickness measurement may demand more frequent echocardiograms and visit times. In the AML trial data (Zhu et al., 2011), patients in a severe disease stage visit the hospital more often than those in a mild disease stage. We call these response-dependent visit times as informative observation times. Thus it is very necessary to determine the relationship between the response process and the observation process so as to take into account or make use of this information for inference about the effect of the covariates on the response process.

For the longitudinal data analysis with informative observation times, two methods have been developed. One is the conditional modeling approach proposed by Lin et al. (2004) and Sun et al. (2005). Lin et al. (2004) constructed their conditional model based on the sequential ignorability assumption (Robins and Rotnitzky, 1992) that the decision to visit at time t did not depend on the current response given the past history, and developed a class of inverse intensity-of-visit process-weighted estimators. Sun et al. (2005) generalized the marginal model given by Lin and Ying (2001) to a conditional model, which obviously characterized the dependence of the response process and the observation times and proposed estimating equation approaches. Another one is the frailty-based approach proposed by Sun et al. (2007), Liang et al. (2009), Zhao et al. (2012), Song et al. (2012) among others. For example, Sun et al. (2007) used a shared latent variable or frailty to characterize the correlations between the response process and the observation times with informative censoring times. Liang et al. (2009) modeled the longitudinal data with informative observation times via two different latent variables that satisfied a linear relationship and some external covariates and the distribution assumption for a latent variable is required. Zhao et al. (2012) considered more general joint models using a completely unspecified link function and a latent variable to characterize the correlations between the response process and the observation process, and developed estimating equation approaches.

As discussed in Section 1.1.1, the different patterns of observation times should also be taken into account in the analysis. The vast majority of research mentioned above were based on a common and key assumption that the observation process follows a Poisson or mixed Poisson with the proportional intensity function (Sun et al., 2005, 2007; Liang et al., 2009; Zhao et al., 2012). However, the fit of the Poisson model may be inadequate when the observation process displays under-dispersion or over-dispersion. Based on such consideration, Song et al. (2012) proposed a new more flexible joint modelling approach for the longitudinal data with informative times via two different latent variables, where the response process was assumed to follow a marginal mean model and the observation process follows a marginal rate model which does not rely on the assumption of a nonhomogeneous Poisson process.

In addition, the relation between the observation and response processes may vary with some covariates. For example, in the bladder cancer study, patients who received the thiotepa treatment may have less superficial bladder tumors, and thus may visit the doctor less often than those in the placebo group, which means that the correlation between the observation times and the tumor recurrent process may be different for different treatment groups. Earlier researchers have not considered this situation, however, ignoring this fact may result in magnify biased estimators. Motivated by the discussions mentioned above, we will develop a new class of semiparametric mean models for the correlated response process and the observation process, which allows for the interaction between the observation history and covariates, leaving patterns of the observation process to be arbitrary in Chapter 2 and also discuss the estimation approach for the models.

1.2.2 Panel Count data

In some longitudinal follow-up studies, each subject may be observed at several distinct times and only the numbers of events between two adjacent times are available. It may be impossible to record the exact event times because of too expensive examination cost or too frequent occurrence of the events for their exact times to be recorded and so on. Moreover, the set of observation times may vary from subject to subject. Such data are called panel count data. For this data, important information including the observation times, the counts of recurrent events, the censoring or follow-up times and the covariates related to the study are recorded for each study subject. Clearly, panel count data is a special case of longitudinal data in which the underlying counting process for recurrent events of interest is regarded as the response process and covariate effects on the underlying recurrent process are often the study of interest. The applications of panel count data including the bladder cancer data mentioned in Section 1.1.1 (Byar, 1980), the incidence of nausea of patients with gallstone disease from the National Cooperative Gallstone Study (NCGS) (Thall and Lachin, 1988; Sun and Kalbleisch, 1995), reliability of nuclear plants (Gaver and O'Muircheartaigh, 1987; Sun and Kalbleisch, 1995) and so on. In panel count data, a special case exists when only one observation is taken for every subject and the survival time of interest is known only to be either less or greater than the observation time. Such data is called current status data (Case 1 interval-censored data). A typical example of current status data can be found in the tumorigenicity experiments (Dewanji and Kalbfleisch, 1986), in which only the death time of animals at study and the status of tumor onset at the death time are observed. Multivariate panel count data arise if more then one kind of recurrent events are to be considered and individuals are only observed repeatedly at intermittent times. In tumorigenicity experiments, this data are commonly exist when several types of tumors occur together. Chen et al. (2005) gave an example of an advanced breast cancer study, in which three types of metastatic bone lesions and related covariates are recorded at distinct examination times for each patient. He et al. (2008) analyzed a cohort study of psoriatic arthritis patients at the University of Toronto Psoriatic Arthritis Clinic, where two types of joint damages (radiologically and functionally joint damage) were considered. Another form related with panel data is the multistate panel data, in which the observations consist of a finite number of states occupied by the individuals under study at a sequence of discrete, irregularly spaced time points, with no information about the exact transition times. These data commonly exist in applications, such as, a study of the smoking habits of school children with three possible "smoking status" for each child (Kalbfleisch and Lawless, 1985), a hepatocellular carcinoma study (Kay, 1986) with three states of the serum alphafetoprotein (AFP) level, a cytomegalovirus (CMV) retinitis clinical trial with five stages of toxicity of the treatments for the Acquired Immune Deficiency Syndrome (AIDS) patients (Lee and Kim, 1998).

A majority of researchers have investigated the analysis of panel count data under the assumption that the observation process is independent of the underlying recurrent event process completely or conditional on covariates. For estimation of the mean function of the underlying recurrent process, many nonparametric methods have been developed. Sun and Kalbleisch (1995) presented a consistent estimator of the mean function based on isotonic regression (Barlow et al., 1972; Robertson et al., 1988), while Wellner and Zhang (2000) derived a nonparametric maximum pseudo-likelihood estimator (NPMPLE) and the nonparametric maximum likelihood estimator (NPMLE) under a nonhomogeneous Poisson process assumption for the underlying recurrent process. Zhang and Jamshidian (2003) introduced the gamma frailty model for the intracorrelated panel counts and constructed an NPMPLE proposed in Wellner and Zhang (2000) with the frailty. Lu et al. (2007) proposed a monotone B-splines-based nonparametric likelihood estimator for the mean function. Hu et al. (2009a) discussed a nonparametric generalized weighted least squares estimator the Sun-Kalbfleisch's estimator and Wellner-Zhang's NPMLE as special cases. Hu et al. (2009b) developed two types of self-consistent estimating equation procedures for the mean function of the underly recurrent process with a Poisson assumption.

In consideration of the covariates effect on the underlying recurrent process, semiparametric analysis of panel count data have drawn considerable attention in survival literatures. Sun and Wei (2000) constructed a proportional means model proposed in Lin et al. (2000) for the underlying recurrent event process with the observation times and follow-up time independent or dependent of the covariates. Under the proportional means model assumption for the underlying recurrent event process, Hu et al. (2003) proposed estimation equation approaches for a general observation process without model restriction and a proportion rate model for observation process, respectively. Zhang (2002) proposed a semiparametric pseudolikelihood estimation method based on a nonhomogeneous Poisson process assumption for the proportional means model. Furthermore, Wellner and Zhang (2007) studied both the semiparametric maximum pseudo-likelihood and maximum likelihood estimators for the proportion means model. Iterative algorithm proposed via profile likelihood approach was used in Zhang (2002) and Wellner and Zhang (2007) to obtain their estimators, however, this algorithm is not efficient, especially for the maximum likelihood estimation method. Thus, an easy implemented generalized Rosen algorithm proposed by Zhang and Jamshidian (2004) was used by Lu et al. (2009) to compute their estimators. Tong et al. (2009) was the first one to consider the variable selection problem in the panel count data, and they developed a non-concave penalized estimating function approach that could select variables and estimate the regression coefficients for the proportional mean model simultaneously. Then Wu and He (2012) explored a fast coordinate ascent algorithm to select relevant predictors for the underlying recurrent event process under a proportional mean model, when the number of predictors far exceeds the number of subjects. Bayesian analysis for panel count data with dependent termination time was proposed by Sinha and Maiti (2004), where they constructed semiparamatric joint models for the underlying recurrent events and the termination time via a frailty and used Markov chain Monte Carlo algorithm to estimate the regression parameters and the unknown function.

When the panel count data consist of independent samples randomly drawn from $k (k \geq 2)$ populations or groups, one important thing is to handle the treatment comparison. Thall and Lachin (1988) suggested to transform the problem to a multivariate comparison problem and then apply a multivariate Wilcoxon-type rank test, while Sun and Fang (2003) proposed a nonparametric approach under the assumption that treatment indicators can be regarded as independent and identically distributed random variables. Also Park et al. (2007) gave a class of nonparametric two-sample tests based on the isotonic regression estimator of the mean function of the underlying recurrent counting process, while Zhang (2006) and Balakrishnan and Zhao (2011) developed some multi-sample nonparametric procedures by using nonparametric maximum pseudo-likelihood (isotonic regression) approach. Furthermore, Balakrishnan and Zhao (2009, 2010) proposed new class of test statistics by using the nonparametric maximum likelihood estimator. Recently, Zhao and Sun (2011) presented nonparametric tests for the comparison of several treatment groups with different observation schemes. In addition, Zhao et al. (2011) provide a relatively complete discussion for the analysis of panel count data wherein more references can be found.

All the references mentioned above are for the univariate panel count data. Other different forms of the panel count data have also attracted a lot of researchers to study them. For the multivariate panel count data analysis, He et al. (2008) presented a class of marginal mean models, leaving the dependence structures of related types of recurrent events completely unspecified. For the current status data analysis, Diamond et al. (1986) extended the proportion hazards model (Cox, 1972) to current status data. Sun and Kalbfleisch (1993) proposed a point process technique to test the equality of mean functions of point processes. Xue et al. (2004) developed a partial linear model for the current status data and proposed a sieve maximum likelihood estimation method. For multistate panel data, Kalbfleisch and Lawless (1985) discussed the fitting of Markov model with homogeneous transition intensities to multistate panel data. Kay (1986) developed a Markov model to assess the dependence of risk of death on disease states. Lee and Kim (1998) proposed a procedure assuming each multistate process marginally to follow a time-homogeneous Markov Model allowing for covariates. These inferences are all based on Markov models, however, in many applications, the Markov assumption is not suitable when the transition intensities depend on the elapsed time in the current state. For this, Kang and Lagakos (2007) developed likelihood-based procedures for multistate panel data from a semi-Markov process, where transition intensities depend on the duration of time in the current state.

In many situations, the underlying recurrent process and the observation process are still related even given covariates, such as an example given by a set of panel count data arising from the bladder cancer follow-up study mentioned in Section 1.1.1. As stated in Section 1.2.1, the number of clinical visits may contain some information about the tumor occurrence rate (Sun and Wei, 2000; Hu et al., 2003; Li et al., 2010; Zhao and Tong, 2011). Another example can be seen in a special case of panel count – current status data in tumorigenicity experiments, where tumor onset time and the death time are usually of interest. If the tumors are lethal, meaning that the tumor onset kills animals instantly, thus the death time may depend on the tumor onset time (Zhang et al., 2005). They developed some statistical analysis of current status data with informative observation times by a random effect to determine the correlation. For the analysis of panel count data with informative observation times, limited research exists. A class of semiparametric transformation models for the recurrent event process was constructed by Li et al. (2010), by incorporating the observation history to the mean model of the recurrent event process to reflect the correlation between these two processes, with a nonhomogeneous Poisson process assumption for the observation times. Zhao and Tong (2011) proposed a joint modeling approach that used an unobserved frailty variable and a completely unspecified link function to characterize the correlation between the recurrent event process and the observation times assuming the observation process to be a nonhomogeneous Poisson process with frailty.

However, just as we have stated for longitudinal data, the inadequacy of the fitting of the Poisson model for the observation process may be yet existed in panel count data when the observation process displays under-dispersion or over-dispersion. Neglecting this under-dispersion or over-dispersion may result in biased estimates and loss of estimation efficiency. Hu et al. (2009a) verified by simulation that Weller-Zhang's MLE is no longer efficient if the over-dispersion exist in the panel count data. Few researchers have considered this under-dispersion or over-dispersion problem in the panel count data. Huang et al. (2006) studied nonparametric and semiparametric models that allow the observation times to be correlated with the event process, where the correlation is induced by a frailty variable and the distributions of the observation times and the frailty were considered as nuisance parameters. Hua and Zhang (2011) established a proportional mean model without any stochastic assumption for the underlying recurrent event process, and developed a spline-based semiparametric projected generalized estimating equation (GEE) method through incorporating a working covariance matrix which accounts for over-dispersion into the GEE so as to improve the estimation efficiency and the variance estimation accuracy.

In the analysis of panel count data with informative observation times, the same situation that the relation between the observation and the recurrent event processes may vary with some covariates may exist as in the analysis of longitudinal data discussed in Section 1.2.1.

In view of these three problems discussed in the previous three paragraphs, in Chapter 3, we will develop a new class of more flexible semiparametric regression models by incorporating the interaction between the observation history and some covariates to the mean model of the recurrent event process, while leaving the patterns of the observation times to be arbitrary. For inference, a B-spline based least square estimation procedure is proposed there.

In addition, in some applications, it would be desirable to develop estimation procedures for panel count data with informative observation times, and also with time-dependent covariates and informative censoring times. For example, in order to assess the effect of the dependence among tumor recurrence on the tumor recurrence rate, Sun et al. (2011) took the natural logarithm of the total number of observed tumors within the last 3 months plus 1 as a time-dependent covariate in their analysis. The underlying recurrent event processes with informative censoring time exist especially in situations where a correlated failure event could potentially terminate the further observation of the recurrent events. Hence, in Chapter 4, we will consider the same models for the underlying recurrent events and the observation times as given in Zhao and Tong (2011) except replacing the time-independent covariates with the time-dependent covariates and removing the assumption of noninformative censoring, and present an estimating equation procedure there.

1.2.3 B-Splines in Survival Analysis

B-splines is a very popular type of polynominal splines in statistical applications, mainly because of their flexibility and numerical properties (de Boor, 1978;

Schumaker, 1981). The definition and fundamental properties of B-splines are presented in Appendix A. Here we just review some references about the applications of B-splines to a variety of aspects in survival analysis.

First of all, B-splines approximation can be used to estimate different nonparametric smooth functions in a variety of survival models. We will summariz some of them here. Whittemore and Keller (1986); Etezadi-Amoli and Ciampi (1987); Rosenberg (1995); Kooperberg et al. (1995); Cai and Betensky (2003) have investigated use of the fixed knots and quadratic or cubic linear splines or B-splines to estimate the hazard function or baseline hazard function of their respective censored survival models. Also, when linear-effect on the log-hazard of the proportional hazards model (Cox, 1972) was not hold, flexible relative risk form (Sleeper and Harrington, 1990; Huang and Liu, 2006) and partially linear single-index (Gray, 1992; Sun et al., 2008) were suggested, where the nonparametric functions involved in these models were estimated by B-splines. Giorgi et al. (2003) proposed to use quadratic B-splines with fixed number of knots to model the hazard ratio for their relative survival regression model proposed by Esteve et al. (1990). More recently, Amorim et al. (2008) used cubic B-splines with fixed number of knots to estimate the time-varying coefficients in the rates model for recurrent event data. A Bayesian estimation procedure that used B-splines for a proportional hazards frailty models was presented in Sharef et al. (2010). Zhang et al. (2010) developed a spline-based semiparametric maximum likelihood method to study the Cox model with interval-censored data. Most of these papers delivered above showed that in moderate and heavily censored samples, the spline-based approaches not only have advantages in rate of computing, but also in accuracy by some simulations or real data analysis.

Furthermore, when the function to be estimated is monotone, isotonic regression estimator (Barlow et al., 1972; Robertson et al., 1988; Wellner and Zhang, 2000), which can be viewed as a special case of monotone I-splines with order one and knots positioned at the distinct data points, was proposed. However, as described in Wellner and Zhang (2000), the computation of the estimator involved the iterative convex minorant algorithm proposed by Jongbloed (1998), which could be computationally demanding when the sample size is large. Therefore, many researchers have developed the monotone B-splines estimation procedure because of the following two reasons. Firstly, it is convenient to impose the monotone constraints on the coefficients of the B-splines bases as B-splines possess the same monotonicity as the coefficients because of the variation-diminishing properties (Schumaker, 1981). Secondly, the splines estimators are less computationally demanding since the number of the B-splines basis functions is often chosen much smaller than the sample size. For example, Ramsay (1988) defined monotone I-splines, and the merits of these monotone splines were showed through a number of statistical applications, including response variable transformation in nonlinear regression and modelling a dose-response function by monotone splines. Kelly and Rice (1990) proposed to use nonparametric smoothing instead of non-adequate parametric modeling procedure to study the dose-response curves under monotonicity constraints. Shen (1998) introduced a spline-based sieve maximum likelihood estimation method to estimate the nondecreasing baseline function and regression parameter in proportional odds model with right-censored and Case 2 interval-censored data. Leitenstorfer and Tutz (2007) developed a fitting procedure based on the monotone B-splines for generalized additive models to investigate the effect of the air pollutant on respiratory mortality. Lu (2010) proposed a monotone B-splines-based sieve maximum likelihood estimator which can be computed by the generalized Rosen algorithm in Jamshidian (2004) for a partly linear model.

As for the nonparametric function estimation in the semiparametric analysis of longitudinal data and panel count data discussed here, monotone B-splines are also widely used. For example, for panel count data, Lu et al. (2007) obtained a monotone I-spline likelihood-based estimator for the mean function of the recurrent event process with panel count data by a generalized Rosen algorithm (Jamshidian, 2004) Then Lu et al. (2009) studied semiparametric likelihood-based method for panel count data by using generalized Rosen algorithm to compute the regression parameters and the underlying mean function approximated by monotone B-splines simultaneously. Hua and Zhang (2011) proposed a proportional mean model without any assumptions for the underlying recurrent counting process and the natural logarithm of the baseline mean function was approximated by a monotone cubic B-spline function, whose coefficients along with regression parameters were obtained by a projected generalized estimating equation method with the working covariance matrix that accounts for overdispersion incorporated. When considering the possible within-cluster heterogeneity existence in panel count data, Nielsen and Dean (2008) assumed that the counts for each individual were generated by mixtures of nonhomogeneous Poisson processes with intensity functions approximated by cubic B-splines. For longitudinal data, Lin and Zhang (1999) developed smoothing spline-based statistical inference in a class of generalized additive mixed models. Huang et al. (2004) proposed to approximate each coefficient function by a polynomial spline and employed a least square method for estimation.

Three components involved in B-splines approximation are the degree of splines, the number of knots and the location of knots. In general, cubic or quadratic splines are sufficient to fit the unknown smooth function well. The number of knots determines the flexibility of the fitted splines since reducing or increasing the number of knots which means reducing or increasing the density of knots in different regions of the observation times will result in reducing or increasing the flexibility within those regions. One way for choosing the number of knots is to let it vary in a relatively large range and define the final number to be the one that maximized the Akaike information criterion (AIC) (Akaike, 1973) or Bayesian information criterion (BIC) as given in Rosenberg (1995) and Huang and Liu (2006). Another way for choosing the number of knots is to set the number of interior knots to be $m_n = O(n^\nu)$ with $0 < \nu < 1/2$ and n being the sample size or the number of the distinct observation times as in Lu et al. (2007, 2009) and Hua and Zhang (2011). Given the number of knots, the location of the knots which determines the shapes of the basis splines thus in turn the shape of the fitted splines is obviously an important problem to be considered. There are mainly two data-driven methods for determining locations of knots – uniform partitions (Lu et al., 2007; Lu, 2010) and partitions according to quantiles of the data (Rosenberg, 1995; Lu et al., 2009; Hua and Zhang, 2011). Given number of the interior knots m_n, for the uniform partitions, the equally spaced knots are given by $t_{\min} + k(t_{\max} - t_{\min})/(m_n + 1), k = 0, 1, \cdots, m_n + 1$, with $t_{\min}$ and $t_{\max}$ being the respective minimum and maximum values of distinct observation times. For the partitions according to quantiles of the data, the $k/(m_n + 1)$ quantiles $(k = 0, 1, \cdots, m_n + 1)$ of the distinct observation times are chosen to be the knots. However, it is showed that the estimation results are rather robust with respect to the number of knots and the location of the knots (Ramsay, 1988; Cai and Betensky, 2003; Lu et al., 2009). The last point which is worth pointing out is that splines composed of linear combinations of exponential, trigonometric, Dirac delta function and some other form of functions are also possible (de Boor, 1978; Schumaker, 1981; Whittemore and Keller, 1986).

1.3 Outline of Thesis

The reminder of this thesis is organized as follows. For correlated response process and observation process in longitudinal data, a new class of semiparametric mean models which allows for the interaction between the observation history and covariates, leaving patterns of the observation process to be arbitrary are developed in Chapter 2. For inference on the regression parameters and the baseline mean function, a spline-based least square estimation approach is proposed, and the consistency, rate of convergence and asymptotic normality of the proposed estimators are established. Simulation studies demonstrate that the proposed inference procedure performs well. Some graphical and numerical techniques are presented to check the adequacy of the fitted model. The analysis of the bladder tumor data is presented to illustrate the proposed method.

Similar to Chapter 2, Chapter 3 presents a new class of semiparametric regression models by incorporating the observation history to the mean model of the recurrent event process, while leaving the patterns of the observation times to be arbitrary. A monotone B-spline-based least-square estimation approach is also proposed to make inference about the regression parameters and the baseline mean function, and asymptotic properties including consistency, rate of convergence and asymptotic normality of the proposed estimators are established. Numerical results including simulation studies and the analysis of the bladder tumor data are also provided.

Chapter 4 extends the joint frailty models proposed by Zhao and Tong (2011) to panel count data with time-dependent covariates and informative observation and censoring times. A novel estimating equation approach that does not depend on distributions of frailty variables and the link function is proposed for estimation of parameters, and the asymptotic properties of the proposed estimators are established. The performance of the proposed inference procedure is demonstrated by some simulation studies and illustrated by the analysis of the bladder tumor data.

In Chapter 5, some conclusions and related future research are presented.

Chapter 2

Longitudinal Data Analysis Using B-Splines

Semiparametric regression analysis of longitudinal data with informative observation times using B-splines is developed in this chapter.

2.1 Introduction

As mentioned in Section 1.2.1, it is desireble to analyze longitudinal data when observation times contain information on the reponse process even given covariates, and overcome the inadequacy of the fitting of the Poisson model assumption for the observation process. In this chapter, motivated by the conditional mean model of the response process given in Sun et al. (2005), we propose a new class of semiparametric regression models which allows for the interaction between the observation history and some covariates, while leaving the patterns of the observation times to be arbitrary. For the nonparametric estimation of the baseline mean function, a B-spline approximation will be used following Lu et al. (2007, 2009).

The remainder of this chapter is organized as follows. We begin in Section 2.2 by introducing some notation and describing our models for longitudinal data. In Section 2.3, a spline-based least square method is proposed for estimation of regression parameters and the baseline unknown mean function invovled in our models. Section 2.4 presents the asymptotic properties of the proposed estimators, including consistency, rate of convergence and asymptotic normality. In order to assess the finite-sample performance of the proposed inference procedure, we present some results obtained from simulation studies in Section 2.5. In Section 2.6, the proposed approaches are illustrated through the analysis of a data set from a bladder tumor study.

2.2 Statistical model

Consider a longitudinal study that consists of a random sample of n subjects. For subject i, let $Y_i(t)$ denote the response variable and $\mathbf{X}_i$ denote a p-dimensional vector of covariates, $i = 1, \ldots, n$. Suppose that $Y_i(t)$ is observed at distinct time points $T_{K_i,1} < T_{K_i,2} < \ldots < T_{K_i,K_i}$, where K_i is the total number of observations on subject i. In the following, we regard these observation times arising from an underlying counting process $H(t)$ characterized by $H_i(t) = \sum_{j=1}^{K_i} I\left(T_{K_i,j} \le t\right)$, where $I(\cdot)$ is the indicator function, and define $\tilde{H}_i(t) = H_i(\min(t, C_i))$, where C_i is the follow-up or censoring time for subject i, $i = 1, \ldots, n$. Then, the process $Y_i(t)$ is observed only at the time points where $\tilde{H}_i(t)$ jumps.

Define $\mathcal{F}_{it} = \{H_i(s), 0 \le s < t\}$ as the observation history just before t. For semiparametric analysis of longitudinal data with informative observation times, Sun et al. (2005) was the first one to propose a conditional modelling approach, and our works in Chapter 2 and Chapter 3 are motivated by their models. Thus, let's first introduce their models.

They assumed that $Y_i(t)$ follows the marginal model

$$E\{Y_i(t)|\mathbf{X}_i, \mathcal{F}_{it}\} = \mu_0(t) + \beta_0'\mathbf{X}_i + \alpha_0'h(\mathcal{F}_{it}), \tag{2.1}$$

where $\mu_0(t)$ is an unspecified smooth function of t, β_0 is a p-dimension covariate effect on the response process, α_0 is a q-dimension regression coefficients, which determines the correlation between the response process and the observation process, and $h(\cdot)$ is a q-dimensional vector of known function. Also $H(t)$ is assumed to follow a nonhomogeneous Poisson process with

$$E\{dH(t)|\mathbf{X}_i\} = e^{\gamma'\mathbf{X}_i} d\Lambda_0(t). \tag{2.2}$$

where γ_0 is a p-dimensional vector of regression parameters, and $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$ is the mean cumulative number of observations by time t.

However, as discussed in Section 1.2.1, the fit of the Poisson model for observation process may not be inadequate when the observation process displays under-dispersion or over-dispersion. In addition, some covariates may influence the relation between the observation and response processes. Thus, by leaving the patterns of the observation times to be arbitrary, we assume that $Y_i(t)$ follows the marginal model

$$E\{Y_i(t)|\mathbf{X}_i, W_i, \mathcal{F}_{it}\} = \mu_0(t) + \beta'\mathbf{X}_i + \alpha' h(\mathcal{F}_{it}, W_i), \tag{2.3}$$

given $\mathbf{X}_i$, $\mathcal{F}_{it}$ and the covariate W_i, which may be a component of the vector $\mathbf{X}_i$ or may be other variables different from $\mathbf{X}_i$, where $\mu_0(t)$ is an unspecified smooth function of t, β is a p-dimensional vector of unknown regression parameters, α is a q-dimensional vector of regression coefficients, and $h(\cdot)$ is a vector of known functions of the counting process $H_i(\cdot)$ up to $t-$ and the covariate W_i, representing the interaction between the observation history and some covariates. Especially, when in some clinical studies with many different treatments, W_i are defined as the group indicators, then $h(\cdot)$ represents the different group effects on the observation times. The main purpose here is to estimate the regression coefficients α, β and the smooth baseline mean function $\mu_0(t)$.

Model (2.3) specifies that the process $Y_i(t)$ depends on the observation process $H_i(t)$ through function h, which can be chosen according to situations. Following the discussion in Sun et al. (2005), a natural and simple choice for h may be $h(\mathcal{F}_{it}, W_i) = H_i(t-)W_i$, which means that $Y_i(t)$ and $\mathcal{F}_{it}$ are related through or all information about $Y_i(t)$ in $\mathcal{F}_{it}$ is given by the total number of observations. An alternative is that $Y_i(t)$ depends on $\mathcal{F}_{it}$ only through a recent number of observations, say, in u time units, and this corresponds to $h(\mathcal{F}_{it}, W_i) = (H_i(t-) - H_i(t-u))W_i$. One could define h as a vector given by the forgoing two choices if both the total and recent numbers of observations may contain information about $H_i(t)$.

In addition, we assume that

$$E\{Y_i(t)|\mathbf{X}_i, H_i(s), 0 \leq s \leq t, C_i\} = E\{Y_i(t)|\mathbf{X}_i, \mathcal{F}_{it}, C_i\}, \tag{2.4}$$

which means that conditional on the covariates $\mathbf{X}_i'$s and C_i's, the mean of response variable at time point t is only related to the observation history before t. The observation for each individual consists of $\mathbf{O} = (K, \bar{T}_K, \bar{Y}_K, \bar{H}_K, \ \mathbf{X}, C)$, with $\bar{T}_K = (T_{K,1}, \cdots, T_{K,K})$, $\bar{Y}_K = (Y(T_{K,1}), \cdots, Y(T_{K,K}))$, $\bar{H}_K = (H(T_{K,1}), \cdots, H(T_{K,K}))$. Throughout this chapter, we will assume that we observe n i.i.d. copies, $\mathbf{O}_1, \cdots, \mathbf{O}_n$ of $\mathbf{O}$.

2.3 Inference procedure

For inference about model (2.3), a B-splines based least sqaure estimation procedure is developed here.

Firstly, define

$$
\begin{aligned}
L_n(\beta, \alpha, \mu) &= \sum_{i=1}^{n} \sum_{j=1}^{K_i} \left[Y_i(T_{K_i,j}) - \mu(T_{K_i,j}) - \beta' \mathbf{X}_i - \alpha' h(\mathcal{F}_{iT_{K_i,j}}, W_i) \right]^2 \xi_i(T_{K_i,j}) \\
&= \sum_{i=1}^{n} \int_0^{\tau} \left\{ Y_i(t) - \mu(t) - \beta' \mathbf{X}_i - \alpha' h(\mathcal{F}_{it}, W_i) \right\}^2 d\tilde{H}_i(t),
\end{aligned}
\tag{2.5}
$$

where $\xi_i(t) = I(C_i \geq t)$.

We propose to use B-splines to approximate $\mu_0(t)$. For a finite closed interval $[0, \tau]$, let $\mathcal{I} = \{t_i\}_1^{m_n+2l}$, with

$$
0 = t_1 = \cdots = t_l < t_{l+1} < \cdots < t_{m_n+l} < t_{m_n+l+1} = \cdots = t_{m_n+2l} = \tau
$$

be a sequence of knots that partition $[0, \tau]$ into $m_n + 1$ subintervals and $m_n = O(n^{\nu})$, for $0 < \nu < 1/2$. Let $\{B_{il}, 1 \leq i \leq q_n\}$ denote the B-spline basis functions with $q_n = m_n + l$. Let $\Psi_{l,\mathcal{I}}$ (with order l and knots $\mathcal{I}$) be the class linearly spanned by the B-spline functions, that is,

$$
\Psi_{l,\mathcal{I}} = \left\{ \sum_{i=1}^{q_n} \theta_i B_{il} : \theta_i \in \mathbb{R}, i = 1, \cdots, q_n \right\}.
$$

Assume that $\mu_0(t) \in \mathcal{F}_r \equiv \{\mu : [0, \infty) \longrightarrow \mathbb{R} \Big| |\mu^{(k)}(s) - \mu^{(k)}(t)| \leq M|s-t|^{\varsigma}\}$, where k is a nonnegative integer, $\varsigma \in (0, 1]$ such that $r = k + \varsigma > 0.5$, M is a positive constant and $f^{(k)}$ is the kth derivative of function f. According to Lemma 5 in

Stone (1985) sketched in Appendix A, there exists a smooth spline $\mu_n(t) \in \Psi_{l,\mathcal{I}}$, with order $l \geq k+1$ such that $\| \mu_n - \mu_0 \|_\infty = \sup_{u\in[0,\tau]} |\mu_n(u) - \mu_0(u)| = O(n^{-\nu r})$. Denote $\mu_n(t) = \theta' B_l(t)$, where $\theta = (\theta_1, \ldots, \theta_{q_n})'$ and $B_l(t) = (B_{1l}(t), \ldots, B_{q_n l}(t))'$. $L_n(\beta, \alpha, \mu)$ in (2.5) is approximate to

$$L_n(\beta,\alpha,\theta) = \sum_{i=1}^{n} \int_0^\tau \{Y_i(t) - \theta' B_l(t) - \beta' \mathbf{X}_i - \alpha' h(\mathcal{F}_{it}, W_i)\}^2 \, d\tilde{H}_i(t).$$

The resulting estimating function for β, α and θ has the form

$$U(\beta,\alpha,\theta) = \sum_{i=1}^{n} \int_0^\tau \begin{pmatrix} \mathbf{X}_i \\ h(\mathcal{F}_{it}, W_i) \\ B_l(t) \end{pmatrix} \times \{Y_i(t) - \theta' B_l(t) - \beta' \mathbf{X}_i - \alpha' h(\mathcal{F}_{it}, W_i)\} \, d\tilde{H}_i(t),$$

The solution to $U(\beta, \alpha, \theta) = 0$ has a closed form

$$\begin{pmatrix} \hat{\beta}_n \\ \hat{\alpha}_n \\ \hat{\theta}_n \end{pmatrix} = \left[\sum_{i=1}^{n} \int_0^\tau \begin{pmatrix} \mathbf{X}_i \\ h(\mathcal{F}_{it}, W_i) \\ B_l(t) \end{pmatrix}^{\otimes 2} d\tilde{H}_i(t) \right]^{-1} \left[\sum_{i=1}^{n} \int_0^\tau \begin{pmatrix} \mathbf{X}_i \\ h(\mathcal{F}_{it}, W_i) \\ B_l(t) \end{pmatrix} Y_i(t) \, d\tilde{H}_i(t) \right].$$

Then the resulting estimator for $\mu_0(t)$ is $\hat{\mu}_n(t) \equiv \sum_{i=1}^{q_n} \hat{\theta}_{ni} B_{il}(t)$.

2.4 Asymptotic theory

To establish the asymptotic properties of the estimators, we need the following regularity conditions.

C1 The maximum spacing of the knots satisfies $\triangle = \max_{l+1<i<m_n+l+1} | t_i - t_{i-1} | = O(n^{-v})$.

C2 The parameter spaces of $(\beta', \alpha')'$, $\mathcal{R}$ is bounded and convex on $\mathbb{R}^{p+q}$, and the true parameter $(\beta_0, \alpha_0, \mu_0) \in \mathcal{R}^\circ \times \mathcal{F}_r$, where $\mathcal{R}^\circ$ is the interior of $\mathcal{R}$.

C3 $P(\|\mathbf{X}\| \leq M_1) = 1$ for a positive constant M_1, that is, the covariate vector is uniformly bounded.

C4 There exists a positive integer M_2 such that $P(K \leq M_2) = 1$, that is, the number of the observation is finite.

C5 If with probability 1, $\mathbf{h}_1'\mathbf{X} + \mathbf{h}_2' h(\mathcal{F}_t, W) + h_3(t) = 0$ for some deterministic function h_3, and $\mathbf{h}_1 \in \mathbb{R}^p$ and $\mathbf{h}_2 \in \mathbb{R}^q$, then $\mathbf{h}_1 = 0$, $\mathbf{h}_2 = 0, h_3(t) = 0$.

Next, we introduce more notation. Let $\mathcal{B}_p$ and $\mathcal{B}$ denote the collection of Borel sets in $\mathbb{R}^p$ and $\mathbb{R}$, respectively, and let $\mathcal{B}_{[0,\tau]} = \{B \cap [0,\tau] : B \in \mathcal{B}\}$. We define measures ν on $(\mathbb{R}^p \times [0,\tau], \mathcal{B}_p \times \mathcal{B}_{[0,\tau]})$ and ν_1 on $([0,\tau], \mathcal{B}_{[0,\tau]})$, as follows: for $B \in \mathcal{B}_{[0,\tau]}$, and $A \in \mathcal{B}_p$,

$$\begin{aligned}
\nu(A \times B) = & \int_{A\times[0,\tau]} \sum_{k=1}^{\infty} P(K = k | \mathbf{X} = \mathbf{x}, C = c)) \\
& \times \sum_{j=1}^{k} P(T_{k,j} \in B \cap [0,c] \big| K = k, \mathbf{X} = \mathbf{x}, C = c) dF(\mathbf{x}, c) \\
= & \int_{A\times[0,\tau]} E\left\{ \sum_{j=1}^{K} I_{B\cap[0,c]}(T_{K,j}) | \mathbf{X} = \mathbf{x}, C = c \right\} dF(\mathbf{x}, c),
\end{aligned}$$

and $\nu_1(B) = \nu(\mathbb{R}^p \times B)$, where F is the joint distribution function of $\mathbf{X}$ and C. Then ν_1 and ν are finite measures under condition C4. Let

$$L_2(\nu_1) = \left\{ f : [0,\infty) \longrightarrow \mathbb{R} \Big| \; ||f||_{L_2(\nu_1)} \equiv \left[\int |f(t)|^2 d\nu_1(t) \right]^{1/2} < \infty \right\}.$$

Clearly,

$$\|f\|_{L_2(\nu_1)} = \left[E\left\{\sum_{j=1}^{K}|f(T_{K,j})|^2\xi(T_{K,j})\right\}\right]^{1/2} = \left[E\left\{\int_0^\tau |f(t)|^2\xi(t)dH(t)\right\}\right]^{1/2}.$$

Let $Z = \{Z(t,W) \equiv h(\mathcal{F}_t, W), 0 \le t \le \tau\}$ represent a q-dimensional bounded random process indexed by t. Here, without loss of generality, we assume that W is one-dimensional. Define $\mathcal{G} \equiv \{z(t,w) : [0,\tau] \times [-M_1, M_1] \longrightarrow \mathcal{M}\}$, where $\mathcal{M}$ is a bounded set on $\mathbb{R}^q$, and for function $f(\mathbf{x}, z, t) : [-M_1, M_1]^p \times \mathcal{G} \times [0,\tau] \longrightarrow \mathbb{R}$, define

$$\|f\|_2 \equiv \left[E\left\{\sum_{j=1}^{K}|f(\mathbf{X}, Z(T_{K,j}, W), T_{K,j})|^2\xi(T_{K,j})\right\}\right]^{1/2}.$$

Set $M_n(g) = n^{-1}L_n(\beta,\alpha,\mu) = \mathbb{P}_n m_g(\mathbf{O})$, where $g(\mathbf{x}, z, t) = \beta'\mathbf{x} + \alpha' z(t,w) + \mu(t)$, $m_g(\mathbf{O}) = \sum_{j=1}^{K}[Y(T_{K,j}) - g(\mathbf{X}, Z(T_{K,j}, W), T_{K,j})]^2\xi(T_{K,j})$, and $M(g) = Pm_g(\mathbf{O})$, where Pf and $\mathbb{P}_n f$ represent $\int f dP$ and $n^{-1}\sum_{i=1}^{n} f(\mathbf{O}_i)$, respectively.

Since $L_2(\nu_1)$ is a Hilbert space, and $\mathcal{F}_r \subset L_2(\nu_1)$, by the Hilbert Projection Theorem (Stakgold, 1998, p. 288), for $x_j \in L_2(\nu_1)$, there is a unique $a_j^* \in \mathcal{F}_r$, s.t. $(x_j - a_j^*) \perp \mathcal{F}_r$, for $j = 1, \cdots, p$. Let $z_l(t,w)$ be the lth component of $h(\mathcal{F}_t, w), l = 1, \cdots, q$. Then for $z_l(t,w) \in L_2(\nu_1)$, there is a unique $b_l^*(t) \in \mathcal{F}_r$, s.t. $(z_l - b_l^*) \perp \mathcal{F}_r$, for $l = 1, \cdots, q$. Let $\mathbf{a}^* = (a_1^*, \cdots, a_p^*)'$ and $\mathbf{b}^* = (b_1^*, \cdots, b_q^*)'$. Furthermore, we need the following condition.

C6 $E\left[\int_0^\tau \begin{pmatrix} \mathbf{X} - \mathbf{a}^* \\ h(\mathcal{F}_t, W) - \mathbf{b}^*(t) \end{pmatrix}^{\otimes 2} d\tilde{H}(t)\right]$ is nonsingular.

In practice, C1 is similar to those required by Stone (1986) and Zhou et al. (1998). C2 is a common assumption in the nonparametric smoothing estimation problem. C3 and C4 are mild conditions. C5 is needed to establish the identifiability of the model. C6 is a technical condition. The asymptotic properties including consistency, rate of convergence and asymptotic normality of the estimators are summarized as follows.

Theorem 2.1 (**Consisitency**). *Under conditions C1 - C4 and C6 ,* $\|\hat{\beta}_n - \beta_0\| \to 0$, $\|\hat{\alpha}_n - \alpha_0\| \to 0$, $\|\hat{\mu}_n - \mu_0\|_{L_2(\nu_1)} \to 0$, *almost surely.*

Proof of Theorem 2.1.

Let $\mu_n(t)$ be the B-spline function approximation of $\mu_0(t)$ with $||\mu_n - \mu_0||_\infty = O(n^{-vr})$, $g_n(\mathbf{x}, z, t) = \beta_0'\mathbf{x} + \alpha_0' z(t, w) + \mu_n(t)$, $\hat{g}_n(\mathbf{x}, z, t) = \hat{\beta}_n'\mathbf{x} + \hat{\alpha}_n' z(t, w) + \hat{\mu}_n(t)$, and $g_0(\mathbf{x}, z, t) = \beta_0'\mathbf{x} + \alpha_0' z(t, w) + \mu_0(t)$. Without loss of generality, we assume that $\mu_n > \mu_0$. Thus $g_n > g_0$, and $\|g_n - g_0\|_\infty = O(n^{-vr})$. Choose a $\phi_n \in \Psi_{l,\mathcal{I}}$ and b_1 and b_2, such that $h_n \equiv {b_1}'\mathbf{x} + {b_2}'z + \phi_n$, and $\|h_n\|_2^2 = O(n^{-vr} + n^{-\frac{1-v}{2}})$. Then for any $\lambda > 0$, $\|g_n - g_0 + \lambda h_n\|_2^2 = O(n^{-vr} + n^{-\frac{1-v}{2}})$. Let

$$
\begin{aligned}
&J_n(\lambda) \equiv M_n(g_n + \lambda h_n) \\
=&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i}[Y_i(T_{K_i,j}) - (g_n + \lambda h_n)(\mathbf{X}_i, Z_i(T_{K_i,j}, W_i), T_{K_i,j})]^2 \xi_i(T_{K_i,j}),
\end{aligned}
$$

then

$$
\begin{aligned}
J_n'(\lambda) =& \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i}[(g_n + \lambda h_n)(\mathbf{X}_i, Z_i(T_{K_i,j}, W_i), T_{K_i,j}) - Y_i(T_{K_i,j})] \\
& \times h_n(\mathbf{X}_i, Z_i(T_{K_i,j}, W_i), T_{K_i,j})\xi_i(T_{K_i,j}),
\end{aligned}
$$

and

$$
{J_n}''(\lambda) = \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i} h_n^2(\mathbf{X}_i, Z_i(T_{K_i,j}, W_i), T_{K_i,j})\xi_i(T_{K_i,j}) \geq 0.
$$

Thus, $J_n'(\lambda)$ is a nondecreasing function. Therefore, to prove the convergence of $\hat{g}_n$ to g_0, it is sufficient to show that $\forall \lambda_0 > 0, J_n'(\lambda_0) > 0$ and $J_n'(-\lambda_0) < 0$ except on an event with probability converging to zero. Then $\hat{g}_n$ must be between $g_n - \lambda_0 h_n$ and $g_n + \lambda_0 h_n$, and so $\|\hat{g}_n - g_n\|_2^2 \leq \lambda_0^2 \|h_n\|_2^2 = O(n^{-vr} + n^{-\frac{1-v}{2}})$.

Next, we show that $J_n'(\lambda_0) > 0$. Define $\mathbb{G}_n = (\mathbb{P}_n - P)$, and

$$
\begin{aligned}
&\frac{1}{2} J_n'(\lambda_0) \\
=&\mathbb{G}_n \sum_{j=1}^{K} [(g_n + \lambda_0 h_n)(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) - Y(T_{K,j})] h_n(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) \xi(T_{K,j}) \\
+&P \sum_{j=1}^{K} [(g_n + \lambda_0 h_n)(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) - Y(T_{K,j})] h_n(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) \xi(T_{K,j}) \\
\equiv& I_{1n} + I_{2n}.
\end{aligned}
$$

By the calculation of Shen and Wong (1994, P. 597), for $\eta > 0$ and any $\varepsilon \leq \eta$,

$$\log N_{[]}(\varepsilon, \Psi_{l,\mathcal{I}}, L_2(\nu_1)) \leq c_1 q_n \log(\eta/\varepsilon),$$

where $q_n = m_n + l$ is the number of spline basis functions and c_1 is a constant.Then

$$
\begin{aligned}
J_{[]}(\eta, \mathcal{M}_\eta, L_2(\upsilon_1)) &= \int_0^\eta \{\log N_{[]}(\varepsilon, \mathcal{M}_\eta, L_2(\upsilon_1))\}^{1/2} d\varepsilon \\
&\leq \int_0^\eta \{c_1 q_n \log \eta/\varepsilon\}^{1/2} d\varepsilon \\
&= -\eta u e^{\frac{-u^2}{c_1 q_n}} \Big|_0^\infty + \eta \int_0^\infty e^{\frac{-u^2}{c_1 q_n}} du \quad (\{c_1 q_n \log \eta/\varepsilon\}^{1/2} = u) \\
&\leq \eta \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\sqrt{\frac{c_1 q_n}{2}}} e^{-\frac{u^2}{2(\sqrt{\frac{c_1 q_n}{2}})^2}} du \sqrt{2\pi} \sqrt{\frac{c_1 q_n}{2}} \\
&\leq \eta \sqrt{c_1 \pi} \sqrt{q_n} \leq c_2 q_n^{1/2} \eta,
\end{aligned}
$$

for a constant c_2. Thus, by Theorem 2.5.2 of Van der Vaart and Wellner (1996, P. 127) (Theorem C.3), $\Psi_{l,\mathcal{I}}$ is a Donsker class. Then given g_n defined before,

$$\mathcal{G}_\eta \equiv \Big\{ \sum_{j=1}^{K} [h(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) - Y(T_{K,j})](h - g_n)(\mathbf{X}, Z(T_{K,j}, W), T_{K,j})\xi(T_{K,j}) : \\ h(\mathbf{x}, z, t) = \beta'\mathbf{x} + \alpha' z(t, w) + \phi(t), \phi \in \Psi_{l,\mathcal{I}}, \|h - g_n\|_2 \le \eta \Big\}$$

is a Donsker class. Thus, $I_{1n} = O_p(n^{-1/2})$.

$$\begin{aligned}
I_{2n} &= E\Big[\int_0^\tau (g_n + \lambda_0 h_n)(\mathbf{X}, Z(t, W), t) h_n(\mathbf{X}, Z(t, W), t)\xi(t) dH(t)\Big] \\
&\quad - E\Big[\int_0^\tau Y(t) h_n(\mathbf{X}, Z(t, W), t)\xi(t) dH(t)\Big] \\
&= E\Big[\int_0^\tau (\lambda_0 h_n + g_n - g_0)(\mathbf{X}, Z(t, W), t) h_n(\mathbf{X}, Z(t, W), t)\xi(t) dH(t)\Big] \\
&\ge E\Big[\int_0^\tau \lambda_0 h_n^2(\mathbf{X}, Z(t, W), t)\xi(t) dH(t)\Big] = \lambda_0 \|h_n\|_2^2.
\end{aligned}$$

The second equality in the above formation is satisfied since

$$\begin{aligned}
& E\Big[\int_0^\tau Y(t) h_n(\mathbf{X}, Z(t, W), t)\xi(t) dH(t)\Big] \\
= \;& E\Big[\int_0^\tau h_n(\mathbf{X}, Z(t, W), t) E\{\xi(t) Y(t) dH(t) | \mathbf{X}, C, \mathcal{F}_t\}\Big] \\
= \;& E\Big[\int_0^\tau h_n(\mathbf{X}, Z(t, W), t) g_0(\mathbf{X}, Z(t, W), t)\xi(t) E\{dH(t) | \mathbf{X}, C, \mathcal{F}_t\}\Big] \\
= \;& E\Big[\int_0^\tau h_n(\mathbf{X}, Z(t, W), t) g_0(\mathbf{X}, Z(t, W), t)\xi(t) dH(t)\Big]
\end{aligned}$$

under the assumption (2.4). Thus, $\frac{1}{2} J_n'(\lambda_0) \ge O_p(n^{-1/2}) + \lambda_0 \|h_n\|_2^2 > 0$, since $\|h_n\|_2^2 = O(p_n^{-1})$ with $p_n^{-1} \equiv n^{-vr} + n^{-\frac{1-v}{2}} \ge n^{-\frac{r}{1+2r}} > n^{-1/2}$ for $0 < v < 1/2$.

For $J_n'(-\lambda_0)$,

$$
\begin{aligned}
&\frac{1}{2}J_n'(-\lambda_0)\\
=&\mathbb{G}_n\sum_{j=1}^{K}[(g_n-\lambda_0 h_n)(\mathbf{X},Z(T_{K,j},W),T_{K,j})-Y(T_{K,j})]h_n(\mathbf{X},Z(T_{K,j},W),T_{K,j})\xi(T_{K,j})\\
+&P\sum_{j=1}^{K}[(g_n-\lambda_0 h_n)(\mathbf{X},Z(T_{K,j},W),T_{K,j})-Y(T_{K,j})]h_n(\mathbf{X},Z(T_{K,j},W),T_{K,j})\xi(T_{K,j})\\
\equiv&I_{1n}^*+I_{2n}^*.
\end{aligned}
$$

Using the same arguments as for $J_n'(\lambda_0)$, $I_{1n}^*=O_P(n^{-1/2})$.

$$
\begin{aligned}
I_{2n}^* &= E\left[\int_0^\tau (g_n-\lambda_0 h_n)(\mathbf{X},Z(t,W),t)h_n(\mathbf{X},Z(t,W),t)\xi(t)dH(t)\right]\\
&\quad - E\left[\int_0^\tau h_n(\mathbf{X},Z(t,W),t)\xi(t)Y(t)dH(t)\right]\\
&= E\left[\int_0^\tau (-\lambda_0 h_n+g_n-g_0)(\mathbf{X},Z(t,W),t)h_n(\mathbf{X},Z(t,W),t)\xi(t)dH(t)\right]\\
&= -\lambda_0 E\left[\int_0^\tau h_n^2(\mathbf{X},Z(t,W),t)\xi(t)dH(t)\right]\\
&\quad + E\left[\int_0^\tau (g_n-g_0)(\mathbf{X},Z(t,W),t)h_n(\mathbf{X},Z(t,W),t)\xi(t)dH(t)\right]\\
&\le -\lambda_0\|h_n\|_2^2+\|g_n-g_0\|_2\|h_n\|_2\\
&\le -c_3 p_n^{-1},
\end{aligned}
$$

for a positive constant c_3. Thus, $\frac{1}{2}J_n'(-\lambda_0)\le O(n^{-1/2})-c_3p_n^{-1}<0$.

Then we have $\|\hat{g}_n - g_0\|_2 \leq \|\hat{g}_n - g_n\|_2 + \|g_n - g_0\|_2 = O(p_n^{-1/2})$, and

$$
\begin{aligned}
\|\hat{g}_n - g_0\|_2 &= \|(\hat{\beta}_n - \beta_0)'\mathbf{x} + (\hat{\alpha}_n - \alpha_0)'z + (\hat{\mu}_n - \mu_0)\|_2 \\
&= \|(\hat{\beta}_n - \beta_0)'(\mathbf{x} - \mathbf{a}^*) + (\hat{\alpha}_n - \alpha_0)'(z - \mathbf{b}^*) \\
&\quad + (\hat{\beta}_n - \beta_0)'\mathbf{a}^* + (\hat{\alpha}_n - \alpha_0)'\mathbf{b}^* + (\hat{\mu}_n - \mu_0)\|_2 \\
&= \|(\hat{\beta}_n - \beta_0)'(\mathbf{x} - \mathbf{a}^*) + (\hat{\alpha}_n - \alpha_0)'(z - \mathbf{b}^*)\|_2 \\
&\quad + \|(\hat{\beta}_n - \beta_0)'\mathbf{a}^* + (\hat{\alpha}_n - \alpha_0)'\mathbf{b}^* + (\hat{\mu}_n - \mu_0)\|_2.
\end{aligned}
$$

By C6, we can get $\|\hat{\beta}_n - \beta_0\| \longrightarrow 0$ and $\|\hat{\alpha}_n - \alpha_0\| \longrightarrow 0$ almost surely from the first term of the right hand side of the above equality and thus it follows that $\|\hat{\mu}_n - \mu_0\|_{L_2(\nu_1)} \longrightarrow 0$. This completes the proof of the theorem.

Theorem 2.2 (**Rate of Convergence**). *Suppose that C1 - C6 hold, then*

$$
\|\hat{\beta}_n - \beta_0\| = O_P(n^{-\frac{1-v}{2}}), \|\hat{\alpha}_n - \alpha_0\| = O_P(n^{-\frac{1-v}{2}}), \|\hat{\mu}_n - \mu_0\|_{L_2(\nu_1)} = O_P(n^{-\frac{1-v}{2}}).
$$

Remark 2.1. *When $v = 1/(1+2r)$, $n^{-\frac{1-v}{2}} = n^{-\frac{r}{1+2r}}$, we conclude from Stone (1980, 1982) that the rate of convergence of the estimator $\hat{\mu}_n$ is the optimal rate in nonparametric regression.*

Proof of Theorem 2.2.

For any $\eta > 0$, let

$$
\mathcal{F}_\eta \equiv \{g = \beta'\mathbf{x} + \alpha'z + \mu : \|\beta - \beta_0\| \leq \eta, \|\alpha - \alpha_0\| \leq \eta, \mu \in \Psi_{l,\mathcal{I}}, \|\mu - \mu_0\|_{L_2(v_1)} \leq \eta\}.
$$

Similar to Lemma A.2 in Huang (1999, P. 1557) given in Appendix A, for any $\varepsilon \leq \eta$, $\log N_{[]}(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2) \leq c_4 q_n \log(\eta/\varepsilon)$ for a constant c_4. Thus, for $\varepsilon > 0$, there exists a set of brackets $\{[g_i^l, g_i^r], i = 1, \cdots, (\frac{\eta}{\varepsilon})^{c_4 q_n}\}$ such that, for each $g \in \mathcal{F}_\eta$, there is a $[g_s^l, g_s^r]$, s.t. $g_s^l(\mathbf{x}, z, t) \le g(\mathbf{x}, z, t) \le g_s^r(\mathbf{x}, z, t)$, for all $\mathbf{x}$, $t \in [0, \tau]$ and $z \in \mathcal{G}$, and $\|g_s^r - g_s^l\|_2^2 \le \varepsilon^2$.

By Theorem 2.1, $\hat{g}_n \in \mathcal{F}_\eta$, for any $\eta > 0$ and sufficiently large n.

Next, consider the class $\mathcal{M}_\eta \equiv \{m_g(\mathbf{O}) - m_{g_0}(\mathbf{O}) : g \in \mathcal{F}_\eta\}$, where

$$m_g(\mathbf{O}) = \sum_{j=1}^{K} \{Y(T_{K,j}) - g(\mathbf{X}, Z(T_{K,j}, W), T_{K,j})\}^2 .$$

For $i = 1, \cdots, (\frac{\eta}{\varepsilon})^{c_4 q_n}$, define

$$\begin{aligned} m_i^l(\mathbf{O}) = \sum_{j=1}^{K} \Big\{ & 2Y(T_{K,j}) g_0(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) - g_0^2(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) \\ & + [\min\{|g_i^l(\mathbf{X}, Z(T_{K,j}, W)|, T_{K,j}), |g_i^r(\mathbf{X}, Z(T_{K,j}, W), T_{K,j})|\}]^2 \\ & - 2Y(T_{K,j})\{g_i^r(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) I(Y \ge 0) \\ & + g_i^l(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) I(Y < 0)\} \Big\} \xi(T_{K,j}), \end{aligned}$$

$$\begin{aligned} m_i^r(\mathbf{O}) = \sum_{j=1}^{K} \Big\{ & 2Y(T_{K,j}) g_0(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) - g_0^2(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) \\ & + [\max\{|g_i^l(\mathbf{X}, Z(T_{K,j}, W), T_{K,j})|, |g_i^r(\mathbf{X}, Z(T_{K,j}, W), T_{K,j})|\}]^2 \\ & - 2Y(T_{K,j})\{g_i^l(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) I(Y \ge 0) \\ & + g_i^r(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) I(Y < 0)\} \Big\} \xi(T_{K,j}) \end{aligned}$$

It is easy to show that $P|m_i^r(\mathbf{O}) - m_i^l(\mathbf{O})|^2 \le c_5\varepsilon^2$ with a constant c_5. In fact,

$$
\begin{aligned}
&P|m_i^r(\mathbf{O}) - m_i^l(\mathbf{O})|^2 \\
=&P|\int_0^\tau (|g_i^l| \wedge |g_i^r|)^2 - (|g_i^l| \vee |g_i^r|)^2 + (g_i^r - g_i^l)2Y[I(Y<0) - I(Y \ge 0)]d\tilde{H}(t)| \\
\le&P\int_0^\tau \left||g_i^r|^2 - |g_i^l|^2 + M_3(g_i^r - g_i^l)\right|^2 d\tilde{H}(t) \\
=&P\int_0^\tau \left|(|g_i^r| + |g_i^l|)(|g_i^r| - |g_i^l|) + M_3(g_i^r - g_i^l)\right|^2 d\tilde{H}(t) \\
\le&c_5\|g_s^r - g_s^l\|_2^2 \le c_5\varepsilon^2.
\end{aligned}
$$

where $a \vee b = \min\{a,b\}$ and $a \wedge b = \max\{a,b\}$, and M_3 is a constant. Thus $\{[m_i^l(\mathbf{O}), m_i^r(\mathbf{O})], i = 1, \cdots, (\frac{\eta}{\varepsilon})^{c_4q_n}\}$ is the set of brackets for $\mathcal{M}_\eta$, which implies that $\log N_{[]}(\varepsilon, \mathcal{M}_\eta, L_2(P)) \le c_4q_n \log(\eta/\varepsilon)$.

Moreover, by some calculations, we can verify that $P|m_g(\mathbf{O}) - m_{g_0}(\mathbf{O})|^2 \le c_6\eta^2$ for any $g \in \mathcal{F}_\eta$ by C4. Therefore, by Lemma 3.4.2 of Van der Vaart and Wellner (1996) (Lemma C.5), we obtain

$$
E\|n^{1/2}(\mathbb{P} - P)\|_{\mathcal{M}_\eta} \le c_7 J_{[]}(\eta, \mathcal{M}_\eta, L_2(P)) \left\{1 + \frac{J_{[]}(\eta, \mathcal{M}_\eta, L_2(P))}{\eta^2 n^{1/2}} M_3\right\}, \quad (2.6)
$$

where M_3 is a constant and $\|n^{1/2}(\mathbb{P} - P)\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} |n^{1/2}(\mathbb{P} - P)f|$, and

$$
\begin{aligned}
\tilde{J}_{[]}(\eta, \mathcal{M}_\eta, L_2(P)) &= \int_0^\eta \{1 + \log N_{[]}(\varepsilon, \mathcal{M}_\eta, L_2(P))\}^{1/2} d\varepsilon \\
&\le \int_0^\eta \{1 + c_4q_n \log \eta/\varepsilon\}^{1/2} d\varepsilon \\
&= -\eta u e^{\frac{1-u^2}{c_4q_n}}\Big|_1^\infty + \eta\int_1^\infty e^{\frac{1-u^2}{c_4q_n}} du \qquad (\{1 + c_4q_n \log \eta/\varepsilon\}^{1/2} = u) \\
&\le \eta + \eta\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\sqrt{\frac{c_4q_n}{2}}} e^{-\frac{u^2}{2(\sqrt{\frac{c_4q_n}{2}})^2}} du\sqrt{2\pi}\sqrt{\frac{c_4q_n}{2}} e^{1/c_4q_n} \\
&\le \eta + \eta\sqrt{c_4\pi}\sqrt{q_n} e^1 \le c_8 q_n^{1/2}\eta.
\end{aligned}
$$

The right hand side of (2.6) yields $\varphi_n(\eta) = c_9(q_n^{1/2}\eta + q_n/n^{1/2})$. It is easy to see that $\varphi_n(\eta)/\eta$ is decreasing in η, and

$$r_n^2\varphi_n(\frac{1}{r_n}) = r_n q_n^{1/2} + r_n^2 q_n/n^{1/2} \leq 2n^{1/2},$$

for $r_n = n^{\frac{1-v}{2}}$ and $0 < v < 1/2$.

Note that

$$\begin{aligned}
&Pm_g(O) - Pm_{g_0}(O) \\
&= P\left[\int_0^\tau \left\{(Y(t) - g(\mathbf{X}, Z(t, W), t))^2 - (Y(t) - g_0(\mathbf{X}, Z(t, W), t))^2\right\}\xi(t)dH(t)\right] \\
&= E\left\{\int_0^\tau (g - g_0)^2(\mathbf{X}, Z(t, W), t)\xi(t)dH(t)\right\} \\
&= \|g - g_0\|_2^2.
\end{aligned}$$

Thus, by Theorem 3.2.5 of Van der Vaart and Wellner (1996) (Theorem C.4), $n^{\frac{1-v}{2}}\|\hat{g}_n - g_0\|_2 = O_p(1)$. Therefore by the similar arguments as those in the proof of consistency of $\hat{\beta}_n$, $\hat{\alpha}_n$, and $\hat{\mu}_n$, we can get the rate of convergence of $\hat{\beta}_n$, $\hat{\alpha}_n$, and $\hat{\mu}_n$, as stated in the Theorem. The choice of $v = 1/(1+2r)$ yields the rate of convergence of $r/(1+2r)$, which completes the proof.

Theorem 2.3 (**Asymptotic Normality**). *Suppose that conditions C1 - C6 hold. Let* $\mathcal{H} \equiv \{(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3) : (\boldsymbol{h}_1', \boldsymbol{h}_2')' \in \mathcal{R}, h_3 \in \mathcal{F}_r, \|\boldsymbol{h}_1\| \leq 1, \|\boldsymbol{h}_2\| \leq 1, \|h_3\|_\infty \leq 1\}$. *Then for any* $(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3) \in \mathcal{H}$,

$$\boldsymbol{h}_1'\sqrt{n}(\hat{\beta}_n - \beta_0) + \boldsymbol{h}_2'\sqrt{n}(\hat{\alpha}_n - \alpha_0) + \int_0^\tau \sqrt{n}(\hat{\mu}_n - \mu_0)(t)dh_3(t)$$

converges in distribution to $N(0, \sigma^2)$, *where* σ^2 *is given in (2.7).*

Since $\hat{\mu}_n(t) = \sum_{i=1}^{q_n} \hat{\theta}_{ni} B_{il}(t)$, for estimation of the covariance matrix of $(\hat{\beta}_n, \hat{\alpha}_n, \hat{\theta}_n)$, we propose to use the following simple bootstrap procedure (Efron, 1979). Let L denote a prespecified positive integer. For each l, where $1 \leq l \leq L$, draw a simple random sample of size n,

$$\begin{aligned}\mathbf{O}^{(l)} = \{&K_i^{(l)}, T^{(l)}_{K_i^{(l)},1}, \cdots, T^{(l)}_{K_i^{(l)},K_i^{(l)}}, Y_i^{(l)}(T^{(l)}_{K_i^{(l)},1}), \cdots, Y_i^{(l)}(T^{(l)}_{K_i^{(l)},K_i^{(l)}}),\\ &H_i^{(l)}(T^{(l)}_{K_i^{(l)},1}), \cdots, H_i^{(l)}(T^{(l)}_{K_i^{(l)},K_i^{(l)}}), \mathbf{X}_i^{(l)}, C_i^{(l)}, i = 1, \cdots, n\}\end{aligned}$$

with replacement from the observed data

$$\begin{aligned}\mathbf{O} = \{&K_i, T_{K_i,1}, \cdots, T_{K_i,K_i}, Y_i(T_{K_i,1}), \cdots, Y_i(T_{K_i,K_i}),\\ &H_i(T_{K_i,1}), \cdots, H_i(T_{K_i,K_i}), \mathbf{X}_i, C_i, i = 1, \cdots, n\}.\end{aligned}$$

Let $(\hat{\beta}_n^{(l)}, \hat{\alpha}_n^{(l)}, \hat{\theta}_n^{(l)})$ be the proposed estimate of $(\beta_0, \alpha_0, \theta_0)$ based on the data set $\mathbf{O}^{(l)}$ defined above. Then according to Appendix D, a natural estimate of the covariance matrix of $(\hat{\beta}_n, \hat{\alpha}_n, \hat{\theta}_n)$ is given by

$$\hat{\Sigma}_L = \frac{1}{L-1} \sum_{l=1}^{L} \left\{ \begin{pmatrix} \hat{\beta}_n^{(l)} \\ \hat{\alpha}_n^{(l)} \\ \hat{\theta}_n^{(l)} \end{pmatrix} - \frac{1}{L} \sum_{l=1}^{L} \begin{pmatrix} \hat{\beta}_n^{(l)} \\ \hat{\alpha}_n^{(l)} \\ \hat{\theta}_n^{(l)} \end{pmatrix} \right\}^{\otimes 2}.$$

Denote the upper left $(p+q) \times (p+q)$ submatrix of $\hat{\Sigma}_L$ by $\hat{\Sigma}_{L,\beta,\alpha}$, which is the consistent estimator for the covariance matrix of $\sqrt{n} \begin{pmatrix} \hat{\beta}_n - \beta_0 \\ \hat{\alpha}_n - \alpha_0 \end{pmatrix}$.

Proof of Theorem 2.3.

We define a sequence of maps S_n mapping a neighborhood of $(\beta_0, \alpha_0, \mu_0)$, denoted by $\mathcal{U}$, in the parameter space for (β, α, μ) into $l^\infty(\mathcal{H})$ as :

$$\begin{aligned}
& S_n(\beta, \alpha, \mu)[\mathbf{h}_1, \mathbf{h}_2, h_3] \\
\equiv\; & n^{-1} \frac{d}{d\varepsilon} L_n(\beta + \varepsilon \mathbf{h}_1, \alpha + \varepsilon \mathbf{h}_2, \mu + \varepsilon h_3)\Big|_{\varepsilon=0} \\
=\; & -\frac{2}{n} \sum_{i=1}^{n} \int_0^\tau [Y_i(t) - \beta' \mathbf{X}_i - \alpha' h(\mathcal{F}_{it}, W_i) - \mu(t)] \\
& \quad \times [\mathbf{h}_1' \ \mathbf{X}_i + \mathbf{h}_2' h(\mathcal{F}_{it}, W_i) + h_3(t)] d\tilde{H}_i(t) \\
\equiv\; & \mathbb{P}_n \psi(\beta, \alpha, \mu)[\mathbf{h}_1, \ \mathbf{h}_2, h_3].
\end{aligned}$$

Correspondingly, we define the limit map $S : \mathcal{U} \longrightarrow l^\infty(\mathcal{H})$ as $S(\beta, \alpha, \mu)[\mathbf{h}_1, \ \mathbf{h}_2, h_3]$, where $l^\infty(\mathcal{H})$ is the space of bounded functionals on $\mathcal{H}$ under the supermum norm $\|f\|_\infty = \sup_{h \in \mathcal{H}} |f(h)|$.

To derive the asymptotic normality of the estimators $(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n)$, motivated by the proof of Theorem 3.3.1 of Van der Vaart and Wellner (1996, p. 310), we first need to verify the following five conditions.

(i) $\sqrt{n}(S_n - S)(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n) - \sqrt{n}(S_n - S)(\beta_0, \alpha_0, \mu_0) = o_p(1)$.

(ii) $\sqrt{n}(S_n - S)(\beta_0, \alpha_0, \mu_0)$ converges in distribution to a tight Gaussian process on $l^\infty(\mathcal{H})$.

(iii) $S(\beta_0, \alpha_0, \mu_0) = 0$ and $S_n(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n) = o_p(n^{-1/2})$.

(iv) $(\beta, \alpha, \mu) \longmapsto S(\beta, \alpha, \mu)$ is Fréchet-differentiable at $(\beta_0, \alpha_0, \mu_0)$ with a continuously invertible derivative $\dot{S}(\beta_0, \alpha_0, \mu_0)$.

(v) $\sqrt{n}\Big(S(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n) - S(\beta_0, \alpha_0, \mu_0)\Big) - \sqrt{n}\dot{S}(\beta_0, \alpha_0, \mu_0)\Big((\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n) - (\beta_0, \alpha_0, \mu_0)\Big) = o_p(1)$.

Note that

$$\begin{aligned}
&\sqrt{n}(S_n - S)(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n) - \sqrt{n}(S_n - S)(\beta_0, \alpha_0, \mu_0) \\
= \; &\sqrt{n}(\mathbb{P}_n - P)(\psi(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n)[\mathbf{h}_1,\ \mathbf{h}_2, h_3] - \psi(\beta_0, \alpha_0, \mu_0)[\mathbf{h}_1,\ \mathbf{h}_2, h_3]).
\end{aligned}$$

Define

$$\rho((\beta_1, \alpha_1, \mu_1), (\beta_2, \alpha_2, \mu_2)) = \left\{||\beta_1 - \beta_2||^2 + ||\alpha_1 - \alpha_2||^2 + ||\mu_1 - \mu_2||^2_{L_2(\nu_1)}\right\}^{1/2}$$

and for $\delta > 0$,

$$\begin{aligned}
\mathcal{F}_\delta = &\Big\{\psi(\beta, \alpha, \mu)[\mathbf{h}_1,\ \mathbf{h}_2, h_3] - \psi(\beta_0, \alpha_0, \mu_0)[\mathbf{h}_1,\ \mathbf{h}_2, h_3] : \\
&\rho\big((\beta, \alpha, \mu) - (\beta_0, \alpha_0, \mu_0)\big) < \delta, (\mathbf{h}_1,\ \mathbf{h}_2, h_3) \in \mathcal{H}\Big\}.
\end{aligned}$$

It is easy to see that $\mathcal{F}_r \subseteq C^r[0, \tau]$ is a Donsker class from Van der Vaart and Wellner (1996, p157), thus $\mathcal{H}$ is a Donsker class and $\psi(\beta, \alpha, \mu)$ is a bounded Lipschitz functional with respect to $\mathcal{H}$, thus $\mathcal{F}_\delta$ is a Donsker class for some $\delta > 0$. And

$$\begin{aligned}
&P\Big|\big[\psi(\beta_1, \alpha_1, \mu_1) - \psi(\beta_2, \alpha_2, \mu_2)\big][\mathbf{h}_1,\ \mathbf{h}_2, h_3]\Big|^2 \\
=&P\Big|2\int_0^\tau \big[(\beta_1 - \beta_2)'\ \mathbf{X} - (\alpha_1 - \alpha_2)'h(\mathcal{F}_t, W) - (\mu_1 - \mu_2)(t)\big] \\
&\quad \times [\mathbf{h}_1'\ \mathbf{X} + \mathbf{h}_2' h(\mathcal{F}_t, W) + h_3(t)]d\tilde{H}(t)\Big|^2 \\
\leq &c\rho^2\big((\beta_1, \alpha_1, \mu_1) - (\beta_2, \alpha_2, \mu_2)\big)
\end{aligned}$$

for a constant c. Thus condition (i) holds by Kosorok (2008, Lemma 13.3) (Lemma C.6).

Condition (ii) is also satisfied since $\{\psi(\beta_0, \alpha_0, \mu_0)[\mathbf{h}_1,\ \mathbf{h}_2, h_3] : (\mathbf{h}_1,\ \mathbf{h}_2, h_3) \in \mathcal{H}\}$ is a Donsker class.

Clearly, $S(\beta_0, \alpha_0, \mu_0) = 0$. For $h_3 \in \mathcal{F}_r$, let h_{3n} be the B-spline function approximation of h_3 with $||h_{3n} - h_3||_\infty = O(n^{-vr})$, then we have

$$S_n(\hat{\alpha}_n, \hat{\beta}_n, \hat{\mu}_n)[\mathbf{h}_1, \mathbf{h}_2, h_{3n}] = 0.$$

Thus, for $(\mathbf{h}_1, \mathbf{h}_2, h_3) \in \mathcal{H}$,

$$\begin{aligned}
& n^{\frac{1}{2}} S_n(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n)[\mathbf{h}_1, \mathbf{h}_2, h_3] \\
= \ & n^{\frac{1}{2}} \left[S_n(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n)[\mathbf{h}_1, \mathbf{h}_2, h_3] - S_n(\hat{\alpha}_n, \hat{\beta}_n, \hat{\mu}_n)[\mathbf{h}_1, \mathbf{h}_2, h_{3n}] \right] \\
= \ & n^{\frac{1}{2}} (\mathbb{P}_n - P) \left[\psi(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n)[\mathbf{h}_1,\ \mathbf{h}_2, h_3] - \psi(\beta_0, \alpha_0, \mu_0)[\mathbf{h}_1,\ \mathbf{h}_2, h_3] \right] \\
& - n^{\frac{1}{2}} (\mathbb{P}_n - P) \left[\psi(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n)[\mathbf{h}_1,\ \mathbf{h}_2, h_{3n}] - \psi(\beta_0, \alpha_0, \mu_0)[\mathbf{h}_1,\ \mathbf{h}_2, h_{3n}] \right] \\
& + n^{\frac{1}{2}} \mathbb{P}_n \left[\psi(\beta_0, \alpha_0, \mu_0)[\mathbf{h}_1,\ \mathbf{h}_2, h_3] - \psi(\beta_0, \alpha_0, \mu_0)[\mathbf{h}_1,\ \mathbf{h}_2, h_{3n}] \right] \\
& + n^{\frac{1}{2}} P \left[\psi(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n)[\mathbf{h}_1,\ \mathbf{h}_2, h_3] - \psi(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n)[\mathbf{h}_1,\ \mathbf{h}_2, h_{3n}] \right] \\
\equiv \ & Q_{1n} - Q_{2n} + Q_{3n} + Q_{4n}.
\end{aligned}$$

It follows from (i) that both Q_{1n} and Q_{2n} are $o_p(1)$. Q_{3n} is also $o_p(1)$ since

$$\begin{aligned}
& P \left[\psi(\beta_0, \alpha_0, \mu_0)[\mathbf{h}_1,\ \mathbf{h}_2, h_3] - \psi(\beta_0, \alpha_0, \mu_0)[\mathbf{h}_1,\ \mathbf{h}_2, h_{3n}] \right]^2 \\
& = P \left[2 \int_0^\tau \{ Y(t) - \beta_0' \mathbf{X} - \alpha_0' h(\mathcal{F}_t, W) - \mu_0(t) \} (h_{3n} - h_3) d\tilde{H}(t) \right]^2 \\
& \leq c \| h_{3n} - h_3 \|_\infty^2 \longrightarrow 0
\end{aligned}$$

for a constant c. Furthermore,

$$
\begin{aligned}
&|Q_{4n}| \\
&= \Big| -2n^{\frac{1}{2}}P\int_0^\tau \Big\{Y(t)-\hat\beta_n'\mathbf{X}-\hat\alpha_n'h(\mathcal{F}_t,W)-\hat\mu_n(t)\Big\}(h_3-h_{3n})d\tilde H(t)\Big| \\
&= \Big|2n^{\frac{1}{2}}P\int_0^\tau [(\hat\beta_n-\beta_0)'\mathbf{X}+(\hat\alpha_n-\alpha_0)'h(\mathcal{F}_t,W)+(\hat\mu_n-\mu_0)(t)](h_3-h_{3n})d\tilde H(t)\Big| \\
&\le cn^{\frac{1}{2}}\rho((\hat\beta,\hat\alpha,\hat\mu)-(\beta_0,\alpha_0,\mu_0))\|h_{3n}-h_3\|_\infty \\
&\le n^{\frac{1}{2}}O(n^{-\frac{1-v}{2}})\cdot O(n^{-vr}) \\
&= o_p(1),
\end{aligned}
$$

for a constant c. Thus, $S_n(\hat\beta_n,\hat\alpha_n,\hat\mu_n)=o_p(n^{-\frac{1}{2}})$.

For the proof of (iv), by the smoothness of $S(\beta,\alpha,\mu)$, the Fr*é*chet differentiability holds and the derivative of $S(\beta,\alpha,\mu)$ at (β_0,α_0,μ_0), denoted by $\dot S(\beta_0,\alpha_0,\mu_0)$, is a map from the space $\{(\beta-\beta_0,\alpha-\alpha_0,\mu-\mu_0):(\beta,\alpha,\mu)\in\mathcal{U}\}$ to $l^\infty(\mathcal{H})$ and

$$
\begin{aligned}
&\dot S(\beta_0,\alpha_0,\mu_0)(\beta-\beta_0,\alpha-\alpha_0,\mu-\mu_0)[\mathbf{h}_1,\mathbf{h}_2,h_3] \\
&= \frac{d}{d\epsilon}S(\beta_0+\varepsilon(\beta-\beta_0),\alpha_0+\varepsilon(\alpha-\alpha_0),\mu_0+\ \varepsilon(\mu-\mu_0))[\mathbf{h}_1,\mathbf{h}_2,h_3]\Big|_{\varepsilon=0} \\
&\equiv \sigma_1(\mathbf{h}_1,\mathbf{h}_2,h_3)'(\beta-\beta_0)+\sigma_2(\mathbf{h}_1,\mathbf{h}_2,h_3)'(\alpha-\alpha_0)+\int_0^\tau(\mu-\mu_0)d\sigma_3(\mathbf{h}_1,\mathbf{h}_2,h_3),
\end{aligned}
$$

where

$$\sigma_1(\mathbf{h}_1,\mathbf{h}_2,h_3)=2P\int_0^\tau[\mathbf{h}_1'\ \mathbf{X}+\mathbf{h}_2'h(\mathcal{F}_t,W)+h_3(t)]\mathbf{X}d\tilde H(t),$$

$$\sigma_2(\mathbf{h}_1,\mathbf{h}_2,h_3)=2P\int_0^\tau[\mathbf{h}_1'\ \mathbf{X}+\mathbf{h}_2'h(\mathcal{F}_t,W)+h_3(t)]h(\mathcal{F}_t,W)d\tilde H(t),$$

and

$$\sigma_3(\mathbf{h}_1,\mathbf{h}_2,h_3)(t)=2P\int_0^t[\mathbf{h}_1'\ \mathbf{X}+\mathbf{h}_2'h(\mathcal{F}_s,W)+h_3(s)]d\tilde H(s).$$

It remains to show that the linear map $\dot{S}(\beta_0, \alpha_0, \mu_0)$ is continuously invertible on its range. Following the proof of Theorem 2 in Zeng et al. (2005), we only need to show that for $h \in \mathcal{H}$, if $\sigma(h) = (\sigma_1(h), \sigma_2(h), \sigma_3(h)) = 0$ almost surely, then $h = 0$. Suppose that $\sigma(h) = 0$, a.s., then $\sigma_1(h)'\mathbf{h}_1 + \sigma_2(h)'\mathbf{h}_2 + \int_0^\tau h_3(t) d\sigma_3(h)(t) = 0$, i. e.

$$
\begin{aligned}
0 = {} & 2P\int_0^\tau [\mathbf{h}_1' \ \mathbf{X} + \mathbf{h}_2' h(\mathcal{F}_t, W) + h_3(t)]\mathbf{X}' d\tilde{H}(t)\mathbf{h}_1 \\
& + 2P\int_0^\tau [\mathbf{h}_1' \ \mathbf{X} + \mathbf{h}_2' h(\mathcal{F}_t, W) + h_3(t)] h(\mathcal{F}_t, W)' d\tilde{H}(t)\mathbf{h}_2 \\
& + 2P\int_0^\tau [\mathbf{h}_1' \ \mathbf{X} + \mathbf{h}_2' h(\mathcal{F}_t, W) + h_3(t)] h_3(t) d\tilde{H}(t) \\
= {} & 2P\int_0^\tau [\mathbf{h}_1' \ \mathbf{X} + \mathbf{h}_2' h(\mathcal{F}_t, W) + h_3(t)]^2 d\tilde{H}(t),
\end{aligned}
$$

which implies that $\mathbf{h}_1' \ \mathbf{X} + \mathbf{h}_2' h(\mathcal{F}_t, W) + h_3(t) = 0, a.s.$ Hence, $\mathbf{h}_1 = 0, \mathbf{h}_2 = 0, h_3 = 0, a.s.$ by C5.

Moreover, condition (v) holds since

$$
\begin{aligned}
& \Big(S(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n) - S(\beta_0, \alpha_0, \mu_0)\Big)[\mathbf{h}_1, \mathbf{h}_2, h_3] \\
= {} & -2P\int_0^\tau [Y(t) - \hat{\beta}_n'\mathbf{X} - \hat{\alpha}_n' h(\mathcal{F}_t, W) - \hat{\mu}_n][\mathbf{h}_1' \ \mathbf{X} + \mathbf{h}_2' h(\mathcal{F}_t, W) + h_3(t)] d\tilde{H}(t) \\
= {} & 2P\int_0^\tau [(\hat{\beta}_n - \beta_0)'\mathbf{X} + (\hat{\alpha}_n - \alpha_0)' h(\mathcal{F}_t, W) + (\hat{\mu}_n - \mu_0)(t)] \\
& \quad \times [\mathbf{h}_1' \ \mathbf{X} + \mathbf{h}_2' h(\mathcal{F}_t, W) + h_3(t)] d\tilde{H}(t).
\end{aligned}
$$

Therefore, by (i)-(v), we have

$$\begin{aligned}
&\sqrt{n}\dot{S}(\beta_0,\alpha_0,\mu_0)(\hat{\beta}_n-\beta_0,\hat{\alpha}_n-\alpha_0,\hat{\mu}_n-\mu_0)[\mathbf{h}_1,\mathbf{h}_2,h_3] \\
&= \sigma_1(\mathbf{h}_1,\mathbf{h}_2,h_3)'\sqrt{n}(\hat{\beta}_n-\beta_0)+\sigma_2(\mathbf{h}_1,\mathbf{h}_2,h_3)'\sqrt{n}(\hat{\alpha}_n-\alpha_0) \\
&\quad + \int_0^\tau \sqrt{n}(\hat{\mu}_n-\mu_0)(t)d\sigma_3(\mathbf{h}_1,\mathbf{h}_2,h_3)(t) \\
&= -\sqrt{n}(S_n-S)(\beta_0,\alpha_0,\mu_0)[\mathbf{h}_1,\mathbf{h}_2,h_3]+o_p(1),
\end{aligned}$$

uniformly in $\mathbf{h}_1$, $\mathbf{h}_2$ and h_3, and for each $(\mathbf{h}_1,\mathbf{h}_2,h_3)\in\mathcal{H}$, there exists unique $(\mathbf{h}_1^*,\mathbf{h}_2^*,h_3^*)\in\mathcal{H}$ such that $\sigma_1(\mathbf{h}_1^*,\mathbf{h}_2^*,h_3^*)=\mathbf{h}_1$, $\sigma_2(\mathbf{h}_1^*,\mathbf{h}_2^*,h_3^*)=\mathbf{h}_2$, $\sigma_3(\mathbf{h}_1^*,\mathbf{h}_2^*,h_3^*)=h_3$. Thus, we have

$$\begin{aligned}
&\mathbf{h}_1'\sqrt{n}(\hat{\beta}_n-\beta_0)+\mathbf{h}_2'\sqrt{n}(\hat{\alpha}_n-\alpha_0)+\int_0^\tau \sqrt{n}(\hat{\mu}_n-\mu_0)(t)dh_3(t) \\
&= \sqrt{n}\dot{S}(\beta_0,\alpha_0,\mu_0)(\hat{\beta}_n-\beta_0,\hat{\alpha}_n-\alpha_0,\hat{\mu}_n-\mu_0)[\mathbf{h}_1^*,\mathbf{h}_2^*,h_3^*] \\
&= -\sqrt{n}(S_n-S)(\beta_0,\alpha_0,\mu_0)[\mathbf{h}_1^*,\mathbf{h}_2^*,h_3^*]+o_p(1) \\
&\to Z \quad \text{in distribution,}
\end{aligned}$$

where Z follows $N(0,\sigma^2)$ with

$$\sigma^2 = E\psi^2(\beta_0,\alpha_0,\mu_0)[\mathbf{h}_1^*,\mathbf{h}_2^*,h_3^*]. \tag{2.7}$$

Find $\mathbf{h}_1^*, \mathbf{h}_2^*, h_3^*$ from $\sigma_1(\mathbf{h}_1^*, \mathbf{h}_2^*, h_3^*) = \mathbf{h}_1, \sigma_2(\mathbf{h}_1^*, \mathbf{h}_2^*, h_3^*) = \mathbf{h}_2, \sigma_3(\mathbf{h}_1^*, \mathbf{h}_2^*, h_3^*) = h_3$.

Solution. Find $(\mathbf{h}_1^*, \mathbf{h}_2^*, h_3^*)$ from the following three equations

$$\sigma_1(\mathbf{h}_1^*, \mathbf{h}_2^*, h_3^*) = 2P\int_0^\tau [\mathbf{h}_1^{*'}\ \mathbf{X} + \mathbf{h}_2^{*'} Z(t,W) + h_3^*(t)]\mathbf{X} d\tilde{H}(t) = \mathbf{h}_1 \quad (2.8)$$

$$\sigma_2(\mathbf{h}_1^*, \mathbf{h}_2^*, h_3^*) = 2P\int_0^\tau [\mathbf{h}_1^{*'}\ \mathbf{X} + \mathbf{h}_2^{*'} Z(t,W) + h_3^*(t)]Z(t,W) d\tilde{H}(t) = \mathbf{h}_2 \quad (2.9)$$

$$\sigma_3(\mathbf{h}_1^*, \mathbf{h}_2^*, h_3^*)(t) = 2P\int_0^t [\mathbf{h}_1^{*'}\ \mathbf{X} + \mathbf{h}_2^{*'} Z(s,W) + h_3^*(s)] d\tilde{H}(s) = h_3(t), \quad (2.10)$$

where $Z(t,W) = h(\mathcal{F}_t, W)$.

Define

$$A_{p\times p} = 2P\int_0^\tau \mathbf{X}\mathbf{X}' d\tilde{H}(t),$$

$$B_{q\times q} = 2P\int_0^\tau Z(t,W)Z'(t,W) d\tilde{H}(t),$$

$$D_{p\times q} = 2P\int_0^\tau \mathbf{X}Z'(t,W) d\tilde{H}(t),$$

$$a(t)_{1\times p} = 2P\int_0^t \mathbf{X}' d\tilde{H}(s),$$

and $b(t)_{1\times q} = 2P\int_0^t Z'(s,W)d\tilde{H}(s)$. We can rewrite (2.8), (2.9) and (2.10) as the following three equations,

$$A\mathbf{h}_1^* + D\mathbf{h}_2^* + 2P\int_0^\tau \mathbf{X}h_3^*(t) d\tilde{H}(t) = \mathbf{h}_1 \quad (2.11)$$

$$D'\mathbf{h}_1^* + B\mathbf{h}_2^* + 2P\int_0^\tau Z(t,W)h_3^*(t) d\tilde{H}(t) = \mathbf{h}_2 \quad (2.12)$$

$$a(t)\mathbf{h}_1^* + b(t)\mathbf{h}_2^* + 2P\int_0^t h_3^*(s) d\tilde{H}(s) = h_3(t), \quad (2.13)$$

By (2.11),

$$\mathbf{h}_1^* = A^{-1}[\mathbf{h}_1 - 2P\int_0^\tau \mathbf{X}h_3^*(t) d\tilde{H}(t) - D\mathbf{h}_2^*].$$

Substitute it into (2.12), we get

$$D'A^{-1}[\mathbf{h}_1 - 2P\int_0^\tau \mathbf{X}h_3^*(t)d\tilde{H}(t) - D\mathbf{h}_2^*] + B\mathbf{h}_2^* + 2P\int_0^\tau Z(t,W)h_3^*(t)d\tilde{H}(t) = \mathbf{h}_2$$

Then

$$\mathbf{h}_2^* = F\left\{\mathbf{h}_2 - 2P\int_0^\tau Z(t,W)h_3^*(t)d\tilde{H}(t) - D'A^{-1}[\mathbf{h}_1 - 2P\int_0^\tau \mathbf{X}h_3^*(t)d\tilde{H}(t)]\right\}, \tag{2.14}$$

where $F = [B - D'A^{-1}D]^{-1}$. Thus

$$\begin{aligned}\mathbf{h}_1^* = A^{-1}\Big[&\mathbf{h}_1 - 2P\int_0^\tau \mathbf{X}h_3^*(t)d\tilde{H}(t)\\ &- DF\big\{\mathbf{h}_2 - 2P\int_0^\tau Z(t,W)h_3^*(t)d\tilde{H}(t) - D'A^{-1}[\mathbf{h}_1 - 2P\int_0^\tau \mathbf{X}h_3^*(t)d\tilde{H}(t)]\big\}\Big].\end{aligned} \tag{2.15}$$

Finally, substitute $\mathbf{h}_1^*$ and $\mathbf{h}_2^*$ into (2.13),

$$\begin{aligned}&\big\{b'(t)FD'A^{-1} - a'(t)A^{-1} - a'(t)A^{-1}DFD'A^{-1}\big\}\,2P\int_0^\tau \mathbf{X}h_3^*(t)d\tilde{H}(t)\\ &+\big\{a'(t)A^{-1}DF - b'(t)F\big\}\,2P\int_0^\tau Z(t,W)h_3^*(t)d\tilde{H}(t) + 2P\int_0^t h_3^*(s)d\tilde{H}(s)\\ =&\big\{b'(t)FD'A^{-1} - a'(t)A^{-1} - a'(t)A^{-1}DFD'A^{-1}\big\}\,\mathbf{h}_1\\ &+\big\{a'(t)A^{-1}DF - b'(t)F\big\}\,\mathbf{h}_2. + h_3^*(t).\end{aligned}$$

Let $A^{*'}(t)_{1\times p} = b'(t)FD'A^{-1} - a'(t)A^{-1} - a'(t)A^{-1}DFD'A^{-1}$, $B^{*'}(t)_{1\times q} = a'(t)A^{-1}DF - b'(t)F$. Then the above equation becomes

$$\begin{aligned}&A^{*'}(t)2P\int_0^\tau \mathbf{X}h_3^*(t)d\tilde{H}(t) + B^{*'}(t)2P\int_0^\tau Z(t,W)h_3^*(t)d\tilde{H}(t) + 2P\int_0^t h_3^*(s)d\tilde{H}(s)\\ =&A^{*'}(t)\mathbf{h}_1 + B^{*'}(t)\mathbf{h}_2 + h_3^*(t).\end{aligned}$$

Thus we can find h_3^* from this equation, then $\mathbf{h}_1^*$ and $\mathbf{h}_2^*$ can be obtained from (2.14) and (2.15) respectively.

2.5 Simulation study

In this section, a simulation study was conducted to assess the finite sample properties of the proposed estimators. We generated the response variable from the following random-effects model:

$$Y_i(t) = \mu_0(t) + \beta_1 X_{1i} + \beta_2 X_{2i} + \alpha H_i(t-) W_i + \epsilon_i(t),$$

where X_{1i} and X_{2i} were generated from Bernoulli distribution with success probability 0.5 and the standard normal distribution, $\epsilon_i(t)$'s were independent standard normal variables, and $W_i = X_{1i}$ or X_{2i}. The follow-up time C_i was generated from the uniform distribution over interval $(\tau/2, \tau)$ with $\tau = 6$. The total number of real observation times for subject i, m_i, was assumed to follow the discrete uniform distribution over $\{1, 2, 3, 4, 5, 6\}$ and the observation times $(T_{m_i,1}, \ldots, T_{m_i,m_i})$ were taken to be the order statistics of a random sample of size m_i from the uniform distribution over $(0, C_i)$ given C_i.

The true parameter values used in our simulation studies are $\beta_0 = (\beta_{10}, \beta_{20})' = (-1, 1)'$, and $\alpha_0 = -1.5, -1, 0, 1$ or 1.5. The smooth function $\mu_0(t)$ was taken as $\sin(t/2)$ or $\log(t+1)$. To estimate $\mu_0(t)$, we considered cubic B-splines and took $m_n = n^\nu$ with $\nu = 1/10$, $1/3$ or $2/5$. For a given number of interior knots m_n, we consider two data-driven methods for determing locations of knots. One is the equally spaced knots, which are given by $T_{\min} + k(T_{\max} - T_{\min})/(m_n + 1), k =$

$0, 1, \cdots, m_n + 1$, with $T_{\min}$ and $T_{\max}$ being the respective minimum and maximum values of distinct observation times. Another is the partitions according to quantiles of the observation times, i. e., the $k/(m_n+1)$ quantiles ($k = 0, 1, \cdots, m_n+1$) of the distinct observation times are chosen to be the knots. We have done simulation for the six combinations of the number and placement of knots and illustrate the estimation results for different combinations with $W = X_1$, $\alpha = 1$, $\mu_0(t) = \log(t+1)$ and $n = 50$ in Table 2.1. From this table, we find that the estimation results are very similar and not sensitive to the selection of number and placement of knots. Thus in the following, we present the overall results with number of interior knots chosen to be $n^{1/10}$ and equally spaced knots.

Tables 2.2 and 2.3 present the simulation results on estimation of β_0 and α_0 with the sample size $n = 50$ or 100 and $W = X_1$, while Tables 2.4 and 2.5 present those with $W = X_2$. In the tables, we compare the proposed method with a competing method developed by Sun et al. (2005) (SPSZ), to demonstrate the robustness of the proposed method. All the tables include the estimated bias (BIAS) given by the average of the estimates minus the true value, the bootstrap standard errors of the estimates (BSE), the sample standard deviation of the estimates (SSE), and thebootstrap 95% coverage probabilities (CP) obtained from 1000 independent runs. Here, we used 200 replications in bootstrap to estimate the standard errors.

Figure 2.1 shows the estimation results of $\mu_0(t) = \log(t+1)$ for observation processes with $\alpha = 1$ and $h(\mathcal{F}_t, W) = H(t-)X_1$, where the sample size n was taken as 50 or 100, respectively. In the figure, the solid line represents the real curve of

$\mu_0(t)$, and the point line represents the B-splines based estimation curve of $\mu_0(t)$.

Based on our simulation results, we have the following findings: (i) For the case of $\alpha = 0$, both the proposed estimators and the estimators based on SPSZ's method are approximately unbiased, and the former are not worse than the latter. (ii) For the case of $\alpha \neq 0$, the proposed estimators are approximately unbiased while the estimators based on SPSZ's method yield biased estimates and the biases could be larger as α diverges from 0. In other words, the proposed estimation procedure seems to be more robust. The possible reason is that our estimation method is model-free for the observation process, while their estimation procedure replies on the model assumption about the observation process. (iii) The estimated curve of $\mu_0(t)$ is close to its real curve with the moderate sample size, indicating that the B-splines estimator for $\mu_0(t)$ works well. (iv) The sample standard errors and the bootstrap standard errors of the proposed estimators are close to each other. Also, the bootstrap 95% coverage rates are close to the nominal level, that is, the proposed spline based semiparametric bootstrap procedure provides reasonable estimates and the normal approximation seems to be appropriate.

Table 2.1. Estimation results of $(\beta_0', \alpha_0')'$ for simulated longitudinal data with non-Poisson observation process, $h(\mathcal{F}_t, W) = H(t-)X_1$, $\alpha = 1$, $\mu_0(t) = \log(t+1)$ and $n = 50$ by chosing different numbers and placements of knots..

ν	Partitions	$\hat{\beta}_1$				$\hat{\beta}_2$				$\hat{\alpha}$			
		BIAS	SSE	BSE	CP	BIAS	SSE	BSE	CP	BIAS	SSE	BSE	CP
$\nu = 1/10$	Q	-0.0101	0.3184	0.3097	0.9450	-0.0002	0.1629	0.1572	0.9490	0.0028	0.1111	0.1067	0.9530
	E	0.0141	0.3098	0.3098	0.9600	-0.0009	0.1647	0.1548	0.9370	0.0017	0.1111	0.1078	0.9520
$\nu = 1/3$	Q	0.0216	0.3021	0.2967	0.9490	-0.0018	0.1634	0.1558	0.9460	-0.0009	0.0960	0.0906	0.9310
	E	0.0154	0.3116	0.3103	0.9590	-0.0009	0.1646	0.1547	0.9380	0.0020	0.1121	0.1099	0.9500
$\nu = 2/5$	Q	-0.0117	0.3078	0.3000	0.9470	-0.0094	0.1700	0.1568	0.9240	0.0015	0.0999	0.0915	0.9310
	E	0.0209	0.3222	0.3156	0.9540	0.0037	0.1608	0.1599	0.9420	-0.0013	0.1131	0.1133	0.9520

Note: Q: Knots partitions according to quantiles of the the observetion times and E: Equally spaced knots based on the observetion times.

Table 2.2. Estimation results of $(\beta_0', \alpha_0')'$ for simulated longitudinal data with non-Poisson observation process, $h(\mathcal{F}_t, W) = H(t-)X_1$ and $\mu_0(t) = \sin(t/2)$

n	α	Method	$\hat{\beta}_1$				$\hat{\beta}_2$				$\hat{\alpha}$			
			BIAS	SSE	BSE	CP	BIAS	SSE	BSE	CP	BIAS	SSE	BSE	CP
50	-1.5	Proposed	-0.0232	0.3209	0.3099	0.9420	0.0001	0.1588	0.1571	0.9450	0.0029	0.1087	0.1077	0.9520
		SPSZ	-0.6804	0.3382			0.0011	0.1620			0.5019	0.1386		
	-1	Proposed	-0.0088	0.3166	0.3105	0.9470	0.0000	0.1593	0.1560	0.9450	0.0013	0.1130	0.1092	0.9430
		SPSZ	-0.4392	0.3183			-0.0012	0.1605			0.3307	0.1098		
	0	Proposed	0.0002	0.3219	0.3142	0.9440	0.0012	0.1640	0.1563	0.9310	0.0005	0.1164	0.1098	0.9380
		SPSZ	0.0151	0.3100			0.0008	0.1637			-0.0120	0.0806		
	1	Proposed	-0.0025	0.3158	0.3138	0.9520	0.0034	0.1614	0.1580	0.9430	-0.0056	0.1118	0.1085	0.9420
		SPSZ	0.4670	0.3193			0.0044	0.1629			-0.3611	0.1092		
	1.5	Proposed	-0.0076	0.3075	0.3112	0.9600	-0.0047	0.1649	0.1559	0.9360	0.0056	0.1114	0.1077	0.9430
		SPSZ	0.6889	0.3339			-0.0039	0.1681			-0.5250	0.1415		
100	-1.5	Proposed	-0.0149	0.2219	0.2195	0.9490	0.0013	0.1115	0.1096	0.9470	0.0008	0.0769	0.0759	0.9520
		SPSZ	-0.6733	0.2387			0.0004		0.1145		-0.0019	0.0966		
	-1	Proposed	0.0021	0.2350	0.2188	0.9330	-0.0004	0.1147	0.1092	0.9340	-0.0028	0.0794	0.0758	0.9420
		SPSZ	-0.4305	0.2401			0.0000	0.1148			0.3236	0.0796		
	0	Proposed	0.0071	0.2286	0.2200	0.9360	-0.0021	0.1133	0.1097	0.9390	0.0019	0.0751	0.0755	0.9540
		SPSZ	0.0270	0.2244			-0.0021	0.1131			-0.0136	0.0541		
	1	Proposed	0.0000	0.2296	0.2199	0.9390	-0.0004	0.1116	0.1102	0.9500	0.0005	0.0751	0.0763	0.9560
		SPSZ	0.4688	0.2372			-0.0006	0.1131			-0.3546	0.0797		
	1.5	Proposed	0.0037	0.2334	0.2199	0.9300	-0.0026	0.1099	0.1097	0.9570	0.0004	0.0801	0.0760	0.9280
		SPSZ	0.7025	0.2529			-0.0020	0.1129			-0.5278	0.1051		

Note: SPSZ: Method in Sun et al. (2005).

Table 2.3. Estimation results of $(\beta_0', \alpha_0')'$ for simulated longitudinal data with non-Poisson observation process, $h(\mathcal{F}_t, W) = H(t-)X_1$ and $\mu_0(t) = \log(t+1)$

n	α	Method	$\hat{\beta}_1$				$\hat{\beta}_2$				$\hat{\alpha}$			
			BIAS	SSE	BSE	CP	BIAS	SSE	BSE	CP	BIAS	SSE	BSE	CP
50	-1.5	Proposed	-0.012	0.3249	0.3109	0.9410	0.0022	0.1572	0.1565	0.9460	-0.0025	0.1133	0.1076	0.9240
		SPSZ	-0.6493	0.3425			0.0020	0.1612			-0.0163	0.1426		
	-1	Proposed	-0.0067	0.3104	0.3115	0.9510	0.0052	0.1592	0.1547	0.9420	-0.0050	0.1153	0.1068	0.9280
		SPSZ	-0.4143	0.3211			0.0047	0.1609			0.3077	0.1112		
	0	Proposed	0.0041	0.3259	0.3107	0.9380	0.0100	0.1656	0.1571	0.9360	0.0000	0.1127	0.1088	0.9380
		SPSZ	0.0465	0.3183			0.0106	0.1659			-0.0336	0.0794		
	1	Proposed	-0.0101	0.3184	0.3097	0.9450	-0.0022	0.1629	0.1572	0.9490	0.0028	0.1111	0.1067	0.9530
		SPSZ	0.4802	0.3246			-0.0023	0.1654			-0.3726	0.1255		
	1.5	Proposed	-0.0174	0.3278	0.3090	0.9240	-0.0004	0.1680	0.1552	0.9250	0.0024	0.1133	0.1076	0.9440
		SPSZ	0.6949	0.3570			-0.0010	0.1728			-0.5429	0.1487		
100	-1.5	Proposed	-0.0029	0.2281	0.2191	0.9400	0.0054	0.1109	0.1102	0.9520	-0.0034	0.0781	0.0755	0.9480
		SPSZ	-0.6319	0.2471			0.0064	0.1125			-0.0260	0.0989		
	-1	Proposed	0.0003	0.2183	0.2193	0.9430	-0.0005	0.1156	0.1097	0.9320	-0.0002	0.0766	0.0755	0.9440
		SPSZ	-0.4053	0.2227			-0.0003	0.1167			0.3044	0.0756		
	0	Proposed	-0.0145	0.2257	0.2217	0.9500	0.0004	0.1159	0.1115	0.9410	0.0047	0.0777	0.0764	0.9460
		SPSZ	0.0324	0.2227			0.0008	0.1169			-0.0317	0.0581		
	1	Proposed	-0.0037	0.1567	0.1548	0.9480	0.0028	0.0783	0.0773	0.9450	0.0011	0.0538	0.0535	0.9400
		SPSZ	0.4948	0.1629			0.0025	0.0792			-0.3743	0.0572		
	1.5	Proposed	0.0051	0.2168	0.2185	0.9560	-0.0040	0.1117	0.1091	0.9480	0.0050	0.0790	0.0754	0.9200
		SPSZ	0.7246	0.2394			-0.0031	0.1151			-0.5352	0.1035		

Note: SPSZ: Method in Sun et al. (2005).

Table 2.4. Estimation results of $(\beta_0', \alpha_0')'$ for simulated longitudinal data with non-Poisson observation process, $h(\mathcal{F}_t, W) = H(t-)X_2$ and $\mu_0(t) = \sin(t/2)$

n	α	Method	$\hat{\beta}_1$				$\hat{\beta}_2$				$\hat{\alpha}$			
			BIAS	SSE	BSE	CP	BIAS	SSE	BSE	CP	BIAS	SSE	BSE	CP
50	-1.5	Proposed	0.0141	0.3176	0.3048	0.9360	0.0017	0.1549	0.1533	0.9420	-0.0020	0.0628	0.0622	0.9350
		SPSZ	0.0092	0.3240			-0.5182	0.1956			0.3975	0.1094		
	-1	Proposed	-0.0019	0.3138	0.3067	0.9420	-0.0014	0.1534	0.1528	0.9510	0.0002	0.0619	0.0623	0.9550
		SPSZ	0.0002	0.3171			-0.3449	0.1705			0.2653	0.0793		
	0	Proposed	-0.0042	0.3158	0.3071	0.9510	-0.0083	0.1558	0.1523	0.9440	-0.0024	0.0607	0.0622	0.9530
		SPSZ	-0.0038	0.3146			0.0041	0.1631			-0.0116	0.0516		
	1	Proposed	-0.0027	0.3169	0.3078	0.9430	0.0056	0.1547	0.1542	0.9470	-0.0014	0.0618	0.0620	0.9460
		SPSZ	-0.0042	0.3205			0.3773	0.1759			-0.2885	0.0851		
	1.5	Proposed	0.0126	0.3239	0.3065	0.9350	-0.0003	0.1554	0.1530	0.9460	-0.0021	0.0639	0.0631	0.9470
		SPSZ	0.0117	0.3366			0.5498	0.1976			-0.4300	0.1190		
100	-1.5	Proposed	0.0045	0.2227	0.2170	0.9450	-0.0039	0.1017	0.1045	0.9570	0.0035	0.0415	0.0410	0.9560
		SPSZ	0.0035	0.2276			-0.5208	0.1275			0.3974	0.0753		
	-1	Proposed	0.0162	0.2207	0.2170	0.9510	0.0018	0.1026	0.1046	0.9570	0.0000	0.0414	0.0413	0.9450
		SPSZ	0.0158	0.2224			-0.3395	0.1159			0.2593	0.0553		
	0	Proposed	-0.0011	0.2277	0.2189	0.9460	0.0018	0.1033	0.1057	0.9590	0.0002	0.0421	0.0413	0.9460
		SPSZ	-0.0017	0.2287			0.0142	0.1046			-0.0094	0.0361		
	1	Proposed	-0.0033	0.2274	0.2174	0.9450	0.0013	0.1064	0.1055	0.9490	-0.0010	0.0410	0.0412	0.9530
		SPSZ	-0.0015	0.2341			0.3658	0.1234			0.7212	0.0610		
	1.5	Proposed	-0.0060	0.2277	0.2178	0.9510	0.0023	0.1032	0.1047	0.9570	-0.0016	0.0418	0.0418	0.9410
		SPSZ	-0.0094	0.2398			0.5462	0.1343			-0.4156	0.0787		

Note: SPSZ: Method in Sun et al. (2005).

Table 2.5. Estimation results of $(\beta_0', \alpha_0')'$ for simulated longitudinal data with non-Poisson observation process, $h(\mathcal{F}_t, W) = H(t-)X_2$ and $\mu_0(t) = \log(t+1)$

n	α	Method	$\hat{\beta}_1$				$\hat{\beta}_2$				$\hat{\alpha}$			
			BIAS	SSE	BSE	CP	BIAS	SSE	BSE	CP	BIAS	SSE	BSE	CP
50	-1.5	Proposed	-0.0254	0.3139	0.3067	0.9530	0.0072	0.1464	0.1526	0.9510	0.0023	0.0604	0.0615	0.9460
		SPSZ	-0.0239	0.3219			-0.5023	0.1785			0.3939	0.1067		
	-1	Proposed	0.0056	0.3118	0.3064	0.9540	0.0039	0.1503	0.1514	0.9510	0.0024	0.0611	0.0623	0.9590
		SPSZ	0.0051	0.3157			-0.3238	0.1720			0.2564	0.0817		
	0	Proposed	0.0023	0.3254	0.3066	0.9370	0.0011	0.1504	0.1527	0.9530	0.0013	0.0599	0.0606	0.9430
		SPSZ	0.0039	0.3260			0.0283	0.1565			-0.0203	0.0524		
	1	Proposed	-0.0016	0.3213	0.3095	0.9390	0.0046	0.1555	0.1541	0.9400	-0.0017	0.0626	0.0631	0.9410
		SPSZ	-0.0033	0.3328			0.3877	0.1828			-0.2999	0.0890		
	1.5	Proposed	0.0095	0.3312	0.3073	0.9290	-0.0093	0.1491	0.1522	0.9590	-0.0033	0.0600	0.0615	0.9540
		SPSZ	0.0076	0.3512			0.5575	0.1945			-0.4369	0.1133		
100	-1.5	Proposed	0.0060	0.2298	0.2182	0.9350	0.0025	0.1049	0.1052	0.9500	0.0007	0.0417	0.0410	0.9420
		SPSZ	0.0079	0.2334			-0.5040	0.1304			0.3855	0.0737		
	-1	Proposed	0.0059	0.2279	0.2184	0.9420	0.0010	0.1010	0.1053	0.9590	-0.0005	0.0420	0.0410	0.9430
		SPSZ	0.0072	0.2284			-0.3273	0.1164			0.2479	0.0527		
	0	Proposed	0.0031	0.2277	0.2181	0.9390	0.0039	0.1012	0.1053	0.9580	-0.0004	0.0410	0.0409	0.9500
		SPSZ	0.0031	0.2282			0.0311	0.1050			-0.0214	0.0367		
	1	Proposed	0.0100	0.2205	0.2182	0.9500	0.0032	0.1064	0.1056	0.9540	0.0003	0.0417	0.0411	0.9520
		SPSZ	0.0109	0.2302			0.3857	0.1244			-0.2910	0.0621		
	1.5	Proposed	0.0179	0.2270	0.2197	0.9390	-0.0015	0.1057	0.1047	0.9560	0.0006	0.0402	0.0412	0.9600
		SPSZ	0.0131	0.2402			0.5583	0.1405			0.0748	0.0827		

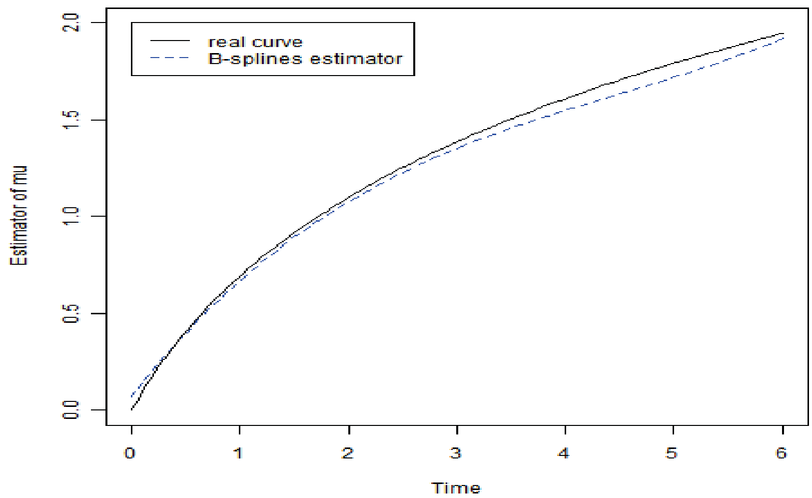
Note: SPSZ: Method in Sun et al. (2005).

2.6 Application

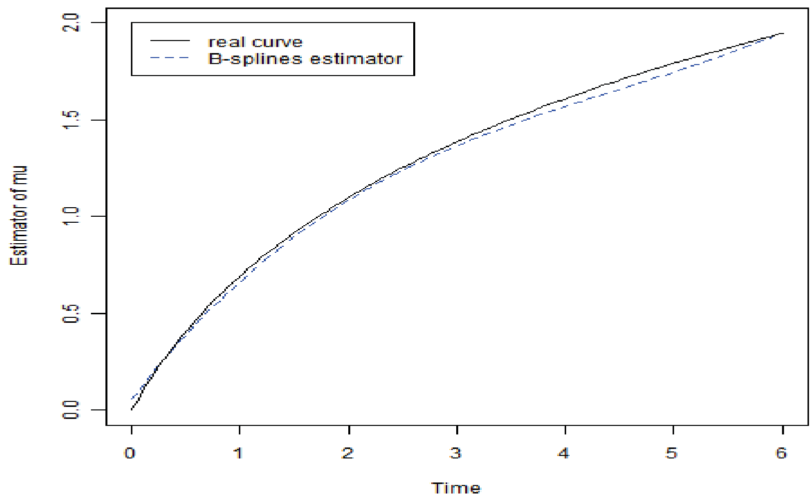
This section presents an analysis of the bladder cancer data by applying our proposed methods. There were 116 subjects with superficial bladder tumors randomized into one of three treatment groups: placebo, thiotepa, and pyridoxine. In the following, we restrict our attention to the placebo and thiotepa groups with respective sizes of 47 and 38 as it has been shown that the pyridoxine treatment had no effect on the recurrence of the bladder tumors (Zhang, 2002). For each patient, the observed information includes times when he or she made clinical visits and the numbers of recurrent tumors between clinical visits. Two baseline covariates were observed and they are the number of initial tumors and the size of the largest initial tumor.

To analyze the data, for patient i, define x_{1i} to be equal to 1 if the ith patient was given the thiotepa treatment and 0 otherwise, x_{2i} the number of initial tumors and x_{3i} the size of the largest initial tumor, $i = 1, \ldots, 85$. We define the response $Y_i(t)$ to be the natural logarithm of the cumulated new tumor numbers of patient i up to time t plus 1 to avoid 0. Let $H_i(\cdot)$ represent the accumulated observation numbers of patient i over the study period. Assume that $\{Y_i(t)\}$ can be described by model (2.3) with $h(\mathcal{F}_{it}, W_i) = H_i(t-)X_{1i}$, meaning that the relation between recurrence rate of bladder tumors and the observation times may vary with different treatments, i.e.,

$$E\{Y_i(t)|X_{1i}, X_{2i}, X_{3i}, \mathcal{F}_{it}\} = \mu_0(t) + \beta_1' X_{1i} + \beta_2' X_{2i} + \beta_3' X_{3i} + \alpha' H_i(t-)X_{1i}.$$



(a) Sample size $n = 50$



(b) Sample size $n = 100$

Figure 2.1. Estimates of $\mu_0(t) = \log(t+1)$ for simulated longitudinal data with non-Poisson observation process, $h(\mathcal{F}_t, W) = H(t-)X_1$ and $\alpha = 1$

Here, we took the last visit time of patient i as C_i in the analysis. For estimation of $\mu_0(t)$, we use the cubic B-spline approximation by taking the number of interior knots m_n as n^v with $v = 0.1$.

The application of the estimation procedure proposed in the previous sections gave $\hat{\beta}_1 = -0.3445, \hat{\beta}_2 = 0.1730, \hat{\beta}_3 = -0.0325$, and $\hat{\alpha} = -0.0288$ with the bootstrap standard errors being 0.1369, 0.0450, 0.0470, and 0.0109, which correspond to p-values of 0.0118, 0.0001, 0.4888, and 0.0079, respectively, based on the asymptotic results of the estimators. Here $\hat{\beta}_1, \hat{\beta}_2$ and $\hat{\beta}_3$ represent the estimated regression coefficients corresponding to the treatment indicator, the number of initial tumors, and the size of the largest initial tumor, respectively, while $\hat{\alpha}$ represents the estimated effect of the interaction between the observation process and the treatment indicator on the tumor recurrence rate. These results indicate that the response process and the interaction between the observation process and the treatment indicator are significantly negatively correlated. Just as explained in Sun et al. (2005), there are two reasons for this finding. One is that the more often the patient visited the clinic, had tumors removed and received treatment, the lower the tumor recurrence rate; another one is that more visits means less time for tumor growth. Furthermore, the thiotepa treatment significantly reduces the occurrence rate of the bladder tumors, and the number of initial tumors has a significant positive effect on the tumor recurrence rate. However, the occurrence rate of the bladder tumors do not seem to be significantly related to the size of the largest initial tumor. These conclusions are consistent with those presented in Sun et al. (2005), Sun et al. (2007) and Liang et al. (2009). Compared to the models in

Sun et al. (2005), Sun et al. (2007) and Liang et al. (2009), our fitted model may provide more information about the correlation between the tumor recurrence rate and observation times over treatment groups and also could be useful to estimate the future recurrence rate based on the observation history.

Chapter 3

Panel Count Data Analysis Using Monotone B-Splines

Semiparametric analysis of panel count data with informative observation times using monotone B-splines is presented in this chapter.

3.1 Introduction

As discussed with the bladder cancer data in Section 1.3.1, the underlying recurrent process and the observation process of panel count data may be dependent even given covariates, and the relation between these two processes may be influenced by some covariates in the study. Also the commonly used Poisson assumption about the observation process may not be true. In this chapter, by generalizing the conditional mean model (2.1) of the response process in Sun et al. (2005) to the underlying recurrent event process, we will develop a new flexible class of semiparametric regression models by incorporating the interaction between the observation history and some covariates to the mean model of the recurrent event process, while leaving the patterns of the observation times to be arbitrary. This weak distributional assumption can provide robustness to model misspecification. For nonparametric estimation of the baseline unknown function, a B-spline approximation will be used following Lu et al. (2007, 2009).

The remainder of this chapter is organized as follows. In Section 3.2, some notation and models for panel count data are presented. For estimation of regression parameters and the unknown baseline function, a spline-based least square method is proposed in Section 3.3. Then the asymptotic properties of the proposed estimators, including the consistency, rate of convergence, and asymptotic normality, are established in Section 3.4. Some simulation results are given in Section 3.5 in order to assess the finite-sample performance of the proposed inference procedure. Finally, by the analysis of a data set from the bladder tumor study, proposed approaches are illustrated in Section 3.6.

3.2 Statistical Models

Consider a study involving n subjects who may experience some recurrent events and suppose that each subject in the study gives rise to a counting process $N_i(t)$, denoting the total number of occurrences of the event of interest up to time t, $0 \leq t \leq \tau$, where τ is a known constant time point, $i = 1, \cdots, n$. Also suppose that for each subject i, $N_i(t)$ is observed only at discrete time points $0 < T_{K_i,1} < T_{K_i,2} < \cdots < T_{K_i,K_i}$, where the total number of observations K_i is an integer-valued random variable. In general, not every subject can be followed until τ and there exists a follow-up time C_i for subject i .That is, $N_i(T_{K_i,j})$ is observed only if $T_{K_i,j} \leq C_i \leq \tau$. Define $\tilde{H}_i(t) = H_i(t \wedge C_i)$,where $H_i(t) = \sum_{j=1}^{K_i} I(T_{K_i,j} \leq t)$ and $I(\cdot)$ is the indicator function, $i = 1, \cdots, n$ and $a \wedge b = \min(a, b)$. That is, $\tilde{H}_i(t)$ is a counting process characterizing the ith subject's observation process and having jumps only at the observation times. Let $\mathbf{X}_i = (X_{i1}, \cdots, X_{ip})'$ denote a p-dimensional vector of covariates that may not depend on $t, i = 1, \cdots, n$. Define $\mathcal{F}_{it} = \{H_i(s) : 0 \leq s < t\}$ as the observation history just before t.

According to the analysis of the panel count data arising from the bladder tumor data in Section 1.2.2, the relation between the observation and the recurrent event processes may be influenced by some covariates in the study. Thus, by relaxing the distributional assumption on the observation process, we assume that given $\mathbf{X}_i$, $\mathcal{F}_{it}$ and the covariate W_i, which may be a component of the vector $\mathbf{X}_i$ or may be other variables different from $\mathbf{X}_i$, the mean function of $N_i(t)$ has the form

$$\mu_i(t) = \exp\{\mu_0(t) + \beta'\mathbf{X}_i + \alpha' h(\mathcal{F}_{it}, W_i)\}, \tag{3.1}$$

where $\mu_0(t)$ is an unspecified smooth, nondecreasing function of t, β is a p-dimensional vector of unknown regression coefficients, and $h(\cdot)$ is a q-dimensional vector of known functions of the counting process H_i up to time $t-$ and the covariates W_i, representing the interaction between the observation history and some covariates, and α is a q-dimensional vector of unknown regression coefficients. Especially, when in some clinical studies with many different treatments, W_i are defined as the group indicators, then $h(\cdot)$ represents the different group effects on the observation times. Here, the right hand side of (3.1) as a whole function of t should be nondecreasing since $N(t)$ is a counting process. The main interest of this paper is to estimate the smooth function $\mu_0(t)$ and the regression parameters β and α. In fact, we can see that if we take $\mu_0(t) = \log \tilde{\mu}_0(t)$, where $\tilde{\mu}_0(t)$ is also an unspecified smooth, increasing, positive function of t, then (3.1) becomes

$$\mu_i(t) = \tilde{\mu}_0(t)\exp\{\beta'\mathbf{X}_i + \alpha' h(\mathcal{F}_{it}, W_i)\},$$

which is the standard form of the proportional means model for panel count data with informative observation times.

Model (3.1) specifies that the process $N_i(t)$ depends on the process $H_i(t)$ through the function h, which can be chosen according to situations. Following the discussion in Sun et al. (2005), a natural and simple choice for h may be $h(\mathcal{F}_{it}, W_i) = H_i(t-)W_i$, which means that $N_i(t)$ and $\mathcal{F}_{it}$ are related through or all information about $N_i(t)$ in $\mathcal{F}_{it}$ is given by the total number of observations. An alternative is that $N_i(t)$ depends on $\mathcal{F}_{it}$ only through a recent number of observations, say, in u time units, and this corresponds to $h(\mathcal{F}_{it}, W_i) = (H_i(t-) - H_i(t-u))W_i$. One could define h as a vector given by the forgoing two choices if both the total and recent numbers of observations may contain information about $N_i(t)$. If $\alpha = 0$, then model (3.1) reduces to the model considered by Sun and Wei (2000), Zhang (2002), and Wellner and Zhang (2007) for regression analysis of panel count data.

In addition, we assume that

$$E\{N_i(t)|X_i, H_i(s), 0 \le s \le t, C_i\} = E\{N_i(t)|X_i, \mathcal{F}_{it}, C_i\}, \tag{3.2}$$

which means that conditional on the covariates $\mathbf{X}_i'$s and C_i's, the number of events at time point t is only related to the observation history before t. The observation for each individual consists of $\mathbf{O} = (K, \bar{T}_K, \bar{N}_K, \bar{H}_K,\ \mathbf{X}, C)$, with $\bar{T}_K =$

$(T_{K,1}, \cdots, T_{K,K})$, $\bar{N}_K = (N(T_{K,1}), \cdots, N(T_{K,K}))$, $\bar{H}_K = (H(T_{K,1}), \cdots, H(T_{K,K}))$. Throughout this paper, we will assume that we observe n i.i.d. copies, $\mathbf{O}_1, \cdots, \mathbf{O}_n$ of $\mathbf{O}$.

3.3 Estimation Procedure

Denote

$$L_n(\beta, \alpha, \mu) = \sum_{i=1}^{n} \int_0^{\tau} \left[N_i(t) - \exp\{\mu(t) + \beta' \mathbf{X}_i + \alpha' h(\mathcal{F}_{it}, W_i)\}\right]^2 d\tilde{H}(t) \qquad (3.3)$$

In this paper, we propose to use B-splines to approximate $\mu(t)$. For a finite closed interval $[0, \tau]$, let $\mathcal{I} = \{t_i\}_1^{m_n+2l}$, with

$$0 = t_1 = \cdots = t_l < t_{l+1} < \cdots < t_{m_n+l} < t_{m_n+l+1} = \cdots = t_{m_n+2l} = \tau$$

be a sequence of knots that partition $[0, \tau]$ into $m_n + 1$ subintervals and $m_n = O(n^{\nu})$, for $0 < \nu < 1/2$. Let $\{B_{il}, 1 \leq i \leq q_n\}$ denote the B-spline basis functions with $q_n = m_n + l$. Let $\Psi_{l,\mathcal{I}}$ (with order l and knots $\mathcal{I}$) be the class linearly spanned by the B-splines functions, that is,

$$\Psi_{l,\mathcal{I}} = \{\sum_{i=1}^{q_n} \eta_i B_{il} : \eta_i \in \mathbb{R}, i = 1, \cdots, q_n\}.$$

We now define a subclass of $\Psi_{l,\mathcal{I}}$, as $\phi_{l,\mathcal{I}} = \{\sum_{i=1}^{q_n} \eta_i B_{il} : \eta_1 \leq \cdots \leq \eta_{q_n}\}$. According to the variation-diminishing properties in Schumaker (1981) which has been sketched in Appendix A, $\phi_{l,\mathcal{I}}$ is a class of monotone nondecreasing splines on $[0, \tau]$ since the monotonicity of the B-splines is guaranteed by the nondecreasing order of coefficients. For estimation of μ_0 specified in (3.1), we approximate the space of μ_0 by a subspace of $\phi_{l,\mathcal{I}}$, defined as

$$\psi_{l,\mathcal{I}} = \left\{ \sum_{i=1}^{q_n} \eta_i B_{il} : \eta_1 \leq \cdots \leq \eta_{q_n} \text{ and } \sum_{i=1}^{q_n} \eta_i^2 \leq \delta^2 \text{ for some constant } \delta \right\}.$$

Then, we approximate the smooth monotone function $\mu_0(t)$ by $\sum_{i=1}^{q_n} \eta_i B_{il}(t)$ and estimate the coefficients $\eta_1 \leq \cdots \leq \eta_{q_n}$ and regression parameters β, and α jointly through minimizing the approximated expression $L_n(\beta, \alpha, \mu)$ subject to nondecreasing constraints.

Since $\mu(t)$ can be approximated by $\sum_{i=1}^{q_n} \eta_i B_{il}(t)$, equation (3.3) becomes

$$L_n(\beta, \alpha, \eta) = \sum_{i=1}^{n} \int_0^{\tau} [N_i(t) - \exp\{\eta' B_l(t) + \beta' \mathbf{X}_i + \alpha' h(\mathcal{F}_{it}, W_i)\}]^2 \, d\tilde{H}(t), \quad (3.4)$$

where $\eta = (\eta_1, \cdots, \eta_{q_n})'$, and $B_l(t) = (B_{1l}(t), \cdots, B_{q_n l}(t))'$.

Let $\hat{\beta}_n, \hat{\alpha}_n, \ \hat{\eta}_n$ be the values that minimize

$$\begin{aligned} &L_n(\beta, \alpha, \eta) \\ =&\sum_{i=1}^{n} \sum_{j=1}^{K_i} \Big[N_i(T_{K_i,j}) - \exp\{\eta' B_l(T_{K_i,j}) + \beta' \mathbf{X}_i + \alpha' h(\mathcal{F}_{iT_{K_i,j}}, W_i)\} \Big]^2 \xi_i(T_{K_i,j}), \end{aligned} \quad (3.5)$$

under constraints $\eta_1 \leq \cdots \leq \eta_{q_n}$, where $\xi_i(t) = I(C_i \geq t)$. Then the monotone splines estimator for $\mu_0(t)$ is $\hat{\mu}_n(t) = \sum_{i=1}^{q_n} \hat{\eta}_{ni} B_{il}(t)$.

The estimation problem is equivalent to a nonlinear programming problem subject to linear inequality constraints. Specifically, the spline estimation problem can be formulated as the linear inequality constrained minimization problem

$$\min_{\theta \in \mathbb{R}^{p+q} \times \Theta_\eta} L_n(\theta), \quad (3.6)$$

where $\theta = (\beta', \alpha', \eta')'$ with $\eta \in \Theta_\eta = \{\eta : \eta_1 \leq \cdots \leq \eta_{q_n}\}$. Jamshidian (2004) proposed a generalized gradient projection algorithm (GP) for optimizing a nonlinear objective function with linear inequality constraints, based on the generalized Euclidean metric $\|x\| = x^T W x$ with W being a positive definite matrix and possibly varying from iteration to iteration. Zhang and Jamshidian (2004) applied the GP algorithm to large-scale nonparametric maximum likelihood estimation problems by choosing $W = D_H$, the matrix containing only the diagonal elements of the negative Hessian matrix H, in order to avoid the storage problem in updating H. However, this will increase the number of iterations and thereby the computing time. Lu et al. (2007) and Lu et al. (2009) used the GP algorithm utilized in Zhang and Jamshidian (2004) with $W = H$ directly because the dimension of unknown parameter space is usually small in their applications due to the use of polynomial splines, which would also substantially reduce the number of iterations. Here we consider the same monotone polynomial splines estimation as that in Lu et al. (2007, 2009), expect that we are solving a constrained minimizing problem and W is not equal to the negative Hessian matrix H here.

Let $\nabla L_n(\theta)$ be the negative gradient of $L_n(\theta)$ with respect to θ and

$$W = \sum_{i=1}^{n} \int_0^\tau \exp\{\theta' Z_{li}(t)\} Z_{li}^{\otimes 2}(t) d\tilde{H}(t),$$

which is a positive definite matrix with $Z_{li}(t) = (\mathbf{X}_i', h(\mathcal{F}_{it}, W_i)', B_l'(t))'$. Let $\mathcal{A} = \{i_1, i_2, \cdots, i_m\}$ denote the index set of active constraints, i.e. $\eta_{i_j} = \eta_{i_j+1}$, for $j = 1, \cdots, m$, during the numerical computation. $\mathcal{A}$ is allowed to be empty when $m = 0$. We define a m by $q_n + p + q$ working matrix corresponding to this set, given as follows:

$$A = \begin{pmatrix} 0 & \cdots & 0 & 0 & \cdots & -1 & 1 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & -1 & 1 \end{pmatrix},$$

i.e., the jth row ($j = 1, \cdots, m$) consists of the unit vector with its $(p + q + i_j)$th and $(p + q + i_j + 1)$th elements equal to -1 and 1 respectively and the remaining components zero. The generalized gradient projection algorithm is implemented in the following steps.

The generalized gradient projection algorithm

Start with a feasible initial value $\theta \in \mathbb{R}^{p+q} \times \Theta_\eta$, and cycle through the following steps until convergence.

S0: (**Computing the feasible search direction**)

$$\underline{d} = \left(I - W^{-1}A^T(AW^{-1}A^T)^{-1}A\right) W^{-1}\nabla L_n(\theta),$$

when there is no active constraint, take $\underline{d} = W^{-1}\nabla L_n(\theta)$.

S1: (**Forcing the updated θ fulfill the constraints**) If the resulted direction $\underline{d}$ is not nondecreasing in its components, compute

$$\gamma = \min_{i \notin \mathcal{A}, d_i > d_{i+1}} \left(-\frac{\eta_{i+1} - \eta_i}{d_{i+1} - d_i}\right).$$

Doing so guarantees that $\eta_{i+1} + \gamma d_{i+1} \geq \eta_i + \gamma d_i$, for $i = 1, \cdots, q_n$.

S2: (**Step-Halving line search**) Find a smallest integer k starting from 0 such that

$$L_n(\theta + (1/2)^k \underline{d}) < L_n(\theta).$$

S3: (**Updating the Solution**) If $\gamma > (1/2)^k$, replace θ by $\tilde{\theta} = \theta + (1/2)^k \underline{d}$ and check the stopping criterion (S5).

S4: (**Updating the active constraint set**) If $\gamma \le (1/2)^k$, in addition to replace θ by $\tilde{\theta} = \theta + \gamma \underline{d}$, modify $\mathcal{A}$ by adding indexes of all the newly active constraints to $\mathcal{A}$ and accordingly modify the working matrix A.

S5: (**Checking the stopping criterion**) If $\|\underline{d}\| \ge \varepsilon$ for a small $\varepsilon > 0$, go to S0. Otherwise, compute the Lagrange multipliers $\lambda = (AW^{-1}A^T)^{-1} AW^{-1}\dot{L}(\theta)$.

(i). If $\lambda_i \le 0$ for all $i \in \mathcal{A}$, set $\hat{\theta} = \theta$ and stop.

(ii). If at least one $\lambda_i > 0$, for $i \in \mathcal{A}$, remove the index corresponding to the largest λ_i from $\mathcal{A}$, and update $\mathcal{A}$ and go to S0.

To initialize the algorithm, we choose $\eta = (1, 2, \cdots, q_n)'$, β and α were all generated from the uniform distribution over interval $(-0.5, 0.5)$.

3.4 Asymptotic Theory

To establish the asymptotic properties of the estimators, we need the following conditions.

C1 The maximum spacing of the knots satisfies $\triangle = \max_{l+1<i<m_n+l+1} \mid t_i - t_{i-1} \mid = O(n^{-\nu})$.

C2 The parameter spaces of $(\beta', \alpha')'$, $\mathcal{R}$ is bounded and convex on $\mathbb{R}^{p+q}$, and the true parameter $(\beta_0, \alpha_0, \mu_0) \in \mathcal{R}^\circ \times \mathcal{F}_r$, where $\mathcal{R}^\circ$ is the interior of $\mathcal{R}$, and

$$\mathcal{F}_r \equiv \{\mu : [0, \infty) \longrightarrow \mathbb{R} \Big| \mu \text{ is monotone and } |\mu^{(k)}(s) - \mu^{(k)}(t)| \leq M|s-t|^\varsigma\},$$

where k is a nonnegative integer, $\varsigma \in (0, 1]$ such that $r = k + \varsigma > 0.5$, M is a positive constant and $f^{(k)}$ is the kth derivative of function f.

C3 $N_i(\tau)(i = 1, \cdots, n)$ are bounded by a constant, and there exist a positive integer M_1, such that $P(\|\mathbf{X}\| \leq M_1) = 1$, that is, the covariate vector is uniformly bounded.

C4 There exists a positive integer M_2 such that $P(K \leq M_2) = 1$, that is, the number of the observation is finite.

C5 If with probability 1, $\mathbf{h}_1'\mathbf{X} + \mathbf{h}_2' h(\mathcal{F}_t, W) + h_3(t) = 0$ for $\mathbf{h}_1 \in \mathbb{R}^p$, $\mathbf{h}_2 \in \mathbb{R}^q$ and some deterministic function h_3, then $\mathbf{h}_1 = 0$, $\mathbf{h}_2 = 0$, $h_3 = 0$.

Next, we introduce more notations. Let $\mathcal{B}_p$ and $\mathcal{B}$ denote the collection of Borel sets in $\mathbb{R}^p$ and $\mathbb{R}$, respectively, and let $\mathcal{B}_{[0,\tau]} = \{B \cap [0, \tau] : B \in \mathcal{B}\}$. We define measures υ on $(\mathbb{R}^p \times [0, \tau], \mathcal{B}_p \times \mathcal{B}_{[0,\tau]})$ and υ_1 on $([0, \tau], \mathcal{B}_{[0,\tau]})$, as follows:

$$\begin{aligned}
\upsilon(A \times B) = & \int\limits_{A\times(0,\infty)} \sum_{k=1}^{\infty} P(K = k|\mathbf{X} = \mathbf{x}, C = c)) \\
& \sum_{j=1}^{k} P(T_{k,j} \in B \cap [0, c] \big| K = k, \mathbf{X} = \mathbf{x}, C = c) dF(\mathbf{x}, c) \\
= & \int_{A\times(0,\infty)} E\left\{\sum_{j=1}^{K} I_{B\cap[0,c]}(T_{K,j})|\mathbf{X} = \mathbf{x}, C = c\right\} dF(\mathbf{x}, c),
\end{aligned}$$

and $\upsilon_1(B) = \upsilon(\mathbb{R}^p \times B)$, for $B \in \mathcal{B}_{[0,\tau]}$, and $A \in \mathcal{B}_p$ set, where F is the joint distribution function of $\mathbf{X}$ and C. Then υ_1 and υ are finite measures under condition C4. Let $L_2(\upsilon_1) \equiv \{f : [0,\infty] \longrightarrow \mathbb{R} \big| ||f||_{L_2(\nu_1)} \equiv \left[\int |f(t)|^2 d\nu_1(t)\right]^{1/2} < \infty\}$. Clearly,

$$\|f\|_{L_2(\upsilon_1)} = \left[E\sum_{j=1}^{K} |f(T_{K,j})|^2 \xi(T_{K,j})\right]^{1/2} = \left[E\int_0^\tau |f(t)|^2 \xi(t) dH(t)\right]^{1/2}.$$

Let $Z = \{Z(t,W) \equiv h(\mathcal{F}_t, W), 0 \le t \le \tau\}$ represent a q-dimensional bounded random process indexed by t. Define

$$\mathcal{G} \equiv \{z(t,w) : [0,\tau] \times [-M_1, M_1] \longrightarrow \mathcal{M}\},$$

where $\mathcal{M}$ is a bounded set on $\mathbb{R}^q$, and for function $f(\mathbf{x}, z, t) : [-M_1, M_1]^p \times \mathcal{G} \times [0,\tau] \longrightarrow \mathbb{R}$, define

$$\|f\|_2 \equiv \left[E\sum_{j=1}^{K} |f(\mathbf{X}, Z(T_{K,j}, W), T_{K,j})|^2 \xi(T_{K,j})\right]^{1/2}.$$

Define $M_n(g) = n^{-1}L_n(\beta, \alpha, \mu) = \mathbb{P}_n m_g(\mathbf{O})$, where $g(\mathbf{x}, z, t) = \exp\{\beta'\mathbf{x} + \alpha' z(t,w) + \mu(t)\}$, $m_g(\mathbf{O}) = \sum_{j=1}^{K}[N(T_{K,j}) - g(\mathbf{X}, Z(T_{K,j}, W), T_{K,j})]^2 \xi(T_{K,j})$, and $M(g) = Pm_g(\mathbf{O})$, where Pf and $\mathbb{P}_n f$ represent $\int f dP$ and $n^{-1}\sum_{i=1}^{n} f(\mathbf{O}_i)$, respectively.

Since $L_2(\upsilon_1)$ is a Hilbert space, and $\mathcal{F}_r \subset L_2(\upsilon_1)$, by the Hilbert Projection Theorem (Stakgold, 1998, P. 288), for $x_j \in L_2(\upsilon_1)$, there is a unique $a_j^* \in \mathcal{F}_r$, s.t. $(x_j - a_j^*) \perp \mathcal{F}_r$, for $j = 1, \cdots, p$. Let $z_l(t,w)$ be the lth component of $h(\mathcal{F}_t, w), l = 1, \cdots, q$. Then for $z_l(t,w) \in L_2(\upsilon_1)$, there is a unique $b_l^*(t) \in \mathcal{F}_r$, s.t. $(z_l - b_l^*) \perp \mathcal{F}_r$, for $l = 1, \cdots, q$. Let $\mathbf{a}^* = (a_1^*, \cdots, a_p^*)'$ and $\mathbf{b}^* = (b_1^*, \cdots, b_q^*)'$.

Then we need another condition

C6 $E\left[\int_0^\tau \begin{pmatrix} \mathbf{X}-\mathbf{a}^* \\ h(\mathcal{F}_t, W)-\mathbf{b}^*(t) \end{pmatrix}^{\otimes 2} d\tilde{H}(t)\right]$ is nonsingular.

In practice, C1 is similar to those required by Stone (1986) and Zhou et al. (1998). C2 is a common assumption in nonparametric smoothing estimation problems. C3 and C4 are mild and easily justified in many applications. C5 is need to establish the identifiability of the model. C6 is a technical condition. The asymptotic properties of the estimators are summarized as follows.

Theorem 3.1 (**Consisitency**). *Under conditions C1 - C4 and C6 ,* $\|\hat{\beta}_n-\beta_0\| \to 0$, $\|\hat{\alpha}_n-\alpha_0\| \to 0$, $\|\hat{\mu}_n-\mu_0\|_{L_2(v_1)} \to 0$, *almost surely.*

Proof of Theorem 3.1.

According to Lemma 5 in Stone (1985) we have sketched in Appendix A, for $\mu_0 \in \mathcal{F}_r$, there exist a $\mu_n \in \psi_{l,\mathcal{I}}$ with order $l \geq k+1$ and knots $\mathcal{I}$ such that $\|\mu_n-\mu_0\|_\infty = O(n^{-vr})$. Let $g_n(\mathbf{x}, z, t) = \exp\{\beta_0'\mathbf{x} + \alpha_0' z(t, w) + \mu_n(t)\}$, $\hat{g}_n(\mathbf{x}, z, t) = \exp\{\hat{\beta}_n'\mathbf{x} + \hat{\alpha}_n' z(t, w) + \hat{\mu}_n(t)\}$, and $g_0(\mathbf{x}, z, t) = \exp\{\beta_0'\mathbf{x} + \alpha_0' z(t, w) + \mu_0(t)\}$. Without loss of generality, we assume that $\mu_n > \mu_0$, thus $g_n > g_0$, and $\|g_n-g_0\|_\infty = O(n^{-\nu r})$. Choose a $\phi_n \in \psi_{l,\mathcal{I}}$ and b_1 and b_2, s.t $h_n \equiv \exp\{{b_1}'\mathbf{x} + {b_2}'z + \phi_n\}$, and $\|h_n\|_2^2 = O(n^{-\nu r} + n^{-\frac{1-\nu}{2}})$. Then for any $\lambda > 0$, $\|g_n - g_0 + \lambda h_n\|_2^2 = O(n^{-\nu r} + n^{-\frac{1-\nu}{2}})$. Let

$$
\begin{aligned}
&J_n(\lambda) \equiv M_n(g_n + \lambda h_n) \\
=&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i}[N_i(T_{K_i,j}) - (g_n + \lambda h_n)(\mathbf{X}_i, Z_i(T_{K_i,j}, W_i), T_{K_i,j})]^2 \xi_i(T_{K_i,j}),
\end{aligned}
$$

then

$$J_n'(\lambda) =\frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i}[(g_n+\lambda h_n)(\mathbf{X}_i, Z_i(T_{K_i,j}, W_i), T_{K_i,j}) - N_i(T_{K_i,j})]$$

$$\times h_n(\mathbf{X}_i, Z_i(T_{K_i,j}, W_i), T_{K_i,j})\xi_i(T_{K_i,j}),$$

and

$$J''_n(\lambda) = \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i} h_n^2(\mathbf{X}_i, Z_i(T_{K_i,j}, W_i), T_{K_i,j})\xi_i(T_{K_i,j}) \geq 0.$$

Thus, $J_n'(\lambda)$ is a nondecreasing function. Therefore, to prove the convergence of $\hat{g}_n$ to g_0, it is sufficient to show that $\forall\lambda_0 > 0, J_n'(\lambda_0) > 0$ and $J_n'(-\lambda_0) < 0$ except on an event with probability converging to zero. Then $\hat{g}_n$ must be between $g_n - \lambda_0 h_n$ and $g_n + \lambda_0 h_n$, so $\|\hat{g}_n - g_n\|_2^2 \leq \lambda_0^2\|h_n\|_2^2 = O(n^{-\nu r} + n^{-\frac{1-\nu}{2}})$.

Next, we'll show that $J_n'(\lambda_0) > 0$.

$$\begin{aligned}
&\frac{1}{2}J_n'(\lambda_0)\\
=&\mathbb{G}_n\sum_{j=1}^{K}[(g_n+\lambda_0 h_n)(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) - N(T_{K,j})]h_n(\mathbf{X}, Z(T_{K,j}, W), T_{K,j})\xi(T_{K,j})\\
+&P\sum_{j=1}^{K}[(g_n+\lambda_0 h_n)(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) - N(T_{K,j})]h_n(\mathbf{X}, Z(T_{K,j}, W), T_{K,j})\xi(T_{K,j})\\
\equiv&I_{1n} + I_{2n}.
\end{aligned}$$

By the calculation of Shen and Wong (1994, p. 597), for $\eta > 0$ and any $\varepsilon \leq \eta$,

$$\log N_{[]}(\varepsilon, \psi_{l,\mathcal{I}}, L_2(\upsilon_1)) \leq c_1 q_n \log(\eta/\varepsilon), \qquad J_{[]}(\eta, \psi_{l,\mathcal{I}}, L_2(\upsilon_1)) \leq c_2 q_n^{1/2}\eta,$$

where $q_n = m_n + l$ is the number of spline basis functions, and c_1 and c_2 are finite constants.

By Theorem 2.5.2 of Van der Vaart and Wellner (1996, p. 127) (Theorem C.3), $\psi_{l,\mathcal{I}}$ is a Donsker class. Then given g_n defined before,

$$\mathcal{G}_\eta \equiv \Big\{ \sum_{j=1}^{K} [h(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) - N(T_{K,j})](h - g_n)(\mathbf{X}, Z(T_{K,j}, W), T_{K,j})\xi(T_{K,j}) :$$

$$h(\mathbf{x}, z, t) = \exp\{\beta'\mathbf{x} + \alpha' z(t, w) + \phi(t)\}, \phi \in \psi_{l,\mathcal{I}}, \|h - g_n\|_2 \le \eta \Big\}$$

is a Donsker class. Thus, $I_{1n} = O_P(n^{-1/2})$.

$$\begin{aligned}
I_{2n} &= E\Big[\int_0^\tau (g_n + \lambda_0 h_n)(\mathbf{X}, Z(t, W), t) h_n(\mathbf{X}, Z(t, W), t)\xi(t) dH(t)\Big] \\
&\quad - E\Big[\int_0^\tau N(t) h_n(\mathbf{X}, Z(t, W), t)\xi(t) dH(t)\Big] \\
&= E\Big[\int_0^\tau (\lambda_0 h_n + g_n - g_0)(\mathbf{X}, Z(t, W), t) h_n(\mathbf{X}, Z(t, W), t)\xi(t) dH(t)\Big] \\
&\ge E\Big[\int_0^\tau \lambda_0 h_n^2(\mathbf{X}, Z(t, W), t)\xi(t) dH(t)\Big] = \lambda_0 \|h_n\|_2^2
\end{aligned}$$

Thus, $\frac{1}{2}J_n'(\lambda_0) > O_P(n^{-1/2}) + \lambda_0\|h_n\|_2^2 > 0$, since $\|h_n\|_2^2 = O(p_n^{-1})$ with $p_n^{-1} \equiv n^{-\nu r} + n^{-\frac{1-\nu}{2}} \ge n^{-\frac{r}{1+2r}} > n^{-1/2}$ for $0 < \nu < 1/2$.

For $J_n'(-\lambda_0)$,

$$\begin{aligned}
&\frac{1}{2}J_n'(-\lambda_0) \\
=&\mathbb{G}_n \sum_{j=1}^{K} [(g_n - \lambda_0 h_n)(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) - N(T_{K,j})] h_n(\mathbf{X}, Z(T_{K,j}, W), T_{K,j})\xi(T_{K,j}) \\
+&P \sum_{j=1}^{K} [(g_n - \lambda_0 h_n)(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) - N(T_{K,j})] h_n(\mathbf{X}, Z(T_{K,j}, W), T_{K,j})\xi(T_{K,j}) \\
\equiv& I_{1n}^* + I_{2n}^*.
\end{aligned}$$

Using the same arguments as for $J_n'(\lambda_0)$, $I_{1n}^* = O_P(n^{-1/2})$.

$$
\begin{aligned}
I_{2n}^* &= E\big[\int_0^\tau (g_n - \lambda_0 h_n)(\mathbf{X}, Z(t,W), t) h_n(\mathbf{X}, Z(t,W), t)\xi(t) dH(t)\big] \\
&\quad - E\big[\int_0^\tau h_n(\mathbf{X}, Z(t,W), t)\xi(t) N(t) dH(t)\big] \\
&= E\big[\int_0^\tau (-\lambda_0 h_n + g_n - g_0)(\mathbf{X}, Z(t,W), t) h_n(\mathbf{X}, Z(t,W), t)\xi(t) dH(t)\big] \\
&= -\lambda_0 E\big[\int_0^\tau h_n^2(\mathbf{X}, Z(t,W), t)\xi(t) dH(t)\big] \\
&\quad + E\big[\int_0^\tau (g_n - g_0)(\mathbf{X}, Z(t,W), t) h_n(\mathbf{X}, Z(t,W), t)\xi(t) dH(t)\big] \\
&\leq -\lambda_0 \|h_n\|_2^2 + \|g_n - g_0\|_2 \|h_n\|_2 \\
&\leq -c_3 p_n^{-1},
\end{aligned}
$$

for a positive constant c_3. Thus, $\frac{1}{2} J_n'(-\lambda_0) \leq O(n^{-1/2}) - c_3 p_n^{-1} < 0$.

Then we have $\|\hat{g}_n - g_0\|_2 \leq \|\hat{g}_n - g_n\|_2 + \|g_n - g_0\|_2 = O(p_n^{-1/2})$, and $\log \hat{g}_n - \log g_0 = \frac{1}{g^*}(\hat{g}_n - g_0)$, with $g^* = (1-\xi) g_0 + \xi \hat{g}_n, 0 \leq \xi \leq 1$. Hence $\|\log \hat{g}_n - \log g_0\|_2 = O(p_n^{-1/2}) \longrightarrow 0$. Also,

$$
\begin{aligned}
\|\log \hat{g}_n - \log g_0\|_2 =& \|(\hat{\beta}_n - \beta_0)'\mathbf{x} + (\hat{\alpha}_n - \alpha_0)' z + (\hat{\mu}_n - \mu_0)\|_2 \\
=& \|(\hat{\beta}_n - \beta_0)'(\mathbf{x} - \mathbf{a}^*) + (\hat{\alpha}_n - \alpha_0)'(z - \mathbf{b}^*) \\
& + (\hat{\beta}_n - \beta_0)'\mathbf{a}^* + (\hat{\alpha}_n - \alpha_0)'\mathbf{b}^* + (\hat{\mu}_n - \mu_0)\|_2 \\
=& \|(\hat{\beta}_n - \beta_0)'(\mathbf{x} - \mathbf{a}^*) + (\hat{\alpha}_n - \alpha_0)'(z - \mathbf{b}^*)\|_2 \\
& + \|(\hat{\beta}_n - \beta_0)'\mathbf{a}^* + (\hat{\alpha}_n - \alpha_0)'\mathbf{b}^* + (\hat{\mu}_n - \mu_0)\|_2.
\end{aligned}
$$

By C6, we can get $\|\hat{\beta}_n - \beta_0\| \longrightarrow 0$, and $\|\hat{\alpha}_n - \alpha_0\| \longrightarrow 0$ from the first term of the right hand side of the above equality, and thus it follows that $\|\hat{\mu}_n - \mu_0\|_{L_2(v_1)} \longrightarrow 0$. This completes the proof of the theorem.

Theorem 3.2 (**Rate of Convergence**). *Suppose that conditions C 1 - C 6 hold, then*

$$\|\hat{\beta}_n - \beta_0\| = O_P(n^{-\frac{1-\nu}{2}}), \|\hat{\alpha}_n - \alpha_0\| = O_P(n^{-\frac{1-\nu}{2}}), \|\hat{\mu}_n - \mu_0\|_{L_2(\upsilon_1)} = O_P(n^{-\frac{1-\nu}{2}}).$$

Remark 3.1. *When $\nu = 1/(1+2r)$, $n^{-\frac{1-\nu}{2}} = n^{-\frac{r}{1+2r}}$, we conclude from Stone (1980, 1982) that the rate of convergence of the estimator $\hat{\mu}_n$ is the optimal rate in nonparametric regression.*

Proof of Theorem 3.2.

For any $\eta > 0$, let

$$\mathcal{F}_\eta \equiv \{g = \exp\{\beta'\mathbf{x} + \alpha' z + \mu\} : \|\beta - \beta_0\| \leq \eta, \|\alpha - \alpha_0\| \leq \eta, \mu \in \psi_{l,\mathcal{I}}, \|\mu - \mu_0\|_{L_2(\upsilon_1)} \leq \eta\}.$$

Similar to Lemma A.2 in Huang (1999, p. 1557) given in Appendix A, for any $\varepsilon \leq \eta$, $\log N_{[]}(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2) \leq c_4 q_n \log(\eta/\varepsilon)$, for a constant c_4. Thus, for $\varepsilon > 0$, there exists a set of brackets $\{[g_i^l, g_i^r], i = 1, \cdots, (\frac{\eta}{\varepsilon})^{c_4 q_n}\}$ such that, for each $g \in \mathcal{F}_\eta$, there is a $[g_s^l, g_s^r]$ with $g_s^l(\mathbf{x}, z, t) \leq g(\mathbf{x}, z, t) \leq g_s^r(\mathbf{x}, z, t)$, for all $\mathbf{x}$, $t \in [0, \tau]$ and $z \in \mathcal{G}$, and $\|g_s^r - g_s^l\|_2^2 \leq \varepsilon^2$.

Then, by Theorem 3.1, $\hat{g}_n \in \mathcal{F}_\eta$, for any $\eta > 0$ and sufficiently large n.

Next, consider the class $\mathcal{M}_\eta \equiv \{m_g(\mathbf{O}) - m_{g_0}(\mathbf{O}) : g \in \mathcal{F}_\eta\}$, where $m_g(\mathbf{O}) = \sum_{j=1}^K [N(T_{K,j}) - g(\mathbf{X}, Z(T_{K,j}, W), T_{K,j})]^2 \xi(T_{K,j})$.

For $i = 1, \cdots, (\frac{\eta}{\varepsilon})^{c_4 q_n}$, define

$$m_i^l(\mathbf{O}) = \sum_{j=1}^K \Big[\{|g_i^l| \vee |g_i^r|\}^2(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) - 2N(T_{K,j}) g_i^r(\mathbf{X}, Z(T_{K,j}, W), T_{K,j})$$
$$+ 2N(T_{K,j}) g_0(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) - g_0^2(\mathbf{X}, Z(T_{K,j}, W), T_{K,j})\Big] \xi(T_{K,j}),$$

$$m_i^r(\mathbf{O}) = \sum_{j=1}^{K} \Big[\{|g_i^l| \wedge |g_i^r|\}^2(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) - 2N(T_{K,j}) g_i^l(\mathbf{X}, Z(T_{K,j}, W), T_{K,j})$$

$$+ 2N(T_{K,j}) g_0(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) - g_0^2(\mathbf{X}, Z(T_{K,j}, W), T_{K,j})\Big] \xi(T_{K,j}),$$

where $a \vee b = \min\{a, b\}$ and $a \wedge b = \max\{a, b\}$. Then, $m_i^l(\mathbf{O}) \leq m_i^r(\mathbf{O})$ and it is easy to show that $P|m_i^r(\mathbf{O}) - m_i^l(\mathbf{O})|^2 \leq c_5 \varepsilon^2$ with a positive constant c_5. Thus $\{[m_i^l(\mathbf{O}), m_i^r(\mathbf{O})], i = 1, \cdots, (\frac{\eta}{\varepsilon})^{c_4 q_n}\}$ is the set of brackets for $\mathcal{M}_\eta$, which implies that

$$\log N_{[]}(\varepsilon, \mathcal{M}_\eta, L_2(P) \leq c_4 q_n \log(\eta/\varepsilon).$$

Moreover, by some calculations, we can verify that $P\|m_g(\mathbf{O}) - m_{g_0}(\mathbf{O})\|^2 \leq c_6 \eta^2$ for any $g \in \mathcal{F}_\eta$ by C4. Therefore, by Lemma 3.4.2 of Van der Vaart and Wellner (1996) (Lemma C.5), we obtain

$$E\|n^{1/2}(\mathbb{P} - P)\|_{\mathcal{M}_\eta} \leq c_7 J_{[]}(\eta, \mathcal{M}_\eta, L_2(P)) \left\{1 + \frac{J_{[]}(\eta, \mathcal{M}_\eta, L_2(P))}{\eta^2 n^{1/2}} M_3\right\}, \quad (3.7)$$

where M_3 is a constant and $\|n^{1/2}(\mathbb{P} - P)\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |n^{1/2}(\mathbb{P} - P)f|$, and

$$\tilde{J}_{[]}(\eta, \mathcal{M}_\eta, L_2(P)) = \int_0^\eta \{1 + \log N_{[]}(\varepsilon, \mathcal{M}_\eta, L_2(P))\}^{1/2} d\varepsilon \leq c_8 q_n^{1/2} \eta.$$

The right hand side of (3.7) yields $\varphi_n(\eta) = c_9(q_n^{1/2}\eta + q_n/n^{1/2})$. It is easy to see that $\varphi_n(\eta)/\eta$ is decreasing in η, and

$$r_n^2 \varphi(\frac{1}{r_n}) = r_n q_n^{1/2} + r_n^2 q_n / n^{1/2} \leq 2n^{1/2},$$

for $r_n = n^{\frac{1-\nu}{2}}$ and $0 < \nu < 1/2$.

Note that

$$
\begin{aligned}
& P[m_g(\mathbf{O}) - m_{g_0}(\mathbf{O})] \\
= & P\Big[\int_0^\tau \left\{[N(t) - g(\mathbf{X}, Z(t,W), t)]^2 - [N(t) - g_0(\mathbf{X}, Z(t,W), t)]^2\right\} \xi(t) dH(t)\Big] \\
= & E\Big[\int_0^\tau (g_0 - g)(\mathbf{X}, Z(t,W), t)[2N(t) - (g + g_0)(\mathbf{X}, Z(t,W), t)]\xi(t)dH(t)\Big] \\
= & E\Big[\int_0^\tau (g - g_0)^2(\mathbf{X}, Z(t,W), t)\xi(t)dH(t)\Big] \\
= & \|g - g_0\|_2^2.
\end{aligned}
$$

Thus, by Theorem 3.2.5 of Van der Vaart and Wellner (1996) (Theorem C.4), $n^{\frac{1-\nu}{2}}\|\hat{g}_n - g_0\|_2 = O_p(1)$. Therefore by the similar arguments as those in the proof of consistency of $\hat{\beta}_n$, $\hat{\alpha}_n$ and $\hat{\mu}_n$, we can get the rate of convergence of $\hat{\mu}_n$, $\hat{\beta}_n$ and $\hat{\alpha}_n$ as stated in the Theorem. The choice of $\nu = 1/(1+2r)$ yields the rate of convergence of $r/(1+2r)$, which completes the proof.

Theorem 3.3 (**Asymptotic Normality**). *Suppose that conditions C 1 - C 6 hold. Let*

$$\mathcal{H} \equiv \{(\boldsymbol{h}_1,\ \boldsymbol{h}_2, h_3) : (\boldsymbol{h}_1', \boldsymbol{h}_2')' \in \mathcal{R}, h_3 \in \mathcal{F}_r, \|\boldsymbol{h}_1\| \leq 1, \|\boldsymbol{h}_2\| \leq 1, \|h_2\|_\infty \leq 1\}.$$

Then for any $(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3) \in \mathcal{H}$,

$$\boldsymbol{h}_1'\sqrt{n}(\hat{\beta}_n - \beta_0) + \boldsymbol{h}_2'\sqrt{n}(\hat{\alpha}_n - \alpha_0) + \int_0^\tau \sqrt{n}(\hat{\mu}_n - \mu_0)(t)dh_3(t)$$

converges in distribution $N(0, \sigma^2)$, *where* σ^2 *is given in (3.8).*

The similar bootstrap covariance matrix estimator for $\sqrt{n}\begin{pmatrix}\hat{\beta}_n - \beta_0 \\ \hat{\alpha}_n - \alpha_0\end{pmatrix}$ can be obtained as in Section 2.4.

Proof of Theorem 3.3.

We define a sequence of maps S_n mapping a neighborhood of $(\alpha_0, \beta_0, \mu_0)$, denoted by $\mathcal{U}$, in the parameter space for (β, α, μ) into $l^\infty(\mathcal{H})$ as :

$$
\begin{aligned}
& S_n(\beta, \alpha, \mu)[\mathbf{h}_1, \mathbf{h}_2, h_3] \\
\equiv \quad & n^{-1}\frac{d}{d\varepsilon}L_n(\beta + \varepsilon\mathbf{h}_1, \alpha + \varepsilon\mathbf{h}_2, \mu + \varepsilon h_3)\Big|_{\varepsilon=0} \\
= \quad & -\frac{2}{n}\sum_{i=1}^{n}\int_0^\tau [N_i(t) - e^{\{\beta'\mathbf{X}_i + \alpha' h(\mathcal{F}_{it}, W_i) + \mu(t)\}}]e^{\{\beta'\mathbf{X}_i + \alpha' h(\mathcal{F}_{it}, W_i) + \mu(t)\}} \\
& \quad \times[\mathbf{h}_1' \ \mathbf{X}_i + \mathbf{h}_2' h(\mathcal{F}_{it}, W_i) + h_3(t)]d\tilde{H}_i(t) \\
\equiv \quad & \mathbb{P}_n\psi(\beta, \alpha, \mu)[\mathbf{h}_1, \ \mathbf{h}_2, h_3],
\end{aligned}
$$

Correspondingly, we define the limit map $S : \mathcal{U} \longrightarrow l^\infty(\mathcal{H})$ as $S(\beta, \alpha, \mu)[\mathbf{h}_1, \ \mathbf{h}_2, h_3]$, where $l^\infty(\mathcal{H})$ is the space of bounded functionals on $\mathcal{H}$ under the supermum norm $\|f\| = \sup_{h\in\mathcal{H}} |f(h)|$.

To derive the asymptotic normality of the estimators $(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n)$, motivated by the proof of Theorem 3.3.1 of Van der Vaart and Wellner (1996, p. 310), we first need to verify the following five conditions.

(i) $\sqrt{n}(S_n - S)(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n) - \sqrt{n}(S_n - S)(\beta_0, \alpha_0, \mu_0) = o_p(1)$.

(ii) $\sqrt{n}(S_n - S)(\beta_0, \alpha_0, \mu_0)$ converges in distribution to a tight Gaussian process on $l^\infty(\mathcal{H})$.

(iii) $S(\beta_0, \alpha_0, \mu_0) = 0$ and $S_n(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n) = o_p(n^{-1/2})$.

(iv) $(\beta, \alpha, \mu) \longmapsto S(\beta, \alpha, \mu)$ is Fréchet-differentiable at $(\beta_0, \alpha_0, \mu_0)$ with a continuously invertible derivative $\dot{S}(\beta_0, \alpha_0, \mu_0)$;

(v) $\sqrt{n}\Big(S(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n) - S(\beta_0, \alpha_0, \mu_0)\Big) - \sqrt{n}\dot{S}(\beta_0, \alpha_0, \mu_0)\Big((\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n) - (\beta_0, \alpha_0, \mu_0)\Big)$ $= o_p(1)$.

Note that

$$
\begin{aligned}
&\sqrt{n}(S_n - S)(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n) - \sqrt{n}(S_n - S)(\beta_0, \alpha_0, \mu_0) \\
=&\sqrt{n}(\mathbb{P}_n - P)\Big(\psi(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n)[\mathbf{h}_1,\ \mathbf{h}_2, h_3] - \psi(\beta_0, \alpha_0, \mu_0)[\mathbf{h}_1,\ \mathbf{h}_2, h_3]\Big).
\end{aligned}
$$

Define

$$
\rho\big((\beta_1, \alpha_1, \mu_1) - (\beta_2, \alpha_2, \mu_2)\big) = \big\{\|\beta_1 - \beta_2\|^2 + \|\alpha_1 - \alpha_2\|^2 + \|\mu_1 - \mu_2\|^2_{L_2(\upsilon_1)}\big\}^{1/2}
$$

and for $\delta > 0$,

$$
\begin{aligned}
\mathcal{F}_\delta =&\Big\{\psi(\beta, \alpha, \mu)[\mathbf{h}_1,\ \mathbf{h}_2, h_3] - \psi(\beta_0, \alpha_0, \mu_0)[\mathbf{h}_1,\ \mathbf{h}_2, h_3] : \\
&\qquad \rho\big((\beta, \alpha, \mu) - (\beta_0, \alpha_0, \mu_0)\big) < \delta, (\mathbf{h}_1,\ \mathbf{h}_2, h_3) \in \mathcal{H}\Big\}.
\end{aligned}
$$

It is easy to see that $\mathcal{F}_r \subseteq C^r[0, \tau]$ is a Donsker class from Van der Vaart and Wellner (1996, p157), thus $\mathcal{H}$ is a Donsker class and

$$
\begin{aligned}
&\big|\psi(\beta, \alpha, \mu)[\mathbf{h}_1,\ \mathbf{h}_2, h_3]\big| \\
=&\big| - 2\int_0^\tau [N(t) - e^{\{\beta'\mathbf{X} + \alpha' h(\mathcal{F}_t, W) + \mu(t)\}}] e^{\{\beta'\mathbf{X} + \alpha' h(\mathcal{F}_t, W) + \mu(t)\}} \\
&\times [\mathbf{h}_1'\ \mathbf{X} + \mathbf{h}_2' h(\mathcal{F}_t, W) + h_3(t)] d\tilde{H}(t)\big| \\
\leq& M_1\|h_1\| + M_2\|h_2\| + M_3\|h_3\|_\infty,
\end{aligned}
$$

for constants M_1, M_2, M_3, which means that $\psi(\beta, \alpha, \mu)$ is a bounded Lipschitz functional with respect to $\mathcal{H}$, thus $\mathcal{F}_\delta$ is a Donsker class for some $\delta > 0$. And

$$
\begin{aligned}
&P\Big|\psi(\beta_1, \alpha_1, \mu_1)[\mathbf{h}_1,\ \mathbf{h}_2, h_3] - \psi(\beta_2, \alpha_2, \mu_2)[\mathbf{h}_1,\ \mathbf{h}_2, h_3]\Big|^2 \\
=&P\Big|2\int_0^\tau \Big[N(t)\big\{e^{\{\beta_1'\mathbf{X}+\alpha_1'h(\mathcal{F}_t,W)+\mu_1(t)\}} - e^{\{\beta_2'\mathbf{X}+\alpha_2'h(\mathcal{F}_t,W)+\mu_2(t)\}}\big\} \\
&\qquad + \big\{e^{\{2(\beta_1'\mathbf{X}+\alpha_1'h(\mathcal{F}_t,W)+\mu_1(t))\}} - e^{\{2(\beta_2'\mathbf{X}+\alpha_2'h(\mathcal{F}_t,W)+\mu_2(t))\}}\big\}\Big] \\
&\qquad \times [\mathbf{h}_1'\ \mathbf{X} + \mathbf{h}_2' h(\mathcal{F}_t, W) + h_3(t)]d\tilde{H}(t)\Big|^2 \\
=&P\Big|2\int_0^\tau \big[N(t)e^{\{\tilde{f}\}} + 2e^{\{2\tilde{f}\}}\big](f_1 - f_2)[\mathbf{h}_1'\ \mathbf{X} + \mathbf{h}_2' h(\mathcal{F}_t, W) + h_3(t)]d\tilde{H}(t)\Big|^2 \\
\leq& c_{10}\rho^2\big((\beta_1, \alpha_1, \mu_1) - (\beta_2, \alpha_2, \mu_2)\big),
\end{aligned}
$$

for a constant c_{10}. The second from the last equation is satisfied since $e^{f_1} - e^{f_2} = e^{\tilde{f}}(f_1 - f_2)$, and $e^{2f_1} - e^{2f_2} = 2e^{2\tilde{f}}(f_1 - f_2)$, for $f_1 = \beta_1'\mathbf{X} + \alpha_1' h(\mathcal{F}_t, W) + \mu_1(t)$, $f_2 = \beta_2'\mathbf{X} + \alpha_2' h(\mathcal{F}_t, W) + \mu_2(t)$, and $\tilde{f} = (1-\xi)f_1 + \xi f_2$, $0 \leq \xi \leq 1$. Thus condition (i) holds by Kosorok (2008, Lemma 13.3) (Lemma C.6).

Condition (ii) is also satisfied since $\{\psi(\beta_0, \alpha_0, \mu_0)[\mathbf{h}_1,\ \mathbf{h}_2, h_3] : (\mathbf{h}_1,\ \mathbf{h}_2, h_3) \in \mathcal{H}\}$ is a Donsker class.

Clearly, $S(\beta_0, \alpha_0, \mu_0) = 0$. For $h_3 \in \mathcal{F}_r$, let h_{3n} be the B-spline function approximation of h_3 with $\|h_{3n} - h_3\|_\infty = O(n^{-\nu r})$, then we have

$$S_n(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n)[\mathbf{h}_1, \mathbf{h}_2, h_{3n}] = 0.$$

Thus, for $(\mathbf{h}_1, \mathbf{h}_2, h_3) \in \mathcal{H}$,

$$
\begin{aligned}
& n^{\frac{1}{2}} S_n(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n)[\mathbf{h}_1, \mathbf{h}_2, h_3] \\
= & n^{\frac{1}{2}} \big\{ S_n(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n)[\mathbf{h}_1, \mathbf{h}_2, h_3] - S_n(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n)[\mathbf{h}_1, \mathbf{h}_2, h_{3n}] \big\} \\
= & n^{\frac{1}{2}} (\mathbb{P}_n - P)\psi(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n)[\mathbf{h}_1,\ \mathbf{h}_2, h_3] - n^{\frac{1}{2}} (\mathbb{P}_n - P)\psi(\beta_0, \alpha_0, \mu_0)[\mathbf{h}_1,\ \mathbf{h}_2, h_3] \\
& - \Big\{ n^{\frac{1}{2}} (\mathbb{P}_n - P)\psi(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n)[\mathbf{h}_1,\ \mathbf{h}_2, h_{3n}] - n^{\frac{1}{2}} (\mathbb{P}_n - P)\psi(\beta_0, \alpha_0, \mu_0)[\mathbf{h}_1,\ \mathbf{h}_2, h_{3n}] \Big\} \\
& + n^{\frac{1}{2}} \mathbb{P}_n \left\{ \psi(\beta_0, \alpha_0, \mu_0)[\mathbf{h}_1,\ \mathbf{h}_2, h_3] - \psi(\beta_0, \alpha_0, \mu_0)[\mathbf{h}_1,\ \mathbf{h}_2, h_{3n}] \right\} \\
& + n^{\frac{1}{2}} P \Big\{ \psi(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n)[\mathbf{h}_1,\ \mathbf{h}_2, h_3] - \psi(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n)[\mathbf{h}_1,\ \mathbf{h}_2, h_{3n}] \Big\} \\
\equiv & Q_{1n} - Q_{2n} + Q_{3n} + Q_{4n}.
\end{aligned}
$$

It follows from (i) that both Q_{1n} and Q_{2n} are $o_p(1)$. And

$$
\begin{aligned}
& |Q_{4n}| \\
= & \Big| 2n^{\frac{1}{2}} P \int_0^\tau [e^{\{\hat{\beta}_n' \mathbf{X} + \hat{\alpha}_n' h(\mathcal{F}_t, W) + \hat{\mu}_n(t)\}} - N(t)] e^{\{\hat{\beta}_n' \mathbf{X} + \hat{\alpha}_n' h(\mathcal{F}_t, W) + \hat{\mu}_n(t)\}} (h_3 - h_{3n}) d\tilde{H}(t) \Big| \\
\leq & c_{11} \Big| n^{\frac{1}{2}} P \int_0^\tau [e^{\{\hat{\beta}_n' \mathbf{X} + \hat{\alpha}_n' h(\mathcal{F}_t, W) + \hat{\mu}_n(t)\}} - e^{\{\beta_0' \mathbf{X} + \alpha_0' h(\mathcal{F}_t, W) + \mu_0(t)\}}](h_3 - h_{3n}) d\tilde{H}(t) \Big| \\
= & c_{11} \Big| n^{\frac{1}{2}} P \int_0^\tau e^{f^*} \{\hat{f}_n - f_0\}(h_3 - h_{3n}) d\tilde{H}(t) \Big| \\
\leq & c_{12} n^{\frac{1}{2}} \rho\big((\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n) - (\beta_0, \alpha_0, \mu_0) \big) \cdot \|h_{3n} - h_3\|_\infty \\
\leq & n^{\frac{1}{2}} O(n^{-\frac{1-\nu}{2}}) \cdot O(n^{-\nu r}) \\
= & o_p(1).
\end{aligned}
$$

for constants c_{11} and c_{12}, where $f_0 = \beta_0' \mathbf{X} + \alpha_0' h(\mathcal{F}_t, W) + \mu_0(t)$, $\hat{f}_n = \hat{\beta}_n' \mathbf{X} + \hat{\alpha}_n' h(\mathcal{F}_t, W) + \hat{\mu}_n(t)$, and $f^* = (1 - \xi) f_0 + \xi \hat{f}_n$ with $0 \leq \xi \leq 1$. Furthermore, Q_{3n}

is also $o_p(1)$ since

$$
\begin{aligned}
& P\left[\psi(\beta_0,\alpha_0,\mu_0)[\mathbf{h}_1,\ \mathbf{h}_2,h_3]-\psi(\beta_0,\alpha_0,\mu_0)[\mathbf{h}_1,\ \mathbf{h}_2,h_{3n}]\right]^2 \\
=& P\Big[2\int_0^\tau [N(t)-e^{\{\beta_0'\mathbf{X}+\alpha_0'h(\mathcal{F}_t,W)+\mu_0(t)\}}]e^{\{\beta_0'\mathbf{X}+\alpha_0'h(\mathcal{F}_t,W)+\mu_0(t)\}}(h_{3n}-h_3)(t)d\tilde{H}(t)\Big]^2 \\
\leq & c_{13}\|h_{3n}-h_3\|_\infty^2 \longrightarrow 0,
\end{aligned}
$$

for a constant c_{13}. Thus $S_n(\hat{\beta}_n,\hat{\alpha}_n,\hat{\mu}_n)=o_p(n^{-1/2})$.

For the proof of (iv), by the smoothness of $S(\beta,\alpha,\mu)$, the Fréchet differentiability holds and the derivative of $S(\beta,\alpha,\mu)$ at (β_0,α_0,μ_0), denoted by $\dot{S}(\beta_0,\alpha_0,\mu_0)$ is a map from the space $\{(\beta-\beta_0,\alpha-\alpha_0,\mu-\mu_0):(\beta,\alpha,\mu)\in\mathcal{U}\}$ to $l^\infty(\mathcal{H})$ and

$$
\begin{aligned}
& \dot{S}(\beta_0,\alpha_0,\mu_0)(\beta-\beta_0,\alpha-\alpha_0,\mu-\mu_0)[\mathbf{h}_1,\mathbf{h}_2,h_3] \\
=& \frac{\partial S(\beta_0+\varepsilon(\beta-\beta_0),\alpha_0+\varepsilon(\alpha-\alpha_0),\mu+\varepsilon(\mu-\mu_0))[\mathbf{h}_1,\ \mathbf{h}_2,h_3]}{\partial\varepsilon}\Big|_{\varepsilon=0} \\
=& 2P\int_0^\tau [2e^{\{\beta_0'\mathbf{X}+\alpha_0'h(\mathcal{F}_t,W)+\mu_0(t)\}}-N(t)]e^{\{\beta_0'\mathbf{X}+\alpha_0'h(\mathcal{F}_t,W)+\mu_0(t)\}} \\
& \qquad \times[(\beta-\beta_0)'\mathbf{X}+(\alpha-\alpha_0)'h(\mathcal{F}_t,W)+(\mu-\mu_0)] \\
& \qquad \times[\mathbf{h}_1'\ \mathbf{X}+\mathbf{h}_2'h(\mathcal{F}_t,W)+h_3(t)]d\tilde{H}(t) \\
=& 2P\int_0^\tau e^{\{2(\beta_0'\mathbf{X}+\alpha_0'h(\mathcal{F}_t,W)+\mu_0(t))\}}[(\beta-\beta_0)'\mathbf{X}+(\alpha-\alpha_0)'h(\mathcal{F}_t,W)+(\mu-\mu_0)] \\
& \qquad \times[\mathbf{h}_1'\ \mathbf{X}+\mathbf{h}_2'h(\mathcal{F}_t,W)+h_3(t)]d\tilde{H}(t) \\
\equiv & \ \sigma_1(\mathbf{h}_1,\mathbf{h}_2,h_3)'(\beta-\beta_0)+\sigma_2(\mathbf{h}_1,\mathbf{h}_2,h_3)'(\alpha-\alpha_0)+\int_0^\tau(\mu-\mu_0)d\sigma_3(\mathbf{h}_1,\mathbf{h}_2,h_3),
\end{aligned}
$$

where

$$
\sigma_1(\mathbf{h}_1,\mathbf{h}_2,h_3)=2P\left\{\int_0^\tau g_0(X,Z(t,W),t)^2[\mathbf{h}_1'\ \mathbf{X}+\mathbf{h}_2'h(\mathcal{F}_t,W)+h_3(t)]\mathbf{X}d\tilde{H}(t)\right\},
$$

$$
\sigma_2(\mathbf{h}_1,\mathbf{h}_2,h_3)=2P\left\{\int_0^\tau g_0(X,Z(t,W),t)^2[\mathbf{h}_1'\ \mathbf{X}+\mathbf{h}_2'h(\mathcal{F}_t,W)+h_3(t)]h(\mathcal{F}_t,W)d\tilde{H}(t)\right\},
$$

and

$$\sigma_3(\mathbf{h}_1, \mathbf{h}_2, h_3)(t) = 2P\left\{\int_0^t g_0(X, Z(s, W), s)^2[\mathbf{h}_1' \ \mathbf{X} + \mathbf{h}_2' h(\mathcal{F}_s, W) + h_3(s)]d\tilde{H}(s)\right\},$$

with $g_0(\mathbf{X}, Z(t, W), t) = e^{\{\beta_0'\mathbf{X}+\alpha_0' Z(t,W)+\mu_0(t)\}}$.

It remains to show that the linear map $\dot{S}(\beta_0, \alpha_0, \mu_0)$ is continuously invertible on its range. Following the proof of Theorem 2 in Zeng et al. (2005), we only need to show that for $h \in \mathcal{H}$, if $\sigma(h) = (\sigma_1(h), \sigma_2(h), \sigma_3(h)) = 0$ almost surely, then $h = 0$. Suppose that $\sigma(h) = 0$, a.s., then $\sigma_1(h)'\mathbf{h}_1+\sigma_2(h)'\mathbf{h}_2+\int_0^\tau h_3(t)d\sigma_3(h)(t) = 0$, i. e.

$$\begin{aligned}
0 =& 2P\int_0^\tau g_0(X, Z(t, W), t)^2[\mathbf{h}_1' \ \mathbf{X} + \mathbf{h}_2' h(\mathcal{F}_t, W) + h_3(t)]\mathbf{X}'d\tilde{H}(t)\mathbf{h}_1 \\
&+2P\int_0^\tau g_0(X, Z(t, W), t)^2[\mathbf{h}_1' \ \mathbf{X} + \mathbf{h}_2' h(\mathcal{F}_t, W) + h_3(t)]h(\mathcal{F}_t, W)'d\tilde{H}(t)\mathbf{h}_2 \\
&+2P\int_0^\tau g_0(X, Z(t, W), t)^2[\mathbf{h}_1' \ \mathbf{X} + \mathbf{h}_2' h(\mathcal{F}_t, W) + h_3(t)]h_3(t)d\tilde{H}(t) \\
&= 2P\int_0^\tau g_0(X, Z(t, W), t)^2[\mathbf{h}_1' \ \mathbf{X} + \mathbf{h}_2' h(\mathcal{F}_t, W) + h_3(t)]^2d\tilde{H}(t)
\end{aligned}$$

which implies that $\mathbf{h}_1' \ \mathbf{X} + \mathbf{h}_2' h(\mathcal{F}_t, W) + h_3(t) = 0, a.s.$. Hence, $\mathbf{h}_1 = 0, \mathbf{h}_2 = 0, h_3 = 0$ by C5.

Moreover, condition (v) holds since

$$
\begin{aligned}
&\Big|\sqrt{n}\Big[S(\hat{\beta}_n,\hat{\alpha}_n,\hat{\mu}_n)-S(\beta_0,\alpha_0,\mu_0)-\\
&\quad -\dot{S}(\beta_0,\alpha_0,\mu_0)(\hat{\beta}_n-\beta_0,\hat{\alpha}_n-\alpha_0,\hat{\mu}_n-\mu_0)\Big][\mathbf{h}_1,\mathbf{h}_2,h_3]\Big|\\
=&\Big|2\sqrt{n}P\int_0^\tau\Big\{(e^{\hat{f}_n}-e^{f_0})e^{\hat{f}_n}-e^{2f_0}(\hat{f}_n-f_0)\Big\}\\
&\quad\times[\mathbf{h}_1'\ \mathbf{X}+\mathbf{h}_2'h(\mathcal{F}_t,W)+h_3(t)]d\tilde{H}(t)\Big|\\
=&\Big|2\sqrt{n}P\int_0^\tau\Big\{\big[e^{f_0}(\hat{f}_n-f_0)+\frac{e^{2f_0}}{2}(\hat{f}_n-f_0)^2+o_p\big((\hat{f}_n-f_0)^2\big)\big]e^{\hat{f}_n}-e^{2f_0}(\hat{f}_n-f_0)\Big\}\\
&\quad\times[\mathbf{h}_1'\ \mathbf{X}+\mathbf{h}_2'h(\mathcal{F}_t,W)+h_3(t)]d\tilde{H}(t)\Big|\\
=&\Big|2\sqrt{n}P\int_0^\tau\Big\{e^{f_0+f^*}(\hat{f}_n-f_0)^2+\big[\frac{e^{2f_0}}{2}(\hat{f}_n-f_0)^2+o_p\big((\hat{f}_n-f_0)^2\big)\big]e^{\hat{f}_n}\Big\}\\
&\quad\times[\mathbf{h}_1'\ \mathbf{X}+\mathbf{h}_2'h(\mathcal{F}_t,W)+h_3(t)]d\tilde{H}(t)\Big|\\
\le& c\Big|2\sqrt{n}P\int_0^\tau[(\hat{f}_n-f_0)^2+o_p\big((\hat{f}_n-f_0)^2\big)][\mathbf{h}_1'\ \mathbf{X}+\mathbf{h}_2'h(\mathcal{F}_t,W)+h_3(t)]d\tilde{H}(t)\Big|\\
\le& c_{14}\sqrt{n}\Big[\rho^2\big((\hat{\beta}_n,\hat{\alpha}_n,\hat{\mu}_n)-(\beta_0,\alpha_0,\mu_0)\big)+o_p\Big(\rho^2\big((\hat{\beta}_n,\hat{\alpha}_n,\hat{\mu}_n)-(\beta_0,\alpha_0,\mu_0)\big)\Big)\Big]\\
=&O_p(n^{\frac{1}{2}-(1-\nu)})+o_p(n^{\frac{1}{2}-(1-\nu)})=o_p(1),
\end{aligned}
$$

for a constant c_{14}.

Therefore, by (i) - (v), we have

$$
\begin{aligned}
&\sqrt{n}\dot{S}(\beta_0,\alpha_0,\mu_0)(\hat{\beta}_n-\beta_0,\hat{\alpha}_n-\alpha_0,\hat{\mu}_n-\mu_0)[\mathbf{h}_1,\mathbf{h}_2,h_3]\\
=&\ \sigma_1(\mathbf{h}_1,\mathbf{h}_2,h_3)'\sqrt{n}(\hat{\beta}_n-\beta_0)+\sigma_2(\mathbf{h}_1,\mathbf{h}_2,h_3)'\sqrt{n}(\hat{\alpha}_n-\alpha_0)\\
&\ +\int_0^\tau\sqrt{n}(\hat{\mu}_n-\mu_0)(t)d\sigma_3(\mathbf{h}_1,\mathbf{h}_2,h_3)\\
=&\ -\sqrt{n}(S_n-S)(\beta_0,\alpha_0,\mu_0)[\mathbf{h}_1,\mathbf{h}_2,h_3]+o_p(1),
\end{aligned}
$$

uniformly in $\mathbf{h}_1$, $\mathbf{h}_2$ and h_3, and for each $(\mathbf{h}_1, \mathbf{h}_2, h_3) \in \mathcal{H}$, there exists unique $(\mathbf{h}_1^*, \mathbf{h}_2^*, h_3^*) \in \mathcal{H}$, such that $\sigma_1(\mathbf{h}_1^*, \mathbf{h}_2^*, h_3^*) = \mathbf{h}_1, \sigma_2(\mathbf{h}_1^*, \mathbf{h}_2^*, h_3^*) = \mathbf{h}_2, \sigma_3(\mathbf{h}_1^*, \mathbf{h}_2^*, h_3^*) = h_3$. Thus, we have

$$
\begin{aligned}
&\mathbf{h}_1'\sqrt{n}(\hat{\beta}_n - \beta_0) + \mathbf{h}_2'\sqrt{n}(\hat{\alpha}_n - \alpha_0) + \int_0^\tau \sqrt{n}(\hat{\mu}_n - \mu_0)(t)dh_3(t) \\
&= \sqrt{n}\dot{S}(\beta_0, \alpha_0, \mu_0)(\hat{\beta}_n - \beta_0, \hat{\alpha}_n - \alpha_0, \hat{\mu}_n - \mu_0)[\mathbf{h}_1^*, \mathbf{h}_2^*, h_3^*] \\
&= -\sqrt{n}(S_n - S)(\beta_0, \alpha_0, \mu_0)[\mathbf{h}_1^*, \mathbf{h}_2^*, h_3^*] + o_p(1) \\
&\longrightarrow Z \quad \text{in distribution},
\end{aligned}
$$

where Z follows $N(0, \sigma^2)$ with

$$
\sigma^2 \equiv E\psi^2(\beta_0, \alpha_0, \mu_0)[\mathbf{h}_1^*, \mathbf{h}_2^*, h_3^*]. \tag{3.8}
$$

3.5 Simulation study

We conducted a simulation study to assess the finite sample properties of the proposed estimators. We considered the situation where there were two covariates and for each subject i, X_{1i}'s and X_{2i}'s were generated from Bernoulli distribution with success probability 0.5 and the standard normal distribution. The follow-up time C_i was from the uniform distribution over interval $(\tau/2, \tau)$ with $\tau = 6$. Given the covariate $\mathbf{X}_i = (X_{1i}, X_{2i})'$, two set-ups for the observation process $H_i(t)$ were considered as follows:

(a). The number of observation times m_i was assumed to follow the Poisson distribution with mean $2C_i/\tau \exp(\gamma'\mathbf{X}_i)$ and the observation times $(T_{i1}, \ldots, T_{im_i})$ were taken to be the order statistics of a random sample of size m_i from the uniform distribution over $(0, C_i)$.

(b). The number of observation times m_i was assumed to follow the uniform distribution over $\{1,2,3,4,5,6\}$ and the observation times $(T_{i1}, \ldots, T_{im_i})$ were generated in the same way as in set-up (a).

Then, given $\mathbf{X}_i$, m_i and the observation times $(T_{m_i,1}, \cdots, T_{m_i,m_i})$, we generated recurrent event counts $\tilde{N}_{m_i} = (N_i(T_{m_i,1}), \cdots, N_i(T_{m_i,m_i}))$ from a Poisson process by taking $N_i(T_{m_i,j}) = N_i(T_{m_i,1}) + \{N_i(T_{m_i,2}) - N_i(T_{m_i,1})\} + \cdots + \{N_i(T_{m_i,j}) - N_i(T_{m_i,j-1})\}$, where

$$\begin{aligned} N_i(t) - N_i(s) \sim & \operatorname{Poisson}(\exp\{\mu_0(t) + \beta_0' \mathbf{X}_i + \alpha_0 H_i(t-) X_{1i}\} \\ & - \exp\{\mu_0(s) + \beta_0' \mathbf{X}_i + \alpha_0 H_i(s-) X_{1i}\}). \end{aligned}$$

Set $\mu_0(t) = \sqrt{t}$ or $\mu_0(t) = \log(t+1)$, $\alpha_0 = 0, 0.3$, or 0.5, representing the different correlations between the panel count process and the observation process, and $\beta_0 = (-0.5, 0.5)$, representing the different effects of the covariate $\mathbf{X}$ on the recurrent event counts. To estimate the smooth function $\mu_0(t)$, we considered cubic B-splines and took $m_n = n^\nu$ with $\nu = 1/10$, $1/3$ or $2/5$. For a given number of interior knots m_n, we consider two data-driven methods for determing locations of knots. One is the equally spaced knots, which are given by $T_{\min} + k(T_{\max} - T_{\min})/(m_n + 1), k = 0, 1, \cdots, m_n + 1$, with $T_{\min}$ and $T_{\max}$ being the respective minimum and maximum values of distinct observation times. Another is the partitions according to quantiles of the observation times, i. e., the $k/(m_n+1)$ quantiles $(k = 0, 1, \cdots, m_n + 1)$ of the distinct observation times are chosen to be the knots. We have done simulation for the six combinations of the number and placement of knots and illustrate the estimation results for different combinations with $W = X_1$, $\alpha = 0.3$, $\mu_0(t) = \log(t + 1)$ and $n = 100$ in Table 3.1. From this table, we find that the estimation results are very similar and not sensitive to the selection of number and placement of knots. Thus in the following, we present the overall results with number of interior knots chosen to be $n^{1/3}$ and the equally spaced knots.

Tables 3.2 and 3.3 present the simulation results on estimation of β_0 and α_0 for Poisson and non-Poisson observation processes with sample size $n = 100$ or 200 and $\mu_0(t) = \sqrt{t}$ and $\log(t + 1)$, respectively. The tables include the estimated bias (BIAS) given by the average of the estimates minus the true value, the sample standard deviation error of estimates (SSE), the mean of the bootstrap standard errors of the estimates (BSE), and the bootstrap 95% coverage probability (CP) obtained from 1000 independent runs. Here we used 100 replications in bootstrap to estimate the standard errors. It can be seen from the tables that the proposed estimators are unbiased for different situations considered, which means that our estimation approach does not rely on the Poisson distributional assumption about the observation process, thus it is more robust than the previous analysis of panel count data with informative observation process under the Poisson assumption, such as Hu et al. (2003), Li et al. (2010) and Zhao and Tong (2011). Also, the SSE and the BSE are quite close to each other and smaller as the sample size increases, which indicates that proposed bootstrap variance estimation procedure provides reasonable estimates. In addition, the 95% bootstrap CP are consistent with the nominal level, which suggests that the normal approximation seems to be appropriate.

Figures 3.1 and 3.2 show the estimation results of $\mu_0(t) = \sqrt{t}$ and $\mu_0(t) = \log(t+1)$, respectively, for simulated panel count data with Possion and Non-Poisson observation processes, $h(\mathcal{F}_t, W) = H(t-)X_1$ and $\alpha = 0.5$. In the figures, the solid line represents the real curve of $\mu_0(t)$, and the point line and the dotted line represent the B-spline based estimation curves of $\mu_0(t)$ for the sample size $n = 100$ and $n = 200$, respectively. Based on the figures, we have the finding that the B-spline based estimation curve of $\mu_0(t)$ is close to its real curve with the moderate sample size and especially closer as the sample size increase in all different situations, indicating that the B-spline estimator for $\mu_0(t)$ works well.

Note that our simulation results for estimation of the regression parameters and the nonparametric function are all reasonable with the moderate sample size even when α diverges far from 0, which is superior to the results in Li et al. (2010), where they proposed a semiparametric transformation model for the underlying recurrent event process, but with a nonhomogeneous Poisson restriction on the informative observation times. Thus, our proposed models and estimation procedure are more flexible and robust.

Table 3.1. Estimation results of $(\beta_0', \alpha_0')'$ for simulated panel count data with non-Poisson observation processes, $h(\mathcal{F}_t, W) = H(t-)X_1$, $\alpha = 0.3$, $\mu_0(t) = \log(t+1)$ and $n = 100$ by chosing different numbers and placements of knots.

ν	Partitions	$\hat{\beta}_1$				$\hat{\beta}_2$				$\hat{\alpha}$			
		BIAS	SSE	BSE	CP	BIAS	SSE	BSE	CP	BIAS	SSE	BSE	CP
$\nu = 1/10$	Q	-0.0004	0.1557	0.1456	0.9330	0.0040	0.0575	0.0548	0.9530	0.0001	0.0406	0.0353	0.9300
	E	-0.0022	0.1492	0.1458	0.9440	0.0047	0.0564	0.0551	0.9450	0.0019	0.0368	0.0357	0.9500
$\nu = 1/3$	Q	-0.0005	0.1567	0.1457	0.9370	0.0048	0.0549	0.0547	0.9460	0.0004	0.0401	0.0356	0.9330
	E	-0.0008	0.1562	0.1458	0.9280	0.0040	0.0577	0.0552	0.9510	0.0000	0.0415	0.0361	0.9300
$\nu = 2/5$	Q	-0.0009	0.1562	0.1455	0.9340	0.0043	0.0542	0.0554	0.9530	0.0006	0.0387	0.0360	0.9320
	E	0.0031	0.1511	0.1453	0.9540	0.0023	0.0640	0.0549	0.9360	-0.0006	0.0444	0.0363	0.9520

Note: Q: Knots partitions according to quantiles of the the observetion times and E: Equally spqced knots based on the observetion times.

Table 3.2. Estimation results of $(\beta_0', \alpha_0')'$ for simulated panel count data with Poisson and non-Poisson observation processes, $h(\mathcal{F}_t, W) = H(t-)X_1$ and $\mu_0(t) = \sqrt{t}$

n	α		$\hat{\beta}_1$				$\hat{\beta}_2$				$\hat{\alpha}$			
			BIAS	SSE	BSE	CP	BIAS	SSE	BSE	CP	BIAS	SSE	BSE	CP
100	0	P	-0.0076	0.1530	0.1456	0.9350	0.0075	0.0791	0.0739	0.9340	-0.0081	0.0927	0.1013	0.9590
		NP	-0.0018	0.1368	0.1295	0.9330	0.0027	0.0530	0.0532	0.9520	-0.0007	0.0395	0.0388	0.9390
	0.3	P	-0.0060	0.1638	0.1482	0.9330	-0.0015	0.0805	0.0731	0.9210	-0.0024	0.0790	0.0863	0.9530
		NP	-0.0011	0.1313	0.1217	0.9320	0.0013	0.0473	0.0455	0.9440	0.0002	0.0325	0.0302	0.9380
	0.5	P	-0.0070	0.1630	0.1511	0.9350	0.0025	0.0808	0.0713	0.9370	-0.0022	0.0797	0.0827	0.9590
		NP	-0.0082	0.1381	0.1200	0.9310	-0.0091	0.0779	0.0515	0.9660	-0.0047	0.0575	0.0385	0.9730
200	0	P	0.0017	0.1061	0.1029	0.9380	0.0002	0.0577	0.0519	0.9250	-0.0018	0.0588	0.0587	0.9530
		NP	-0.0039	0.1033	0.0925	0.9330	0.0010	0.0419	0.0375	0.9250	0.0003	0.0307	0.0274	0.9180
	0.3	P	-0.0077	0.1119	0.1063	0.9410	0.0013	0.0591	0.0519	0.9420	0.0001	0.0502	0.0506	0.9560
		NP	0.0024	0.0879	0.0863	0.9360	0.0018	0.0323	0.0327	0.9490	0.0000	0.0216	0.0224	0.9540
	0.5	P	0.0054	0.1178	0.1098	0.9520	-0.0059	0.0770	0.0583	0.9640	-0.0064	0.0640	0.0512	0.9660
		NP	0.0004	0.1035	0.0959	0.9500	-0.0115	0.0755	0.0573	0.9690	-0.0094	0.0569	0.0449	0.9680

Note: P and NP represent observation times generated from Poisson and non-Poisson processes, respectively.

Table 3.3. Estimation results of $(\beta_0', \alpha_0')'$ for simulated panel count data with Poisson and non-Poisson observation processes, $h(\mathcal{F}_t, W) = H(t-)X_1$ and $\mu_0(t) = \log(t+1)$

n	α		$\hat{\beta}_1$				$\hat{\beta}_2$				$\hat{\alpha}$			
			BIAS	SSE	BSE	CP	BIAS	SSE	BSE	CP	BIAS	SSE	BSE	CP
100	0	P	-0.0102	0.1883	0.1779	0.9320	0.0080	0.1030	0.0942	0.9280	-0.0066	0.1085	0.1268	0.9670
		NP	-0.0098	0.1661	0.1555	0.9320	0.0021	0.0686	0.0652	0.9310	0.0004	0.0518	0.0467	0.9220
	0.3	P	-0.0101	0.1794	0.1800	0.9420	0.0028	0.0934	0.0873	0.9360	-0.0025	0.0918	0.1011	0.9510
		NP	-0.0008	0.1562	0.1458	0.9280	0.0040	0.0577	0.0552	0.9510	0.0000	0.0415	0.0361	0.9300
	0.5	P	-0.0057	0.1904	0.1778	0.9350	0.0017	0.1053	0.0883	0.9420	-0.0024	0.0925	0.0947	0.9630
		NP	0.0023	0.1460	0.1430	0.9460	-0.0066	0.0731	0.0597	0.9620	-0.0068	0.0546	0.0448	0.9680
200	0	P	-0.0031	0.1264	0.1249	0.9480	0.0068	0.0689	0.0636	0.9330	-0.0004	0.0738	0.0715	0.9320
		NP	-0.0074	0.1118	0.1103	0.9380	0.0034	0.0475	0.0456	0.9420	-0.0010	0.0335	0.0328	0.9490
	0.3	P	-0.0016	0.1274	0.1266	0.9540	-0.0040	0.0819	0.0636	0.9350	-0.0024	0.0637	0.0591	0.9380
		NP	0.0038	0.1087	0.1054	0.9420	0.0014	0.0458	0.0402	0.9510	-0.0011	0.0323	0.0276	0.9560
	0.5	P	-0.0033	0.1409	0.1315	0.9370	-0.0051	0.0940	0.0694	0.9490	-0.0035	0.0622	0.0598	0.9590
		NP	0.0014	0.1195	0.1125	0.9510	-0.0133	0.0712	0.0646	0.9610	-0.0121	0.0533	0.0514	0.9560

Note: P and NP represent observation times generated from Poisson and non-Poisson processes, respectively.

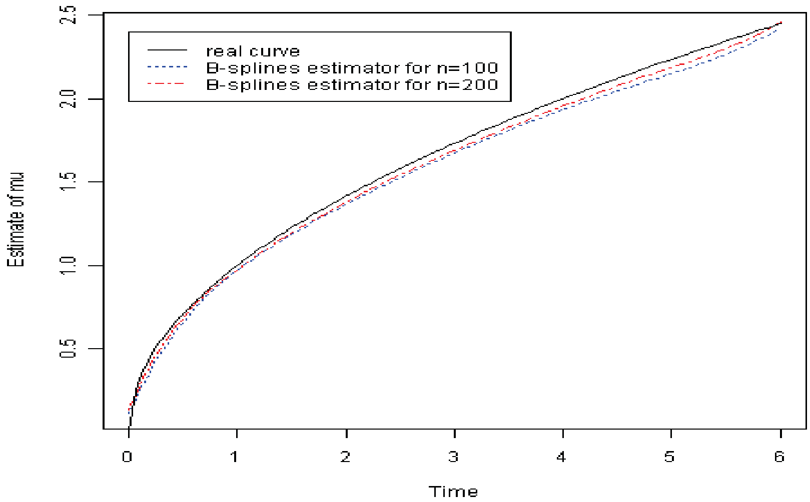
3.6 Application

This section presents an analysis of the bladder cancer data by applying our proposed methods. There were 116 subjects with superficial bladder tumors randomized into one of three treatment groups: placebo, thiotepa, and pyridoxine. In the following, we restrict our attention to the placebo and thiotepa groups with respective sizes of 47 and 38 as it has been shown that the pyridoxine treatment had no effect on the recurrence of the bladder tumors (Zhang, 2002). For each patient, the observed information includes times when he or she made clinical visits and the numbers of recurrent tumors between clinical visits. Two baseline covariates were observed and they are the number of initial tumors and the size of the largest initial tumor.

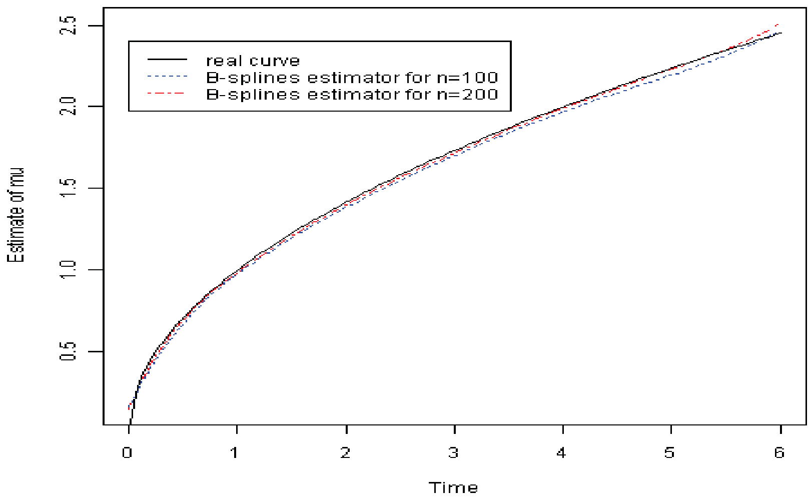
To analyze the data, for patient i, define x_{1i} to be equal to 1 if the ith patient was given the thiotepa treatment and 0 otherwise, x_{2i} the number of initial tumors and x_{3i} the size of the largest initial tumor, $i = 1, \ldots, 85$. We define the response $N_i(t)$ to be the cumulated new tumor numbers of patient i up to time t. Let $H_i(\cdot)$ represent the accumulated observation numbers of patient i over the study period. Assume that $\{N_i(t)\}$ can be described by model (3.1) with $h(\mathcal{F}_{it}) = H_i(t-)X_{1i}$, meaning that the relation between recurrence rate of bladder tumors and the observation times are related through the total number of observations., i.e.,

$$E\{N_i(t)|X_{1i}, X_{2i}, X_{3i}, \mathcal{F}_{it}\} = \exp\{\mu_0(t) + \beta_1' X_{1i} + \beta_2' X_{2i} + \beta_3' X_{3i} + \alpha' H_i(t-)X_{1i}\}.$$

Here, we took the last visit time of patient i as C_i in the analysis. For estimation

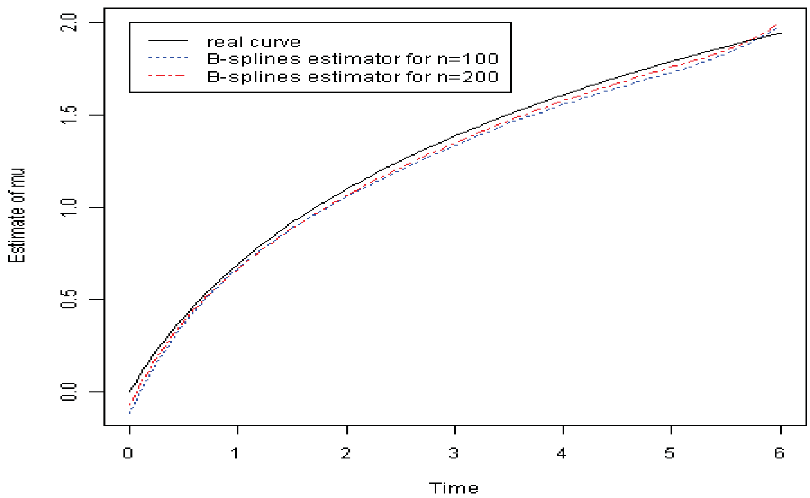


(a) Poisson observation process

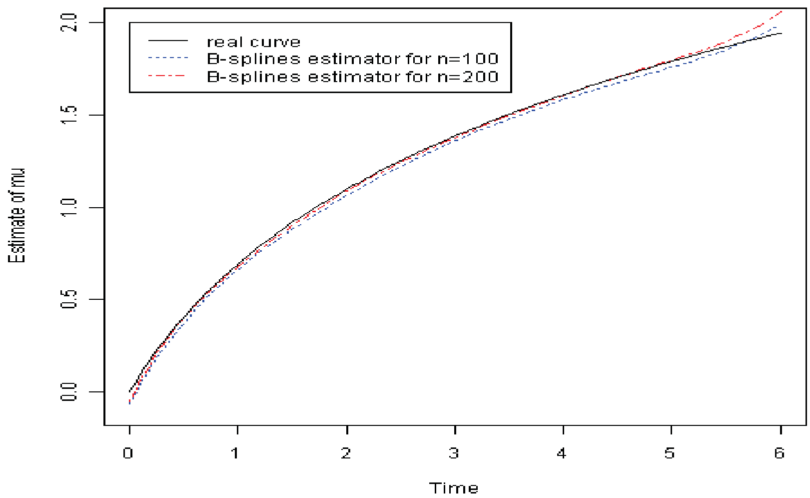


(b) Non-Poisson observation process

Figure 3.1. Estimates of $\mu_0(t) = \sqrt{t}$ for simulated panel count data with Poisson and non-Poisson observation processes, $h(\mathcal{F}_t, W) = H(t-)X_1$ and $\alpha = 0.5$



(a) Poisson observation process



(b) Non-Poisson observation process

Figure 3.2. Estimates of $\mu_0(t) = \log(t+1)$ for simulated panel count data with Poisson and non-Poisson observation processes, $h(\mathcal{F}_t, W) = H(t-)X_1$ and $\alpha = 0.5$

of $\mu_0(t)$, we use the cubic B-spline approximation by taking the number of interior knots m_n as n^v with $v = 1/3$ and the equally spaced knots.

The application of the estimation procedure proposed in the previous sections gave $\hat{\beta}_1 = -0.9006, \hat{\beta}_2 = 0.1980, \hat{\beta}_3 = -0.0658$, and $\hat{\alpha} = -0.4076$ with the bootstrap standard errors being 0.5051, 0.1009, 0.2054, and 0.1510, which correspond to p-values of 0.0746, 0.0497, 0.7486, and 0.0069, respectively, based on the asymptotic results of the estimators. Here $\hat{\beta}_1, \hat{\beta}_2$ and $\hat{\beta}_3$ represent the estimated regression coefficients corresponding to the treatment indicator, the number of initial tumors, and the size of the largest initial tumor, respectively, while $\hat{\alpha}$ represents the estimated effect of the interaction between the observation process and the treatment indicator on the tumor recurrence rate.

These results indicate that the recurrent event process and the interaction between the observation process and the treatment indicator are significantly negatively correlated, which is consistent with the analysis results in Sun et al. (2005) and Section 2.6. Furthermore, the thiotepa treatment significantly reduces the occurrence rate of the bladder tumors, and the number of initial tumors has a significant positive effect on the tumor recurrence rate. However, the occurrence rate of the bladder tumors do not seem to be significantly related to the size of the largest initial tumor. These conclusions are roughly consistent with those presented in Li et al. (2010), and Zhao and Tong (2011). Compared to the models in Li et al. (2010), and Zhao and Tong (2011), our proposed procedure could be useful to estimate the future recurrence rate based on the observation history.

Chapter 4

Panel Count Data Analysis with Time-Dependent Covariates

In this chapter, we consider some semiparametric regression analysis of panel count data with time-dependent covariates and information observation and censoring times.

4.1 Introduction

In many situations, the underlying recurrent process and the observation process are still dependent even given covariates. For this, Zhao and Tong (2011) proposed a joint modeling approach that used an unobserved frailty variable and a completely unspecified link function to characterize the correlation between the recurrent event process and the observation times with time-independent covariates. However, in some applications, panel count data with informative observation times, and also with time-dependent covariates and informative censoring times may exist, when a failure time is correlated to the censoring mechanismand some associated covariates vary with time. Thus it is desirable to develop estimation procedures for panel count data with informative observation and censoring times, and also with time-dependent covariates. For this, we considered the same models for the underlying recurrent events and the observation times as given in Zhao and Tong (2011) except replacing the time-independent covariates with the time-dependent covariates and removing the assumption of noninformative censoring.

The remainder of this chapter is organized as follows. We begin in Section 4.2 by introducing some notation and describing statistical models for the underlying recurrent event process and the observation process. In Section 4.3, a novel estimation procedure that does not depend on the distribution of frailty variables and the link function is proposed for estimation of regression parameters and the asymptotic properties including consistency and asymptotic normality of the proposed estimators are established in Section 4.4. In order to assess the finite-sample properties of the proposed inference procedure, we present some results obtained from simulation studies in Section 4.5. In Section 4.6, the proposed approaches are illustrated through the analysis of a data set from the bladder tumor study.

4.2 Statistical Models

Consider a recurrent event study that consists of n independent subjects, and let $N_i(t)$ denote the number of occurrences of the recurrent event of interest before or at time t for subjects i. Suppose that for each subject, there exist a p-dimensional possibly time-dependent covariates, denoted by $\mathbf{X}_i(t)$, and Z_i is an unobserved positive random variable that is independent of the covariates. Then, for subject i, given $\mathbf{X}_i(t)$ and Z_i, the mean function of $N_i(t)$ is assumed to have the form

$$E\{N_i(t)|\mathbf{X}_i(t), Z_i\} = \mu_0(t)g(Z_i)\exp\{\mathbf{X}_i'(t)\beta_0\}, \tag{4.1}$$

where $\mu_0(\cdot)$ is a completely unknown continuous baseline mean function, β_0 is a vector of unknown regression parameters, and $g(\cdot)$ is a completely unspecified function with $E(g(Z)) = 1$. Since $N_i(t)$ is a counting process, the choice of time-dependent covariates should be constrained by the fact that $E\{N_i(t)|\mathbf{X}_i(t), Z_i\}$ is a nondecreasing function of time. Also the covariate histories $\{\mathbf{X}_i(t) : 0 \leq t \leq C_i\}(i = 1, \cdots, n)$ are assumed to be observed.

For subject i, suppose that $N_i(\cdot)$ is observed only at finite time points $T_{i1} < \cdots < T_{iK_i}$, where K_i denotes the potential number of observation times, $i = 1, \cdots, n$. That is, only the values of $N_i(t)$ at these observation times are known and we have panel count data on the $N_i(t)$'s. Let C_i be the censoring time and thus $N_i(t)$ is observed only at these T_{ij}'s with $T_{ij} \leq C_i$, $i = 1, \cdots, n$. Define $\tilde{H}_i(t) = H_i\{\min(t, C_i)\}$, where $H_i(t) = \sum_{j=1}^{K_i} I\{T_{ij} \leq t\}, i = 1, \cdots, n$, and $I(\cdot)$ is a indicator function. Then $\tilde{H}_i(\cdot)$ is a point process characterizing the ith subject's observation process and jumps only at the observation times.

In the following, we assume that given $\mathbf{X}_i(t)$ and Z_i, $H_i(\cdot)$ is a nonhomogeneous Poisson process with the intensity function

$$\lambda(t\ |\mathbf{X}_i(t), Z_i) = \lambda_0(t) Z_i \exp\{\mathbf{X}_i'(t)\gamma_0\}, \tag{4.2}$$

where $\lambda_0(\cdot)$ is a completely unknown continuous baseline intensity function and γ_0 denotes a vector of regression parameters. Here, we assume that $E(Z) = 1$ for identifiability. Let $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$. In addition, we assume that conditional on the covariates $\mathbf{X}_i(t)$'s and Z_i's, N_i's, H_i's and C_i's are mutually independent, and $\{H_i(t), N_i(t), \mathbf{X}_i(t), C_i, Z_i, 0 \leq t \leq \tau\}, i = 1, \cdots, n$, are independent and identically distributed, where τ is the length of study.

The special cases of models (4.1) and (4.2) have been studied individually by earlier researchers. For example, model (4.1) with $g(Z_i) = 1$ and time-independent covariates was considered by Sun and Wei (2000), Zhang (2002), and Wellner and Zhang (2007) for regression analysis of panel count data; Huang et al. (2010) considered model (4.2) with time-dependent and time-independent covariates, and Wang et al. (2001) and Huang and Wang (2004) considered model (4.2) with time-independent covariates for recurrent event data; Furthermore, Zhao and Tong (2011) developed the joint analysis of the two models with time-independent covariates.

In the following, we study the joint analysis of the two models together. The proposed models allow the underlying recurrent event process and the observation process to be correlated through their connections with the link function of the frailty; moreover, both the link function and the distribution of the frailty are considered as nuisance parameters. Our main goal here is to make inference about β. Toward this end, we develop a novel estimation procedure that depends neither on the form of the link function nor on the distribution of the frailty in the next section.

4.3 Estimation Procedure

For estimation of β_0 along with other parameters, define $\tilde{N}_i(t) = \int_0^t N_i(s) d\tilde{H}_i(s)$, then this newly defined process only has possible jumps at the observation time points $\{T_{ij} \wedge C_i : j = 1, \cdots, K_i\}$ with respective jump sizes $N_i(T_{ij}), i = 1, \cdots, n$.

Thus we have

$$
\begin{aligned}
& E\{d\tilde{N}_i(t)|\mathbf{X}_i(t), C_i\} \\
= \;& E\{\xi_i(t)E[N_i(t)dH_i(t)|\mathbf{X}_i(t), C_i, z_i]|\mathbf{X}_i(t), C_i\} \\
= \;& E\{\xi_i(t)E[N_i(t)|\mathbf{X}_i(t), z_i]E[dH_i(t)|\mathbf{X}_i(t), z_i]|\mathbf{X}_i(t), C_i\} \\
= \;& E\{\xi_i(t)\mu_0(t)g(Z_i)\exp\{\mathbf{X}_i'(t)\beta_0\}Z_i\exp\{\mathbf{X}_i'(t)\gamma_0\}d\Lambda_0(t)|\mathbf{X}_i(t), C_i\} \\
= \;& \exp\{\mathbf{X}_i'(t)(\beta_0+\gamma_0)\}E[g(Z_i)Z_i]\xi_i(t)\mu_0(t)d\Lambda_0(t) \\
= \;& \exp\{\mathbf{X}_i'(t)\theta_0\}\xi_i(t)d\phi_0(t).
\end{aligned}
$$

where $\theta_0 = \beta_0 + \gamma_0$, $\xi_i(t) = I(C_i \geq t)$ and $\phi_0(t) = \int_0^t E[g(Z)Z]\mu_0(s)d\Lambda_0(s)$.

Similar to Hu et al. (2003), borrowing the structure of the Cox partial likelihood score function of the Andersen-Gill proportional intensity model (Andersen and Gill, 1982), which is also asymptotically unbised for a more general non-Poisson process (Lawless and Nadeau, 1995), we construct an estimating equation of θ_0 in the form of

$$U(\theta; \tilde{N}) = \sum_{i=1}^{n} \int_0^{\tau} W(t)\{\mathbf{X}_i(t) - \bar{\mathbf{X}}(t;\theta)\}d\tilde{N}_i(t) = 0$$

where $\bar{\mathbf{X}}(t;\theta) = S^{(1)}(t;\theta)/S^{(0)}(t;\theta)$, and

$$S^{(k)}(t;\theta) = n^{-1}\sum_{i=1}^{n}\xi_i(t)\mathbf{X}_i(t)^{\otimes k}\exp\{\mathbf{X}_i'(t)\theta\}, \qquad k = 0, 1, 2,$$

where $a^{\otimes 0} = 1, a^{\otimes 1} = a, a^{\otimes 2} = aa'$ for a vector a .

It can be shown that this estimating equation $U(\theta;\tilde{N}) = 0$ is unbiased for θ (i.e., $E[U(\theta_0;\tilde{N})] = 0$). Solving the estimating equation provides us with an estimator of θ_0, denoted by $\hat{\theta}$, and thus, given γ_0, β_0 can be estimated by $\hat{\theta} - \gamma_0$.

But γ_0 is unknown, we need to find an estimator for it.

Since

$$\begin{aligned} E\{dH_i(t)|\mathbf{X}_i(t)\} &= E\{E[dH_i(t)|\mathbf{X}_i(t), Z_i]|\mathbf{X}_i(t)\} \\ &= \exp\{\mathbf{X}_i'(t)\gamma_0\}d\Lambda_0(t) \end{aligned}$$

and C_i's are independent of (N_i, H_i)'s conditional on covariate and the frailty, as in Liang et al. (2009), the methods proposed by Lin et al. (2000) for the proportional rate model can be used to consistently estimate γ_0 and $\Lambda_0(\cdot)$. To be specific, γ_0 can be consistently estimated from the following estimating equation

$$U_2(\gamma;\tilde{H}) = \sum_{i=1}^{n}\int_0^{\tau}\{\mathbf{X}_i(t) - \bar{\mathbf{X}}(t;\gamma)\}d\tilde{H}_i(t) = 0,$$

where $\bar{\mathbf{X}}(t;\gamma) = S^{(1)}(t;\gamma)/S^{(0)}(t;\gamma)$, and

$$S^{(k)}(t;\gamma) = n^{-1}\sum_{i=1}^{n}\xi_i(t)\mathbf{X}_i(t)^{\otimes k}\exp\{\mathbf{X}_i'(t)\gamma\}, \qquad k = 0,\ 1,\ 2.$$

The resulting estimator is denoted by $\hat{\gamma}$. In addition, $\Lambda_0(t)$ can be consistently estimated by the Aalen-Breslow-type estimator $\hat{\Lambda}_0(t) = \hat{\Lambda}_0(t;\hat{\gamma})$, where

$$\hat{\Lambda}_0(t;\gamma) = \sum_{i=1}^{n}\int_0^t \frac{d\tilde{H}_i(s)}{nS^{(0)}(s\ ;\gamma)}.$$

4.4 Asymptotic Theory

Let

$$s^{(k)}(t;\mu) = \lim_{n\to\infty} S^{(k)}(t;\mu) = E[\xi_1(t)\exp\{\mathbf{X}_1'(t)\mu\}\mathbf{X}_1(t)^{\otimes k}], \qquad k = 0,\ 1,\ 2,$$

and define $\bar{\mathbf{x}}(t;\mu) = s_1^{(1)}(t;\mu)/s_1^{(0)}(t;\mu)$.

To establish the asymptotic properties of $\hat{\theta}$, we need the following regularity conditions.

(C.1.) $P(C \geq \tau) > 0$.

(C.2.) $\mathbf{X}_i(t), i = 1, \cdots, n$ have bounded total variations, i.e. $|X_{ji}(0)| + \int_0^\tau |X_{ji}(t)| \leq M_0$ for all $j = 1, \cdots, p$ and $i = 1, \cdots, n$, where X_{ji} is the jth component of $\mathbf{X}_i$ and M_0 is a constant.

(C.3.) $\Lambda_0(\tau) \leq M_1, \mu_0(\tau) \leq M_2$, where M_1, M_2 are constants.

(C.4.) $N_i(\tau)$ $(i = 1, \cdots, n)$ are bounded by a constant and the K_i's are bounded; $W(\cdot)$ is nonnegative and have bounded total variations with $W(\cdot) \to w(\cdot)$, as $n \to \infty$.

(C.5.)

$$A_\theta(\theta_0) \equiv E\left[\int_0^\tau w(t)\{\mathbf{X}_1(t) - \bar{\mathbf{x}}(t;\theta_0)\}^{\otimes 2}\xi_1(t)\exp\{\mathbf{X}_1'(t)\theta_0\}d\phi_0(t)\right],$$

and

$$A_\gamma(\gamma_0) \equiv E\left[\int_0^\tau \{\mathbf{X}_1(t) - \bar{\mathbf{x}}(t;\gamma_0)\}^{\otimes 2}\xi_1(t)\exp\{\mathbf{X}_1'(t)\gamma_0\}d\Lambda_0(t)\right]$$

are positive definite.

In practice, condition (C.1) can be enforced simple by not choosing τ to be greater than the maximum observation time. The boundedness conditions in (C.2), (C.3) and (C.4) simplify the derivation of the asymptotic results. Condition (C.5) can be interpreted that the sample covariance is asymptotically non-singular. The asymptotic properties are summarized as follows.

Theorem 4.1 (Consistency of $\hat{\theta}$). *Under conditions (C.1 – C.5),* $\hat{\theta} \longrightarrow \theta_0, a.s.$

Proof of Theorem 4.1.

By the strong law of large numbers, for each $t \in [0, \tau]$, $S^{(k)}(t;\theta)$ converges almost surely to $s^{(k)}(t;\theta)$, for every θ, $k = 0,\ 1,\ 2$. Define

$$Y_n(\theta) \equiv n^{-1} \sum_{i=1}^{n} \int_0^{\tau} W(t) \left[(\theta - \theta_0)' \mathbf{X}_i(t) - \log\{S^{(0)}(t;\theta)/S^{(0)}(t;\theta_0)\} \right] d\tilde{N}_i(t)$$

and

$$\mathcal{Y}(\theta) \equiv E\left[\int_0^{\tau} w(t) \left[(\theta - \theta_0)' \mathbf{X}_1(t) - \log\{s^{(0)}(t;\theta)/s^{(0)}(t;\theta_0)\} \right] d\tilde{N}_1(t) \right].$$

We can see that $Y_n(\theta)$ converges almost surely to $\mathcal{Y}(\theta)$, for every θ and

$$\partial Y_n(\theta)/\partial\theta = n^{-1} U(\theta; \tilde{N}).$$

Note that

$$\begin{aligned} & \partial^2 Y_n(\theta)/\partial\theta\partial\theta' \\ &= -n^{-1} \sum_{i=1}^{n} \int_0^{\tau} W(t)\xi_i(t) \exp(\mathbf{X}_i'(t)\theta)[\mathbf{X}_i(t) - \bar{\mathbf{X}}(t;\theta)]^{\otimes 2} d\left[\frac{n^{-1}\sum_{j=1}^{n} \tilde{N}_j(t)}{S^{(0)}(t;\theta)} \right] \\ &\equiv -\hat{A}_\theta(\theta) \end{aligned}$$

is negative semidefinite. Thus, $Y_n(\theta)$ is concave, which implies that the convergence of $Y_n(\theta)$ to $\mathcal{Y}(\theta)$ is uniform on any compact set of θ (Rockafellar, 1970, Th 10.8). In particular, letting $\mathcal{A}_\epsilon(\theta_0) = \{\theta : \| \theta - \theta_0 \| \le \varepsilon\}$, we have

$$\sup_{\theta \in \mathcal{A}_\epsilon(\theta_0)} \| Y_n(\theta) - \mathcal{Y}(\theta) \| \longrightarrow 0 \tag{4.3}$$

almost surely. It is easy to show that $\partial \mathcal{Y}(\theta_0)/\partial\theta = 0$ and

$$\partial^2 \mathcal{Y}(\theta_0)/\partial\theta\partial\theta' = -A_\theta(\theta_0),$$

where $A_\theta(\theta_0)$ is positive definite (condition C.5). Thus $\mathcal{Y}(\theta)$ has a unique maximizer θ_0.

In particular, $\sup_{\theta\in\partial\mathcal{A}_\epsilon(\theta_0)} \mathcal{Y}(\theta) < \mathcal{Y}(\theta_0)$, where

$$\partial\mathcal{A}_\epsilon(\theta_0) = \{\theta : \| \theta - \theta_0 \| = \varepsilon\}.$$

This fact, together with (4.3) implies that $Y_n(\theta) < Y_n(\theta_0)$ for all $\theta \in \partial\mathcal{A}_\epsilon(\theta_0)$ and all large n. Therefore, there must be a maximizer of $Y_n(\theta)$, i.e., a solution to $\partial Y_n(\theta)/\partial\theta = 0$, say $\hat{\theta}$, in the interior of $\mathcal{A}_\epsilon(\theta_0)$.

On the other hand, $\partial^2 Y_n(\theta)/\partial\theta\partial\theta'$ converges almost surely to $\partial^2\mathcal{Y}(\theta_0)/\partial\theta\partial\theta'$. This along with the fact that $\partial^3 Y_n(\theta)/\partial\theta\partial\theta'\partial\theta'$ is bounded ensures existence of ε, such that $\partial^2 Y_n(\theta)/\partial\theta\partial\theta'$ is negative definitive for $\theta \in \mathcal{A}_\epsilon(\theta_0)$, when n is large enough. Thus the fact that $\partial^2 Y_n(\hat{\theta})/\partial\theta\partial\theta'$ is negative definitive implies that $\hat{\theta}$ is the unique global maximizer of $Y_n(\theta)$ in $\mathcal{A}_\epsilon(\theta_0)$, i.e., the unique solution to $U(\theta; \tilde{N}) = 0$.

Finally, since ϵ can be chosen arbitrarily small, $\hat{\theta}$ must converge to θ_0 almost surely, as $n \longrightarrow \infty$.

Since $\hat{\gamma}$ is consistent as in Lin et al. (2000), then $\hat{\beta} = \hat{\theta} - \hat{\gamma}$ is a consistent estimator of β_0.

A consistent Aalen-Breslow-type estimator for $\phi_0(t)$ can be obtained as follows,

$$\hat{\phi}_0(t) = \hat{\phi}_0(t;\hat{\theta}) = \int_0^t \frac{\sum_{i=1}^n d\tilde{N}_i(s)}{nS^{(0)}(s;\hat{\theta})}, \qquad t \in [0,\tau].$$

To establish the asymptotic normality of $\hat{\beta}$, define

$$\begin{aligned}
\hat{\tilde{M}}_i(t;\hat{\theta}) &= \tilde{N}_i(t) - \int_0^t \xi_i(s)\exp(\mathbf{X}_i'(s)\hat{\theta})d\hat{\phi}_0(s),\\
\hat{M}_i(t;\hat{\gamma}) &= \tilde{H}_i(t) - \int_0^t \xi_i(s)\exp\{\mathbf{X}_i'(s)\hat{\gamma}\}d\hat{\Lambda}_0(s),\\
\hat{A}_\theta &= \hat{A}_\theta(\hat{\theta}) = n^{-1}\sum_{i=1}^n \int_0^\tau W(t)\xi_i(t)\exp(\mathbf{X}_i'(t)\hat{\theta})[\mathbf{X}_i(t) - \bar{\mathbf{X}}(t;\hat{\theta})]^{\otimes 2}d\hat{\phi}_0(t),\\
\hat{A}_\gamma &= \hat{A}_\gamma(\hat{\gamma}) = n^{-1}\sum_{i=1}^n \int_0^\tau \{\mathbf{X}_i(t) - \bar{\mathbf{X}}(t;\hat{\gamma})\}^{\otimes 2}\xi_i(t)\exp\{\mathbf{X}_i'(t)\hat{\gamma}\}d\hat{\Lambda}_0(t),\\
\hat{a}_i &= \hat{A}_\theta^{-1}\int_0^\tau W(t)[\mathbf{X}_i(t) - \bar{\mathbf{X}}(t;\hat{\theta})]d\hat{\tilde{M}}_i(t;\hat{\theta}),\\
\hat{b}_i &= \hat{A}_\gamma^{-1}\int_0^\tau [\mathbf{X}_i(t) - \bar{\mathbf{X}}(t;\hat{\gamma})]d\hat{M}_i(t;\hat{\gamma}),
\end{aligned}$$

and $\hat{c}_i = \hat{a}_i - \hat{b}_i$.

Theorem 4.2 (**Asymptotic normality of** $\hat{\beta}$). *Under conditions (C.1 – C.5), $n^{1/2}(\hat{\beta} - \beta_0)$ is asymptotically zero-mean normal, with covariance matrix $\Sigma_\beta = E[c_1^{\otimes 2}]$, which can be consistently estimated by*

$$\hat{\Sigma}_\beta = n^{-1}\sum_{i=1}^n \hat{c}_i^{\otimes 2},$$

where c_1 is given in the proof of this theorem.

Proof of Theorem 4.2.

Notice that

$$U(\theta;\tilde{N}) = \sum_{i=1}^n \int_0^\tau W(t)\{\mathbf{X}_i(t) - \bar{\mathbf{X}}(t;\theta)\}d\tilde{M}_i(t;\theta),$$

where

$$\tilde{M}_i(t;\theta) = \tilde{N}_i(t) - \int_0^t \xi_i(s)\exp(\mathbf{X}_i'(t)\theta)d\phi_0(s).$$

Similar to the arguments of Lin and Wei (1989), we can show that

$$n^{-1/2}U(\theta_0;\tilde{N}) = n^{-1/2}\sum_{i=1}^n \int_0^\tau W(t)\{\mathbf{X}_i(t) - \bar{\mathbf{x}}(t;\theta_0)\}d\tilde{M}_i(t;\theta_0) + o_p(1).$$

By the Taylor expansion,

$$\begin{aligned}
& n^{1/2}(\hat{\theta} - \theta_0) \\
&= \left[-n^{-1}\partial U(\theta;\tilde{N})/\partial\theta\Big|_{\theta=\theta_0}\right]^{-1}\left[n^{-1/2}U(\theta_0;\tilde{N})\right] + o_p(1) \\
&= A_\theta(\theta_0)^{-1}n^{-1/2}\sum_{i=1}^n \int_0^\tau W(t)\{\mathbf{X}_i(t) - \bar{\mathbf{x}}(t;\theta_0)\}d\tilde{M}_i(t;\theta_0) + o_p(1) \\
&\equiv n^{-1/2}\sum_{i=1}^n a_i + o_p(1).
\end{aligned}$$

By (A.5) of Lin et al. (2000),

$$\begin{aligned}
n^{1/2}(\hat{\gamma} - \gamma_0) &= A_\gamma(\gamma_0)^{-1}n^{-1/2}\sum_{i=1}^n \int_0^\tau \{\mathbf{X}_i(t) - \bar{\mathbf{x}}(t;\gamma_0)\}dM_i(t;\gamma_0) + o_p(1) \\
&\equiv n^{-1/2}\sum_{i=1}^n b_i + o_p(1),
\end{aligned}$$

where $A_\gamma(\gamma_0)$ is given in (C.5) and

$$M_i(t;\gamma) = \tilde{H}_i(t) - \int_0^t \xi_i(t)\exp\{\mathbf{X}_i'(t)\gamma\}d\Lambda_0(t).$$

Thus,

$$n^{1/2}(\hat{\beta} - \beta_0) = n^{-1/2}\sum_{i=1}^n c_i + o_p(1).$$

where $c_i = a_i - b_i, i = 1, \cdots, n$. Then, by the multivariate central limit theorem, we conclude that $n^{1/2}(\hat{\beta} - \beta_0)$ is asymptotically zero-mean normal with covariance matrix $\Sigma_\beta = E[c_1^{\otimes 2}]$.

Next, we'll verify that Σ_β can be consistently estimated by $\hat{\Sigma}_\beta$ as defined in Theorem 4.2.

By the uniform strong law of large numbers (Pollard, 1990, p. 4), $n^{-1} \sum_{i=1}^{n} \tilde{N}_i(t)$ converges almost surely to $E\{\tilde{N}_1(t)\}$ uniformly in t and $S^{(0)}(t;\theta)$ converges almost surely to $s^{(0)}(t;\theta)$ uniformly in t and θ. This entails the uniform convergence of

$$\hat{\phi}_0(t;\theta) = \int_0^t \frac{\sum_{i=1}^{n} d\tilde{N}_i(s)}{nS^{(0)}(s;\theta)} \quad \text{to} \quad \int_0^t \frac{s^{(0)}(s;\theta_0)}{s^{(0)}(s;\theta)} d\phi_0(s),$$

under models (4.1) and (4.2). The derivative of $\hat{\phi}_0(t;\theta)$ with respect to θ is uniformly bounded in t for all large n and θ in a bounded region. Therefore, the strong consistency of $\hat{\theta}$ implies that $\hat{\phi}_0(t) = \hat{\phi}_0(t;\hat{\theta})$ converges almost surely to $\phi_0(t)$ uniformly in t.

Since we have shown that $\hat{A}_\theta(\theta_0)$ converges almost surely to $A_\theta(\theta_0)$, then by the strong consistency of $\hat{\theta}$ and the continuity of $\hat{A}_\theta(\cdot)$ with respect to θ, we can obtain the almost surely convergence of $\hat{A}_\theta(\hat{\theta})$ to $A(\theta_0)$. Then, $\hat{a}_i$ is the consistent estimator of a_i.

According to the argument in the A.3 of Lin et al. (2000), we can see that $\hat{b}_i$ is a consistent estimator for b_i, and thus $\hat{c}_i$ is consistent, which ensures the consistency of $\hat{\Sigma}_\beta$. This completes the proof.

4.5 Simulation Study

We conducted Monte Carlo simulation studies to evaluate the finite-sample properties of the proposed estimators. To generate the simulated data, we first generated z_i from the gamma distribution with mean 1 and variance σ^2, and let $g(z_i) = z_i^\alpha$. We assume that the time-dependent covariate $x_i(t)$ takes the form $u_i \log(t)$, where u_i has a uniform distribution over $[0, 0.5]$, and the follow-up times C_i's were generated from the uniform distribution over $(\tau/2, \tau)$ with $\tau = 18$. Here the symbol of α characterizes the relationship between the observation process and the recurrent event process. When $\alpha > 0$, a subject with more frequent observations would have a higher occurrence rate of the recurrent event and the two processes are positively correlated; when $\alpha = 0$, the two processes have no correlation given the covariates; when $\alpha < 0$, a subject with more frequent observations would have a lower occurrence rate of the recurrent event and the two processes are negatively correlated.

For observation process, we assume that H_i is a homogeneous Poisson process with $\lambda_0(t) = 1$. Then, given $x_i, C_i,\ z_i,\ K_i^* = \xi_i(C_i)H_i(C_i)$, the total number of real observation times for subjects i, follows the Poisson distribution with mean

$$\Lambda_0(C_i \mid x_i, z_i) = \int_0^{C_i} z_i \exp\{x_i(t)\gamma_0\}\lambda_0(t)dt = z_i \frac{C_i^{u_i\gamma_0+1}}{u_i\gamma_0 + 1},$$

$i = 1, \cdots, n$. In this case, the observation times $(T_{i1}, \cdots, T_{iK_i^*})$ are the order statistics of a random sample of size K_i^* from the uniform distribution over $(0, C_i)$. Finally, given K_i^* and $(T_{i1}, \cdots, T_{iK_i^*})$, we generate $N_i(T_{ij})'s$ by taking

$N_i(T_{ij}) = N_i(T_{i1}) + \{N_i(T_{i2}) - N_i(T_{i1})\} + \cdots + \{N_i(T_{ij}) - N_i(T_{ij-1})\}$, where

$$N_i(t) - N_i(s) \sim \text{Poisson}(0.5t^2 g(z_i) \exp\{x_i(t)\beta_0\} - 0.5s^2 g(z_i) \exp\{x_i(s)\beta_0\}),$$

$j = 1, \cdots, K_i^*, i = 1, \cdots, n.$

Set $\gamma_0 = 1$ and $\beta_0 = -1, 0, 1$, representing the different effect of the covariate $x(t)$ on the panel counts. On one hand, in order to check the effect of the estimators with time-independent covariates, we performed Monte Carlo studies when the time-independent covariate x_i follows a Bernoulli distribution with success probability 0.5. On the other hand, we also considered the situation that the observation process H_i follows a nonhomogeneous Poisson process with $\lambda_0(t) = (t+1)/(\tau/2+1)$ to verify that whether the different forms of the observation process H_i will affect the estimation of β or not. For each setting, we consider the sample size $n = 100$. All the results reported here are based on 500 Monte Carlo replications using R software.

Tables 4.1 presents the simulation results on estimation of β with time-independent and time-dependent covariates respectively under the homogeneous poisson observation process with $n = 100$, while Table 4.2 presents those under the nonhomogeneous poisson observation process. The tables include the bias (Bias) given by the sample means of the point estimates $\hat{\beta}$ minus the true values, the sample standard deviations of the estimates (SSD), the means of the estimated standard deviations (ESD), and the empirical 95% coverage probabilities (CP) for β. These results indicate that the estimate $\hat{\beta}$ seems to be unbiased and the proposed variance estimation procedure provides reasonable estimates. Also the results on the empirical coverage probabilities indicate that the normal approximation seems to be appropriate.

In addition, one can see from Tables 4.1 and 4.2 that the biases of the estimators of β, the SSD and ESD of the estimators of β with time-independent covariates are smaller than those with time-dependent covariates, which means that estimators with time-independent covariates are more precise and more stable than those with time-dependent covariates since there are more nondeterminacy with the time-varying covariates. Furthermore, one can see that the effect of the estimators with time-dependent covariates worsens rapidly as the variance of the frailty increases as discussed in Lin et al. (2000).

Table 4.3 shows the results of the estimators of β under the homogeneous and nonhomogeneous poisson observation process respectively with $n = 200$ and time-independent covariates. Compared with the corresponding results in Tables 4.1 and 4.2, we can see that the SSD and ESD of the estimators decreases when the sample size increases. As shown in Tables 4.1 and 4.2, the variance seems underestimated; a possible reason is that the simulated data were generated from the joint model including random effects, and the estimating equation only involves the means of random effects. The results in Table 4.3 indicate that this does not seem to be a problem for large sample size.

Table 4.1. Estimation of β with time-independent and time-dependent covariates respectively and $n = 100$ under the homogeneous Poisson observation process

$\alpha = -0.5$: H and N are negatively correlated						
β_0	1	0	-1	1	0	-1
	Time-indep covariates			Time-dep covariates		
$\text{Bias}_{\hat{\beta}}$	0.0022	-0.0040	0.0021	-0.0339	-0.0222	-0.0293
SSD	0.0746	0.0775	0.1031	0.1226	0.1079	0.1262
ESD	0.0738	0.0744	0.0963	0.1148	0.1044	0.1247
CP	0.9380	0.9300	0.9260	0.9100	0.9440	0.9480
$\alpha = 0$: H and N have no correlation						
β_0	1	0	-1	1	0	-1
	Time-indep covariates			Time-dep covariates		
$\text{Bias}_{\hat{\beta}}$	0.0002	0.0048	0.0016	-0.0292	-0.0239	-0.0207
SSD	0.0614	0.0665	0.0714	0.1117	0.0843	0.1103
ESD	0.0605	0.0622	0.0668	0.1008	0.0801	0.1012
CP	0.9380	0.9160	0.9220	0.9000	0.9280	0.9360
$\alpha = 0.5$: H and N are positively correlated						
β_0	1	0	-1	1	0	-1
	Time-indep covariates			Time-dep covariates		
$\text{Bias}_{\hat{\beta}}$	-0.0039	0.0035	-0.0051	-0.0249	-0.0321	- 0.0239
SSD	0.1006	0.0991	0.0800	0.1791	0.1354	0.1473
ESD	0.0927	0.0932	0.0793	0.1649	0.1234	0.1379
CP	0.9280	0.9280	0.9340	0.9160	0.9100	0.9360

Table 4.2. Estimation of β with time-independent and time-dependent covariates respectively and $n = 100$ under the nonhomogeneous Poisson observation process

$\alpha = -0.5$: H and N are negatively correlated						
β_0	1	0	-1	1	0	-1
	Time-indep covariates			Time-dep covariates		
$\text{Bias}_{\hat{\beta}}$	0.0036	0.0004	0.0062	-0.0305	-0.0411	-0.0262
SSD	0.0811	0.0798	0.0844	0.1232	0.1253	0.1395
ESD	0.0777	0.0786	0.0833	0.1202	0.1176	0.1324
CP	0.9500	0.9260	0.9240	0.9240	0.9200	0.9320
$\alpha = 0$: H and N have no correlation						
β_0	1	0	-1	1	0	-1
	time-indep covariates			time-dep covariates		
$\text{Bias}_{\hat{\beta}}$	0.0060	0.0025	0.0049	-0.0324	-0.0337	-0.0315
SSD	0.0669	0.0714	0.0746	0.1138	0.0905	0.1113
ESD	0.0638	0.0657	0.0706	0.1041	0.0814	0.1065
CP	0.9280	0.9240	0.9280	0.9160	0.9080	0.9120
$\alpha = 0.5$: H and N are positively correlated						
β_0	1	0	-1	1	0	-1
	Time-indep covariates			Time-dep covariates		
$\text{Bias}_{\hat{\beta}}$	0.0072	-0.0021	-0.0047	-0.0349	-0.0330	- 0.0290
SSD	0.1011	0.0995	0.1161	0.1847	0.1452	0.1611
ESD	0.0974	0.0953	0.1022	0.1642	0.1281	0.1412
CP	0.9380	0.9280	0.9160	0.8740	0.8900	0.9220

Table 4.3. Estimation of β under the homogeneous and nonhomogeneous Poisson observation process respectively with n=200 and time-independent covariates

$\alpha = -0.5$: H and N are negatively correlated						
β_0	1	0	-1	1	0	-1
	Homogeneous			Nonhomogeneous		
$\text{Bias}_{\hat{\beta}}$	0.0001	0.00391	0.0042	0.0080	0.0048	0.0028
SSD	0.0553	0.0527	0.0558	0.0555	0.0564	0.0585
ESD	0.0525	0.0534	0.0562	0.0554	0.0562	0.0591
CP	0.9460	0.9440	0.9420	0.9420	0.9500	0.9540
$\alpha = 0$: H and N have no correlation						
β_0	1	0	-1	1	0	-1
	Homogeneous			Nonhomogeneous		
$\text{Bias}_{\hat{\beta}}$	0.0009	-0.0004	0.0002	0.0043	0.0008	-0.0005
SSD	0.0449	0.0471	0.0502	0.0468	0.0483	0.0507
ESD	0.0436	0.0451	0.0486	0.0459	0.0472	0.0513
CP	0.9440	0.9360	0.9420	0.9380	0.9440	0.9440
$\alpha = 0.5$: H and N are positively correlated						
β_0	1	0	-1	1	0	-1
	Homogeneous			Nonhomogeneous		
$\text{Bias}_{\hat{\beta}}$	-0.0023	0.0078	-0.0007	0.0057	0.0049	0.0056
SSD	0.0703	0.0766	0.0755	0.0747	0.0744	0.0775
ESD	0.0667	0.0684	0.0711	0.0722	0.0722	0.0749
CP	0.9420	0.9120	0.9380	0.9320	0.9440	0.9420

4.6 An Application

This section presents an analysis of the bladder cancer data by applying our proposed methods. There were 121 subjects with superficial bladder tumors randomized into one of three treatment groups: placebo, thiotepa, and pyridoxine. In the following, we restrict our attention to the placebo and thiotepa groups with respective sizes of 47 and 38 as it has been shown that the pyridoxine treatment had no effect on the recurrence of the bladder tumors (Zhang, 2002). For each patient, the observed information includes times when he or she made clinical visits and the numbers of recurrent tumors between clinical visits. Two baseline covariates were observed and they are the number of initial tumors and the size of the largest initial tumor.

To analysis the data, for patient i, define x_{i1} to be equal to 1 if the ith patient was given the thiotepa treatment and 0 otherwise, x_{i2} to be the number of initial tumors and x_{i3} to be the size of the largest initial tumor, $i = 1, \cdots, 85$. Assume that the occurrence process of the bladder tumors and the clinical visit process can be described by joint models (4.1) and (4.2). Let $N_i(\cdot)$ represent the accumulated new tumor numbers of patient i over study period. We took the last visit time of the subject to approximate C_i in the analysis.

The application of the estimation procedure proposed in the previous sections gave $\hat{\gamma}_1 = 0.5071, \hat{\gamma}_2 = -0.0049, \hat{\gamma}_3 = 0.0321, \hat{\beta}_1 = -1.4905, \hat{\beta}_2 = 0.2867, \hat{\beta}_3 = -0.0821$ with the estimated standard errors being 0.1175, 0.0343, 0.0359, 0.3287, 0.0615 and 0.1056, which correspond to p-values of 1.5905e-05, 0.8864, 0.3712, 5.7732e-06, 3.1347e-06 and 0.4369, respectively based on the asymptotic results of the estimators. Here γ_1 and β_1, γ_2 and β_2, and γ_3 and β_3 represent regression coefficients corresponding to the treatment indicator, the number of initial tumors, and the size of the largest initial tumor, respectively. These results indicate that the thiotepa treatment significantly reduces the occurrence rate of the bladder tumors and the number of initial tumors has a significant positive effect on the tumor recurrence rate but no significant effect on the visit process. However, both the occurrence rate of the bladder tumors and the visit times do not seem to be significantly related to the size of the largest initial tumor. These conclusions are consistent with the analysis results presented in Sun and Wei (2000), Hu et al. (2003) and Zhao and Tong (2011). Furthermore, one can see that our proposed approach yields the smallest standard deviations except that the standard deviation of $\hat{\beta}_3$ is slightly higher than that of Zhao and Tong (2011), which suggests that our approach works well in applications.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

In Chapter 2, for the statistical analysis of longitudinal data, we have proposed a new semiparametric model for the situations where the observation times may be correlated with the response process even given the covariates, including Sun et al. (2005)'s conditional model as a special case. The new model allows for the interaction between the observation history and some components of the covariates and is different from Sun et al. (2007)'s and Liang et al. (2009)'s joint models through latent variables to characterize the correlation between the response process and the observation times. For inference about model parameters, a spline-based least square estimation procedure has been proposed. Another key difference between the approach developed here and those presented in Sun et al. (2005), Sun et al. (2007) and Liang et al. (2009) is that the patterns of the observation times are left arbitrary in our method, whereas their estimation procedures rely on the model specification for observation processes. As demonstrated in the simulation analysis, the proposed approaches are more flexible and robust.

Time-varying coefficient models with longitudinal data have been considered by many authors, such as Wu et al. (1998), Hoover et al. (1998), and Lin and Ying

(2001) among others. Motivated by these models, we can also extend our model to a class of conditional time-varying coefficient models as follows:

$$E\{Y_i(t)|\mathbf{X}_i, W_i, \mathcal{F}_{it}\} = \mu_0(t) + \beta(t)'\mathbf{X}_i + \alpha(t)'h(\mathcal{F}_{it}, W_i).$$

For inference about the above model, B-spline function approximations can be used to estimate the time-varying coefficients and the smooth baseline mean function simultaneously, and then the asymptotic properties of spline-based estimators could be established by using the similar arguments.

Chapter 3 considered a marginal conditional model for the underlying recurrent event process of the panel count data which allows for the interaction between the informative observation times and covariates, leaving the distributional form of the observation process to be arbitrary and proposed to use the easy implemented B-splines based method to estimate the regression parameters and the unknown smooth monotone function in the model simultaneously. As demonstrated by simulation and application that our proposed model and procedure are more flexible, robust and applicative since they can overcome the under-dispersion or over-dispersion problem resulting from the model specification for the observation process. We established the asymptotical results including consistency, rate of convergence for the estimators of the regression parameters and the unknown monotone function and asymptotic normality for the estimators of the regression parameters in Section 3.4. However, the asymptotic normality for the unknown function has not been obtained, which may be reserved as a problem to be solved in the future.

In some longitudinal studies, informative observation times and a dependent terminal event such as death that stops the follow-up may simultaneously exist. For example, in the bladder cancer study we have mentioned in Section 1.1.1, further observation of a patient during a particular clinic visit would be terminated probably because of his/her clinically significant improvement in the disease symptoms. If a patient who is very prone to superficial bladder tumors will visit the doctors more often to install the treatment (thiotepa) in the bladder, thus he/she would take longer than usual time to termination. Motivated by this fact, it is desirable to investigate the analysis of panel count data with informative observation times and dependent termination such as Liu et al. (2008) wherein a joint random effects model of longitudinal data with informative observation times and a dependent terminal event was considered.

Motivated by Li et al. (2010), our proposed models can also be extended to a class of transformation models as follows,

$$E\{N_i(t)|\mathbf{X}_i, W_i, \mathcal{F}_{it}\} = g\{\mu_0(t) + \beta'\mathbf{X}_i + \alpha' h(\mathcal{F}_{it}, W_i)\},$$

with a given monotone smooth function g. Then for inference of the models, the same algorithm as in Chapter 3 can be used to obtain the estimators for the regression parameters β and α and B-splines approximation with monotone nondecreasing estimated coefficients for the nonparametric monotone function $\mu_0(t)$, and the asymptotic properties of the spline-based estimators could be established by using the similar arguments.

In Chapter 4, we have generalized Zhao and Tong (2011)'s joint modeling approach for the analysis of panel count data to the situations where the covariates are time-dependent and the observation and censoring times are informative. For estimation of the covariate effect on the underlying recurrent process, we have developed a novel estimating equation-based procedure, which depends on neither the form of the link function of the frailty nor the distribution of the frailty, and established the consistency and asymptotic normality of the resulting estimates.

By using the approach proposed by Huang et al. (2010), one can obtain the estimators of the parameter γ and $\Lambda_0(\cdot)$ in model (4.2), which are different from the approach proposed in Lin et al. (2000). Then, by replacing $\hat{\gamma}$ and $\hat{\lambda}_0(\cdot)$ with those given in Huang et al. (2010), one can get another estimator for β_0, which is different from our proposed estimator. Thus, it is desirable to compute the efficiency of these two different estimators.

In practice, it is important to predict the mean of panel counts. However, it is hard to estimate the baseline mean function $\mu_0(t)$ in the current setting. Further research is needed to address this issue.

Just as Zhao and Tong (2011) mentioned, the time-dependent frailty, the non-poisson observation process are also important issues to be studied.

5.2 Further Research

5.2.1 Proportional partial linear intensity model for recurrent event data

During some relatively long follow-up studies, each individual may experience the same event repeatedly. The events are called recurrent events in survival analysis. One main difference between recurrent event data and panel count data is that the former posseses a observation process in the whole follow-up, while the later involves a sequence of consecutive observation. It is natural and convenient to represent the recurrent event times as a counting process. The most popular counting-process model is the proportional intensity model studied by Andersen and Gill (1982). Let $N^*(t)$ denote the number of events that the subjects has experienced by time t, and let $\mathbf{X}(t)$ be a vector of possibly time-dependent covariates. The proportional intensity model specifies that the intensity function for $N^*(t)$ associated with $\mathbf{X}$ takes the form

$$\lambda(t|\mathbf{X}) = \lambda_0(t) \exp\{\beta^T \mathbf{X}(t)\}, \tag{5.1}$$

where $\lambda_0(t)$ is an unspecified baseline intensity function and β is a vector of unknown regression parameters.

Much research had been studied based on this model, where the covariate effects on the logarithm of the hazard function are assumed to be linear. However, true covariate effects may be more complex than the log-linear effect and studying nonlinear effects is a challenging problem. Huang (1999) considered a partly linear additive Cox model with right-censored data and proposed the maximum partial likelihood estimators by using polynomial splines to approximate the nonparametric component. Fan et al. (2006) extended the proportional hazards model by adding a nonlinear term in the logarithm of the hazard function for lifetime data and proposed a local partial-likelihood technique to estimate the nonlinear term and also established its asymptotic properties. Cai et al. (2007) put forward a partially linear hazard regression model for multivariate survival data and proposed a profile pseudo-partial likelihood estimation method under the marginal hazard model framework. In addition, the additive hazards model and the accelerated failure time model had been extended by researchers through adding nonlinear covariate terms, such as Yin et al. (2008), Lu and Zhang (2010), among others. To our knowledge, no partial linear covariate effect on the logarithm of intensity function of the recurrent counting process have been considered to handle recurrent event data. Thus, we propose a proportional partial linear intensity model as follows:

$$\lambda(t|\mathbf{X}, Z) = \lambda_0(t) \exp\{\beta^T \mathbf{X}(t) + g(Z(t))\}, \tag{5.2}$$

where g is an unknown smooth function with $g(0) = 0$, Z is an univariate covariate whose effects on the logarithm of the intensity function is non-linear.

Recurrent event times are commonly subject to right censoring. Let C denote the censoring time. We assume that there are n subjects, and the data consist of $\{X_i(\cdot), Z_i(\cdot), N_i(\cdot), Y_i(\cdot)\}, i = 1, \cdots, n$, where $N_i(t) = N_i^*(t \wedge C_i), Y_i(t) = I(C_i \geq t)$, and $I(\cdot)$ is the indicator function. $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$ is the baseline cumulative intensity function. Let τ denote the terminal time of the study, we assume that the conditional probability of $C > t$ given $\{\mathbf{X}(s), Z(s), N^*(s), s \in [0, \tau]\}$ is noninformative about (λ_0, β, g). In addition, we assume that the conditional distribution of $\{\mathbf{X}(t), Z(t)\}$ given $\{\mathbf{X}(s), Z(s), N(s), Y(s); s \in [0, \tau]\}$ is noninformative about (λ_0, β, g).

Assume that Z takes values in $[a, b]$, where a and b are finite numbers . Let

$\mathcal{I} = \{\xi_i\}_1^{m_n+2l}$, with

$$a = \xi_1 = \cdots = \xi_l < \xi_{l+1} < \cdots < \xi_{m_n+l} < \xi_{m_n+l+1} = \cdots = \xi_{m_n+2l} = b,$$

be a sequence of knots that partition $[a, b]$ into m_n+1 subintervals $J_i = [\xi_{l+i}, \xi_{l+i+1}]$, $i = 0, \cdots, m_n$. Denote by $\Psi_{l,\mathcal{I}}$ the class of polynomial splines of order $l \geq 1$ with the knot sequence $\mathcal{I}$. For each $s \in \Psi_{l,\mathcal{I}}$, s is a polynomial of order l in J_i for $0 \leq i \leq m_n$, and s is l' times continuously differentiable on $[a, b]$, for $l \geq 2$, and $0 \leq l' \leq l-2$. (Schumaker, 1981, p.108, Def 4.1).

For any $g \in \Psi_{l,\mathcal{I}}$, there exist $\alpha_1, \cdots, \alpha_{q_n}$ such that

$$g(z) = \sum_{i=1}^{q_n} \alpha_i B_{il}(z),$$

where $\{B_{il}, 1 \leq i \leq q_n\}$ with $q_n = m_n + l$ is the B-spline basis functions of $\Psi_{l,\mathcal{I}}$. (Schumaker, 1981, P.117 Corollary 4.10).

Thus, replacing $g(\cdot)$ by its B-spline approximation in the model (5.2), we have

$$\lambda(t|\mathbf{X}, Z) = \lambda_0(t) \exp\{\beta^T \mathbf{X}(t) + \alpha_n^T \tilde{B}_n(Z(t))\},$$

where $\alpha_n = (\alpha_1, \cdots, \alpha_{q_n})^T$, and $\tilde{B}_n(z) = (B_{1l}(z), \cdots, B_{q_n l}(z))^T$.

Then, the estimates of the parameters (β, α_n) are obtained by maximizing the following log-partial likelihood:

$$l(\beta, \alpha_n) = \sum_{i=1}^{n} \int_0^{\tau} \{\beta^T X_i(t) + \alpha_n^T \tilde{B}_n(Z_i(t)) - \log[n S^{(0)}(\beta, \alpha_n)]\} dN_i(t) \qquad (5.3)$$

where $S^{(0)}(\beta, \alpha_n, t) = n^{-1} \sum_{i=1}^{n} Y_i(t) \exp\{\beta^T X_i(t) + \alpha_n^T \tilde{B}_n(Z_i(t))\}$ and $dN_i(t)$ denotes the numbers of events in a small time interval $[t, t + dt)$.

Let $\hat{\alpha}_n = (\hat{\alpha}_1, \hat{\alpha}_2, \cdots, \hat{\alpha}_{q_n})^T$ and $\hat{\beta}_n$ be the estimators. Then we denote the spline estimator of $g(\cdot)$ by $\hat{g}_n(\cdot) = \sum_{i=1}^{q_n} \hat{\alpha}_i B_{il}(\cdot) = \hat{\alpha}_n^T \tilde{B}_n(\cdot)$. Our main purpose is to find out the consistency and convergency of the estimators $\hat{\beta}_n$ and $\hat{g}_n(\cdot)$, which may be verified by using the empirical process theory in Van der Vaart and Wellner (1996), Huang (1999), Lu et al. (2007) and Lu et al. (2009).

Remark 5.1. ***Proof of (5.3)***: *According to Cook and Lawless (2007), p.77, (3.25), the partial likelihood is*

$$\begin{aligned}
L(\beta, \alpha_n) &= \prod_{i=1}^{n} \prod_{j=1}^{n_i} \frac{\exp\left\{[\beta^T X_i(T_{ij}) + \alpha_n^T \tilde{B}_n(Z_i(T_{ij}))]\right\}}{\sum_{k=1}^{n} Y_k(T_{ij}) \exp\{\beta^T X_k(T_{ij}) + \alpha_n^T \tilde{B}_n(Z_k(T_{ij}))\}} \\
&= \prod_{i=1}^{n} \prod_{j=1}^{n_i} \exp\left\{\left[\beta^T X_i(T_{ij}) + \alpha_n^T \tilde{B}_n(Z_i(T_{ij}))\right] - \log[nS^{(0)}(\beta, \alpha_n, T_{ij})]\right\} \\
&= \prod_{i=1}^{n} \exp\left[\int_0^{\tau} \{\beta^T X_i(t) + \alpha_n^T \tilde{B}_n(Z_i(t)) - \log[nS^{(0)}(\beta, \alpha_n, t)]\} dN_i(t)\right].
\end{aligned}$$

where $\{T_{ij}, j = 1, \cdots, n_i; i = 1, \cdots, n\}$ are the observed event times and n_i is the number of observed events on the ith subject.

5.2.2 New nonparametric tests for panel count data

In the analysis of panel count data, we assume that each subject in the study gives rise to a point process $N(t)$, denoting the total number of occurrences of the event of interest up to time t, and the data consist of independent samples of panel count data randomly drawn from $k (k \geq 2)$ populations or groups. $\Lambda_l(t) = E(N(t))$ is the mean function of $N(t)$ corresponding to the lth group for $l = 1, \cdots, k$. As noted in Section 1.2.2, many researchers have studied the testing problem on the hypothesis $H_0 : \Lambda_1(t) = \cdots = \Lambda_k(t)$, such as Thall and Lachin (1988), Sun and Fang (2003), Zhang (2006), Park et al. (2007), and Balakrishnan and Zhao

(2009, 2010, 2011). Among them, Sun and Fang (2003), Park et al. (2007), and Balakrishnan and Zhao (2011) proposed the tests based on the isotonic regression estimator of the mean functions (Sun and Kalbleisch, 1995; Wellner and Zhang, 2000), which will be recounted here for our use.

Suppose there are n independent subjects and n_l in the lth group with $n_1 + \cdots + n_k = n$. Let $N_i(t)$ denote the point process arising from subject i ($i = 1, \cdots, n$ and each subject be observed only at discrete time points $0 < T_{i,1} < \cdots < T_{i,K_i}$. Let $n_{i,j} = N_i(T_{i,j})$ be the observed value of N_i at $T_{i,j}, j = 1, \cdots, K_i, i = 1, \cdots, n$.

For simplicity, assume that H_0 is true, and let $\Lambda_0(t)$ denote the common mean function of $N_i(t)$'s. Further, let $s_1, \cdots, s_m$ denote the ordered distinct observation times in the set $\{T_{i,j} : j = 1, \cdots, K_i, i = 1, \cdots, n\}$ and ω_l and $\bar{n}_l$ be the number and mean value, respectively, of observations made at time $s_l, l = 1, \cdots, m$. Then the isotonic regression estimator, denoted by $\hat{\Lambda}_n(t)$, is defined as a nondecreasing step function with possible jumps at the s_l's, and is given by

$$\hat{\Lambda}_n(s_l) = \max_{r \leq l} \min_{s \geq l} \frac{\sum_{v=r}^{s} \omega_v \bar{n}_v}{\sum_{v=r}^{s} \omega_v} = \min_{s \geq l} \max_{r \leq l} \frac{\sum_{v=r}^{s} \omega_v \bar{n}_v}{\sum_{v=r}^{s} \omega_v}, \quad l = 1, \cdots, m,$$

the isotonic regression of the n_l's with weights ω_l's (Robertson et al., 1988).

Let $\hat{\Lambda}_{n_l}$ denote the isotonic regression estimate of Λ_l based on samples from all the subjects in the lth group. To test the hypothesis H_0, one of the two classes of test statistics given by Balakrishnan and Zhao (2011) is as follows:

$$V_n^{(l)} = \sqrt{n} \int_0^\tau W_n^{(l)}(t) \{\hat{\Lambda}_{n_1}(t) - \hat{\Lambda}_{n_l}(t)\} dG_n(t), \quad l = 2, \cdots, k, \tag{5.4}$$

where τ is the largest observation time, $W_n^{(l)}(t)$ are bounded weight processes, and

$$G_n(t) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i} I\{T_{i,j} \leq t\},$$

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

In contrast to the above hypothesis and the corresponding tests, Cook et al. (1996) mentioned that tests regarding the performance of certain drug combinations versus others are interesting in trials with multiple arms and multiple drug therapies. For example, if one treatment arm consists of a combination therapy of drugs given in other arms, one might plan to investigate if the treatments prove more beneficial in combination than individually. In such a situation, one might specify a hypothesis of the form

$$H_0 : \mathbf{L}^T\mathbf{\Lambda}(t) = 0, \qquad\qquad t > 0, \tag{5.5}$$

where $\mathbf{L} = (L_1, \cdots, L_k)^T$ is a fixed vector of coefficients forming the contrast, and $\mathbf{\Lambda}(t) = (\Lambda_1(t), \cdots, \Lambda_k(t))^T$.

For this hypothesis, we can construct more general statistics of the form

$$U_n = \sqrt{n}\int_0^\tau W_n(t)\mathbf{L}^T\hat{\mathbf{\Lambda}}(\mathbf{t})dG_n(t), \tag{5.6}$$

where $\hat{\mathbf{\Lambda}}(t) = (\hat{\Lambda}_{n_1}(t), \cdots, \hat{\Lambda}_{n_k}(t))^T$.

Since the above statistics are a generalization of the statistics

$$\sqrt{n}\int_0^\tau W_n(t)\{\hat{\Lambda}_{n_1}(t) - \hat{\Lambda}_{n_2}(t)\}dG_n(t),$$

given in Balakrishnan and Zhao (2011) for the special case $k = 2$, proposed for testing $H_0 : \Lambda_1(t) = \Lambda_2(t) \quad t > 0$, we could obtain the asymptotic distribution of U_n similar to $V_n^{(2)}$ in Balakrishnan and Zhao (2011).

Appendix A. B-Splines

B-splines, firstly introduced by de Boor (1978), are a popular type of polynominal splines in statistical applications, mainly because of their flexibility and numerical properties.

Define $\mathcal{P}_l$ as the space of polynomials of order l (degree $(l-1)$), then the basis functions span $\mathcal{P}_l$ are $\{1, t, t^2, \cdots, t^{l-1}\}$ and an element in $\mathcal{P}_l$ can be written as $p(t) = a_1 + a_2 t + \cdots + a_l t^{l-1} = \sum_{j=1}^{l} a_j t^{j-1}$.

For a finite closed interval $[a, b]$, let $\mathcal{I} = \{t_i\}_1^{m_n+2l}$, with

$$a = t_1 = \cdots = t_l < t_{l+1} < \cdots < t_{m_n+l} < t_{m_n+l+1} = \cdots = t_{m_n+2l} = b$$

be a sequence of knots that partition $[a, b]$ into m_n+1 subintervals $I_i = [t_{l+i}, t_{l+i+1})$, for $i = 0, 1, \cdots, m_n$. Denote by $\Psi_{l,\mathcal{I}}$ the class of polynominal splines of order $l \geq 1$ with the knot sequence $\mathcal{I}$, i. e.,

$$\Psi_{l,\mathcal{I}} = \{s \in C^{l-2}[a, b] \text{ for } l \geq 2 : s|_{I_i} \in \mathcal{P}_l, i = 0, 1, \cdots, m_n\},$$

where $C^{l-2}[a, b] = \{f :$ the $(l-2)$th derivative $f^{(l-2)}$ is continuous on $[a, b]\}$. A spline for $l = 4$ is a piecewise-cubic polynomial with continuous second-order derivative. As a special case, the spline with $l = 1$ is a step function which is discontinuous at each knot.

In fact, the class $\Psi_{l,\mathcal{I}}$ is linearly spanned by the B-spline basis functions $\{B_{il}, 1 \leq i \leq q_n\}$; that is, for any $s \in \Psi_{l,\mathcal{I}}$, there exist $c_1, \cdots, c_{q_n}$ such that $s(t) = \sum_{i=1}^{q_n} c_i B_{il}(t)$ (Schumaker, 1981), where $q_n = m_n + l$ is the number of basis functions. An recursive relation that is very useful in practice (Schumaker, 1981) can be summarized as follows:

Firstly,

$$B_{i1}(t) = \begin{cases} 1, & t \in [t_i, t_{i+1}) \\ 0, & \text{otherwise} \end{cases}$$

for $i = 0, 1, \cdots, m_n + 2l - 1$, then

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

for $i = 1, \cdots, m_n + 2l - l = q_n$. The fact that $B_{il}(t) > 0$ only when $t_i \le t < t_{i+q_n}$ and is zero otherwise is a very important property. Furthermore, another property, the variation-diminishing property (Schumaker, 1981, page 117) is

$$S^- \left(\sum_{i=1}^{q_n} c_i B_{il}(t) \right) \le S^-(c_1, \cdots, c_{q_n}), \qquad \text{any } c_1, \cdots, c_{q_n} \text{ not all } 0,$$

where $S^-(\mathbf{v})$ is the number of sign changes in the sequence $v_1, \cdots, v_n$ (zeros are ignored) with $\mathbf{v} = (v_1, \cdots, v_n)$. When the unknown function is nonnegative or monotone, this property is very practical when using the nonnegative or monotone B-splines estimator, since the B-splines approximation $\sum_{i=1}^{q_n} c_i B_{il}(t)$ possesses the same nonnegative and monotonicity as $c_i, i = 1, \cdots, q_n$, which can be obtained according to Examples 4.74, 4.75, 4.76 in Schumaker (1981).

Furthermore, a monotone I-splines are proposed by Ramsay (1988), which can be defined as

$$I_{il}(t) = \int_a^t B_{il}(s) ds,$$

then these I-splines have degree of l. For knots $\mathcal{I}$, the I-splines can be obtain in the more convenient forms

$$I_{il}(t) = \begin{cases} 0, & i > j, \\ \frac{1}{l+1} \sum_{m=i}^{j} (t_{m+l+1} - t_m) B_{m(l+1)}(t), & j - l + 1 \leq i \leq j, \\ 1, & i < j - l + 1. \end{cases}$$

for $t_j \leq t < t_{j+1}$.

Finally, here we'll summarize two useful lemmas used in the proof of consistency and rate of convergence in Chapters 2 and 3, which are the following Lemma 5 in Stone (1985) and Lemma A.2 in Huang (1999).

Lemma 5 in Stone (1985) *For each $h \in \mathcal{F}_r$, and $n \geq 1$, there is an $s \in \Psi_{l,\mathcal{I}}$ with*

$$\|s - h\|_\infty \leq M m_n^r,$$

here M is some fixed positive constant, m_n is the number of interior knots and $\mathcal{F}_r$ is defined in Section 2.

Lemma A.2 in Huang (1999) *For any $\eta > 0$, let*

$$\Theta_n = \{x'\beta + \phi : \|\beta - \beta_0\| \leq \eta, \phi \in \Psi_{l,\mathcal{I}}, \|\phi - \phi_0\|_\infty \leq \eta\}.$$

Then for any $\varepsilon \leq \eta$,

$$\log N_{[]}(\varepsilon, \Theta_n, L_2(P)) \leq c q_n \log(\eta/\varepsilon).$$

where $q_n = m_n + l$ is the number of spline basis functions and $N_{[]}(\varepsilon, \Theta_n, L_2(P))$ is the bracketing number we will introduce in Appendix C.

Appendix B. Generalized Gradient Projected Algorithm

Consider the problem (A1)

$$\text{Maximize } : l(\theta)$$

$$\text{Subject to } : \mathbf{a}_i'\theta = b_i, i \in I_1, \ \mathbf{a}_i'\theta \le b_i, i \in I_2,$$

where $l(\theta)$ is a sufficiently smooth objective function, $\mathbf{a}_i$ is a given $p \times 1$ vector, b_i is a given scalar, and equality constraints and inequality constraints are indexed by index sets I_1 and I_2 respectively.

A well-known gradient projection algorithm was firstly proposed by Rosen (1960) to optimizing such a nonlinear programming problem subject to linear constraints. Rosen's algorithm is based on the ordinary Euclidian metric. Jamshidian (2004) developed a general algorithm based on the generalized Euclidian metric $\|x\|_W = x'Wx$, where W is a positive matrix and can vary from iteration to iteration. Here we'll sketch Jamshidian's generalized gradient projection algorithm as follows.

A constraints is said to be active if it holds with equality. Let $\mathcal{A}$ be an initial working set of active constraints, that is,

$$\mathcal{A} = \{i \in I_1 \cup I_2 | \mathbf{a}_i'\theta = b_i\} \supseteq I_1,$$

and let A be an $m \times p$ working matrix whose rows consist of $\mathbf{a}'$ for all $i \in \mathcal{A}$, b denote the corresponding vector of b_i's. Rosen's gradient projection method is based on projecting the search direction into the subspace tangent to the active constraints. Active set method is a procedure that determines optimal active constraints by moving among several working sets of potential optimal active constraints (Fletcher, 1987). Jamshidian (2004) proposed a gradient projection - active set algorithm.

The generalized gradient of l in the metric $\|\cdot\|_W$ is given by $\tilde{g}(\theta) = W^{-1}g(\theta) \equiv W^{-1}\nabla l(\theta)$. Start from a feasible initial point $\theta_r \in \Omega \equiv \{\theta \in \mathbf{R}^p | A\theta = b\}$, then get a new point $\tilde{\theta}_r = \theta_r + d$, through a direction d. Then $\tilde{\theta}_r \in \Omega \Longleftrightarrow d \in \mathcal{N} \equiv \{d \in \mathcal{M} | Ad = 0\}$ with $\mathcal{M}$ be defined as the p dimensional Euclidean space with a norm defined by $\|x\|_W = x'Wx$ and $\mathcal{N}$ is called the space of feasible space. Gradient projection method generates a sequence of feasible points by moving along feasible directions that converges to a solution of (A1). The feasible direction at a point $\theta_r \in \Omega$ is obtained by projecting $\tilde{g}(\theta_r)$ onto $\mathcal{N}$ in the metric $\|\cdot\|_W$. Some reduction as in Jamshidian (2004, p.139-140) can result in that

$$\lambda = (AW^{-1}A')^{-1}A\tilde{g}(\theta_r),$$

and

$$d = I - W^{-1}A'(AW^{-1}A')^{-1}A\tilde{g}(\theta_r),$$

where I is the identity matrix. And it can verify that d is a generalized gradient of l in $\mathcal{N}$ in the metric of $\|\cdot\|_W$ since

$$\begin{aligned}
\nabla l(\nabla\theta) &= \langle \nabla\theta, g(\theta) \rangle \\
&= \langle \nabla\theta, \tilde{g}(\theta) \rangle_W \\
&= \langle P_W \nabla\theta, \tilde{g}(\theta) \rangle_W \\
&= \langle \nabla\theta, P_W \tilde{g}(\theta) \rangle_W,
\end{aligned}$$

where $P_W = I - W^{-1}A'(AW^{-1}A')^{-1}A$ is idempotent and self-adjoint and a projection onto $\mathcal{N}$ in the metric of $\|\cdot\|_W$. Then if $d \neq 0$, a small enough step from θ_r in the direction of d results in a new feasible point $\tilde{\theta}_r$ such that $l(\tilde{\theta}_r) > l(\theta_r)$. In fact, d is a steepest ascent direction with respect to l since

$$\begin{aligned}
\langle d, g\rangle &= \langle d, \tilde{g}\rangle_W \\
&= \langle d, \tilde{g} - d + d\rangle_W \\
&= \langle d, \tilde{g} - d\rangle_W + \langle d, d\rangle_W \\
&> \langle P_W\tilde{g}, (I - P_W)\tilde{g}\rangle_W \\
&> \langle P_W(I - P_W)\tilde{g}, \tilde{g}\rangle_W \\
&> 0,
\end{aligned}$$

where $\langle u, v\rangle_W = u'Wv$. Then the largest step length α_1 is obtained by $\arg\max_\alpha\{\alpha|\theta_r + \alpha d \in \Omega\}$, and a new point $\tilde{\theta}_r$ is obtained by performing the line search $\arg\max_{0<\alpha\leq\alpha_1}\{l(\theta_r+\alpha d)\}$, and then add indexes of newly active constraints, if any, to the working set $\mathcal{A}$, and A and Ω are redefined accordingly. If $d = 0$, and all components of λ are nonnegative, then $\tilde{\theta}_r$ satisfies the first order necessary Karush-Kuhn-Tucker Conditions (Luenberger, 1984, Chap. 2) which can be stated as the existence of a vector λ such that

(1) $\lambda \geq 0$;

(2) $\lambda W^{-1}(A\theta_r - b) = 0$;

(3) $\tilde{g}(\theta_r) - W^{-1}A'\lambda = 0$

for being a constrained optimum. On the other hand, if $d = 0$ and at least one component of λ is negative, then drop a constraint corresponding to this negative λ_i from the working set, and calculate a new nonzero direction d which leads to a new feasible improved point. The detailed gradient projection-active set algorithm are summarized in Algorithm 1,

---

Algorithm 1 The Gradient Projection - Active Set Algorithm

Start with an initial point θ, that satisfies $A\theta = b$, and cycle through the following steps until convergence:

S1: Compute
$$\underline{d} = \left(I - W^{-1}A^T(AW^{-1}A^T)^{-1}A\right)W^{-1}\nabla l(\theta),$$
when there is no active constraint, take $\underline{d} = W^{-1}\nabla l(\theta)$.

S2: If $\underline{d} = 0$, compute the Lagrange multiplier $\lambda = (AW^{-1}A')^{-1}A\tilde{g}(\theta_r)$. Let λ_i denote the ith component of λ.

a . If $\lambda_i \geq 0$ for all $i \in \mathcal{A} \cap I_2$, Stop. The current point satisfies the Karush-Kuhn-Tucker Conditions.

b . If there is at least one negative λ_i for $i \in \mathcal{A} \cap I_2$, determine the index corresponding to the smallest such λ_i, and delete this index from $\mathcal{A}$. Modify A and b, by dropping a row from each accordingly. and go to S1.

S3: If $\underline{d} \neq 0$, obtain $\alpha_1 = arg\max_{\alpha}\{\alpha|\theta_r + \alpha\underline{d} \in \Omega\}$. Then search for $\alpha_2 = arg\max_{0<\alpha\leq\alpha_1}\{l(\theta_r+\alpha\underline{d})\}$. Set $\tilde{\theta} = \theta+\alpha_2\underline{d}$. Add indexes of new coordinates, if any, of $\tilde{\theta}$ that are newly on the boundary to the working set $\mathcal{A}$. Modify A and b, by adding new rows, accordingly.

S4: Replace θ, by $\tilde{\theta}$ and go to S1.

As said in Jamshidian (2004), the sufficient condition for $\hat{\theta}$ to be a local maximum of $l(\theta)$ in Ω is that $H(\hat{\theta})$, the Hessian of $l(\theta)$ at $\hat{\theta}$, be negative definite on $\mathcal{N}$. Theoretically, gradient projection algorithm converges from almost any arbitrary feasible point and for any positive definite W. The choice of W, however, is important because the local rate of convergence of gradient projection algorithm depends on the ratio of the smallest to the largest eigenvalue of the Hessian of $\hat{\theta}$

in the metric of $\|\cdot\|_W$, i.e. $W^{-1}H(\hat{\theta})$, restricted to $\mathcal{N}$. Accurately, the closer the ratio to one, the faster the rate of convergence.

Appendix C. Empirical Process

Empirical process technique has become an increasingly important tool for statistical inference in semiparametric or nonparametric models, which is also our main theoretical background for inference of the asymptotic properties of our estimators. Thus we will sketch some commonly used conclusions here.

Glivenko-Cantelli and Donsker Classes

An empirical process is a stochastic process based on a random sample. Consider a random sample $X_1, \cdots, X_n$ from a probability measure P on an arbitrary measure space $(\mathcal{X}, \mathcal{A})$. The empirical measure is defined as $\mathbb{P}_n = n^{-1}\sum_{i=1}^n \delta_{X_i}$, where δ_x is the measure which assigns mass 1 at x and zero elsewhere. Denote $\mathbb{P}_n f = n^{-1}\sum_{i=1}^n f(X_i)$ and $Pf = \int f dP$, for a measurable function $f : \mathcal{X} \longrightarrow \mathbb{R}$. Then an empirical process

$$\mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n f - Pf) = \frac{1}{\sqrt{n}}\sum_{i=1}^n [f(X_i) - E_P f(X_i)]$$

for any class $\mathcal{F}$ of measurable functions $f : \mathcal{X} \longrightarrow \mathbb{R}$. The envelope function $F : \mathcal{X} \longrightarrow \mathbb{R}$ of the class $\mathcal{F}$ is the function such that $|f(x)| \leq F(x) < \infty$ for every $x \in \mathcal{X}$ and $f \in \mathcal{F}$.

By the law of large numbers and the central limit theorem, for each $f \in \mathcal{F}$,

$$\mathbb{P}_n f \xrightarrow{a.s.} Pf \text{ and } \mathbb{G}_n f \xrightarrow{d} N(0, P(f - Pf)^2).$$

provided Pf exists and $Pf^2 < \infty$, respectively, where $\xrightarrow{d}$ means converge in distribution.

When investigating the properties, the uniform convergence and asymptotic normality are more desirable, which can be defined as follows. A class $\mathcal{F}$ of measurable functions $f : \mathcal{X} \longrightarrow \mathbb{R}$ is said to be a P-Glivenko-Cantelli class, if

$$\sup_{f \in \mathcal{F}} |\mathbb{P}_n f - Pf| \xrightarrow{a.s.} 0.$$

And $\mathcal{F}$ is said to be a P-Donsker class, if the process $\{\mathbb{G}_n f : f \in \mathcal{F}\}$ converges in distribution to a tight limit processes in $l^\infty(\mathcal{F})$, which is the space of bounded functionals on $\mathcal{F}$ under the supermum norm $\|f\| = \sup_{h \in \mathcal{F}} |f(h)|$.

Whether a class of functions is a Glivenko-Cantelli or a Donsker class (hereafter, we'll drop the P if the context is clear) is mainly determined by the "size" of the class. A relatively simple way to measure the size of a class is in terms of entropy including entropy with bracketing and entropy with covering. We will mainly introduce the entropy with the $L_r(P)$-norm, $\|f\|_{r,P} = (\int |f|^r dP)^{1/r}$.

We need to introduce the ϵ-bracket in $L_r(P)$ firstly. A pair of functions $\{l, u\} \in L_r(P)$ is an ϵ-bracket if they are satisfying $P(l(X) \leq u(X)) = 1$ and $\|l - u\|_{r,P} \leq \epsilon$. A function $f \in \mathcal{F}$ lies in the bracket $\{l, u\}$ if $P(l(X) \leq f(X) \leq u(X)) = 1$. Then the bracketing number $N_{[]}(\epsilon, \mathcal{F}, L_r(P))$ be defined as the minimum number of ϵ-brackets in $L_r(P)$ needed to cover $\mathcal{F}$. The logarithm of the bracketing number is the entropy with bracketing. Here it is required that both l and u are of finite norm in terms of $\|\cdot\|_{r,P}$ but need not necessarily belong to $\mathcal{F}$. And the covering number $N(\epsilon, \mathcal{F}, L_r(P))$ is the minimum number of $L_r(P)$ ϵ-balls needed to cover $\mathcal{F}$, where an $L_r(P)$ ϵ-ball around a function $g \in L_r(P)$ is the set $\{h \in L_r(P) : \|h-g\|_{r,P} < \epsilon\}$. Similarly, the centers of the balls to cover $\mathcal{F}$ are not necessary to belong to $\mathcal{F}$. The entropy (without bracketing) is the logarithm of the covering number.

Remark C.1 *If f is in the ϵ-bracket $\{l, u\}$, then it is in the $\epsilon/2$-ball round midpoint $(l+u)/2$. Thus it follows that*

$$N(\epsilon/2, \mathcal{F}, L_r(P)) \leq N_{[]}(\epsilon, \mathcal{F}, L_r(P)).$$

Glivenko-Cantelli and Donsker Theorems

In the following, we'll sketch two important theorems in modern empirical process.

Firstly, the simplest Glivenko-Cantelli theorem based on entropy with bracketing is given in Van der Vaart and Wellner (1996, Th 2.4.1), which is presented as follows,

Theorem C.2 (Glivenko-Cantelli Theorem) *Let $\mathcal{F}$ be a class of measurable functions such that $N_{[]}(\epsilon, \mathcal{F}, L_1(P)) < \infty$ for every $\epsilon > 0$. Then $\mathcal{F}$ is Glivenko-Cantelli.*

Donsker theorems based on entropy with bracketing require more stringent conditions on the number of brackets needed to cover $\mathcal{F}$. For most classes of interest, the entropy goes to infinity as $\epsilon \downarrow 0$. The sufficient condition for a class to be a Donsker is that the bracketing integral

$$J_{[]}(\delta, \mathcal{F}, L_r(P)) \equiv \int_0^\delta \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, L_r(P))} d\epsilon$$

needs to be bounded for $r = 2$ and $\delta = \infty$, which can be derived from Van der Vaart and Wellner (1996, Th 2.5.2), we'll summarize it here.

Theorem C.3 (Donsker Theorem) *Let $\mathcal{F}$ be a class of measurable functions with $J_{[]}(\infty, \mathcal{F}, L_2(P)) < \infty$. Then $\mathcal{F}$ is Donsker.*

The following theorem indicate that the class of uniformly bounded, monotone functions on the real line is Donsker, which is the Van der Vaart and Wellner (1996, Th 2.7.5)

Class of bounded monotone functions *The $\mathcal{F}$ of monotone functions $f : \mathbb{R} \longrightarrow [0,1]$ satisfies*

$$\log N_{[]}(\epsilon, \mathcal{F}, L_r(P)) \leq K(\frac{1}{\epsilon}),$$

for every probability measure P, every $r \geq 1$, and a constant K that depends on r only.

M Estimators : Rate of Convergence

A M-estimator $\hat{\theta}_n$ is the approximate maximum of a data-dependent function $\theta \longmapsto \mathbb{M}_n(\theta)$ with θ belongs to a semimetric space Θ with a semimetric d.

The rate of convergence for a estimator $\hat{\theta}_n$ is r_n, if $r_n(\hat{\theta}_n - \theta_0) = O_p(1)$.

Van der Vaart and Wellner (1996, Th 3.2.5) is commonly used to obtain the rate of convergence for the infinite-dimensional parametric estimators, which is also used in this thesis to deduce the rate of convergence for the regression parameters and B-splines based nonparametric estimators, thus we will summarize it here.

Theorem C.4 (Rate of Convergence) *Let $\mathbb{M}_n$ be stochastic processes indexed* *by a semimetric space* Θ*, and* $\mathbb{M} : \Theta \longrightarrow \mathbb{R}$ *a determinstic function, such that for every* θ *in a neighborhood of* θ_0*,*

$$\mathbb{M}(\theta) - \mathbb{M}(\theta_0) \lesssim -d^2(\theta, \theta_0).$$

Suppose that, for every n *and sufficiently small* δ*, the centered process* $\mathbb{M}_n - \mathbb{M}$ *satisfies*

$$E \sup_{d(\theta,\theta_0)<\delta} \left|(\mathbb{M}_n - \mathbb{M})(\theta) - (\mathbb{M}_n - \mathbb{M})(\theta_n)\right| \lesssim \frac{\phi_n(\delta)}{\sqrt{n}},$$

for function ϕ_n*, such that* $\delta \longrightarrow \phi_n(\delta)/\delta^\alpha$ *is decreasing for some* $\alpha < 2$ *(not depending on* n*). Let*

$$r_n^2 \phi_n(\frac{1}{r_n}) \leq \sqrt{n}, \qquad \textit{for every } n.$$

If the sequence $\hat{\theta}_n$ *satisfies* $\mathbb{M}_n(\hat{\theta}_n) \geq \mathbb{M}_n(\theta_0) - O_p(r_n^{-2})$ *and converges in probability to* θ_0*, then* $r_n d(\hat{\theta}_n, \theta_0) = O_p(1)$*. If the displayed conditions are valid for every* θ *and* δ*, then the condition that* $\hat{\theta}_n$ *is consistent is unnecessary.*

Note: The notation $\lesssim$ means "is bounded above up to a universal constant".

In the case of i.i.d. data and criterion functions of the form $\mathbb{M}_n(\theta) = \mathbb{P}_n m_\theta$, the centered and scaled process $\sqrt{n}(\mathbb{M}_n - \mathbb{M}) = \mathbb{G}_n m_\theta$. The second condition of the theorem involves the suprema of the empirical process indexed by classes of function

$$\mathcal{M}_\delta = \{m_\theta - m_{\theta_0} : d(\theta, \theta_0) < \delta\}.$$

It is not unreasonable to assume that these suprema are bounded uniformly in n. This leads to the Van der Vaart and Wellner (1996, Corollary 3.2.6) as follows.

Corollary 3.2.6 in Van der Vaart and Wellner (1996). *In the i.i.d. case assume that, for every θ in a neighborhood of θ_0,*

$$P(m_\theta - m_{\theta_0}) \lesssim -d^2(\theta, \theta_0).$$

Furthermore, assume that there exists a function ϕ such that $\delta \longrightarrow \phi(\delta)/\delta^\alpha$ is decreasing for some $\alpha < 2$ and, for every n,

$$E\|\mathbb{G}_n\|_{\mathcal{M}_\delta} \lesssim \phi(\delta).$$

If the sequence $\hat{\theta}_n$ satisfies $\mathbb{P}_n m_{\hat{\theta}_n} \geq \mathbb{P}_n m_{\theta_0} - O_p(r_n^{-2})$ and converges in probability to θ_0, then $r_n d(\hat{\theta}_n, \theta_0) = O_p(1)$ for every sequense r_n such that $r_n^2 \phi_n(\frac{1}{r_n}) \leq \sqrt{n}$, for every n.

The following lemma is Van der Vaart and Wellner (1996, Lemma 3.4.2), which is used in our thesis to prove the rate of convergence.

Lemma C.5 *Let $\mathcal{F}$ be class of measurable functions such that $Pf^2 < \delta^2$ and $\|f\|_\infty \leq M$ for every f in $\mathcal{F}$. Then*

$$E_P\|\mathbb{G}\|_{\mathcal{F}} \lesssim \tilde{J}_{[]}(\delta, \mathcal{F}, L_2(P)) \left(1 + \frac{\tilde{J}_{[]}(\delta, \mathcal{F}, L_2(P))}{\delta^2\sqrt{n}} M\right),$$

for a constant M, and where $\tilde{J}_{[]}(\delta, \mathcal{F}, L_2(P)) = \int_0^\delta \sqrt{1 + N_{[]}(\varepsilon, \mathcal{F}, L_2(P))} d\varepsilon$.

Z Estimators : Asymptotic Normality

A Z-estimator $\hat{\theta}_n$ is the approximate zero of a data-dependent function $\Psi_n : \Theta \longrightarrow \mathcal{L}$, where Θ is a subset of a Banach space, and $\mathcal{L}$ is another Banach space and $\Psi : \Theta \longrightarrow \mathcal{L}$ is a fixed map.

If $\mathcal{L}$ is an $l^\infty(\mathcal{H})$-space, as can be assumed without loss of generality, the equation $\Psi_n(\hat{\theta}_n) = 0$ is equivalent to the collection of (real-valued) estimating equations $\Psi_n(\hat{\theta}_n)h = 0$, when h run through $\mathcal{H}$.

In the case of i.i.d. observations, $\Psi_n(\theta)h = \mathbb{P}_n\psi_{\theta,h}$ and $\Psi(\theta)h = P\psi_{\theta,h}$ for given measurable functions $\psi_{\theta,h}$ indexed by Θ and an arbitrary index set $\mathcal{H}$. In this case $\sqrt{n}(\Psi_n - \Psi)(\theta) = \{\mathbb{G}_n\psi_{\theta,h} : h \in \mathcal{H}\}$ is the empirical process indexed by the class of functions $\{\psi_{\theta,h} : h \in \mathcal{H}\}$. Then the condition(i) needed for the proof of Theorem 2.3 and Theroem 3.3 is

$$\sqrt{n}(\Psi_n - \Psi)(\hat{\theta}_n) - \sqrt{n}(\Psi_n - \Psi)(\theta_0) = o_p(1). \tag{C3}$$

which can be satisfied under the sufficient conditions in the following Kosorok (2008, Lemma 13.3).

Lemma C.6 *Suppose that the class of functions*

$$\{\psi_{\theta,h} - \psi_{\theta_0,h} : \|\theta - \theta_0\| < \delta, h \in \mathcal{H}\}$$

is P-Donsker for some $\delta > 0$ and that

$$\sup_{h\in\mathcal{H}} P(\psi_{\theta,h} - \psi_{\theta_0,h})^2 \longrightarrow 0, \qquad \theta \longrightarrow \theta_0.$$

Then if $\Psi_n(\hat{\theta}_n) = o_p(n^{1/2})$ and $\hat{\theta}_n \xrightarrow{P} \theta_0$ then (C3) is satisfied.

Appendix D. The Bootstrap Estimate of Standard Error

The bootstrap was introduced in Efron (1979) as a computer-based method for estimating the standard error of estimators. The bootstrap estimate of standard error requires no theretical calculations, and is available no matter how mathematically complicated the estimators may be.

Assume that $X_1, \cdots, X_n$ independently sampled from an unknown probability distribution F, an estimate for the parameter of interest θ is $\hat{\theta} = t(\mathbf{X})$, where $\mathbf{X} = (X_1, \cdots, X_n)$. The standard error of $\hat{\theta}$ defined as

$$\mathrm{se}\{\hat{\theta}; F\} = [\mathrm{Var}_F\{t(\mathbf{X})\}]^{1/2} \qquad (D1)$$

is a commomly used measure of the accuracy for estimators $\hat{\theta}$. Genarate the bootstrap sample of size n, $\mathbf{X}^* = (X_1^*, \cdots, X_n^*)$ from the emipirical distribution $\hat{F}_n$, which is defined as $\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^n I\{X_i \le x\}$. Then substituting $\hat{F}_n$ for F in (D1) gives a reasonable estimate of the standard error for $\hat{\theta}$, namely

$$\mathrm{se}_{\mathrm{boot}}\{\hat{\theta}^*\} \equiv \mathrm{se}\{\hat{\theta}^*; \hat{F}_n\} = [\mathrm{Var}_{\hat{F}_n}\{t(\mathbf{X}^*)\}]^{1/2},$$

where $\hat{\theta}^* = t(\mathbf{X}^*)$. If there is no explicit formula to compute $\mathrm{se}_{\mathrm{boot}}\{\hat{\theta}^*\}$, the Monte Carlo approximation is proposed. That is, generate B independent bootstrap samples $\mathbf{X}^{*b}, \cdots, \mathbf{X}^{*b}$ $i.i.d. \sim \hat{F}_n$. Evaluate $\hat{\theta}^{*b} = t(\mathbf{X}^{*b})$, $b = 1, \cdots, B$. Then estimate the standard error $\mathrm{se}\{\hat{\theta}; F\}$ by the sample standar deviation of the B bootstrap samples

$$\hat{\mathrm{se}}_B = \left\{ \frac{1}{B-1} \sum_{b=1}^{B} [\hat{\theta}^{*b} - \hat{\theta}^{*\cdot}]^2 \right\}^{1/2},$$

where $\hat{\theta}^{*\cdot} = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^{*b}$, since $\lim\limits_{B\longrightarrow\infty}\hat{\text{se}}_B = \text{se}_{\text{boot}}\{\hat{\theta}^*\}$ and $\text{se}_{\text{boot}}\{\hat{\theta}^*\}$ is a plug-in estimate for $\text{se}\{\hat{\theta}; F\}$ (The plug-in estimate of a parameter $\theta = t(F)$ is defined to be $\hat{\theta} = t(\hat{F}_n)$, Efron and Tibshirani (1993)).

Remark D.1 *(1). Another way to say $\boldsymbol{X}^* = (X_1^*, \cdots, X_n^*)$ i.i.d. $\sim \hat{F}_n$: $X_1^*, \cdots, X_n^*$ are a random sample of size n drawn with replacement from the polulation of n objects $X_1, \cdots, X_n$. Here, the points $X_1, \cdots, X_n$ are treated as a population, with distribution $\hat{F}_n$. (2). Easy way to implement bootstrap sampling on the computer: Randomly select integers $i_1, \cdots, i_n$, each of which equals any value of $1, \cdots, n$ with probability $1/n$, then $X_1^* = X_{i_1}, \cdots, X_n^* = X_{i_n}$.*

Remark D.2 *As Efron and Tibshirani (1993) disscussed in Section 6.4, the number B will ordinarily be in the range 25-200 for estimation a standard error.*

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Second International Symposium on Information Theory* **1,** 267–281.

Albert, P. S. (1991). A two-state markov mixture model for a time series of epileptic seizure counts. *Biometrics* **47,** 1371–1381.

Amorim, L. D., Cai, J., Zeng, D., and Barreto, M. (2008). Regression splines in the time-dependent coefficient rates model for recurrent event data. *Statistics in Medicine* **27,** 5890–5906.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics* **10,** 1100–1120.

Andrews, D. F. and Herzberg, A. M. (1985). *Data : a collection of problems from many fields for the student and research worker*. New York: Springer-Verlag.

Balakrishnan, N. and Zhao, X. (2009). New multi-sample nonparametric tests for panel count data. *The Annals of Statistics* **37,** 1112 – 1149.

Balakrishnan, N. and Zhao, X. (2010). A nonparametric test for the equality of counting processes with panel count data. *Computational Statistics and Data Analysis* **54,** 135 – 142.

Balakrishnan, N. and Zhao, X. (2011). A class of multi-sample nonparametric tests for panel count data. *Annals of the Institute of Statistical Mathematics* **63,** 135–156.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. (1972). *Statistical Inference Under Order Restrictions.* John Wiely, New York.

Byar, D. P. (1980). The veterans administration study of chemoprophylaxis for recurrent stage 1 bladder tumors: comparisons of placebo, pyridoxine, and topical thiotepa. In *Bladder Tumors and Other Topics in Urological Oncology.* (eds Pavane-Macaluso, M., Smith, P. H. and Edsmyr, F.), pages 363–370, New York.

Cai, J., Fan, J., Jiang, J., and Zhou, H. (2007). Partially linear hazard regression for multivariate survival data. *Journal of the American Statistical Association* **102,** 538–551.

Cai, T. and Betensky, R. A. (2003). Hazard regression for interval-censored data with penalized spline. *Biometrics* **59,** 570–579.

Chen, B. E., Cook, R. J., Lawless, J. F., and Zhan, M. (2005). Statistical methods for multivariate interval-censored recurrent events. *Statistics in Medicine* **24,** 671–691.

Cook, R. J. and Lawless, J. F. (2007). *The Analysis of Recurrent Event Data.* Springer - Verlag, New York.

Cook, R. J., Lawless, J. F., and Nadeau, C. (1996). Robust tests for treatment comparisons based on recurrent event responses. *Biometrics* **52,** 557–571.

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

de Boor, C. (1978). *A Practical Guide to Splines.* Springer-Verlag.

Deng, S. (2012). Semiparametric regression analysis of panel count data with time-dependent covariates and informative observation and censoring times. in Press.

Dewanji, A. and Kalbfleisch, J. D. (1986). Nonparametric methods for survival/sacrifice experiments. *Biometrics* **42,** 325–341.

Diamond, I., McDonald, J., and Shah, I. (1986). Proportional hazards models for current status data: Application to the study of differentials in age at weaning in pakistan. *Demography* **23,** 607–620.

Diggle, P. J., Heagerty, P., Liang, K. Y., and Zeger, S. L. (1994). *Analysis of Longitudinal Data.* U. K. Oxford University Press, Oxford.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* **7,** 1–26.

Efron, B. and Tibshirani, R. (1993). *An introduction to the bootstrap.* Chapman and Hall (New York).

Esteve, J., Benhamou, E., Croasdale, M., and Raymond, L. (1990). Relative survival and the estimation of net survival: Elements for further discussion. *Statistics in Medicine* **9,** 529–538.

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

Fan, J., Lin, H., and Zhou, Y. (2006). Local partial-likelihood estimation for lifetime data. *The Annals of Statistics* **34,** 290–325.

Fletcher, R. (1987). *Practical methods of optimization, Volume 1.* Wiley.

Gaver, D. P. and O'Muircheartaigh, I. G. (1987). Robust empirical bayes analyses of event rates. *Technometrics* **29,** 1–15.

Giorgi, R., Abrahamowicz, M., Quantin, C., Bolard, P., Esteve, J., Gouvernet, J., and Faivre, J. (2003). A relative survival regression model using b-spline functions to model non-proportional hazards. *Statistics in Medicine* **22,** 2767–2784.

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

He, X., Tong, X., Sun, J., and Cook, R. J. (2008). Regression analysis of multivariate panel count data. *Biostatistics* **9,** 234–248.

He, X., Zhu, Z., and Fung, W. (2002). Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika* **89,** 579–590.

Hoover, D. R., Rice, J. A., Wu, C. O., and Yang, L. P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85,** 809–822.

Hu, X. J., Lagakos, S. W., and Lockhart, R. A. (2009a). Generalized least squares estimation of the mean function of a counting process based on panel counts. *Statistica Sinica* **19,** 561– 580.

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

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

Hua, L. and Zhang, Y. (2011). Spline-based semiparametric projected generalized estimating equation method for panel count data. *Biostatistics* **0,** 1–15.

Huang, C. Y., Qin, J., and Wang, M. C. (2010). Semiparametric analysis for recurrent event data with time-dependent covariates and informative censoring. *Biometrics* **66,** 39–49.

Huang, C. Y. and Wang, M. C. (2004). Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association* **99,** 1153–1165.

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

Huang, J. (1999). Efficient estimation of the partly linear additive cox model. *Journal of the American Statistical Association* **27,** 1536–1563.

Huang, J. Z. and Liu, L. (2006). Polynomial spline estimation and inference of proportional hazards regression models with flexible relative risk form. *Biometrics* **62,** 793–802.

Huang, J. Z., Wu, C. O., and Zhou, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statistica Sinca* **14,** 763–788.

Jamshidian, M. (2004). On algorithms for restricted maximum likelihood estimation. *Computational Statistics and Data Analysis* **45,** 137 – 157.

Jongbloed, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *Journal of Computational and Graphical Statistics* **7,** 310–321.

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

Kang, M. and Lagakos, S. W. (2007). Statistical methods for panel data from a semi-markov process, with application to hpv. *Biostatistics* **8,** 252–264.

Kay, R. (1986). A markov model for analysing cancer markers and disease states in survival studies. *Biometrics* **42,** 855–865.

Kelly, C. and Rice, J. (1990). Monotone smoothing with application to dose-response curves and the assessment of synergism. *Biometrics* **46,** 1071–1085.

Kooperberg, C., Stone, C. J., and Truong, Y. K. (1995). Hazard regression. *Journal of the American Statistical Association* **90,** 78–94.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer series in statistics. Springer.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38,** 963–974.

Lawless, J. F. and Nadeau, J. C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics* .

Lee, E. W. and Kim, M. Y. (1998). The analysis of correlated panel data using a continuous-time markov model. *Biometrics* **54,** 1638–1644.

Leitenstorfer, F. and Tutz, G. (2007). Generalized monotonic regression based on b-splines with an application to air pollution data. *Biostatistics* **8,** 654–673.

Li, N., Sun, L., and Sun, J. (2010). Semiparametric transformation models for panel count data with dependent observation process. *Statistics in Biosciences* **2,** 191–210.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73,** 13–22.

Liang, Y., Lu, W., and Ying, Z. (2009). Joint modeling and analysis of longitudinal data with informative observation times. *Biometrics* **65,** 377–384.

Lin, D. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association* **96,** 103–126.

Lin, D. Y. and Wei, L. J. (1989). The robust inference for the cox proportional hazards model. *Journal of the American Statistical Association* **84,** 1074–1078.

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

Lin, H., Scharfstein, D. O., and Rosenheck, R. A. (2004). Analysis of longitudinal data with irregular, outcome-dependent follow-up. *Journal of the Royal Statistical Society, Series B* **66,** 791–813.

Lin, X. and Zhang, D. (1999). Inference in generalized additive mixed modelsby using smoothing splines. *Journal of the Royal Statistical Society, Series B* **61,** 381–400.

Lipshultz, S. E., Lipsitz, S. R., Mone, S. M., Goorin, A. M., Sallan, S. E., Sanders, S. P., Orav, E. J., Gelber, R. D., and Colan, S. D. (1995). Female sex and higher drug dose as risk factors for late cardiotoxic effects of doxorubicin therapy for childhood cancer. *New England Journal of Medicine* **332,** 1738–1744.

Lipsitz, S. R., Fitzmaurice, G. M., Ibrahim, J. G., Gelber, R., and Lipshultz, S. (2002). Parameter estimation in longitudinal studies with outcome-dependent follow-up. *Biometrics* **58,** 621–630.

Liu, L., Huang, X., and O'Quigley, J. (2008). Analysis of longitudinal data in the presence of informative observational times and a dependent terminal event, with application to medical cost data. *Biometrics* **64,** 950–958.

Lu, M. (2010). Spline-based sieve maximum likelihood estimation in the partly linear model under monotonicity constraints. *Journal of Multivariate Analysis* **101,** 2528 – 2542.

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

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

Lu, W. and Zhang, H. H. (2010). On estimation of partially linear transformation models. *Journal of the American Statistical Association* **105,** 683–691.

Luenberger, D. G. (1984). *Linear and nonlinear programming.* Addison-Wesley.

Nielsen, J. D. and Dean, C. B. (2008). Clustered mixed nonhomogeneous poisson process spline models for the analysis of recurrent event panel data. *Biometrics* **64,** 751–761.

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

Pollard, D. (1990). *Empirical processes: Theory and applications.* Hayward: Institute of Mathematical Statistics.

Ramsay, J. O. (1988). Monotone regression splines in action. *Statistical Science* **3,** 459–461.

Robertson, T., Wright, F. T., and Dykstra, R. (1988). *Order restricted statistical inference.* Wiley.

Robins, J. M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiolog - Methodological Issues,* eds N. Jewell, K. Dietz and V. Farewell. Boston : Birkh*ä*user.

Rockafellar, R. T. (1970). *Convex Analysis.* Princeton University Press.

Rosen, J. B. (1960). The gradient projection method for nonlinear programming. *Journal of the Society for Industrial and Applied Mathematics* **8,** 181–217.

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

Rubnitz, J. E., Inaba, H., Dahl, G., Ribeiro, R. C., Bowman, W. P., Taub, J., Pounds, S., Razzouk, B. I., Lacayo, N. J., Cao, X., Meshinchi, S., Degar, B., Airewele, G., Raimondi, S. C., Onciu, M., Coustan-Smith, E., Downing, J. R., Leung, W., Pui, C., and Campana, D. (2010). Minimal residual disease-directed therapy for childhood acute myeloid leukaemia: results of the aml02 multicentre trial. *The Lancet Oncology* **11,** 543 – 552.

Schumaker, L. L. (1981). *Spline Functions: Basic Theory*. John Wiley and Sons.

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

Shen, X. (1998). Propotional odds regression and sieve maximum likelihood estimation. *Biometrika* **85,** 165–177.

Shen, X. and Wong, W. H. (1994). Convergence rate of sieve estimates. *The Annals of Statistics* **22,** 580–615.

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

Sleeper, L. A. and Harrington, D. P. (1990). Regression splines in the cox model with application to covariate effects in liver disease. *Journal of the American Statistical Association* **85,** 941–949.

Song, X., Mu, X., and Sun, L. (2012). Regression analysis of longitudinal data with time-dependent covariates and informative observation times. *Scandinavian Journal of Statistics* .

Stakgold, I. (1998). *Green's functions and boundary value problems*. Toronto :John Wiley, New York.

Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics* **8,** 1348–1360.

Stone, C. J. (1982). Optimal globle rates of convergence for nonparametric estimators. *The Annals of Statistics* **10,** 1040–1053.

Stone, C. J. (1985). Additive regression and other nonparametric models. *The Annals of Statistics* **13,** 689–705.

Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. *The Annals of Statistics* **14,** 590–606.

Sun, J. (2006). *The Statistical Analysis of Interval-censored Failure Time Data.* New York, NY : Springer Science+Business Media, Inc.

Sun, J. and Fang, H. (2003). A nonparametric test for panel count data. *Biometrika* **90,** 199–208.

Sun, J. and Kalbfleisch, J. D. (1993). The analysis of current status data on point processes. *Journal of the American Statistical Association* **88,** pp. 1449–1454.

Sun, J. and Kalbleisch, J. D. (1995). Estimation of the mean function of point processes based on panel count data. *Statistica Sinca* **5,** 279–290.

Sun, J., Kopciuk, K. A., and Lu, X. (2008). Polynomial spline estimation of partially linear single-index proportional hazards regression models. *Computational Statistics and Data Analysis* **53,** 176 – 188.

Sun, J., Park, D. H., Sun, L., and Zhao, X. (2005). Semiparametric regression analysis of longitudinal data with informative observation times. *Journal of the American Statistical Association* **100,** 882–889.

Sun, J., Sun, L., and Liu, D. (2007). Regression analysis of longitudinal data in the presence of informative observation and censoring times. *Journal of the American Statistical Association* **102,** 1397–1406.

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

Sun, L., Song, X., and Zhou, J. (2011). Regression analysis of longitudinal data with time-dependent covariates in the presence of informative observation and censoring times. *Journal of Statistical Planning and Inference* **141,** 2902 – 2919.

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

Thall, P. F. and Vail, S. C. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics* **46,** 657–671.

Tong, X., He, X., Sun, L., and Sun, J. (2009). Variable selection for panel count data via non-concave penalized estimating function. *Scandinavian Journal of Statistics* **36,** 620–635.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.

Verbeke, G. and Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. Springer, New York.

Wang, M. C., Qin, J., and Chiang, C. T. (2001). Analyzing recurrent event data with informative censoring. *Journal of the American Statistical Association* **96,** 1057–1065.

Wang, Y. (1998). Mixed effects smoothing spline analysis of variance. *Journal of the Royal Statistical Society, Series B* **60,** 159–174.

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

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

Whittemore, A. S. and Keller, J. B. (1986). Survival estimation using splines. *Biometrics* **42,** 495–506.

Wu, C. O., Chiang, C. T., and Hoover, D. R. (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *Journal of the American Statistical Association* **93,** 1388–1402.

Wu, T. and He, X. (2012). Coordinate ascent for penalized semiparametric regression on high-dimensional panel count data. *Computational Statistics and Data Analysis* **56,** 25 – 33.

Xue, H., Lam, K. F., and Li, G. (2004). Sieve maximum likelihood estimator for semiparametric regression models with current status data. *Journal of the American Statistical Association* **99,** 346–356.

Yin, G., Li, H., and Zeng, D. (2008). Partially linear additive hazards regression with varying coefficients. *Journal of the American Statistical Association* **103,** 1200–1213.

Zeger, S. L. and Diggle, P. J. (1994). Semiparametric models for longitudinal data with application to cd4 cell numbers in hiv seroconverters. *Biometrics* **50,** 689–699.

Zeger, S. L., Liang, K. Y., and Albert, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics* **44,** 1049–1060.

Zeng, D., Lin, D. Y., and Yin, G. (2005). Maximum likelihood estimation for the proportional odds model with random effects. *Journal of the American Statistical Association* **100,** 470–483.

Zhang, D., Lin, X., Raz, J., and Sowers, M. (1998). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association* **93,** 710–719.

Zhang, Y. (2002). A semiparametric pseudolikelihood estimation method for panel count data. *Biometrika* **89,** 39–48.

Zhang, Y. (2006). Nonparametric k-sample tests with panel count data. *Biometrika* **93,** 777–790.

Zhang, Y., Hua, L., and Huang, J. (2010). A spline-based semiparametric maximum likelihood estimation method for the cox model with interval-censored data. *Scandinavian Journal of Statistics* **37,** 338–354.

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

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

Zhang, Z., Sun, J., and Sun, L. (2005). Statistical analysis of current status data with informative observation times. *Statistics in Medicine* **24,** 1399–1407.

Zhao, X., Balakrishnan, N., and Sun, J. (2011). Nonparametric inference based on panel count data. *Test* **20,** 1–42.

Zhao, X. and Sun, J. (2011). Nonparametric comparison for panel count data with unequal observation processes. *Biometrics* **67,** 770–779.

Zhao, X. and Tong, X. (2011). Semiparametric regression analysis of panel count data with informative observation times. *Computational Statistics and Data Analysis* **55,** 291 – 300.

Zhao, X., Tong, X., and Sun, L. (2012). Joint analysis of longitudinal data with dependent observation times. *Statistica Sinica* **22,** 317–336.

Zhou, S., Shen, X., and Wolfe, D. A. (1998). Local asymptotics for regression splines and confidence regions. *Computational Statistics and Data Analysis* **26,** 1760–1782.

Zhu, L., Sun, J., Tong, X., and Pounds, S. (2011). Regression analysis of longitudinal data with informative observation times and application to medical cost data. *Statistics in Medicine* **30,** 1429–1440.