SEMIPARAMETRIC REGRESSION ANALYSIS OF RECURRENT EVENTS

XUENAN FENG

Ph.D

The Hong Kong Polytechnic University

2016

The Hong Kong Polytechnic University
Department of Applied Mathematics

Semiparametric Regression Analysis of Recurrent Events

Xuenan Feng

A thesis submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

April 2016

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)

________FENG Xuenan_______(Name of student)

Dedicate to my family.

Abstract

In this Dissertation, we mainly concern the effects of covariates on the underlying recurrent event process. Two topics are considered:

Recurrent event data are data in which the event of interest can occur repeatedly and the successive event times are available. We study the semiparametric regression model with random effects for recurrent event data in the presence of informative censoring times. For inference, we propose using the maximum likelihood approach for estimation of the underlying baseline intensity function and regression parameters. The proposed estimates are consistent and have asymptotically a normal distribution. Also the maximum likelihood estimators of regression parameters are asymptotically efficient. The finite sample properties of the proposed estimates are investigated through simulation studies. An illustrative example from a clinical trial is provided.

Panel count data deal with the recurrent events in discrete times. We study the semiparametric regression analysis of panel count data when certain covariate effects may be much more complex than linear effects. To explore the nonlinear interactions between covariates, we propose a class of partially linear models with possibly varying coefficients for the mean function of the counting processes with panel count data. The functional coefficients are estimated by B-spline function approximations. The estimation procedures are based on maximum pseudo-likelihood and likelihood approaches and they are easy to implement. The asymptotic properties of the resulting estimators are established, and their finite-sample performance is assessed by Monte Carlo simulation studies. We also demonstrate the value of the proposed method by the analysis of a cancer data set, where the new modeling approach provides more comprehensive information than the usual proportional mean model.

Acknowledgements

Nietzsche once said "the charm of knowledge would be small if so much shame did not have to be overcome on the road to it". It was lucky for me to obtain this precious opportunity to conquer shame. The endeavor for studying and thinking is for lasting.

First and foremost, I wish to express my heartfelt thanks to my supervisor, Prof. Xingqiu Zhao, for her guidance and encouragement without which the completion of this thesis would have been impossible.

I'm also grateful to Prof. Heung Wong and Prof. Xun Li for their advice and kind help through the course of my study.

Furthermore, I would like to thank AMA/PolyU for their financial support and facinating environment. And my peers are smart and open to share brilliant ideas. Nice memories to be cherished forever.

Last but not least, I am indebted to my family for their unconditional love and steadfast support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

The objective of statistical inference is to estimate or to predict unknowns based on given data or information. The performance of the models is mainly based on how close the assumptions made to the true mechanism from which the data is generated. Cross-sectional study and longitudinal study are two types of observational studies that observe the state of interest with no interventions. Data are collected at a static time point from a cross-sectional study, on either individual or group level within a specified population. While longitudinal study is conducted over time to detect possible effects for the observed changes. Memory ability of different age groups can be compared with a cross-sectional study, but the effect of aging towards memory loss needs to be examined through a longitudinal study. Prospective studies and retrospective studies are two types of longitudinal studies. The prospective studies look forward in time. For example, we select two groups of subjects and follow them for years. It may take even decades to accumulate sufficient data to draw any strong conclusions. Reversely, retrospective studies work backwards. Like, we find people infected with some disease and try to uncover the potential risk factors. The biggest problem in a retrospective study is that some of the information needed may not be observed. We usually use it when the event takes a long time to appear.

Terminal event or recurrent events are usually confronted in longitudinal studies. Time-to-event data modeling is a branch of statistics that examines the time point or duration the event of interest occurs and explores the underlying effects of risk factors (covariates) on the time. In event history analysis, the states changing could be characterized as stochastic processes with discrete state space and either discrete or continuous time. The event of interest could be a single event for each subject, such as death or the first infection of some disease. It could be multiple events of one or more types that occur repeatedly (recurrent events), e.g. infection occurrences and relapse of cancer.

The statistical field of dealing with the single event is called survival analysis or failure time analysis, having only two states "alive or normal" and "dead or failure". The object of survival analysis is the hazard rate function for the survival time distribution, with either censored or complete data. Lawless (1983) summarizes the developments made during 1959-1983 in this area and gives suggestions on further problems to be solved. Cox and Oaks (1984), Kalbfleisch and Prentice (2002) and Klein and Moeschberger (2003) give explicit details and examples in survival analysis. The censoring mechanism is crucial for survival data modeling, right-censored and interval-censored data are commonly encountered.

The case of recurrent events occur in continuous time is referred to as *recurrent event data*, where the successive event times are available. Two processes involved are the underlying counting process that characterizes the recurrent process of interest and the follow-up process. Most existing methods have been developed for the analysis of such data by assuming that the follow-up times are independent of the occurrence of the events completely or given covariates. Early works from Aalen (1975, 1978) pave the way for a general framework of conducting event history analysis in terms of counting process by means of its intensity process. The study of event history data has been going on for decades based on counting processes, martingale theory and stochastic integrals. Renowned models in survival analysis are extended to fit for event history data, e.g. the counting process with Cox type intensity model in Andersen and Gill (1982), which is frequently used for the analysis of recurrent event data. One may refer to Andersen and Borgan (1985) for an extensive review without any prior knowledge. Cook and Lawless (2007) focuses mainly on recurrent event data and thoroughly organizes important models and methods in detail; and illustrative examples are bladder cancer data, bowel motility data, pulmonary exacerbations and rhDNase, software debugging data, and artificial field repair data.

Panel count data deal with the recurrent events in discrete times. The counts of the occurrences of the events between observation times are provided without specific recurrent times. Some clinical trial examples in Thall and Lachin (1988) include gallstones patients with nausea, diabetes patients with hypoglycemia, coronary disease patients with angina pectoris, and epileptic patients with seizures. For all the cases, the recurrence situation of the symptom corresponding to each disease is required at each visit to the hospital for every individuals. Patients may enter late, miss some prescribed visits, or drop out of the study early for different reasons. The number of observations and observation times are allowed to be distinct from individual to individual. If complete records with exact occurrence times are impractical to be prepared, but the event counts between two subsequent time points are available, panel count data analysis is then conducted. A special case is the so-called *current status data* that each subject is observed only once and there are only the total counts of occurrences by the observation time available. Examples and more discussions can be found in Sun and Kalbfleisch (1993). Recurrent events of multiple types that observe at discrete times are referred to as *multivariate penal count data*, and the relationships among those types of events may not be ignored. For instance, basal cell carcinoma and squamous cell carcinoma are two types of non-melanoma skin cancer. It was diagnosed that some patients develop one of these two types of cancer, but some other patients have both under discrete recurrence times. Different from recurrent event data, in addition to the recurrent event process and the follow-up process, the observation process with incomplete occurrence times needs to be considered for panel count data. The dependence or relationship between the event process and the observation process, and even the correlations among the three processes make panel count data special. Sun and Zhao (2013) suggests that one may understand the structure difference between recurrent event data and panel count data via seeing the difference between right-censored and interval-censored data in survival analysis.

1.2 Literature review

As we know, the intensity process fully specifies the corresponded univariate counting process. Aalen (1978) first introduces stochastic integrals and the martingale-based counting process theory into statistical inference. Andersen and Gill (1982) extends the Cox model to a multivariate counting process set-up and studies the asymptotic properties by means of martingale central limit theorem. But in many situations, the dependence structures among recurrent events cannot be modeled properly via intensity models. Pepe and Cai (1993) advocates a more flexible idea of using rate functions. Lawless and Nadeau (1995) studies mean functions and gives asymptotic results for discrete time cases. Lawless (1995) reviews the methods based on intensity function as well as the marginal rate and mean functions for recurrent event data. The intensity model requires that the counting process has jumps of size one, while the rate and mean functions work well for characterizing the counting process with arbitrary jump sizes. Since martingale central limit theorem fails for arbitrary continuous counting process with rates and mean model, Lin et al. (2000) provides rigorous proofs through empirical process theory and establishes a class of graphical and numerical techniques for model checking. Thereafter, one may model the intensity process or the mean function for the analysis of recurrent event data based on their research questions. But for panel count data modeling, the focus is only on the rate or mean function of the underlying recurrent event process.

In clinical trial studies, patients are often classified into groups in terms of medical treatment or some categorical covariates (e.g. gender). For panel count data analysis, the null hypothesis could be formulated by the mean functions of the underlying recurrent event processes in different groups. Estimation of mean functions is the foundation for comparing two or more recurrent event processes in the light of constructing test statistics.

Isotonic regression involves finding a weighted least-squares subject to some order restriction. Sun and Kalbfleisch (1995) presents the isotonic regression estimator (IRE) of the mean function with the monotonicity constraint using the max-min formula in Barlow et al. (1972). Under the non-homogeneous Poisson assumption of the underlying counting process, Wellner and Zhang (2000) shows that the IRE is identical with the nonparametric maximum pseudo-likelihood estimator (NPMPLE), which ignores the dependence of successive counts and treats them as independent. While their proposed nonparametric maximum likelihood estimator (NPMLE) allows for the dependence structure and utilizes the independence of the increments of the process under Poisson assumption. The NPMPLE and the NPMLE of the mean function are both consistent. The robustnesses of the two estimators against the Poisson assumption are revealed through simulation studies. They are identical when it comes to the current status data, the NPMLE is more efficient when the data has numerous observation times but more intractable in computing. Groeneboom and Wellner (1992) considers the expectation-maximization (EM) algorithm and the iterative convex minorant (ICM) algorithm in solving the NPMLE for interval censored data. Jongbloed (1998) modifies the ICM algorithm (MICM) by inserting a line search procedure and proves the global convergence. Wellner and Zhang (2000) follows this idea and utilizes the MICM algorithm in acquiring the NPMLE of the mean function for panel count data. Cheng et al. (2011) proposes a projected Newton-Raphson algorithm to compute the NPMLE that is faster than the ICM algorithm in the literature. Asymptotic properties of the NPMPLE and the NPMLE without any model assumption for the counting process are given in Wellner and Zhang (2000) via empirical process theory. The assumption of independent counts within each subject is too restrictive. To make the NPMPLE more efficient without sacrificing simplicity in computation, one may take the correlation of sequential counts into account by importing some frailty variable. Zhang and Jamshidian (2003) generalizes the non-homogeneous Poisson assumption about the underlying counting process to conditional Poisson given the gamma-frailty variable (mixed-Poisson), in which the mean function and the gamma parameter are estimated by the EM algorithm, but leave the study of asymptotic behaviors an open problem. To obtain a smooth estimator of the mean function, Lu et al. (2007) presents the spline-based estimator by involving the monotone cubic spline (I-spline) proposed by Ramsay (1988) in the NPMLE and the NPMPLE of Wellner and Zhang (2000). Instead of giving only one estimator, Hu et al. (2009a) presents a class of isotonic regression-based estimators depending on the weight function, which include the IRE when the weight function is the identity matrix. Different choice of the weight function results in diverse estimator of the mean function with distinct efficiency. Simulation studies show that the estimator with GEE weights is close to the efficiency of the NPMLE for Poisson event processes and is more efficient for non-Poisson processes. Rather than direct methods of estimating the mean function, Thall and Lachin (1988) suggests to estimate the rate function by partitioning the study period into fixed number of consecutive time intervals and to compare two treatment groups based on the empirical rate difference. One may derive the mean function accordingly by the integral of the estimator of the rate function. Hu et al. (2009b) derives the mean function by a summation of the estimates of the rate function at finite observation times, and provides two types of self-consistent estimating equations in nonparametric estimation. All the methods so far are based on the assumption that the observation process and the underlying recurrent event process are independent. Panel count data with informative observation process is dealing with the case that the two processes are not independent. Many research studies about this problem are discussed in the context of regression analysis.

For two-sample comparison problems in the analysis of panel count data, the test statistic is commonly constructed using the mean functions based on the above mentioned NPMLE and NPMPLE/IRE. For simplicity, Thall and Lachin (1988) transforms the problem of rate comparison into a multivariate comparison one by partitioning the whole length of study into fixed and non-overlapped consecutive intervals and proposes a Wilcoxon-like rank test, but the efficiency would be highly dependent upon the partition settings. Sun and Kalbfleisch (1993) accommodates the IRE of the mean functions in the test statistic motivated by the log-rank test for right-censored survival comparison, in the context of current status data both with same and different observation times among all subjects. Sun and Fang (2003) generalizes this idea to panel count data and proves the asymptotic normal properties of the test statistic that represents the integrated weighted difference between the IREs of one group and the overall of the common mean under the assumption that subjects from different population or treatment group share identical observation processes. One important type of experiment in medical research is randomized clinical trials (RCTs), where subjects under study are randomly allocated to different treatment groups, hence meets the requirement of the test statistic. Park et al. (2007) suggests a test statistic representing the integrated weighted difference between the IRE's of the two mean functions for the two different groups with no assumption of independent and identically distributed group indicators in Sun and Fang (2003) and shows its asymptotic normality. Motivated by the efficiency of the NPMLE, Balakrishnan and Zhao (2010) follows the idea from Sun and Fang (2003) and constructs a different test statistic based on the NPMLE, and compares the power (the probability that the test appropriately reject the null hypothesis when the alternative hypothesis is true) of the hypothesis tests based on both the NPMLE and the NPMPLE/IRE through comprehensive simulation studies, where the test statistics in Sun and Fang (2003) and Park et al. (2007) with three different weight functions are all involved. For more details on simulation studies, one may refer to Balakrishnan and Zhao (2010) and Zhao et al. (2011). For multivariate panel count data, the null hypothesis accommodates all types of recurrent events that the mean functions in different groups being equal. Zhao et al. (2014) generalizes the test statistic in Park et al. (2007) leaving the correlation between different types of recurrent events arbitrary and proves its asymptotic normality. This can be extended to multi-sample situation.

For comparing mean functions of k-sample, Zhang (2006) constructs a class of nonparametric tests based on the asymptotic normality of a smooth functional of the NPMPLEs of two groups under monotonic condition of the limit weight process. Balakrishnan and Zhao (2011) generalizes the existing test statistics based on the NPMPLE to the multi-sample cases with both the assumption of iid treatment indicator holds and not. As suggested, more weight processes can be used in their test statistics by relaxing the assumption about the weight function in Zhang (2006). One could certainly substitute the NPMLE for the NPMPLE/IRE, but it is more complicated both in theory and computation. Balakrishnan and Zhao (2009) considers two classes of nonparametric test statistics based on the integrated weighted difference between the increased rates of the NPMLEs. So far, the assumption of identical observation processes in different groups is made for the aforementioned test procedures. To further consider the difference of the observation processes in different groups, Zhao and Sun (2011) formulates their test statistics allowing unequal observation processes based on the IRE and establishes asymptotic normality. The weight functions work for imposing different emphases on different time intervals either involving the group difference or not. One may consider the NPMLE instead and construct new nonparametric test procedures. But it's not always better that the NPMLE-based procedures perform. Zhao et al. (2011) shows that the choices of both the test statistic and the weight process become compliated when the underlying mean functions between groups cross, although in many cases we have no idea whether it is true or not.

Regression analysis of panel count data is mainly to estimate the conditional expectation of the underlying recurrent event process given covariates (either time-invariant or time-varying). The fundamental model is the so-called proportional mean model that the baseline mean function and the regression parameters are to be determined. Early studies are built upon this assumption of the model. Zhang (2002) follows the idea of mean function estimation with the IRE and recommends the semiparametric maximum pseudo-likelihood estimation (SPMPLE) based on the non-homogeneous Poisson process assumption. The two-step iterative algorithm for the SPMPLE is to give an initial set of the regression parameters for updating the mean via the IRE, and then with the updated mean to get a new estimator of the regression parameters using Newton-Raphson algorithm. Convergence is then measured by the relative difference between the log pseudo-likelihoods of two consecutive iterations, which is robust against the Poisson assumption. Although the SPMPLE has computational advantages, it is less efficient when the number of observation times is heavy-tailed as indicated in Wellner et al. (2004). A feasible algorithm for the semiparametric maximum likelihood estimation (SPMLE) is further required. Wellner and Zhang (2007) replaces the IRE with the NPMLE by the MICM algorithm and utilizes the same two-step algorithm to get the SPMLE, and the SPMPLE of the regression parameters is chosen as the initial value. But this algorithm for the SPMLE is time-consuming due to large amount of iterations for both MICM and Newton-Raphson. Asymptotic normality has been established for both the SPMPLE and the SPMLE of the regression parameters. But asymptotic variances have very complicated forms and hence being estimated by bootstrap procedures. Rather than treating the baseline mean function unspecified, one may consider a smooth function (e.g. kernel or splines) approximation of the baseline mean function. As in Lu et al. (2007), the spline-based estimator of the mean function performs better than the NPMPLE and the NPMLE in Wellner and Zhang (2000). Lu et al. (2009) proposes the spline-based sieve SPMPLE and SPMLE using the monotone cubic B-splines to approximate the logarithm of the baseline mean function. The spline coefficients and the regression parameters are jointly estimated by the generalized Rosen algorithm of Jamshidian (2004) in the modified form. The dimension has greatly reduced for the estimation procedure and the computation is less intensive than the estimators of Wellner and Zhang (2007). For possible improvements, Lu et al. (2009) suggests that the penalized likelihood method could be used to select the amount and location of the knots rather than prespecifying partition for monotone B-splines, and additive or additive-multiplicative mean model could be considered instead of the restrictive proportional mean model. One may also assume the underlying counting process to be mixed Poisson and involve a latent variable in the mean model, similar procedures can be made to estimate the baseline mean or spline coefficients and regression parameters. Poisson assumption implies that the variance and the mean are equal for the recurrent event process given covariates, and the variance is no less than the mean for the mixed Poisson assumption. Both are too confining and may not hold in practice.

Generally our main interest is the effects of covariates on the underlying recurrent event process, but sometimes the covariate effects on the observation process and even the follow-up process cannot be ignored. For instance, patients in different treatment group may have different frequencies for visiting their doctor. The observation process may depend on the treatment indicator as a covariate. The likelihood-based methods model the underlying recurrent event process condition on the observation process. Sun and Wei (2000) leads an alternative way to solve regression problem utilizing estimating equation methods that further model the covariate effects on the observation process by the proportional mean model and the follow-up process by the proportional hazards model. This approach does not depend on the assumption of the underlying recurrent event process, but it requires that the three processes are independent given covariates. The three estimators of the regression parameters are proved consistent and unique. The asymptotic normality and a closed form asymptotic variance are given for all of them, respectively. Estimating equation methods do not need to estimate the unknown baseline mean function, and hence less intensive in computation. But the likelihood-based methods are more efficient when the distributional assumptions make clear sense. More on estimating equation methods, one may refer to Cheng and Wei (2000) and Hu et al. (2003). Sun and Zhao (2013) gives a comprehensive review of the up-to-date methodologies on regression analysis of panel count data, primarily based on the assumptions of time-independent covariate effects for regression models and no missing or censoring data on covariates, with either independent or informative observation times. Sun et al. (2009) generalized the proportional mean model to incorporate both time-independent and time-varying regression parameters with independent observation times.

Aforementioned approaches require that the observation process is independent of the underlying recurrent event process given covariates. The marginal approach is often used for inference about the recurrent event process. But in situations we have to consider their correlations, one may model them jointly or model one marginally and the other conditionally given the former. To accommodate the follow-up process as well, they are commonly referred to as panel count data with informative observation times and informative censoring. Huang et al. (2006) and Sun et al. (2007) investigate the case when the observation process and the underlying recurrent event process are related, while the follow-up process is independent of the other two processes given covariates. Huang et al. (2006) assumes a non-homogeneous Poisson process for the underlying recurrent event process and proposes the nonparametric and semiparametric proportional rate models with a multiplicative frailty to characterize the correlation between the two processes. No distributional assumptions are made for the frailty and the observation process. The conditional likelihood method is employed to estimate the baseline rate function and the regression parameters. Sun et al. (2007) assumes that the observation process is a non-homogeneous Poisson process and proposes another semiparametric model for solving this problem. He et al. (2009) considers more complicated situation that the three processes are correlated. They conduct a joint analysis with some shared frailty models and introduce two latent variables to account for the correlations among the three processes. A three-step estimation procedure is presented and the estimating equation approach is used. Instead of constructing a specified latent variable, Zhao and Tong (2011) proposes a joint model in characterizing the correlation between the two processes through a latent variable along with a completely unspecified link function, and the latent variable does not need to be estimated in the estimation procedure. They also address that the proposed method could be generalized to the case with informative censoring by utilizing the approach in Sun et al. (2007). Deng (2013) follows this idea and generalizes to the situation with time-dependent covariates and informative censoring. To further relax the concerned assumptions, Zhao et al. (2013) puts forward a more general and robust estimation approach combining the idea from Zhao and Tong (2011). In addition, the preceding methods all assume that the covariate effects on the underlying recurrent event process could be captured by the proportional mean model. Li et al. (2010) presents a more flexible estimation approach to allow for a variety of dependence patterns, which gives a class of semiparametric transformation models to include the proportional mean model as a special case.

For the analysis of multivariate panel count data, one may separately apply methods for univariate panel count data to each type of the event. But it would be more efficient to conduct a joint or multivariate analysis if the different types of recurrent events are related. To model the correlation between different types of recurrent event processes, one way is to introduce some frailty variables as with informative follow-ups. The analysis of multivariate panel count data is beyond the scope of this dissertation. The literatures that discuss about this topic include Chen et al. (2005), He et al. (2008), Lee (2008), Li et al. (2011), Zhang et al. (2013), Zhao et al. (2013b), Zhao et al. (2014) and Li et al. (2015).

In practice, one may encounter the situation that some subjects are observed continuously and the others are observed only at discrete times. It could also happen that each subject is observed continuously over certain time periods but at discrete times over other time periods. A mixture of two data structures makes this problem more complicated. A naive solution is to avoid one structure and model only the recurrent event data by imputation or the panel count data by data grouping. But the results could be biased or less efficient. Zhu et al. (2013) defines this problem and presents a marginal mean model, and an estimating equation-based approach is used to estimate the regression parameters. For further considerations of this topic, one may refer to Zhu et al. (2014) and Zhu et al. (2015).

1.3 Outline of the Dissertation

In this Dissertation, we discuss about two special situations, one with recurrent event data and the other with panel count data.

In Chapter 2, we study the semiparametric regression model with random effects for recurrent event data in the presence of informative censoring times. For inference, we propose using the maximum likelihood approach for estimation of the underlying baseline intensity function and regression parameters. The proposed estimates are consistent and have a asymptotic normal distribution. And the maximum likelihood estimators of the regression parameters are asymptotically efficient. The finite sample properties of the proposed estimates are investigated through simulation studies. An illustrative example from a clinical trial is provided.

In Chapter 3, we conduct semiparametric regression analysis of panel count data. To explore the nonlinear interactions between covariates, we propose a class of partially linear models with possibly varying coefficients for the mean function of the counting processes with panel count data. The functional coefficients are estimated by B-spline function approximations. The estimation procedures are based on the maximum pseudo-likelihood and likelihood approaches, and they are easy to implement.The asymptotic properties of the resulting estimators are established, and their finite-sample performance is assessed by Monte Carlo simulation studies. We also demonstrate the value of the proposed method by the analysis of a cancer data set, where the new modeling approach provides more comprehensive information than the usual proportional mean model.

And in Chapter 4, we finalize with possible directions for future research.

Chapter 2

Maximum Likelihood Estimation for Recurrent Event Data with Informative Censoring

2.1 Introduction

Recurrent event data are data in which the event of interest can occur repeatedly. Areas that often produce such data include clinical and longitudinal follow-up studies, reliability experiments, and sociological studies (Andersen et al. 1993; Cook and Lawless 1996, 2007; Cook et al. 1996). Examples include hospitalizations, infections, and tumor metastases. For recurrent event data, two processes are involved: the underlying counting process that characterizes the recurrent process of interest and the follow-up process. Most existing methods have been developed for the analysis of such data by assuming that the follow-up times are independent of the reccurrence of the events completely or given covariates. For example, Andersen and Gill (1982) and Prentice et al. (1981) develope some intensity-based methods, while Lawless and Nadeau (1995) and Lin et al. (2000) propose some marginal mean and rate-based approaches. In particular, Cook and Lawless (2007) provides an excellent review of the existing methods about the analysis of recurrent events.

In practice, the assumption that the recurrent process and the follow-up process are independent may not be true. For example, the follow-up times may be times to some terminal events such as death, which are related to the recurrent events of interest. Wang et al. (2001) consideres a study of AIDS patients in which the recurrent and censoring events are hospitalization and death, respectively. In the presence of dependent censoring events, several approaches have been developed for the analysis of recurrent event data. For instance, Wang et al. (2001), Huang and Wang (2004), Liu et al. (2004) and Ye et al. (2007) develope frailty-based joint modeling procedures that model the recurrent event process and the censoring process together. Ghosh and Lin (2000, 2002) also studies the same problem by using marginal models instead of the frailty model. For estimation of regression parameters in these models, an estimating equation-based approach is used and the resulting estimators may not be efficient. In this chapter, we consider a semiparametric model which allows the censoring process to be correlated with the recurrent event process through an unobserved frailty, and develop an efficient estimation procedure for the proposed model.

The remainder of this chapter is organized as follows. Section 2.2 introduces notation and describes the model for the underlying recurrent event process. To characterize the correlation between the recurrent events and the follow-up or censoring times, we employ an unobserved latent variable in the model. The maximum likelihood estimators (MLE) of regression parameters and the baseline intensity function are obtained and an EM algorithm for these estimators is presented. In Section 2.3, we study the large sample properties of the proposed estimators. The consistency and asymptotic distribution of the MLEs are established. Section 2.4 presents some results obtained from a simulation study for assessing the finite-sample properties of the proposed estimates. In Section 2.5, we apply the proposed methodology to a set of recurrent event data arising from a study of leukemia patients. Some research findings are summarized in Section 2.6.

2.2 Maximum Likelihood Estimation and Algorithm

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 subject i. Suppose that for each subject i, there exists a vector of p-dimensional covariates denoted by X_i, and given X_i and one latent variable U_i, $N_i(t)$ is a non-stationary Poisson process with the intensity function given by

$$\lambda(t|X_i, U_i) = \lambda_0(t)\, \exp(X_i'\gamma + U_i)\,. \tag{2.1}$$

In the model above, $\lambda_0(t)$ is a completely unknown continuous baseline intensity function and γ denotes the vector of regression parameters. Here the frailties $U_1, \ldots, U_n$ are identically and independently distributed with density function $h(u)$. The Gamma frailty is commonly used to model $\exp(U_i)$. The advantage of this setting is that the conditional distribution of $\exp(U_i)$ given the observed data is still a Gamma distribution. Zhang and Jamshidian (2003) uses the Gamma frailty variable to characterize the intracorrelation between the panel counts of the counting process and obtained a maximum pseudo-likelihood estimate. However, its asymptotic properties are unknown. We prefer to use the zero mean normal distribution. That is, $h(u) = \phi(u; \sigma)$, where $\phi(\cdot; \sigma)$ is the density function of the zero-mean normal distribution with variance σ^2. Let $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$ be the cumulative intensity function. For identifiability, we assume that $\Lambda_0(\tau) = 1$ with τ being the largest follow-up time or study time.

For subject i, $i = 1, \ldots, n$, let C_i denote the observed censoring time and let $T_{i,1} < \cdots < T_{i,K_i}$ be the observed event times where $K_i = N_i(C_i \wedge \tau)$ is the total number of the observed recurrent event times. The censoring time C_i is assumed to be independent of $N_i(\cdot)$ conditional on (X_i, U_i). Then the observed data consist of $O_i = (\tilde{T}_i, X_i, C_i), i = 1, \ldots, n$, with $\tilde{T}_i = (T_{i,1}, \ldots, T_{i,K_i})$. Let γ_0 be the true value of

γ. Now we consider the maximum likelihood estimation of γ_0 along with σ and the function Λ_0.

Assume that the observations O_i's from the n subjects are independently and identically distributed realizations of $O = (\tilde{T}, X, C)$ with $\tilde{T} = (T_1, \ldots, T_K)$. Given the observed data $O_i, i = 1, \ldots, n$, the likelihood function of $\theta = (\gamma, \sigma, \Lambda)$ is

$$L_n(\theta|O_i's) \propto \prod_{i=1}^{n} \int f(\tilde{T}_i|X_i, C_i, u_i, \gamma, \Lambda)\phi(u_i; \sigma)du_i, \tag{2.2}$$

where

$$\begin{aligned} & f(\tilde{T}_i|X_i, C_i, u_i, \gamma, \Lambda) \\ & = f(\tilde{T}_i|N_i(C_i), X_i, C_i, u_i, \gamma, \Lambda) f(N_i(C_i)|X_i, C_i, u_i, \gamma, \Lambda) \\ & = \left\{ \prod_{j=1}^{K_i} \frac{\lambda(T_{i,j})}{\Lambda(C_i)} \right\} \frac{\exp\{-\Lambda(C_i)\exp(X_i'\gamma + u_i)\}\{\Lambda(C_i)\exp(X_i'\gamma + u_i)\}^{K_i}}{K_i!}. \end{aligned}$$

Removing the terms that are not related to the parameter θ, we have

$$\begin{aligned} L_n(\theta|O_i's) \propto \quad & \left\{ \prod_{i=1}^{n} \prod_{j=1}^{K_i} \lambda(T_{i,j}) \right\} \times \prod_{i=1}^{n} \int \exp\{-\Lambda(C_i)\exp(X_i'\gamma + u_i)\} \\ & \times \{\exp(X_i'\gamma + u_i)\}^{K_i} \phi(u_i; \sigma) du_i. \end{aligned}$$

Then the maximum likelihood estimator of $\theta_0 = (\gamma_0, \sigma_0, \Lambda_0)$ can be obtained by maximizing the above likelihood function subject to the monotonicity of $\Lambda(\cdot)$, that is, the function $\Lambda(\cdot)$ must be non-decreasing. It would seem natural to calculate the maximum likelihood estimators. However, the maximum of the foregoing likelihood function is infinity, because we can always choose some function $\Lambda(\cdot)$ with fixed values at the $T_{i,j}$ while letting $\lambda(t) = d\Lambda(t)/dt$ go to infinity. Here we relax $\Lambda(\cdot)$ to be right-continuous and allow it to have jumps at $T_{i,j}$. Set $0 < t_1 < \cdots < t_M$ be the ordered distinct time points of $\{T_{i,j}\}$. For $m = 1, \ldots, M$, set $N_m = \sum_{i=1}^{n} \sum_{j=1}^{K_i} I(T_{i,j} = t_m)$, which is the total number that the observation time is exactly t_m. Then the likelihood function can be rewritten as

$$L_n(\theta|O_i's) \propto \left\{\prod_{i=1}^{n}\prod_{j=1}^{K_i}\Delta\Lambda(T_{i,j})\right\} \times \prod_{i=1}^{n}\int \exp\{-\Lambda(C_i)\exp(X_i'\gamma + u_i)\} \times \exp(K_i X_i'\gamma + K_i u_i)\phi(u_i;\sigma)du_i, \tag{2.3}$$

where $\Delta\Lambda(t) = \Lambda(t) - \Lambda(t-)$ and $\Lambda(t) = \sum_{m=1}^{M}\Delta\Lambda(t_m)I(t \geqslant t_m)$. The resulting estimator is referred to as the MLE, denoted by $\hat{\theta}_n = (\hat{\gamma}_n, \hat{\sigma}_n, \hat{\Lambda}_n)$.

EM Algorithm. It is obvious that there exist some difficulties in this maximization procedure. The first is that the number $(M + p + 1)$ of estimated parameters are so large since M is very large, where M is the number of the distinct time points $T_{i,j}'s$. Secondly, the integral has no closed form. We appeal to the EM algorithm. For this, we treat the $U_i's$ as the missing data and the complete likelihood function is given by

$$L_n(\theta|O_i's, U_i's) = \prod_{m=1}^{M}\{\Delta\Lambda(t_m)\}^{N_m}\prod_{i=1}^{n}\exp\{-\Lambda(C_i)\exp(X_i'\gamma + U_i) + K_i(X_i'\gamma + U_i)\}\phi(U_i;\sigma).$$

To implement the EM algorithm, we first consider the E-step, which computes the conditional expectation of the log-likelihood function given the current estimate of θ and the observed data O_i's. To this end, note that the log-likelihood function can be written as

$$\begin{aligned} & l_n(\theta|O_i's, U_i's) \\ &= \sum_{i=1}^{n} l_{ni}(\theta|O_i, U_i) \\ &= \textstyle\sum_{m=1}^{M} N_m \log\Delta\Lambda(t_m) \\ &\quad + \sum_{i=1}^{n}\left\{-\Lambda(C_i)\exp(X_i'\gamma + U_i) + K_i(X_i'\gamma + U_i) + \log\phi(U_i;\sigma)\right\}. \end{aligned} \tag{2.4}$$

Now we consider the maximization of the above expectation, which is M-step. Suppose that at the kth step, the current parameter is $\theta^{(k)} = (\gamma^{(k)}, \sigma^{(k)}, \Lambda^{(k)})$, then the conditional expectation of the "complete" log-likelihood function is

$$\begin{aligned} l_n(\theta|O_i's,\theta^{(k)}) = & \sum_{m=1}^{M} N_m \log \Delta\Lambda(t_m) + \sum_{i=1}^{n}\Big[-\Lambda(C_i)\exp(X_i'\gamma)E\{\exp(U_i)|O_i,\theta^{(k)}\} \\ & +K_i\{X_i'\gamma + E(U_i|O_i,\theta^{(k)})\} + E\{\log\phi(U_i;\sigma)|O_i,\theta^{(k)}\}\Big]. \end{aligned}$$

Therefore, maximizing the above log-likelihood function, one obtains the $(k+1)$th step estimates of θ given by

$$[\Delta\Lambda(t_m)]^{(k+1)} = N_m\Big[\sum_{i=1}^{n} I(C_i \geqslant t_m)\exp(X_i'\gamma^{(k)})E\{\exp(U_i)|O_i,\theta^{(k)}\}\Big]^{-1}, \qquad (2.5)$$

for $m = 1, \cdots, M$,

$$[\sigma^2]^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n} E(U_i^2|O_i,\theta^{(k)}) \qquad (2.6)$$

and $\gamma^{(k+1)}$ satisfying the following equation

$$\sum_{i=1}^{n} X_i\Big[K_i - [\Lambda(C_i)]^{(k)}\exp(X_i'\gamma^{(k+1)})E\{\exp(U_i)|O_i,\theta^{(k)}\}\Big] = 0. \qquad (2.7)$$

Then the MLE of θ can be obtained by the following steps:

Step 1. Provide one initial value $\theta^{(0)}$;

Step 2. At the kth step, we calculate the $(k+1)$th step $\theta^{(k+1)}$ through (2.5), (2.6) and (2.7);

Iterate Step 2 until the desired convergence is achieved.

It is clear that, to calculate (2.5)-(2.7), one needs to calculate $E(U_i^2|O_i,\theta^{(k)})$ and $E(e^{U_i}|O_i,\theta^{(k)})$. Generally,

$$E\{g(U_i)|O_i,\theta^{(k)}\} = \int g(u_i)\, f(u_i|O_i,\theta^{(k)})\, du_i\,,$$

for some function g, where

$$f(u_i|O_i,\theta) \;=\; \frac{\exp\{-\Lambda(C_i)\exp(X_i'\gamma+u_i)\}\exp(K_iu_i)\phi(u_i;\sigma)}{\displaystyle\int \exp\{-\Lambda(C_i)\exp(X_i'\gamma+u_i)\}\exp(K_iu_i)\phi(u_i;\sigma)du_i}\,, \tag{2.8}$$

is the conditional density of U_i given O_i and θ. It is apparent that this integration has no closed form. For this, with $\theta = \theta^{(k)}$, let $\{U_{il}; i=1,...,n, l=1,...,L\}$ be L i.i.d. samples from $N(0,\{\sigma^{(k)}\}^2)$, where L is sufficiently large. Then one can approximate $E\{g(U_i)|O_i,\theta^{(k)}\}$ by

$$\hat{E}\{g(U_i)|O_i,\theta^{(k)}\} \;=\; \frac{\sum_{l=1}^{L} g(U_{il})\exp\{-[\Lambda(C_i)]^{(k)}\exp(X_i'\gamma^{(k)}+U_{il})\}\exp(K_iU_{il})}{\sum_{l=1}^{L}\exp\{-[\Lambda(C_i)]^{(k)}\exp(X_i'\gamma^{(k)}+U_{il})\}\exp(K_iU_{il})}\,. \tag{2.9}$$

2.3 Large Sample Behaviors

In this section, we study the asymptotical properties for the maximum likelihood estimator. First, we impose the following regularity conditions.

C1. The covariate X is uniformly bounded.

C2. There exists some positive constant δ_0 such that $P(C_i \geqslant \tau|X_i,U_i) \geqslant \delta_0$.

C3. $\Lambda_0(\cdot)$ is a strictly increasing function on $[0,\tau]$ and has continuous first derivative function λ_0, satisfying $\Lambda_0(\tau)=1$ and $\lambda_0(0)>0$.

C4. The true value $(\gamma_0',\sigma_0)'$ lies in the interior of a known compact subset Γ of $R^p \times R^+$, where

$$\Gamma = \{(\gamma,\sigma) : |\gamma| \leqslant M_0 \text{ for some constant } M_0,\ \sigma > \sigma_a \text{ for some positive } \sigma_a\}.$$

C5. If there exists a vector γ and a constant c such that

$$X'\gamma = c$$

almost surely, then $\gamma = 0$ and $c = 0$.

Conditions C1-C2 are commonly seen in recurrent event data literature. Condition C3 is made to make the model identifiable. Condition C5 is referred as the identifiability condition, which can guarantee the estimates are unique and consistent. Denote the function class Υ, which satisfies condition C3. Let $\Theta = \Gamma \times \Upsilon$ denote the parameter space. Then the maximum likelihood estimator $\hat{\theta}_n$ is obtained by maximizing (2.3) over $\theta \in \Theta$.

Theorem 2.1. Under conditions C1-C5, $\|\hat{\gamma}_n - \gamma_0\| \to 0$, $|\hat{\sigma}_n^2 - \sigma_0^2| \to 0$ and $\sup_{t\in[0,\tau]} |\hat{\Lambda}_n(t) - \Lambda_0(t)| \to 0$ almost surely.

Theorem 2.1 states that the maximum likelihood estimator is consistent. Furthermore, we have the following theorem about the asymptotical distribution.

Theorem 2.2. Under conditions C1-C5, $n^{1/2}(\hat{\gamma}_n' - \gamma_0', \hat{\sigma}_n^2 - \sigma_0^2, \hat{\Lambda}_n(t) - \Lambda_0(t))'$ converges weakly to a zero mean Gaussian process in the metric space $R^{p+1} \times l^\infty[0,\tau]$, where $l^\infty[0,\tau]$ is a normed space consisting of all the functions in Υ, where the norm is taken as the supremum norm on $[0,\tau]$. In addition, $\hat{\gamma}_n$ and $\hat{\sigma}_n^2$ are asymptotically efficient.

Theorems 2.1 and 2.2 are proven in Appendix A.

Now we turn to consider estimation of variance of $(\hat{\gamma}_n, \hat{\sigma}_n)$. Intuitively, using the idea similar to Zeng et al. (2005, 2006), we can treat the distinct jump size $\Delta\Lambda(t_m)$ as the parameters and then the observed likelihood function is regarded as the function of these parameters and (γ, σ). Then the asymptotical covariance matrix can be estimated by the inverse of the observed information matrix for all parameters. The observed information matrix can be calculated via the Louis (1982) formula.

2.4 Numerical results

This section will report some results obtained from a simulation study conducted for assessing the performance of the inference procedures proposed in the previous sections. In the study, we considered situations where there exists one covariate X_i. It follows the Bernoulli distribution with the success probability 0.5, or follows the uniform distribution U(0, 1). The latent variable U_i was generated from the zero mean normal distribution with the variance of 0.5^2. The censoring time C_i was generated from the minimum of $U(2, \tau + 1)$ and τ, with $\tau = 4$. With given X_i and U_i, we generated the observation process $N_i(t)$ under model (2.1) with different values of γ and $\lambda_0(t)$. Here γ can be -0.2, 0, 0.5 or 1, and $\lambda_0(t) = 2.5$ or $\lambda_0(t) = 1.25t$. The results reported below are based on 1000 replications with sample size $n = 100$ or 200.

Tables 2.1 and 2.2 present the simulation results. Each table includes the averages of the estimates $\hat{\gamma}$ and $\hat{\sigma}$, the sample standard deviations of the estimates (SSD), the averages of the estimated standard errors (ESE), and the 95% empirical coverage probabilities (CP). These results indicate that the proposed estimation procedure seems to work well and in addition, the normal approximation to the distributions of $\hat{\gamma}$ and $\hat{\sigma}$ seems to perform well. We also considered some other set-ups and obtained similar results.

2.5 Analysis of AML Data

Now we apply the methodology developed in the previous sections to a set of recurrent event data arising from a clinical trial on the patients with acute myeloid leukemia (AML) (Rubnitz et al., 2005). The data set consists of 201 subjects who were consecutively treated for their AML from October 2002 through October 2008. During the chemotherapy course, the patients may experience repeated bacterial, viral or fungal infections, and one of the objectives of the study was to investigate the infection rates and its relationship with various predictor variables. In the study we combine the three types of infections and Table 2.3 provides a summary of the frequency of all infections among the 201 patients. In total, there were 391 events during their chemotherapy with a median risk time as 160 days. During the study, 38 patients experienced the relapse, transplant or death, which will be treated as informative censoring. For the study, there are six predictor variables or baseline covariates of interest. They are gender (male = 0; female = 1), the leukemia risk level (low = 0; standard = 1; high = 2), the dose of cytarabine given for the first course of chemotherapy (standard = 0; high = 1), and race (white = 0; others = 1) along with the white blood count (WBC) and the age both at the diagnosis of AML. Table 2.4 presents the analysis results obtained by applying the proposed estimation procedure to the data here. It includes the estimated effect of each covariate on the occurrence rate of the infections, the estimated standard error (ESE), and the p-value for testing no significant covariate effect. These results indicate that age seems to have significant effect on the infection rate. More specifically, younger patients suffer more infections than older patients. All other covariates do not have a significant effect on the infection rate. Figure 2.1 presents the estimates of the cumulative baseline intensities along with the pointwise 95% confidence bands.

For comparison, we also apply the method given in Wang et al. (WQC) (2001), which accounts for the informative censoring also by a frailty-based model but estimates parameters with estimating equations. The results are also included in Table 2.4.

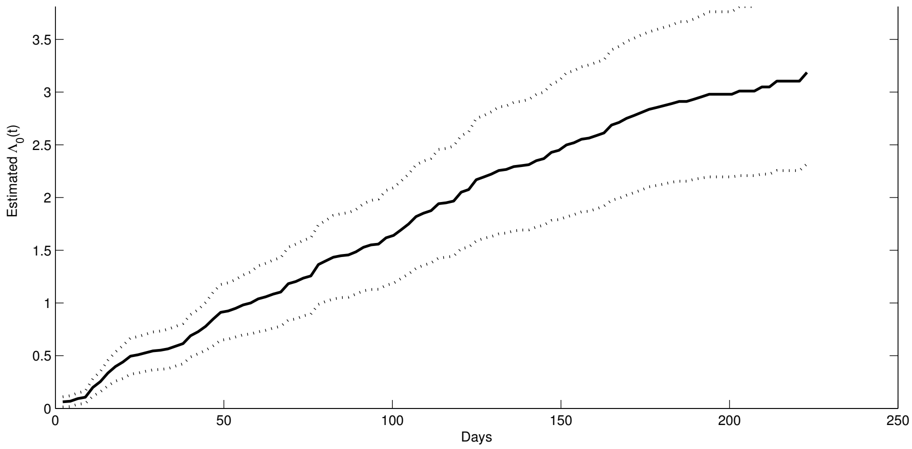

Figure 2.1: Estimates of $\Lambda_0(t)$ with the pointwise 95% confidence bands.

2.6 Summary

In this Chapter, we considered regression analysis of recurrent event data when the underlying recurrent event process and the follow-up or censoring times involved may be related conditional on covariates. For the purpose, some shared frailty models were proposed. For inference, the maximum likelihood approach and an EM algorithm were developed for estimation of regression parameters representing covariate effects and the baseline intensity function of the underlying recurrent event process. The proposed estimators for regression parameters achieve the asymptotic efficiency.

Table 2.1: Estimation of γ and σ for n=100

$\lambda(t)$	X	γ	$\hat{\gamma}$	SSD	ESE	CP	$\hat{\sigma}$	SSD	ESE	CP
2.5	0/1	-0.2	-0.186	0.127	0.121	0.932	0.475	0.055	0.054	0.924
		0	0.012	0.122	0.119	0.939	0.477	0.054	0.053	0.932
		0.5	0.513	0.117	0.116	0.942	0.484	0.049	0.050	0.929
		1	1.021	0.122	0.114	0.929	0.487	0.048	0.047	0.928
	U(0,1)	-0.2	-0.183	0.213	0.211	0.953	0.477	0.055	0.054	0.926
		0	0.024	0.205	0.208	0.951	0.481	0.051	0.052	0.936
		0.5	0.535	0.207	0.203	0.935	0.490	0.050	0.050	0.935
		1	1.045	0.201	0.197	0.941	0.490	0.047	0.047	0.940
1.25t	0/1	-0.2	-0.173	0.136	0.129	0.929	0.486	0.057	0.058	0.941
		0	0.004	0.131	0.126	0.940	0.495	0.057	0.056	0.941
		0.5	0.502	0.131	0.125	0.935	0.510	0.054	0.054	0.957
		1	1.002	0.125	0.122	0.945	0.514	0.051	0.051	0.955
	U(0,1)	-0.2	-0.178	0.231	0.220	0.929	0.489	0.055	0.057	0.959
		0	0.021	0.230	0.219	0.937	0.495	0.055	0.057	0.955
		0.5	0.524	0.218	0.213	0.941	0.505	0.053	0.053	0.958
		1	1.026	0.221	0.211	0.931	0.516	0.052	0.051	0.946

Table 2.2: Estimation of γ and σ for n=200

$\lambda(t)$	X	γ	$\hat{\gamma}$	SSD	ESE	CP	$\hat{\sigma}$	SSD	ESE	CP
2.5	0/1	-0.2	-0.187	0.086	0.085	0.947	0.479	0.038	0.038	0.929
		0	0.010	0.088	0.085	0.941	0.487	0.037	0.037	0.932
		0.5	0.509	0.083	0.082	0.934	0.489	0.036	0.035	0.921
		1	1.022	0.078	0.081	0.941	0.496	0.035	0.033	0.949
	U(0,1)	-0.2	-0.171	0.154	0.148	0.935	0.482	0.036	0.038	0.925
		0	0.031	0.151	0.147	0.944	0.485	0.038	0.038	0.933
		0.5	0.505	0.156	0.138	0.926	0.493	0.036	0.036	0.936
		1	1.039	0.148	0.139	0.930	0.495	0.032	0.033	0.949
1.25t	0/1	-0.2	-0.185	0.094	0.090	0.938	0.494	0.041	0.041	0.951
		0	0.005	0.092	0.089	0.946	0.502	0.040	0.040	0.946
		0.5	0.505	0.088	0.088	0.950	0.511	0.038	0.038	0.942
		1	1.005	0.097	0.094	0.933	0.519	0.036	0.039	0.937
	U(0,1)	-0.2	-0.173	0.155	0.155	0.948	0.493	0.041	0.041	0.941
		0	0.017	0.157	0.154	0.938	0.500	0.040	0.040	0.951
		0.5	0.503	0.158	0.146	0.921	0.513	0.039	0.038	0.936
		1	1.028	0.154	0.149	0.942	0.519	0.035	0.036	0.935

Table 2.3: Frequencies of infections in the AML study

Infection	0	1	2	3	$\geqslant 4$
Frequncy	44	55	35	28	39

Table 2.4: Estimates of covariate effects for the AML study

	Proposed method			WQC's method		
Covariate	Estimate	ESE	p-value	Estimate	ESE	p-value
Gender	-0.0929	0.1170	0.4273	-0.0481	0.1195	0.6875
Risk	-0.0827	0.0764	0.2791	-0.1122	0.0799	0.1604
Dose	0.0630	0.1168	0.5897	0.0895	0.1216	0.4617
Race	0.0239	0.1285	0.8522	0.0533	0.1330	0.6887
WBC	-0.0006	0.0007	0.3947	-0.0009	0.0008	0.2547
Age	-0.0212	0.0100	0.0330	-0.0206	0.0113	0.0685

Chapter 3

Semiparametric Partially Linear Varying Coefficient Models with Panel Count Data

3.1 Introduction

This chapter considers regression analysis of panel count data when certain covariate effects may be much more complex than linear effects. By panel count data, we mean the data that concern occurrence rates of certain recurrent events and give only the numbers of the events that occur between the observation times, but not their occurrence times. Such data naturally occur in longitudinal follow-up studies on recurrent events in which study subjects can be observed only at discrete time points rather than continuously. Many authors have discussed the analysis of panel count data by using nonparametric and semiparametric methods. For example, Sun and Kalbfleisch (1995), Wellner and Zhang (2000), Zhang and Jamshidian (2003), Lu et al. (2007), and Hu et al. (2009a) study nonparametric estimation for the mean function of the counting process with panel count data; Thall and Lachin (1988), Sun and Fang (2003), Zhang (2006), and Balakrishnan and Zhao (2009) propose some nonparametric tests for the problem of nonparametric comparison of treatment groups based on panel count data. Sun and Wei (2000), Cheng and Wei (2000), Hu et al. (2003), and Hua and Zhang (2011) discuss regression analysis of panel count data by the estimating equation-based approaches, while Zhang (2002), Wellner and Zhang (2007), and Lu et al. (2009) present more efficient inference procedures for joint estimation of parametric and nonparametric components in the proportional mean model by the likelihood-based approaches. In addition, Huang et al. (2006) and Sun et al. (2007) consider the analysis of panel count data with informative observation times.

All these semiparametric regression methods mentioned above have focused on parametric modeling of covariate effects on the recurrent event process. In many applications, a covariate effect may be nonlinear and vary with another covariate. To investigate both linear effects and nonlinear interaction effects between covariates, we propose a class of semiparametric partially linear varying-coefficient models for panel count data. Suppose that $N(t)$ is a counting process arising from a recurrent event study. Let Z be a d-dimensional vector of covariates, and V and W be p-dimensional vectors of covariates. We assume that given (Z, V, W), $N(t)$ is a non-homogeneous Poisson process with the mean function $\Lambda(t|Z, V, W) = E\{N(t)|Z, V, W\}$ having the following form

$$\Lambda(t|Z, V, W) = \Lambda_0(t) \exp\left\{ Z'\beta + \sum_{r=1}^{p} V_r \phi_r(W_r) \right\}, \tag{3.1}$$

where $\Lambda_0(\cdot)$ is a completely unknown continuous baseline mean function, $\phi_r(\cdot)$ ($r = 1, \ldots, p$) are completely unspecified smooth functions, and β is a d-dimensional vector of unknown regression parameters. When $V_r = 0$ ($r = 1, \ldots, p$), the model reduces to linear regression model with panel count data, which has been well studied by Wellner and Zhang (2007) and Lu et al. (2009), among others. When $V_r = 1$ ($r = 1, \ldots, p$), the model reduces to partly linear regression model for panel count data, which has not been studied in the literature. There are many investigations about nonlinear effects of covariates on response variables for censored data and longitudinal data. For example, Zhang et al. (2014) studies a proportional hazards model with varying coefficients for right-censored and length-biased data; Lindqvist et al. (2015) examines the functional form for covariates in parametric accelerated failure time models with right-censored data by using residual plots; Cheng et al. (2014) provides a simultaneous variable selection and structure identification procedure for ultra-high dimensional longitudinal data. For notational simplicity, we consider the case with $p = 1$, that is,

$$\Lambda(t|Z, V, W) = \Lambda_0(t) \exp\{Z'\beta + V\phi(W)\}. \tag{3.2}$$

For inference about model (3.2), we propose to use likelihood-based methods, where the functional coefficient is estimated by the B-spline function approximation, and the baseline mean function is still directly estimated with parametric components because its B-spline function approximation has some nonlinear restriction that can cause more complicated computing. For this reason, we develop a new algorithm which can be easily implemented.

The remainder of this chapter is organized as follows. In Section 3.2, we present two semiparametric methods including maximum pseudo-likelihood and maximum likelihood approaches for joint estimation of parametric and nonparametric components in the model, and also provide corresponding algorithms about computation of the estimates. The asymptotic properties of the resulting estimators are established in Section 3.3, while the proofs are given in Appendix B. Section 3.4 reports some simulation results obtained for assessing the finite sample properties of the proposed estimates and an illustrative example is given in Section 3.5. Some remarks are made in Section 3.6.

3.2 Semiparametric Likelihood Approaches

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 subject i. Suppose that for each subject, given covariates (Z_i, V_i, W_i), $N_i(t)$ is a non-homogeneous Poisson process with the mean function given by (3.2), that is,

$$P\{N_i(t) = k | Z_i, V_i, W_i\} = \exp\{-\Lambda_i(t|Z_i, V_i, W_i)\} \frac{\{\Lambda_i(t|Z_i, V_i, W_i)\}^k}{k!},$$

where $\Lambda_i(t|Z_i, V_i, W_i) = \Lambda_0(t) \exp\{Z_i'\beta + V_i\phi(W_i)\}$. For subject i, suppose that $N_i(\cdot)$ is observed only at finite time points $T_{K_i,1} < \cdots < T_{K_i,K_i} \leqslant \tau$, where K_i denotes the potential number of observation times, $i = 1, \ldots, n$, and τ is the length of the study. 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.

In the following, we will assume that given (Z_i, V_i, W_i), $(K_i; T_{K_i,1}, ..., T_{K_i,K_i})$ are independent of the counting processes N_i's. Let $\mathbf{X} = (K, \mathbf{T}, \mathbf{N}, Z, V, W)$, where $\mathbf{T} = (T_{K,1}, ..., T_{K,K})$ and $\mathbf{N} = (N(T_{K,1}), \ldots, N(T_{K,K}))$. Then $\{\, \mathbf{X}_i = (K_i, \mathbf{T}_i, \mathbf{N}_i), Z_i, V_i, W_i\; i = 1, ..., n\}$ is a random sample of size n from the distribution of $\mathbf{X}$, where $\mathbf{T}_i = (T_{K_i,1}, ..., T_{K_i,K_i})$ and $\mathbf{N}_i = (N_i(T_{K_i,1}), \ldots, N_i(T_{K_i,K_i}))$.

Without loss of generality, assume that W has support on $[0, 1]$. For estimation of the smooth function ϕ, we use B-spline function approximation. We first introduce some notation (Huang, 1999). Let $\mathcal{T} = \{s_i, i = 1, \ldots, m_n + 2l\}$, with

$$0 = s_1 = \cdots = s_l < s_{l+1} < \cdots < s_{m_n+l} < s_{m_n+l+1} = \cdots = s_{m_n+2l} = 1,$$

be a sequence of knots that partition $[0, 1]$ into $m_n + 1$ subintervals $I_i = [s_{l+i}, s_{l+i+1}]$, for $i = 0, 1, \ldots, m_n$. Define Φ_n the class of polynomial splines of order $l \geqslant 1$ with the knot sequence $\mathcal{T}$. Then Φ_n can be linearly spanned by the normalized B-spline basis functions $\{b_i, i = 1, \ldots, b_{q_n}\}$ with $q_n = m_n + l$ (Schumaker, 1981). Let $B_n = (b_1, \cdots, b_{q_n})'$. Then we can approximate ϕ by $\phi_n = B_n'\alpha$, where α is a q_n-dimensional vector of unknown coefficients.

3.2.1 Maximum Pseudo-likelihood Approach

The log pseudo-likelihood function for β, Λ, and ϕ is

$$\begin{aligned} l_n^{ps}(\beta, \Lambda, \phi) &= \sum_{i=1}^{n}\sum_{j=1}^{K_i}[N_i(T_{K_i,j})\log\{\Lambda(T_{K_i,j})\} + N_i(T_{K_i,j})\{Z_i'\beta + V_i\phi(W_i)\} \\ &\qquad -\Lambda(T_{K_i,j})\exp\{Z_i'\beta + V_i\phi(W_i)\}] \end{aligned}$$

after omitting the parts independent of β, Λ, and ϕ.

Let $t_1 < \cdots < t_m$ denote the ordered distinct observation time points in the set of all observation time points $\{T_{K_i,j}, j = 1, \ldots, K_i, i = 1, \ldots, n\}$. Let w_ℓ and $\bar{N}_\ell$ be the number and mean value, respectively, of the observations made at time t_ℓ, $\ell = 1, ..., m$, that is,

$$w_\ell = \sum_{i=1}^{n}\sum_{j=1}^{K_i} I(T_{K_i,j} = t_\ell) \quad \text{and} \quad \bar{N}_\ell = \frac{1}{w_\ell}\sum_{i=1}^{n}\sum_{j=1}^{K_i} N_i(T_{K_i,j})I(T_{K_i,j} = t_\ell).$$

Define

$$\bar{A}_\ell(\beta, \phi) = \frac{1}{w_\ell}\sum_{i=1}^{n}\sum_{j=1}^{K_i}\exp\{Z_i'\beta + V_i\phi(W_i)\}I(T_{K_i,j} = t_\ell)$$

and

$$\bar{B}_\ell(\beta, \phi) = \frac{1}{w_\ell}\sum_{i=1}^{n}\sum_{j=1}^{K_i} N_i(T_{K_i,j})\{Z_i'\beta + V_i\phi(W_i)\}I(T_{K_i,j} = t_\ell).$$

Then $l_n^{ps}(\beta, \Lambda, \phi)$ can be expressed as

$$l_n^{ps}(\beta, \Lambda, \phi) = \sum_{\ell=1}^{m} w_\ell\{\bar{N}_\ell \log\Lambda_\ell - \bar{A}_\ell(\beta, \phi)\Lambda_\ell + \bar{B}_\ell(\beta, \phi)\},$$

where $\Lambda_\ell = \Lambda(t_\ell)$, $\ell = 1, \ldots, m$.

Let $\mathcal{R} \subset \mathbb{R}^d$ be a bounded closed set, and let

$$\mathcal{F} = \{\Lambda : \Lambda \text{ is a nondecreasing function over } [0, \tau], \Lambda(0) = 0\},$$

and

$$\Psi_n = \{\phi : \phi = B_n' \alpha \in \Phi_n, ||\phi||_\infty \leqslant M_0\}$$

where τ is the maximum follow-up time of the study and M_0 is a constant. Let $\theta = (\beta, \Lambda, \phi)$, and $\hat{\theta}_n^{ps} = (\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}, \hat{\phi}_n^{ps})$ be the value that maximizes $l_n^{ps}(\theta)$ with respect to $\theta \in \Theta_n = \{\mathcal{R} \times \mathcal{F} \times \Psi_n\}$. Following Wellner and Zhang (2000, 2007), we define the estimator $\hat{\Lambda}_n^{ps}$ to have jumps only at the observation time points to meet with uniqueness since $l_n^{ps}(\beta, \Lambda, \phi)$ depends on Λ only at the observation time points.

We denote the estimator of ϕ by $\hat{\phi}_n^{ps} = B_n' \hat{\alpha}_n^{ps}$. Following Zhang (2002), one can find the solution of $(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}, \hat{\alpha}_n^{ps})$.

Step 1. Choose an initial $(\beta^{(0)}, \alpha^{(0)})$.

Step 2. For given $(\beta^{(k)}, \alpha^{(k)})$ $(k = 0, 1, 2, ...)$, compute

$$\Lambda_\ell^{(k)} = \max_{i \leqslant \ell} \min_{j \geqslant \ell} \frac{\sum_{i \leqslant r < j} w_r \bar{N}_r}{\sum_{i \leqslant r < j} w_r \bar{A}_r(\beta^{(k)}, \alpha^{(k)})}, \quad \ell = 1, \ldots, m.$$

Step 3. Update (β, α) by finding

$$(\beta^{(k+1)}, \alpha^{(k+1)}) = \text{argmax}_{(\beta,\alpha) \in \mathbb{R}^{d+q_n}} \hat{l}_n^{ps}(\beta, \alpha, \Lambda^{(k)})$$

through the Newton-Raphson algorithm, where

$$\hat{l}_n^{ps}(\beta, \alpha, \Lambda) = \sum_{\ell=1}^m w_\ell \{\bar{B}_\ell(\beta, \alpha) - \bar{A}_\ell(\beta, \alpha) \Lambda_\ell\}.$$

Step 4. Repeat Steps 2 and 3 until the convergence is achieved.

3.2.2 Maximum Likelihood Approach

The log-likelihood function for β, Λ, and ϕ is

$$
\begin{aligned}
l_n(\beta,\Lambda,\phi) \quad = \quad & \sum_{i=1}^{n}\sum_{j=1}^{K_i}\left[\{N_i(T_{K_i,j}) - N_i(T_{K_i,j-1})\}\log\{\Lambda(T_{K_i,j}) - \Lambda(T_{K_i,j-1})\}\right. \\
& + \{N_i(T_{K_i,j}) - N_i(T_{K_i,j-1})\}\{Z_i'\beta + V_i\phi(W_i)\} \\
& \left. -\{\Lambda(T_{K_i,j}) - \Lambda(T_{K_i,j-1})\}\exp\{Z_i'\beta + V_i\phi(W_i)\}\right]
\end{aligned}
$$

after omitting the parts independent of β, Λ, and ϕ, where $T_{K_i,0} = 0$.

Let $(\hat{\beta}_n, \hat{\Lambda}_n, \hat{\phi}_n)$ be the value that maximizes $l_n(\beta,\Lambda,\phi)$ with respect to $(\beta,\Lambda,\phi) \in \Theta_n$. Similarly, the estimator $\hat{\Lambda}_n$ is defined to have jumps only at the observation time points. This estimator can be computed by the algorithm proposed by Wellner and Zhang (2007), but it is computationally expensive. Here, we propose a new algorithm by using the self-consistent algorithm (Hu et al., 2009b).

Define $\lambda_\ell = \Lambda(t_\ell) - \Lambda(t_{\ell-1})$, $\Delta N_i(t_\ell) = N_i(t_\ell) - N_i(t_{\ell-1})$, and $Y_i(t) = I(t \leqslant T_{K_i,K_i})$. Let

$$R_i(t_\ell) = \min\{T_{K_i,j}, j = 1,\ldots,K_i; T_{K_i,j} \geqslant t_\ell\}$$

and

$$L_i(t_\ell) = \max\{T_{K_i,j}, j = 1,\ldots,K_i; T_{K_i,j} < t_\ell\}$$

denote the most recent observation times of individual i not before and before t_ℓ, respectively. Here $R_i(t_\ell) = t_{m+1} = \infty$ if $t_\ell > T_{K_i,K_i}$. Define $\tilde{\Delta}N_i(t_\ell) = N_i(R_i(t_\ell)) - N_i(L_i(t_\ell))$ and $\tilde{\Delta}\Lambda_i(t_\ell) = \Lambda(R_i(t_\ell)) - \Lambda(L_i(t_\ell))$, that is, $\tilde{\Delta}\Lambda_i(t_\ell) = \sum_{r:\, L_i(t_\ell) < t_r \leqslant R_i(t_\ell)} \lambda_r$. For given β and ϕ, we have the following estimating equation for Λ_0:

$$\sum_{i=1}^{n} Y_i(t_\ell)\left[\lambda_\ell \frac{\tilde{\Delta}N_i(t_\ell)}{\tilde{\Delta}\Lambda_i(t_\ell)} - \lambda_\ell \exp\{Z_i'\beta + V_i\phi(W_i)\}\right] = 0, \quad \ell = 1,\ldots,m.$$

As Hu et al. (2009b) points out, the estimating functions are unbiased and also can be viewed as the expectation of the likelihood estimating functions conditional on panel counts.

We denote the estimators of ϕ by $\hat{\phi}_n = B_n'\hat{\alpha}_n$. To find out the solution of $(\hat{\beta}_n, \hat{\alpha}_n, \hat{\Lambda}_n)$, we propose to implement the following algorithm.

Step 1. Choose the initial $(\beta^{(0)}, \alpha^{(0)}) = (\hat{\beta}_n^{ps}, \hat{\alpha}_n^{ps})$.

Step 2. For given $(\beta^{(k)}, \alpha^{(k)})$, obtain $\lambda_\ell^{(k)}$ $(\ell = 1, \ldots, m)$ by computing

$$\lambda_\ell^{(k,u)} = \frac{\sum_{i=1}^n Y_i(t_\ell)\lambda_\ell^{(k,u-1)}\tilde{\Delta}N_i(t_\ell)/\tilde{\Delta}\Lambda_i^{(k,u-1)}(t_\ell)}{\sum_{i=1}^n Y_i(t_\ell)\exp\{Z_i'\beta^{(k)} + V_iB_n(W_i)'\alpha^{(k)}\}}$$

for $u = 1, 2, \ldots$ until the convergence is achieved. Here we choose $\Lambda^{(0,0)} = \hat{\Lambda}_n^{ps}$ and $\Lambda^{(k,0)} = \Lambda^{(k-1)}$ for $k \geqslant 1$.

Step 3. Update (β, α) by finding

$$(\beta^{(k+1)}, \alpha^{(k+1)}) = \text{argmax}_{(\beta,\alpha)\in\mathbb{R}^{d+q_n}}\hat{l}_n(\beta, \alpha, \Lambda^{(k)})$$

through the Newton-Raphson algorithm, where

$$\begin{aligned}\hat{l}_n(\beta, \alpha, \Lambda) &= \sum_{i=1}^n\sum_{j=1}^{K_i}\left[\{N_i(T_{K_i,j}) - N_i(T_{K_i,j-1})\}\{\beta'Z_i + V_iB_n(W_i)'\alpha\}\right.\\ &\quad \left. -\{\Lambda(T_{K_i,j}) - \Lambda(T_{K_i,j-1})\}\exp\{\beta'Z_i + V_iB_n(W_i)'\alpha\}\right]\end{aligned}$$

Step 4. Repeat Steps 2 and 3 until the convergence is achieved.

3.3 Asymptotic Results

In this section, we study the asymptotic properties of the estimators $\hat{\theta}_n^{ps} = (\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}, \hat{\phi}_n^{ps})$ and $\hat{\theta}_n = (\hat{\beta}_n, \hat{\Lambda}_n, \hat{\phi}_n)$ of $\theta = (\beta, \Lambda, \phi)$. Let $\mathcal{B}_{d+2}$ and $\mathcal{B}$ denote the collection of Borel sets in $\mathbb{R}^{d+2}$ and $\mathbb{R}$, respectively, and let $\mathcal{B}_{[0,\tau]} = \{A \cap [0, \tau] : A \in \mathcal{B}\}$. Let $Y = (Z', V, W)'$ with distribution function $F(y)$. Following Wellner and Zhang

(2007), define the measures $\mu_1, \mu_2, \nu_1, \nu_2$, and γ as follows: for $A, A_1, A_2 \in \mathcal{B}_{[0,\tau]}$, and $A_3 \in \mathcal{B}_{d+2}$,

$$\nu_1(A \times A_3) = \int_{A_3} \sum_{k=1}^{\infty} P(K = k|Y = y) \sum_{j=1}^{k} P(T_{k,j} \in A|K = k, Y = y) dF(y),$$

$$\mu_1(A) = \nu_1(A \times \mathbb{R}^{d+2}),$$

$$\nu_2(A_1 \times A_2 \times A_3)$$
$$= \int_{A_3} \sum_{k=1}^{\infty} \{P(K = k|Y = y) \sum_{j=1}^{k} P(T_{k,j-1} \in A_1, T_{k,j} \in A_2|K = k, Y = y)\} dF(y),$$

$$\mu_2(A_1 \times A_2) = \nu_2(A_1 \times A_2 \times \mathbb{R}^{d+2}),$$

$$\gamma(A) = \int_{\mathbb{R}^{d+2}} \sum_{k=1}^{\infty} P(K = k|Y = y) \sum_{j=1}^{k} P(T_{k,k} \in A|K = k, Y = y) dF(y).$$

We also define the L_2-metrics d_1 and d_2 as

$$d_1(\theta_1, \theta_2) = \left\{ ||\beta_1 - \beta_2||^2 + \int |\Lambda_1(t) - \Lambda_2(t)|^2 d\mu_1(t) + E|\phi_1(W) - \phi_2(W)|^2 \right\}^{1/2},$$

and

$$d_2(\theta_1, \theta_2) = \left\{ ||\beta_1 - \beta_2||^2 + \int\int |(\Lambda_1(u) - \Lambda_1(v)) - (\Lambda_2(u) - \Lambda_2(v))|^2 d\mu_2(u, v) + E|\phi_1(W) - \phi_2(W)|^2 \right\}^{1/2}.$$

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

C1. The maximum spacing of the knots, $\max_{l+1 \leqslant i \leqslant m_n+l+1} |s_i - s_{i-1}| = O(n^{-v})$ with $m_n = O(n^v)$ for $0 < v < 0.5$.

C2. The true parameter $\theta_0 = (\beta_0, \Lambda_0, \phi_0) \in \mathcal{R}^0 \times \mathcal{F} \times \mathcal{F}_r$ with $\Lambda_0(\tau) \leqslant M$ for a constant $M > 0$, and $r = l + a > 0.5$, where

$$\mathcal{F}_r = \{g(\cdot) : |g^{(l)}(w_1) - g^{(l)}(w_2)| \leqslant M_0|w_1 - w_2|^a \text{ for all } 0 \leqslant w_1, w_2 \leqslant 1\}$$

and $g^{(l)}$ is the lth derivative function of g.

C3. The measure $\mu_i \times F$ is absolutely continuous with respect to ν_i, for $i = 1, 2$.

C4. The function M_0^{ps} defined by $M_0^{ps}(X) = \sum_{j=1}^{K} N(T_{K,j}) \log(N(T_{K,j}))$ satisfies $PM_0^{ps}(X) < \infty$.

C5. The function M_0 defined by $M_0(X) = \sum_{j=1}^{K} \Delta N(T_j) \log(\Delta N(T_{K,j}))$ satisfies $PM_0(X) < \infty$.

C6. $\mathcal{C} = \text{supp}(F)$, is a bound set in $\mathbb{R}^{d+2}$. Thus there exist z_0 and v_0 such that $P(|Z| \leqslant z_0) = 1$ and $P(|V| \leqslant v_0) = 1$. That is, the covariates Z and V are uniformly bounded.

C7. If with probability 1, $Z'b + V\psi(W) + \zeta(T_{K,K}) = 0$ for some b, ψ and ζ, then $b = 0$, $\psi = 0$ and $\zeta = 0$.

C8. There exists a positive integer K_0 such that $P(K \leqslant K_0)$=1.

Conditions C1 and C2 are common assumptions in semiparametric estimation problems. Conditions C4-C6 and C8 similar to those required by Wellner and Zhang (2007). Conditions C3 and C7 are needed for identifiability of the model.

Theorem 3.1. (Consistency). Suppose that conditions C1-C8 hold.

(i) If $\mu_1([b, \tau]) > 0$ for $0 < b < \tau$, then

$$\lim_{n\to\infty} d_1((\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps} 1_{[0,b]}, \hat{\phi}_n^{ps}), (\beta_0, \Lambda_0 1_{[0,b]}, \phi_0)) = 0 \quad \text{in Probability.}$$

If $\mu_1(\{\tau\}) > 0$,

$$\lim_{n\to\infty} d_1(\hat{\theta}_n^{ps}, \theta_0) = 0 \quad \text{in Probability.}$$

(ii) If $\gamma([b,\tau]) > 0$ for $0 < b < \tau$, then

$$\lim_{n\to\infty} d_2((\hat{\beta}_n, \hat{\Lambda}_n 1_{[0,b]}, \hat{\phi}_n), (\beta_0, \Lambda_0 1_{[0,b]}, \phi_0)) = 0 \quad \text{in Probability.}$$

If $\gamma(\{\tau\}) > 0$, then

$$\lim_{n\to\infty} d_2(\hat{\theta}_n, \theta_0) = 0 \quad \text{in Probability.}$$

To establish the rate of convergence and the asymptotic normality, we need additional conditions.

C9. For some positive constant c_0, $E[\exp\{c_0 N(\tau)\}] < \infty$.

C10. $P(\cap_{j=1}^K \{T_{K,j} \in [\tau_0, \tau]\}) = 1$ with $\tau_0 > 0$ and $\Lambda_0(\tau_0) > 0$.

C11. There exists a positive constant s_0 such that

$$P(\min_{1\leqslant j\leqslant K}\{T_{K,j} - T_{K,j-1}\} \geqslant s_0) = 1.$$

C12. Λ_0 is differentiable and the derivative has a positive and finite lower and upper bounds in $[\tau_0, \tau]$.

C13. There exists $\eta_1, \eta_2 \in (0,1)$ such that $a'Var(Z|U,V,W)a \geqslant \eta_1 a'E(Z'Z|U,V,W)a$ a.s. for all $a \in \mathbb{R}^d$, and $Var(V|U,W) \geqslant \eta_2 E(V^2|U,W)$, where (U,Y) has distribution $\nu_1/\nu_1(\mathbb{R}^+ \times \mathcal{C})$.

C14. There exists $\eta_1, \eta_2 \in (0,1)$ such that

$$a'Var(Z|U_1,U_2,Y)a \geqslant \eta_1 E(Z'Z|U_1,U_2,Y)a, \; a.s.$$

for all $a \in \mathbb{R}^d$, and $Var(V|U_1, U_2, W) \geqslant \eta_2 E(V^2|U_1, U_2, W)$, where (U_1, U_2, Y) has distribution $\nu_2/\nu_2(\mathbb{R}^{+2} \times \mathcal{C})$.

Conditions C9-C14 and their justifications are similar to those given in Wellner and Zhang (2007).

Theorem 3.2. (Rate of Convergence). Suppose that conditions C1-C10 hold.

(i) If condition C13 holds, then $n^{\frac{1-v}{3}} d_1(\hat{\theta}_n^{ps}, \theta_0) = O_p(1)$.

(ii) If conditions C11, C12 and C14 hold, then $n^{\frac{1-v}{3}} d_2(\hat{\theta}_n, \theta_0) = O_p(1)$.

Theorem 3.3. (Asymptotic Normality). Suppose that $\frac{1}{6r-2} < v < \frac{1}{4}$ with $r > 1$ and the conditions C1-C12 hold. Define

$$\mathcal{H} = \left\{ (\mathbf{h}_1, h_2, h_3) : \begin{array}{l} \mathbf{h}_1 \in \mathbb{R}^d, ||\mathbf{h}_1|| \leqslant 1, \ h_3 \in \mathcal{F}_r \\ h_2 \text{ is a function with bounded total variation in } [0, \tau] \\ h_2(0) = 0 \end{array} \right\}.$$

(i) If condition C13 holds, then for $(\mathbf{h}_1, h_2, h_3) \in \mathcal{H}$,

$$\mathbf{h}_1' \sqrt{n}(\hat{\beta}_n^{ps} - \beta_0) + \sqrt{n} \int \{\hat{\Lambda}_n^{ps}(t) - \Lambda_0(t)\} dh_2(t) + \sqrt{n} \int \{\hat{\phi}_n^{ps}(w) - \phi_0(w)\} dh_3(w)$$

$$\rightarrow_d N(0, \sigma_{ps}^2),$$

where σ_{ps}^2 is given in (B.5).

(ii) If condition C14 holds, then for $(\mathbf{h}_1, h_2, h_3) \in \mathcal{H}$,

$$\mathbf{h}_1' \sqrt{n}(\hat{\beta}_n - \beta_0) + \sqrt{n} \int \{\hat{\Lambda}_n(t) - \Lambda_0(t)\} dh_2(t) + \sqrt{n} \int \{\hat{\phi}_n(w) - \phi_0(w)\} dh_3(w)$$

$$\rightarrow_d N(0, \sigma^2),$$

where σ^2 is given in (B.6).

Remark: The proofs of these theorems are given in Appendix B. In particular, Theorem 3.3.1 of van der Vaart and Wellner (1996, page 310) cannot be directly applied to prove Theorem 3.3 because the rate of convergence for the proposed estimators is no longer $n^{-1/2}$. We will show the theorem by modifying the conditions required by Theorem 3.3.1 of van der Vaart and Wellner (1996, page 310).

3.4 Simulation Studies

To access the performance of the proposed estimation procedure, we conducte simulation studies under various situations with the focus on the estimation of β. In the study, we consider a bivariate covariate $Z = (Z_1, Z_2)'$, where $Z_1 \sim N(1, 1)$ and $Z_2 \sim \text{Uniform}(-1, 1)$. The covariates V and W followed a Bernoulli distribution with success probability 0.5 and a standard uniform distribution over $[0, 1]$. The follow-up time C_i were generated by $\min(\tilde{C}_i, \tau)$, where $\tilde{C}_i \sim \text{Uniform}(2, 9)$ and $\tau = 8$.

For the observation process, we consider two scenarios. One is to assume that the observation times are independent of covariates and the other is to suppose that the observation process $H(t)$ depends on the covariate Z. For the i-th subject, the number of real observation times K_i^* was generated from a discrete uniform distribution between 1 and 5 for the former setup, and it followed a Poisson distribution with mean $\{C_i \exp(Z_{1i} + Z_{2i})/\tau\}$ for the latter one. Furthermore, the observation times $(T_{K_i,1}, \ldots, T_{K_i,K_i^*})$ were the order statistics of a random sample of size K_i^* from the uniform distribution over $(0, C_i)$.

Given K_i^* and $(T_{K_i,1}, \ldots, T_{K_i,K_i^*})$, we generated the panel counts $N_i(T_{K_i,j})$ from

$$N_i(T_{K_i,j}) = N_i(T_{K_i,1}) + \{N_i(T_{K_i,2}) - N_i(T_{K_i,1})\} + \cdots + \{N_i(T_{K_i,j}) - N_i(T_{i,j-1})\} ,$$

for $j = 1, \ldots, K_i^*$ and $i = 1, \ldots, n$. In the above, $N_i(t)$ follows a Poisson distribution with mean $t^2 \exp\{Z_{1i}\beta_1 + Z_{2i}\beta_2 + V_i\phi(W_i)\}/2$, where $\phi(w) = 2\sin(2w +$

$0.1) + \exp(-0.5w)$. The results given below are based on $n = 100$ or 200, and 500 replications with a bootstrap sample size 100.

Table 3.1 presents the simulation results by using the proposed maximum pseudo-likelihood and maximum likelihood approaches for the situation where the observation process is independent of covariates and $(\beta_1, \beta_2) = (1, 1)$, $(1, 0)$, $(1, -1)$, $(0, 1)$, or $(0, 0)$. The table includes the estimated bias (BIAS) given by the averages of the point estimates $(\hat{\beta}_1, \hat{\beta}_2)$ minus the true value of (β_1, β_2), the sample standard errors of the estimates (SSE), the means of the bootstrap standard error estimates (BSE), and the empirical 95% coverage probabilities (CP) for (β_1, β_2). It can be seen that the estimates $(\hat{\beta}_1, \hat{\beta}_2)$ seem to be unbiased and the two standard error estimates are quite close to each other, indicating that the bootstrap variance estimation procedure provides reasonable estimates. In particular, the maximum likelihood method yields smaller standard error estimates than the maximum pseudo-likelihood approach. Moreover, the empirical coverage probabilities suggest that the normal approximation seems to be appropriate.

The results for the situation where the observation times depend on the covariates Z are given in Table 3.2 in which other setups are the same as those in Table 3.1. As shown in Table 3.2, the conclusions are similar to those from Table 3.1 and indicate that the proposed estimation procedure seems to perform well for the scenarios consider here.

Table 3.3 presents the simulation results of nonparametric estimates for $(\beta_1, \beta_2) = (1, 1)$ indicating that the estimated $\phi(W)$ seems to be unbiased. The conclusions are similar when $(\beta_1, \beta_2) = (1, 0)$, $(1, -1)$, $(0, 1)$, or $(0, 0)$. Our proposed estimation procedure for ϕ by the B-spline function approximation performs well for all the scenarios in the simulation study.

In addition, we have investigated the computation time of our simulation programs in MATLAB using a PC with Intel Xeon CPU E5520 2.27 GHz. For 500 replications with n=200, it would take about 100 hours for the pseudo-likelihood approach and 15 hours for the likelihood approach.

3.5 Reanalysis of Bladder Cancer Data

To illustrate the proposed methodology given in the previous sections, we apply it to the bladder cancer study conducted by the Veterans Administration Cooperative Urological Research Group (Byar, 1980; Andrews and Herzberg, 1985; Sun and Wei, 2000). In the original study, patients with superficial bladder tumors were randomly divided into three treatment groups (placebo, thiotepa and pyridoxine) and followed for 53 months. At the beginning of the study, two important baseline characteristics, the number of initial bladder tumors and the size of the largest initial tumor, were observed for each patient. After removing all the initial tumors, many patients had multiple recurrences of tumors during the study. At each clinical follow-up visit, the visit time and the number of recurrent tumors between visits were recorded, and then the recurrent tumors were removed. Following Sun and Wei (2000), we focus on patients in the thiotepa (38) and placebo (47) groups.

For the analysis, we define Z to be 1 if the patient was given the thiotepa treatment and 0 otherwise. Let V denote the number of initial bladder tumors, and W be the natural logarithm of the size of the largest initial tumor plus 1. Assume that the occurrence process of the bladder tumors can be described by model (3.2). Our model specification regarding $(Z; V; W)$ is based on the previous literature. Both the number of initial tumors and the size of the largest initial tumor have been widely used as important diagnostic factors in cancer studies. Among others, Sun and Wei (2000) and Zhang (2002) conclude that the number of initial bladder tumors is significantly positively related with the tumor recurrence rate but the size of the largest initial tumor does not have a significant effect. Therefore, we examine the size of the largest initial tumor (W) as a potential moderator (effect modifier) of the association between the tumor recurrence and the number of initial bladder tumors (V). With a bootstrap sample size 1000, the application of the maximum pseudo-likelihood procedure yieldes $\hat{\beta}^{ps} = -1.2957$ with an estimated standard error of 0.3713, while we obtaine $\hat{\beta} = -0.8271$ with an estimated standard error of 0.3828 by applying the maximum likelihood approach. Both results suggest that the thiotepa treatment significantly reduced the recurrence rate of the bladder tumors.

Table 3.4 presents the estimated $\phi(W)$ and its 95% pointwise bootstrap confidence interval for the flexible effect of the number of initial tumors on the tumor recurrence rate based on both maximum pseudo-likelihood and maximum likelihood approaches. The results indicate that the number of initial tumors seems to be positively associated with the tumor recurrence rate only when the size of the largest initial tumor is 1 or 3, while the association is insignificant elsewhere. The difference is related to the unbalanced sample sizes for the stratified subgroups defined by W. In particular, the observed sample sizes are $n = 48$ (for $W = 1$), $n = 10$ (for $W = 2$), $n = 16$ (for $W = 3$), $n = 5$ (for $W = 4$), $n = 2$ (for $W = 5$), $n = 3$ (for $W = 6$), and $n = 1$ (for $W = 7$). Therefore, the statistical power to reject $H_0 : \phi(W) = 0$ would be relatively low due to the small sample size at most values of W except for $W = 1$ or 3. A 95% bootstrap confidence band could be an alternative approach, but it would be wide and not as informative as the 95% pointwise bootstrap confidence interval due to the small sample size in this application example.

The conclusion is comparable with those given by Sun and Wei (2000), Zhang (2002), Wellner and Zhang (2007) and Lu et al. (2009) among others, but the proposed model reveals more insight on how the effect of the number of initial tumors is moderated by the size of the largest initial tumor. In practice, one may specify $(Z; V; W)$ under a conceptual model according to research questions in which W is a possible moderator (effect modifier) of the association between the recurrent event process and covariate V.

3.6 Summary

In this chapter, we consider regression analysis of panel count data when certain covariates have nonlinear effects on recurrent events. For estimation of the constant and functional coefficients and the baseline mean function, we develop spline-based pseudo-likelihood/likelihood approaches that yield the consistency and asymptotical normality of the estimates, and propose a new algorithm for computing the spline-based maximum likelihood estimators. The proposed inference procedures are robust because the obtained asymptotic results do not rely on the Poisson assumption on the panel counts at all.

It is important to mention that Theorem 3.3 shows not only the asymptotic normality of the parametric estimators but also the asymptotic normality of the functionals of the nonparametric estimators, which can be useful for hypothesis testing problems, while Weller and Zhang (2007) and Lu et al. (2009) focus on the asymptotic distributions of the parametric estimators. Similar to Theorem 3.3, we can establish the asymptotic normality for the functionals of the estimators of the baseline mean function in the proportional mean model proposed by Weller and Zhang (2007) and Lu et al. (2009). In addition, we can also derive the asymptotic normality of the functionals of the spline likelihood-based estimators proposed by Lu et al. (2007), and thus construct a new class of nonparametric tests, which could be more powerful than the existing nonparametric tests for nonparametric comparison of several treatment groups with panel count data.

Table 3.1: Simulation results for covariate-independent observation processes

Method	n	(β_1, β_2)	BIAS	SSE	BSE	CP
Maximum pseudo-likelihood	100	(1,1)	(0.0001,0.0005)	(0.0094,0.0202)	(0.0110,0.0233)	(0.966,0.980)
		(1,0)	(0.0003,0.0000)	(0.0099,0.0209)	(0.0114,0.0230)	(0.976,0.968)
		(1,-1)	(0.0007,-0.0007)	(0.0091,0.0201)	(0.0116,0.0232)	(0.974,0.974)
		(0,1)	(0.0013,0.0000)	(0.0185,0.0354)	(0.0205,0.0375)	(0.954,0.948)
		(0,0)	(0.0004,0.0013)	(0.0195,0.0334)	(0.0210,0.0376)	(0.954,0.962)
	200	(1,1)	(0.0003,-0.0004)	(0.0055,0.0118)	(0.0061,0.0129)	(0.970,0.964)
		(1,0)	(-0.0002,0.0002)	(0.0057,0.0123)	(0.0063,0.0132)	(0.960,0.960)
		(1,-1)	(0.0002,0.0000)	(0.0057,0.0121)	(0.0062,0.0132)	(0.958,0.962)
		(0,1)	(-0.0002,0.0009)	(0.0112,0.0239)	(0.0122,0.0230)	(0.960,0.930)
		(0,0)	(0.0006,0.0002)	(0.0122,0.0219)	(0.0125,0.0225)	(0.936,0.954)
Maximum likelihood	100	(1,1)	(0.0000,-0.0003)	(0.0069,0.0146)	(0.0079,0.0166)	(0.976,0.970)
		(1,0)	(0.0002,-0.0001)	(0.0071,0.0142)	(0.0082,0.0165)	(0.962,0.970)
		(1,-1)	(0.0000,0.0002)	(0.0070,0.0142)	(0.0078,0.0164)	(0.962,0.972)
		(0,1)	(-0.0002,0.0021)	(0.0148,0.0285)	(0.0163,0.0305)	(0.964,0.954)
		(0,0)	(-0.0013,0.0002)	(0.0162,0.0283)	(0.0170,0.0302)	(0.960,0.954)
	200	(1,1)	(0.0000,0.0000)	(0.0044,0.0093)	(0.0047,0.0099)	(0.964,0.952)
		(1,0)	(0.0000,-0.0002)	(0.0047,0.0093)	(0.0050,0.0101)	(0.956,0.962)
		(1,-1)	(-0.0003,-0.0002)	(0.0043,0.0093)	(0.0046,0.0099)	(0.956,0.964)
		(0,1)	(-0.0004,0.0012)	(0.0097,0.0179)	(0.0100,0.0190)	(0.950,0.972)
		(0,0)	(-0.0007,0.0002)	(0.0101,0.0172)	(0.0106,0.0191)	(0.954,0.962)

Table 3.2: Simulation results for covariate-dependent observation processes

Method	n	(β_1, β_2)	BIAS	SSE	BSE	CP
Maximum pseudo-likelihood	100	(1,1)	(0.0003,-0.0010)	(0.0087,0.0177)	(0.0087,0.0201)	(0.940,0.966)
		(1,0)	(0.0004,-0.0002)	(0.0093,0.0196)	(0.0096,0.0217)	(0.946,0.952)
		(1,-1)	(0.0001,0.0009)	(0.0093,0.0208)	(0.0099,0.0225)	(0.956,0.952)
		(0,1)	(0.0003,-0.0041)	(0.0224,0.0406)	(0.0227,0.0417)	(0.946,0.972)
		(0,0)	(-0.0012,-0.0011)	(0.0252,0.0386)	(0.0229,0.0421)	(0.936,0.958)
	200	(1,1)	(0.0001,0.0004)	(0.0054,0.0114)	(0.0052,0.0121)	(0.930,0.958)
		(1,0)	(0.0002,0.0005)	(0.0062,0.0119)	(0.0058,0.0128)	(0.910,0.958)
		(1,-1)	(-0.0001,-0.0001)	(0.0062,0.0140)	(0.0059,0.0134)	(0.922,0.932)
		(0,1)	(-0.0002,0.0006)	(0.0146,0.0267)	(0.0137,0.0261)	(0.938,0.936)
		(0,0)	(-0.0002,-0.0020)	(0.0163,0.0269)	(0.0145,0.0259)	(0.934,0.936)
Maximum likelihood	100	(1,1)	(-0.0004,0.0001)	(0.0059,0.0138)	(0.0071,0.0155)	(0.972,0.988)
		(1,0)	(-0.0002,0.0008)	(0.0066,0.0134)	(0.0075,0.0156)	(0.972,0.964)
		(1,-1)	(-0.0001,0.0006)	(0.0063,0.0141)	(0.0074,0.0160)	(0.968,0.960)
		(0,1)	(0.0000,0.0010)	(0.0146,0.0278)	(0.0160,0.0311)	(0.964,0.962)
		(0,0)	(-0.0009,-0.0018)	(0.0156,0.0290)	(0.0170,0.0310)	(0.956,0.968)
	200	(1,1)	(0.0001,0.0001)	(0.0036,0.0081)	(0.0041,0.0093)	(0.968,0.966)
		(1,0)	(0.0000,0.0005)	(0.0042,0.0084)	(0.0045,0.0093)	(0.952,0.962)
		(1,-1)	(0.0003,0.0002)	(0.0041,0.0084)	(0.0043,0.0093)	(0.946,0.970)
		(0,1)	(0.0000,-0.0002)	(0.0088,0.0174)	(0.0097,0.0192)	(0.962,0.956)
		(0,0)	(0.0000,-0.0005)	(0.0103,0.0181)	(0.0105,0.0190)	(0.942,0.954)

Table 3.3: Simulation results of the estimated flexible effect $\phi(W)$ for $(\beta_1, \beta_2) = (1, 1)$

		Covariate-independent observation processes				Covariate-dependent observation processes			
		$n = 100$		$n = 200$		$n = 100$		$n = 200$	
W	$\phi(W)$	$\hat{\phi}^{ps}(W)$	$\hat{\phi}(W)$	$\hat{\phi}^{ps}(W)$	$\hat{\phi}(W)$	$\hat{\phi}^{ps}(W)$	$\hat{\phi}(W)$	$\hat{\phi}^{ps}(W)$	$\hat{\phi}(W)$
1/21	1.364	1.366	1.361	1.363	1.364	1.364	1.366	1.361	1.366
2/21	1.526	1.528	1.524	1.525	1.526	1.525	1.526	1.525	1.527
3/21	1.684	1.685	1.681	1.682	1.683	1.682	1.683	1.684	1.684
4/21	1.834	1.835	1.832	1.833	1.835	1.833	1.834	1.835	1.835
5/21	1.977	1.979	1.976	1.976	1.978	1.976	1.977	1.978	1.977
6/21	2.111	2.112	2.110	2.111	2.111	2.110	2.111	2.111	2.111
7/21	2.234	2.235	2.233	2.234	2.234	2.234	2.235	2.234	2.234
8/21	2.345	2.345	2.344	2.345	2.345	2.344	2.345	2.344	2.345
9/21	2.442	2.442	2.441	2.443	2.442	2.441	2.443	2.442	2.442
10/21	2.525	2.525	2.524	2.526	2.525	2.524	2.525	2.525	2.525
11/21	2.593	2.593	2.592	2.593	2.593	2.592	2.593	2.593	2.593
12/21	2.645	2.645	2.645	2.645	2.645	2.643	2.645	2.645	2.645
13/21	2.680	2.680	2.680	2.679	2.680	2.679	2.680	2.680	2.680
14/21	2.698	2.698	2.698	2.697	2.698	2.697	2.698	2.698	2.697
15/21	2.698	2.698	2.698	2.698	2.697	2.697	2.699	2.699	2.698
16/21	2.680	2.680	2.680	2.680	2.680	2.679	2.681	2.681	2.680
17/21	2.645	2.645	2.644	2.645	2.645	2.643	2.646	2.645	2.645
18/21	2.592	2.591	2.591	2.592	2.593	2.591	2.593	2.591	2.592
19/21	2.522	2.520	2.520	2.522	2.524	2.521	2.524	2.521	2.521
20/21	2.436	2.432	2.431	2.436	2.437	2.435	2.437	2.435	2.433

Table 3.4: Results of the estimated flexible effect of the number of the initial tumors on the tumor recurrence rate

	Maximum pseudo-likelihood		Maximum likelihood	
Size	$\hat{\phi}^{ps}(W)$	95% pointwise bootstrap CI	$\hat{\phi}(W)$	95% pointwise bootstrap CI
1	0.2200*	(0.0526, 0.4227)	0.2325*	(0.0356, 0.4526)
2	0.0296	(-11.4648, 0.4819)	0.0673	(-11.9842, 0.4854)
3	0.2923*	(0.0045, 0.8261)	0.2708*	(0.0540, 0.8248)
4	-0.5134	(-6.5228, 0.9392)	-0.1731	(-8.0308, 1.2608)
5	0.1993	(-5.4838, 1.4465)	0.3540	(-6.5023, 1.4312)
6	-0.2365	(-3.6525, 1.0312)	-0.3196	(-4.0047, 0.6203)
7	-2.6712	(-10.3878, 10.4999)	-2.8510	(-12.2190, 11.4039)

* P-value $\leqslant 0.05$.

Chapter 4

Future Work

In this Dissertation, our main interest is the effects of covariates on the underlying recurrent event process. And our approach depend on the distributional assumption of the underlying recurrent event process. In this chapter, we discuss some possible directions for future research based on our studies of the recurrent event data and the panel count data.

In Chapter 2, we only considered effects of time-independent covariates on the recurrent event process. It would be desirable to extend the proposed procedure to handle both time-dependent and time-independent covariates.

The example in section 2.5 involved multiple types of recurrent events. Rather than ignoring the potential correlations, it would be worthwhile to develop the maximum likelihood estimation procedure for multivariate recurrent event data in the presence of informative censoring.

The proposed approach is based on the normality assumption of the latent variable. It may not be satisfied in some applications and it would be of great interest to relax this assumption.

In Chapter 3, we assumed that the recurrent event process is independent of the observation times given covariates. To relax this assumption, we could consider the observation history as a covariate in the model and thus, the proposed method can be generalized to the dependent case. Clearly, the proposed estimation procedures work under independent censoring. However, if the censoring time is informative, a joint modeling approach needs to be developed for further research.

In addition, developing an appropriate model-checking procedure for our proposed method is an important direction for future research. Another research direction may involve high-dimensional partially linear proportional mean model with panel count data.

Appendix

Appendix A:

For $i = 1, \cdots, n$ and $s \in [0, \tau]$, define the at-risk function $Y_i(s) = I(C_i \geqslant s)$, and two functions

$$\Psi_1(X, C, K, U; \theta) = \exp\{-\Lambda(C) \exp(X'\gamma + U)\} \exp(KX'\gamma + KU)\phi(U; \sigma)$$

and

$$\Psi_2(X, C, K; \theta) = \left\{ \int \Psi_1(X, C, K, u; \theta) du \right\}^{-1} \int \Psi_1(X, C, K + 1, u; \theta) du.$$

A.1. Proof of Theorem 2.1

For $n = 1, 2, \ldots$, since the $\hat{\Lambda}_n(\cdot)$ is a bounded and nondecreasing right continuous function and the $(\hat{\gamma}_n, \hat{\sigma}_n)$ is bounded, then it follows from Helly's selection theorem that there exists a subsequence, still indexed by n, such that $\hat{\Lambda}_n(\cdot)$ uniformly converges to $\Lambda^*(\cdot)$ for some function $\Lambda^*(\cdot)$ and $\hat{\gamma}_n \to \gamma^*$ and $\hat{\sigma}_n \to \sigma^*$ for some constants γ^* and σ^*.

Next we prove that $\theta^* = \theta_0$, where $\theta^* = (\gamma^*, \sigma^*, \Lambda^*)$. Taking derivatives to the log-likelihood function (2.3) with respect to $\Delta\Lambda(t_m)$, one obtains that

$$\hat{\Lambda}(t) = \int_0^t \frac{\frac{1}{n} \sum_{i=1}^n dN_i(s)}{\frac{1}{n} \sum_{i=1}^n Y_i(s) \Psi_2(X_i, C_i, K_i; \hat{\theta}_n)}. \qquad (A.1)$$

Then we define

$$\tilde{\Lambda}_n(t) = \int_0^t \frac{\frac{1}{n} \sum_{i=1}^n dN_i(s)}{\frac{1}{n} \sum_{i=1}^n Y_i(s) \Psi_2(X_i, C_i, K_i; \theta_0)}. \qquad (A.2)$$

We now show that $\tilde{\Lambda}_n(t)$ converges to $\Lambda_0(t)$ uniformly in $t \in [0, \tau]$ almost surely. By the Glinvenko-Cantelli Theorem, the denominator of the fraction in the integral part of (A.2) converges uniformly to $E\{Y(s)\Psi_2(X, C, K; \theta_0)\}$.

Note that given (X_i, U_i, C_i), K_i follows a Poisson distribution with intensity $\Lambda_0(C_i)\exp(X_i'\gamma_0 + U_i)$, therefore the above expectation is

$$
\begin{aligned}
& E\{Y_i(s)\Psi_2(X_i, C_i, K_i; \theta_0)\} \\
= \; & E\Big\{Y_i(s)\sum_{k=0}^{\infty}\Psi_2(X_i, C_i, k; \theta_0)\int P(K_i = k|X_i, C_i, u_i)\phi(u_i; \sigma_0)du_i\Big\} \\
= \; & E\Big\{Y_i(s)\sum_{k=0}^{\infty}\int \frac{\Lambda^k(C_i)}{k!}\Psi_1(X_i, C_i, k+1, u_i; \theta_0)du_i\Big\} \\
= \; & E\Big\{Y_i(s)\int\sum_{k=0}^{\infty}P(K_i = k|X_i, C_i, u_i)\exp(X_i'\gamma_0 + u_i)\phi(u_i; \sigma_0)du_i\Big\} \\
= \; & E\Big\{Y_i(s)\int\exp(X_i'\gamma_0 + u_i)\phi(u_i; \sigma_0)du_i\Big\},
\end{aligned}
$$

which immediately yields that $\tilde{\Lambda}_n(t)$ uniformly converges to $\Lambda_0(t)$ for $t \in [0, \tau]$.

By (A.1) and (2.3), one obtains that

$$
\begin{aligned}
\frac{1}{n}l_n(\hat{\theta}_n) = \; & -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i}\log\Big\{\sum_{k=1}^{n}Y_k(T_{ij})\Psi_2(X_k, C_k, K_k; \hat{\theta}_n)\Big\} \\
& +\frac{1}{n}\sum_{i=1}^{n}\log\left\{\int\Psi_1(X_i, C_i, K_i, u_i; \hat{\theta}_n)du_i\right\}.
\end{aligned}
$$

Likewise, by (A.2) and (2.3), one also obtains that

$$
\begin{aligned}
\frac{1}{n}l_n(\gamma_0, \sigma_0, \tilde{\Lambda}_n) = \; & -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i}\log\Big\{\sum_{k=1}^{n}Y_k(T_{ij})\Psi_2(X_k, C_k, K_k; \theta_0)\Big\} \\
& +\frac{1}{n}\sum_{i=1}^{n}\log\left\{\int\Psi_1(X_i, C_i, K_i, u_i; \theta_0)du_i\right\}.
\end{aligned}
$$

By (A.1) and (A.2), one can see that $\hat{\Lambda}_n(\cdot)$ is absolutely continuous with respect to $\tilde{\Lambda}_n(\cdot)$, and satisfies

$$
\hat{\Lambda}_n(t) = \int_0^t \frac{\frac{1}{n}\sum_{i=1}^{n}Y_i(s)\Psi_2(X_i, C_i, K_i; \theta_0)}{\frac{1}{n}\sum_{i=1}^{n}Y_i(s)\Psi_2(X_i, C_i, K_i; \hat{\theta}_n)}d\tilde{\Lambda}_n(s).
$$

Therefore $\Lambda^*(t)$ is absolutely continuous with respect to $\Lambda_0(t)$ by taking limits on both sides above, and hence it is differentiable with respect to t. In addition,

$d\hat{\Lambda}_n(t)/d\Lambda_0(t)$ converges uniformly to $d\Lambda^*(t)/d\Lambda_0(t)$. Note that since $(\hat{\gamma}_n, \hat{\sigma}_n, \hat{\Lambda}_n)$ maximizes the log-likelihood function, therefore

$$
\begin{aligned}
0 \leqslant\ & \frac{1}{n}\{l_n(\hat{\gamma}_n, \hat{\sigma}_n^2, \hat{\Lambda}_n) - l_n(\gamma_0, \sigma_0, \tilde{\Lambda}_n)\} \\
=\ & \frac{1}{n}\sum_{i=1}^{n} \log\left\{\int \Psi_1(X_i, C_i, K_i, u_i; \hat{\theta}_n) du_i\right\} - \frac{1}{n}\sum_{i=1}^{n} \log\left\{\int \Psi_1(X_i, C_i, K_i, u_i; \tilde{\theta}_n) du_i\right\} \\
& + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i} \log\left\{\Delta\hat{\Lambda}_n(T_{ij})/\Delta\tilde{\Lambda}_n(T_{ij})\right\},
\end{aligned}
$$

where $\tilde{\theta}_n = (\gamma_0, \sigma_0, \tilde{\Lambda}_n)$. Letting n tend to be infinity, one obtains that

$$
\begin{aligned}
0 \leqslant\ & E\Bigg[\log\Bigg\{\Bigg(\prod_{j=1}^{K_i} \lambda^*(T_{ij}) \int \Psi_1(X_i, C_i, K_i, u_i; \tilde{\theta}) du_i\Bigg) \\
& \times \Bigg(\prod_{j=1}^{K_i} \lambda_0(T_{ij}) \int \Psi_1(X_i, C_i, K_i, u_i; \theta_0) du\Bigg)^{-1}\Bigg\}\Bigg].
\end{aligned}
$$

Note that the right-hand side of the above inequality is the Kullback-Leibler information. Therefore

$$
\prod_{j=1}^{K_i} \lambda^*(T_{ij}) \int \Psi_1(X_i, C_i, K_i, u_i; \theta^*) du_i = \prod_{j=1}^{K_i} \lambda_0(T_{ij}) \int \Psi_1(X_i, C_i, K_i, u_i; \theta_0) du_i
$$

almost surely. That is,

$$
\begin{aligned}
& \int \prod_{j=1}^{K_i} \lambda^*(T_{ij}) \exp\{-\Lambda^*(C_i)\exp(X_i'\gamma^* + u_i)\} \exp\{K_i(X_i'\gamma^* + u_i)\}\phi(u_i; \sigma^*) du_i \\
=\ & \int \prod_{j=1}^{K_i} \lambda_0(T_{ij}) \exp\{-\Lambda_0(C_i)\exp(X_i'\gamma_0 + u_i)\} \exp\{K_i(X_i'\gamma_0 + u_i)\}\phi(u_i; \sigma_0) du_i.
\end{aligned}
\tag{A.3}
$$

Letting $T_{i,j} = 0$ and $C_i \to 0$, from condition (C3), one obtains that

$$
K_i \log \lambda^*(0) + K_i X_i'\gamma^* + \frac{1}{2}K_i^2(\sigma^*)^2 = K_i \log \lambda_0(0) + K_i X_i'\gamma_0 + \frac{1}{2}K_i^2\sigma_0^2.
$$

Since K_i can be positive, the above equality implies that $\sigma^* = \sigma_0$ and

$$X_i'(\gamma^* - \gamma_0) = \log \lambda_0(0) - \log \lambda^*(0), \quad \text{almost surely}$$

which immediately yields that $\gamma^* = \gamma_0$ and $\lambda^*(0) = \lambda_0(0)$ from condition (C5).

Again using (A3) and letting $T_{i,j} = 0$, one can conclude that for any $t \in [0, \tau]$,

$$\begin{aligned} & \int \exp\{-\Lambda^*(t) \exp(X_i'\gamma_0 + u_i)\} \exp\{K_i(X_i'\gamma_0 + u_i)\}\phi(u_i; \sigma_0) du_i \\ = & \int \exp\{-\Lambda_0(t) \exp(X_i'\gamma_0 + u_i)\} \exp\{K_i(X_i'\gamma_0 + u_i)\}\phi(u_i; \sigma_0) du_i \end{aligned}$$

yielding that $\Lambda^*(t) = \Lambda_0(t)$ since both sides of the above equation are strictly monotone in $\Lambda^*(t)$ and $\Lambda_0(t)$. Combining all the results, we prove Theorem 2.1. □

A.2. Proof of Theorem 2.2

We show this theorem by verifying the four conditions in Theorem 3.3.1 of van der Vaart & Wellner (1996). Now we define the mappings Ψ and Ψ_n. For $o = (x, c, k, \tilde{t})$ with $\tilde{t} = (t_1, \cdots, t_k)$, define the following three functions

$$\begin{aligned} \pi_1(o; \theta) = & \left\{\int \Psi_1(x, c, k, u; \theta) du\right\}^{-1} \int \Psi_1(x, c, k, u; \theta) \exp(x'\gamma + u) du, \\ \pi_2(o; \theta) = & \left\{\int \Psi_1(x, c, k, u; \theta) du\right\}^{-1} \int \Psi_1(x, c, k, u; \theta)\{k - \Lambda(c) \exp(x'\gamma + u)\} x du, \\ \text{and} & \\ \pi_3(o; \theta) = & \left\{\sigma \int \Psi_1(x, c, k, u; \theta) du\right\}^{-1} \int \Psi_1(x, c, k, u; \theta) \left(\frac{u^2}{\sigma^2} - 1\right) du. \end{aligned}$$

Define a random map Ψ as follows: for any function $g_\Lambda(\cdot)$ on $[0, \tau]$ of bounded total variation, $g_\gamma \in R^p$ with $\|g_\gamma\| \leqslant 1$ and scalar $g_\sigma \in [-1, 1]$,

$$\Psi(\theta)[g_\gamma, g_\sigma, g_\Lambda] = -\pi_1(O; \theta) \int_0^C g_\Lambda(t) d\Lambda(t) + \pi_2(O; \theta)' g_\gamma + \pi_3(O; \theta) g_\sigma + \sum_{j=1}^K g_\Lambda(T_j),$$

which is the score function along the path $(\gamma + a g_\gamma, \sigma + a g_\sigma, \Lambda + a \int g_\Lambda d\Lambda)$. Specifically,

$$\Psi(\theta)[g_\gamma, g_\sigma, g_\Lambda] = \frac{\partial}{\partial a} l \left(\gamma + a g_\gamma, \sigma + a g_\sigma, \Lambda + a \int g_\Lambda d\Lambda | O\right)\Big|_{a=0},$$

where $l(\cdot|O)$ is the log-likelihood function of θ, given by taking the logarithm of (2.3) for one observation. Then the mappings $\Psi_n(\theta)[g_\gamma, g_\sigma, g_\Lambda]$ is defined as the mean of the i.i.d. sample of $\Psi(\theta)[g_\gamma, g_\sigma, g_\Lambda]$.

Firstly, one can show that the score functions Ψ are $P-$ Donsker. Therefore the first conditions hold and

$$\sqrt{n}\left\{\Psi_n(\theta_0) - E\Psi(\theta_0)\right\} \to N(0, \eta^{-2}),$$

where

$$\eta^2 = 1/E[\{\Psi(\theta_0)[g_\gamma, g_\sigma, g_\Lambda]\}^2]. \tag{A.4}$$

This implies that the second condition holds.

Following the proof of Theorem 2 in Zeng, et al. (2005), we verify the third condition that $\dot{\Psi}_{\theta_0}$ is continuously invertible. It is enough to prove that, if $\Psi(\theta_0)[g_\gamma, g_\sigma, g_\Lambda] = 0$ almost surely, then $g_\gamma = 0, g_\sigma = 0$ and $g_\Lambda = 0$. Letting $T_j = 0$ and $C = 0$, one obtains that

$$KX'g_\gamma + \pi_3(O;\theta)g_\sigma + \sum_{j=1}^{K} g_\Lambda(0) = 0.$$

After some calculations, one obtains that

$$Kg_\Lambda(0) + KX'g_\gamma + K^2\sigma_0 g_\sigma = 0,$$

which immediately yields that $g_\Lambda(0) = 0, g_\gamma = 0$ and $g_\sigma = 0$ from condition (C5). Using this result and $\Psi(\theta_0) = 0$, one obtains that $\int_0^C g_\Lambda(t)d\Lambda_0(t) = 0$ and thus $g_\Lambda(t) = 0$ from conditions (C1) and (C3).

Observe that $\Psi_n(\hat{\theta}_n)[g_\gamma, g_\sigma, g_\Lambda] = 0$ and $E(\Psi(\theta_0)[g_\gamma, g_\sigma, g_\Lambda]) = 0$, which implies that the forth condition is satisfied.

The asymptotic efficiency follows from Proposition 1 of Bickel et al. (1993, p. 65), that is, $(\hat{\gamma}_n', \hat{\sigma}_n)'$ is asymptotically efficient in the semiparametric sense. This completes the proof. □

Appendix B:

B.1. Proof of Theorem 3.1

Here, we only present the proof of part (i) since part (ii) can be verified similarly. Let

$$m_\theta^{ps}(X) = \sum_{j=1}^{K} \left[N(T_{K,j}) \log\{\Lambda(T_{K,j}) \exp\left(\beta' Z + V\phi(W)\right)\} - \Lambda(T_{K,j}) \exp\left(\beta' Z + V\phi(W)\right)\right],$$

$$M_n^{ps}(\theta) = P_n m_\theta^{ps}(X), \quad \text{and} \quad M^{ps}(\theta) = P m_\theta^{ps}(X),$$

where P and P_n denote the probability measure and the empirical measure, respectively. Let $h(x) = x \log x - x + 1$. Note that $h(x) \geqslant (x-1)^2/4$ for $0 \leqslant x \leqslant 5$. For any θ in a sufficiently small neighborhood of θ_0,

$$\begin{aligned} &M^p(\theta_0) - M^p(\theta) \\ &= \int \Lambda(u) \exp\{Z'\beta + v\phi(w)\} h\left(\frac{\Lambda_0(u)\exp(z'\beta_0 + v\phi_0(w))}{\Lambda(u)\exp(z'\beta + v\phi(w))}\right) d\nu_1(u, z, v, w) \\ &\geqslant \frac{1}{4} \int \Lambda(u) \exp\{Z'\beta + v\phi(w)\} \left\{\frac{\Lambda_0(u)\exp(z'\beta_0 + v\phi_0(w))}{\Lambda(u)\exp(z'\beta + v\phi(w))} - 1\right\}^2 d\nu_1(u, z, v, w). \end{aligned} \tag{$B.1$}$$

Then, using (B.1) and the arguments similar to those in Wellner and Zhang (2007), we can show that $M^{ps}(\theta_0) = M^{ps}(\theta)$ if and only if $\beta = \beta_0$, $\Lambda(t) = \Lambda_0(t)$ a.e. with respect to μ_1, and $\phi = \phi_0$ by C3 and C7.

By the similar arguments as those used in Wellner and Zhang (2007) again, we can also show that $\hat{\Lambda}_n^{ps}(t)$ is uniformly bounded in probability for $t \in [0, b]$ if $\mu_1([b, \tau]) > 0$ for some $0 < b < \tau$ or $t \in [0, \tau]$ if $\mu_1(\{\tau\}) > 0$.

By Helly-Selection Theorem and compactness of Θ_n, it follows that $\hat{\theta}_n^{ps} = (\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}, \hat{\phi}_n^{ps})$ has a subsequence $\hat{\theta}_{n_k}^{ps} = (\hat{\beta}_{n_k}^{ps}, \hat{\Lambda}_{n_k}^{ps}, \hat{\phi}_{n_k}^{ps})$ converging to $\theta^+ = (\beta^+, \Lambda^+, \phi^+)$, where Λ^+

is a nondecreasing bound function on $[0, b]$ for $0 < b < \tau$ and it can be defined on $[0, \tau]$ if $\mu_1(\{\tau\}) > 0$.

Note that Θ_n is compact, and the function $m_\theta^{ps}(x)$ is upper semicontinuous in θ for almost all x. Furthermore, $m_\theta^{ps}(X) \leqslant M_0^{ps}(X) < \infty$ with $PM_0^{ps}(X) < \infty$ by C4. Thus, by Theorem A.1 of Wellner and Zhang (2000), we have

$$\limsup_{n\to\infty} \sup_{\theta\in\Theta_n} (P_n - P) m_\theta^{ps}(X) \leqslant 0 \qquad a.s. \tag{B.2}$$

By the Dominated Convergence Theorem and C4, $M^{ps}(\theta)$ is continuous in θ. Therefore, for any $\varepsilon > 0$, there exists $\phi_0^* \in \Psi_n$ such that

$$M^{ps}(\beta_0, \Lambda_0, \phi_0) - \varepsilon \leqslant M^{ps}(\beta_0, \Lambda_0, \phi_0^*) \quad \text{with } ||\phi_0 - \phi_0^*||_\infty = o(1).$$

Clearly,

$$M_n^{ps}(\beta_0, \Lambda_0, \phi_0^*) - M^{ps}(\beta_0, \Lambda_0, \phi_0^*) = o_p(1)$$

and

$$M_n^{ps}(\beta_0, \Lambda_0, \phi_0^*) \leqslant M_n^{ps}(\hat\beta_n^{ps}, \hat\Lambda_n^{ps}, \hat\phi_n^{ps}).$$

Then, using (B.2) and the arguments similar to those used in Lu et al. (2009), we can show that $M^{ps}(\theta^+) = M^{ps}(\theta_0)$, which yields $\beta^+ = \beta_0$, $\Lambda^+ = \Lambda_0$, a.e., and $\phi^+ = \phi_0$. Therefore, we obtain the weak consistency of $(\hat\beta_n^{ps}, \hat\Lambda_n^{ps}, \hat\phi_n^{ps})$ in the metric d_1.

B.2. Proof of Theorem 3.2

To obtain the rate of convergence, we will apply Theorem 3.2.5 of van der Vaart and Wellner (1996). Let $m_\theta^{ps}(X)$, $M_n^{ps}(\theta)$, and $M^{ps}(\theta)$ be as defined in Appendix B.1. Let $\mu(u, v, w) = \Lambda(u)\exp\{v\phi(w)\}$, $\mu_0(u, v, w) = \Lambda_0(u)\exp\{v\phi_0(w)\}$ and $g(t) = \mu_t(U, Z, V, W)\exp(Z'\beta_t)$, where $(U, Z, V, W) \sim \nu_1$, $\mu_t = t\mu + (1-t)\mu_0$, $\beta_t = t\beta + (1-t)\beta_0$ for $0 \leqslant t \leqslant 1$. Then,

$$\Lambda(U)e^{Z'\beta + V\phi(W)} - \Lambda_0(U)e^{Z'\beta_0 + V\phi_0(W)} = g(1) - g(0).$$

By the mean value theorem, there exits a $0 \leqslant \xi \leqslant 1$ such that $g(1) - g(0) = g'(\xi)$. Since

$$\begin{aligned} g'(\xi) &= \exp(Z'\beta_\xi)[(\mu-\mu_0)(U,V,W) + \{\mu_0 + \xi(\mu-\mu_0)\}(U,V,W)Z'(\beta-\beta_0)] \\ &= \exp(Z'\beta_\xi)[(\mu-\mu_0)(U,V,W)\{1+\xi Z'(\beta-\beta_0)\} + \mu_0(U,V,W)Z'(\beta-\beta_0)], \end{aligned}$$

then from (B.1) we have

$$\begin{aligned} & M^{ps}(\theta_0) - M^{ps}(\theta) \\ & \geqslant c_1 \int \{\Lambda(u)\exp(z'\beta + v\phi(w)) - \Lambda_0(u)\exp(z'\beta_0 + v\phi_0(w))\}^2 \, d\nu_1(u,z,v,w) \\ & = c_1 \int [(\mu-\mu_0)(u,v,w)\{1+\xi z'(\beta-\beta_0)\} + \mu_0(u,v,w)z'(\beta-\beta_0)]^2 d\nu_1(u,z,v,w) \\ & = c_1 \nu_1(g_1 h_1 + g_2)^2 \end{aligned}$$

for a constant c_1, where $g_1(U,Z,V,W) = Z'(\beta-\beta_0)\mu_0(U,V,W)$, $g_2(U,V,W) = (\mu-\mu_0)(U,V,W)$, and $h_1(U,Z,V,W) = 1 + \xi\frac{(\mu-\mu_0)(U,V,W)}{\mu_0(U,V,W)}$ in the notation of Lemma 8.8 of van der Vaart (2002, page 432). To apply van der Vaart's Lemma, we need to show that

$$\{\nu_1(g_1 g_2\}^2 \leqslant c\nu_1(g_1^2)\nu_1(g_2^2) \qquad (B.3)$$

for a constant $c < 1$. By the Cauchy-Schwarz inequality and condition C13, we can show that (B.3) holds for $c = 1 - \eta_1$. Let

$$\Lambda_t = t\Lambda + (1-t)\Lambda_0, \phi_t = t\phi + (1-t)\phi_0, Q(t) = \Lambda_t(U)e^{V\phi_t(W)}.$$

Then

$$g_2(U,V,W) = Q(1) - Q(0) = Q'(\zeta) \quad \text{for } 0 \leqslant \zeta \leqslant 1,$$

and

$$\nu_1(g_2^2) = \nu_1((h_2 g_3 + g_4)^2)$$

where $g_3(U,V,W) = V(\phi(W) - \phi_0(W))\Lambda_0(U)$, $g_4(U) = (\Lambda - \Lambda_0)(U)$, and $h_2(U,V,W) = 1 + \zeta\frac{(\Lambda-\Lambda_0)(U)}{\Lambda_0(U)}$. Similarly, we can show that

$$\{\nu_1(g_3 g_4\}^2 \leqslant (1-\eta_2)\nu_1(g_3^2)\nu_1(g_4^2).$$

So, by van der Vaart's Lemma, we have

$$\nu_1(g_1 h + g_2)^2 \geqslant c d_1^2(\theta, \theta_0).$$

To derive the rate of convergence, next we need to find a $\varphi_n(\delta)$ such that

$$E\left\{\sup_{d_1(\theta,\theta_0)<\delta} \sqrt{n}|(P_n - P)(m_\theta^{ps}(X) - m_{\theta_0}^{ps}(X))|\right\} \leqslant c\varphi_n(\delta).$$

Let

$$\mathcal{F}_\delta^{ps} = \left\{m_\theta^{ps}(X) - m_{\theta_0}^{ps}(X) : d_1(\theta,\theta_0) \leqslant \delta\right\}.$$

From the result of Theorem 2.7.5 of van der Vaart and Wellner (1996) and Lemma B.2 of Huang (1999), for any $\epsilon \leqslant \delta$, we have

$$\log N_{[]}(\epsilon, \mathcal{F}_\delta^{ps}, ||\cdot||_{P,B}) \leqslant c\left(\frac{1}{\epsilon} + q_n \log\frac{\delta}{\epsilon}\right),$$

where $||\cdot||_{P,B}$ is the Bernstein Norm defined as $||f||_{P,B} = \{2P(e^{|f|} - 1 - |f|)\}^{1/2}$ by van der Vaart and Wellner (1996, page 324). Moreover, we can show that

$$||m_\theta^{ps}(X) - m_{\theta_0}^{ps}(X)||_{P,B}^2 \leqslant c\delta^2,$$

for any $m_\theta^{ps}(X) - m_{\theta_0}^{ps}(X) \in \mathcal{F}_\delta^{ps}$. Therefore, by Lemma 3.4.3 of van der Vaart and Wellner (1996), we obtain

$$E||n^{1/2}(P_n - P)||_{\mathcal{F}_\delta^{ps}} \leqslant cJ_{[]}(\delta, \mathcal{F}_\delta^{ps}, ||\cdot||_{P,B})\left\{1 + \frac{J_{[]}(\delta, \mathcal{F}_\delta^{ps}, ||\cdot||_{P,B})}{\delta^2 n^{1/2}}\right\}$$

where

$$J_{[]}(\delta, \mathcal{F}_\delta^{ps}, ||\cdot||_{P,B}) = \int_0^\delta \{1+\log N_{[]}(\epsilon, \mathcal{F}_\delta^{ps}, ||\cdot||_{P,B})\}^{1/2} d\epsilon \leqslant c q_n^{\frac{1}{2}} \int_0^\delta \left(1 + \frac{1}{\epsilon} + \log\frac{\delta}{\epsilon}\right)^{1/2} d\epsilon \leqslant c q_n^{\frac{1}{2}} \delta^{\frac{1}{2}}.$$

Thus,

$$\varphi_n(\delta) = c q_n^{\frac{1}{2}} \delta^{\frac{1}{2}} \left(1 + \frac{c q_n^{1/2} \delta^{1/2}}{\delta^2 n^{1/2}}\right) = c(q_n^{\frac{1}{2}} \delta^{\frac{1}{2}} + \frac{q_n}{\delta n^{1/2}}).$$

It is easy to see that $\varphi_n(\delta)/\delta$ is decreasing in δ, and

$$r_n^2 \varphi_n(\frac{1}{r_n}) = r_n^2 (q_n^{\frac{1}{2}} r_n^{-\frac{1}{2}} + \frac{q_n}{r_n^{-1} n^{1/2}}) = r_n^{\frac{3}{2}} q_n^{\frac{1}{2}} + r_n^3 q_n n^{-\frac{1}{2}} \leqslant c n^{\frac{1}{2}}$$

for $r_n = n^{\frac{1-v}{3}}$ and $0 < v < 1/2$. Hence, it follows from Theorem 3.2.5 of van der Vaart and Wellner (1996) that $n^{\frac{1-v}{3}} d_1(\hat{\theta}_n^{ps}, \theta_0) = O_p(1)$. Similarly, we can obtain the rate of convergence for $\hat{\theta}_n$.

B.3. Proof of Theorem 3.3

First, we prove part (i). Recall that

$$\begin{aligned} l_n^{ps}(\beta, \Lambda, \phi) &= \sum_{i=1}^n \sum_{j=1}^{K_i} \left[N_i(T_{K_i,j}) \log\{\Lambda(T_{K_i,j})\} + N_i(T_{K_i,j})\{Z_i'\beta + V_i\phi(W_i)\} \right. \\ &\qquad \left. - \Lambda(T_{K_i,j}) \exp\{Z_i'\beta + V_i\phi(W_i)\} \right]. \end{aligned}$$

We define a sequence of maps S_n^{ps} mapping a neighborhood of $(\beta_0, \Lambda_0, \phi_0)$, denoted by $\mathcal{U}$, in the parameter space for (β, Λ, ϕ) into $l^\infty(\mathcal{H})$ as :

$$
\begin{aligned}
& S_n^{ps}(\theta)[\mathbf{h}_1, h_2, h_3] \\
& = n^{-1}\frac{d}{d\varepsilon} l_n^{ps}(\beta + \varepsilon \mathbf{h}_1, \Lambda + \varepsilon h_2, \phi + \varepsilon h_3)\Big|_{\varepsilon=0} \\
& = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{K_i}[\{N_i(T_{K_i,j}) - \Lambda(T_{K_i,j})\exp(\beta' Z_i + V_i\phi(W_i))\}\mathbf{h}_1' Z_i \\
& \qquad + \left\{\frac{N_i(T_{K_i,j})}{\Lambda(T_{K_i,j})} - \exp(\beta' Z_i + V_i\phi(W_i))\right\} h_2(T_{K_i,j}) \\
& \qquad + \{N_i(T_{K_i,j}) - \Lambda(T_{K_i,j})\exp(\beta' Z_i + V_i\phi(W_i))\}V_i h_3(W_i)] \\
& \equiv A_{n1}^{ps}(\theta)[\mathbf{h}_1] + A_{n2}^{ps}(\theta)[h_2] + A_{n3}^{ps}(\theta)[h_3] \\
& \equiv P_n(\mathbf{h}_1' \dot{l}_\beta^{ps}) + P_n(\dot{l}_\Lambda^{ps}[h_2]) + P_n(\dot{l}_\phi^{ps}[h_3]) \\
& \equiv P_n\psi_{ps}(\theta)[\mathbf{h}_1, h_2, h_3].
\end{aligned}
$$

Correspondingly, we define the limit map $S^{ps} : \mathcal{U} \longrightarrow l^\infty(\mathcal{H})$ as

$$
S^{ps}(\theta)[\mathbf{h}_1, h_2, h_3] = A_1^{ps}(\theta)[\mathbf{h}_1] + A_2^{ps}(\theta)[h_2] + A_3^{ps}(\theta)[h_3],
$$

where

$$
A_1^{ps}(\theta)[\mathbf{h}_1] = P\left[\sum_{j=1}^{K}\{N(T_{K,j}) - \Lambda(T_{K,j})\exp(\beta' Z + V\phi(W))\}\mathbf{h}_1' Z\right],
$$

$$
A_2^{ps}(\theta)[h_2] = P\left[\sum_{j=1}^{K}\left\{\frac{N(T_{K,j})}{\Lambda(T_{K,j})} - \exp(\beta' Z + V\phi(W))\right\} h_2(T_{K,j})\right],
$$

and

$$
A_3^{ps}(\theta)[h_3] = P\left[\sum_{j=1}^{K}\{N(T_{K,j}) - \Lambda(T_{K,j})\exp(\beta' Z + V\phi(W))\}V h_3(W)\right].
$$

To derive the asymptotic normality of the estimators $(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}, \hat{\phi}_n^{ps})$, motivated by the proofs of Theorem 3.3.1 of Van der Vaart and Wellner (1996, page 310) and Theorem 2 of Zeng et al. (2005), we need to verify the following five conditions.

(a1) $\sqrt{n}(S_n^{ps} - S^{ps})(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}, \hat{\phi}_n^{ps}) - \sqrt{n}(S_n^{ps} - S^{ps})(\beta_0, \Lambda_0, \phi_0) = o_p(1)$.

(a2) $S^{ps}(\beta_0, \Lambda_0, \phi_0) = 0$ and $S_n^{ps}(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}, \hat{\phi}_n^{ps}) = o_p(n^{-1/2})$.

(a3) $\sqrt{n}(S_n^{ps} - S^{ps})(\beta_0, \Lambda_0, \phi_0)$ converges in distribution to a tight Gaussian process on $l^\infty(\mathcal{H})$.

(a4) $S^{ps}(\beta, \Lambda, \phi)$ is Fréchet-differentiable at $(\beta_0, \Lambda_0, \phi_0)$ with a continuously invertible derivative $\dot{S}^{ps}(\beta_0, \Lambda_0, \phi_0)$.

(a5) $S^{ps}(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}, \hat{\phi}_n^{ps}) - S^{ps}(\beta_0, \Lambda_0, \phi_0) - \dot{S}^{ps}(\beta_0, \Lambda_0, \phi_0)(\hat{\beta}_n^{ps} - \beta_0, \hat{\Lambda}_n^{ps} - \Lambda_0, \hat{\phi}_n^{ps} - \phi_0) = o_p(n^{-1/2})$.

Condition (a1) holds since

$$\Big\{\psi_{ps}(\beta, \Lambda, \phi)[\mathbf{h}_1, h_2, h_3] - \psi_{ps}(\beta_0, \Lambda_0, \phi_0)[\mathbf{h}_1, h_2, h_3] : \\ d_1((\beta, \Lambda, \phi), (\beta_0, \Lambda_0, \phi_0)) < \delta, (\mathbf{h}_1, h_2, h_3) \in \mathcal{H}\Big\}$$

is a Donsker class for some $\delta > 0$, and that

$$\sup_{(\mathbf{h}_1, h_2, h_3) \in \mathcal{H}} P\big[\psi_{ps}(\beta, \Lambda, \phi)[\mathbf{h}_1, h_2, h_3] - \psi_{ps}(\beta_0, \Lambda_0, \phi_0)[\mathbf{h}_1, h_2, h_3]\big]^2 \longrightarrow 0,$$

as $(\beta, \Lambda, \phi) \longrightarrow (\beta_0, \Lambda_0, \phi_0)$ in d_1.

Clearly, $S^{ps}(\beta_0, \Lambda_0, \phi_0) = 0$. For $h_3 \in \mathcal{F}_r$, let h_{3n} be the B-spline function approximation of h_3 with $||h_3 - h_{3n}||_\infty = O(n^{-vr})$ by Corollary 6.21 of Schumaker (1981, page 227). Then we have $S_n^{ps}(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}, \hat{\phi}_n^{ps})[\mathbf{h}_1, h_2, h_{3n}] = 0$. Thus, for $(\mathbf{h}_1, h_2, h_3) \in \mathcal{H}$,

$$\begin{aligned}
&\sqrt{n}\{S_n^{ps}(\hat{\beta}_n^{ps}, \hat{\Lambda}_n^{ps}, \hat{\phi}_n^{ps})[\mathbf{h}_1, h_2, h_3]\} \\
&= \sqrt{n}P_n\psi_{ps}(\hat{\theta}_n^{ps})[\mathbf{h}_1, h_2, h_3] - \sqrt{n}P_n\psi_{ps}(\hat{\theta}_n^{ps})[\mathbf{h}_1, h_2, h_{3n}] \\
&= I_{n1} - I_{n2} + I_{n3} + I_{n4}
\end{aligned}$$

where

$$I_{n1} = \sqrt{n}(P_n - P)\left\{\psi_{ps}(\hat{\theta}_n^{ps})[\mathbf{h}_1, h_2, h_3] - \psi_{ps}(\theta_0)[\mathbf{h}_1, h_2, h_3]\right\},$$

$$I_{n2} = \sqrt{n}(P_n - P)\left\{\psi_{ps}(\hat{\theta}_n^{ps})[\mathbf{h}_1, h_2, h_{3n}] - \psi_{ps}(\theta_0)[\mathbf{h}_1, h_2, h_{3n}]\right\},$$

$$I_{n3} = \sqrt{n}P_n\left\{\psi_{ps}(\theta_0)[\mathbf{h}_1, h_2, h_3] - \psi_{ps}(\theta_0)[\mathbf{h}_1, h_2, h_{3n}]\right\},$$

and

$$I_{n4} = \sqrt{n}P\left\{\psi_{ps}(\hat{\theta}_n^{ps})[\mathbf{h}_1, h_2, h_3] - \psi_{ps}(\hat{\theta}_n^{ps})[\mathbf{h}_1, h_2, h_{3n}]\right\}.$$

From (a1), we have $I_{n1} = o_p(1)$ and $I_{n2} = o_p(1)$. Next we need to show $I_{n3} = o_p(1)$ and $I_{n4} = o_p(1)$. Note that

$$E(I_{n3}^2) = P\left\{\psi_{ps}(\theta_0)[\mathbf{h}_1, h_2, h_3] - \psi_{ps}(\theta_0)[\mathbf{h}_1, h_2, h_{3n}]\right\}^2 \leqslant c||h_{3n} - h_3||_\infty^2 \to 0,$$

and

$$\begin{aligned}
|I_{n4}| &= \Bigg|\sqrt{n}P\Bigg[\sum_{j=1}^{K}\{\Lambda_0(T_{K,j})\exp(Z'\beta_0 + V\phi_0(W)) \\
&\qquad -\hat{\Lambda}_n^{ps}(T_{K,j})\exp(Z'\hat{\beta}_n^{ps} + V\hat{\phi}_n^{ps}(W))\Big\}V(h_3(W) - h_{3n}(W)\Bigg]\Bigg| \\
&\leqslant c\sqrt{n}d_1(\hat{\theta}_n^{ps}, \theta_0)||h_3 - h_{3n}||_\infty \\
&= O(n^{-\frac{1-v}{3} - vr + \frac{1}{2}})
\end{aligned}$$

by Theorem 3.2. Thus (a2) holds for $\frac{1}{6r-2} < v < \frac{1}{2}$.

Condition (a3) holds because $\mathcal{H}$ is a Donsker class and the functionals $A_{n1}^{ps}, A_{n2}^{ps}, A_{n3}^{ps}$ are bounded Lipschitz functions with respect to $\mathcal{H}$.

For (a4), by the smoothness of $S^{ps}(\beta, \Lambda, \phi)$, the Fréchet differentiability holds and the derivative of $S^{ps}(\beta, \Lambda, \phi)$ at $(\beta_0, \Lambda_0, \phi_0)$, denoted by $\dot{S}^{ps}(\beta_0, \Lambda_0, \phi_0)$, is a map from the space $\{(\beta-\beta_0, \Lambda-\Lambda_0, \phi-\phi_0) : (\beta, \Lambda, \phi) \in \mathcal{U}\}$ to $l^\infty(\mathcal{H})$ and

$$
\begin{aligned}
&\dot{S}^{ps}(\beta_0, \Lambda_0, \phi_0)(\beta-\beta_0, \Lambda-\Lambda_0, \phi-\phi_0)[\mathbf{h}_1, h_2, h_3] \\
&= \frac{d}{d\varepsilon}\left\{A_1^{ps}(\theta_0+\varepsilon(\theta-\theta_0))[\mathbf{h}_1]\right\}\Big|_{\varepsilon=0} + \frac{d}{d\varepsilon}\left\{A_2^{ps}(\theta_0+\varepsilon(\theta-\theta_0))[h_2]\right\}\Big|_{\varepsilon=0} \\
&\quad + \frac{d}{d\varepsilon}\left\{A_3^{ps}(\theta_0+\varepsilon(\theta-\theta_0))[h_3]\right\}\Big|_{\varepsilon=0} \\
&= -P\sum_{j=1}^{K}\exp(\beta_0'Z+V\phi_0(W))\mathbf{h}_1'Z\left[\{\Lambda(T_{K,j})-\Lambda_0(T_{K,j})\}\right. \\
&\qquad\qquad \left. +\Lambda_0(T_{K,j})\left\{(\beta-\beta_0)'Z+V(\phi(W)-\phi_0(W))\right\}\right] \\
&\quad -P\sum_{j=1}^{K}\exp(\beta_0'Z+V\phi_0(W))h_2(T_{K,j})\left[\left\{\frac{\Lambda(T_{K,j})-\Lambda_0(T_{K,j})}{\Lambda_0(T_{K,j})}\right.\right. \\
&\qquad\qquad \left. + \left\{(\beta-\beta_0)'Z+V(\phi(W)-\phi_0(W))\right\}\right] \\
&\quad -P\sum_{j=1}^{K}\exp(\beta_0'Z+V\phi_0(W))Vh_3(W)\left[\{\Lambda(T_{K,j})-\Lambda_0(T_{K,j})\}\right. \\
&\qquad\qquad \left. +\Lambda_0(T_{K,j})\left\{(\beta-\beta_0)'Z+V(\phi(W)-\phi_0(W))\right\}\right].
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
&\dot{S}^{ps}(\beta_0, \Lambda_0, \phi_0)(\beta-\beta_0, \Lambda-\Lambda_0, \phi-\phi_0)[\mathbf{h}_1, h_2, h_3] \\
&= (\beta-\beta_0)'Q_1^{ps}(\mathbf{h}_1, h_2, h_3) + \int(\Lambda(t)-\Lambda_0(t))dQ_2^{ps}(\mathbf{h}_1, h_2, h_3)(t) \qquad (B.4) \\
&\quad + \int(\phi(w)-\phi_0(w))dQ_3^{ps}(\mathbf{h}_1, h_2, h_3)(w)
\end{aligned}
$$

where

$$
\begin{aligned}
&Q_1^{ps}(\mathbf{h}_1, h_2, h_3) \\
&= -E\left[Z\exp\{\beta_0'Z+V\phi_0(W)\}\sum_{j=1}^{K}\left\{\Lambda_0(T_{K,j})\mathbf{h}_1'Z+h_2(T_{K,j})+\Lambda_0(T_{K,j})Vh_3(W)\right\}\right],
\end{aligned}
$$

$dQ_2^{ps}(\mathbf{h}_1, h_2, h_3)(t)$

$= -E\left[\exp\{\beta_0' Z + V\phi_0(W)\} \sum_{j=1}^K \frac{1}{\Lambda_0(t)} \{\Lambda_0(t)\mathbf{h}_1' Z + h_2(t) + \Lambda_0(t) V h_3(W)\} \, dP(T_{K,j} \leqslant t | K, Y)\right],$

and

$dQ_3^{ps}(\mathbf{h}_1, h_2, h_3)(w)$

$= -E\left[V \exp\{\beta_0' Z + V\phi_0(w)\} \sum_{j=1}^K \{\Lambda_0(T_{K,j})\mathbf{h}_1' Z + h_2(T_{K,j}) + \Lambda_0(T_{K,j}) V h_3(w)\} | W = w\right] dF_W(w)$

where F_W denotes the cumulative distribution of W.

Next, we show that $Q^{ps} = (Q_1^{ps}, Q_2^{ps}, Q_3^{ps})$ is one-to-one, that is, for $h \in \mathcal{H}$, if $Q^{ps}(\mathbf{h}_1, h_2, h_3) = 0$, then $\mathbf{h}_1 = \mathbf{0}, h_2 = 0, h_3 = 0$.

Suppose that $Q^{ps}(\mathbf{h}_1, h_2, h_3) = 0$. Then $\dot{S}^{ps}(\beta_0, \Lambda_0, \phi_0)(\beta - \beta_0, \Lambda - \Lambda_0, \phi - \phi_0)[\mathbf{h}_1, h_2, h_3] = 0$ for any (β, Λ, ϕ) in the neighborhood $\mathcal{U}$. In particular, we take $\beta = \beta_0 + \epsilon \mathbf{h}_1, \Lambda = \Lambda_0 + \epsilon h_2, \phi = \phi_0 + \epsilon h_3$ for a small constant ϵ. Thus we have

$$\begin{aligned} 0 &= \dot{S}^{ps}(\beta_0, \Lambda_0, \phi_0)(\beta - \beta_0, \Lambda - \Lambda_0, \phi - \phi_0)[\mathbf{h}_1, h_2, h_3] \\ &= -\epsilon E\left[\exp\{\beta_0' Z + V\phi_0(W)\} \sum_{j=1}^K \Lambda_0(T_{K,j}) \left\{\mathbf{h}_1' Z + V h_3(W) + \frac{h_2(T_{K,j})}{\Lambda_0(T_{K,j})}\right\}^2\right], \end{aligned}$$

which yields

$$\mathbf{h}_1' Z + V h_3(W) + \frac{h_2(T_{K,j})}{\Lambda_0(T_{K,j})} = 0, \; j = 1, \ldots, K, \;\; a.s.$$

and so $\mathbf{h}_1 = \mathbf{0}, h_2 = 0, h_3 = 0$ by C7.

Next we show that (a5) holds. Write

$$\begin{aligned} & S^{ps}(\hat{\theta}_n^{ps})[\mathbf{h}_1, h_2, h_3] - S^{ps}(\theta_0)[\mathbf{h}_1, h_2, h_3] \\ & - \dot{S}(\beta_0, \Lambda_0, \phi_0)(\hat{\beta}_n^{ps} - \beta_0, \hat{\Lambda}_n^{ps} - \Lambda_0, \hat{\phi}_n^{ps} - \phi_0)[\mathbf{h}_1, h_2, h_3] \\ & = B_{n1} + B_{n2} + B_{n3} \end{aligned}$$

where

$$B_{n1} = A_1^{ps}(\hat{\theta}_n^{ps})[\mathbf{h}_1] - \frac{d}{d\varepsilon}\left\{A_1^{ps}(\theta_0 + \varepsilon(\hat{\theta}_n^{ps} - \theta_0))[\mathbf{h}_1]\right\}\Big|_{\varepsilon=0},$$

$$B_{n2} = A_2^{ps}(\hat{\theta}_n^{ps})[h_2] - \frac{d}{d\varepsilon}\left\{A_2^{ps}(\theta_0 + \varepsilon(\hat{\theta}_n^{ps} - \theta_0))[h_2]\right\}\Big|_{\varepsilon=0},$$

and

$$B_{n3} = A_3^{ps}(\hat{\theta}_n^{ps})[h_3] - \frac{d}{d\varepsilon}\left\{A_3^{ps}(\theta_0 + \varepsilon(\hat{\theta}_n^{ps} - \theta_0))[h_3]\right\}\Big|_{\varepsilon=0}$$

It is easy to show that $B_{n1} = O_p(d_1^2(\hat{\theta}_n^{ps}, \theta_0))$, $B_{n2} = O_p(d_1^2(\hat{\theta}_n^{ps}, \theta_0))$, and $B_{n3} = O_p(d_1^2(\hat{\theta}_n^{ps}, \theta_0))$. Thus, by Theorem 3.2, (a5) holds for $0 < v < 1/4$.

It follows from (B.4), (a1), (a2) and (a5) that

$$\begin{aligned}
&\sqrt{n}(\hat{\beta}_n^{ps} - \beta_0)' Q_1^{ps}(\mathbf{h}_1, h_2, h_3) + \sqrt{n}\int\{\hat{\Lambda}_n^{ps}(t) - \Lambda_0(t)\} dQ_2^{ps}(\mathbf{h}_1, h_2, h_3)(t) \\
&+\sqrt{n}\int\{\hat{\phi}_n^{ps}(w) - \phi_0(w)\} dQ_3^{ps}(\mathbf{h}_1, h_2, h_3)(w) \\
&= -\sqrt{n}(S_n^{ps} - S^{ps})(\beta_0, \Lambda_0, \phi_0)[\mathbf{h}_1, h_2, h_3] + o_p(1),
\end{aligned}$$

uniformly in $\mathbf{h}_1$, h_2 and h_3.

For each $(\mathbf{h}_1, h_2, h_3) \in \mathcal{H}$, since Q^{ps} is invertible, there exists $(\mathbf{h}_1^{ps}, h_2^{ps}, h_3^{ps}) \in \mathcal{H}$ such that

$$Q_1^{ps}(\mathbf{h}_1^{ps}, h_2^{ps}, h_3^{ps}) = \mathbf{h}_1, Q_2^{ps}(\mathbf{h}_1^{ps}, h_2^{ps}, h_3^{ps}) = h_2, Q_3^{ps}(\mathbf{h}_1^{ps}, h_2^{ps}, h_3^{ps}) = h_3.$$

Therefore, we have

$$\begin{aligned}
&\mathbf{h}_1'\sqrt{n}(\hat{\beta}_n^{ps} - \beta_0) + \sqrt{n}\int\{\hat{\Lambda}_n^{ps}(t) - \Lambda_0(t)\} dh_2(t) \\
&+\sqrt{n}\int\{\hat{\phi}_n^{ps}(w) - \phi_0(w)\} dh_3(w) \\
&= -\sqrt{n}(S_n^{ps} - S^{ps})(\beta_0, \Lambda_0, \phi_0)[\mathbf{h}_1^{ps}, h_2^{ps}, h_3^{ps}] + o_p(1) \\
&\to_d N(0, \sigma_{ps}^2),
\end{aligned}$$

where

$$\sigma_{ps}^2 = E\{\psi_{ps}^2(\beta_0, \Lambda_0, \phi_0)[\mathbf{h}_1^{ps}, h_2^{ps}, h_3^{ps}]\}. \qquad (B.5)$$

To prove part (ii), we define a sequence of maps S_n mapping a neighborhood of $(\beta_0, \Lambda_0, \phi_0)$, $\mathcal{U}$, in the parameter space for (β, Λ, ϕ) into $l^\infty(\mathcal{H})$ as:

$$S_n(\theta)[\mathbf{h}_1, h_2, h_3] = n^{-1} \frac{d}{d\varepsilon} l_n(\beta + \varepsilon \mathbf{h}_1, \Lambda + \varepsilon h_2, \phi + \varepsilon h_3)\Big|_{\varepsilon=0}.$$

Write $\Delta N_i(T_{K_i,j}) = N_i(T_{K_i,j}) - N_i(T_{K_i,j-1})$, $\Delta\Lambda(T_{K_i,j}) = \Lambda(T_{K_i,j}) - \Lambda(T_{K_i,j-1})$, and $\Delta h(T_{K_i,j}) = h(T_{K_i,j}) - h(T_{K_i,j-1})$.

Then, we have

$$\begin{aligned}
& S_n(\theta)[\mathbf{h}_1, h_2, h_3] \\
&= n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{K_i} [\{\Delta N_i(T_{K_i,j}) - \Delta\Lambda(T_{K_i,j}) \exp(\beta' Z_i + V_i \phi(W_i))\} \mathbf{h}_1' Z_i \\
&\qquad + \left\{ \frac{\Delta N_i(T_{K_i,j})}{\Delta\Lambda(T_{K_i,j})} - \exp(\beta' Z_i + V_i \phi(W_i)) \right\} \Delta h_2(T_{K_i,j}) \\
&\qquad + \{\Delta N_i(T_{K_i,j}) - \Delta\Lambda(T_{K_i,j}) \exp(\beta' Z_i + V_i \phi(W_i))\} V_i h_3(W_i)] \\
&\equiv A_{n1}(\theta)[\mathbf{h}_1] + A_{n2}(\theta)[h_2] + A_{n3}(\theta)[h_3] \\
&\equiv P_n(\mathbf{h}_1' \dot{l}_\beta) + P_n(\dot{l}_\Lambda[h_2]) + P_n(\dot{l}_\phi[h_3]) \\
&\equiv P_n \psi(\theta)[\mathbf{h}_1, h_2, h_3].
\end{aligned}$$

Correspondingly, we define the limit map $S : \mathcal{U} \longrightarrow l^\infty(\mathcal{H})$ as

$$S(\theta)[\mathbf{h}_1, h_2, h_3] = A_1(\theta)[\mathbf{h}_1] + A_2(\theta)[h_2] + A_3(\theta)[h_3],$$

where

$$A_1(\theta)[\mathbf{h}_1] = E\left[\sum_{j=1}^{K} \{\Delta N(T_{K,j}) - \Delta\Lambda(T_{K,j}) \exp(\beta' Z + V\phi(W))\} \mathbf{h}_1' Z \right],$$

$$A_2(\theta)[h_2] = E\left[\sum_{j=1}^{K} \left\{ \frac{\Delta N(T_{K,j})}{\Delta\Lambda(T_{K,j})} - \exp(\beta' Z + V\phi(W)) \right\} \Delta h_2(T_{K,j}) \right],$$

and

$$A_3(\theta)[h_3] = E\left[\sum_{j=1}^{K}\{\Delta N(T_{K,j}) - \Delta\Lambda(T_{K,j})\exp(\beta' Z + V\phi(W))\}Vh_3(W)\right].$$

Furthermore, the derivative of $S(\beta, \Lambda, \phi)$ at $(\beta_0, \Lambda_0, \phi_0)$, denoted by $\dot{S}(\beta_0, \Lambda_0, \phi_0)$, is a map from the space $\{(\beta - \beta_0, \Lambda - \Lambda_0, \phi - \phi_0) : (\beta, \Lambda, \phi) \in \mathcal{U}\}$ to $l^\infty(\mathcal{H})$ and

$$\begin{aligned}
&\dot{S}(\beta_0, \Lambda_0, \phi_0)(\beta - \beta_0, \Lambda - \Lambda_0, \phi - \phi_0)[\mathbf{h}_1, h_2, h_3] \\
&= (\beta - \beta_0)' Q_1(\mathbf{h}_1, h_2, h_3) + \int \{\Lambda(t) - \Lambda_0(t)\} dQ_2(\mathbf{h}_1, h_2, h_3)(t) \\
&+ \int \{\phi(w) - \phi_0(w)\} dQ_3(\mathbf{h}_1, h_2, h_3)(w)
\end{aligned}$$

where

$$\begin{aligned}
&Q_1(\mathbf{h}_1, h_2, h_3) \\
&= -E\left[Z\exp\{\beta_0' Z + V\phi_0(W)\}\right. \\
&\quad \left. \times \sum_{j=1}^{K}\{\Delta\Lambda_0(T_{K,j})\mathbf{h}_1' Z + \Delta h_2(T_{K,j}) + \Delta\Lambda_0(T_{K,j})Vh_3(W)\}\right],
\end{aligned}$$

$$\begin{aligned}
&dQ_2(\mathbf{h}_1, h_2, h_3)(t) \\
&= -E\left[\exp\{\beta_0' Z + V\phi_0(W)\}\right. \\
&\quad \times \sum_{j=1}^{K}\left\{\left(\mathbf{h}_1' Z + \frac{h_2(t) - h_2(T_{K,j-1})}{\Lambda_0(t) - \Lambda_0(T_{K,j-1})} + Vh_3(W)\right) dP(T_{K,j} \leqslant t | K, T_{K,j-1}, Y)\right. \\
&- \left.\left.\left(\mathbf{h}_1' Z + \frac{h_2(T_{K,j}) - h_2(t)}{\Lambda_0(T_{K,j}) - \Lambda_0(t)} + Vh_3(W)\right) dP(T_{K,j-1} \leqslant t | K, T_{K,j}, Y)\right\}\right],
\end{aligned}$$

and

$$
\begin{aligned}
& dQ_3(\mathbf{h}_1, h_2, h_3)(w) \\
& = -E\left[V \exp\{\beta_0' Z + V\phi_0(w)\}\right. \\
& \quad \left. \times \sum_{j=1}^{K} \{\Delta\Lambda_0(T_{K,j})\mathbf{h}_1' Z + \Delta h_2(T_{K,j}) + \Delta\Lambda_0(T_{K,j}) V h_3(w)\} \,|W = w\right] dF_W(w).
\end{aligned}
$$

Next, we show that $Q = (Q_1, Q_2, Q_3)$ is one-to-one, that is, for $h \in \mathcal{H}$, if $Q(\mathbf{h}_1, h_2, h_3) = 0$, then $\mathbf{h}_1 = \mathbf{0}, h_2 = 0, h_3 = 0$

Suppose that $Q(\mathbf{h}_1, h_2, h_3) = 0$. Then $\dot{S}(\beta_0, \Lambda_0, \phi_0)(\beta - \beta_0, \Lambda - \Lambda_0, \phi - \phi_0)[\mathbf{h}_1, h_2, h_3] = 0$ for any (β, Λ, ϕ) in the neighborhood $\mathcal{U}$. In particular, we take $\beta = \beta_0 + \epsilon\mathbf{h}_1, \Lambda = \Lambda_0 + \epsilon h_2, \phi = \phi_0 + \epsilon h_3$ for a small constant ϵ. Thus we have

$$
\begin{aligned}
0 &= \dot{S}(\beta_0, \Lambda_0, \phi_0)(\beta - \beta_0, \Lambda - \Lambda_0, \phi - \phi_0)[\mathbf{h}_1, h_2, h_3] \\
&= -\epsilon E\left[\exp\{\beta_0' Z + V\phi_0(W)\} \sum_{j=1}^{K} \Delta\Lambda_0(T_{K,j}) \left\{\mathbf{h}_1' Z + V h_3(W) + \frac{\Delta h_2(T_{K,j})}{\Delta\Lambda_0(T_{K,j})}\right\}^2\right],
\end{aligned}
$$

which yields

$$
\mathbf{h}_1' Z + V h_3(W) + \frac{\Delta h_2(T_{K,j})}{\Delta\Lambda_0(T_{K,j})} = 0, \; j = 1, \ldots, K, \;\; a.s.
$$

and so $\mathbf{h}_1 = \mathbf{0}, h_2 = 0, h_3 = 0$ by C7.

Similarly, we can show that $S(\beta_0, \Lambda_0, \phi_0) = 0$, $S_n(\hat{\beta}_n, \hat{\Lambda}_n, \hat{\phi}_n) = o_p(n^{-1/2})$, and

$$
\begin{aligned}
& S(\hat{\theta}_n)[\mathbf{h}_1, h_2, h_3] - S(\theta_0)[\mathbf{h}_1, h_2, h_3] \\
& = \dot{S}(\beta_0, \Lambda_0, \phi_0)(\hat{\beta}_n - \beta_0, \hat{\Lambda}_n - \Lambda_0, \hat{\phi}_n - \phi_0)[\mathbf{h}_1, h_2, h_3] + O_p(d_2^2(\hat{\theta}_n, \theta_0)) \\
& = \dot{S}(\beta_0, \Lambda_0, \phi_0)(\hat{\beta}_n - \beta_0, \hat{\Lambda}_n - \Lambda_0, \hat{\phi}_n - \phi_0)[\mathbf{h}_1, h_2, h_3] + o_p(n^{-1/2}).
\end{aligned}
$$

for $0 < v < 1/4$. Thus it follows that

$$\sqrt{n}(\hat{\beta}_n - \beta_0)' Q_1(\mathbf{h}_1, h_2, h_3) + \sqrt{n} \int \{\hat{\Lambda}_n(t) - \Lambda_0(t)\} dQ_2(\mathbf{h}_1, h_2, h_3)(t)$$

$$+\sqrt{n} \int \{\hat{\phi}_n(w) - \phi_0(w)\} dQ_3(\mathbf{h}_1, h_2, h_3)(w)$$

$$= -\sqrt{n}(S_n - S)(\beta_0, \Lambda_0, \phi_0)[\mathbf{h}_1, h_2, h_3] + o_p(1),$$

uniformly in $\mathbf{h}_1$, h_2 and h_3.

For each $(\mathbf{h}_1, h_2, h_3) \in \mathcal{H}$, since Q is invertible, there exists $(\mathbf{h}_1^*, \mathbf{h}_2^*, h_3^*) \in \mathcal{H}$ such that

$$Q_1^*(\mathbf{h}_1^*, h_2^*, h_3^*) = \mathbf{h}_1, Q_2^*(\mathbf{h}_1^*, h_2^*, h_3^*) = 0, Q_3^*(\mathbf{h}_1^*, h_2^*, h_3^*) = h_3.$$

Thus, we have

$$\mathbf{h}_1' \sqrt{n}(\hat{\beta}_n - \beta_0) + \sqrt{n} \int \{\hat{\Lambda}_n(t) - \Lambda_0(t)\} dh_2(t)$$

$$+\sqrt{n} \int \{\hat{\phi}_n(w) - \phi_0(w)\} dh_3(w)$$

$$= -\sqrt{n}(S_n - S)(\beta_0, \Lambda_0, \phi_0)[\mathbf{h}_1^*, h_2^*, h_3^*] + o_p(1)$$

$$\rightarrow_d N(0, \sigma^2),$$

where

$$\sigma^2 = E\{\psi^2(\beta_0, \Lambda_0, \phi_0)[\mathbf{h}_1^*, h_2^*, h_3^*]\}. \qquad (B.6)$$

Bibliography

[1] Aalen, O. (1975). Statistical inference for a family of counting presses. Ph.D. dissertation, University of California, Berkeley.

[2] Aalen, O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics* 6(4): 701726.

[3] Andersen, P. K. (1982). Testing goodness of fit of Cox's regression and life model. *Biometrics*, 67-77.

[4] Andersen, P. K. and Borgan, Ø.(1985). Counting process models for life history data: A review (with discussion). *Scandinavian Journal of Statistics*, 12: 97-158.

[5] Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1982). Linear nonparametric tests for comparison of counting processes, with applications to censored survival data (with discussion). *Int. Statist.Rev.* 50: 219-258. Correction 52: 225.

[6] Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.

[7] 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.

[8] 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.

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

[10] Balakrishnan, N. and Zhao, X. (2010). Nonparametric test for the equality of counting processes with panel count data. *Computational Statistics and Data Analysis* 54, 135-142.

[11] Balshaw, R. F. and Dean, C. B. (2002). A semiparametric model for the analysis of recurrent event panel data. *Biometrics*, 58(2): 324-331.

[12] Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. (1972). *Statistical Inference under Order Restrictions*. New York: Wiley.

[13] Bickel, P. J., Klassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: Johns Hopkins University Press.

[14] Breslow, N. E. (1972). Contribution to the discussion of the paper by D. R. Cox. *Journal of the Royal Statistical Society*, Series B 34(2): 216-217.

[15] Byar, D. P. (1980). The Veterans Administration study of chemoprophylaxis for recurrent stage I bladder tumors: comparisons of placebo, pyridoxine, and topical thiotepa. In *Bladder tumors and other topics in urological oncology*, M. Pavane-Macaluso, P. H. Smith, F. Edsmyr (eds), 363-370. New York: Plenum.

[16] 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.

[17] Cheng, G., Zhang, Y. and Lu, L. (2011). Efficient algorithms for computing the non and semi-parametric maximum likelihood estimates with panel count data. *Journal of Nonparametric Statistics* 23(2): 567-579.

[18] Cheng, M. Y., Honda, T., Li, J. and Peng, H. (2014). Nonparametric independence screening and structure identification for ultra-high dimensional longitudinal data. *The Annals of Statistics* 42: 1819-1849.

[19] Cheng, S. C. and Wei, L. J. (2000). Inferences for a semiparametric model with panel data. *Biometrika* 87: 89-97.

[20] Cook, R. J. and Lawless, J. F. (1996). Interim monitoring of longitudinal comparative studies with recurrent event responses. *Biometrics* 52: 1311-1323.

[21] Cook, R. J., and Lawless, J. F. (2007). *The statistical analysis of recurrent events*. Springer Science + Business Media.

[22] Cook, R. J., Lawless, J. F. and Nadeau, C. (1996). Robust tests for treatment comparisons based on recurrent event response. *Biometrics* 52: 557-571.

[23] Cox, D. R. and Oakes, D. (1984). *Analysis of survival data* (Vol. 21). Chapman & Hall/CRC.

[24] Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis* (Vol. 169). John Wiley & Sons.

[25] Ghosh, D. and Lin, D. Y. (2000). Nonparametric analysis of recurrent events and death. *Biometrics* 56: 554-562.

[26] Ghosh, D. and Lin, D. Y. (2002). Marginal regression models for recurrent and terminal events. *Statistica Sinica* 12: 663-688.

[27] Groeneboom, P. and Wellner, J. A. (1992). *Information bounds and nonparametric maximum likelihood estimation*. Springer Basel AG.

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

[29] Hu, X.J., Sun, J. and Wei, L.J. (2003). Regression parameter estimation from panel counts. *Scandinavian Journal of Statistics*, 30(1): 25-43.

[30] Hu, X. J., Lagakos, S. W., and Lockhart, R. A. (2009a). Marginal analysis of panel counts through estimating functions. *Biometrika* 96(2): 445-456.

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

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

[33] 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.

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

[35] Huang, J. (1999). Efficient estimation of the partly linear additive Cox model. *The Annals of Statistics* 27: 1536-1563.

[36] Jongbloed, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *Journal of Computational and Graphical Statistics* 7(3): 310-321.

[37] Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data* (Vol. 360). John Wiley & Sons.

[38] Klein, J. P. and Moeschberger, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer-Verlag, New York.

[39] Lawless, J. F. (1983). Statistical methods in reliability. *Technometrics*, 25(4): 305-316.

[40] Lawless, J. F. (1995). The analysis of recurrent events for multiple subjects. *Applied Statistics*, 487-498.

[41] Lawless, J. F. and Nadeau, J. C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics* 37:158-168.

[42] Lee, L-Y. (2008). Nonparametric and semiparametric models for multivariate panel count data. Ph.D. Dissertation, University of Wisconsin-Madison.

[43] Li, N., Park, D-H., Sun, J. and Kim, K. (2011). Semiparametric transformation models for multivariate panel count data with dependent observation process. *The Canadian Journal of Statistics*, 39: 458-474.

[44] Li, Y., He, X., Wang, H., Zhang, B., and Sun, J. (2015). Semiparametric regression of multivariate panel count data with informative observation times. *Journal of Multivariate Analysis* 140: 209-219.

[45] 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 Royal Statistical Society* Series B 62:711-730.

[46] Lin, D. Y., Wei, L. J., and Ying, Z. (2001). Semiparametric transformation models for point processes. *Journal of the American Statistical Association* 96(454): 620-628.

[47] Lindqvist, B. H., Kvalϕy, J. T. and Aaserud, S. (2015). Residual plots to reveal the functional form for covariates in parametric accelerated failure time models. *Lifetime Data Analysis* 21: 353-378.

[48] Liu, L., Wolfe, R. A. and Huang, X. (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics* 60: 747-756.

[49] Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of Royal Statistical Society* Series B 44: 226-233.

[50] Lu, M., Zhang, Y. and Huang, J. (2007). Estimation of the mean function with panel count data using monotone polynomial splines. *Biometrika*, 94(3): 705-718.

[51] 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(487): 1060-1070.

[52] Mazroui, Y., Mauguen, A., Mathoulin-Plissier, S., MacGrogan, G., Brouste, V. and Rondeau, V. (2015). Time-varying coefficients in a multivariate frailty model: Application to breast cancer recurrences of several types and death. *Lifetime data analysis*, 1-25.

[53] Pénichoux, J., Moreau, T. and Latouche, A. (2015). Simulating recurrent events that mimic actual data: a review of the literature with emphasis on event-dependence. arXiv:1503.05798.

[54] Pepe, M. S. and Cai, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *Journal of the American Statistical Association*, 88(423): 811-820.

[55] Prentice, R. L., Williams, B. J. and Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika* 68: 373-379.

[56] Ramsay, J. O. (1988). Monotone regression splines in action. *Statist. Sci.* 3, 425-441.

[57] Rubnitz, J., Razzouk, B., Bowman, P., Dahl, G., Lacayo, N., Taub, J., Ravindranath, Y., Meshinichi, S., Airewele, G., Lensing, S., Pounds, S., Ribeiro, R., Pui, C. H., Coustan-Smith, E. and Campana, D. (2005). Improved remission induction rate of childhood AML: preliminary results of the AML02 trial. *Blood (ASH Annual Meeting Abstracts 2005)* 106, Abstract 275.

[58] Schumaker, L. (2007). *Spline functions: basic theory (3rd Ed.)*. Cambridge University Press.

[59] Staniswalis, J.G., Thall, P.F. and Salch, J. (1997). Semiparametric regression analysis for recurrent event interval counts. *Biometrics*, 1334-1353.

[60] Stukel, T.A. (1993). Comparison of methods for the analysis of longitudinal interval count data. *Statistics in medicine*, 12(14): 1339-1351.

[61] Sun, J. (1999). A nonparametric test for current status data with unequal censoring. *Journal of the Royal Statistical Society* Series B: 243-250.

[62] Sun, J. and Kalbfleisch, J.D. (1993). The analysis of current status data on point processes. *Journal of the American Statistical Association* 88: 1449-1454

[63] Sun, J. and Kalbfleisch, J.D. (1995). Estimation of the mean function of point processes based on panel count data. *Statistica Sinica*, 5: 279-290.

[64] Sun, J. and Wei, L.J. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *Journal of Royal Statistical Society* Series B 62(2): 293-302.

[65] Sun, J. and Zhao, X. (2013). *Statistical Analysis of Panel Count Data*. Springer Science + Business Media.

[66] Sun, J., Tong, X. and He, X. (2007). Regression analysis of panel count data with dependent observation times. *Biometrics* 63: 1053-1059.

[67] Sun, L., Guo, S. and Chen, M. (2009). Marginal regression model with time-varying coefficients for panel data. *Communications in Statistics -Theory and Methods* 38: 1241-1261.

[68] Sun, L., Zhou, X. and Guo, S. (2011). Marginal regression models with time-varying coefficients for recurrent event data. *Statistics in Medicine* 30(18): 2265-2277.

[69] Sun, L., Zhu, L. and Sun, J. (2009). Regression analysis of multivariate recurrent event data with time-varying covariate effects. *Journal of Multivariate Analysis* 100: 2214-2223.

[70] Thall, P. F. (1988). Mixed Poisson likelihood regression models for longitudinal interval count data. *Biometrics*, 44: 197-209.

[71] Thall, P. F. and Lachin, J. M. (1988). Analysis of recurrent events: Nonparametric methods for random-interval count data. *Journal of America Statistical Association*, 83: 339-347.

[72] Tsiatis, A. A. (1981). A large sample study of Cox's regression model. *The Annals of Statistics*, 93-108.

[73] Van der Vaart, A. W. (2002). Semiparametric Statistics. In *Lectures on Probability Theory and Statistics, Ecole dEté de Probabilités de Saint-Flour XXIX-1999, Lecture Notes in Mathematics* 1781: 330-457.

[74] Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer-Verlag.

[75] Wang, M. C., Qin J. and Chiang, C. T. (2001). Analyzing recurrent event data with informative censoring. *Journal of America Statistical Association*, 96: 1057-1065.

[76] Wellner, J. A. and Zhang, Y. (2000). Two Estimators of the Mean of a Counting Process with Panel Count Data. *The Annals of Statistics*, 779-814.

[77] Wellner, J. A. and Zhang, Y. (2007). Two likelihood-based semiparametric estimation methods for panel count data with covariates. *The Annals of Statistics*, 35(5): 2106-2142.

[78] Wellner, J. A., Zhang, Y. and Liu, H. (2004). A semiparametric regression model for panel count data: When do pseudo-likelihood estimators become badly inefficient?, 143-174. *Proceedings of the Second Seattle Symposium in Biostatistics: Analysis of Correlated Data*, Lecture Notes in Statistics. Springer, New York.

[79] Ye, Y., Kalbfleisch, J. D. and Schaubel D. E. (2007). Semiparametric analysis of correlated recurrent and terminal events. *Biometrics* 63: 78-87.

[80] Yu, Z., Liu, L., Bravata, D. M., Williams, L. S. and Tepper, R. S. (2013). A semiparametric recurrent events model with time-varying coefficients. *Statistics in medicine* 32(6): 1016-1026.

[81] Zeng, D. L. and Lin, D. Y. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika* 93: 627-640.

[82] Zeng, D. L., Lin, D. Y. and Yin, G. S. (2005). Maximum likelihood estimation for the proportional odds model with random effects. *Journal of America Statistical Association*, 100: 470-482.

[83] Zhang, F., Chen, X. and Zhou, Y. (2014). Proportional hazards model with varying coefficients for length-biased data. *Lifetime Data Analysis*, 20: 132-157.

[84] Zhang, H. Zhao, H., Sun, J., Wang, D. and Kim, K. M. (2013). Regression analysis of multivariate panel count data with an informative observation process. *Journal of Multivariate Analysis*, 119: 71-80.

[85] Zhang, Y. (1998). Estimation for counting processes based on incomplete data. Unpublished Ph.D. dissertation, Univ. Washington.

[86] Zhang, Y. (2002). A semiparametric pseudo-likelihood estimation method for panel count data. *Biometrika*, 89(1): 39-48.

[87] Zhang, Y. (2006). Nonparametric k-sample tests with panel count data. *Biometrika*, 93(4): 777-790.

[88] Zhang, Y. and Jamshidian, M. (2003). The gamma-frailty Poisson model for the nonparametric estimation of panel count data. *Biometrics* 59: 1099-1106.

[89] Zhao, H., Li, Y. and Sun, J. (2013a). Analyzing panel count data with dependent observation process and a terminal event. *The Canadian Journal of Statistics* 41: 174-191.

[90] Zhao, H., Li, Y. and Sun, J. (2013b). Semiparametric analysis of multivariate panel count data with dependent observation process and terminal event. *Journal of Nonparametric Statistics* 25: 379-394.

[91] Zhao, H., Virkler, K. and Sun, J. (2014). Nonparametric comparison for multivariate panel count data. *Communications in Statistics - Theory and Methods* 43(3): 644-655.

[92] Zhao, X., Balakrishnan, N. and Sun, J. (2011). Nonparametric inference based on panel count data. *Test* 20, 1-42.

[93] Zhao, X. and Sun, J. (2011). Nonparametric comparison for panel count data with unequal observation processes. *Biometrics* 67(3): 770-779.

[94] Zhu, L., Tong, X. W., Zhao, H., Sun, J., Srivastava, D. K., Leisenring, W., and Robison, L. (2013). Statistical analysis of mixed recurrent event data with application to cancer survivor study. *Statistics in Medicine* 32: 1954-1963.

[95] Zhu, L., Tong, X. W., Sun, J., Chen, M., Srivastava, D. K., Leisenring, W. and Robison, L. L. (2014). Regression analysis of mixed recurrent-event and panel-count data. *Biostatistics* 15(3): 555-568.

[96] Zhu, L., Zhao, H., Sun, J., Leisenring, W. and Robison, L. L. (2015). Regression Analysis of Mixed Recurrent-Event and Panel-Count Data with Additive Rate Models. *Biometrics* 71: 71-79.