DYNAMIC ANALYSIS OF COMPLEX PANEL COUNT DATA

A Dissertation
presented to
the Faculty of the Graduate School
at the University of Missouri-Columbia

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

---

by
YUANYUAN GUO
Dr. (Tony) Jianguo Sun, Dissertation Supervisor
MAY 2021

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled:

DYNAMIC ANALYSIS OF COMPLEX PANEL
COUNT DATA

presented by Yuanyuan Guo,

a candidate for the degree of Doctor of Philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. (Tony) Jianguo Sun

Dr. Shih-Kang Chao

Dr. Yushu Shi

Dr. Shawn Ni

ACKNOWLEDGMENTS

First and foremost, I owe my deepest gratitude to my esteemed advisor Dr. (Tony) Jianguo Sun for his wonderful direction and extensive support throughout the development of this dissertation. His critical encouragement and inspiring instruction were invaluable for my doctoral studies and guaranteed the accomplishment of this work.

I extend my gratitude to my advisory committee members, Drs. Shih-Kang Chao, Xiaoguang Ni, and Yushu Shi, for their insightful comments and advice on my work. Special thanks are due to Dr. Yang Li for her generous academic support and Dr. Peter Tonellato for his great advice on my collaboration work at CBMI.

I also owe a debt of gratitude to all faculty in the Department of Statistics, especially those who have taught me, for their inspiration and encouragement in my graduate studies. I am grateful to Dr. Larry Ries, who provided me with helpful guidance and precious advice for being a better instructor. I also appreciate the plenty of help by our great staff Judy Dooley and Abbie Van Nice Booher. In addition, I want to take the opportunity to thank all my classmates and friends I met here. Because of them, my life in Mizzou is so enjoyable and memorable.

In the end, I especially owe my gratitude to my parents and parents-in-law for their love and support throught my life. Thank you all for giving me strength to pursue an academic career. I also want to express my appreciation for my husband Dayu, for his endless love, best care, and generous support, without whom I could not have accomplished what I have today.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure | Page

DYNAMIC ANALYSIS OF COMPLEX PANEL COUNT DATA

Yuanyuan Guo

Dr. (Tony) Jianguo Sun, Dissertation Supervisor

ABSTRACT

Panel count data occur in many fields including clinical, demographical and industrial studies and an extensive literature has been established for their regression analysis. However, most of the existing methods apply only to the situations where both covariates and their effects are constant or one of them may be time-dependent. In the first part of this dissertation, we consider a situation where both covariates and their effects may be time-dependent and an estimating equation-based approach is developed for estimating those time-varying effects. In the method, B-spline functions are employed to approximate time-dependent coefficients and the asymptotic properties of the proposed estimators are established. To assess the performance of the proposed approach, an extensive simulation study is conducted and suggests that it works well in practical situations. An application to the China Health and Nutrition Survey (CHNS) study is provided.

In practice, there could exist more than one type of event of interest, such as two types of tumor recurrence, leading to multivariate panel count data. The second part of this dissertation considers marginal mean model for multivariate panel count data with time-dependent coefficient and covariate effects, which has limited previous

research. Based on the conditional estimating equation method developed for time-dependent covariates, we approximate the coefficients by B-splines, hence allow both coefficients and covariates to be time-dependent. Simulation studies show that the proposed estimation procedures work well for practical situations. The methodology is again applied to the China Health and Nutrition Survey (CHNS) study.

When we consider time-varying covariates and coefficients effects, most of the previous study focused on the proportional mean model because the likelihood function under the rate model involves intractable integration. However, the rate model is more realistic and efficient. Hence, in the third part of this dissertation, we propose a semi-parametric MLE method under the rate model for panel count data with time-dependent covariates and time-varying effects. B-spline functions are employed again to approximate time-dependent coefficients and an efficient Expectation-Maximization-type algorithm is developed to overcome the computational difficulty. The resulting estimators are shown to be consistent and asymptotically efficient. Monte Carlo simulation studies demonstrate that the proposed method enjoys desirable finite-sample properties. An application to The Young Women's Project (YWP) is provided.

Chapter 1

Introduction

1.1 Introduction to Panel Count Data

Event history studies concerning some recurrent events are often conducted in many fields, including clinical, demographical and industrial studies. For the situation, two types of data commonly occur, recurrent event data and panel count data (Cook and Lawless (2007); Sun and Zhao (2013)). The former means that all study subjects can be observed or followed continuously and thus one has complete data on the occurrences of the event of interest. In contrast, the latter means that study subjects can be observed only at discrete time points and only incomplete information is available on the occurrences.

A large literature has been established for the analysis of panel count data and in particular, the pseudo-likelihood approach is often used for their regression analysis. For example, Zhang (2002), Wellner and Zhang (2007) and Zhu et al. (2017)

developed such procedures under the nonhomogeneous Poisson process assumption for the underlying point process. Zhang and Jamshidian (2004) considered the same problem with the use of the spline-based approach. Also Hua et al. (2014) proposed a sieve maximum likelihood method under the Gamma-Frailty nonhomogeneous Poisson process assumption for the underlying point process.

1.2 Time-varying Coefficients and Covariates

In reality, it is apparent that one may face the situation that involves time-dependent covariates and in which there also exist some time-varying covariate effects. A common example is that treatment effects may take some time to be effective and then gradually disappear after some time. We can identify the similar pattern from the China Health and Nutrition Survey (CHNS), in which it is naturally to assume that mothers' fertility desire may be affected by their income, locations, education levels and health statues. Furthermore, those effects may change with ages of mothers and may disappear after a certain age.

Although some estimation procedures have been developed for either time-varying coefficients or covariates, they mainly focus on failure time data situations. For example, Perperoglou (2013) and Perperoglou (2012) gave some B-spline-based methods, and Tian et al. (2005), Cai et al. (2007), Yu and Lin (2010) and Lin et al. (2015) proposed some kernel-weighted likelihood methods. In addition, Sun et al. (2009) discussed regression analysis of multivariate recurrent event data with time-dependent covariate effects. Both Zhao et al. (2018) and Wang and Yu (2021b) considered the situation of time-varying coefficients and proposed some pseudo-likelihood methods for estimation for panel count data. In their methods, the former used B-spline function approximation and the latter employed local polynomials.

One drawback of these methods is that they only apply to the situation where either covariates or their effects are constant. To address this, in this dissertation, we will discuss regression analysis of panel count data that involves both time-dependent covariates and time-varying covariate effects and propose some spline-based approaches.

1.3 Multivariate Panel Count Data

Multivariate panel count data arise in studies involving several types of recurrent events in which patients are examined only at periodic follow-up assessments. In such settings, observations are taken at several distinct time points and only the number of different events that occur between observation times is known; no information is available on subjects between the observation time points. This frequently happens in prospective cohort studies, population-based epidemiological studies, reliability studies, and tumorigenicity experiments, in which, it is either impossible or not practical to maintain continuous observation of subjects. The second part of this dissertation discusses analysis of multivariate panel count data.

An common example arises in tumorigenicity experiments when several types of tumors can occur together and are of interest. In He et al. (2007), they considered another example arising from a cohort study of patients with psoriatic arthritis conducted at the University of Toronto Psoriatic Arthritis Clinic where the event of interest is the development of joint damage. Clinicians are interested in damage as measured by radiographic changes as well as loss in function as detected by functional examination, and these constitute the 2 types of events. We consider a third example arising from the China health and nutrition study(CHNS), which is an international collaborative project between the Carolina Population Center at the University of North Carolina at Chapel Hill and the National Institute for Nutrition and Health at the Chinese Center for Disease Control and Prevention (CCDC). The number of pregnancy and marriage are recorded for each patient longitudinally, which yield multivariate panel count data.

Futhermore, as mentioned before, the time-varying coefficient and covariate effects situation should be considered. Still take the CHNS data as the example, when we estimate the effect of mothers' education levels, they would change with age of the mother, and their effects may disappreare after some ages. To deal with this problem, we develop a marginal mean model of multivariate panel count data to involve time-dependent covariates and in which there also exist some time-varying covariate effects, which will be introduced in Chapter 3 in detail. By results in Section 3.6, we can found the effect of locations and education levels are varying with age of mothers, and their covariate effects disappeared after some ages.

For univariate panel count data, as introduced before, a more comprehensive review of the existing statistical methods for panel count data can be referred to the book of Sun and Zhao (2013). Recently, more models and estimation procedures have been developed including Lu et al. (2007); Lu et al. (2009); He et al. (2008); Zhao et al. (2013); Zhao et al. (2012). For multivariate panel count data, He et al. (2007) presented a class marginal mean models and developed estimating equation methods. Li et al. (2011) developed semiparametric transformation models for multivariate panel count data with dependent observation process. Zhao et al. (2013) considered regression analysis of such multivariate data in the presence of a terminal event. Futhermore, Zhang et al. (2013) proposed a robust joint model for multivariate panel count data with informative observation processes. Besides, nonparametric comparison procedures of multivariate panel count data were developed by Zhao et al. (2014).

However, the aforementioned approaches for panel count data are based on constant coeffcients and covariates assumption. As mentioned before, this assumption may be often unrealistic in practice. To deal with such problems, for univariate panel count data, He et al. (2017) developed a semiparametric partially linear varying coeffcient models. Zhao et al. (2018) and Wang and Yu (2021b) analyzed the time-varying coeffcient model by B-splines and local linear expansion respectively. To our best knowledge, there exists no related literature about both time-varying coeffcients and covariates situation for multivariate panel count data. Thus, it is necessary to develop a B-spline approximation based method for multivariate panel count data in Chapter 3.

1.4 Two Different Fomulations: Mean Model vs Rate Model

Statistical methods for panel count data usually focus on studying the relationship of covariates and the underlying recurrent event processes $N(t)$, which represents the cumulative number of events occurrence up to time t. Many studies considered a semi-parametric proportional model of the mean function of $N(t)$ conditionally on a

p-dimensional possibly time-dependent covariates $X(t)$ as

$$E\left[N(t)\,|X(t)\right] = \Lambda(t)\exp\left(\beta^T Z\right). \tag{1.1}$$

In the above, β is a p-dimensional vector of regression coefficients, and $\Lambda(t)$ is an unspecified increasing function. As introduced before, Hu et al. (2003) and Sun and Wei (2000) proposed to estimate β by an estimating equation method based on model (1.1). Lu et al. (2009) used the monotone B-spline as basis to estimate $\Lambda(t)$ by the two likelihood-based methods proposed in Wellner and Zhang (2007). He et al. (2017) discussed nonlinear interactions between covariates, which is in the form of the possibly varying coefficients for the mean function of the counting processes.

Futhermore, as claimed in Sections 1.2 and 1.3, all the studies mentioned above are based on mean model assumed that covariates and/or their corresponding effects are time-independent. Nevertheless, it is possible that they vary along with time. For example, there exist interactions between time and baseline covariates or biomarkers as covariates change over time, and their effects may vary over time, too. Based on model (1.1), Li et al. (2010) and Li et al. (2013) considered a semi-parametric transformation model with time-dependent covariates and proposed estimating equation methods; both Zhao et al. (2018) and Wang and Yu (2021b) considered the situation of time-varying coefficients in panel count data and proposed some pseudo-likelihood methods for estimation. In their methods, the former used B-spline function approximation and the latter employed local polynomials.

To the best of our knowledge, all methods analyzing panel count data with time-dependent effects are based on model (1.1). However, one drawback of model (1.1) is that, when β or Z fluctuate, it is hard to satisfy the non-decreasing property of the mean function in model (1.1). Moreover, even though the valid statistical methods can always induce estimates of β and μ such that the left hand side of model (1.1) is non-decreasing, when we predict the mean function of a new subject, there is no guarantee that the predicted mean function is non-decreasing if β or Z is time-dependent. Therefore, we consider an alternative similar semi-parametric model based on the rate or intensity function of $N(t)$, that is

$$E\left[dN(t)\,|W(t), Z(t)\right] = \exp\left(\gamma^T W(t) + \beta^T(t) Z(t)\right) d\Lambda(t). \tag{1.2}$$

Here, $\Lambda(t)$ is an unspecified non-decreasing function , $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_{p_1})^T$ and $\boldsymbol{\beta}(t) =$ $(\beta_1(t), ...,$
$\beta_{p_2}(t))^T$ represent constant and time-dependent coefficient effects, respectively. Model (1.2) only requires its right hand side to be positive. Hence it is a more flexible and realistic model to incorporate time-dependent effects in inference. Model (1.1) and model (1.2) are equivalent when covariates and their effects are time-independent over time while they are generally different when covariates and/or their effects are time-varying.

To efficiently estimate γ and β, in the third part of this dissertation, we consider a maximum likelihood estimation method under the rate model (1.2). B-spline functions are used again to approximate time-varying coefficient effects $\beta(t)$. Here, to overcome difficulties in calculation and maximize the likelihood function efficiently, we develop an Expectation-Maximization (EM) algorithm (Dempster et al., 1977) which circumvents the direct integration and reduces computational burden.

1.5 Outline of the Dissertation

The remainder of this dissertation will be organized as follows. In Chapter 2, we consider the estimating equation method with B-spline approxiamtion to deal with the time-varying coefficients and covariates for univariate panel count data. We first introduce model and notation for the proposed method. Then we will introduce the detailed estimation procedure. As discussed above, most previous methods cannot be directly applied to the desired case. We therefore propose a B-spline approximation based estimating equation method to enable estimation. We rigorously establish the consistency, rate of convergence and asymptotic normality of the proposed method. A simulation study shows the performance of the proposed method. An application to the China Health and Nutrition Survey(CHNS) study illustrates the proposed method in practice.

Chapter 3 still considers the time-varying coefficient and covariates, but for multivariate panel count data. Similarly to the method in Chapter 2 for univariate panel count data, an marginal estimating equation is constructed to combine with the B-spline approximation, which can be used to infer the dynamic feature of the multivariate panel count data. The consistency, rate of convergence and asymptotic normality of the proposed method are provided. Besides, numerical studies are carried out to illustrate the finite-sample properties of the proposed method. We then discuss the application of the proposed method to the CHNS.

Chapter 4 discusses the time-varying coefficient and covariate effects situation. However, we utilize a more flexible and efficient semi-parametric MLE method for inference. We develop the asymptotic properties of the estimators from the proposed method. An extensive simulation shows the finite-sample properties of the estimators. An application to The Young Women's Project (YWP) demonstrates that the proposed method works well in practice.

Chapter 2

Dynamic Analysis of Univariate Panel Count Data with Mean Model

2.1 Introduction

As described in Section 1.2, panel count data are common in many areas. However, there exists little research on the estimation of both time-varying coefficients and covariates. Most previous research on time-varying effects mainly focus on failure time data situations. For example, Perperoglou (2013) and Perperoglou (2012) gave some B-spline-based methods, and Tian et al. (2005), Cai et al. (2007), Yu and Lin (2010) and Lin et al. (2015) proposed some kernel-weighted likelihood methods. For panel count data, both Zhao et al. (2018) and Wang and Yu (2021b) considered the situation of time-varying coefficients and proposed some pseudo-likelihood methods for estimation. In their methods, the former used B-spline function approximation and the latter employed local polynomials.

One drawback of these methods is that they only apply to the situation where either covariates or their effects are constant. To address this, in this chapter, we will discuss regression analysis of univariate panel count data that involves both time-dependent covariates and time-varying covariate effects and propose some spline-based approaches.

The remainder of this chapter is organized as follows. After introducing some notation and the assumptions that will be used throughout the chapter, an estimating equation procedure is proposed in Section 2.2 for estimation of covariate effects. In the method, the conditional mean model is employed for the underlying recurrent event process and B-spline functions are used to approximate time-varying covariate effects. The asymptotic properties of the proposed estimators, including the consistency and asymptotic distribution, are established in Section 2.3. Section 2.4 presents some results obtained from an extensive simulation study conducted to assess the finite sample performance of the proposed method and they suggest that it works well for practical situations. In Section 2.5, we apply the proposed approach to the data arising from the aforementioned China Health and Nutrition Survey(CHNS), and Section 2.6 gives some discussions and concluding remarks.

2.2 Estimation of Time-varying Covariate Effects

Consider a recurrent event study consisting of n independent subjects and let $N_i(t)$ denote the underlying recurrent event process representing the total number of the occurrences of the recurrent event of interest up to time t for subject i. Assume that

$N_i(t)$ is potentially observed only at $0 < t_{i,1} < \cdots < t_{i,m_i} < \tau$ and define $H_i^*(t) = \sum_{j=1}^{m_i} I(t_{i,j} \leq t)$, the underlying observation process. In practice, for each subject, there usually exists a following or stopping time C_i. Define $H_i(t) = H_i^*\{\min(C_i, t)\}$, the real observation process on the ith subject. That is, we only have panel count data and $N_i(t)$ is observed only at the time points where $H_i(t)$ jumps, $i = 1, ..., n$.

For each subject, suppose that there exist two vectors of covariates denoted by $\mathbf{W}_i = (W_{i1}, ..., W_{ip_1})^T$ and $\mathbf{Z}_i = (Z_{i1}, ..., Z_{ip_2})^T$, which may be time-dependent. The former represents the covariates that only have constant effects, while the latter denotes the covariates that may have time-varying effects. To describe the covariate effects, we assume that given $\mathbf{W}_i$ and $\mathbf{Z}_i$, the recurrent event process $N_i(t)$ follows the conditional multiplicative mean model

$$E\left\{N_i(t)|\mathbf{W}_i(t), \mathbf{Z}_i(t)\right\} = \Lambda(t) \exp\left\{\boldsymbol{\gamma}^T \mathbf{W}_i(t) + \boldsymbol{\beta}^T(t) \mathbf{Z}_i(t)\right\} . \tag{2.1}$$

In the above, $\Lambda(t)$ denotes an unknown baseline mean function, and $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_{p_1})^T$ and $\boldsymbol{\beta}(t) = (\beta_1(t), ..., \beta_{p_2}(t))^T$ represent constant and time-dependent coefficients, respectively. In the following, we will assume that given $\mathbf{W}_i$ and $\mathbf{Z}_i$, $N_i(t)$ and $H_i(t)$ are independent and some comments on this will be given below.

For estimation of $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}(t)$, let $\mathcal{B}$ and $\mathcal{M}_j$ denote the parameter spaces for $\boldsymbol{\gamma}$ and β_j, respectively, $j = 1, \ldots, p_2$, and assume that $\mathcal{B}$ is a compact subset of $\mathbb{R}^{p_1}$ and $\mathcal{M}_{0j} \subseteq \mathcal{L}_2([0, \tau])$. Define $\mathcal{M} = \prod_{j=1}^{p_2} \mathcal{M}_{0j}$ and $\Theta = \mathcal{B} \times \mathcal{M}$. Note that due to the dimension of $\mathcal{M}$, the estimation may not be easy and to deal with this, by following others, we propose to employ the sieve approach to first approximate $\boldsymbol{\beta}(t)$ by B-spline functions. Specifically, let $\mathcal{T} = \{t_j\}_{j=1}^{m_n+2l}$ with $0 = t_1 = \cdots = t_l < t_{l+1} < \cdots < t_{m_n+l} < t_{m_n+l+1} = \cdots = t_{m_n+2l} = \tau$ being a sequence of knots that partition

$[0, \tau]$ into $K_n + 1$ subintervals $[t_{l+j}, t_{l+j+1}]$ for $j = 0, \ldots, K_n$ with $K_n = O(n^{\nu})$ and $\max_{0<j<m_n} |t_{j+1} - t_j| = O(n^{-\nu})$ for $\nu \in (0, 0.5)$. Define

$$\mathcal{M}_{nj} = \left\{ \beta_{nj}(t) = \alpha_{j0} + \sum_{k=1}^{q_n} \alpha_{jk} B_k(t) = \mathbf{B}_n^T(t)\, \boldsymbol{\alpha}_j, \|\boldsymbol{\alpha}_j\|_1 < M_n \right\},$$

the class of B-splines of order l with the knots sequence $\mathcal{T}$. In the above, M_n is some large number with $M_n \to \infty$ as $n \to \infty$, $q_n = K_n + l$, $\mathbf{B}_n(t) = \{1, B_1(t), \ldots, B_k(t)\}^T$ is a class of B-spline basis, and $\boldsymbol{\alpha}_{nj} = (\alpha_{nj0}, \alpha_{nj1}, \ldots \alpha_{njq_n})$. Define $\mathcal{M}_n = \prod_{j=1}^{p_2} \mathcal{M}_{nj}$. Then $\Theta_n = \mathcal{B} \times \mathcal{M}_n$ is a sieve space for the original parameter space Θ.

Under the sieve space Θ_n, by replacing $\boldsymbol{\beta}(t)$ by $\boldsymbol{\beta}_n(t)$, model (2.1) can be rewritten as

$$\begin{aligned} E\{N_i(t)|\mathbf{W}_i(t), \mathbf{Z}_i(t)\} &= \Lambda_0(t)\, \exp\left\{\boldsymbol{\gamma}^T \mathbf{W}_i(t) + \sum_{j=1}^{p_2} \left(\mathbf{B}_n^T(t)\, \boldsymbol{\alpha}_{nj}\right) Z_{ij}(t)\right\} \\ &= \Lambda_0(t)\, \exp\left\{\boldsymbol{\gamma}^T \mathbf{W}_i(t) + \boldsymbol{\alpha}_n^T \tilde{\mathbf{Z}}_i(t)\right\} = \Lambda_0(t)\, \exp\left\{\boldsymbol{\theta}^T \mathbf{X}_i(t)\right\}. \end{aligned} \tag{2.2}$$

Here,

$$\tilde{\mathbf{Z}}_i(t) = \left(Z_{i1}(t)\mathbf{B}_n^T(t), Z_{i2}(t)\mathbf{B}_n^T(t), ..., Z_{ip_2}(t)\mathbf{B}_n^T(t)\right)^T,$$

$\boldsymbol{\alpha}_n = (\boldsymbol{\alpha}_{n1}^T, \boldsymbol{\alpha}_{n2}^T, ..., \boldsymbol{\alpha}_{np_2}^T)^T$, $\mathbf{X}_i(t) = \left(\mathbf{W}_i^T(t), \tilde{\mathbf{Z}}_i^T(t)\right)^T$, and $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \boldsymbol{\alpha}_n)^T$. Note that model (2.2) involves only time-independent covariate effects. Thus for estimation of $\boldsymbol{\theta}$, motivated by Hu et al. (2003), we propose to employ the estimating equation

$$\frac{1}{n} \sum_{i=1}^{n} \int_0^{\tau} Y_i(t) N_i(t) \left\{\mathbf{X}_i(t) - \bar{\mathbf{X}}(t; \boldsymbol{\theta})\right\} dH_i(t) = 0\,. \tag{2.3}$$

In the above, $Y_i(t) = I(C_i \geq t)$ is the at-risk indicator and $\bar{\mathbf{X}}(t; \boldsymbol{\theta}) = \mathbf{S}_1(t; \boldsymbol{\theta})/S_0(t; \boldsymbol{\theta})$,

where

$$S_u(t;\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\mathbf{X}_i^{\otimes u}(t)\exp\left(\boldsymbol{\theta}^T X_i(t)\right) dH_i(t) \ ,$$

$u = 0, 1, 2$ for $0 \le t \le \tau$, with $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$ and $a^{\otimes 2} = aa^T$ for some vector a.

Let $\hat{\boldsymbol{\theta}}_n$ denote the estimator of $\boldsymbol{\theta}$ given by the solution to the equation (2.3). Then one can estimate $\beta_j(t)$ by $\hat{\beta}_j(t) = \mathbf{B}_n^T(t)\,\hat{\boldsymbol{\alpha}}_{nj}$. In practice, sometimes one may also be interested in estimating the baseline mean function $\Lambda(t)$ and for this, it is apparent that one natural estimator is given by the Breslow-type estimator

$$\hat{\Lambda}(t,\hat{\boldsymbol{\theta}}_n) = \sum_{i=1}^{n} \frac{Y_i(t)N_i(t)dH_i(t)}{nS_0(t;\hat{\boldsymbol{\theta}}_n)} \ .$$

2.3 Asymptotic Properties

Now we will establish the asymptotic properties of the estimators proposed in the previous section, including the consistency, convergence rate and asymptotic normality. For this, let $\boldsymbol{\vartheta} = (\boldsymbol{\gamma}, \boldsymbol{\beta}, \Lambda)$ and $\boldsymbol{\vartheta}_0 = (\boldsymbol{\gamma}_0, \boldsymbol{\beta}_0, \Lambda_0)$ denote the true value of $\boldsymbol{\vartheta}$. Based on $\hat{\boldsymbol{\theta}}_n$, the estimator for $\boldsymbol{\vartheta}$ is $\hat{\boldsymbol{\vartheta}}_n = \left(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\beta}}_n, \hat{\boldsymbol{\Lambda}}\right)$. Also for convenience, let $\mathbf{V}(t) = \left(\mathbf{W}^T(t), \mathbf{Z}^T(t)\right)^T$ and redefine the parameter space $\Theta = \mathcal{A}\times\mathcal{M}\times\mathcal{F}$, where $\mathcal{F}$ denotes the parameter space of Λ. Let $\mathfrak{B}^d$ denote the collection of Borel sets in $\mathbb{R}^d$ and $\mathfrak{L}_2[0,\tau]$ the collection of Borel sets in $\mathcal{L}_2$ on $[0,\tau]$. Define $\mathfrak{B}^1[0,\tau] = \{B\cap[0,\tau] : B \in \mathfrak{B}^1\}$, $\mathfrak{B}^d = \mathfrak{B}^1[0,\tau]\times\ldots\times\mathfrak{B}^1[0,\tau]$ and $\mathfrak{L}_2^d[0,\tau] = \mathfrak{L}^1[0,\tau]\times\ldots\times\mathfrak{L}^1[0,\tau]$. Also define the measure

$$\upsilon_1(B_1\times B_2\times B_3) = \int_{B_3\times B_2}\int_{B_1} dE\left[Y(t)\,H(t)\right] d\mu_Z d\mu_W$$

for $B_1 \in \mathfrak{B}^1[0,\tau]$, $B_2 \in \mathfrak{L}_2^{p_2}[0,\tau]$ and $B_3 \in \mathfrak{L}_2^{p_1}[0,\tau]$, where μ_W and μ_Z are the measures for $\mathbf{W}$ and $\mathbf{Z}$; and $\mu_1(B_1 \times B_2) = \upsilon_1(B_1 \times B_2 \times \mathfrak{L}_2^{p_2}[0,\tau])$. Alternatively, let $\mu_V = \mu_Z \times \mu_W$, we can rewrite $\upsilon_1(B_1 \times B_2 \times B_3)$ as

$$\upsilon_1(B_1 \times B_4) = \int_{B_4 \times B_1} dE[Y(t)\,H(t)]\,d\mu_V$$

for $B_1 \in \mathfrak{B}^1[0,\tau]$ and $B_4 \in \mathfrak{L}_2^p[0,\tau]$; and $\mu_1(B) = \upsilon_1(B \times \mathfrak{L}_2^p[0,\tau])$. Define the L_2 metric $d(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2)$ on Θ as

$$d(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2) = \left(\|\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2\|_2^2 + \int \|\boldsymbol{\beta}_1(u) - \boldsymbol{\beta}_2(u)\|_2^2\, d\mu_1(u) + \|\Lambda_1 - \Lambda_2\|_{L_2(\mu_1)}^2 \right)^{1/2}.$$

To establish the asymptotic results, we need the following regularity conditions.

(C1) The observation process has the rate function $E[dH^*(t)\,|\mathbf{W}(t), \mathbf{Z}(t), C] = \omega(t)$ dt, where $\omega(t)$ is a bounded, nonnegative and continuous function on $[0,\tau]$. There exists a positive integer M_0 such that $\Pr(H(\tau) < M_0) = 1$. That is, the total observation number is finite. Moreover, the support of $\omega(t)$ is $[\tau_0, \tau]$ with $\tau_0 > 0$ and $\Lambda_0(\tau_0) > 0$ for some constant τ_0.

(C2) The measure $\mu_1 \times \mu_V$ is absolutely continuous with respect to υ_1 and $\mu_1(\{\tau\}) > 0$.

(C3) The parameters space of Λ, $\mathcal{F}$, consists of bounded non-decreasing functions in $\mathcal{L}_2$ over $[0,\tau]$.

(C4) The parameters space of $\boldsymbol{\beta}$, $\mathcal{M}$, is bounded and convex in $\mathcal{L}_2([0,\tau])$. Each component of the true value of $\boldsymbol{\beta}(t)$, denoted by $\beta_{0j}(t)$, $j = 1, \ldots, p_2$, is continuously rth differentiable in $[0,\tau]$.

(C5) The parameter space of $\boldsymbol{\gamma}$, $\mathcal{A}$, is bounded and convex in $\mathbb{R}^d$.

(C6) The covariate vector $\mathbf{V}(t) = \left(\mathbf{W}^T(t), \mathbf{Z}^T(t)\right)^T$ is uniformly bounded over $[0, \tau]$ with the distribution μ_V.

(C7) Given $\mathbf{V}(t)$, $t \in [0, \tau]$, C and N are independent. Besides, with probability 1,

$$\inf_{\mathbf{V}(t), t\in[0,\tau]} \Pr\left(C \geq \tau | \mathbf{V}(t) = \mathbf{v}(t), t \in [0, \tau]\right)$$
$$= \inf_{\mathbf{V}(t), t\in[0,\tau]} \Pr\left(C = \tau | \mathbf{V}(t) = \mathbf{v}(t), t \in [0, \tau]\right) > 0.$$

(C8) If $\boldsymbol{\gamma}^T\mathbf{W}(t) + \boldsymbol{\beta}^T(t)\mathbf{Z}(t) \equiv 0$, $t \in [0, \tau]$ with probability 1 for some $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$, then $\boldsymbol{\gamma} = 0$ and $\boldsymbol{\beta}(t) = 0$ for $t \in [0, \tau]$.

(C9) The function $M_0(\mathbf{V}) = \int N(t) \log(N(t))\, dH(t)$ satisfies $\mathbf{P}M_0(\mathbf{V}) < \infty$.

(C10) $E\left[\exp\left(C_0 N(t)\right)\right]$ is bounded in $[0, \tau]$ for some constant C_0.

(C11) The true baseline mean function Λ_0 is differentiable in $[\tau_0, \tau]$. Moreover, its first order derivative has a positive and finite lower and upper bound in $[\tau_0, \tau]$.

(C12) There exist $\eta_1 \in (0, 1)$ such that

$$a^T \mathrm{Var}\left(\mathbf{V}(U)|U\right) a \geq \eta_1 a^T E\left(\mathbf{V}^T(U)\mathbf{V}(U)|U\right) a,$$

a.s. for all $a \in \mathbb{R}^{p_1+p_2}$, where $(U, \mathbf{V})$ has distribution $\nu_1/\nu_1(\mathbb{R}^+ \times \mathcal{V})$.

Note that conditions (C1) and (C7) are common on the observation schemes and similar to the combination of C8, C10 and C11 in Lu et al. (2009). Condition (C2) comes from the condition in Theorem 1 of Wellner and Zhang (2007) and Theorem

1 of Lu et al. (2009), ensuring $\hat{\Lambda}$ is bounded, and conditions (C6)-(C11) are common assumptions in the semiparametric estimation. Also conditions (C2) and (C8) ensure the identifiability of the semiparametric model and conditions (C9), (C10) and (C11) are from conditions C4, C10 and C12 in Wellner and Zhang (2007). Condition (C12) is needed to prove the convergence rate and can be justified by the arguments similar to those in Wellner and Zhang (2007).

Theorem 1 (Consistency)**.** *Assume that the regularity conditions (C1)-(C9) given above hold. Then we have that* $d\left(\hat{\boldsymbol{\vartheta}}_n, \boldsymbol{\vartheta}_0\right) \to 0$ *in probability as* $n \to \infty$.

Theorem 2 (Rate of Convergence)**.** *Assume that the regularity conditions (C1)-(C12) given above hold. Then we have that*

$$n^{\min\left\{n^{\frac{1-\nu}{3}}, n^{r\nu}\right\}} d\left(\hat{\boldsymbol{\vartheta}}_n, \boldsymbol{\vartheta}_0\right) = O_p\left(1\right)$$

with the optimal rate $O_p(n^{-r/(3r+1)})$ *achieved at* $\nu = 1/\left(1+3r\right)$.

Note that the order of the optimal rate $n^{-r/(3r+1)}$ is slower than $n^{-r/(2r+1)}$ in Lu et al. (2009) because the nonparametric parameter Λ is estimated by a step function though $\boldsymbol{\beta}(t)$ is estimated by B-splines. Nevertheless, we can still derive the asymptotic distribution of $\hat{\boldsymbol{\gamma}}$ with rate of convergence $n^{-1/2}$. The next theorem establishes the asymptotic normality of $\hat{\boldsymbol{\theta}}$ in the form similar to He et al. (2017).

Theorem 3 (Asymptotic Normality)**.** *Assume that the regularity conditions (C1)-(C12) given above hold and also* $(4r)^{-1} < \nu < 2^{-1}$ *with* $r > 1$. *Define* $\mathcal{H}_1 = \{\boldsymbol{h}_1 : \boldsymbol{h}_1 \in \mathcal{A}, \|\boldsymbol{h}_1\| \leq 1\}$, $\mathcal{H}_2 = \{\boldsymbol{h}_2 : \boldsymbol{h}_2 \in \mathcal{M},$ *each component of* $\boldsymbol{h}_2$ *is of bounded total variation.*$\}$, *and* $\mathcal{H}_3 = \{h_3 : h_3$ *is a fucntion with bouned total variation in* $[0, \tau]$ *and*

$h_3(0) = 0\}$ *Then for some* $(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3) \in \mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$, *we have that*

$$\sqrt{n}\left(\boldsymbol{\gamma}-\gamma_0\right)^T \boldsymbol{h}_1+\sqrt{n} \int_0^\tau\left(\boldsymbol{\beta}(t)-\boldsymbol{\beta}_0(t)\right)^T d \boldsymbol{h}_2(t) \\ +\sqrt{n} \int_0^\tau\left(\Lambda(t)-\Lambda_0(t)\right) d h_3(t) \rightarrow_d N\left(0, \sigma^2\right),$$

where σ^2 *is given in the Appendix.*

The proof of the results above is sketched in the Appendix. Note that similar to He et al. (2017), we cannot find the explicit form of the asymptotic distribution because the explicit forms of $(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3)$ cannot be solved though exist. On the other hand, by following Amorim et al. (2008), we can have the following *ad hoc* estimators for the asymptotic covariance matrix of $\hat{\boldsymbol{\gamma}}$ and the pointwise asymptotic variance of $\boldsymbol{\beta}(t)$ in $t \in [0, \tau]$. First the asymptotic covariance matrix of $\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma_0}$ can be consistently estimated by $\hat{A}_{\hat{\boldsymbol{\gamma}}}(t)^{-1}\hat{B}_{\hat{\boldsymbol{\gamma}}}(t)\hat{A}_{\hat{\boldsymbol{\gamma}}}(t)^{-1}$. Here $\hat{A}_{\hat{\boldsymbol{\gamma}}}(t)$ and $\hat{B}_{\hat{\boldsymbol{\gamma}}}(t)$ are the top-left $p_1 \times p_1$ sub-matrices of

$$\begin{aligned} \hat{A}(t) &= \frac{\partial \sum_{i=1}^n \int_0^\tau Y_i(t) N_i(t)\left\{(\mathbf{X}(t))-\bar{\mathbf{X}}(t ; \boldsymbol{\theta})\right\} d H_i(t)}{\partial \boldsymbol{\theta}} \\ &= \sum_{i=1}^n \int_0^\tau Y_i(t) N_i(t)\left\{-\frac{\partial \bar{\mathbf{X}}(t ; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\} d H_i(t), \end{aligned}$$

and

$$\hat{B}(t)=\left[\sum_{i=1}^n \int_0^\tau Y_i(t)\left(N_i(t)-\mu_0(t) \exp \left(\boldsymbol{\theta}^T \mathbf{X}_i(t)\right)\right)\left\{(\mathbf{X}(t))-\bar{\mathbf{X}}(t ; \boldsymbol{\theta})\right\} d H_i(t)\right]^{\otimes 2},$$

respectively, where

$$\frac{\partial \bar{\mathbf{X}}(t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\partial \tilde{\mathbf{S}}_1(t;\boldsymbol{\theta})/\tilde{S}_0(t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \tilde{\mathbf{S}}_2(t;\boldsymbol{\theta})/\tilde{S}_0(t;\boldsymbol{\theta}) - (\tilde{\mathbf{S}}_1(t;\boldsymbol{\theta})/\tilde{S}_0(t;\boldsymbol{\theta}))^2 .$$

Furthermore, the asymptotic variance of $\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)$ for a given t can be estimated by

$$\hat{\Omega}(t) = \mathbf{B}^T(t) \left(\hat{A}_{\hat{\boldsymbol{\alpha}}}(t)^{-1} \hat{B}_{\hat{\boldsymbol{\alpha}}}(t) \hat{A}_{\hat{\boldsymbol{\alpha}}}(t)^{-1} \right) \mathbf{B}(t),$$

where $\hat{A}_{\hat{\boldsymbol{\alpha}}}(t)$ and $\hat{B}_{\hat{\boldsymbol{\alpha}}}(t)$ are the bottom-right $p_2 \times p_2$ sub-matrices of $\hat{A}(t)$ and $\hat{B}(t)$, respectively. The numerical study in the next Section shows that these variance estimators work well.

2.4 A Simulation Study

In this section, we present some results obtained from an extensive simulation study conducted to evaluate the finite sample performance of the estimation procedure proposed in the previous sections. In the study, we consider the situation of four covariates with two having constant effects and two having time-varying effects. That is, we have that $p_1 = p_2 = 2$. More specifically, we assume that $W_1(t)$ and $Z_1(t)$ are time-dependent variables and generate them independently by setting them equal to $B_1 I(t \leq V) + B_2 I(t > V)$. Here B_1, B_2 and V are generated independently from the uniform distributions over $(0, 0.5)$, $(0.5, 1)$ and $(0, \tau)$ with $\tau = 1$, respectively. Furthermore, $W_2(t)$ and $Z_2(t)$ are assumed to be time-independent and generated from the uniform distribution over $(0, 1)$ independently.

To generate the observed panel count data, we first generate the observation times

$t_{i,j}$'s from the non-homogeneous Poisson process with the mean function $3\,t+4$ and the follow-up times C_i's from the uniform distribution over $(0.9\tau,\,\tau)$. Then given the covariates and the real observation times, the observed data is generated by assuming that

$$N_i\left(t_{i,j}\right) = N_i^*\left(t_{i,1}\right) + N_i^*\left(t_{i,2}-t_{i,1}\right) + \cdots + N_i^*\left(t_{i,j}-t_{i,j-1}\right)\,,$$

where $N_i^*(t_{i,1})$ and $N_i^*(t_{i,j}-t_{i,j-1})$ follow the Poisson distributions with the means

$$v_i\,\Lambda_0\left(t_{i,1}\right) e^{\left(\boldsymbol{\gamma}^T\mathbf{W}_i(t_{i,1})+\boldsymbol{\beta}^T(t_{i,1})\mathbf{Z}_i(t_{i,1})\right)}\,,$$

and

$$v_i\,\Lambda_0\left(t_{i,j}\right) e^{\left(\boldsymbol{\gamma}^T\mathbf{W}_i(t_{i,j})+\boldsymbol{\beta}^T(t_{i,j})\mathbf{Z}_i(t_{i,j})\right)} - v_i\,\Lambda_0\left(t_{i,j-1}\right) e^{\left(\boldsymbol{\gamma}^T\mathbf{W}_i(t_{i,j-1})+\boldsymbol{\beta}^T(t_{i,j-1})\mathbf{Z}_i(t_{i,j-1})\right)}\,,$$

respectively. In the above, the v_i's are assumed to follow the gamma distribution with mean 1 and variance σ^2. That is, $N_i(t)$'s are mixed Poisson processes. The results given below are based on $n=300$ with 1000 replications.

Table 2.1 presents the results on estimation of two time-independent coefficient effects γ_1 and γ_2 obtained with $\Lambda_0(t)=2t+3$, the use of cubic B-splines with 3 interior knots, and $\sigma^2=0$ or 1. Here for the time-dependent coefficient effects, we considere two settings with setting 1 being $\beta_1(t)=t$ and $\beta_2(t)=t^2$ and setting 2 being $\beta_1(t)=(\sin(4\pi t)+4\pi t)/12$ and $\beta_2(t)=(\cos(4\pi t)+4\pi t)/12$. The results include the empirical bias (BIAS) given by the average of the estimates minus the true value, the sampling standard deviation (ESD), the average of the estimated standard errors (SE) and the 95% empirical coverage probability (CP). One can see that they suggest that the proposed estimators seem to be unbiased and the variance estimates also appear to be appropriate. In addition, the results on CP indicate that the normal approximation to the distribution of the proposed estimator $\hat{\boldsymbol{\gamma}}$ seems to be reasonable too. Figure 2.1 gives the averages of the estimated $\beta_1(t)$ and $\beta_2(t)$ over 1000 equal-spaced grid points on the time axis for the situation of $\sigma^2 = 1$. For comparison, the true curves are presented too and the results indicate that the proposed procedure seems to yield unbiased estimates again. Furthermore, Figure 2.2 shows the average of the estimated point-wise standard errors along with the point-wise sample standard errors and indicates that the proposed method appears to give reasonable variance estimates.

To assess the possible effects of the baseline mean function on the proposed estimation procedure, we repeat the study with $\sigma^2 = 0$ except assuming setting 3 $\Lambda_0(t)$ =$(\sin(4\pi t) + 4\pi t)/2$ or setting 4 $\Lambda_0(t)$ =$(\cos(4\pi t) + 4\pi t)/2$, and the results obtained on estimation of the two time-independent covariate effects γ_1 and γ_2 are given in Table 2.2. Figures 2.3 and 2.4, similar to Figures 2.1 and 2.2, display the averages of the estimated $\beta_1(t)$ and $\beta_2(t)$ and the estimated point-wise standard errors, respectively, along with the true curves and the point-wise sample standard errors. Again they suggest that the proposed estimation procedure seems to work well. We also consider other set-ups, including different degrees of B-spline functions, different numbers of interior knots and other types of covariates, and obtain similar results.

2.5 Analysis of China Health and Nutrition Survey

In this section, we apply the methodology proposed in the previous sections to the China Health and Nutrition Survey (CHNS), an international collaborative project between the Carolina Population Center at the University of North Carolina at Chapel Hill and the National Institute for Nutrition and Health at the Chinese Center for Disease Control and Prevention (CCDC). Initiated in 1985, the survey was conducted every 2 to 4 years and designed to examine the effects of the health, nutrition, and family planning policies and programs implemented by national and local governments and to see how the social and economic transformation of Chinese society was affecting the health and nutritional status of its population. The survey took place over a 7-day period using a multistage, random cluster process to draw a sample of over 11000 households with over 42,000 individuals participated in 15 provinces and municipal cities that vary substantially in geography, economic development, public resources, and health indicators. Villages and townships within the counties and urban/suburban neighborhoods within the cities were selected randomly.

One objective of the survey is to assess the relationship between the number of children and parents' long-term income and wealth (Tian (2018); Oliveira (2016)). For this, the information on the pregnancy of the female participants, that is, the number of pregnancy, was collected. It is easy to see that due to the periodic follow-up nature of the survey, only panel count data are available on the pregnancy process. In addition, the information is available on four factors or covariates: whether the mother came from urban or rural areas (urban = 0, rural = 1), the average monthly wage last year, the completed years of formal education in regular school (0: No school

- 36: 6 yrs. college or more), and the current health status (1 - Excellent, 2 - Good, 3 - Fair, 4 - Poor). For the analysis below, we will focus on the 2537 female participants with complete information on the four covariates described above, after removing some subjects with apparent record errors. Among them, the average pregnancy count is 1.416.

To apply the proposed estimation procedure, we first assume that all of the four covariates have time-varying effects and Figure 2.5 presents the estimated covariate effects with the use of 3 interior knots. The results suggest that the mothers' location seems to have a significantly positive relationship with the fertility and the effect appears to increase along with the mothers' age. In other words, the mothers from rural areas are more likely to have more children compared to those from urban areas. On the average monthly wage, the effects seem to change directions along with or depend on the age of a mother. Specifically, it appears to have positive effects on the fertility for young mothers but has no significant effects on middle age mothers and then seems to have significantly negative effects on the fertility of old mothers. In contrast, the mother's education level and health status seem to have no significant or only constant effects on the fertility. Note that one may not want to pay much attention to the estimated effects at the end of study due to the sparsity of the observed data.

Now we assume that only the location and average monthly wage have time-varying effects, while the education level and health status have constant effects on the fertility, based on previous results. Table 2.3 gives the estimated covariate effects for the education level and health status with the use of 3, 5 or 7 interior knots. The estimated time-dependent effects of the location and the average monthly wage based on 3 interior knots are presented in Figure 2.6 and the results with the use of 5 or 7 interior knots are similar. One can see from Table 2.3 that the mothers' education levels are significantly negatively correlated with the fertility rate and the mothers with lower education levels tend to have more children. In contrast, the health status level has positive effects on the fertility rate, and these results are consistent with respect to the number of interior knots. The results given in Figure 2.6 are apparently similar to those given in Figure 2.5 and again indicate that the mothers at rural areas have a higher fertility rate than those at urban areas and the effect of the average monthly wage the fertility depend on the age of a mother.

2.6 Discussions and Concluding Remarks

In this chapter, we discussed regression analysis of panel count data with both time-dependent covariates and time-varying covariate effects. For the problem, a spline-based estimating equation procedure was developed and the asymptotic properties of the proposed estimators were established. In the method, B-splines functions were used to approximate the time-varying covariate effects and the method can be easily implemented. An extensive numerical study was conducted and suggested that the proposed method works well in practical situations. Also the usefulness of the proposed estimation procedure was illustrated by applying it to the China Health and Nutrition Survey through identifying some time-varying covariate effects.

Table 2.1: Simulation results on estimation of γ_1 and γ_2 with $\Lambda_0(t) = 2t + 3$.

Setting	Parameters	True value	BIAS	ESD	SE	CP
$\sigma^2 = 0$						
Setting 1	γ_1	0.5	-0.0001	0.0506	0.0481	0.945
	γ_2	0.5	0.0013	0.0529	0.0481	0.917
Setting 2	γ_1	0.5	0.0011	0.0491	0.0450	0.921
	γ_2	0.5	0.0005	0.0489	0.0450	0.926
$\sigma^2 = 1$						
Setting 1	γ_1	0.5	0.0094	0.2279	0.2086	0.931
	γ_2	0.5	0.0096	0.2366	0.2092	0.917
Setting 2	γ_1	0.5	-0.0008	0.2188	0.2091	0.940
	γ_2	0.5	0.0032	0.2261	0.2090	0.914

Table 2.2: Simulation results on estimation of γ_1 and γ_2 with $\sigma^2 = 0$.

Parameters	True value	BIAS	ESD	SE	CP
$\Lambda_0(t) = (\sin(4\pi t) + 4\pi t)/2$					
γ_1	0.5	-0.0052	0.2514	0.2223	0.917
γ_2	0.5	-0.0025	0.2537	0.2210	0.911
$\Lambda_0(t) = (\cos(4\pi t) + 4\pi t)/2$					
γ_1	0.5	0.0094	0.2407	0.2252	0.929
γ_2	0.5	0.0116	0.2468	0.2238	0.925

# of interior knots	Parameter	Estimated effect	SD	95% CI
3	γ_1	-0.0211	0.0018	(-0.0246,-0.0177)
	γ_2	0.0322	0.0117	(0.0093. 0.0551)
5	γ_1	-0.0208	0.0018	(-0.0243,-0.0173)
	γ_2	0.0330	0.0117	(0.0101, 0.0558)
7	γ_1	-0.0209	0.0018	(-0.0244,-0.0174)
	γ_2	0.0330	0.0117	(0.0102, 0.0559)

Table 2.3: Estimated constant effects of the education level and health status for CHNS.

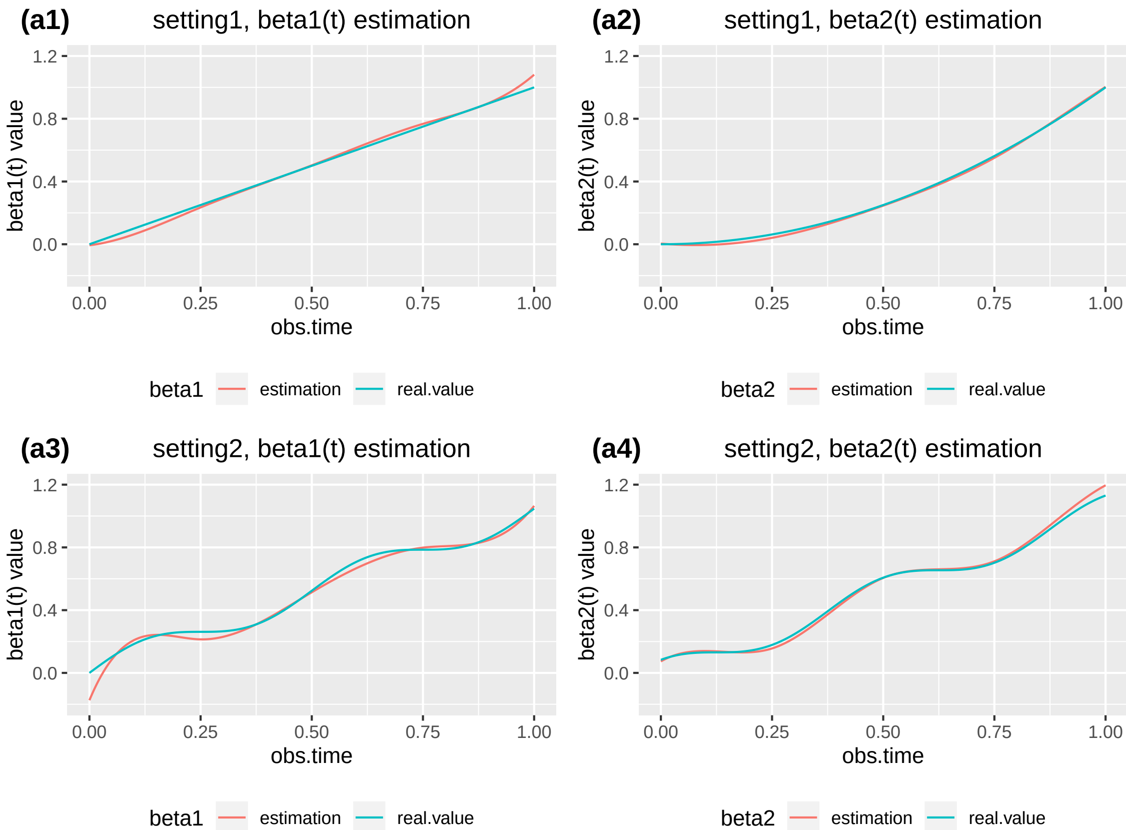

Figure 2.1: Simulation results on estimation of $\hat{\beta}_1(t)$ and $\hat{\beta}_2(t)$ with $\Lambda_0(t) = 2t+3$ and $\sigma^2 = 1$, where, setting 1: $\beta_1(t) = t$, $\beta_1(t) = t^2$; setting 2: $\beta_1(t) = (\sin(4\pi t) + 4\pi t)/12$, $\beta_2(t) = (\cos(4\pi t) + 4\pi t)/12$.

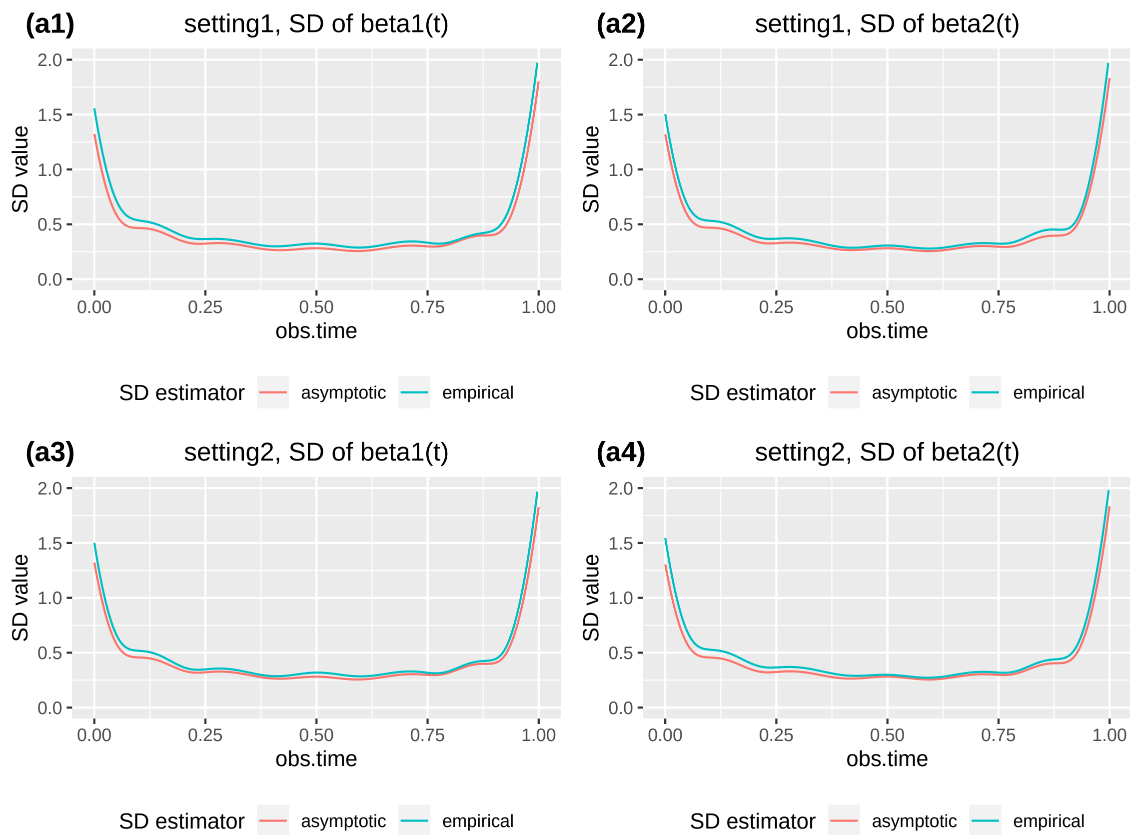

Figure 2.2: Simulation results on standard error estimation with $\Lambda_0(t) = 2t + 3$ and $\sigma^2 = 1$, where, setting 1: $\beta_1(t) = t$, $\beta_1(t) = t^2$; setting 2: $\beta_1(t) = (\sin(4\pi t) + 4\pi t)/12$, $\beta_2(t) = (\cos(4\pi t) + 4\pi t)/12$.

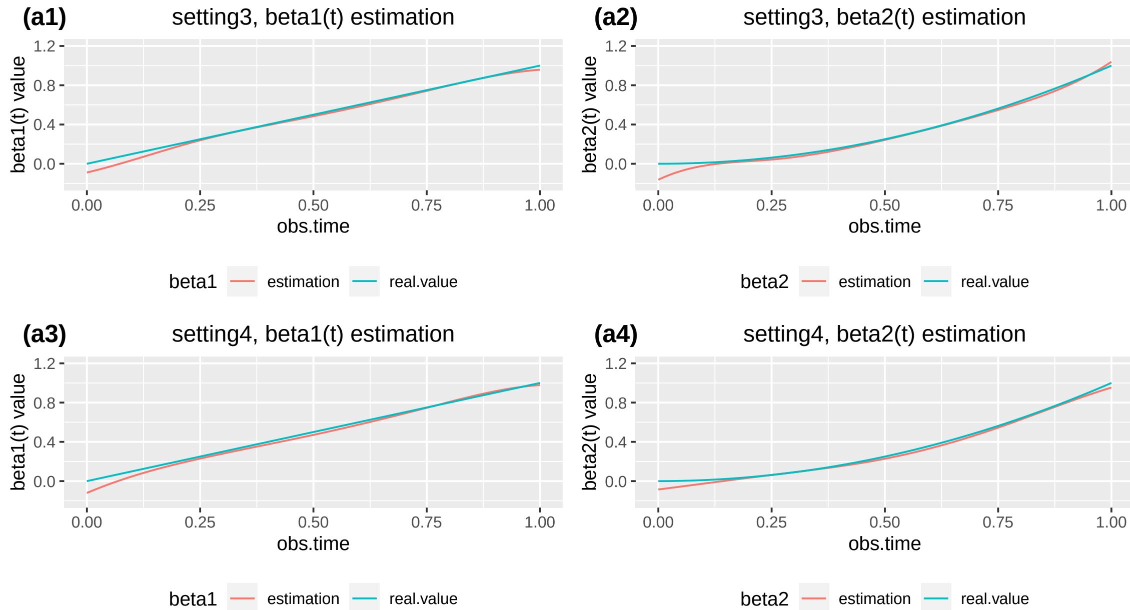

Figure 2.3: Simulation results on estimation of $\hat{\beta}_1(t)$ and $\hat{\beta}_2(t)$ with $\sigma^2 = 0$, where, setting 3: $\Lambda_0(t) = (\sin(4\pi t) + 4\pi t)/2$; setting 4: $\Lambda_0(t) = (\cos(4\pi t) + 4\pi t)/2$.

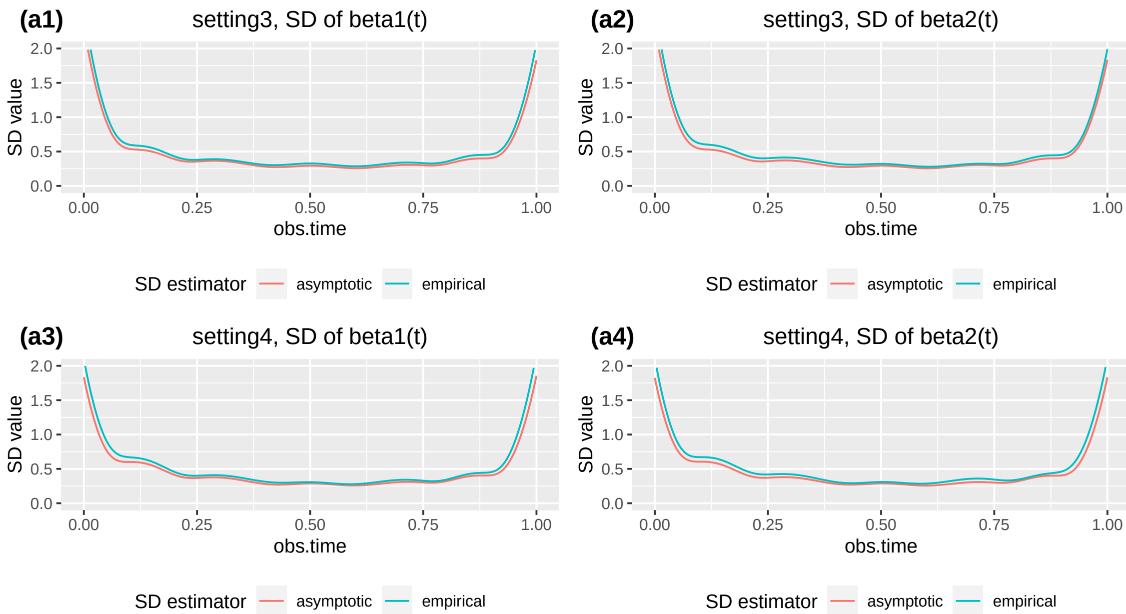

Figure 2.4: Simulation results on standard error estimation with $\sigma^2 = 0$, where, setting 3: $\Lambda_0(t) = (\sin(4\pi t) + 4\pi t)/2$; setting 4: $\Lambda_0(t) = (\cos(4\pi t) + 4\pi t)/2$.

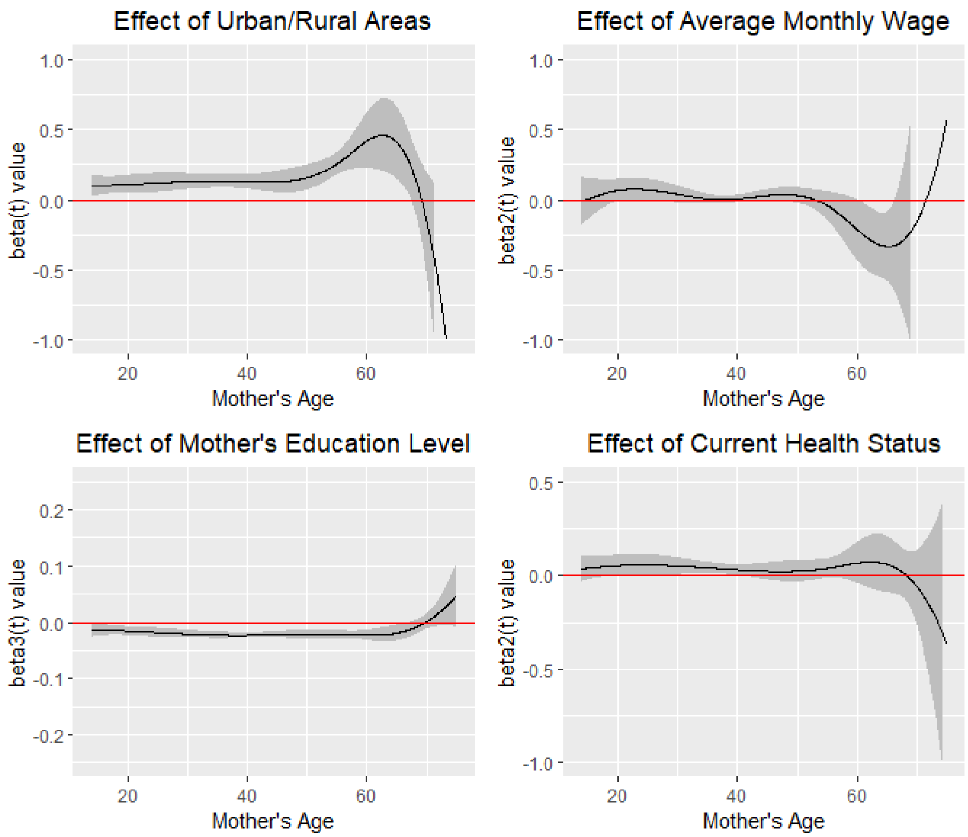

Figure 2.5: Estimated time-varying effects of all four covariates (solid curves) and corresponding pointwise 95% confidence intervals (grey curves).

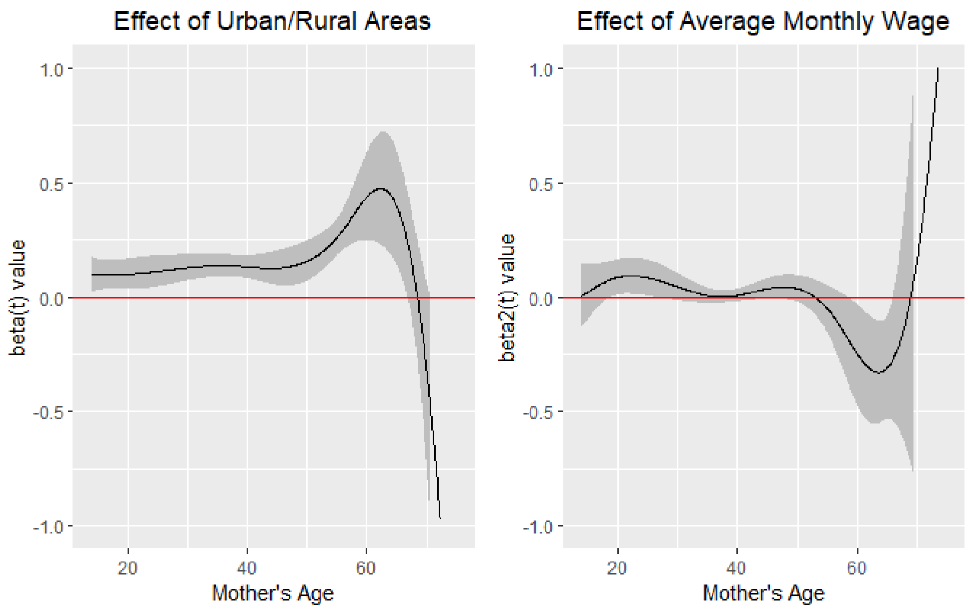

Figure 2.6: Estimated time-varying effects of the location and average monthly wage (solid curves) and corresponding pointwise 95% confidence intervals (grey curves).

Chapter 3

Dynamic Analysis of Multivariate Panel Count Data with Mean Model

3.1 Introduction

Multivariate panel count data arise in studies involving several types of recurrent events in which patients are examined only at periodic follow-up assessments. This chapter will discuss analysis of multivariate panel count data. Futhermore, as mentioned in chapter 2, the time-varying coefficient and covariate effects situation would also be considered.

The remainder of this chapter is organized as follows. After introducing some notation and the assumptions that will be used throughout this chapter, an estimating equation procedure is proposed in Section 3.3 for estimation when coefficients and covariates effects are time-dependent simultanously for multivariate panel count data. In particular, marginal mean models are employed for the underlying counting processes that characterize multivariate panel count data. B-spline functions are used to approximate time-varying covariate effects. The corresponding asymptotic results are established in Section 3.4. Section 3.5 presents some results obtained from an extensive simulation study conducted to assess the finite sample performance of the proposed method and they suggest that it works well for practical situations. In Section 3.6, we apply the proposed approach to the data arising from the aforementioned China Health and Nutrition Survey(CHNS), and Section 3.7 gives some discussions and concluding remarks.

3.2 Model and Notation

Consider a recurrent event study that involves n independent subjects and suppose that each subject may experience K different types of events. For subject i , let $N_{ik}(t)$ denote the total number of type k events that have occurred up to time $t, 0 \leq t \leq \tau$, where τ denotes a known constant representing the study length, $i = 1, ..., n, k = 1, ..., K$. Assume that $N_{ik}(t)$ is potentially observed only at $0 < t_{i,1} < \cdots < t_{i,m_i} < \tau$ and define $H^*_{ik}(t) = \sum_{j=1}^{m_i} I(t_{i,j,k} \leq t)$, the underlying observation process. Also for each i, suppose that there exists a positive random variable C_i representing the censoring or follow-up time on subject i. Define $H_{ik}(t) = H^*_{ik}\{\min(C_i, t)\}$, the real observation process on the ith subject. That is, we only have multivariate panel count data and $N_{ik}(t)$ is observed only at the time points where $H_{ik}(t)$ jumps, $i = 1, ..., n, k = 1, ..., K$.

For each subject, suppose that there may exist vectors of covariates denoted by $\mathbf{W}_i = (W_{i1}, ..., W_{ip_1})^T$ and $\mathbf{Z}_i = (Z_{i1}, ..., Z_{ip_2})^T$ which may be time-dependent. The former represents the covariates that only have constant effects, while the latter denotes the covariates that may have time-varying effects. They may affect the rate of occurrence of type k events. Here, for the simplicity of presentation, we assume that the follow-up time or observation period and the covariates that may affect $N_{ik}(t)$ are the same for different types of recurrent events. The inference approach presented below can be easily generalized to situations where C_i, $\mathbf{W}_i$ and $\mathbf{Z}_i$ may differ for different types of recurrent events. Define $Y_i(t) = I(t \leq C_i)$, indicating if subject i is at risk of experiencing the recurrent events at time $t, i = 1, ..., n$.

3.3 Estimation

For the effects of covariates on $N_{ik}(t)$, we assume that given $\mathbf{W}_i(t)$ and $\mathbf{Z}_i(t)$, the marginal mean function of $N_{ik}(t)$ has the form

$$E\left\{N_{ik}(t)|\mathbf{W}_i(t), \mathbf{Z}_i(t)\right\} = \Lambda_k(t)\exp\left(\boldsymbol{\gamma}^T\mathbf{W}_i(t) + \boldsymbol{\beta}^{\boldsymbol{T}}(t)\mathbf{Z}_i(t)\right) \tag{3.1}$$

In the model above, $\Lambda_k(t)$ is an unknown continuous baseline mean function, $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_{p_1})^T$ and $\boldsymbol{\beta}(t) = (\beta_1(t), ..., \beta_{p_2}(t))^T$ represent constant and time-dependent coefficients, respectively. In the following, we will assume that given $\mathbf{W}_i$ and $\mathbf{Z}_i$, $N_{ik}(t)$ and $H_{ik}(t)$ are independent and some comments on this will be given below. By following He et al. (2007), model (3.1) assumes that the baseline mean functions can be different for different types of recurrent events, but the effects of covariates on different types of recurrent events are common. The goal here is to estimate regression parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\beta(t)}$.

For estimation of $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}(t)$, let $\mathcal{B}$ and $\mathcal{M}_j$ denote the parameter spaces for $\boldsymbol{\gamma}$

and β_j, respectively, $j = 1, \ldots, p_2$, and assume that $\mathcal{B}$ is a compact subset of $\mathbb{R}^{p_1}$ and $\mathcal{M}_{0j} \subseteq \mathcal{L}_2([0, \tau])$. Define $\mathcal{M} = \prod_{j=1}^{p_2} \mathcal{M}_{0j}$ and $\Theta = \mathcal{B} \times \mathcal{M}$. Note that due to the dimension of $\mathcal{M}$, the estimation may not be easy and to deal with this, by following others, we propose to employ the sieve approach to first approximate $\boldsymbol{\beta}(t)$ by B-spline functions. Specifically, let $\mathcal{T} = \{t_j\}_{j=1}^{m_n+2l}$ with $0 = t_1 = \cdots = t_l < t_{l+1} < \cdots < t_{m_n+l} < t_{m_n+l+1} = \cdots = t_{m_n+2l} = \tau$ being a sequence of knots that partition $[0, \tau]$ into $K_n + 1$ subintervals $[t_{l+j}, t_{l+j+1}]$ for $j = 0, \ldots, K_n$ with $K_n = O(n^{\nu})$ and $\max_{0<j<m_n} |t_{j+1} - t_j| = O(n^{-\nu})$ for $\nu \in (0, 0.5)$. Define

$$\mathcal{M}_{nj} = \left\{ \beta_{nj}(t) = \alpha_{j0} + \sum_{k=1}^{q_n} \alpha_{jk} B_k(t) = \mathbf{B}_n^T(t)\, \boldsymbol{\alpha}_j, \|\boldsymbol{\alpha}_j\|_1 < M_n \right\},$$

the class of B-splines of order l with the knots sequence $\mathcal{T}$. In the above, M_n is some large number with $M_n \to \infty$ as $n \to \infty$, $q_n = K_n + l$, $\mathbf{B}_n(t) = \{1, B_1(t), \ldots, B_k(t)\}^T$ is a class of B-spline basis, and $\boldsymbol{\alpha}_{nj} = (\alpha_{nj0}, \alpha_{nj1}, \ldots \alpha_{njq_n})$. Then $\beta_j(t)$ can be approximated by $\beta_{nj}(t)$. Define $\mathcal{M}_n = \prod_{j=1}^{p_2} \mathcal{M}_{nj}$. Then $\Theta_n = \mathcal{B} \times \mathcal{M}_n$ is a sieve space for the original parameter space Θ.

Under the sieve space Θ_n, by replacing $\boldsymbol{\beta}(t)$ by $\boldsymbol{\beta}_n(t)$, model (3.1) can be rewritten as

$$\begin{aligned} E\{N_{ik}(t)|\mathbf{W}_i(t), \mathbf{Z}_i(t)\} &= \Lambda_k(t) \exp\left\{ \boldsymbol{\gamma}^T \mathbf{W}_i(t) + \sum_{j=1}^{p_2} \left(\mathbf{B}_n^T(t)\, \boldsymbol{\alpha}_{nj}\right) Z_{ij}(t) \right\} \\ &= \Lambda_k(t) \exp\left\{ \boldsymbol{\gamma}^T \mathbf{W}_i(t) + \boldsymbol{\alpha}_n^T \tilde{\mathbf{Z}}_i(t) \right\} = \Lambda_k(t) \exp\left\{ \boldsymbol{\theta}^T \mathbf{X}_i(t) \right\}. \end{aligned} \quad (3.2)$$

Here,

$$\tilde{\mathbf{Z}}_i(t) = \left(Z_{i1}(t)\mathbf{B}_n^T(t), Z_{i2}(t)\mathbf{B}_n^T(t), ..., Z_{ip_2}(t)\mathbf{B}_n^T(t) \right)^T,$$

$\boldsymbol{\alpha}_n = \left(\boldsymbol{\alpha}_{n1}^T, \boldsymbol{\alpha}_{n2}^T, ..., \boldsymbol{\alpha}_{np_2}^T\right)^T$, $\mathbf{X}_i(t) = \left(\mathbf{W}_i^T(t), \tilde{\mathbf{Z}}_i^T(t)\right)^T$ and $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \alpha)^T$. Note that this model does not include time-dependent coefficients now.

Thus for estimation of $\boldsymbol{\theta}$, motivated by Hu et al. (2003) and He et al. (2007), we propose to employ the estimating equation

$$\mathbf{U}_T(\boldsymbol{\theta}) = \frac{1}{n}\sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)N_{ik}(t)\left\{\mathbf{X}_i(t) - \bar{\mathbf{X}}_k(t;\boldsymbol{\theta})\right\} dH_{ik}(t) = 0\,. \tag{3.3}$$

In the above, $Y_i(t) = I(C_i \geq t)$ is the at-risk indicator and

$$\bar{\mathbf{X}}_{\mathbf{k}}(t;\boldsymbol{\theta}) = \mathbf{S}_{1k}(t;\boldsymbol{\theta})/S_{0k}(t;\boldsymbol{\theta}),$$

where

$$S_{uk}(t;\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\mathbf{X}_i^{\otimes u}(t)\exp\left(\boldsymbol{\theta}^T X_i(t)\right) dH_{ik}(t)\ ,$$

$u = 0, 1, 2$ for $0 \leq t \leq \tau$, with $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$ and $a^{\otimes 2} = aa^T$ for some vector a.

Let $\hat{\boldsymbol{\theta}}_n$ denote the estimator of $\boldsymbol{\theta}$ given by the solution to the equation (3.3). Then one can estimate $\beta_j(t)$ by $\hat{\beta}_j(t) = \mathbf{B}_n^T(t)\,\hat{\boldsymbol{\alpha}}_{nj}$. In practice, sometimes one may also be interested in estimating the baseline mean function $\Lambda_k(t)$ and for this, it is apparent that one natural estimator is given by the Breslow-type estimator

$$\hat{\Lambda}_k(t, \hat{\boldsymbol{\theta}}_n) = \sum_{i=1}^{n}\frac{Y_i(t)N_{ik}(t)dH_{ik}(t)}{nS_{0k}(t;\hat{\boldsymbol{\theta}}_n)}\,.$$

3.4 Asymptotic Results

In this section, we establish the asymptotic properties of the proposed estimator, namely, the consistency, rate of convergence and asymptotic normality. For convenience, let

$$(N_{\cdot k}(t), H_{\cdot k}(t), \mathbf{W}(t), \mathbf{Z}(t))$$

be the population version of $(N_{ik}(t), H_{ik}(t), \mathbf{W}_i(t), \mathbf{Z}_i(t))$ for $k = 1, \ldots, K$. Let $\boldsymbol{\Lambda} = (\Lambda_1, \Lambda_2, \ldots, \Lambda_K)$ denote the $K-$tuple of all K baseline functions. Then let $\boldsymbol{\vartheta} = (\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\Lambda})$ and $\boldsymbol{\vartheta}_0 = (\boldsymbol{\gamma}_0, \boldsymbol{\beta}_0, \boldsymbol{\Lambda}_0)$ denote the true value of $\boldsymbol{\vartheta}$ with $\boldsymbol{\Lambda}_0 = (\Lambda_{01}, \Lambda_{02}, \ldots, \Lambda_{0K})$. Based on $\hat{\boldsymbol{\theta}}_n$, the estimator for $\boldsymbol{\vartheta}$ is $\hat{\boldsymbol{\vartheta}}_n = \left(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\beta}}_n, \hat{\boldsymbol{\Lambda}}\right)$ where $\hat{\boldsymbol{\Lambda}} = \left(\hat{\Lambda}_1(t, \hat{\boldsymbol{\theta}}_n), \ldots, \hat{\Lambda}_K(t, \hat{\boldsymbol{\theta}}_n)\right)$. Also for convenience, let $\mathbf{V}(t) = \left(\mathbf{W}^T(t), \mathbf{Z}^T(t)\right)^T$ and redefine the parameter space $\Theta = \mathcal{A} \times \mathcal{M} \times \mathcal{F}$, where $\mathcal{F} = \mathcal{F}_1 \times \cdots \times \mathcal{F}_K$ denotes the parameter space of $\boldsymbol{\Lambda}$ with $\mathcal{F}_k$ being the parameter space for Λ_k.

Let $\mathfrak{B}^d$ denote the collection of Borel sets in $\mathbb{R}^d$ and $\mathfrak{L}_2[0, \tau]$ the collection of Borel sets in $\mathcal{L}_2$ on $[0, \tau]$. Define $\mathfrak{B}^1[0, \tau] = \{B \cap [0, \tau] : B \in \mathfrak{B}^1\}$, $\mathfrak{B}^d = \mathfrak{B}^1[0, \tau] \times \ldots \times \mathfrak{B}^1[0, \tau]$ and $\mathfrak{L}_2^d[0, \tau] = \mathfrak{L}^1[0, \tau] \times \ldots \times \mathfrak{L}^1[0, \tau]$. For $k = 1, \ldots, K$, define the measure

$$\upsilon_{1k}(B_1 \times B_2 \times B_3) = \int_{B_3 \times B_2} \int_{B_1} dE\left[Y(t) H_{\cdot k}(t)\right] d\mu_Z \times \mu_W$$

for $B_1 \in \mathfrak{B}^1[0, \tau]$, $B_2 \in \mathfrak{L}_2^{p_2}[0, \tau]$ and $B_3 \in \mathfrak{L}_2^{p_1}[0, \tau]$, where μ_W and μ_Z are the measures for $\mathbf{W}$ and $\mathbf{Z}$. Alternatively, let $\mu_V = \mu_Z \times \mu_W$, we can rewrite $\upsilon_{1k}(B_1 \times B_2 \times B_3)$ as

$$\upsilon_{1k}(B_1 \times B_4) = \int_{B_4 \times B_1} dE\left[Y(t) H_{\cdot k}(t)\right] d\mu_V$$

for $B_1 \in \mathfrak{B}^1[0, \tau]$ and $B_4 \in \mathfrak{L}_2^p[0, \tau]$; and $\mu_{0k}(B_1) = \upsilon_{1k}(B_1 \times \mathfrak{L}_2^p[0, \tau])$, $k = 1, \ldots K$.

We further define

$$\tilde{\upsilon}_1 \left(B_1 \times B_4 \right) = \int_{B_4 \times B_1} \sum_{k=1}^{K} \Lambda_{0k}^2 \left(u \right) d\upsilon_{1k} \left(u, v \right)$$

and correspondingly $\tilde{\mu}_0 \left(B_1 \right) = \tilde{\upsilon}_1 \left(B_1 \times \mathfrak{L}_2^p \left[0, \tau \right] \right)$. Define the L_2 metric $d \left(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2 \right)$ on Θ as

$$d \left(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2 \right) = \left(\| \boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2 \|_2^2 + \| \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2 \|_{L_2(\tilde{\mu}_0)}^2 + \sum_{k=1}^{K} \| \Lambda_{1k} - \Lambda_{2k} \|_{L_2(\mu_{0k})}^2 \right)^{1/2} .$$

To establish the asymptotic results, we need the following regularity conditions.

(C1) The observation process has the rate function

$$E \left[dH_{\cdot k}^* \left(t \right) | \mathbf{W}(t), \mathbf{Z}(t), C \right] = \omega_k \left(t \right) dt,$$

where $\omega_k \left(t \right)$ is a bounded, nonnegative and continuous function on $[0, \tau]$, for $k = 1, \ldots, K$. There exists a positive integer M_0 such that $\Pr \left(H_{\cdot k} \left(\tau \right) < M_0 \right) = 1$ for all $k = 1, \ldots, K$. That is, the total observation number is finite. Moreover, the support of all $\omega_k \left(t \right)$, $k = 1, \ldots, K$, is $[\tau_0, \tau]$ with $\tau_0 > 0$ and $\Lambda_{0k} \left(\tau_0 \right) > 0$ for some constant τ_0.

(C2) The measure $\mu_{0k} \times \mu_V$ is absolutely continuous with respect to υ_{1k} as well as $\mu_{0k}(\{\tau\}) > 0$, for $k = 1, \ldots, K$.

(C3) The parameters space of Λ_k, $\mathcal{F}_k$, consists of bounded non-decreasing functions in $\mathcal{L}_2$ over $[0, \tau]$, for $k = 1, \ldots, K$.

(C4) The parameters space of $\boldsymbol{\beta}$, $\mathcal{M}$, is bounded and convex in $\mathcal{L}_2([0, \tau])$. Each component of the true value of $\boldsymbol{\beta}(t)$, denoted by $\beta_{0j}(t)$, $j = 1, \ldots, p_2$, is continuously rth differentiable in $[0, \tau]$.

(C5) The parameter space of $\boldsymbol{\gamma}$, $\mathcal{A}$, is bounded and convex in $\mathbb{R}^d$.

(C6) The covariate vector $\mathbf{V}(t) = \left(\mathbf{W}^T(t), \mathbf{Z}^T(t)\right)^T$ is uniformly bounded over $[0, \tau]$ with the distribution μ_V.

(C7) Given $\mathbf{V}(t)$, $t \in [0, \tau]$, C and N are independent. Besides, with probability 1,

$$\begin{aligned} &\inf_{\mathbf{V}(t), t\in[0,\tau]} \Pr\left(C \geq \tau | \mathbf{V}(t) = \mathbf{v}(t), t \in [0, \tau]\right) \\ &\qquad\qquad = \inf_{\mathbf{V}(t), t\in[0,\tau]} \Pr\left(C = \tau | \mathbf{V}(t) = \mathbf{v}(t), t \in [0, \tau]\right) > 0. \end{aligned}$$

(C8) If $\boldsymbol{\gamma}^T\mathbf{W}(t) + \boldsymbol{\beta}^T(t)\mathbf{Z}(t) \equiv 0$, $t \in [0, \tau]$ with probability 1 for some $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$, then $\boldsymbol{\gamma} = 0$ and $\boldsymbol{\beta}(t) = 0$ for $t \in [0, \tau]$.

(C9) The function $M_{0k}(\mathbf{V}) = \int N_{\cdot k}(t) \log\left(N_{\cdot k}(t)\right) dH_{\cdot k}(t)$ satisfies $\mathbf{P}M_{0k}(\mathbf{V}) < \infty$, for $k = 1, \ldots K$.

(C10) $E\left[\exp\left(C_0 N_{\cdot k}(t)\right)\right]$ is bounded in $[0, \tau]$ for some constant C_0 for $k = 1, \ldots K$.

(C11) The true baseline mean function Λ_{0k} is differentiable in $[\tau_0, \tau]$ for $k = 1, \ldots, K$. Moreover, the first order derivative of Λ_{0k} has a positive and finite lower and upper bound in $[\tau_0, \tau]$, $k = 1, \ldots, K$.

(C12) There exist $\eta_1 \in (0, 1)$ such that

$$a^T \mathrm{Var}\left(\mathbf{V}(U) | U\right) a \geq \eta_1 a^T E\left(\mathbf{V}^T(U)\, \mathbf{V}(U) | U\right) a,$$

a.s. for all $a \in \mathbb{R}^{p_1+p_2}$, where $(U, \mathbf{V})$ has distribution $\upsilon_{1k}/\upsilon_{1k}(\mathbb{R}^+ \times \mathcal{V})$, $k = 1, \ldots, K$.

Note that conditions (C1) and (C7) are common on the observation schemes and similar to the combination of C8, C10 and C11 in Lu et al. (2009) for univariate case. Condition (C2) comes from the condition in Theorem 1 of Wellner and Zhang (2007) and Theorem 1 of Lu et al. (2009), ensuring each $\hat{\Lambda}_k$ is bounded, $k = 1, \ldots K$. Conditions (C6)-(C11) are common assumptions in the semiparametric estimation with adpation to the multivariate panel count data. Also conditions (C2) and (C8) ensure the identifiability of the semiparametric model and conditions (C9), (C10) and (C11) are adapted from conditions C4, C10 and C12 in Wellner and Zhang (2007). Condition (C12) is needed to prove the convergence rate and can be justified by the arguments similar to those in Wellner and Zhang (2007) as it does not involve the multivariate responses. Similar to Remark 3.4 in Wellner and Zhang (2007), Condition (C8) and (C12) imply that $E\left(\mathbf{V}^T(U)\,\mathbf{V}(U)\,|U\right)$ and $\mathrm{Var}\left(\mathbf{V}(U)\,|U\right)$ are positive definite a.e. υ_{1k}, which is crucial to establish the rate of convergence.

Theorem 4 (Consistency)**.** *Assume that the regularity conditions (C1)-(C9) given above hold. Then we have that* $d\left(\hat{\boldsymbol{\vartheta}}_n, \boldsymbol{\vartheta}_0\right) \to 0$ *in probability as* $n \to \infty$.

Theorem 5 (Rate of Convergence)**.** *Assume that the regularity conditions (C1)-(C12) given above hold. Then we have that*

$$n^{\min\left\{n^{\frac{1-\nu}{3}}, n^{r\nu}\right\}} d\left(\hat{\boldsymbol{\vartheta}}_n, \boldsymbol{\vartheta}_0\right) = O_p(1)$$

with the optimal rate $O_p(n^{-r/(3r+1)})$ *achieved at* $\nu = 1/(1+3r)$.

The metric $d(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2)$ in Theorems 4 and 5 has different measures for $\boldsymbol{\beta}$ and $\boldsymbol{\Lambda}$.

The convergence of Λ_k is characterized by the observation process $H_{\cdot k}$ only while the convergence of $\boldsymbol{\beta}$ are defined on a measure depending on both $H_{\cdot k}$ and Λ_k. This property of the estimator is different from the time-constant coefficient cases for the panel count data.

Theorem 6 (Asymptotic Normality)**.** *Assume that the regularity conditions (C1)-(C12) given above hold and also* $(4r)^{-1} < \nu < 2^{-1}$ *with* $r > 1$*. Define* $\mathcal{H}_1 = \{\boldsymbol{h}_1 : \boldsymbol{h}_1 \in \mathcal{A}, \|\boldsymbol{h}_1\| \leq 1\}$*,* $\mathcal{H}_2 = \{\boldsymbol{h}_2 : \boldsymbol{h}_2 \in \mathcal{M},$ *each component of* $\boldsymbol{h}_2$ *is of bounded total variation.*$\}$*, and* $\mathcal{H}_3 = \{h_3 = (h_{31}, \ldots, h_{3K}) : h_{3k}$ *is a fucntion with bouned total variation in* $[0, \tau]$*, and* $h_{3k}(0) = 0, k = 1, \ldots, K\}$ *Then for some* $(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3) \in \mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$*, we have that*

$$\begin{aligned}\sqrt{n}\left(\boldsymbol{\gamma} - \gamma_0\right)^T \boldsymbol{h}_1 + \sqrt{n} \int_0^\tau \left(\boldsymbol{\beta}\left(t\right) - \boldsymbol{\beta}_0\left(t\right)\right)^T d\boldsymbol{h}_2\left(t\right) \\ + \sqrt{n} \sum_{k=1}^K \int_0^\tau \left(\Lambda_k\left(t\right) - \Lambda_{k0}\left(t\right)\right) dh_{3k}\left(t\right) \\ \to_d N\left(0, \sigma^2\right),\end{aligned}$$

where σ^2 *is given in the Appendix.*

The proof of the results above are provided in the Appendix. Similar to He et al. (2017), we cannot find the explicit form of the asymptotic distribution because the explicit forms of $(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3)$ cannot be solved though exist. On the other hand, by following Amorim et al. (2008), we can have the following *ad hoc* estimators for the asymptotic covariance matrix of $\hat{\boldsymbol{\gamma}}$ and the pointwise asymptotic variance of $\boldsymbol{\beta}(t)$ in $t \in [0, \tau]$. First the asymptotic covariance matrix of $\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma_0}$ can be consistently estimated by $\hat{A}_{\hat{\boldsymbol{\gamma}}}(t)^{-1}\hat{B}_{\hat{\boldsymbol{\gamma}}}(t)\hat{A}_{\hat{\boldsymbol{\gamma}}}(t)^{-1}$. Here $\hat{A}_{\hat{\boldsymbol{\gamma}}}(t)$ and $\hat{B}_{\hat{\boldsymbol{\gamma}}}(t)$ are the top-left $p_1 \times p_1$ sub-matrices of

$$\begin{aligned}
\hat{A}(t) &= \frac{\partial \mathbf{U}_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\partial \sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^\tau Y_i(t)N_{ik}(t)\left\{(\mathbf{X}\,(t)) - \bar{\mathbf{X}}_k(t;\boldsymbol{\theta})\right\} dH_{ik}(t)}{\partial \boldsymbol{\theta}} \\
&= \sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^\tau Y_i(t)N_{ik}(t)\left\{-\frac{\partial \bar{\mathbf{X}}_k(t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\} dH_{ik}(t)\,,
\end{aligned}$$

and

$$\begin{aligned}
&\hat{B}(t) \\
&= \sum_{i=1}^{n}\left[\int_0^\tau \sum_{k=1}^{K} Y_i(t)\left(N_{ik}(t) - \Lambda_k(t)\exp(\boldsymbol{\theta}^T\mathbf{X}_i\,(t))\right)\left\{\mathbf{X}\,(t) - \bar{\mathbf{X}}_k(t;\boldsymbol{\theta})\right\} dH_{ik}(t)\right]^{\otimes 2},
\end{aligned}$$

respectively, where

$$\frac{\partial \bar{\mathbf{X}}_k(t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\partial \tilde{\mathbf{S}}_{1k}(t;\boldsymbol{\theta})/\tilde{S}_{0k}(t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \tilde{\mathbf{S}}_{2k}(t;\boldsymbol{\theta})/\tilde{S}_{0k}(t;\boldsymbol{\theta}) - (\tilde{\mathbf{S}}_{1k}(t;\boldsymbol{\theta})/\tilde{S}_{0k}(t;\boldsymbol{\theta}))^2\,.$$

Furthermore, the asymptotic variance of $\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}_0(t)$ for a given t can be estimated by

$$\hat{\Omega}(t) = \mathbf{B}^T(t)\left(\hat{A}_{\hat{\boldsymbol{\alpha}}}(t)^{-1}\hat{B}_{\hat{\boldsymbol{\alpha}}}(t)\hat{A}_{\hat{\boldsymbol{\alpha}}}(t)^{-1}\right)\mathbf{B}(t),$$

where $\hat{A}_{\hat{\boldsymbol{\alpha}}}(t)$ and $\hat{B}_{\hat{\boldsymbol{\alpha}}}(t)$ are the bottom-right $p_2 \times p_2$ sub-matrices of $\hat{A}(t)$ and $\hat{B}(t)$, respectively. The numerical study in the next section shows these variance estimators work well.

3.5 Simulation

3.5.1 Data Generation

To evaluate the operational characteristics of the proposed method, we conduct a simulation study. In the simulation, we set $K = 2$. We first generate the follow-up time C_i from the uniform distribution over $(0.6\tau,\ \tau)$. The observation time t_{ik}'s are generated from nonhomogeneous Poisson process with means $\tilde{\Lambda}_1(t)$ $=3t + 4$ and $\tilde{\Lambda}_2(t)$ $=2t + 3$. We assume that baseline hazard functions $\Lambda_1(t) = 4t, \Lambda_2(t) = 4t^2$. The maximum follow-up time $\tau = 1$.

For the degree of B-spline functions, three different setting are considered: linear or piecewise B-spline basis, quadratic B-spline basis and the cubic B-spline models. For each combination defined by the model complexity and the shape of the true coefficient functions, 1000 samples of size 500 or 800 are generated. For each configuration, we present the sampling bias, sampling/empirical standard deviation (ESD), standard error (SE) and the coverage probability (CP) of the 95% confidence interval.

3.5.2 Simulation Results

We assume that $p_1 = 1$, $p_2 = 1$, $\gamma_1 = 1, 2$. To generate covariate process, we assume that $W_1(t)$ is time dependent variable and generate $W_1(t)$ by imitating two-stage randomization:

$$W_1(t) = B_1 I(t \leq V) + B_2 I(t > V),$$

where B_1 and B_2 are independent $U(0, 0.5)$ and $U(0.5, 1)$, and $V \sim U(0, \tau)$ with the maximum follow up time $\tau = 1$. We generate $Z_1(t)$ independently by the similar way.

To make our methods more general, we add a frailty in our model to simulate a mixed-Possion process. We still assume that

$$N_{ik}\left(T_{ik,l}\right) = N_{ik}^{*}\left[\Lambda_k\left(T_{ik,1}\right)\right] + N_{ik}^{*}\left[\Lambda_k\left(T_{ik,2}\right) - \Lambda_k\left(T_{ik,1}\right)\right] + \cdots + N_{ik}^{*}\left[\Lambda_k\left(T_{ik,l}\right) - \Lambda_k\left(T_{ik,l-1}\right)\right], \quad (3.4)$$

$l = 1, ..., m_{ik}, k = 1, 2, i = 1, ..., n, T_{ik,0} = 0$. However, $N_{ik}^{*}\left[\Lambda_k\left(t\right)\right]$ and

$$N_{ik}^{*}\left[\Lambda_k\left(T_{ik,l}\right) - \Lambda_k\left(T_{ik,l-1}\right)\right]$$

here are assumed to follow a distribution with the mean functions

$$Q_i \Lambda_k\left(t_{i,1}\right) e^{\left(\boldsymbol{\gamma}^T \mathbf{W}_i(t) + \boldsymbol{\beta}^{\boldsymbol{T}}(t) \mathbf{Z}_i(t)\right)}$$

and

$$Q_i \Lambda_k\left(t_{i,j}\right) e^{\left(\boldsymbol{\gamma}^T \mathbf{W}_i(t) + \boldsymbol{\beta}^{\boldsymbol{T}}(t) \mathbf{Z}_i(t)\right)} - Q_i \Lambda_k\left(t_{i,j-1}\right) e^{\left(\boldsymbol{\gamma}^T \mathbf{W}_i(t) + \boldsymbol{\beta}^{\boldsymbol{T}}(t) \mathbf{Z}_i(t)\right)}$$

respectively, where Q_i is an unobserved random effect (or frailty), independent of covariates, following a standard gamma distribution with mean 1 and variance $\sigma_Q^2 = 1$ or 2. Note that σ_Q^2 measures the correlation between the 2 recurrent processes N_{i1} and N_{i2}.

Estimation of Coefficients

Under cubic B-splines models, Table 3.1 presents some important summary statistics for $\boldsymbol{\gamma}$ with 3 interior knots, where setting 1: $\beta_1(t) = t/2$; setting 2: $\beta_1(t) = t^2/2$ and setting 3: $\beta_1(t) = (sin(4\pi t) + 4\pi t)/24$ under different samples sizes, real values of $\boldsymbol{\gamma}$ and σ_Q^2 values. Figure 3.1 presents the true and the average estimated time-dependent coefficients $\boldsymbol{\beta(t)}$ curves under $\sigma_Q^2 = 1$, sample sizes 800, over 1000 iterations, over 1000 equal-spaced grid points on the time axis under setting 1 and setting 2.

Figure 3.1 (a1) and (a2) show the true and the average estimated $\boldsymbol{\beta}(t)$ curves. The estimated curve is generally close to the true curve. We present the variances of $\boldsymbol{\beta}(t)$ and the ESD of $\boldsymbol{\beta(t)}$ under sample size 800 in Figure 3.1 (a3,a4). The figures shows that empirical variance of $\boldsymbol{\beta}(t)$ is very close to the variance of $\boldsymbol{\beta}(t)$ by the asymptotic result. The variations of spline estimators increase at the boundary.

Figure 3.2 presents time-dependent coefficients results for $\sigma_Q^2 = 2$, where other settings are similar to Figure 3.1. Figure 3.2 (a1) and (a2) show that the average curves for $\boldsymbol{\beta}(t)$ estimated under sample sizes 800 are close to the true curve, except for a small bias near the end of the time window. The coverage probabilities of the *ad hoc* confidence intervals are generally close to the nominal level. Similar to Figure 3.1, we also provide the variances of $\boldsymbol{\beta}(t)$ and the ESD of $\boldsymbol{\beta}(t)$ under sample size 800 in Figure 3.2 (a3,a4). The figures shows that the emprical standard deviation(ESD) of $\boldsymbol{\beta}(t)$ is very close to standard error (SE) of $\boldsymbol{\beta}(t)$.

In summary, the proposed estimation procedure produce unbiased estimators. The asymptotic and empirical standard deviation estimates are quite similar. As the theory indicate, the coverage probabilities of the 95% confidence intervals of estimators approach its nominal level as the sample size increases.

Diffrent Baseline Mean Function $\Lambda(t)$ and $\tilde{\Lambda}(t)$

Here we consider different values of $\Lambda(t)$ and $\tilde{\Lambda}(t)$, Table 3.2 shows the estimation results for $\boldsymbol{\gamma}$'s under the same set-ups as for Table 3.1 under $\sigma^2 = 1$ except that for here, setting 4: $\Lambda_1(t) =(\sin(4\pi t) + 4\pi t)/\pi$ and $\Lambda_2(t) =(\cos(4\pi t) + 4\pi t)/\pi$ or setting 5: $\tilde{\Lambda}_1(t) =4t + 5$ and $\tilde{\Lambda}_2(t) =3t + 4$. The proposed methods are robust to different $\Lambda(t)$ and $\tilde{\Lambda}(t)$ functions.

Degree of Splines

In this subsection, we provide the results for piecewise/linear, quadratic and cubic B-splines models with sample sizes 800. We show the results for the simulation studies considering the aforementioned B-splines models for γ_1 and $\beta_1(t)$ under setting 1 in Table 3.3. Over all, the estimation procedure has satisfying results under all spline models.

3.6 An Application

In this section, we again apply the methodology proposed in the previous sections to the previous CHNS data. The China Health and Nutrition Survey (CHNS) was conducted from 1989 to 2015 for every 2-4 years. The pregnancy information and number of marriage times of the participants was collected during this time points, therefore, we only have incomplete panel count data over the whole follow-up period or some follow-up periods. More details about this survey could be found in Section 2.5. Here we still focus on the four interesting variables: average monthly wage last year, whether the mother came from urban or rural areas (urban = 0, rural

= 1), the completed years of formal education in regular school (0: No school; 1: Primary school; 2: Middle school; 3: Technical school; 4: College), and the current health status (1: Excellent, 2: Good, 3: Fair, 4: Poor). Specifically, we will focus on a subgroup of the female participants with the required covariate and response information($n = 2402$).

The final subgroup includes 2402 subjects. For the analysis, define $N_{i1}(t)$ and $N_{i2}(t)$ as the cumulative numbers of pregnancy and marriage times up to time t for patient i , respectively, $i = 1, ..., 2402, k = 1, 2$. γ and $\beta(t)$ denote the corresponding parameters. To decrease the effect of magnitude due to the covariate range, we normalized the average monthly wage last year. For the observed data, the average pregnancy counts are 1.4610, the average marriage times are 1.0190.

To apply the proposed estimation procedure, we first assume that all of the four covariates have time-varying effects and Figure 3.3 presents the estimated covariate effects with the use of 3 interior knots under multivariate analysis. Figure 3.4 and Figure 3.5 presents the corresponding results under seperate univariate analysis of pregnancy numbers and marriage times. The results based on the joint analysis suggest that all 4 covariates have significant effects on the rates of both pregnancy and marriage times, although the time-varying pattern of wage and current health status seems to be not very obvious. Specifically, the mothers from rural areas with higher salaries and lower education levels seem to have higher rates. The univariate analyses gave similar conclusions for most of the covariates effects except that they significantly underestimated the effect of wage on the rate of interested events. It can be seen that the estimated effects of all covariates based on the joint analysis are intermediate between those based on the 2 univariate analyses.

We also consider the case when all covariates are assumed to be time-independent. For comparison, we also perform univariate analyses which involves separate modeling of covariate effects for pregnancy number and marriage counts and include the results in Table 3.4. Table 3.4 shows the results of joint and univariate analysis results for locations, wage, education levels and current health status, correspondingly for $\gamma_1, \gamma_2, \gamma_3$, and γ_4. It still can be seen that the estimated effects of all covariates based on the multivariate analysis are intermediate between those based on the 2 univariate analyses. Futhermore, the joint analysis shows the education level is a significant variable, which, didn't be identified under univariate analysis for marriage times.

Last, according to our previous results, we assume that only the location and education levels have time-varying coefficient effects, while the wage and health status have constant effects on the events. Table 3.5 gives the estimated covariate effects for the wage and current health status under multivariate and univariate analysis with different number of knots. Figure 3.6 presents the corresponding results under multivariate analysis of pregnancy numbers and marriage times with 3 interior knots. The multivariate analysis results with the use of 5 interior knots are shown in Figure 3.7 and similar to Figure 3.6, which again shows the locations and education levels have significant positive and negative effects on events.

3.7 Concluding Remarks

In this chapter, we have discussed regression analysis of multivariate panel count data when both the covariates and their effects are varying with time. Such data naturally occur when a recurrent event study involves several related types of recurrent events.

The areas in which one often faces such data include clinical trials and medical or social follow-up studies. For the problem, we presented a flexible marginal mean model for the underlying recurrent event processes and an estimating equation-based inference procedure was developed for estimation of regression parameters to combine with the B-splines. Both finite sample and asymptotic properties of the proposed estimates have been established and the simulation results indicated that the procedure work well for practical situations. The methodology was applied to a set of bivariate panel count data arising from CHNS.

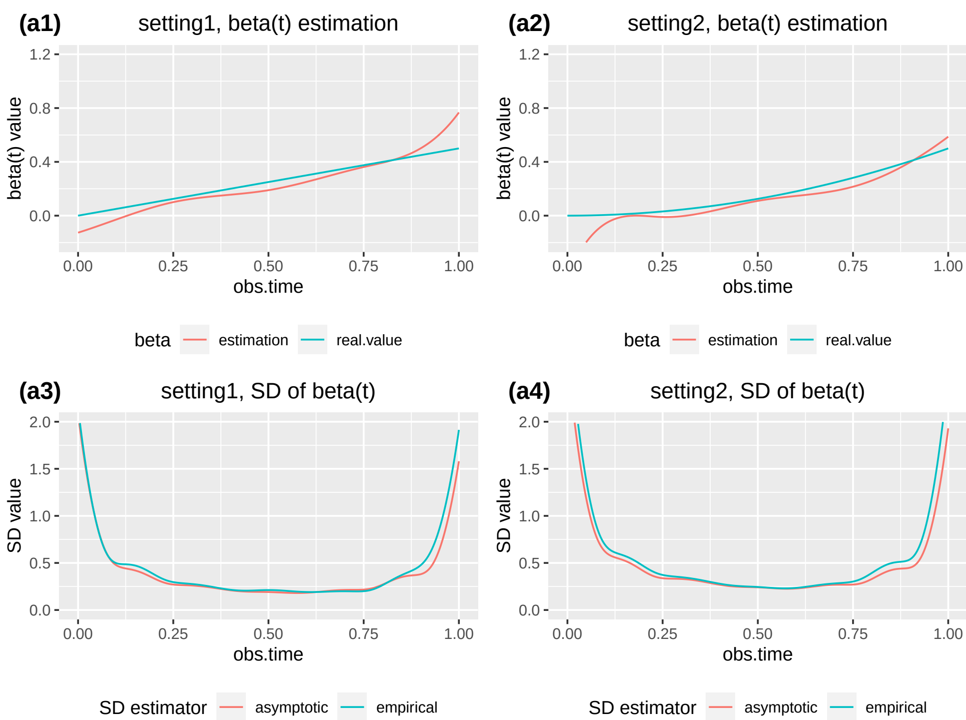

Figure 3.1: Simulation results on estimation of $\beta(t)$ with $\sigma^2 = 1$, where, setting 1: $\beta(t) = t/2$; setting 2: $\beta(t) = t^2/2$.

	Sample Size	Para	Real Value	Bias	ESD	SE	CP
$\sigma_Q^2 = 1$							
setting1	500	γ_1	1	-0.0243	0.2200	0.2030	0.922
	800	γ_1	1	-0.0153	0.1765	0.1648	0.937
	500	γ_1	2	0.0044	0.2526	0.2234	0.948
	800	γ_1	2	0.0060	0.1801	0.1764	0.949
setting2	500	γ_1	1	0.0034	0.2056	0.1982	0.937
	800	γ_1	1	0.0089	0.1827	0.1695	0.945
	500	γ_1	2	0.0141	0.2339	0.2197	0.933
	800	γ_1	2	-0.0030	0.1768	0.1760	0.968
setting3	500	γ_2	1	-0.0308	0.2372	0.2126	0.961
	800	γ_1	1	-0.0004	0.1766	0.1671	0.921
	500	γ_2	2	0.0141	0.2339	0.2197	0.933
	800	γ_1	2	0.0356	0.1962	0.1859	0.950
$\sigma_Q^2 = 2$							
setting1	500	γ_1	1	0.0324	0.3097	0.2732	0.911
	800	γ_1	1	-0.0052	0.2364	0.2282	0.975
	500	γ_1	2	0.0493	0.3388	0.2942	0.907
	800	γ_1	2	-0.0108	0.2240	0.2402	0.973
setting2	500	γ_1	1	0.0387	0.2581	0.2798	0.980
	800	γ_1	1	0.0151	0.2490	0.2226	0.926
	500	γ_1	2	-0.0157	0.3201	0.2957	0.917
	800	γ_1	2	-0.0169	0.2443	0.2456	0.946
setting3	500	γ_2	1	0.0165	0.3021	0.2623	0.920
	800	γ_1	1	-0.0287	0.2236	0.2217	0.930
	500	γ_2	1	-0.0048	0.3179	0.2736	0.917
	800	γ_2	1	-0.0037	0.2537	0.2268	0.918
	500	γ_2	2	-0.0056	0.3283	0.2956	0.915
	800	γ_1	2	-0.0033	0.2674	0.2429	0.920

Table 3.1: Simulation results on estimation of γ for different sample sizes, $\beta(t)$ functions and σ_Q^2's wtih 3 knots.

$\sigma_Q^2 = 1$	Sample Size	Para	Real Value	Bias	ESD	SE	CP
setting4	500	γ_1	1	-0.0225	0.2295	0.1962	0.915
	800	γ_1	1	-0.0038	0.1675	0.1603	0.931
setting5	500	γ_1	1	-0.0025	0.2233	0.2057	0.937
	800	γ_1	1	0.0027	0.1774	0.1686	0.939

Table 3.2: Simulation results on estimation of γ for different Λ and $\tilde{\Lambda}$ functions under $\sigma_Q^2 = 1$ with 3 knots.

Spline Model	Real value	Sample size	BIAS	ESD	SE	ECP
Piecewise/linear spline	1	800	-0.0013	0.1738	0.1667	0.936
Quadratic spline	1	800	-0.0061	0.1805	0.1684	0.935
Cubic spline	1	800	-0.0153	0.1765	0.1648	0.937

Table 3.3: Simulation results on estimation of γ for different degrees of spline.

Multivariate analysis	$\hat{\gamma}$	$SE(\hat{\gamma})$	CI($\hat{\gamma}$)
location	0.0924	0.0120	(0.0689,0.1158)
wage	-0.0119	0.0062	(-0.0240,0.0003)
education level	-0.0851	0.0070	(-0.0988,-0.0713)
current health status	0.0181	0.0068	(0.0049,0.0314)
Univariate analysis - pregnancy number	$\hat{\gamma}$	$SE(\hat{\gamma})$	CI($\hat{\gamma}$)
location	0.1638	0.0210	(0.1226,0.2050)
wage	-0.0227	0.0135	(-0.0491,0.0038)
education level	-0.1503	0.0123	(-0.1744,-0.1262)
current health status	0.0269	0.0119	(0.0035,0.0503)
Univariate analysis - marriage number	$\hat{\gamma}$	$SE(\hat{\gamma})$	CI($\hat{\gamma}$)
location	0.0018	0.0039	(-0.0059,0.0095)
wage	-0.0005	0.0007	(-0.0018,0.0008)
education level	-0.0032	0.0021	(-0.0074,0.0010)
current health status	0.0079	0.0029	(0.0023,0.0135)

Table 3.4: When considering all coefficients to be time-independent. Results of joint and univariate analyses of pregnancy and marriage numbers from CHNS data.

nknots	Multivariate analysis	$\hat{\gamma}$	$SE(\hat{\gamma})$	CI($\hat{\gamma}$)
3	wage	-0.0118	0.0062	(-0.0240,0.0004)
3	current health status	0.0170	0.0068	(0.0036,0.0304)
5	wage	-0.0118	0.0062	(-0.0238,0.0004)
5	current health status	0.0170	0.0068	(0.0036,0.0303)
7	wage	-0.0118	0.0062	(-0.0240,0.0004)
7	current health status	0.0170	0.0068	(0.0036,0.0304)
nknots	Univariate analysis pregnancy number	$\hat{\gamma}$	$SE(\hat{\gamma})$	CI($\hat{\gamma}$)
3	wage	-0.0228	0.0137	(-0.0496,0.0040)
3	current health status	0.0257	0.0120	(0.0022,0.0492)
nknots	Univariate analysis marriage number	$\hat{\gamma}$	$SE(\hat{\gamma})$	CI($\hat{\gamma}$)
3	wage	-0.0004	0.0007	(-0.0017,0.0009)
3	current health status	0.0078	0.0029	(0.0021,0.0134)

Table 3.5: When considering the coefficients of location and education levels to be time-dependent. Results of joint and univariate analyses of pregnancy and marriage numbers from CHNS data under different numbers of knots.

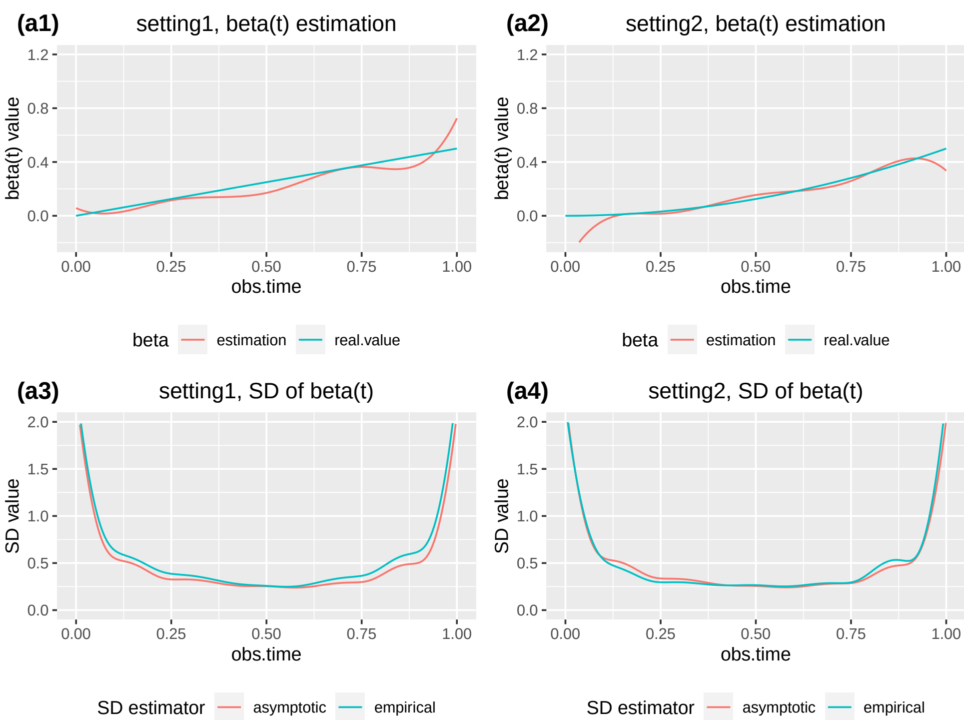

Figure 3.2: Simulation results on estimation of $\beta(t)$ with $\sigma^2 = 2$, where, setting 1: $\beta(t) = t/2$; setting 2: $\beta(t) = t^2/2$.

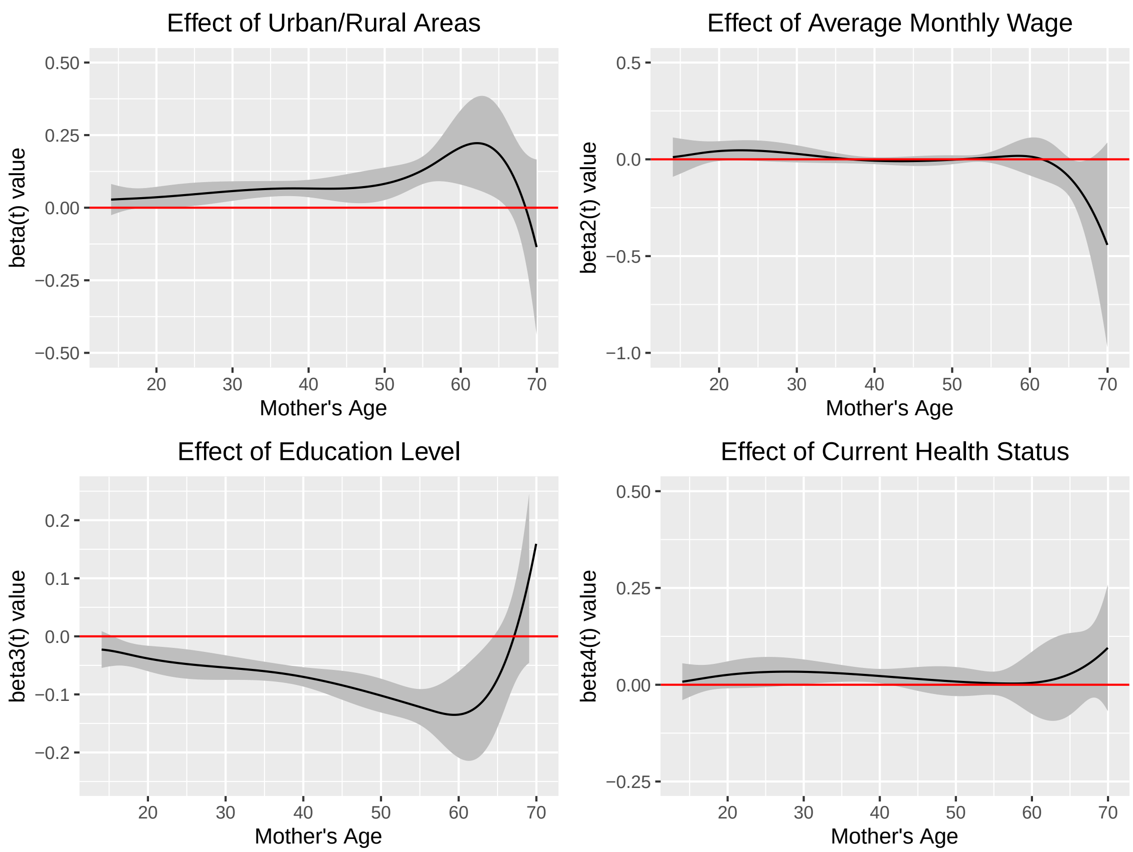

Figure 3.3: To consider all coefficients to be time-dependent. Multivariate analysis of pregnancy numbers and marriage times. Estimated time-varying effects (solid curves) and corresponding pointwise 95% confidence intervals (grey curves).

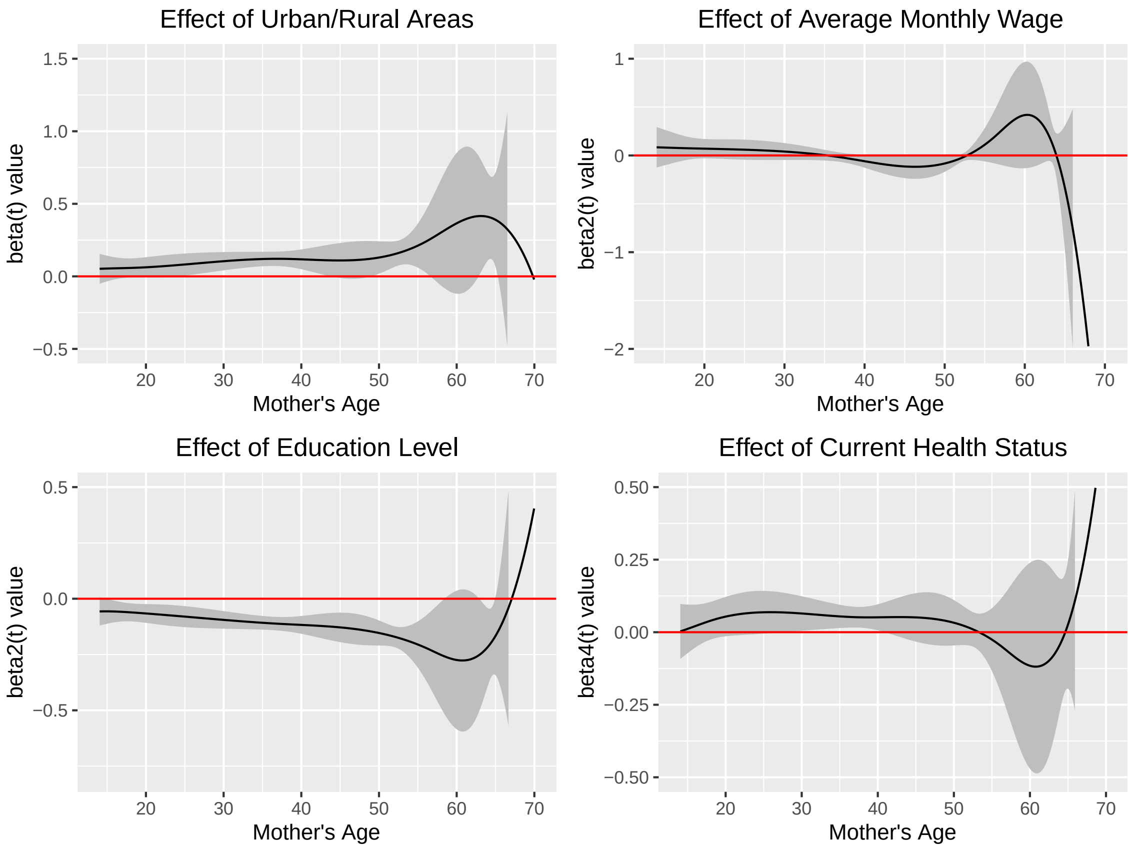

Figure 3.4: To consider all coefficients to be time-dependent. Univariate analysis of pregnancy numbers. Estimated time-varying effects(solid curves) and corresponding pointwise 95% confidence intervals (grey curves).

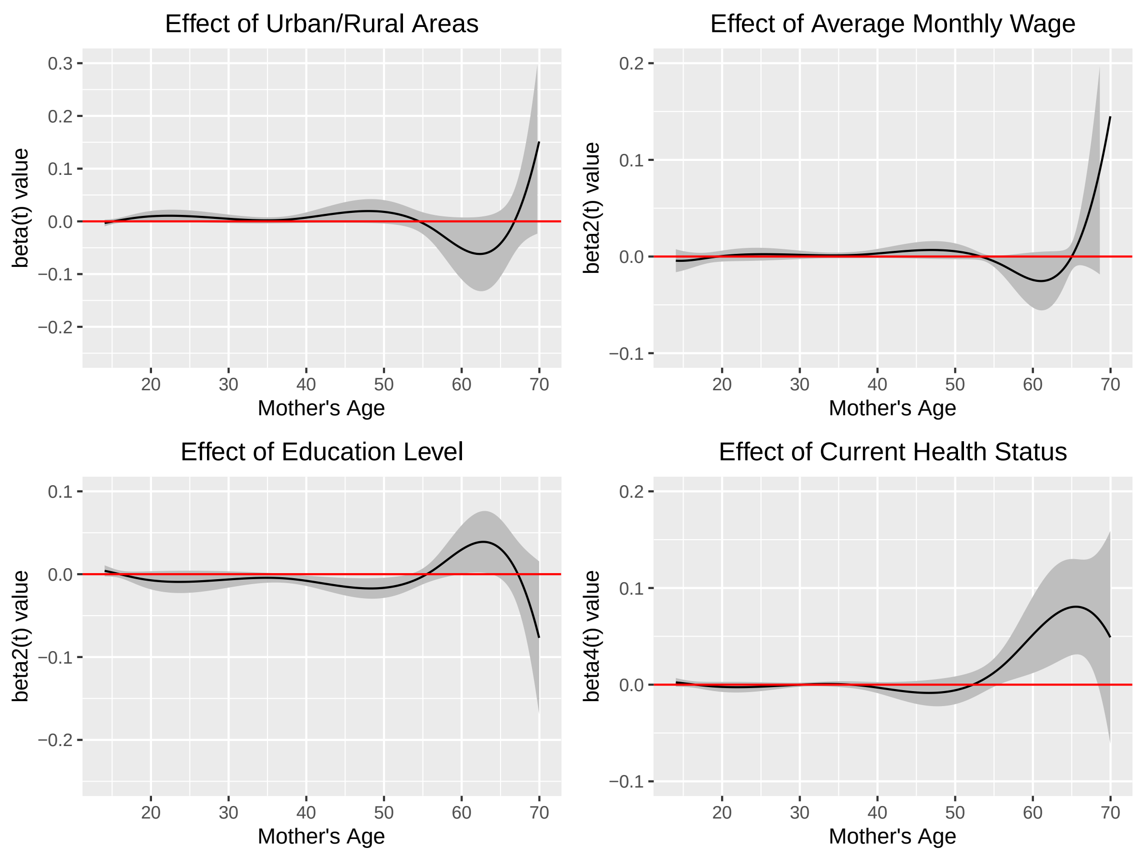

Figure 3.5: To consider all covariates to be time-dependent. Univariate analysis of marriage times. Estimated time-varying effects(solid curves) and corresponding pointwise 95% confidence intervals (grey curves).

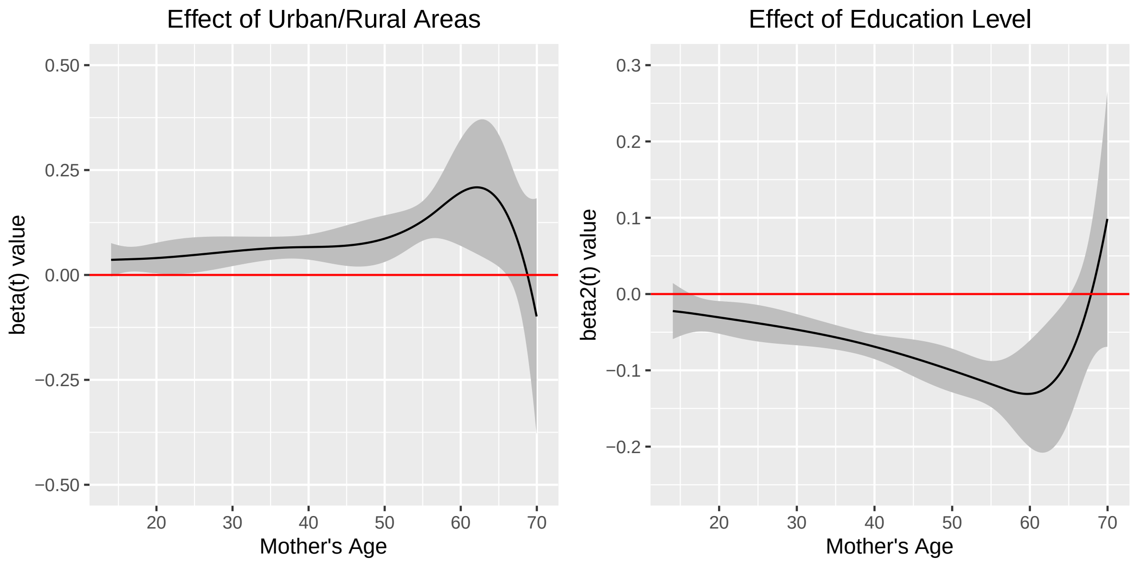

Figure 3.6: To consider the coefficients of location and education level to be time-dependent. 3 interior knots. Multivariate analysis of pregnancy numbers and marriage times. Estimated time-varying effects(solid curves) and corresponding pointwise 95% confidence intervals (grey curves).

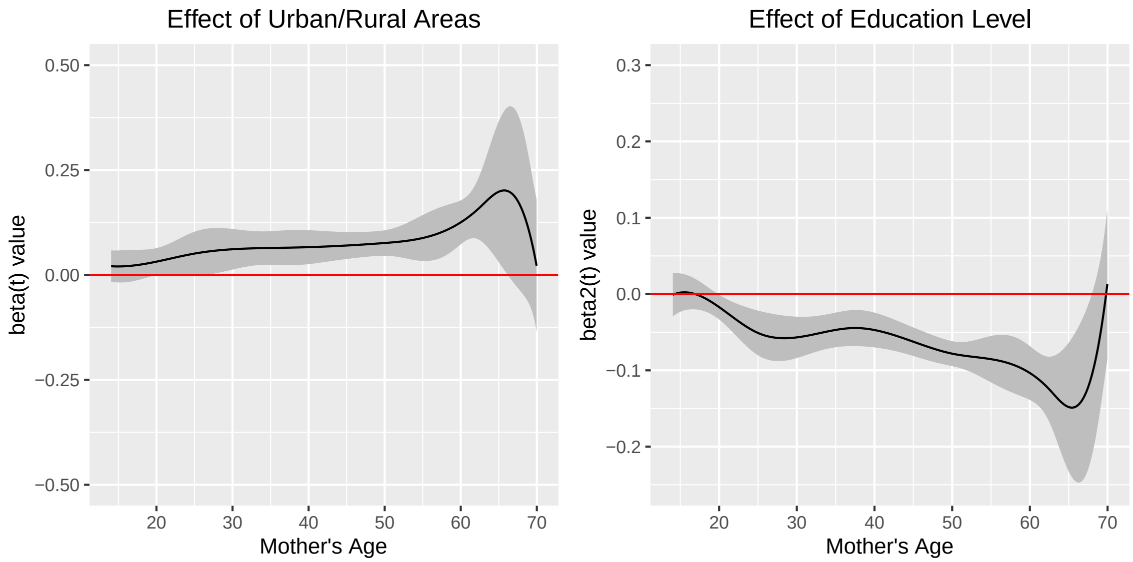

Figure 3.7: To consider coefficients of location and education level to be time-dependent. 5 interior knots. Multivariate analysis of pregnancy numbers and marriage times. Estimated time-varying effects(solid curves) and corresponding pointwise 95% confidence intervals (grey curves).

Chapter 4

Dynamic Analysis of Univariate Panel Count Data with Rate Model

4.1 Introduction

Statistical methods for panel count data usually focus on studying the relationship of covariates and the underlying recurrent event processes $N(t)$. To the best of our knowledge, all methods analyzing panel count data with time-dependent effects are based on mean model, like we did in Chapters 2 and 3. However, as claimed in Section 1.4, one drawback of mean model is that, when coefficients or covariates fluctuate, it is hard to satisfy the non-decreasing property of the mean function. Moreover, when we predict the mean function of a new subject, there is no guarantee that the predicted mean function is non-decreasing if coefficients or covariates are time-dependent. Therefore, we consider an alternative semi-parametric model based on the rate or intensity function in this chapter, which is a more flexible and realistic model to incorporate time-dependent effects in inference.

The rest of this chapter is organized as follows. Section 4.2 presents the proposed likelihood-based method with the proposed B-spline approximation and the EM algorithm to maximize the log-likelihood function. Section 4.3 proves that the estimators are consistent, efficient and asymptotically normally distributed. A simulation study illustrates the finite sample properties of the proposed method is provided in Section 4.4. In Section 4.5, we apply the proposed approach to the data arising from The Young Women's Project (YWP). Some concluding remarks are given in Section 4.6.

4.2 Estimation Method

4.2.1 Spline Approximation

As introduced in Section 1.4, here we consider a semi-parametric model based on the rate or intensity function of $N(t)$, that is

$$E\left[dN(t)\,|W(t), Z(t)\right] = \exp\left(\gamma^T W(t) + \beta^T(t) Z(t)\right) d\Lambda(t)\,. \tag{4.1}$$

Here, $\Lambda(t)$ is an unspecified non-decreasing baseline function , $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_{p_1})^T$ and $\boldsymbol{\beta}(t) = (\beta_1(t), ..., \beta_{p_2}(t))^T$ represent constant and time-dependent coefficients, respectively.

For estimation of γ and $\beta(t)$, let $\mathcal{A}$ and $\mathcal{B}_j$ denote the parameter spaces for γ and β_j, respectively, $j = 1, \ldots, p_2$, and assume that $\mathcal{A}$ is a compact subset of $\mathbb{R}^{p_1}$ and $\mathcal{B}_{0j} \subseteq \mathcal{L}_2([0, \tau])$. Define $\mathcal{B} = \prod_{j=1}^{p_2} \mathcal{B}_{0j}$ and $\Theta = \mathcal{A} \times \mathcal{B}$. Note that due to the dimension of $\mathcal{B}$, the estimation may not be easy and to deal with this, by following others, we propose to employ the sieve approach to first approximate $\beta(t)$ by B-spline functions. Specifically, let $\mathcal{T} = \{t_j\}_{j=1}^{m_n+2l}$ with $0 = t_1 = \cdots = t_l < t_{l+1} < \cdots < t_{m_n+l} < t_{m_n+l+1} = \cdots = t_{m_n+2l} = \tau$ being a sequence of knots that partition $[0, \tau]$ into $K_n + 1$ subintervals $[t_{l+j}, t_{l+j+1}]$ for $j = 0, \ldots, K_n$ with $K_n = O(n^{\nu})$ and $\max_{0<j<m_n} |t_{j+1} - t_j| = O(n^{-\nu})$ for $\nu \in (0, 0.5)$. Define

$$\mathcal{B}_{nj} = \left\{ \beta_{nj}(t) = \alpha_{j0} + \sum_{k=1}^{q_n} \alpha_{jk} B_k(t) = \mathbf{B}_n^T(t)\,\alpha_j, \|\alpha_j\|_1 < M_n \right\},$$

the class of B-splines of order l with the knots sequence $\mathcal{T}$. In the above, M_n is some large number with $M_n \to \infty$ as $n \to \infty$, $q_n = K_n + l$, $\mathbf{B}_n(t) = \{1, B_1(t), \ldots, B_k(t)\}^T$ is a class of B-spline basis, and $\alpha_{nj} = (\alpha_{nj0}, \alpha_{nj1}, \ldots \alpha_{njq_n})$. Then $\beta_j(t)$ can be approximated by $\beta_{nj}(t)$. Define $\mathcal{B}_n = \prod_{j=1}^{p_2} \mathcal{B}_{nj}$, then $\Theta_n = \mathcal{A} \times \mathcal{B}_n$ is a sieve space for the original parameter space Θ.

Under the sieve space Θ_n, by replacing $\beta(t)$ by $\beta_n(t)$, model (4.1) can be rewritten as

$$\begin{aligned} E\left[dN(t)\,|W_i(t), X_i(t)\right] &= \exp\left(\gamma^T W_i(t) + \beta^T(t) Z_i(t)\right) d\Lambda(t) \\ &= \exp\left\{\gamma^T W_i(t) + \alpha_n^T \tilde{Z}_i(t)\right\} d\Lambda(t) \\ &= \exp\left\{\theta^T X_i(t)\right\} d\Lambda(t). \end{aligned} \tag{4.2}$$

Here,

$$\tilde{Z}_i(t) = (Z_{i1}(t)\mathbf{B}_n^T(t), Z_{i2}(t)\mathbf{B}_n^T(t), ..., Z_{ip_2}(t)\mathbf{B}_n^T(t))^T,$$

$\alpha_n = (\alpha_{n1}^T, \alpha_{n2}^T, ..., \alpha_{np_2}^T)^T$, $X_i(t) = \left(W_i^T(t), \tilde{Z}_i^T(t)\right)^T$ and $\theta = (\gamma, \alpha)^T$. Note that this model does not include time-varying coefficients now.

4.2.2 Likelihood Function

For subject i, $i = 1, \ldots, n$, the underlying counting process is denoted by

$$N_i(t) = \sum_{k=1}^{\infty} I(R_{ik} \leq t),$$

where $\{R_{ik}\}_{k=1}^{\infty}$ are event occurrence times before the censoring time C_i. However, one cannot observe $\{R_{ik}\}_{k=1}^{\infty}$ directly and can only observe $N_i(t)$ at $0 < T_{i1} < \ldots < T_{iJ_i} = C_i$. Let $X_i(t)$ denote the time-dependent covariates mentioned before for subject i. The rate function $\lambda_i(t)$ of $N_i(t)$ conditional on $X_i(\cdot)$ follows

$$\lambda_i(t)\, dt = \exp\left(\theta^T X_i(t)\right) d\Lambda(t),$$

where $\Lambda(t)$ is an unknown non-decreasing function.

Let $\Delta N_{ij} = N_i(T_{ij}) - N_i\left(T_{i(j-1)}\right)$ denote the total number of event occurrence between T_{ij} and $T_{i(j-1)}$ for subject i with $T_{i0} = 0$. Let $\Delta_i(t) = I(C_i > t)$ denote the censoring indicator. Let $0 = t_0 < t_1 < \ldots < t_K$ denotes the ordered unique values of T_{ij} for all i and j. We assume that $N_i(t)$ follows a Poisson process given covariates $X_i(t)$. Then the likelihood function is

$$L(\theta,\Lambda)=\prod_{i=1}^{n}\prod_{j=1}^{J_i}\frac{1}{\Delta N_{ij}!}\exp\left(-\int_{T_{i(j-1)}}^{T_{ij}}\exp\left(\theta^T X_i(t)\right)d\Lambda(t)\right)$$
$$\times\left\{\int_{T_{i(j-1)}}^{T_{ij}}\exp\left(\theta^T X_i(t)\right)d\Lambda(t)\right\}^{\Delta N_{ij}}. \quad (4.3)$$

In (4.3), the integral $\int_{T_{i(j-1)}}^{T_{ij}}\exp\left(\theta^T X_i(t)\right)d\Lambda(t)$ cannot be calculated easily because we do not know the exact form of $\Lambda(t)$. Many numerical integration methods, such as Gaussian quadrature, require a certain degree of smoothness of integrands. However, in practice, the form of $X_i(t)$ could be very rough leading to inaccuracy of numerical integration results. These difficulties in calculation complicates the maximization of (4.3).

To make the integration in (4.3) tractable, we assume the estimator of Λ to be a step function with nonnegative jump size λ_k at t_k with $\lambda_0 = 0$. Hence, we maximize

$$L(\theta,\Lambda)=\prod_{i=1}^{n}\prod_{j=1}^{J_i}\frac{1}{\Delta N_{ij}!}\exp\left(-\sum_{T_{i(j-1)}<t_k\le T_{ij}}\exp\left(\theta^T X_i(t_k)\right)\lambda_k\right)$$
$$\left\{\sum_{T_{i(j-1)}<t_k\le T_{ij}}\exp\left(\theta^T X_i(t_k)\right)\lambda_k\right\}^{\Delta N_{ij}} \quad (4.4)$$
$$=\prod_{i=1}^{n}\prod_{j=1}^{J_i}\Pr\left(\sum_{T_{i(j-1)}<t_k\le T_{ij}}W_{ik}=\Delta N_{ij}\right). \quad (4.5)$$

In the above, W_{ik}'s are independent Poisson random variables with means $\exp(\theta^T X_i(t_k))\lambda_k$ for all i and k. The maximization of (4.5) is still not straightforward but

(4.5) can be regarded as a likelihood function based on W_{ik} given that one can only observe $\sum_{T_{i(j-1)}<t_k\leq T_{ij}} W_{ik} = \Delta N_{ij}$ for all i, j, k. Hence, we can develop the following EM algorithm to maximize (4.5) by treating W_{ik} as missing data and $\sum_{T_{i(j-1)}<t_k\leq T_{ij}} W_{ik}$ as observed data.

4.2.3 The EM Algorithm

Let $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_K)$. The complete-data likelihood based on W_{ik} can be written as

$$L_C(\theta, \lambda) \propto \prod_{i=1}^{n} \prod_{k=1}^{K} \exp\left(-\exp\left(\theta^T X_i(t_k)\right)\lambda_k\right) \left\{\exp\left(\theta^T X_i(t_k)\right)\lambda_k\right\}^{W_{ik} I\left(t_k \leq T_{iJ_i}\right)},$$

and the complete-data log-likelihood is then

$$l_C(\theta, \lambda) = \sum_{i=1}^{n} \sum_{k=1}^{K} I\left(t_k \leq T_{iJ_i}\right) \left\{W_{ik}\left(\theta^T X_i(t_k) + \log(\lambda_k)\right) - \exp\left(\theta^T X_i(t_k)\right)\lambda_k\right\}. \tag{4.6}$$

In the E-step, we need to find the posterior expectation of W_{ik} given the observed data and the last iteration of estimators for θ and λ. Let j_{ik} be the observation time index for subject i such that $T_{i(j-1)} < t_k \leq T_{ij}$ for any k and $A_{ik} = \left\{l : T_{i(j_{ik}-1)} < t_l \leq T_{ij_{ik}}\right\}$. Denote the qth iteration of θ and λ as $\theta^{(q)}$ and $\lambda^{(q)} = \left(\lambda_1^{(q)}, \ldots, \lambda_K^{(q)}\right)$. In the qth iteration, the posterior mean of W_{ik} is

$$\begin{aligned}\hat{E}\left[W_{ik}|\theta^{(q)},\lambda^{(q)}\right] &= E\left[W_{ik}|\sum_{T_{i(j_{ik}-1)}<t_k\leq T_{ij_{ik}}} W_{ik}=\Delta N_{ij_{ik}},\theta^{(q)},\lambda^{(q)}\right] \\ &= \frac{\exp\left(\theta^{(q)\top}X_i\left(t_k\right)\right)\lambda_k^{(q)}}{\sum_{l\in A_{ik}}\exp\left(\theta^{(q)\top}X_i\left(t_k\right)\right)\lambda_l^{(q)}}\Delta N_{ij_{ik}}.\end{aligned}$$

Clearly, $\hat{E}\left[W_{ik}|\theta^{(q)},\lambda^{(q)}\right]$ can be interpreted as a weighted mean of $\triangle N_{ijik}$.

In the M-step, we update λ by

$$\lambda_k^{(q+1)} = \frac{\sum_{i=1}^{n}\Delta_{ik}\hat{E}\left[W_{ik}|\theta^{(q)},\lambda^{(q)}\right]}{\sum_{i=1}^{n}\Delta_{ik}\exp\left(\theta^{(q)\top}X_i\left(t_k\right)\right)}, \tag{4.7}$$

for $k=1,\ldots,K$. Plugging in (4.7) into the posterior expectation of (4.6) given the observed data, we update θ by solving the following equation

$$U^{(q)}\left(\theta\right) = \sum_{i=1}^{n}\sum_{k=1}^{K}\Delta_{ik}\hat{E}\left[W_{ik}|\theta^{(q)},\lambda^{(q)}\right]\left\{X_i\left(t_k\right)-\frac{S^{(1)}\left(t_k,\theta\right)}{S^{(0)}\left(t_k,\theta\right)}\right\}=0.$$

Here,

$$S^{(u)}\left(t,\theta\right)=\sum_{i=1}^{n}X_l\left(t\right)^{\otimes u}I\left(C_i>t\right)\exp\left(\theta^{\top}X_i\left(t\right)\right),$$

where $a^{\otimes 0}=1$, $a^{\otimes 1}=a$ and $a^{\otimes 2}=aa^{\top}$ for a vector a. To ease computational burden, we can update θ by one-step Newton-Raphson at each M-step as

$$\theta^{(q+1)}=\theta^{(q)}-\left[\dot{U}^{(q)}\left(\theta^{(q)}\right)\right]^{-1}\left[U^{(q)}\left(\theta^{(q)}\right)\right],$$

where

$$\dot{U}^{(q)}\left(\theta^{(q)}\right)=-\sum_{i=1}^{n}\sum_{k=1}^{K}\Delta_{ik}\hat{E}\left[W_{ik}|\theta^{(q)},\lambda^{(q)}\right]\left\{\frac{S^{(2)}\left(t_k,\theta\right)}{S^{(0)}\left(t_k,\theta\right)}-\left[\frac{S^{(1)}\left(t_k,\theta\right)}{S^{(0)}\left(t_k,\theta\right)}\right]^{\otimes 2}\right\}.$$

We then iterate between the E- and M-steps until a stopping criterion is met. One example criterion is that the sum of the relative differences of two consecutive iterations of estimates is smaller than 10^{-4}. The resulting estimators are denoted by $\left(\hat{\theta}_n,\hat{\Lambda}\right)$. The proposed EM algorithm enjoys some advantages. First, we can avoid the intractable integration in (4.3). This fills the gap that model (1.2) has not been widely used and investigated. Second, we only need the values of $X_i\left(t\right)$ at $t_1,\ldots,t_K$ instead of the whole trajectories of covariates $X_i\left(t\right)$ over time which are usually unavailable in practice. Finally, as pointed by Zeng et al. (2017), we can avoid inversion of high-dimensional matrices since we can update the high-dimensional parameters λ by one explicit formula.

4.2.4 Estimation of Variance of $\hat{\theta}_n$

To perform hypothesis testing and construct confidence intervals of $\hat{\theta}_n$, we also need to estimate the variance of $\hat{\theta}_n$. Zeng et al. (2017) proposed to use Corollary 3 in Murphy and Vaart (2000) to estimate the variance of $\hat{\theta}_n$. By following them, we define the profile log-likelihood

$$\mathrm{pl}\left(\theta\right)=\max_{\Lambda\in\mathcal{C}}\log L_n\left(\theta,\Lambda\right),$$

where $\mathcal{C}$ is the set of step functions with nonnegative jumps at t_k. Then, the covariance matrix of $\hat{\theta}_n$ is estimated by the negative inverse of a matrix whose (j,k)th element is

$$\frac{\mathrm{pl}_n\left(\hat{\theta}_n\right) - \mathrm{pl}_n\left(\hat{\theta}_n + h_n e_k\right) - \mathrm{pl}_n\left(\hat{\theta}_n + h_n e_j\right) + \mathrm{pl}_n\left(\hat{\theta}_n + h_n e_k + h_n e_j\right)}{h_n^2}. \tag{4.8}$$

where e_j is the jth canonical vector in $\mathbb{R}^p$ and h_n is a constant of order $n^{-1/2}$. This is to estimate the Hessian matrix of $\mathrm{pl}_n(\theta)$ at $\hat{\theta}_n$. The profile log-likelihood $\mathrm{pl}_n(\theta)$ can calculated by reusing the EM algorithm to update λ with θ held fixed.

However, Zeng et al. (2017) argued that the estimated Hessian matrix of $\mathrm{pl}_n(\theta)$ at $\hat{\theta}_n$ may be negative definite, especially in small samples. They then proposed to estimate the covariance matrix of $\hat{\theta}_n$ by $\left(n\hat{V}\right)^{-1}$ with

$$\hat{V}_n = n^{-1}\sum_{i=1}^{n}\left[\left\{\frac{\partial}{\partial\theta} l_i\left(\theta, \hat{\mathcal{A}}_\theta\right)\Big|_{\theta=\hat{\theta}_n}\right\}^{\otimes 2}\right],$$

where $\hat{\mathcal{A}}_\theta = \arg\max_{\mathcal{A}} \log L_n(\theta, \mathcal{A})$ for $\theta \in \Theta$ and $l_i(\theta, \mathcal{A})$ is the log-likelihood function for the ith subject. Here, $n\hat{V}$ is actually the empirical covariance matrix of the gradient of $l_i\left(\theta, \hat{\mathcal{A}}_{\theta_n}\right)$. This gradient is approximated by a first-order numerical difference of $l_i\left(\theta, \hat{\mathcal{A}}_\theta\right)$ as

$$\frac{l_i\left(\theta, \hat{\mathcal{A}}_\theta\right)\Big|_{\theta=\hat{\theta}_n + h_n e_k} - l_i\left(\theta, \hat{\mathcal{A}}_\theta\right)\Big|_{\theta=\hat{\theta}_n}}{h_n}, \tag{4.9}$$

for the kth component of the gradient. In this study, we found the Hessian matrix of $\mathrm{pl}_n(\theta)$ at $\hat{\theta}_n$ calculated by the first method through (4.8) is sometimes negative-definite in the simulation study. Thus we adopt (4.9) to calculate the variance estimation.

4.3 Asymptotic Properties

Now we will establish the asymptotic properties of the estimators proposed in the previous section, including the consistency, convergence rate and asymptotic normality. For this, let $\boldsymbol{\vartheta} = (\gamma, \beta, \Lambda)$ and $\boldsymbol{\vartheta}_0 = (\gamma_0, \beta_0, \Lambda_0)$ denote the true value of $\boldsymbol{\vartheta}$. We denote the estimator for $\boldsymbol{\vartheta}$ to be $\hat{\boldsymbol{\vartheta}}_n = \left(\hat{\gamma}, \hat{\beta}_n, \hat{\Lambda}\right)$. Also for convenience, let $V(t) = \left(W^T(t), Z^T(t)\right)^T$ and redefine the parameter space $\Theta = \mathcal{A} \times \mathcal{B} \times \mathcal{F}$, where $\mathcal{F}$ denotes the parameter space of Λ.

Let $\mathfrak{B}^d$ denote the collection of Borel sets in $\mathbb{R}^d$ and $\mathfrak{L}_2[0,\tau]$ the collection of Borel sets in $\mathcal{L}_2$ on $[0,\tau]$. Define $\mathfrak{B}^1[0,\tau] = \{B \cap [0,\tau] : B \in \mathfrak{B}^1\}$, $\mathfrak{B}^d = \mathfrak{B}^1[0,\tau] \times \ldots \times \mathfrak{B}^1[0,\tau]$ and $\mathfrak{L}_2^d[0,\tau] = \mathfrak{L}^1[0,\tau] \times \ldots \times \mathfrak{L}^1[0,\tau]$. Also define the measure

$$\upsilon_1(B_1 \times B_2 \times B_3) = \int_{B_3}\int_{B_2}\int_{B_1} dE\left[Y(t)H(t)\right] d\mu_Z d\mu_W$$

for $B_1 \in \mathfrak{B}^1[0,\tau]$, $B_2 \in \mathfrak{L}_2^{p_2}[0,\tau]$ and $B_3 \in \mathfrak{L}_2^{p_1}[0,\tau]$, where μ_W and μ_Z are the measures for W and Z; and $\mu_1(B_1 \times B_2) = \upsilon_1(B_1 \times B_2 \times \mathfrak{L}_2^{p_2}[0,\tau])$. Alternatively, let $\mu_V = \mu_Z \times \mu_W$, we can rewrite $\upsilon_1(B_1 \times B_2 \times B_3)$ as

$$\upsilon_1(B_1 \times B_4) = \int_{B_4}\int_{B_1} dE\left[Y(t)H(t)\right] d\mu_V$$

for $B_1 \in \mathfrak{B}^1[0,\tau]$ and $B_4 \in \mathfrak{L}_2^p[0,\tau]$; and $\mu_1(B) = \upsilon_1(B \times \mathfrak{L}_2^p[0,\tau])$. Define the L_2 metric $d(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2)$ on Θ as

$$d(\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2) = \left(\|\gamma_1 - \gamma_2\|_2^2 + \|\beta_1 - \beta_2\|_{L(\tilde{\mu}_0)}^2 + \|\Lambda_1 - \Lambda_2\|_{L(\mu_0)}^2\right)^{1/2}.$$

We need the following conditions for prove the consistency and convergence rate of $\hat{\boldsymbol{\vartheta}}_n$ as well as the asymptotic normality of $\hat{\beta}$.

(C1) The true value of $\beta(t)$, denoted by $\beta_0(t)$, lies in the interior of the compact set $\mathcal{B}$ such that $\beta_j(t) \in \mathcal{B}$ is rth differentiable. The true value of $\Lambda \in \mathcal{F}$, denoted by Λ_0, is a continuously differentiable with positive and bounded derivative Λ_0^T in some interval $O[T] = [\xi, \tau]$ where $\Pr\left(\cap_{j=1}^{J}\{T_{\cdot j} \in [\xi, \tau]\}\right) = 1$.

(C2) The covariates $W(t)$ and $Z(t)$ are uniformly bounded with uniformly bounded total variation over $[\xi, \tau]$, and its left limit exists for any t.

(C3) If there exists some vector β such that $\gamma^\top(t) W(t) + \beta^\top(t) Z(t) = a(t)$ for some deterministic function $a(t)$ on $t \in [\xi, \tau]$ with probability 1, then $\beta(t) = 0$ and $a(t) = 0$.

(C4) The number of observation times, J, is positive and finite with probability one. In addition, $\nu_3(\{\tau\}) > 0$.

(C5) The observation times are s_0-separated, that is, there exist a constant $s_0 > 0$, such that $\Pr\left(T_{\cdot j} - T_{\cdot (j-1)} \geq s_0 | J, X\right) = 1$ for all $j = 1, \ldots J$. In addition, μ_1 is absolutely continuous with respect to Lebesgue measure on $[0, \tau]^2$.

(C6) The function $m_0 = \sum_{j=1}^{J} \Delta N_{\cdot j} \log\{\Delta N_{\cdot j}\}$ satisfies $\mathbf{P} m_0 < 0$.

(C7) For some $\eta \in (0, 1)$,

$$a^\top \mathrm{Cov}\left(V(t), V(s) | U_1, U_2\right) a \geq \eta a^\top E\left[V(t) V^\top(s) | U_1, U_2\right] a$$

a.s. for all $a \in \mathbb{R}^p$ and $t, s \in [0, \tau]$ where (U_1, U_2, V) has distribution $\nu_1/\nu_1(\mathbb{R}^+ \times \mathbb{R}^+ \times \boldsymbol{\mathcal{X}})$.

(C8) For some $c_0 \in (0, \infty)$, the function $V \to E\left[\exp\left(c_0 N\left(\tau\right)\right)|V\right]$ uniformly bounded for $V \in \mathcal{X}$.

We establish the weak consistency of $\hat{\boldsymbol{\vartheta}}_n$ by the next theorem.

Theorem 7 (Consistency)**.** *Under regularity conditions (C1)-(C6),* $d\left(\hat{\boldsymbol{\vartheta}}_n, \boldsymbol{\vartheta}_0\right) \to 0$ *in probability as* $n \to \infty$.

One concern in construction of the proposed EM algorithm is that we impose the Poisson assumption on the underlying counting process which may be often violated in practice. However, the proof of Theorem 7 does not actually require the Poisson assumption. This implies that we can obtain consistent estimators from the EM algorithm as long as (1.2) holds, indicating the generality of the proposed method.

Theorem 8 (Rate of Convergence)**.** *Assume that the regularity conditions (C1)-(C8) given above hold. Then we have that*

$$n^{\min\left\{n^{\frac{1-\nu}{3}}, n^{r\nu}\right\}} d\left(\hat{\boldsymbol{\vartheta}}_n, \boldsymbol{\vartheta}_0\right) = O_p\left(1\right)$$

with the optimal rate $O_p(n^{-r/(3r+1)})$ *achieved at* $\nu = 1/\left(1+3r\right)$.

Next theorem shows the asymptotic normality of $\hat{\boldsymbol{\vartheta}}$.

Theorem 9 (Asymptotic Normality)**.** *if* $1/\left(4r\right) < \nu < 1/4$, *under regularity conditions (C1)-(C8), we have*

$$\sqrt{n}\left(\hat{\gamma} - \gamma_0\right)^T h_1 + \sqrt{n}\int_0^\tau \left(\hat{\beta}_n\left(t\right)\beta_0\left(t\right)\right)^T dh_2\left(t\right) + \sqrt{n}\int_0^\tau \left(\hat{\Lambda}\left(t\right) - \Lambda_0\left(t\right)\right) dh_3\left(t\right)$$

$$\to_d N\left(0, \sigma^2\right).$$

where σ^2 is given in the appendix.

The proofs of Theorem 7 - 9 are provided in the appendix.

4.4 A Simulation Study

To investigate the finite sample properties of the proposed method, we carry out a simulation study. We consider the underlying recurrent process follows a non-homogeneous Poisson process with the rate function

$$\lambda(t)\exp(\gamma_1 W_1 + \gamma_2 W_2 + \beta_1 Z_1 + \beta_2 Z_2)\, d\Lambda(t).$$

We set the maximum follow-up time $\tau = 2$. We generate the piecewise time-dependent covariate $W_1(t)$ by

$$W_1(t) = W_{11} I(t \leq U_1) + W_{12} I(t \geq U_1),$$

where $W_{1j} \sim Unif(0,1)$, $U_1 \sim Unif(0,\tau)$ and generate Z_1 in the similar way. We consider time-independent covariates for W_2 and Z_2, where $W_2 \sim Unif(0,1)$ and $Z_2 \sim Bern(0.5)$.

We set $\gamma_1 = 1$, $\gamma_2 = -1$ and consider the following two scenarios for $\beta_1(t)$ and $\beta_2(t)$:

Scenario 1 $\beta_1(t) = t$ and $\beta_2(t) = t^2/2$;

Scenario 2 $\beta_1(t) = (\sin(4\pi t) + 4\pi t)/6$ and $\beta_2(t) = (\cos(4\pi t) + 4\pi t)/6$.

We consider two cases of λ as $\lambda(t) = 4/(1+t)$ or $\lambda(t) = 4\log(t+1)$. We assume that the censoring time $C \sim Unif(0.9\tau, \tau)$. The total number of observation is generated from a zero-truncated Poisson distribution with mean 4. Next, the observation times are generated as sorted uniform random variables on $(0, C)$. Simulated observation times are then rounded to two decimals so that they could be tied. We set the sample size $n = 200$ and 400, and simulate 1000 replicates for each sample size. To estimate the variance of $\hat{\beta}$, we use $h_n = 5n^{-1/2}$ as recommended by Zeng et al. (2017). We set the maximum number of iterations to 10000 and the convergence threshold to 10^{-5}. The nominal level of the confidence intervals is 95%. In practice, we use a cubic nature spline, because it is visually smooth, and the shape beyond the two end knots is constrained to be linear, precluding any erratic tail behavior. We set the number of knots as 3 and the degree as 3.

Table 4.1 shows the summary statistics of the Monte Carlo simulation results for the two scenarios. In Table 4.1, the biases of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are all very small. The standard errors of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ decreases when n increases and the estimators of the standard errors are closed to the true standard errors. The coverage probabilities of the confidence intervals constructed from the variance estimators are close to the nominal level 95%. The biases and standard errors of $\hat{\gamma}_1$ are larger than those of $\hat{\gamma}_2$. One possible reason is that the change in $W_1(t)$ induces more variation to the estimation of $\hat{\gamma}_1$.

Figure 4.1 summarizes the simulation results under Scenario 1, the estimation of $\hat{\beta}_1(t) = t$ and $\hat{\beta}_2(t) = t^2/2$ under $\lambda_0(t) = 4/(1+t)$. For comparison, the true curves are presented too and the results indicate that the proposed procedure seems to yield unbiased estimates again. It also shows the average of the estimated pointwise standard errors along with the point-wise sample standard errors and indicates that the proposed method appears to give reasonable variance estimates.

Furthermore, Figure 4.2 shows the simulation results under scenario 2, the estimation of $\beta_1(t) = (\sin(4\pi t) + 4\pi t)/6$ and $\beta_2(t) = (\cos(4\pi t) + 4\pi t)/6$ when $\lambda(t) = 4/(1+t)$ and indicates that the proposed method appears to give similar reasonable estimation and vaiance estimates.

4.5 An Application

The Young Women's Project (YWP) is a longitudinal, observational study conducted during 1999-2009 to study factors of sexually transmitted pathogens such as Chlamydia trachomatis, Neisseria gonorrhoeae, and Trichomonas vaginalis. A convenience sample of young women aged 14 to 17 were identified by clinic schedule and those who were eligible and agreed to participate were enrolled at the current or subsequent clinical visit. The eligibility was irrelevant to sexual experience, and the maximum follow-up length was 8.2 years. Baseline demographic, family and sexual behavioral information were collected at enrollment. At each quarterly follow-up visit, participants completed face-to-face interviews and the summaries of sexual behaviors during the previous 3 months were recorded. Cervical and vaginal specimens were collected by a research nurse practitioner for testing sexually transmitted infection (STI), which was treated by antibiotic medication once detected. Besides, the exact STI onset times were interval-censored between the quarterly visits and panel count data arose. Repeated infections may occur before a woman was re-interviewed; however, STI was rarely noticed or self-treated for the minor symptoms were almost unnoticeable. Therefore, in this research, STI detected at each visit will be treated as a new infection since last tested.

We focus on 271 young women who provided at least one complete quarterly interview, and our primary objective is to study the association between STI and possible impacts from social, family relationships and risky behaviors represented by use of marijuana. Let $N_i(t)$ represents the cumulative infection number of Chlamydia trachomatis (CT), Neisseria gonorrhoeae (NG), or Trichomonas vaginalis (TV) at time t based on model 4.1. Our covariates of interest consist of number of sexual partners in the last quarter (NSP), living with parents (1) or not (0) (LWP), use of marijuana (THC) and having or not received sex education from parents (EDP) (1 for yes, 0 for no).

We assume that all four covariates have time-varying coefficients. We implement the proposed methods to do analysis for all three responses. Figure 4.3, Figure 4.4, and Figure 4.5 present the corresponding results for the cumulative infection number of Chlamydia trachomatis (CT), Neisseria gonorrhoeae (NG), or Trichomonas vaginalis (TV) with 2 interior knots. Our analytical results show that for CT, NSP appears to have significant positive effect in the early stage; in contrast, LWP and EDP have significant negative effects in the early stage. However, THC do not have significant effect over the whole time period. As for NG, NSP and THC are identified to have significant positive effects in the early and late time periods, respectively. One comment here is that the relatively large coefficient effects of THC under NG may come from the sparsity of observation points in the end of study. Last, for TV, only NSP is identified to have significant positive effects. This suggests that for young women with larger number of sexual partners, the risk of infection of TV increases than those with a lower number of NSP.

4.6 Discussions

In this chapter, we proposed a likelihood-based method based on the semi-parametric rate model with time-dependent covariates and time-varying effects. The advantage of this model avoids the unrealistic monotonicity of the mean model. The B-spline allows the coeffecients to vary with time and the EM algorithm provides an efficient way to maximize the log-likelihood function. The simulation and the application show that the proposed method works very well in practice.

		Scenario I				Scenario II			
		Bias ($\times 10^{-3}$)	SE	SDE	CP	Bias ($\times 10^{-3}$)	SE	SDE	CP
$\lambda_0(t) = 4/(1+t)$									
$n = 200$	β_1=1	11.7016	0.0823	0.0852	96.1	0.9969	0.0724	0.0729	95.0
	β_2=-1	-0.8000	0.0759	0.0788	96.2	-1.5073	0.0661	0.0672	95.2
$n = 400$	β_1=1	12.7960	0.0568	0.0576	93.9	-1.7134	0.0511	0.0489	94.1
	β_2=-1	0.0096	0.0526	0.0532	95.7	2.1053	0.0460	0.0452	94.0
$\lambda(t) = 4\log(t+1)$									
$n = 200$	β_1=1	13.4114	0.0592	0.0616	95.6	15.0466	0.0510	0.0524	94.7
	β_2=-1	-0.4376	0.0570	0.0597	95.5	-2.4870	0.0500	0.0503	95.0
$n = 400$	β_1=1	10.0640	0.0429	0.0417	93.0	14.8291	0.0367	0.0353	94.6
	β_2=-1	-1.1485	0.0403	0.0402	95.0	-2.8583	0.0357	0.0340	93.8

Table 4.1: The summary statistics of the simulation results, including the biases of the point estimators (Bias), standard errors of the point estimators (SE), empirical average of the standard deviation estimators (SDE) and the empirical coverage probabilities of the confidence intervals with the nominal level 95%.

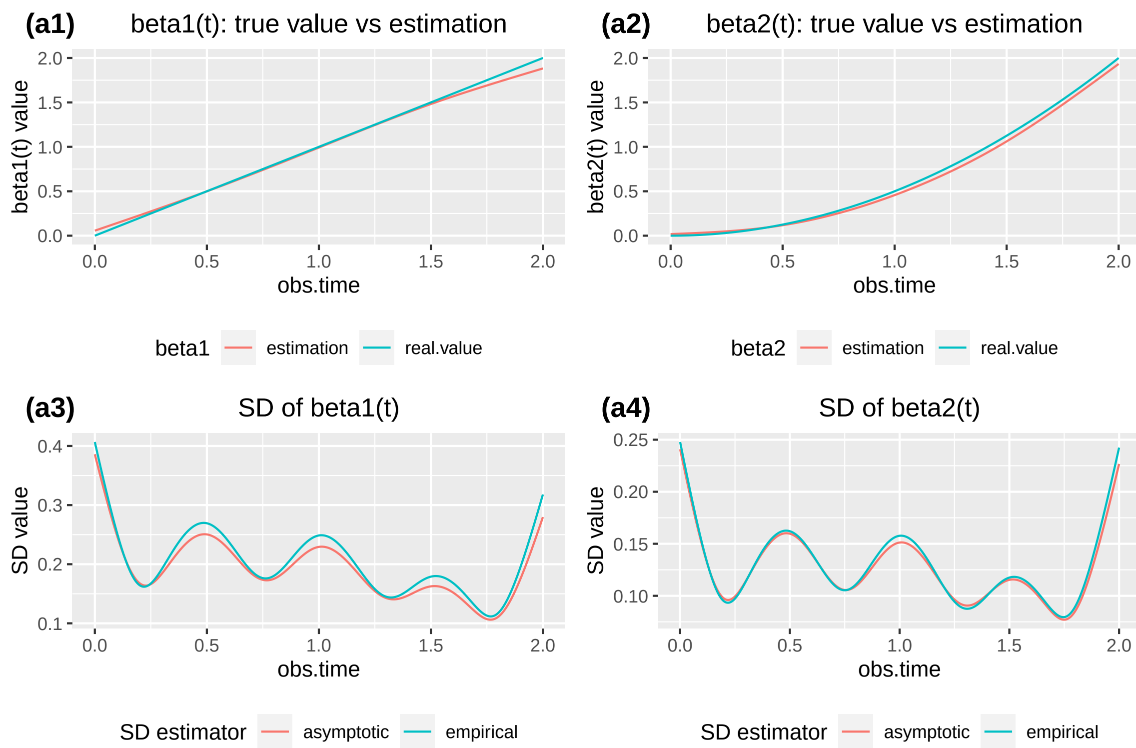

Figure 4.1: Simulation results on estimation of $\hat{\beta}_1(t)$ and $\hat{\beta}_2(t)$ with $\lambda_0(t) = 4/(1+t)$, where, $\beta_1(t) = t$, $\beta_1(t) = t^2/2$.

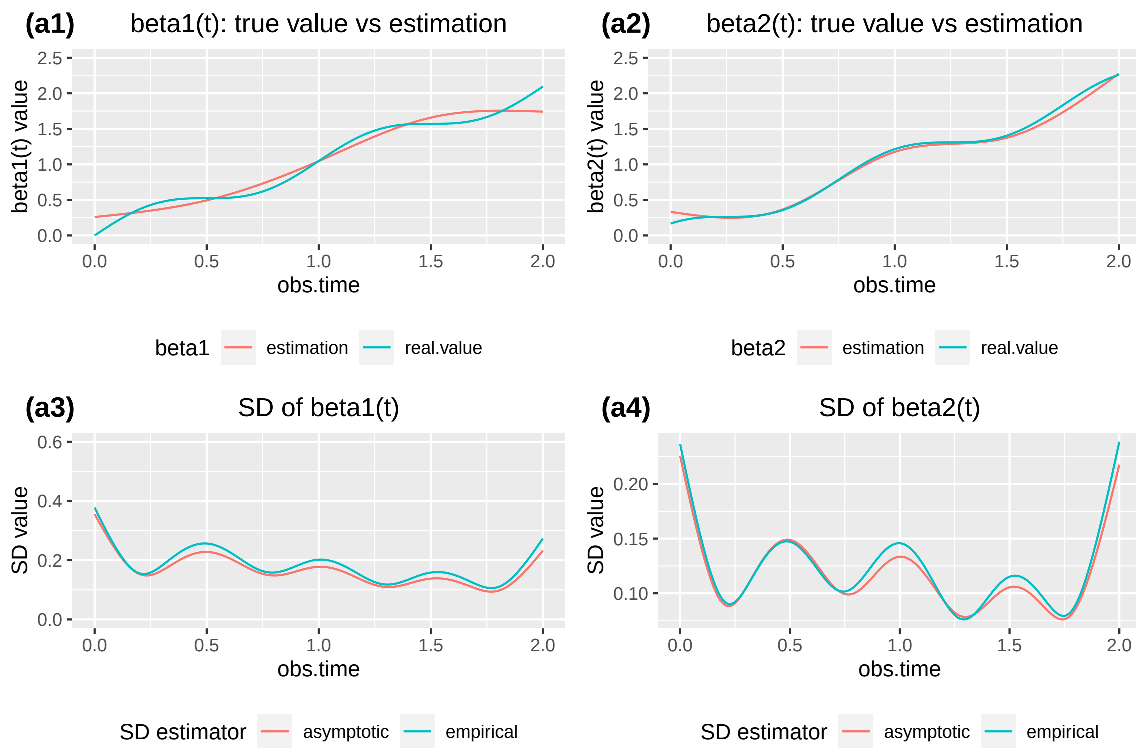

Figure 4.2: Simulation results on estimation of $\hat{\beta}_1(t)$ and $\hat{\beta}_2(t)$ with $\lambda_0(t) = 4/(1+t)$, where, $\beta_1(t) = (\sin(4\pi t) + 4\pi t)/6$ and $\beta_2(t) = (\cos(4\pi t) + 4\pi t)/6$.

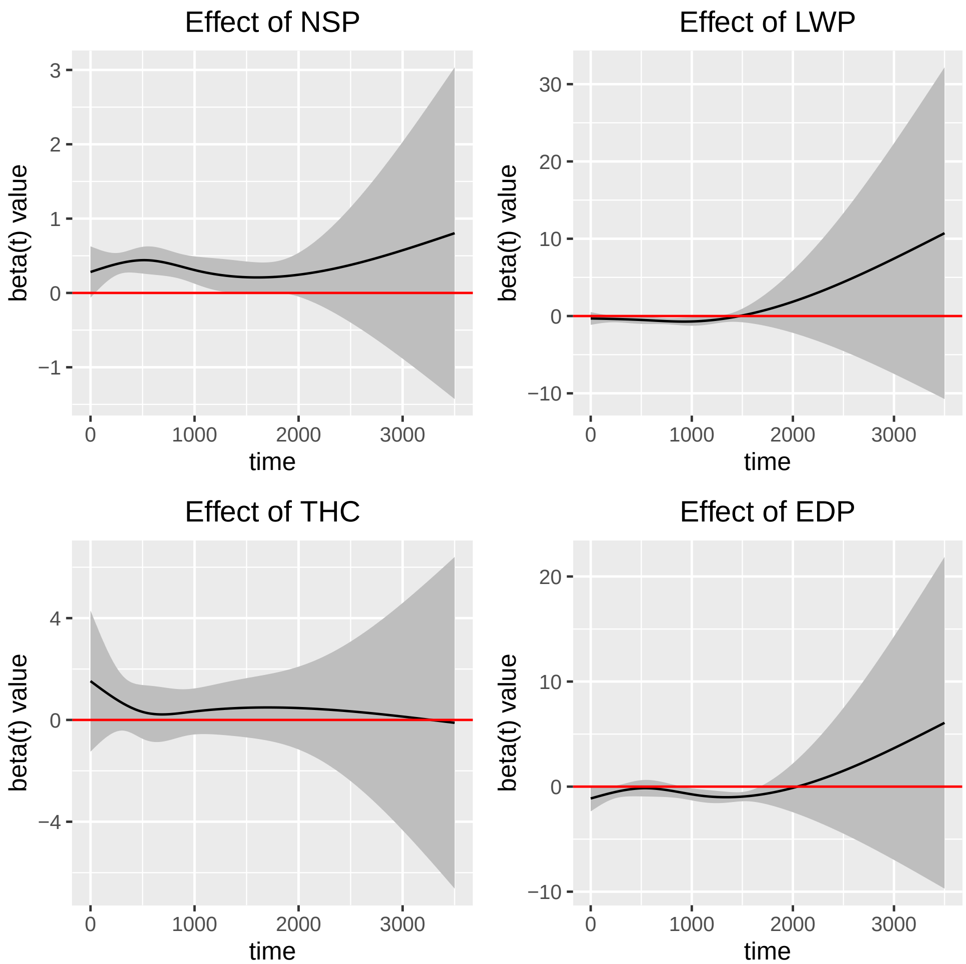

Figure 4.3: To consider all coefficients to be time-varying. Estimated time-varying effects (solid curves) and corresponding pointwise 95% confidence intervals (grey curves) for the cumulative infection number of Chlamydia Trachomatis (CT).

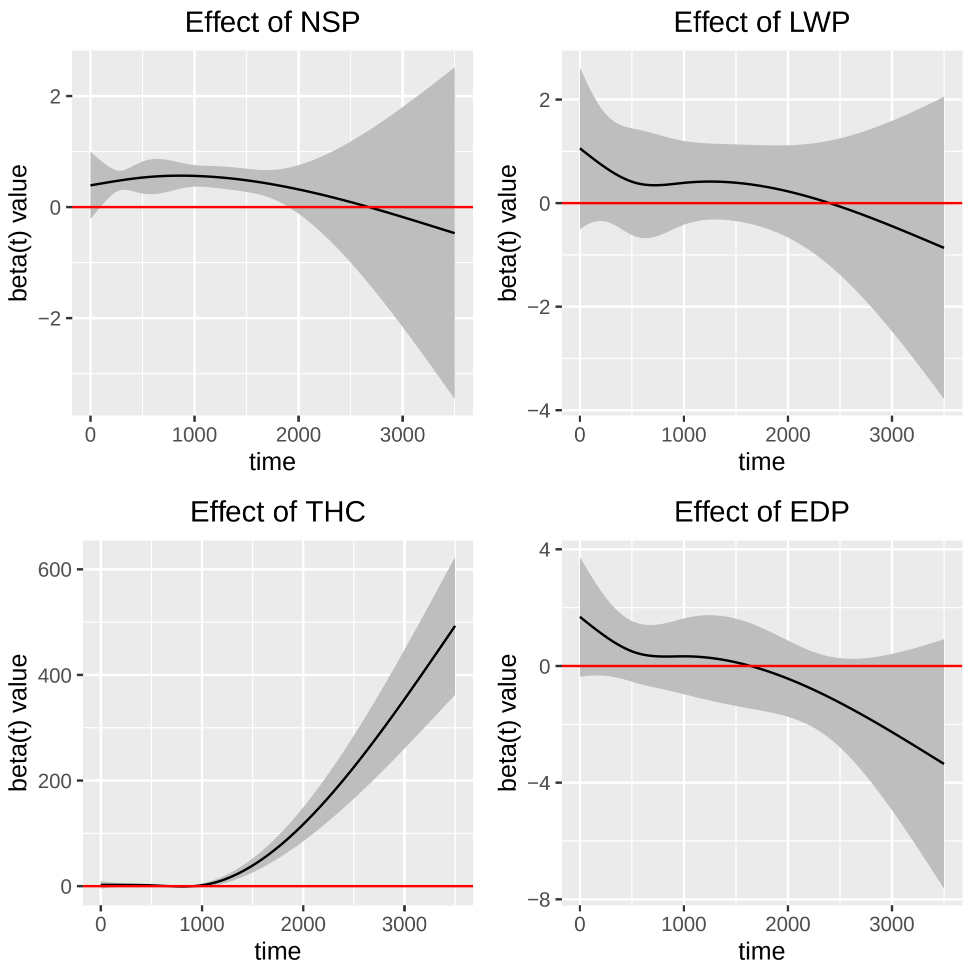

Figure 4.4: To consider all coefficients to be time-varying. Estimated time-varying effects (solid curves) and corresponding pointwise 95% confidence intervals (grey curves) for the cumulative infection number of Neisseria Gonorrhoeae (NG).

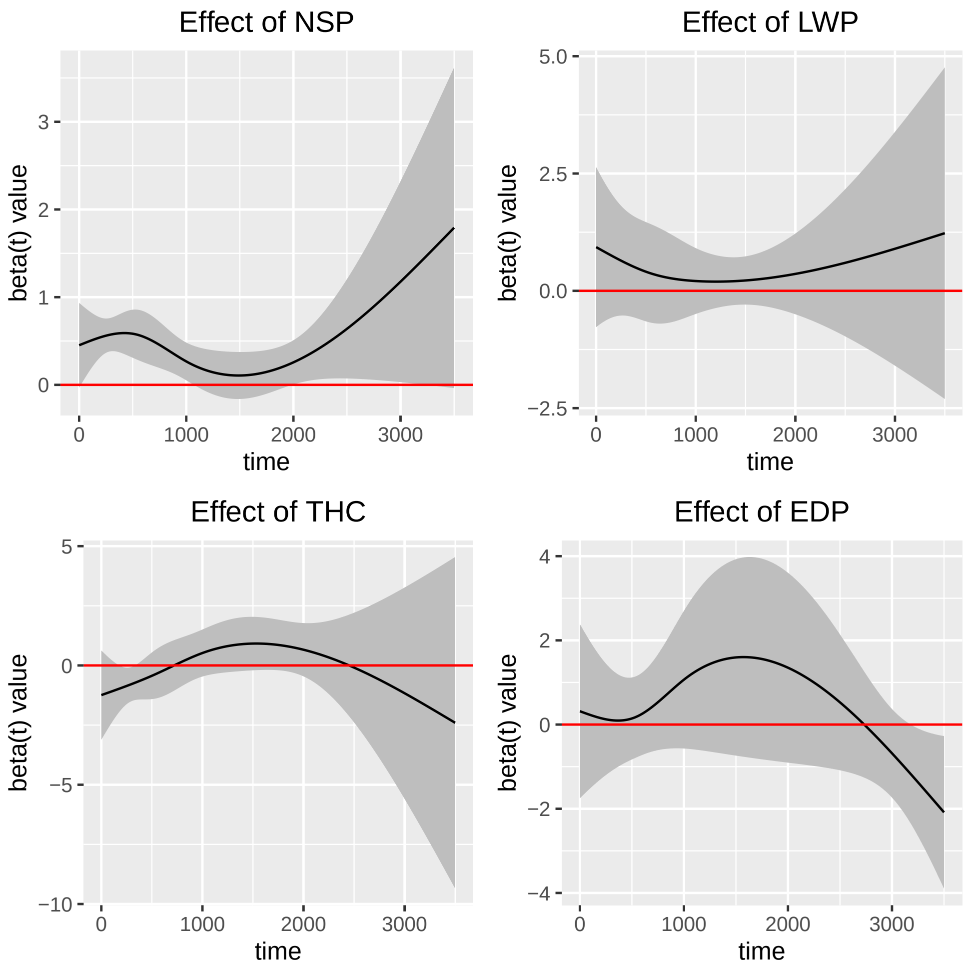

Figure 4.5: To consider all coefficients to be time-varying. Estimated time-varying effects (solid curves) and corresponding pointwise 95% confidence intervals (grey curves) for the cumulative infection number of Trichomonas Vaginalis (TV).

Chapter 5

Future Research

5.1 Dynamic Analysis of Panel Count Data with Mean Model

In chapters 2 and 3, we focused on the situation where both coefficients and covariates are varying with time for univariate and multivariate panel count data under mean model. We utilized the conditional estimating equation method with B-spline approximation.

There exist several directions for future research for univariate panel count data. One is that in the proposed method, we have assumed that the observation process does not depend on covariates and is also independent of the underlying recurrent event process of interest. In practice, it is apparent that this may not be true (Sun and Zhao (2013)). To develop a valid estimation procedure, one may need to model the three processes together. Also in the previous chapters, we have assumed that the underlying recurrent event process follows the proportional mean model, meaning that the mean functions associated with any two sets of covariate values are proportional over time. Sometimes this may be too restrictive (Lin et al. (2001)) and correspondingly, one may want to consider other models such as the class of semiparametric transformation mean models and develop some estimation procedures.

A third assumption behind the proposed estimation procedure is that one knows which covariates have time-varying effects and which covariates have constant effects. It is clear that this is generally not true and a simple approach is to try different combinations as shown in the application in Chapters 2 and 3. On the other hand, it would be useful to develop a data-driven procedure for the separation of the different types of covariates.

There also exist several other directions for future research under multivariate panel count data. One is that in the proposed methodology, a semi-parametric proportional model of the mean function of $N_k(t)$ was utilized. However, one drawback of model (3.1) is that, when β or Z fluctuates, it is hard to satisfy the non-decreasing property of the mean function. Therefore, we considered an alternative similar semi-parametric model based on the rate or intensity function of $N_k(t)$ in Chapter 4.

In addition, for Chapter 3, it is worth noting that model (3.1) is the marginal model and we took the marginal approach for the problem considered here, as the main focus of the chapter was on estimation of covariate effects. An advantage of the proposed approach, as many other marginal approaches for multivariate data, is that it leaves the correlation between different types of recurrent events arbitrary. An alternative is to directly model the correlation structure, which would be appealing if the correlation is of main interest.

Another direction is again the selection of the time-varying coefficient under the multivariate panel count data. In Chapter 3, the analysis of the CHNS data showed that locations and education levels may have time-varying effects (see Section 3.6). In such a modeling context, all of the existing methods require analysts to pre-specify each variable either as time-varying or time-invariant. Misspecification could result in biased estimation or reduced efficiency (Xie et al. (2019)). We could consider data-driven approaches for determining a covariate effect as "no effect", "time constant effect", or "time-varying effect".

5.2 Dynamic Analysis of Panel Count Data with Rate Model

In chapter 4, we discussed the semiparametric maximul likelihood method for univariate panel count data with time-varying coeffficients and time-dependent covariates. There exist several directions for future research. As discussed in chapter 3, multivariate panel count data is of interest. A likelihood function could also be constructed similarly with a shared frailty model based on model (4.1).

Besides, another assumption behind the proposed estimation procedure is that one knows which covariates have time-varying effects and which covariates have constant effects. It is clear that this is generally not true and a simple approach is to try different combinations as shown in the application above. On the other hand, it would be useful to develop a data-driven procedure for the separation of the two types of covariates. Also, under the semiparametric MLE method, we could consider some penalized methods that could select between time-independent and time-varying specifications of their presence in the model (Xie et al., 2019; Yan and Huang, 2012).

Last, note that all the above methods assumed the parametric covariate effect on the mean or rate function for panel count data, which may not be realistic in some studies. For example, in the China Health and Nutrition Survey mentioned above in Chapters 2 and 3, the education level was assessed as having a significant effect on marriage times and pregnancy numbers (see Section 3.6). Furthermore, its effect was not linear and cannot be analyzed appropriately using existing methods. Hence, it is crucial to explore the nonlinear effects of covariates (Wang and Yu, 2021a). A model with nonparametric covariate functions for panel count data may reveal the nonlinear effects of the education level at young ages of mothers, thereby facilitate our analysis. Therefore, statistical methods that can deal with nonlinear covariate effects of panel count data are desired.

Appendix A

Theoretical Proofs

A.1 Proofs of Theorems 1, 2, and 3

In this Section, we will sketch the proofs of the consistency and asymptotic properties of the proposed estimator $\hat{\boldsymbol{\vartheta}}$ described in Theorems 1, 2, and 3 of Chapter 2.

It is not straightforward to study the the asymptotic properties of $\hat{\boldsymbol{\vartheta}}_n$ based on the sieve estimating equation (2.3). We will show that solving (2.3) is equivalent to maximize

$$l_n(\boldsymbol{\vartheta}) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)(N_i(t)\log(\Lambda(t)) + N_i(t)\boldsymbol{\theta}^T(t)\mathbf{V}_i(t) - \exp\left(\boldsymbol{\theta}^T(t)\mathbf{V}_i(t)\right)\Lambda(t))dH_i(t) \quad \text{(A.1)}$$

with respect to $\boldsymbol{\vartheta}$ over $\Theta_n = \mathcal{A} \times \mathcal{M}_n \times \mathcal{F}$. With slight abuse of notation, here $\boldsymbol{\theta}^T(t)\mathbf{V}_i(t) = \boldsymbol{\gamma}^T\mathbf{W}_i(t) + \boldsymbol{\beta}^T(t)\mathbf{Z}_i(t)$ and $\boldsymbol{\theta}(t) = (\boldsymbol{\gamma}, \boldsymbol{\beta}(t))$. Let $\mathbb{P}_n$ be the empirical measure and $\mathbf{P}$ be the true measure. Let $\mathbb{M}_n(\boldsymbol{\vartheta}) = l_n(\boldsymbol{\vartheta}) = \mathbb{P}_n m_{\boldsymbol{\vartheta}}(\mathbf{V})$ and $\mathbf{M}(\boldsymbol{\vartheta}) = \mathbf{P} m_{\boldsymbol{\vartheta}}(\mathbf{V})$, where

$$\begin{aligned} m_{\boldsymbol{\vartheta}}(\mathbf{O}) = \int_0^\tau & Y(t)(N(t)\log(\Lambda(t)) + N(t)\boldsymbol{\theta}^T(t)\,\mathbf{V}(t) - \\ & \exp\left(\boldsymbol{\theta}^T(t)\,\mathbf{V}(t)\right)\Lambda(t))dH(t) \end{aligned} \quad \text{(A.2)}$$

and $\mathbf{O} = (\mathbf{V}, H, Y)$.

Let

$$\hat{\Lambda}(t, \boldsymbol{\theta}(t)) = \frac{\sum_{j=1}^n Y_j(t) N_j(t) dH_j(t)}{\sum_{j=1}^n Y_j(t)\exp\left(\boldsymbol{\theta}^T(t)\,\mathbf{V}_j(t)\right) dH_j(t)}.$$

We first show $\hat{\Lambda}(t, \boldsymbol{\theta}(t))$ maximizes ((A.1)) for a fixed $\boldsymbol{\theta}(t)$. After some algebra,

$$\begin{aligned} & \mathbb{M}_n(\boldsymbol{\theta}, \Lambda) - \mathbb{M}_n\left(\boldsymbol{\theta}, \hat{\Lambda}\right) \\ = & \frac{1}{n}\sum_{i=1}^n \int_0^\tau Y_i(t)\hat{\Lambda}(t, \boldsymbol{\theta}(t))\exp\left(\boldsymbol{\theta}^T(t)\,\mathbf{V}_i(t)\right) \\ \times & \left(\frac{N_i(t)}{\hat{\Lambda}(t, \boldsymbol{\theta}(t))\exp\left(\boldsymbol{\theta}^T(t)\,\mathbf{V}_i(t)\right)}\log\left(\frac{\Lambda(t)}{\hat{\Lambda}(t, \boldsymbol{\theta}(t))}\right) - \frac{\Lambda(t)}{\hat{\Lambda}(t, \boldsymbol{\theta}(t))} + 1\right) dH_i(t). \end{aligned}$$

Plug in $\hat{\Lambda}(t, \boldsymbol{\theta}(t))$, and by Fubini's theorem, we have

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\hat{\Lambda}(t,\boldsymbol{\theta}(t))\exp\left(\boldsymbol{\theta}^T(t)\,\mathbf{V}_i(t)\right)\left(\frac{\Lambda(t)}{\hat{\Lambda}(t,\boldsymbol{\theta}(t))}-1\right)dH_i(t)\\
=&\frac{1}{n}\int_0^{\tau}\left(1-\frac{\Lambda(t)}{\hat{\Lambda}(t,\boldsymbol{\theta}(t))}\right)\left(\sum_{j=1}^{n}Y_j(t)N_j(t)dH_j(t)\right)\\
&\times\sum_{i=1}^{n}\frac{Y_i(t)\exp\left(\boldsymbol{\theta}^T(t)\,\mathbf{V}_i(t)\right)dH_i(t)}{\sum_{j=1}^{n}Y_j(t)\exp\left(\boldsymbol{\theta}^T(t)\,\mathbf{V}_i(t)\right)dH_j(t)}\\
=&\frac{1}{n}\int_0^{\tau}\sum_{j=1}^{n}Y_j(t)N_j(t)\left(1-\frac{\Lambda(t)}{\hat{\Lambda}(t,\boldsymbol{\theta}(t))}\right)dH_j(t)\,.
\end{aligned}
$$

Therefore, since $\log(x) - x + 1 \leq -(x-1)^2$ for all positive x and the equality holds iff $x = 1$,

$$
\begin{aligned}
&\mathbb{M}_n(\boldsymbol{\theta},\Lambda)-\mathbb{M}_n\left(\boldsymbol{\theta},\hat{\Lambda}\right)\\
=&\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}Y_i(t)N_i(t)\left\{\log\left(\frac{\Lambda(t)}{\hat{\Lambda}(t,\boldsymbol{\theta}(t))}\right)-\frac{\Lambda(t)}{\hat{\Lambda}(t)}+1\right\}dH_i(t)\\
\leq&-\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}Y_i(t)N_i(t)\left(\frac{\Lambda(t)}{\hat{\Lambda}(t,\boldsymbol{\theta}(t))}-1\right)^2dH_i(t)\\
\leq&0
\end{aligned}
$$

This implies $\mathbb{M}_n(\boldsymbol{\theta},\Lambda) \leq \mathbb{M}_n\left(\boldsymbol{\theta},\hat{\Lambda}\right)$ for any $\boldsymbol{\theta}$. The equality holds iff $\Lambda(t) = \hat{\Lambda}(t,\boldsymbol{\theta}(t))$ at points where $\sum_{i=1}^{n} Y_i(t)H_i(t)$ jumps. Since $l_n(\boldsymbol{\vartheta})$ are only determined by the value of $\hat{\Lambda}(t,\boldsymbol{\theta}(t))$ at points where $\sum_{i=1}^{n} Y_i(t)H_i(t)$ jumps, $\hat{\Lambda}(t,\boldsymbol{\theta}(t))$ is the unique maximizer of $l_n(\boldsymbol{\theta},\Lambda)$ with respect to Λ. Then, to maximize $l_n\left(\boldsymbol{\theta},\hat{\Lambda}(t,\boldsymbol{\theta})\right)$ with respect to $\boldsymbol{\theta}$ over $\mathcal{A}\times\mathcal{M}_n$, by the idea of profile likelihood in Wellner and Zhang (2007), we need to maximize $l_n\left(\boldsymbol{\theta}_n,\hat{\Lambda}(t,\boldsymbol{\theta}_n)\right)$ with respect to $\boldsymbol{\theta}_n$. After some algebra,

$\partial l_n\left(\boldsymbol{\theta}_n, \hat{\Lambda}(t, \boldsymbol{\theta}_n)\right)/\partial \boldsymbol{\theta}_n$ equals the left hand side of (2.3). Obviously, $l_n\left(\boldsymbol{\theta}_n, \hat{\Lambda}(t, \boldsymbol{\theta}_n)\right)$ is convex with respect to $\boldsymbol{\theta}_n$, implying maximizing $l_n(\boldsymbol{\vartheta})$ over $\mathcal{A}\times\mathcal{M}_n\times\mathcal{F}$ is equivalent to solving (2.3).

After showing the equivalence of solving the estimating equation and maximizing $l_n(\boldsymbol{\vartheta})$, the estimator is actually an M-estimator and its (asymptotic) behavior of the estimator can be investigated though $m_\theta(\mathbf{O})$ with the empirical process theory. Moreover, (A.1) coincides the pseudo-likelihood function for panel count data proposed in Wellner and Zhang (2007) which has been extensively investigated, similar to He et al. (2017). We can then use many conclusions in existing literature to facilitate our theoretical justification.

A.1.1 Proof of Consistency

We can prove the consistency of $\hat{\boldsymbol{\theta}}_n$ by Theorem 3.1 and Remark 3.1 in Chen (2007).

We first show $\boldsymbol{\vartheta}_0$ is the unique maximizer of $\mathbf{M}(\boldsymbol{\vartheta})$. After some calculation based on the conditional expectation on $\mathbf{V}$, we have

$$\mathbf{M}(\boldsymbol{\vartheta}) = \int \exp\left(\boldsymbol{\theta}^T(t)\, v(t)\right)\Lambda(t)\left\{\log(\Lambda(t)) + \boldsymbol{\theta}^T(t)\, v(t) - 1\right\} d\upsilon_1(t, v).$$

and therefore

$$\mathbf{M}(\boldsymbol{\vartheta}_0) - \mathbf{M}(\boldsymbol{\vartheta}) = \int \Lambda(u)\exp\left(\boldsymbol{\theta}^T(u)\, v(u)\right) h\left\{\frac{\Lambda_0(u)\exp\left(\boldsymbol{\theta}_0^T(u)\, v(u)\right)}{\Lambda(u)\exp\left(\boldsymbol{\theta}^T(u)\, v(u)\right)}\right\} d\upsilon_1(t, v) \tag{A.3}$$

where $h(x) = x\log(x) - x + 1$. Note that $h(x) \geq 0$ for all $x > 0$ and the equality holds only when $x = 1$. Therefore, by similar argument in Wellner and Zhang (2007), under condition (C2) and (C8), $\mathbf{M}(\boldsymbol{\vartheta}_0) = \mathbf{M}(\boldsymbol{\vartheta})$ if and only if $\boldsymbol{\theta}(t) = \theta_0(t)$ and

$\Lambda(t) = \Lambda_0(t)$ a.e. with respect to μ_1. In this manner, $\boldsymbol{\vartheta}_0$ is the unique maximizer of $\mathbf{M}(\boldsymbol{\vartheta})$.

By the similar arguments used in Wellner and Zhang (2007), by condition (C1)-(C5), we can show that $\hat{\Lambda}(t)$ is uniformly bounded in probability for $t \in [0, \tau]$ with $\mu_1(\{\tau\}) > 0$.

By Helly-Selection Theorem and compactness of Θ_n, it follows that $\hat{\boldsymbol{\vartheta}}_n = \left(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\beta}}_n, \hat{\Lambda}\right)$ has a subsequence $\hat{\boldsymbol{\vartheta}}_{n_k}$ converging to $\boldsymbol{\vartheta}^+ = \left(\boldsymbol{\gamma}^+, \boldsymbol{\beta}^+, \Lambda^+\right)$ with $\boldsymbol{\vartheta}^+ \in \Theta$. By the compactness of Θ_n as well as the fact that $m_{\boldsymbol{\vartheta}}(O)$ is upper semicontinuous in $\boldsymbol{\vartheta}$ for almost all O. Furthermore, $m_{\boldsymbol{\vartheta}} \leq M_0 < \infty$ with $\mathbf{P}M_0(\mathbf{V}) < \infty$ by (C9). Thus, by Theorem A.1 of Wellner and Zhang (2007), we have

$$\lim_{n\to\infty} \sup \sup_{\boldsymbol{\vartheta}\in\Theta_n} (\mathbb{P}_n - \mathbf{P})\, m_{\boldsymbol{\vartheta}}(\mathbf{V}) \leq 0 \tag{A.4}$$

almost surely. By the Dominated Convergence Theorem and (C9), $M(\boldsymbol{\vartheta})$ is continuous in $\boldsymbol{\vartheta}$. By the Corollary 6.21 of Schumaker (1981), there exists a spline approximation $\beta_{n0j}(t) \in \mathcal{M}_{nj}$ to β_{0j} such that

$$\sup_{t\in[0,\tau]} |\beta_{0j}(t) - \beta_{n0j}(t)| = O\left(K_n^{-r}\right) = O\left(n^{-vr}\right) \tag{A.5}$$

for $j = 1, \ldots, p_2$. Therefore, for any $\epsilon > 0$, there exists $\boldsymbol{\beta}_0^* \in \mathcal{M}_n$ such that

$$\mathbf{M}(\boldsymbol{\vartheta}_0) - \epsilon \leq \mathbf{M}(\boldsymbol{\gamma}_0, \boldsymbol{\beta}_0^*, \Lambda_0)$$

with $\max_{j=1,\ldots,p_2} \left\|\beta_{0j}(t) - \beta_{0j}^*(t)\right\|_\infty = o(1)$. Also by the similar argument in Lu et al. (2009), we have

$$\mathbb{M}_n\left(\boldsymbol{\gamma}_0, \boldsymbol{\beta}_0^*, \Lambda_0\right) - \mathbf{M}\left(\boldsymbol{\gamma}_0, \boldsymbol{\beta}_0^*, \Lambda_0\right) = o_p\left(1\right)$$

and

$$\mathbb{M}_n\left(\boldsymbol{\gamma}_0, \boldsymbol{\beta}_0^*, \Lambda_0\right) \leq \mathbb{M}_n\left(\hat{\gamma}, \hat{\boldsymbol{\beta}}_n, \hat{\Lambda}\right).$$

Then by (A.4) and the arguments similar to those used in Lu et al. (2009), we can show that $\mathbf{M}\left(\boldsymbol{\vartheta}_0\right) = \mathbf{M}\left(\boldsymbol{\vartheta}^+\right)$, implying $\boldsymbol{\beta}^+ = \boldsymbol{\beta}_0$ and $\Lambda^+ = \Lambda_0$ a.e. with respect to μ_1. Since this holds for any convergent subsequence, we conclude that all the limits of subsequence of $\hat{\boldsymbol{\vartheta}}_{n_k}$ are $\boldsymbol{\vartheta}_0$. Therefore, due to the uniform boundedness of $\hat{\Lambda}\left(t\right)$, we obtain the weak consistency of $\hat{\boldsymbol{\vartheta}}_n$ in the metric d.

A.1.2 Proof of the Rate of Convergence

In (A.3), since $h\left(x\right) \geq \left(1/4\right)\left(x-1\right)^2$ for $0 \leq x \leq 5$, for $\boldsymbol{\theta}$ in a sufficiently small neighborhood of θ_0

$$\begin{aligned} &\mathbf{M}\left(\boldsymbol{\vartheta}_0\right) - \mathbf{M}\left(\boldsymbol{\vartheta}\right) \qquad \text{(A.6)} \\ \geq &\frac{1}{4}\int \Lambda\left(u\right)\exp\left(\boldsymbol{\theta}^T\left(u\right)v(u)\right)\left\{\frac{\Lambda_0\left(u\right)\exp\left(\boldsymbol{\theta}_0^T\left(u\right)v(u)\right)}{\Lambda\left(u\right)\exp\left(\boldsymbol{\theta}^T\left(u\right)v(u)\right)} - 1\right\}^2 dv_1\left(u,v\right) \\ \geq &C\int\left\{\Lambda\left(u\right)\exp\left(\boldsymbol{\theta}^T\left(u\right)v(u)\right) - \Lambda_0\left(u\right)\exp\left(\boldsymbol{\theta}_0^T\left(u\right)v(u)\right)\right\}^2 dv_1\left(u,v\right) \qquad \text{(A.7)} \end{aligned}$$

Let $\rho\left(u,z\right) = \Lambda\left(u\right)\exp\left(\boldsymbol{\beta}^T\left(u\right)z\left(u\right)\right)$ and $\rho_0\left(u,z\right) = \Lambda_0\left(u\right)\exp\left(\boldsymbol{\beta}_0^T\left(u\right)z\left(u\right)\right)$. We also define $\rho_s = s\rho + \left(1-s\right)\rho_0$, $\Lambda_s = s\Lambda + \left(1-s\right)\Lambda_0$, $\boldsymbol{\gamma}_s = s\boldsymbol{\gamma} + \left(1-s\right)\boldsymbol{\gamma}_0$, $\boldsymbol{\beta}_s =$

$s\boldsymbol{\beta} + (1-s)\,\boldsymbol{\beta}_0$, $\boldsymbol{\theta}_s = s\boldsymbol{\theta} + (1-s)\,\boldsymbol{\theta}_0$ for $s \in (0,1)$. Then let

$$g\,(s) = \rho_s\,(U,\mathbf{Z})\exp\left(\boldsymbol{\gamma}_s^T\mathbf{W}\,(U)\right).$$

and clearly

$$\Lambda\,(U)\exp\left(\boldsymbol{\theta}^T\,(U)\,\mathbf{V}(U)\right) - \Lambda_0\,(U)\exp\left(\boldsymbol{\theta}_0^T\,(U)\,\mathbf{V}(U)\right) = g\,(1) - g\,(0)\,.$$

By the mean value theorem, there exists a $0 \le \xi \le 1$ such that $g\,(1) - g\,(0) = g'\,(\xi)$ where

$$\begin{aligned}
&g'\,(\xi)\\
=&\exp\left(\boldsymbol{\gamma}_\xi^T\mathbf{W}(U)\right)\left\{(\rho-\rho_0)\,(U,\mathbf{Z}) + (\xi\rho + (1-\xi)\,\rho_0)\,(U,\mathbf{Z})\,(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)^T\,\mathbf{W}(U)\right\}\\
=&\exp\left(\boldsymbol{\gamma}_\xi^T\mathbf{W}(U)\right)\left\{(\rho-\rho_0)\,(U,\mathbf{Z})\left\{1+\xi\,(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)^T\,\mathbf{W}(U)\right\}\right.\\
&\left.+\ \rho_0\,(U,\mathbf{Z})\,(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)^T\,\mathbf{W}(U)\right\}
\end{aligned}$$

From (A.6), we have

$$\begin{aligned}
&\mathbf{M}\,(\boldsymbol{\vartheta}_0) - \mathbf{M}\,(\boldsymbol{\vartheta})\\
\ge&C\int\left\{(\rho-\rho_0)\,(u,z)\left\{1+\xi\,(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)^T\,w(u)\right\} + \rho_0\,(u,z)\,(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)^T\,\mathbf{w}(u)\right\}^2\\
&\times\,d\nu_1\,(u,z,w)\\
=&C\nu_1\left\{g_1h+g_2\right\}^2
\end{aligned}$$

where $g_1(U,\mathbf{V}) = \left((\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)^T\mathbf{W}(U)\right)\rho_0(U,\mathbf{Z})$, $g_2(U,\mathbf{Z}) = (\rho-\rho_0)(U,\mathbf{Z})$ and

$$h(U,\mathbf{Z}) = 1+\xi(\rho-\rho_0)(U,\mathbf{Z})/\rho_0(U,\mathbf{Z}).$$

By the similar method in Wellner and Zhang (2007) and He et al. (2017), under condition (C12)

$$\begin{aligned}\mathbf{M}(\boldsymbol{\vartheta}_0)-\mathbf{M}(\boldsymbol{\vartheta}) &\geq C\nu_1\{g_1h+g_2\}^2\\ &\geq C_1\left\{\nu_1\left(g_1^2\right)+\nu_1\left(g_2^2\right)\right\}.\end{aligned}$$

Similarly, by the mean value theorem and condition (C12),

$$\begin{aligned}\nu_1\left(g_2^2\right) &= \nu_1\left((h_2g_3+g_4)^2\right)\\ &\geq C_2\left\{\nu_1\left(g_3^2\right)+\nu_1\left(g_4^2\right)\right\}\end{aligned}$$

where $g_3(U,\mathbf{Z}) = \left((\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T(U)\,\mathbf{Z}(U)\right)\Lambda_0(U)$, $g_4(U) = (\Lambda-\Lambda_0)(U)$ and $h_2(U) = 1+\zeta(\Lambda-\Lambda_0)(U)/\Lambda_0(U)$ for some $\zeta\in(0,1)$. Therefore,

$$\begin{aligned}\mathbf{M}(\boldsymbol{\vartheta}_0)-\mathbf{M}(\boldsymbol{\vartheta}) &\geq C_2\left\{\nu_1\left(g_1^2\right)+\nu_1\left(g_3^2\right)+\nu_1\left(g_4^2\right)\right\}\\ &= C_2\left(\|\boldsymbol{\gamma}-\gamma_0\|_2^2+\int\|\boldsymbol{\beta}(u)-\boldsymbol{\beta}_0(u)\|_2^2\,d\mu_1(u)+\|\Lambda-\Lambda_0\|_{L_2(\mu_1)}^2\right)\\ &\gtrsim d(\boldsymbol{\vartheta}_0,\boldsymbol{\vartheta}).\end{aligned}$$

Next, we need to find $\varphi_n(\delta)$ such that

$$E\left[\sup_{d(\boldsymbol{\vartheta},\boldsymbol{\vartheta}_0)<\delta}\sqrt{n}\left|(\mathbb{P}_n-\mathbf{P})\left(m_{\boldsymbol{\vartheta}}(O)-m_{\boldsymbol{\vartheta}_0}(O)\right)\right|\right]\leq c\varphi_n(\delta).$$

Let

$$\mathcal{F}_\delta = \{m_{\boldsymbol{\vartheta}}(O) - m_{\boldsymbol{\vartheta}_0}(O) : d(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_0) \le \delta\}$$

From the result of Theorem 2.7.5 of (van der Vaart and Wellner, 1996) and Lemma A.2 of Lu et al. (2009), for any $\epsilon \le \delta$, we have

$$\log N_{[]}\left(\epsilon, \mathcal{F}_\delta, \|\cdot\|_{\mathbf{P},B}\right) \le c\left(\frac{1}{\epsilon} + (p_1 + p_2 q_n)\log\left(\frac{\delta}{\epsilon}\right)\right),$$

where $\|\cdot\|_{P,B}$ is the Bernstein norm defined as $\|f\|_{P,B} = \left\{2\mathbf{P}\left(e^{|f|} - 1 - |f|\right)\right\}^{1/2}$ by (van der Vaart and Wellner, 1996, page 324). Similar to the argument in Lu et al. (2009) and Wellner and Zhang (2007), under conditions (C6) and (C10), we have

$$\|m_{\boldsymbol{\vartheta}}(O) - m_{\boldsymbol{\vartheta}_0}(O)\|^2_{\mathbf{P},B} \le c\delta^2,$$

for any $m_{\boldsymbol{\vartheta}}(O) - m_{\boldsymbol{\vartheta}_0}(O) \in \mathcal{F}_\delta$. Therefore, by Lemma 3.4.3 in (van der Vaart and Wellner, 1996), we can show a maximal inequality

$$E\left\|\sqrt{n}(\mathbb{P}_n - \mathbf{P})\right\|_{\mathcal{F}_\delta} \le cJ_{[]}\left(\delta, \mathcal{F}_\delta, \|\cdot\|_{\mathbf{P},B}\right)\left\{1 + \frac{J_{[]}\left(\delta, \mathcal{F}_\delta, \|\cdot\|_{\mathbf{P},B}\right)}{\delta^2 n^{1/2}}\right\}$$

where

$$\begin{aligned} J_{[]}\left(\delta, \mathcal{F}_\delta, \|\cdot\|_{\mathbf{P},B}\right) &= \int_0^\delta \left\{1 + \log N_{[]}\left(\epsilon, \mathcal{F}_\delta, \|\cdot\|_{\mathbf{P},B}\right)\right\}^{1/2} d\epsilon \\ &\le c_1 q_n^{1/2} \int_0^\delta \left\{1 + \frac{1}{\epsilon} + \log\left(\frac{\delta}{\epsilon}\right)\right\}^{1/2} d\epsilon \\ &\le q_n^{1/2}\delta^{1/2} \end{aligned}$$

Thus,

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

It is not hard to show that $\varphi_n(\delta)/\delta$ is decreasing in δ and therefore

$$r_n^2\varphi_n\left(\frac{1}{r_n}\right) = r_n^{3/2}q_n^{1/2} + r_n^3 q_n n^{-1/2} \lesssim n^{1/2}$$

if $r_n = \min\left\{n^{\frac{1-\nu}{3}}, n^{r\nu}\right\}$ and $0 < \nu < 1/2$.

Moreover, using similar argument in Lu et al. (2009), we can show $\mathbb{M}_n\left(\hat{\boldsymbol{\vartheta}}_n\right) - \mathbb{M}_n\left(\boldsymbol{\vartheta}_0\right) > -O_p\left(n^{-2r\nu}\right) \geq O_p\left(r_n^2\right)$. Then, by Theorem 3.2.5 of Wellner and Zhang (2007), we have $r_n d\left(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_0\right) = O_p\left(1\right)$. If ν is chosen as $1/\left(3r+1\right)$, we obtain the optimal rate $n^{r/(3r+1)}$ because $(1-\nu)/3 = r\nu$.

A.1.3 Proof of the Asymptotic Normality

We mainly use the method in He et al. (2017). We define a sequence of maps S_n mapping a neighborhood of $\boldsymbol{\vartheta}_0$, denoted by $\mathcal{U}$, in the parameter space for $\boldsymbol{\vartheta}$ into $l^\infty\left(\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3\right)$ as

$$\begin{aligned}
S_n\left(\boldsymbol{\vartheta}\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right] =& \frac{d}{d\epsilon}\, l_n\left(\boldsymbol{\gamma} + \epsilon\boldsymbol{h}_1, \boldsymbol{\beta} + \epsilon\boldsymbol{h}_2, \Lambda + \epsilon h_3\right)\Big|_{\epsilon=0} \\
=& n^{-1}\sum_{i=1}^{n}\int_0^\tau Y_i(t)\left\{\frac{N_i(t)}{\Lambda\left(t\right)}h_3\left(t\right) + N_i(t)\left(\boldsymbol{h}_1^T\mathbf{W}_i(t) + \boldsymbol{h}_2^T\left(t\right)\mathbf{Z}_i(t)\right)\right. \\
&- \left(\boldsymbol{h}_1^T\mathbf{W}_i(t) + \boldsymbol{h}_2^T\left(t\right)\mathbf{Z}_i(t)\right)\exp\left(\boldsymbol{\gamma}^T\mathbf{W}_i(t) + \boldsymbol{\beta}^T\left(t\right)\mathbf{Z}_i(t)\right)\Lambda\left(t\right) \\
&\left. - \exp\left(\boldsymbol{\gamma}^T\mathbf{W}_i(t) + \boldsymbol{\beta}^T\mathbf{Z}_i(t)\right)h_3\left(t\right)\right\}dH_i(t) \\
=& A_{n1}\left(\boldsymbol{\vartheta}\right)\left[\boldsymbol{h}_1\right] + A_{n2}\left(\boldsymbol{\vartheta}\right)\left[\boldsymbol{h}_2\right] + A_{n3}\left(\boldsymbol{\vartheta}\right)\left[h_3\right] \\
=& \mathbb{P}_n\psi\left(\boldsymbol{\vartheta}\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right],
\end{aligned}$$

where

$$A_{n1}\left(\boldsymbol{\vartheta}\right)\left[\boldsymbol{h}_1\right] =$$
$$n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\boldsymbol{h}_1^T\mathbf{W}_i(t)\left\{N_i(t) - \exp\left(\boldsymbol{\gamma}^T\mathbf{W}_i(t) + \boldsymbol{\beta}^T\left(t\right)\mathbf{Z}_i(t)\right)\Lambda\left(t\right)\right\}dH_i(t),$$

$$A_{n2}\left(\boldsymbol{\vartheta}\right)\left[\boldsymbol{h}_2\right] =$$
$$n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\boldsymbol{h}_2^T\left(t\right)\mathbf{Z}_i(t)\left\{N_i(t) - \exp\left(\boldsymbol{\gamma}^T\mathbf{W}_i(t) + \boldsymbol{\beta}^T\left(t\right)\mathbf{Z}_i(t)\right)\Lambda\left(t\right)\right\}dH_i(t),$$

and

$$A_{n3}\left(\boldsymbol{\vartheta}\right)\left[h_3\right] = n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)h_3\left(t\right)\left\{\frac{N_i(t)}{\Lambda\left(t\right)} - \exp\left(\boldsymbol{\gamma}^T\mathbf{W}_i(t) + \boldsymbol{\beta}^T\mathbf{Z}_i(t)\right)\right\}dH_i(t).$$

Correspondingly, we define the limit map $S : \mathcal{U} \to l^{\infty}\left(\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3\right)$ as

$$S\left(\boldsymbol{\vartheta}\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right] = A_1\left(\boldsymbol{\vartheta}\right)\left[\boldsymbol{h}_1\right] + A_2\left(\boldsymbol{\vartheta}\right)\left[\boldsymbol{h}_2\right] + A_3\left(\boldsymbol{\vartheta}\right)\left[h_3\right]$$

where

$$A_1\left(\boldsymbol{\vartheta}\right)\left[\boldsymbol{h}_1\right] = \mathbf{P}\int_0^{\tau} Y(t)\boldsymbol{h}_1^T\mathbf{W}(t)\left\{N(t) - \exp\left(\boldsymbol{\gamma}^T\mathbf{W}(t) + \boldsymbol{\beta}^T\left(t\right)\mathbf{Z}(t)\right)\Lambda\left(t\right)\right\}dH(t),$$

$$A_2\left(\boldsymbol{\vartheta}\right)\left[\boldsymbol{h}_2\right]$$
$$= \mathbf{P}\int_0^{\tau} Y(t)\boldsymbol{h}_2^T\left(t\right)\mathbf{Z}(t)\left\{N(t) - \exp\left(\boldsymbol{\gamma}^T\mathbf{W}(t) + \boldsymbol{\beta}^T\left(t\right)\mathbf{Z}(t)\right)\Lambda\left(t\right)\right\}dH(t),$$

and

$$A_3(\boldsymbol{\vartheta})[h_3] = \mathbf{P}\int_0^\tau Y(t)h_3(t)\left\{\frac{N(t)}{\Lambda(t)} - \exp\left(\boldsymbol{\gamma}^T\mathbf{W}(t) + \boldsymbol{\beta}^T\mathbf{Z}(t)\right)\right\}dH(t).$$

To derive the asymptotic normality of $\hat{\boldsymbol{\vartheta}}_n$, we need to verify the following five conditions in He et al. (2017).

(a1) $\sqrt{n}(S_n - S)\left(\hat{\boldsymbol{\vartheta}}_n\right) - \sqrt{n}(S_n - S)(\boldsymbol{\vartheta}_0) = o_p(1)$.

(a2) $S(\boldsymbol{\vartheta}_0) = 0$ and $S_n\left(\hat{\boldsymbol{\vartheta}}_n\right) = o_p\left(n^{-1/2}\right)$.

(a3) $\sqrt{n}(S_n - S)(\theta_0)$ converges in distribution to a tight Gaussian process on $l^\infty(\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3)$.

(a4) $S(\boldsymbol{\vartheta})$ is Fréchet-differentiable at $\boldsymbol{\vartheta}_0$ denoted by $\dot{S}(\boldsymbol{\vartheta}_0)$.

(a5) $S\left(\hat{\boldsymbol{\vartheta}}_n\right) - S(\boldsymbol{\vartheta}_0) - \dot{S}(\boldsymbol{\vartheta}_0)\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right) = o_p\left(n^{-1/2}\right)$.

Using similar argument in Lu et al. (2009), it is not hard to show

$$\left\{\psi(\boldsymbol{\vartheta})[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3] - \psi(\boldsymbol{\vartheta}_0)[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3] : d(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_0) < \delta, (\boldsymbol{h}_1, \boldsymbol{h}_2, h_3) \in \mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3\right\}$$

is a Donkser class for some δ. Therefore,

$$\sup_{(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3) \in \mathcal{A}\times\mathcal{M}\times\mathcal{F}} \mathbf{P}\left\{\psi\left(\hat{\boldsymbol{\vartheta}}_n\right)[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3] - \psi(\boldsymbol{\vartheta}_0)[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3]\right\}^2 \to 0$$

as $d\left(\hat{\boldsymbol{\vartheta}}_n, \boldsymbol{\vartheta}_0\right) \to 0$ in probability and thus (a1) holds.

For (a2), clearly, $S(\boldsymbol{\vartheta}_0) = 0$. For $\boldsymbol{h}_2 \in \mathcal{H}_2$, let $\boldsymbol{h}_{2n}$ be the B-spline function approximation of $\boldsymbol{h}_2$ with $\max_{j=1,\ldots,p_2}\|h_{2j} - h_{2nj}\|_\infty = O\left(n^{-\nu r}\right)$ by (A.13). Then we have $S_n\left(\hat{\boldsymbol{\vartheta}}_n\right)[\boldsymbol{h}_1,\boldsymbol{h}_2,h_{3n}]=0$. Thus,

$$\begin{aligned}S_n\left(\hat{\boldsymbol{\vartheta}}_n\right)[\boldsymbol{h}_1,\boldsymbol{h}_2,h_3] &= \sqrt{n}\mathbb{P}_n\psi\left(\hat{\boldsymbol{\vartheta}}_n\right)[\boldsymbol{h}_1,\boldsymbol{h}_2,h_3]-\sqrt{n}\mathbf{P}\psi\left(\hat{\boldsymbol{\vartheta}}_n\right)[\boldsymbol{h}_1,\boldsymbol{h}_2,h_{3n}]\\ &= I_{n1}-I_{n2}+I_{n3}+I_{n4}\end{aligned}$$

where

$$I_{n1}=\sqrt{n}\left(\mathbb{P}_n-\mathbf{P}\right)\left\{\psi\left(\hat{\boldsymbol{\vartheta}}_n\right)[\boldsymbol{h}_1,\boldsymbol{h}_2,h_3]-\psi\left(\boldsymbol{\vartheta}_0\right)[\boldsymbol{h}_1,\boldsymbol{h}_2,h_3]\right\}$$

$$I_{n2}=\sqrt{n}\left(\mathbb{P}_n-\mathbf{P}\right)\left\{\psi\left(\hat{\boldsymbol{\vartheta}}_n\right)[\boldsymbol{h}_1,\boldsymbol{h}_{2n},h_3]-\psi\left(\boldsymbol{\vartheta}_0\right)[\boldsymbol{h}_1,\boldsymbol{h}_{2n},h_3]\right\}$$

$$I_{n3}=\sqrt{n}\mathbb{P}_n\left\{\psi\left(\boldsymbol{\vartheta}_0\right)[\boldsymbol{h}_1,\boldsymbol{h}_2,h_3]-\psi\left(\boldsymbol{\vartheta}_0\right)[\boldsymbol{h}_1,\boldsymbol{h}_{2n},h_3]\right\}$$

and

$$I_{n4}=\sqrt{n}\mathbf{P}\left\{\psi\left(\hat{\boldsymbol{\vartheta}}_n\right)[\boldsymbol{h}_1,\boldsymbol{h}_2,h_3]-\psi\left(\hat{\boldsymbol{\vartheta}}_n\right)[\boldsymbol{h}_1,\boldsymbol{h}_{2n},h_3]\right\}.$$

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

$$\begin{aligned}I_{n3}=&\sqrt{n}\left(\mathbb{P}_n-\mathbf{P}\right)\left\{\psi\left(\boldsymbol{\vartheta}_0\right)[\boldsymbol{h}_1,\boldsymbol{h}_2,h_3]-\psi\left(\boldsymbol{\vartheta}_0\right)[\boldsymbol{h}_1,\boldsymbol{h}_{2n},h_3]\right\}\\ &+\sqrt{n}\mathbf{P}\left\{\psi\left(\boldsymbol{\vartheta}_0\right)[\boldsymbol{h}_1,\boldsymbol{h}_2,h_3]-\psi\left(\boldsymbol{\vartheta}_0\right)[\boldsymbol{h}_1,\boldsymbol{h}_{2n},h_3]\right\}\\ =&I_{n31}+I_{n32}.\end{aligned}$$

Similarly to proving (a1), $I_{n31}=o_p(1)$ and $I_{n32}=0$ since $S\left(\boldsymbol{\vartheta}_0\right)=0$ for any $h_2, h_{2n}\in$

$\mathcal{H}_2$. For I_{n4},

$$
\begin{aligned}
|I_{n4}| &\leq \sqrt{n} d\left(\hat{\boldsymbol{\vartheta}}_n, \theta_0\right)\left(\max_{j=1,\ldots,p_2} \|h_{2j} - h_{2nj}\|_\infty\right) \\
&= O_p\left(\max\left\{n^{-(1-\nu)/3}, n^{-r\nu}\right\} n^{-rv+1/2}\right) \\
&= o_p\left(1\right)
\end{aligned}
$$

if $1/\left(4r\right) < \nu < 1/2$. Thus (a2) holds.

Condition (a3) holds because $\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$ is a Donsker class and the functionals $A_1\left(\boldsymbol{\vartheta}\right)\left[\boldsymbol{h}_1\right]$, $A_2\left(\boldsymbol{\vartheta}\right)\left[\boldsymbol{h}_2\right]$ and $A_3\left(\boldsymbol{\vartheta}\right)\left[h_3\right]$ are bounded Lipschitz functions with respect to $\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$ due the compactness of $\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$.

For (a4), by the smoothness of $S\left(\boldsymbol{\vartheta}\right)$ the Fréchet differentiability holds and the derivative of $S\left(\boldsymbol{\vartheta}\right)$ at $\boldsymbol{\vartheta}_0$, denoted by $\dot{S}\left(\boldsymbol{\vartheta}_0\right)$ is a map from the space $\left\{\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0 : \boldsymbol{\vartheta} \in \mathcal{U}\right\}$

to $l^\infty(\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3)$. Now we calculate $\dot{S}\left(\boldsymbol{\vartheta}_0\right)$ as

$$
\begin{aligned}
& \dot{S}\left(\boldsymbol{\vartheta}_0\right)\left(\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right] \\
=& \frac{d}{d\epsilon}\left\{A_1\left(\boldsymbol{\vartheta}_0+\epsilon\left(\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0\right)\right)\left[\boldsymbol{h}_1\right]\right\}\Big|_{\epsilon=0} \\
& +\frac{d}{d\epsilon}\left\{A_2\left(\boldsymbol{\vartheta}_0+\epsilon\left(\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0\right)\right)\left[\boldsymbol{h}_1\right]\right\}\Big|_{\epsilon=0} \\
& +\frac{d}{d\epsilon}\left\{A_3\left(\boldsymbol{\vartheta}_0+\epsilon\left(\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0\right)\right)\left[\boldsymbol{h}_1\right]\right\}\Big|_{\epsilon=0} \\
=& -\mathbf{P} \int_0^\tau Y(t) \boldsymbol{h}_1^T \mathbf{W}(t) \exp\left(\boldsymbol{\gamma}^T \mathbf{W}(t)+\boldsymbol{\beta}^T(t) \mathbf{Z}(t)\right) \\
& \times\left\{\left(\left(\boldsymbol{\gamma}-\gamma_0\right)^T \mathbf{W}(t)+\left(\boldsymbol{\beta}(t)-\boldsymbol{\beta}_0(t)\right)^T \mathbf{Z}(t)\right) \Lambda_0(t)+\left(\Lambda(t)-\Lambda_0(t)\right)\right\} dH(t) \\
& -\mathbf{P} \int_0^\tau Y(t) \boldsymbol{h}_2^T(t) \mathbf{Z}(t) \exp\left(\boldsymbol{\gamma}^T \mathbf{W}(t)+\boldsymbol{\beta}^T(t) \mathbf{Z}(t)\right) \\
& \times\left\{\left(\left(\boldsymbol{\gamma}-\gamma_0\right)^T \mathbf{W}(t)+\left(\boldsymbol{\beta}(t)-\boldsymbol{\beta}_0(t)\right)^T \mathbf{Z}(t)\right) \Lambda_0(t)+\left(\Lambda(t)-\Lambda_0(t)\right)\right\} dH(t) \\
& -\mathbf{P} \int_0^\tau Y(t) h_3(t) \exp\left(\boldsymbol{\gamma}^T \mathbf{W}(t)+\boldsymbol{\beta}^T \mathbf{Z}(t)\right) \\
& \times\left\{\frac{\Lambda(t)-\Lambda_0(t)}{\Lambda_0(t)}+\left(\boldsymbol{\gamma}-\gamma_0\right)^T \mathbf{W}(t)+\left(\boldsymbol{\beta}(t)-\boldsymbol{\beta}_0(t)\right)^T \mathbf{Z}(t)\right\} dH(t).
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
& \dot{S}\left(\boldsymbol{\vartheta}_0\right)\left(\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right] \\
=& \left(\boldsymbol{\gamma}-\gamma_0\right)^T Q_1\left(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right) \\
& +\int_0^\tau\left(\boldsymbol{\beta}(t)-\boldsymbol{\beta}_0(t)\right)^T dQ_2\left(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right)(t) \\
& +\int_0^\tau\left(\Lambda(t)-\Lambda_0(t)\right) dQ_3\left(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right)(t)
\end{aligned}
$$

where

$$Q_1(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3) = -\mathbf{P}\int_0^{\tau} \mathbf{W}(t)Y(t)\left(\boldsymbol{h}_1^T\mathbf{W}(t)\Lambda_0(t) + \boldsymbol{h}_2^T(t)\mathbf{Z}(t)\Lambda_0(t) + h_3(t)\right) \times \exp\left(\boldsymbol{\gamma}^T\mathbf{W}(t) + \boldsymbol{\beta}^T(t)\mathbf{Z}(t)\right)dH(t),$$

$$dQ_2(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3)(t) = -\mathbf{P}\mathbf{Z}(t)Y(t)\left(\boldsymbol{h}_1^T\mathbf{W}(t)\Lambda_0(t) + \boldsymbol{h}_2^T(t)\mathbf{Z}(t)\Lambda_0(t) + h_3(t)\right) \times \exp\left(\boldsymbol{\gamma}^T\mathbf{W}(t) + \boldsymbol{\beta}^T(t)\mathbf{Z}(t)\right)dH(t),$$

and

$$dQ_3(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3)(t) = -\mathbf{P}\frac{Y(t)}{\Lambda_0(t)}\left(\boldsymbol{h}_1^T\mathbf{W}(t)\Lambda_0(t) + \boldsymbol{h}_2^T(t)\mathbf{Z}(t)\Lambda_0(t) + h_3(t)\right) \times \exp\left(\boldsymbol{\gamma}^T\mathbf{W}(t) + \boldsymbol{\beta}^T(t)\mathbf{Z}(t)\right)dH(t).$$

We can also show $Q = (Q_1, Q_2, Q_3)$ is one-to-one by the similar method in He et al. (2017).

For (a5), we have

$$S\left(\hat{\boldsymbol{\vartheta}}_n\right) - S(\boldsymbol{\vartheta}_0) - \dot{S}(\boldsymbol{\vartheta}_0)\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right) = B_{n1} + B_{n2} + B_{n3}$$

where

$$B_{n1} = A_1\left(\hat{\boldsymbol{\vartheta}}_n\right)[\boldsymbol{h}_1] - \frac{d}{d\varepsilon}\left\{A_1\left(\boldsymbol{\vartheta}_0 + \varepsilon\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right)\right)[\boldsymbol{h}_1]\right\}\Big|_{\varepsilon=0},$$

$$B_{n2} = A_2\left(\hat{\boldsymbol{\vartheta}}_n\right)[\boldsymbol{h}_2] - \frac{d}{d\varepsilon}\left\{A_2\left(\boldsymbol{\vartheta}_0 + \varepsilon\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right)\right)[\boldsymbol{h}_2]\right\}\Big|_{\varepsilon=0},$$

and

$$B_{n3} = A_3\left(\hat{\boldsymbol{\vartheta}}_n\right)[h_3] - \frac{d}{d\varepsilon}\left\{A_3\left(\boldsymbol{\vartheta}_0 + \varepsilon\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right)\right)[h_3]\right\}\Bigg|_{\varepsilon=0}.$$

It is not hard to see

$$\begin{aligned}B_{n1} = \mathbf{P}\int_0^\tau & Y(t)\boldsymbol{h}_1^T\mathbf{W}(t)\exp\left(\boldsymbol{\gamma}_0^T\mathbf{W}(t) + \boldsymbol{\beta}_0^T(t)\,\mathbf{Z}(t)\right)\Lambda_0(t) \\ & \times q_1\left(\left(\hat{\boldsymbol{\gamma}} - \gamma_0\right)^T\mathbf{W}(t) - \left(\hat{\boldsymbol{\beta}}_n(t) - \boldsymbol{\beta}_0(t)\right)^T\mathbf{Z}(t)\right)dH(t),\end{aligned}$$

where $q_1(x) = 1 - \exp(y)(1-y)$ and $q_1(x) \leq x^2$ when x is in a neighborhood of 0. Thus

$$\begin{aligned}B_{n1} \leq & \mathbf{P}\int_0^\tau Y(t)\boldsymbol{h}_1^T\mathbf{W}(t)\exp\left(\boldsymbol{\gamma}_0^T\mathbf{W}(t) + \boldsymbol{\beta}_0^T(t)\,\mathbf{Z}(t)\right)\Lambda_0(t) \\ & \times\left\{\left(\hat{\boldsymbol{\gamma}} - \gamma_0\right)^T\mathbf{W}(t) - \left(\hat{\boldsymbol{\beta}}_n(t) - \boldsymbol{\beta}_0(t)\right)^T\mathbf{Z}(t)\right\}^2 dH(t). \\ = & O\left(d^2\left(\hat{\boldsymbol{\vartheta}}_n, \boldsymbol{\vartheta}\right)\right).\end{aligned}$$

Similarly, we can show $B_{n2} \leq O\left(d^2\left(\hat{\boldsymbol{\vartheta}}_n, \boldsymbol{\vartheta}\right)\right)$ and $B_{n3} \leq O\left(d^2\left(\hat{\boldsymbol{\vartheta}}_n, \boldsymbol{\vartheta}\right)\right)$ and hence

$$S\left(\hat{\boldsymbol{\vartheta}}_n\right) - S\left(\boldsymbol{\vartheta}_0\right) - \dot{S}\left(\boldsymbol{\vartheta}_0\right)\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right) \leq O\left(d^2\left(\hat{\boldsymbol{\vartheta}}_n, \boldsymbol{\vartheta}\right)\right).$$

Since $n^{1/2}d^2\left(\hat{\boldsymbol{\vartheta}}_n, \boldsymbol{\vartheta}\right) = O_p\left(n^{1/2}\max\left\{n^{-2(1-\nu)/3}, n^{-2r\nu}\right\}\right) = o_p(1)$ if $1/(4r) < v < 1/4$, we can conclude that $S\left(\hat{\boldsymbol{\vartheta}}_n\right) - S\left(\boldsymbol{\vartheta}_0\right) - \dot{S}\left(\boldsymbol{\vartheta}_0\right)\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right) = o_p\left(n^{-1/2}\right)$ and (a5) holds.

If (a1)-(a5) hold, according to He et al. (2017), we have

$$-\sqrt{n}\dot{S}\left(\boldsymbol{\vartheta}_0\right)\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right)[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3] = \sqrt{n}\left(S_n - S\right)\left(\boldsymbol{\vartheta}_0\right)[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3] + o_p(1),$$

uniformly in $\boldsymbol{h}_1, \boldsymbol{h}_2, h_3$. For each $(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3) \in \mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$, Q is invertible by the similar arguement in He et al. (2017). Then there exists $(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3) \in \mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$ such that

$$Q_1\left(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right) = \boldsymbol{h}_1, Q_2\left(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right) = \boldsymbol{h}_2, Q_3\left(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right) = \boldsymbol{h}_3.$$

Therefore, we have

$$\begin{aligned}
&\left(\hat{\gamma} - \gamma_0\right)^T \boldsymbol{h}_1 + \int_0^\tau \left(\hat{\boldsymbol{\beta}}_n\left(t\right) - \boldsymbol{\beta}_0\left(t\right)\right)^T d\boldsymbol{h}_2\left(t\right) + \int_0^\tau \left(\hat{\Lambda}\left(t\right) - \Lambda_0\left(t\right)\right) dh_3\left(t\right) \\
=&\sqrt{n}\left(S_n - S\right)\left(\boldsymbol{\vartheta}_0\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right] + o_p\left(1\right) \rightarrow_d N\left(0, \sigma^2\right)
\end{aligned}$$

where $\sigma^2 = E\left[\psi^2\left(\boldsymbol{\vartheta}_0\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right]\right]$ because of (a3).

To find the asymptotic distribution of γ only, we can find $\boldsymbol{h}_1$, $\boldsymbol{h}_2$ and h_3 as a solution of $Q_2 = 0$ and $Q_3 = 0$. Unfortunately, we cannot find the explicit forms of $\boldsymbol{h}_1$, $\boldsymbol{h}_2$ and h_3 as He et al. (2017). Hence, we adopt the variance estimation method in the Chapter 2.

A.2 Proofs of Theorems 4, 5, and 6

In this section, we will sketch the proofs of the consistency, rate of convergence, and asymptotic normality of $\hat{\boldsymbol{\vartheta}}_n$ described in Theorems 4 - 6.

When splines are involved in estimating equation-based methods, a more straightforward method to derive the asymptotic properties of $\hat{\boldsymbol{\vartheta}}_n$ is to find a equivalent optimization problem to solving equation (3.3) and follow the M-estimator theory (van der Vaart and Wellner, 1996). We will show that solving (3.3) is equivalent to maximize

$$l_n(\boldsymbol{\vartheta}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau Y_i(t)$$
$$\left(N_{ik}(t)\log\left(\Lambda_k(t)\right) + N_{ik}(t)\boldsymbol{\theta}^T(t)\,\mathbf{V}_i(t) - \exp\left(\boldsymbol{\theta}^T(t)\,\mathbf{V}_i(t)\right)\Lambda_k(t)\right)dH_{ik}(t) \quad \text{(A.8)}$$

with respect to $\boldsymbol{\vartheta}$ over $\Theta_n = \mathcal{A}\times\mathcal{M}_n\times\mathcal{F}$. Here $\boldsymbol{\theta}^T(t)\,\mathbf{V}_i(t) = \boldsymbol{\gamma}^T\mathbf{W}_i(t) + \boldsymbol{\beta}^T(t)\mathbf{Z}_i(t)$ and $\boldsymbol{\theta}(t) = (\boldsymbol{\gamma}, \boldsymbol{\beta}(t))$. Let $\mathbb{P}_n$ be the empirical measure and $\mathbf{P}$ be the true measure. Let $\mathbb{M}_n(\boldsymbol{\vartheta}) = l_n(\boldsymbol{\vartheta}) = \mathbb{P}_n m_{\boldsymbol{\vartheta}}(\mathbf{V})$ and $\mathbf{M}(\boldsymbol{\vartheta}) = \mathbf{P}m_{\boldsymbol{\vartheta}}(\mathbf{V})$, where

$$m_{\boldsymbol{\vartheta}}(\mathbf{O}) = \sum_{k=1}^{K}\int_0^\tau Y(t)\left(N_{\cdot k}(t)\log\left(\Lambda_k(t)\right) + N_{\cdot k}(t)\boldsymbol{\theta}^T(t)\,\mathbf{V}(t)\right.$$
$$\left. - \exp\left(\boldsymbol{\theta}^T(t)\,\mathbf{V}(t)\right)\Lambda_k(t)\right)dH_{\cdot k}(t)$$

and $\mathbf{O}$ denotes the data.

Follow the definition $\hat{\Lambda}_k(t, \hat{\boldsymbol{\theta}}_n)$, we define

$$\hat{\Lambda}_k(t, \boldsymbol{\theta}(t)) = \frac{\sum_{j=1}^n Y_j(t)N_{jk}(t)dH_{jk}(t)}{\sum_{j=1}^n Y_j(t)\exp\left(\boldsymbol{\theta}^T(t)\,\mathbf{V}_j(t)\right)dH_{jk}(t)}.$$

First, we show $\hat{\Lambda}_k(t, \boldsymbol{\theta}(t))$ is a unique maximize of (A.8) when $\boldsymbol{\theta}(t)$ is given. After some calculation,

$$
\begin{aligned}
&\mathbb{M}_n\left(\boldsymbol{\theta}, \boldsymbol{\Lambda}\right) - \mathbb{M}_n\left(\boldsymbol{\theta}, \hat{\boldsymbol{\Lambda}}\right) \\
=&\frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^\tau Y_i(t) \hat{\Lambda}_k(t, \boldsymbol{\theta}\left(t\right)) \exp\left(\boldsymbol{\theta}^T\left(t\right) \mathbf{V}_i(t)\right) \\
&\times \left\{ \frac{N_{ik}(t)}{\hat{\Lambda}_k(t, \boldsymbol{\theta}\left(t\right)) \exp\left(\boldsymbol{\theta}^T\left(t\right) \mathbf{V}_i(t)\right)} \log\left(\frac{\Lambda_k\left(t\right)}{\hat{\Lambda}_k(t, \boldsymbol{\theta}\left(t\right))}\right) \right. \\
&+ \left. \left(1 - \frac{\Lambda_k\left(t\right)}{\hat{\Lambda}_k(t, \boldsymbol{\theta}\left(t\right))}\right) \right\} dH_{ik}(t).
\end{aligned}
$$

Focusing on the the second term in the bracket, by plugging in $\hat{\Lambda}_k(t, \boldsymbol{\theta}\left(t\right))$, Fubini's theorem yields that

$$
\begin{aligned}
&\frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^\tau Y_i(t) \hat{\Lambda}_k(t, \boldsymbol{\theta}\left(t\right)) \exp\left(\boldsymbol{\theta}^T\left(t\right) \mathbf{V}_i(t)\right) \left(1 - \frac{\Lambda_k\left(t\right)}{\hat{\Lambda}_k(t, \boldsymbol{\theta}\left(t\right))}\right) dH_{ik}(t) \\
=&\frac{1}{n} \sum_{k=1}^{K} \int_0^\tau \left(1 - \frac{\Lambda_k\left(t\right)}{\hat{\Lambda}_k(t, \boldsymbol{\theta}\left(t\right))}\right) \left(\sum_{j=1}^{n} Y_j(t) N_{jk}(t) dH_{jk}\left(t\right)\right) \qquad \text{(A.9)} \\
&\times \sum_{i=1}^{n} \frac{Y_i(t) \exp\left(\boldsymbol{\theta}^T\left(t\right) \mathbf{V}_i(t)\right) dH_{ik}(t)}{\sum_{j=1}^{n} Y_j(t) \exp\left(\boldsymbol{\theta}^T\left(t\right) \mathbf{V}_i(t)\right) dH_{jk}\left(t\right)} \\
=&\frac{1}{n} \int_0^\tau \sum_{k=1}^{K} \sum_{j=1}^{n} Y_j(t) N_{jk}(t) \left(1 - \frac{\Lambda_k\left(t\right)}{\hat{\Lambda}_k(t, \boldsymbol{\theta}\left(t\right))}\right) dH_{jk}\left(t\right). \qquad \text{(A.10)}
\end{aligned}
$$

The last equation holds because

$$
\begin{aligned}
&\sum_{i=1}^{n} \frac{Y_i(t) \exp\left(\boldsymbol{\theta}^T\left(t\right) \mathbf{V}_i(t)\right) dH_{ik}(t)}{\sum_{j=1}^{n} Y_j(t) \exp\left(\boldsymbol{\theta}^T\left(t\right) \mathbf{V}_i(t)\right) dH_{jk}\left(t\right)} \\
=&\frac{\sum_{i=1}^{n} Y_i(t) \exp\left(\boldsymbol{\theta}^T\left(t\right) \mathbf{V}_i(t)\right) dH_{ik}(t)}{\sum_{j=1}^{n} Y_j(t) \exp\left(\boldsymbol{\theta}^T\left(t\right) \mathbf{V}_i(t)\right) dH_{jk}\left(t\right)} = 1.
\end{aligned}
$$

Hence, by (A.10), $\mathbb{M}_n(\boldsymbol{\theta}, \boldsymbol{\Lambda}) - \mathbb{M}_n\left(\boldsymbol{\theta}, \hat{\boldsymbol{\Lambda}}\right)$ reduces to

$$\frac{1}{n}\sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)N_{ik}(t)\left\{\log\left(\frac{\Lambda_k(t)}{\hat{\Lambda}_k(t,\boldsymbol{\theta}(t))}\right) - \frac{\Lambda_k(t)}{\hat{\Lambda}_k(t)} + 1\right\} dH_{ik}(t).$$

Then, since $\log(x) - x + 1 \leq -(x-1)^2$ for all positive x and the equality holds iff $x = 1$,

$$\begin{aligned}
&\mathbb{M}_n(\boldsymbol{\theta}, \boldsymbol{\Lambda}) - \mathbb{M}_n\left(\boldsymbol{\theta}, \hat{\boldsymbol{\Lambda}}\right) \\
=&\frac{1}{n}\sum_{k=1}^{K}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)N_{ik}(t)\left\{\log\left(\frac{\Lambda_k(t)}{\hat{\Lambda}_k(t,\boldsymbol{\theta}(t))}\right) - \frac{\Lambda_k(t)}{\hat{\Lambda}_k(t)} + 1\right\} dH_{ik}(t) \\
\leq& -\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau} Y_i(t)N_{ik}(t)\left(\frac{\Lambda_k(t)}{\hat{\Lambda}_k(t,\boldsymbol{\theta}(t))} - 1\right)^2 dH_{ik}(t) \\
\leq& 0
\end{aligned}$$

as $Y_i(t)$ and $N_{ik}(t)$ are all nonnegative. This suggested $\mathbb{M}_n(\boldsymbol{\theta}, \boldsymbol{\Lambda}) \leq \mathbb{M}_n\left(\boldsymbol{\theta}, \hat{\boldsymbol{\Lambda}}\right)$ for arbitrary $\boldsymbol{\theta}$ and the equality holds iff $\Lambda_k(t) = \hat{\Lambda}_k(t, \boldsymbol{\theta}(t))$ a.e. on μ_{0k}, $k = 1, \ldots, K$. Thus, $\hat{\Lambda}_k(t, \boldsymbol{\theta}(t))$ is the unique maximizer of $l_n(\boldsymbol{\vartheta}) = l_n\left(\boldsymbol{\theta}, \hat{\boldsymbol{\Lambda}}(t,\boldsymbol{\theta})\right)$ for given $\boldsymbol{\theta}$. Then, to maximize $l_n(\boldsymbol{\vartheta})$ over $\mathcal{A} \times \mathcal{M}_n \times \mathcal{F}$, by the idea of profile likelihood in Wellner and Zhang (2007), we need to maximize $l_n\left(\boldsymbol{\theta}_n, \hat{\boldsymbol{\Lambda}}(t,\boldsymbol{\theta}_n)\right)$ with respect to $\boldsymbol{\theta}_n$. After some algebra, $\partial l_n\left(\boldsymbol{\theta}_n, \hat{\boldsymbol{\Lambda}}(t,\boldsymbol{\theta}_n)\right)/\partial\boldsymbol{\theta}_n = 0$ is equivalent to (3.3). Obviously, $l_n\left(\boldsymbol{\theta}_n, \hat{\boldsymbol{\Lambda}}(t,\boldsymbol{\theta}_n)\right)$ is convex with respect to $\boldsymbol{\theta}_n$, implying maximizing $l_n(\boldsymbol{\vartheta})$ over $\mathcal{A} \times \mathcal{M}_n \times \mathcal{F}$ is equivalent to solving (3.3).

Since $\hat{\boldsymbol{\vartheta}}_n = \arg\max_{\boldsymbol{\vartheta}\in\mathcal{A}\times\mathcal{M}_n\times\mathcal{F}} l_n(\boldsymbol{\vartheta})$, the the asymptotic properties of estimator $\hat{\boldsymbol{\vartheta}}_n$ can be investigated through $m_{\vartheta}(\mathbf{O})$ with the empirical process theory. An insight from this equivalence is that (A.8) can be regarded as an extension of the pseudo-likelihood function for panel count data proposed in Wellner and Zhang (2007) to the multivariate case. We can then use many conclusions regarding the pseudo-likelihood function in existing literature to facilitate our theoretical justification.

A.2.1 Proof of Consistency

We first show $\boldsymbol{\vartheta}_0$ is the unique maximizer of $\mathbf{M}(\boldsymbol{\vartheta})$ which concerns the identifiability of the parameter. After some calculation based on the conditional expectation on $\mathbf{V}$, we have

$$\mathbf{M}(\boldsymbol{\vartheta}) = \sum_{k=1}^{K} \int \exp\left(\boldsymbol{\theta}^T(t)\, v(t)\right) \Lambda_k(t) \left\{\log\left(\Lambda_k(t)\right) + \boldsymbol{\theta}^T(t)\, v(t) - 1\right\} d\upsilon_{1k}(t, v).$$

and therefore

$$\begin{aligned}\mathbf{M}(\boldsymbol{\vartheta}_0) - \mathbf{M}(\boldsymbol{\vartheta}) = \sum_{k=1}^{K} \int & \Lambda_k(u) \exp\left(\boldsymbol{\theta}^T(u)\, v(u)\right) \\ & \times h\left\{\frac{\Lambda_{0k}(u) \exp\left(\boldsymbol{\theta}_0^T(u)\, v(u)\right)}{\Lambda_k(u) \exp\left(\boldsymbol{\theta}^T(u)\, v(u)\right)}\right\} d\upsilon_{1k}(t, v) \quad \text{(A.11)}\end{aligned}$$

where $h(x) = x \log(x) - x + 1$. Note that $h(x) \geq 0$ for all $x > 0$ and the equality holds only when $x = 1$. Therefore, by similar argument in Wellner and Zhang (2007), under conditions (C2) and (C8), $\mathbf{M}(\boldsymbol{\vartheta}_0) = \mathbf{M}(\boldsymbol{\vartheta})$ if and only if $\boldsymbol{\theta}(t) = \theta_0(t)$ and $\Lambda_k(t) = \Lambda_{0k}(t)$ a.e. with respect to υ_{1k}, for $k = 1, \ldots, K$. In this manner, $\boldsymbol{\vartheta}_0$ is the unique maximizer of $\mathbf{M}(\boldsymbol{\vartheta})$.

By the similar arguments used in Wellner and Zhang (2007), by conditions (C1)-(C5), we can show that $\hat{\Lambda}_k(t)$, $k = 1, \ldots, K$, is uniformly bounded in probability for $t \in [0, \tau]$ with $\mu_{0k}(\{\tau\}) > 0$. By Helly-Selection Theorem and compactness of Θ_n, it follows that $\hat{\boldsymbol{\vartheta}}_n = \left(\hat{\gamma}, \hat{\boldsymbol{\beta}}_n, \hat{\Lambda}\right)$ has a subsequence $\hat{\boldsymbol{\vartheta}}_{n_l}$ converging to $\boldsymbol{\vartheta}^+ = \left(\boldsymbol{\gamma}^+, \boldsymbol{\beta}^+, \Lambda^+\right)$ with $\boldsymbol{\vartheta}^+ \in \Theta$. By the compactness of Θ_n as well as the fact that $m_{\boldsymbol{\vartheta}}(O)$ is upper semicontinuous in $\boldsymbol{\vartheta}$ for almost all O. Furthermore, $m_{\boldsymbol{\vartheta}} \leq \sum_{k=1}^{K} M_{0k} < \infty$ by condition (C9). Thus, by Theorem A.1 of Wellner and Zhang (2007), we have

$$\limsup_{n\to\infty} \sup_{\boldsymbol{\vartheta}\in\Theta_n} \left(\mathbb{P}_n - \mathbf{P}\right) m_{\boldsymbol{\vartheta}}(\mathbf{V}) \leq 0 \tag{A.12}$$

almost surely. By the Dominated Convergence Theorem and (C9), $M(\boldsymbol{\vartheta})$ is continuous in $\boldsymbol{\vartheta}$.

By the Corollary 6.21 of Schumaker (1981), there exists a spline approximation $\beta_{n0j}(t) \in \mathcal{M}_{nj}$ to β_{0j} such that

$$\sup_{t\in[0,\tau]} \left|\beta_{0j}(t) - \beta_{n0j}(t)\right| = O\left(K_n^{-r}\right) = O\left(n^{-vr}\right) \tag{A.13}$$

for $j = 1, \ldots, p_2$. Therefore, for any $\epsilon > 0$, there exists $\boldsymbol{\beta}_0^* \in \mathcal{M}_n$ such that

$$\mathbf{M}(\boldsymbol{\vartheta}_0) - \epsilon \leq \mathbf{M}(\boldsymbol{\gamma}_0, \boldsymbol{\beta}_0^*, \boldsymbol{\Lambda}_0)$$

with $\max_{j=1,\ldots,p_2} \left\|\beta_{0j}(t) - \beta_{0j}^*(t)\right\|_\infty = o(1)$. Also by the similar argument in Lu et al. (2009), we have

$$\mathbb{M}_n(\boldsymbol{\gamma}_0, \boldsymbol{\beta}_0^*, \boldsymbol{\Lambda}_0) - \mathbf{M}(\boldsymbol{\gamma}_0, \boldsymbol{\beta}_0^*, \boldsymbol{\Lambda}_0) = o_p(1)$$

and

$$\mathbb{M}_n(\boldsymbol{\gamma}_0, \boldsymbol{\beta}_0^*, \boldsymbol{\Lambda}_0) \leq \mathbb{M}_n\left(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\beta}}_n, \hat{\boldsymbol{\Lambda}}\right).$$

Then by (A.12) and the arguments similar to those used in Lu et al. (2009), we can show that $\mathbf{M}(\boldsymbol{\vartheta}_0) = \mathbf{M}\left(\boldsymbol{\vartheta}^+\right)$, that is $\boldsymbol{\beta}^+ = \boldsymbol{\beta}_0$ and $\boldsymbol{\Lambda}_k^+ = \boldsymbol{\Lambda}_{0k}$ a.e. with respect to υ_{1k} as $\boldsymbol{\vartheta}_0$ is the unique maximizer of $\mathbf{M}(\boldsymbol{\vartheta})$ for $k = 1, \ldots, K$. Since this holds for any convergent subsequence, we conclude that all the limits of subsequence of $\hat{\boldsymbol{\vartheta}}_{n_l}$ are $\boldsymbol{\vartheta}_0$. Therefore, due to the uniform boundedness of $\hat{\Lambda}_k(t)$, $k = 1, \ldots, K$, we obtain the weak consistency of $\hat{\boldsymbol{\vartheta}}_n$ in the metric d.

A.2.2 Proof of the Rate of Convergence

In (A.11), since $h(x) \geq (1/4)(x-1)^2$ for $0 \leq x \leq 5$, for $\boldsymbol{\theta}$ in a sufficiently small neighborhood of θ_0

$$\begin{aligned}
\mathbf{M}(\boldsymbol{\vartheta}_0) - \mathbf{M}(\boldsymbol{\vartheta}) \geq & \frac{1}{4}\sum_{k=1}^{K}\int \Lambda_k(u)\exp\left(\boldsymbol{\theta}^T(u)\,v(u)\right) && \text{(A.14)}\\
& \times \left\{\frac{\Lambda_{0k}(u)\exp\left(\boldsymbol{\theta}_0^T(u)\,v(u)\right)}{\Lambda_k(u)\exp\left(\boldsymbol{\theta}^T(u)\,v(u)\right)} - 1\right\}^2 d\upsilon_{1k}(u,v) \\
\geq & C\int \left\{\Lambda_k(u)\exp\left(\boldsymbol{\theta}^T(u)\,v(u)\right)\right. && \text{(A.15)}\\
& \left. - \Lambda_{0k}(u)\exp\left(\boldsymbol{\theta}_0^T(u)\,v(u)\right)\right\}^2 d\upsilon_1(u,v) && \text{(A.16)}
\end{aligned}$$

Let $\rho_k(u,z) = \Lambda_k(u)\exp\left(\boldsymbol{\beta}^T(u)\,z(u)\right)$ and $\rho_{0k}(u,z) = \Lambda_{0k}(u)\exp\left(\boldsymbol{\beta}_0^T(u)\,z(u)\right)$. We also define $\rho_{sk} = s\rho_k + (1-s)\,\rho_{0k}$, $\Lambda_{sk} = s\Lambda_k + (1-s)\,\Lambda_{0k}$, $\boldsymbol{\gamma}_s = s\boldsymbol{\gamma} + (1-s)\,\boldsymbol{\gamma}_0$, $\boldsymbol{\beta}_s = s\boldsymbol{\beta} + (1-s)\,\boldsymbol{\beta}_0$, $\boldsymbol{\theta}_s = s\boldsymbol{\theta} + (1-s)\,\boldsymbol{\theta}_0$ for $s \in (0,1)$. Then let

$$g_k(s) = \rho_{sk}(U,\mathbf{Z})\exp\left(\boldsymbol{\gamma}_s^T\mathbf{W}(U)\right).$$

and clearly

$$\Lambda_k\left(U\right)\exp\left(\boldsymbol{\theta}^T\left(U\right)\mathbf{V}(U)\right)-\Lambda_{0k}\left(U\right)\exp\left(\boldsymbol{\theta}_0^T\left(U\right)\mathbf{V}(U)\right)=g_k\left(1\right)-g_k\left(0\right).$$

By the mean value theorem, there exists a $0\leq\xi\leq1$ such that $g_k\left(1\right)-g_k\left(0\right)=g_k'\left(\xi\right)$ where

$$\begin{aligned}
&g_k'\left(\xi\right)\\
=&\exp\left(\boldsymbol{\gamma}_\xi^T\mathbf{W}(U)\right)\left\{\left(\rho_k-\rho_{0k}\right)\left(U,\mathbf{Z}\right)+\left(\xi\rho_k+\left(1-\xi\right)\rho_{0k}\right)\left(U,\mathbf{Z}\right)\left(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\right)^T\mathbf{W}(U)\right\}\\
=&\exp\left(\boldsymbol{\gamma}_\xi^T\mathbf{W}(U)\right)\\
&\times\left\{\left(\rho_k-\rho_{0k}\right)\left(U,\mathbf{Z}\right)\left\{1+\xi\left(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\right)^T\mathbf{W}(U)\right\}+\rho_{0k}\left(U,\mathbf{Z}\right)\left(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\right)^T\mathbf{W}(U)\right\}
\end{aligned}$$

From (A.16), we have

$$\begin{aligned}
&\mathbf{M}\left(\boldsymbol{\vartheta}_0\right)-\mathbf{M}\left(\boldsymbol{\vartheta}\right)\\
\geq&C\sum_{k=1}^K\int\left\{\left(\rho_k-\rho_{0k}\right)\left(u,z\right)\left\{1+\xi\left(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\right)^T w(U)\right\}\right.\\
&\left.+\rho_{0k}\left(u,z\right)\left(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\right)^T\mathbf{w}(u)\right\}^2 d\upsilon_{1k}\left(u,z,w\right)\\
=&C\sum_{k=1}^K\upsilon_{1k}\left\{g_{1k}h_k+g_{2k}\right\}^2
\end{aligned}$$

where $g_{1k}\left(U,\mathbf{V}\right)=\left(\left(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\right)^T\mathbf{W}(U)\right)\rho_{0k}\left(U,\mathbf{Z}\right)$, $g_{2k}\left(U,\mathbf{Z}\right)=\left(\rho_k-\rho_{0k}\right)\left(U,\mathbf{Z}\right)$ and

$$h\left(U,\mathbf{Z}\right)=1+\xi\left(\rho_k-\rho_{0k}\right)\left(U,\mathbf{Z}\right)/\rho_{0k}\left(U,\mathbf{Z}\right).$$

By the similar method in Wellner and Zhang (2007) and He et al. (2017), under condition (C12),

$$\begin{aligned}\mathbf{M}\left(\boldsymbol{\vartheta}_0\right)-\mathbf{M}\left(\boldsymbol{\vartheta}\right) &\geq C\sum_{k=1}^{K}\upsilon_{1k}\left\{g_{1k}h_k+g_{2k}\right\}^2\\ &\geq C_1\sum_{k=1}^{K}\left\{\upsilon_{1k}\left(g_{1k}^2\right)+\upsilon_{1k}\left(g_{2k}^2\right)\right\}.\end{aligned}$$

Similarly, by the mean value theorem and condition (C12),

$$\begin{aligned}\sum_{k=1}^{K}\upsilon_{1k}\left(g_{2k}^2\right) &= \sum_{k=1}^{K}\upsilon_{1k}\left(\left(h_{2k}g_{3k}+g_{4k}\right)^2\right)\\ &\geq C_2\sum_{k=1}^{K}\left\{\upsilon_{1k}\left(g_{3k}^2\right)+\upsilon_{1k}\left(g_{4k}^2\right)\right\}\end{aligned}$$

where $g_{3k}\left(U,\mathbf{Z}\right)=\left(\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right)^T\left(U\right)\mathbf{Z}(U)\right)\Lambda_{0k}\left(U\right)$, $g_{4k}\left(U\right)=\left(\Lambda_k-\Lambda_{0k}\right)\left(U\right)$ and $h_{2k}\left(U\right)=1+\zeta\left(\Lambda_k-\Lambda_{0k}\right)\left(U\right)/\Lambda_{0k}\left(U\right)$ for some $\zeta\in(0,1)$. Specifically,

$$\begin{aligned}\sum_{k=1}^{K}\upsilon_{1k}\left(g_{3k}^2\right) &= \sum_{k=1}^{K}\int\left(\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right)^T\left(u\right)\mathbf{Z}(u)\right)^2\Lambda_{0k}^2\left(u\right)d\upsilon_{1k}\left(u,v\right)\\ &= \int\left(\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right)^T\left(u\right)\mathbf{Z}(u)\right)^2\sum_{k=1}^{K}\Lambda_{0k}^2\left(u\right)d\upsilon_{1k}\left(u,v\right)\\ &= \int\left(\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right)^T\left(u\right)\mathbf{Z}(u)\right)^2\tilde{\upsilon}_1\left(u,v\right).\end{aligned}$$

Therefore, under conditions (C8) and (C12), since $K \geq 1$,

$$
\begin{aligned}
\mathbf{M}(\boldsymbol{\vartheta}_0) - \mathbf{M}(\boldsymbol{\vartheta}) &\geq C_2 \sum_{k=1}^{K} \left\{ \upsilon_{1k}\left(g_{1k}^2\right) + \upsilon_{1k}\left(g_{3k}^2\right) + \upsilon_{1k}\left(g_{4k}^2\right) \right\} \\
&\geq C_3 \left(K \left\| \boldsymbol{\gamma} - \gamma_0 \right\|_2^2 + \left\| \beta_1 - \beta_2 \right\|_{L_2(\tilde{\mu}_0)}^2 + \sum_{k=1}^{K} \left\| \Lambda_k - \Lambda_{0k} \right\|_{L_2(\mu_{0k})}^2 \right) \\
&\geq C_3 \left(\left\| \boldsymbol{\gamma} - \gamma_0 \right\|_2^2 + \left\| \beta_1 - \beta_2 \right\|_{L_2(\tilde{\mu}_0)}^2 + \sum_{k=1}^{K} \left\| \Lambda_k - \Lambda_{0k} \right\|_{L_2(\mu_{0k})}^2 \right) \\
&\gtrsim d^2(\boldsymbol{\vartheta}_0, \boldsymbol{\vartheta}).
\end{aligned}
$$

Next, we need to find $\varphi_n(\delta)$ such that

$$
E\left[\sup_{d(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_0) < \delta} \sqrt{n} \left| (\mathbb{P}_n - \mathbf{P}) \left(m_{\boldsymbol{\vartheta}}(O) - m_{\boldsymbol{\vartheta}_0}(O) \right) \right| \right] \leq c\varphi_n(\delta).
$$

Let

$$
\mathcal{F}_\delta = \left\{ m_{\boldsymbol{\vartheta}}(O) - m_{\boldsymbol{\vartheta}_0}(O) : d(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_0) \leq \delta \right\}
$$

By the result of Theorem 2.7.5 of van der Vaart and Wellner (1996) and Lemma A.2 of Lu et al. (2009), for any $\epsilon \leq \delta$, after some calculation, we have

$$
\log N_{[]}\left(\epsilon, \mathcal{F}_\delta, \|\cdot\|_{\mathbf{P},B} \right) \leq c \left(\frac{K}{\epsilon} + (p_1 + p_2 q_n) \log\left(\frac{\delta}{\epsilon} \right) \right),
$$

where $\|\cdot\|_{P,B}$ is the Bernstein norm defined as $\|f\|_{P,B} = \left\{ 2\mathbf{P}\left(e^{|f|} - 1 - |f| \right) \right\}^{1/2}$ by van der Vaart and Wellner (1996, page 324). Similar to the argument in Lu et al. (2009) and Wellner and Zhang (2007), under conditions (C6) and (C10), we have

$$
\left\| m_{\boldsymbol{\vartheta}}(O) - m_{\boldsymbol{\vartheta}_0}(O) \right\|_{\mathbf{P},B}^2 \leq c\delta^2,
$$

for any $m_{\boldsymbol{\vartheta}}(O) - m_{\boldsymbol{\vartheta}_0}(O) \in \mathcal{F}_\delta$. Therefore, according to Lemma 3.4.3 in van der Vaart and Wellner (1996), we can show a maximal inequality

$$E\left\|\sqrt{n}\left(\mathbb{P}_n - \mathbf{P}\right)\right\|_{\mathcal{F}_\delta} \leq cJ_{[]}\left(\delta, \mathcal{F}_\delta, \|\cdot\|_{\mathbf{P},B}\right)\left\{1 + \frac{J_{[]}\left(\delta, \mathcal{F}_\delta, \|\cdot\|_{\mathbf{P},B}\right)}{\delta^2 n^{1/2}}\right\}$$

where

$$\begin{aligned} J_{[]}\left(\delta, \mathcal{F}_\delta, \|\cdot\|_{\mathbf{P},B}\right) &= \int_0^\delta \left\{1 + \log N_{[]}\left(\epsilon, \mathcal{F}_\delta, \|\cdot\|_{\mathbf{P},B}\right)\right\}^{1/2} d\epsilon \\ &\leq c_1 q_n^{1/2} \int_0^\delta \left\{1 + \frac{K}{\epsilon} + \log\left(\frac{\delta}{\epsilon}\right)\right\}^{1/2} d\epsilon \\ &\leq q_n^{1/2}\delta^{1/2} \end{aligned}$$

Thus,

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

Obviously, $\varphi_n(\delta)/\delta$ is decreasing in δ as the leading term is $\delta^{-1/2}$. Therefore

$$r_n^2 \varphi_n\left(\frac{1}{r_n}\right) = r_n^{3/2} q_n^{1/2} + r_n^3 q_n n^{-1/2} \lesssim n^{1/2}$$

if $r_n = \min\left\{n^{\frac{1-\nu}{3}}, n^{r\nu}\right\}$ and $0 < \nu < 1/2$. Moreover, using similar argument in Lu et al. (2009), we can show $\mathbb{M}_n\left(\hat{\boldsymbol{\vartheta}}_n\right) - \mathbb{M}_n\left(\boldsymbol{\vartheta}_0\right) > -O_p\left(n^{-2r\nu}\right) \geq O_p\left(r_n^2\right)$. Then, by Theorem 3.2.5 of Wellner and Zhang (2007), we have $r_n d\left(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_0\right) = O_p(1)$. If ν is chosen as $1/(3r+1)$, we obtain the optimal rate $n^{r/(3r+1)}$ because $(1-\nu)/3 = r\nu$.

A.2.3 Proof of the Asymptotic Normality

We mainly use the method in He et al. (2017). We define a sequence of maps S_n

mapping a neighborhood of $\boldsymbol{\vartheta}_0$, denoted by $\mathcal{U}$, in the parameter space for $\boldsymbol{\vartheta}$ into $l^\infty(\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3)$ as

$$
\begin{aligned}
& S_n(\boldsymbol{\vartheta})[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3] \\
=& \frac{d}{d\epsilon}\, l_n(\boldsymbol{\gamma} + \epsilon \boldsymbol{h}_1, \boldsymbol{\beta} + \epsilon \boldsymbol{h}_2, \boldsymbol{\Lambda} + \epsilon h_3)\big|_{\epsilon=0} \\
=& n^{-1} \sum_{k=1}^{K} \sum_{i=1}^{n} \int_0^\tau Y_i(t) \left\{ \frac{N_{ik}(t)}{\Lambda_k(t)} h_{3k}(t) + N_{ik}(t) \left(\boldsymbol{h}_1^T \mathbf{W}_i(t) + \boldsymbol{h}_2^T(t) \mathbf{Z}_i(t) \right) \right. \\
& - \left(\boldsymbol{h}_1^T \mathbf{W}_i(t) + \boldsymbol{h}_2^T(t) \mathbf{Z}_i(t) \right) \exp\left(\boldsymbol{\gamma}^T \mathbf{W}_i(t) + \boldsymbol{\beta}^T(t) \mathbf{Z}_i(t) \right) \Lambda_k(t) \\
& \left. - \exp\left(\boldsymbol{\gamma}^T \mathbf{W}_i(t) + \boldsymbol{\beta}^T \mathbf{Z}_i(t) \right) h_{3k}(t) \right\} dH_{ik}(t) \\
=& A_{n1}(\boldsymbol{\vartheta})[\boldsymbol{h}_1] + A_{n2}(\boldsymbol{\vartheta})[\boldsymbol{h}_2] + A_{n3}(\boldsymbol{\vartheta})[h_3] \\
=& \mathbb{P}_n \psi(\boldsymbol{\vartheta})[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3],
\end{aligned}
$$

where

$$
\begin{aligned}
A_{n1}(\boldsymbol{\vartheta})[\boldsymbol{h}_1] = n^{-1} \sum_{k=1}^{K} \sum_{i=1}^{n} \int_0^\tau & Y_i(t) \boldsymbol{h}_1^T \mathbf{W}_i(t) \{ N_{ik}(t) \\
& - \exp\left(\boldsymbol{\gamma}^T \mathbf{W}_i(t) + \boldsymbol{\beta}^T(t) \mathbf{Z}_i(t) \right) \Lambda_k(t) \} dH_{ik}(t),
\end{aligned}
$$

$$
\begin{aligned}
A_{n2}(\boldsymbol{\vartheta})[\boldsymbol{h}_2] = n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_0^\tau & Y_i(t) \boldsymbol{h}_2^T(t) \mathbf{Z}_i(t) \{ N_{ik}(t) \\
& - \exp\left(\boldsymbol{\gamma}^T \mathbf{W}_i(t) + \boldsymbol{\beta}^T(t) \mathbf{Z}_i(t) \right) \Lambda_k(t) \} dH_{ik}(t),
\end{aligned}
$$

and

$$A_{n3}(\boldsymbol{\vartheta})[h_3] = n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau} Y_i(t)h_{3k}(t)\left\{\frac{N_{ik}(t)}{\Lambda_k(t)} - \exp\left(\boldsymbol{\gamma}^T\mathbf{W}_i(t) + \boldsymbol{\beta}^T\mathbf{Z}_i(t)\right)\right\} dH_{ik}(t).$$

Correspondingly, we define the limit map $S : \mathcal{U} \to l^{\infty}(\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3)$ as

$$S(\boldsymbol{\vartheta})[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3] = A_1(\boldsymbol{\vartheta})[\boldsymbol{h}_1] + A_2(\boldsymbol{\vartheta})[\boldsymbol{h}_2] + A_3(\boldsymbol{\vartheta})[h_3]$$

where

$$A_1(\boldsymbol{\vartheta})[\boldsymbol{h}_1] = \mathbf{P}\sum_{k=1}^{K}\int_0^{\tau} Y(t)\boldsymbol{h}_1^T\mathbf{W}(t)\left\{N_{\cdot k}(t) - \exp\left(\boldsymbol{\gamma}^T\mathbf{W}(t) + \boldsymbol{\beta}^T(t)\mathbf{Z}(t)\right)\Lambda_k(t)\right\} dH_{\cdot k}(t),$$

$$A_2(\boldsymbol{\vartheta})[\boldsymbol{h}_2] = \mathbf{P}\sum_{k=1}^{K}\int_0^{\tau} Y(t)\boldsymbol{h}_2^T(t)\mathbf{Z}(t) \left\{N_{\cdot k}(t) - \exp\left(\boldsymbol{\gamma}^T\mathbf{W}(t) + \boldsymbol{\beta}^T(t)\mathbf{Z}(t)\right)\Lambda_k(t)\right\} dH_{\cdot k}(t),$$

and

$$A_3(\boldsymbol{\vartheta})[h_3] = \mathbf{P}\sum_{k=1}^{K}\int_0^{\tau} Y(t)h_{3k}(t)\left\{\frac{N_{\cdot k}(t)}{\Lambda_k(t)} - \exp\left(\boldsymbol{\gamma}^T\mathbf{W}(t) + \boldsymbol{\beta}^T\mathbf{Z}(t)\right)\right\} dH_{\cdot k}(t).$$

To derive the asymptotic normality of $\hat{\boldsymbol{\vartheta}}_n$, we need to verify the following five conditions in He et al. (2017).

(a1) $\sqrt{n}\left(S_n - S\right)\left(\hat{\boldsymbol{\vartheta}}_n\right) - \sqrt{n}\left(S_n - S\right)\left(\boldsymbol{\vartheta}_0\right) = o_p\left(1\right)$.

(a2) $S\left(\boldsymbol{\vartheta}_0\right) = 0$ and $S_n\left(\hat{\boldsymbol{\vartheta}}_n\right) = o_p\left(n^{-1/2}\right)$.

(a3) $\sqrt{n}\left(S_n - S\right)\left(\theta_0\right)$ converges in distribution to a tight Gaussian process on $l^\infty(\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3)$.

(a4) $S\left(\boldsymbol{\vartheta}\right)$ is Fréchet-differentiable at $\boldsymbol{\vartheta}_0$ denoted by $\dot{S}\left(\boldsymbol{\vartheta}_0\right)$.

(a5) $S\left(\hat{\boldsymbol{\vartheta}}_n\right) - S\left(\boldsymbol{\vartheta}_0\right) - \dot{S}\left(\boldsymbol{\vartheta}_0\right)\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right) = o_p\left(n^{-1/2}\right)$.

Using similar argument in Lu et al. (2009), it is not hard to show

$$\left\{\psi\left(\boldsymbol{\vartheta}\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right] - \psi\left(\boldsymbol{\vartheta}_0\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right] : d\left(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_0\right) < \delta, \left(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right) \in \mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3\right\}$$

is a Donkser class for some δ. Therefore,

$$\sup_{\left(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right) \in \mathcal{A} \times \mathcal{M} \times \mathcal{F}} \mathbf{P}\left\{\psi\left(\hat{\boldsymbol{\vartheta}}_n\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right] - \psi\left(\boldsymbol{\vartheta}_0\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right]\right\}^2 \to 0$$

as $d\left(\hat{\boldsymbol{\vartheta}}_n, \boldsymbol{\vartheta}_0\right) \to 0$ in probability and thus (a1) holds.

For (a2), clearly, $S\left(\boldsymbol{\vartheta}_0\right) = 0$. For $\boldsymbol{h}_2 \in \mathcal{H}_2$, let $\boldsymbol{h}_{2n}$ be the B-spline function approximation of $\boldsymbol{h}_2$ with $\max_{j=1,\ldots,p_2} \left\|h_{2j} - h_{2nj}\right\|_\infty = O\left(n^{-\nu r}\right)$ by (A.13). Then we have $S_n\left(\hat{\boldsymbol{\vartheta}}_n\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_{2n}, h_3\right] = 0$. Thus,

$$\begin{aligned} S_n\left(\hat{\boldsymbol{\vartheta}}_n\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right] &= \sqrt{n}\mathbb{P}_n\psi\left(\hat{\boldsymbol{\vartheta}}_n\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right] - \sqrt{n}\mathbf{P}\psi\left(\hat{\boldsymbol{\vartheta}}_n\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_{2n}, h_3\right] \\ &= I_{n1} - I_{n2} + I_{n3} + I_{n4} \end{aligned}$$

where

$$I_{n1} = \sqrt{n}\left(\mathbb{P}_n - \mathbf{P}\right)\left\{\psi\left(\hat{\boldsymbol{\vartheta}}_n\right)[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3] - \psi\left(\boldsymbol{\vartheta}_0\right)[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3]\right\}$$

$$I_{n2} = \sqrt{n}\left(\mathbb{P}_n - \mathbf{P}\right)\left\{\psi\left(\hat{\boldsymbol{\vartheta}}_n\right)[\boldsymbol{h}_1, \boldsymbol{h}_{2n}, h_3] - \psi\left(\boldsymbol{\vartheta}_0\right)[\boldsymbol{h}_1, \boldsymbol{h}_{2n}, h_3]\right\}$$

$$I_{n3} = \sqrt{n}\mathbb{P}_n\left\{\psi\left(\boldsymbol{\vartheta}_0\right)[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3] - \psi\left(\boldsymbol{\vartheta}_0\right)[\boldsymbol{h}_1, \boldsymbol{h}_{2n}, h_3]\right\}$$

and

$$I_{n4} = \sqrt{n}\mathbf{P}\left\{\psi\left(\hat{\boldsymbol{\vartheta}}_n\right)[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3] - \psi\left(\hat{\boldsymbol{\vartheta}}_n\right)[\boldsymbol{h}_1, \boldsymbol{h}_{2n}, h_3]\right\}.$$

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

$$\begin{aligned}
I_{n3} =& \sqrt{n}\left(\mathbb{P}_n - \mathbf{P}\right)\left\{\psi\left(\boldsymbol{\vartheta}_0\right)[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3] - \psi\left(\boldsymbol{\vartheta}_0\right)[\boldsymbol{h}_1, \boldsymbol{h}_{2n}, h_3]\right\} \\
&+ \sqrt{n}\mathbf{P}\left\{\psi\left(\boldsymbol{\vartheta}_0\right)[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3] - \psi\left(\boldsymbol{\vartheta}_0\right)[\boldsymbol{h}_1, \boldsymbol{h}_{2n}, h_3]\right\} \\
=& I_{n31} + I_{n32}.
\end{aligned}$$

Similarly to proving (a1), $I_{n31} = o_p(1)$ and $I_{n32} = 0$ since $S\left(\boldsymbol{\vartheta}_0\right) = 0$ for any $h_{2,} h_{2n} \in \mathcal{H}_2$. For I_{n4},

$$\begin{aligned}
|I_{n4}| &\leq \sqrt{n} d\left(\hat{\boldsymbol{\vartheta}}_n, \theta_0\right)\left(\max_{j=1,\ldots,p_2}\|h_{2j} - h_{2nj}\|_\infty\right) \\
&= O_p\left(\max\left\{n^{-(1-\nu)/3}, n^{-r\nu}\right\} n^{-rv+1/2}\right) \\
&= o_p(1)
\end{aligned}$$

if $1/(4r) < \nu < 1/2$. Thus (a2) holds.

Condition (a3) holds because $\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$ is a Donsker class and the functionals

$A_1(\boldsymbol{\vartheta})[\boldsymbol{h}_1]$, $A_2(\boldsymbol{\vartheta})[\boldsymbol{h}_2]$ and $A_3(\boldsymbol{\vartheta})[h_3]$ are bounded Lipschitz functions with respect to $\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$ due the compactness of $\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$.

For (a4), by the smoothness of $S(\boldsymbol{\vartheta})$ the Fréchet differentiability holds and the derivative of $S(\boldsymbol{\vartheta})$ at $\boldsymbol{\vartheta}_0$, denoted by $\dot{S}(\boldsymbol{\vartheta}_0)$ is a map from the space $\{\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0 : \boldsymbol{\vartheta} \in \mathcal{U}\}$ to $l^\infty(\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3)$. Now we calculate $\dot{S}(\boldsymbol{\vartheta}_0)$ as

$$
\begin{aligned}
&\dot{S}(\boldsymbol{\vartheta}_0)(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0)[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3] \\
=&\frac{d}{d\epsilon}\left\{A_1(\boldsymbol{\vartheta}_0 + \epsilon(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0))[\boldsymbol{h}_1]\right\}|_{\epsilon=0} \\
&+ \frac{d}{d\epsilon}\left\{A_2(\boldsymbol{\vartheta}_0 + \epsilon(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0))[\boldsymbol{h}_2]\right\}|_{\epsilon=0} \\
&+ \frac{d}{d\epsilon}\left\{A_3(\boldsymbol{\vartheta}_0 + \epsilon(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0))[h_3]\right\}|_{\epsilon=0} \\
=&-\mathbf{P}\sum_{k=1}^{K}\int_0^\tau Y(t)\boldsymbol{h}_1^T\mathbf{W}(t)\exp\left(\boldsymbol{\gamma}^T\mathbf{W}(t) + \boldsymbol{\beta}^T(t)\,\mathbf{Z}(t)\right) \\
&\times\left\{\left((\boldsymbol{\gamma} - \gamma_0)^T\mathbf{W}(t) + (\boldsymbol{\beta}(t) - \boldsymbol{\beta}_0(t))^T\mathbf{Z}(t)\right)\Lambda_{0k}(t) + (\Lambda_k(t) - \Lambda_{0k}(t))\right\}dH_{\cdot k}(t) \\
&-\mathbf{P}\sum_{k=1}^{K}\int_0^\tau Y(t)\boldsymbol{h}_2^T(t)\,\mathbf{Z}(t)\exp\left(\boldsymbol{\gamma}^T\mathbf{W}(t) + \boldsymbol{\beta}^T(t)\,\mathbf{Z}(t)\right) \\
&\times\left\{\left((\boldsymbol{\gamma} - \gamma_0)^T\mathbf{W}(t) + (\boldsymbol{\beta}(t) - \boldsymbol{\beta}_0(t))^T\mathbf{Z}(t)\right)\Lambda_{0k}(t) + (\Lambda_k(t) - \Lambda_{0k}(t))\right\}dH_{\cdot k}(t) \\
&-\mathbf{P}\sum_{k=1}^{K}\int_0^\tau Y(t)h_{3k}(t)\exp\left(\boldsymbol{\gamma}^T\mathbf{W}(t) + \boldsymbol{\beta}^T\mathbf{Z}(t)\right) \\
&\times\left\{\frac{\Lambda_k(t) - \Lambda_{0k}(t)}{\Lambda_{0k}(t)} + (\boldsymbol{\gamma} - \gamma_0)^T\mathbf{W}(t) + (\boldsymbol{\beta}(t) - \boldsymbol{\beta}_0(t))^T\mathbf{Z}(t)\right\}dH_{\cdot k}(t).
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
&\dot{S}\left(\boldsymbol{\vartheta}_0\right)\left(\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right] \\
= &\left(\boldsymbol{\gamma}-\gamma_0\right)^T Q_1\left(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right) \\
&+\int_0^\tau\left(\boldsymbol{\beta}(t)-\boldsymbol{\beta}_0(t)\right)^T dQ_2\left(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right)(t) \\
&+\sum_{k=1}^K \int_0^\tau\left(\Lambda_k(t)-\Lambda_{0k}(t)\right) dQ_{3k}\left(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right)(t)
\end{aligned}
$$

where

$$
\begin{aligned}
&Q_1\left(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right) \\
&\quad=-\mathbf{P} \sum_{k=1}^K \int_0^\tau \mathbf{W}(t) Y(t)\left(\boldsymbol{h}_1^T \mathbf{W}(t) \Lambda_{0k}(t)+\boldsymbol{h}_2^T(t) \mathbf{Z}(t) \Lambda_{0k}(t)+h_{3k}(t)\right) \\
&\qquad\qquad \times \exp\left(\boldsymbol{\gamma}^T \mathbf{W}(t)+\boldsymbol{\beta}^T(t) \mathbf{Z}(t)\right) dH_{\cdot k}(t),
\end{aligned}
$$

$$
\begin{aligned}
&dQ_2\left(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right)(t)= \\
&\quad-\mathbf{P} \sum_{k=1}^K \mathbf{Z}(t) Y(t)\left(\boldsymbol{h}_1^T \mathbf{W}(t) \Lambda_{0k}(t)+\boldsymbol{h}_2^T(t) \mathbf{Z}(t) \Lambda_{0k}(t)+h_{3k}(t)\right) \\
&\qquad\qquad \times \exp\left(\boldsymbol{\gamma}^T \mathbf{W}(t)+\boldsymbol{\beta}^T(t) \mathbf{Z}(t)\right) dH_{\cdot k}(t),
\end{aligned}
$$

and

$$dQ_{3k}\left(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right)(t) = \\ - \mathbf{P}\frac{Y(t)}{\Lambda_{0k}(t)}\left(\boldsymbol{h}_1^T\mathbf{W}(t)\Lambda_{0k}(t) + \boldsymbol{h}_2^T(t)\mathbf{Z}(t)\Lambda_{0k}(t) + h_{3k}(t)\right) \\ \times \exp\left(\boldsymbol{\gamma}^T\mathbf{W}(t) + \boldsymbol{\beta}^T(t)\mathbf{Z}(t)\right) dH_{\cdot k}(t).$$

We can also show $Q = (Q_1, Q_2, Q_3)$ is one-to-one by the similar method in He et al. (2017).

For (a5), we have

$$S\left(\hat{\boldsymbol{\vartheta}}_n\right) - S\left(\boldsymbol{\vartheta}_0\right) - \dot{S}\left(\boldsymbol{\vartheta}_0\right)\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right) = B_{n1} + B_{n2} + B_{n3}$$

where

$$B_{n1} = A_1\left(\hat{\boldsymbol{\vartheta}}_n\right)[\boldsymbol{h}_1] - \frac{d}{d\varepsilon}\left\{A_1\left(\boldsymbol{\vartheta}_0 + \varepsilon\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right)\right)[\boldsymbol{h}_1]\right\}\Big|_{\varepsilon=0},$$

$$B_{n2} = A_2\left(\hat{\boldsymbol{\vartheta}}_n\right)[\boldsymbol{h}_2] - \frac{d}{d\varepsilon}\left\{A_2\left(\boldsymbol{\vartheta}_0 + \varepsilon\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right)\right)[\boldsymbol{h}_2]\right\}\Big|_{\varepsilon=0},$$

and

$$B_{n3} = A_3\left(\hat{\boldsymbol{\vartheta}}_n\right)[h_3] - \frac{d}{d\varepsilon}\left\{A_3\left(\boldsymbol{\vartheta}_0 + \varepsilon\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right)\right)[h_3]\right\}\Big|_{\varepsilon=0}.$$

It is not hard to see

$$B_{n1} = \mathbf{P}\sum_{k=1}^{K}\int_0^\tau Y(t)\boldsymbol{h}_1^T\mathbf{W}(t)\exp\left(\boldsymbol{\gamma}_0^T\mathbf{W}(t) + \boldsymbol{\beta}_0^T(t)\mathbf{Z}(t)\right)\Lambda_{0k}(t) \\ \times q_1\left(\left(\hat{\boldsymbol{\gamma}} - \gamma_0\right)^T\mathbf{W}(t) - \left(\hat{\boldsymbol{\beta}}_n(t) - \boldsymbol{\beta}_0(t)\right)^T\mathbf{Z}(t)\right)dH_{\cdot k}(t),$$

where $q_1(x) = 1 - \exp(y)(1-y)$ and $q_1(x) \leq x^2$ when x is in a neighborhood of 0. Thus

$$\begin{aligned}
B_{n1} \leq & \mathbf{P} \sum_{k=1}^{K} \int_0^{\tau} Y(t) \boldsymbol{h}_1^T \mathbf{W}(t) \exp\left(\boldsymbol{\gamma}_0^T \mathbf{W}(t) + \boldsymbol{\beta}_0^T(t) \mathbf{Z}(t)\right) \Lambda_{0k}(t) \\
& \times \left\{ (\hat{\boldsymbol{\gamma}} - \gamma_0)^T \mathbf{W}(t) - \left(\hat{\boldsymbol{\beta}}_n(t) - \boldsymbol{\beta}_0(t)\right)^T \mathbf{Z}(t) \right\}^2 dH_{\cdot k}(t). \\
= & O\left(d^2\left(\hat{\boldsymbol{\vartheta}}_n, \boldsymbol{\vartheta}\right)\right).
\end{aligned}$$

Similarly, we can show $B_{n2} \leq O\left(d^2\left(\hat{\boldsymbol{\vartheta}}_n, \boldsymbol{\vartheta}\right)\right)$ and $B_{n3} \leq O\left(d^2\left(\hat{\boldsymbol{\vartheta}}_n, \boldsymbol{\vartheta}\right)\right)$ and hence

$$S\left(\hat{\boldsymbol{\vartheta}}_n\right) - S(\boldsymbol{\vartheta}_0) - \dot{S}(\boldsymbol{\vartheta}_0)\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right) \leq O\left(d^2\left(\hat{\boldsymbol{\vartheta}}_n, \boldsymbol{\vartheta}\right)\right).$$

Since $n^{1/2} d^2\left(\hat{\boldsymbol{\vartheta}}_n, \boldsymbol{\vartheta}\right) = O_p\left(n^{1/2} \max\left\{n^{-2(1-\nu)/3}, n^{-2r\nu}\right\}\right) = o_p(1)$ if $1/(4r) < v < 1/4$, we can conclude that $S\left(\hat{\boldsymbol{\vartheta}}_n\right) - S(\boldsymbol{\vartheta}_0) - \dot{S}(\boldsymbol{\vartheta}_0)\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right) = o_p\left(n^{-1/2}\right)$ and (a5) holds.

If (a1)-(a5) hold, according to He et al. (2017), we have

$$-\sqrt{n}\dot{S}(\boldsymbol{\vartheta}_0)\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right)[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3] = \sqrt{n}(S_n - S)(\boldsymbol{\vartheta}_0)[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3] + o_p(1),$$

uniformly in $\boldsymbol{h}_1, \boldsymbol{h}_2, h_3$. For each $(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3) \in \mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$, Q is invertible by the similar arguement in He et al. (2017). Then there exists $(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3) \in \mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$ such that

$$Q_1(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3) = \boldsymbol{h}_1, Q_2(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3) = \boldsymbol{h}_2, Q_{3k}(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3) = h_{3k}.$$

Therefore, we have

$$\sqrt{n}\left(\hat{\gamma}-\gamma_0\right)^T \boldsymbol{h}_1+\sqrt{n}\int_0^\tau\left(\hat{\boldsymbol{\beta}}_n(t)-\boldsymbol{\beta}_0(t)\right)^T d\boldsymbol{h}_2(t)$$

$$+\sqrt{n}\sum_{k=1}^K\int_0^\tau\left(\hat{\Lambda}_k(t)-\Lambda_{0k}(t)\right)dh_{3k}(t)$$

$$=\sqrt{n}\left(S_n-S\right)\left(\boldsymbol{\vartheta}_0\right)\left[\boldsymbol{h}_1,\boldsymbol{h}_2,h_3\right]+o_p(1)\rightarrow_d N\left(0,\sigma^2\right)$$

where $\sigma^2=E\left[\psi^2\left(\boldsymbol{\vartheta}_0\right)\left[\boldsymbol{h}_1,\boldsymbol{h}_2,h_3\right]\right]$ because of (a3). To find the asymptotic distribution of γ only, we can find $\boldsymbol{h}_1$, $\boldsymbol{h}_2$ and h_3 as a solution of $Q_2=0$ and $Q_3=0$. Unfortunately, we cannot find the explicit forms of $\boldsymbol{h}_1$, $\boldsymbol{h}_2$ and h_3 as He et al. (2017). Hence, we adopt the variance estimation method in the main body.

A.3 Proofs of Theorems 7, 8, and 9

In this section, we will sketch the proofs of the consistency and asymptotic properties of the proposed estimator $\hat{\boldsymbol{\vartheta}}$ described in Theorems 7, 8, and 9 of Chapter 4.

A.3.1 Proof of Consistency

We first prove that the true parameter $\vartheta_0=\left(\gamma_0,\beta_0,\Lambda_0\right)$ is the unique maximizer of $\mathbf{M}(\vartheta)$. From the definition of $\mathbf{M}(\vartheta)$,

$$\mathbf{M}\left(\vartheta_0\right)-\mathbf{M}(\vartheta)=\int\int_{u_1}^{u_2}\exp\left(\gamma^\top w(t)+\beta^\top(t)\,x(t)\right)d\Lambda(t)$$

$$h\left[\frac{\int_{u_1}^{u_2}\exp\left(\gamma_0^\top w(t)+\beta_0^\top(t)\,x(t)\right)d\Lambda_0(t)}{\int_{u_1}^{u_2}\exp\left(\gamma_0^\top w(t)+\beta^\top(t)\,x(t)\right)d\Lambda(t)}\right]d\nu_1\left(u_1,u_2,x\right)$$

where $h(z)=z\log z-z+1$. The function $h(z)$ is nonnegative for all $z>0$ with equality holding only at $z = 1$. Therefore, $\mathbf{M}(\vartheta_0) \geq \mathbf{M}(\vartheta)$ and $\mathbf{M}(\vartheta_0) = \mathbf{M}(\vartheta)$ if and only if

$$\int_{u_1}^{u_2} \exp\left(\gamma_0^\top w(t) + \beta_0^\top(t)\, x(t)\right) d\Lambda_0(t) = \int_{u_1}^{u_2} \exp\left(\gamma^\top w(t) + \beta^\top(t)\, x(t)\right) d\Lambda(t) \tag{A.17}$$

almost everywhere (a.e.) with respect to ν_1. Under condition (C1), (A.17) is equivalent to

$$\begin{aligned}\int\int_{u_1}^{u_2} \exp\left(\gamma_0^\top w(t) + \beta_0^\top(t)\, x(t)\right) \Lambda_0'(t)\, dt d\nu_1(u_1, u_2, x) = \\ \int\int_{u_1}^{u_2} \exp\left(\gamma^\top w(t) + \beta^\top(t)\, x(t)\right) \Lambda'(t)\, dt d\nu_1(u_1, u_2, x) \\ \int \exp\left(\gamma_0^\top w(t) + \beta_0^\top(t)\, x(t)\right) \Lambda_0'(t)\, d\nu_2(t, x) = \\ \int \exp\left(\gamma^\top w(t) + \beta^\top(t)\, x(t)\right) \Lambda'(t)\, d\nu_2(t, x)\end{aligned}$$

that is

$$\exp\left(\gamma_0^\top w(t) + \beta_0^\top(t)\, x(t)\right) \Lambda_0'(t) = \exp\left(\gamma^\top w(t) + \beta^\top(t)\, x(t)\right) \Lambda'(t)$$

a.e. with respect to ν_2. By the similar argument in Wellner and Zhang (2007, page 2123), we have

$$\left\{1 - \exp\left((\gamma_0 - \gamma)^\top w(t) + (\beta_0(t) - \beta(t))^\top x(t)\right)\right\}^2 = \left(\frac{\Lambda'(t)}{\Lambda_0'(t)} - 1\right)^2$$

a.e. with respect to ν_2. This leads to $\gamma_0 = \gamma$, $\beta_0(t) = \beta(t)$ and $\Lambda_0'(t) = \Lambda'(t)$ in light of (C3) a.e. with respect to ν_2. The latter equality also implies $\Lambda_0(t) = \Lambda(t)$

a.e. with respect to ν_2. Hence, we can conclude that $\vartheta_0 = (\gamma_0, \beta_0, \Lambda_0)$ is the unique maximizer of $\mathbf{M}(\vartheta)$.

Next, we need to show $\hat{\Lambda}(t)$ is uniformly bounded almost surely for $t \in [0, \tau]$. For any given $\epsilon > 0$, let $\tilde{\vartheta}_\epsilon = \left(\hat{\gamma}, \hat{\beta}, (1-\epsilon)\hat{\Lambda} + \epsilon\Lambda_0\right) = \hat{\vartheta} + \epsilon\left(0, \Lambda_0 - \hat{\Lambda}_n\right)$. Since $\hat{\vartheta}_n$ maximizes $\mathbb{M}_n$, we have $\mathbb{M}_n\left(\hat{\vartheta}\right) \geq \mathbb{M}_n\left(\tilde{\vartheta}_\epsilon\right) = \mathbb{M}_n\left(\hat{\vartheta} + \epsilon\left(0, 0, \Lambda_0 - \hat{\Lambda}_n\right)\right)$. Therefore,

$$
\begin{aligned}
0 \geq & \lim_{\epsilon \to 0} \frac{\mathbb{M}_n\left(\hat{\vartheta} + \epsilon\left(0, 0, \Lambda_0 - \hat{\Lambda}_n\right)\right) - \mathbb{M}_n\left(\hat{\vartheta}\right)}{\epsilon} \\
& = \mathbb{P}_n \sum_{j=1}^{J} \left\{ \frac{\Delta N_{\cdot j}}{\int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\hat{\gamma}^\top W(t) + \hat{\beta}^\top(t) X(t)\right) d\hat{\Lambda}(t)} - 1 \right\} \\
& \int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\hat{\gamma}^\top W(t) + \hat{\beta}^\top(t) X(t)\right) d\left(\Lambda_0 - \hat{\Lambda}\right)(t).
\end{aligned}
$$

This implies, by conditions (C1), (C2) and (C4),

$$
\begin{aligned}
& \mathbb{P}_n \sum_{j=1}^{J} \left\{ \frac{\Delta N_{\cdot j} \int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\hat{\gamma}^T W(t) + \hat{\beta}^\top(t) X(t)\right) d\Lambda_0(t)}{\int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\hat{\gamma}^T W(t) + \hat{\beta}^\top(t) X(t)\right) d\hat{\Lambda}(t)} + \right. \\
& \left. \int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\hat{\gamma}^T W(t) + \hat{\beta}^\top(t) X(t)\right) d\hat{\Lambda}(t) \right\} \\
\leq & \mathbb{P}_n \sum_{j=1}^{J} \left\{ \Delta N_{\cdot j} + \int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\hat{\gamma}^T W(t) + \hat{\beta}^\top(t) X(t)\right) d\Lambda_0(t) \right\} \\
\leq & C\mathbb{P}_n \sum_{j=1}^{J} \left\{ \Delta N_{\cdot j} + \Lambda_0(T_{\cdot j}) - \Lambda_0\left(T_{\cdot(j-1)}\right) \right\} \\
= & C\mathbb{P}_n \left\{ \sum_{j=1}^{J} \Delta N_{\cdot j} + \Lambda_0(T_{\cdot J}) \right\} \to_{\text{a.s.}} C\mathbf{P} \left\{ \sum_{j=1}^{J} \Delta N_{\cdot j} + \Lambda_0(T_{\cdot J}) \right\}
\end{aligned}
$$

for some finite constant C. The limit on the right hand is finite under conditions

(C1), (C2) and (C4). On the other hand,

$$
\begin{aligned}
&\limsup_{n\to\infty} \mathbb{P}_n \sum_{j=1}^{J} \left\{ \frac{\Delta N_{\cdot j} \int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\hat{\gamma}^T W(t) + \hat{\beta}^\top(t) X(t)\right) d\Lambda_0(t)}{\int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\hat{\gamma}^T W(t) + \hat{\beta}^\top(t) X(t)\right) d\hat{\Lambda}(t)} + \right. \\
&\left. \int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\hat{\gamma}^T W(t) + \hat{\beta}^\top(t) X(t)\right) d\hat{\Lambda}(t) \right\} \\
\geq &\limsup_{n\to\infty} \mathbb{P}_n \sum_{j=1}^{J} \int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\hat{\gamma}^T W(t) + \hat{\beta}^\top(t) X(t)\right) d\hat{\Lambda}(t) \\
= &\limsup_{n\to\infty} \mathbb{P}_n \int_0^{T_{\cdot J}} \exp\left(\hat{\gamma}^T W(t) + \hat{\beta}^\top(t) X(t)\right) d\hat{\Lambda}(t).
\end{aligned}
$$

Because $\Delta N_{\cdot j}$ and $\int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\hat{\gamma}^T W(t) + \hat{\beta}^\top(t) X(t)\right) d\Lambda_0(t)$ are both nonnegative. Under conditions (C1), (C2) and (C4), for any $0 < \xi < \tau$,

$$
\begin{aligned}
&\limsup_{n\to\infty} \mathbb{P}_n \int_0^{T_{\cdot J}} \exp\left(\hat{\gamma}^T W(t) + \hat{\beta}^\top(t) X(t)\right) d\hat{\Lambda}(t) \\
\gtrsim &\limsup_{n\to\infty} \mathbb{P}_n \int_0^{T_{\cdot J}} d\hat{\Lambda}(t) \\
= &\limsup_{n\to\infty} \mathbb{P}_n \hat{\Lambda}(T_{\cdot J}) \\
\geq &\limsup_{n\to\infty} \mathbb{P}_n I\{\xi \leq T_{\cdot J} \leq \tau\} \hat{\Lambda}(T_{\cdot J}) \\
\geq &\limsup_{n\to\infty} \hat{\Lambda}(\xi) \mathbb{P}_n I\{\xi \leq T_{\cdot J} \leq \tau\} \\
= &\limsup_{n\to\infty} \hat{\Lambda}(\xi) \nu_3([\xi, \tau]).
\end{aligned}
$$

The right hand side of the last equality is obviously finite and hence $\hat{\Lambda}(t)$ is uniformly bounded almost surely for $t \in [0, \tau]$ if $\nu_3(\{\tau\}) > 0$ from (C4).

Since $\hat{\Lambda}(t)$ is non-decreasing, $\hat{\Lambda}(t)$ is also of totally bounded variation. Then, by the Helly selection theorem and the compactness of $\mathcal{B} \times \mathcal{F}$, for any subsequence of $\hat{\vartheta}$, there exists a further subsequence $\hat{\vartheta}_{n_l}$ such that $\hat{\vartheta}_{n_l}$ converges to $\vartheta^+ = (\gamma^+, \beta^+, \Lambda^+)$ as $l \to \infty$, where Λ^+ is an nondecreasing bounded function defined on $[0, \tau]$.

Under (C6), $m_\vartheta (O) \leq m_0 (O)$ with $\mathbf{P} m_0 < \infty$. Moreover, m_ϑ is upper semi-continous in $\mathcal{F}$ for each O and $\mathcal{B} \times \mathcal{F}$ is compact. Therefore, by the one-sided Gilvenko-Cantelli theorem in Wellner, Zhang, et al. (2000),

$$\limsup_{n \to \infty} \sup_{\vartheta \in \mathcal{B} \times \mathcal{F}} (\mathbb{P} - \mathbf{P}) m_\vartheta \leq 0.$$

almost surely. This inequality implies, with probability 1,

$$\limsup_{n_l \to \infty} \mathbb{M}_n \left(\hat{\vartheta}_{n_l} \right) \leq \limsup_{n_l \to \infty} \mathbf{M} \left(\hat{\vartheta}_{n_l} \right) = \mathbf{M} \left(\vartheta^+ \right)$$

by the continuity of $\mathbf{M}$ in ϑ and the Dominated Convergence Theorem. By the law of large numbers, $\mathbb{M}_n (\vartheta_0) \to \mathbf{M} (\vartheta_0)$ a.s. and by the fact that $\mathbb{M}_n (\vartheta_0) \leq \mathbb{M}_n \left(\hat{\vartheta} \right)$, we have, with probability 1,

$$\begin{aligned} \mathbf{M} (\vartheta_0) &\leq \liminf_{n \to \infty} \mathbb{M}_n \left(\hat{\vartheta} \right) \\ &\leq \liminf_{n \to \infty} \mathbb{M}_n \left(\hat{\vartheta}_{n_l} \right) \\ &\leq \limsup_{n_l \to \infty} \mathbb{M}_n \left(\hat{\vartheta}_{n_l} \right) \\ &\leq \mathbf{M} \left(\vartheta^+ \right). \end{aligned}$$

However, $\mathbf{M} (\vartheta_0) \geq \mathbf{M} (\vartheta^+)$ implying $\mathbf{M} (\vartheta_0) = \mathbf{M} (\vartheta^+)$. By the uniqueness of ϑ_0, we can deduce $\gamma_0 = \gamma^+$, $\beta_0 = \beta^+$ and $\Lambda_0 = \Lambda^+$ a.e. on ν_2. Since this is true for any convergent subsequence of $\hat{\vartheta}$, we conclude that all the limits of subsequence of $\hat{\vartheta}_{n_l}$ are equal to ϑ_0. This also implies the pointwise convergence of $\hat{\Lambda} (t)$ to $\Lambda_0 (t)$ for $t \in [0, \tau]$

a.s.. Due to the uniform boundedness of $\hat{\Lambda}$, the Dominated Convergence Theorem yields the strong consistency of $\hat{\vartheta}$ in the metric $d(\cdot,\cdot)$.

A.3.2 Proof of Convergence Rate

We mainly use Theorem 3.2.5 of van der Vaart and Wellner (1996) to derive the convergence rate of $\hat{\vartheta}$. We first verify $\mathbf{M}(\vartheta_0) - \mathbf{M}(\vartheta_0) \gtrsim d^2(\vartheta, \vartheta_0)$. Since $h(z) \geq (1/4)(z-1)^2$ for z in a small enough neighborhood of 1, for any ϑ in a sufficiently small neighborhood of ϑ_0,

$$
\begin{aligned}
\mathbf{M}(\vartheta_0) - \mathbf{M}(\vartheta) \geq & \int \int_{u_1}^{u_2} \exp\left(\gamma^\top w(t) + \beta^\top(t) x(t)\right) d\Lambda(t) \\
& \left\{ \frac{\int_{u_1}^{u_2} \exp\left(\gamma_0^\top w(t) + \beta_0^\top(t) x(t)\right) d\Lambda_0(t)}{\int_{u_1}^{u_2} \exp\left(\gamma^\top w(t) + \beta^\top(t) x(t)\right) d\Lambda(t)} - 1 \right\}^2 d\nu_1(u_1, u_2, v) \\
\gtrsim & \int \left\{ \int_{u_1}^{u_2} \exp\left(\gamma_0^\top w(t) + \beta_0^\top(t) x(t)\right) d\Lambda_0(t) \right. \\
& \left. - \int_{u_1}^{u_2} \exp\left(\gamma^\top w(t) + \beta^\top(t) x(t)\right) d\Lambda(t) \right\}^2 d\nu_1(u_1, u_2, v)
\end{aligned}
$$

by (C1) and (C2).

Let $g(\xi) = \int I(U_1 < t < U_2) \exp\left(\gamma_\xi^\top W(t) + \beta_\xi^\top X(t)\right) d\Lambda_\xi(t)$ where $\gamma_\xi = \xi\gamma + (1-\xi)\gamma_0$, $\beta_\xi = \xi\beta + (1-\xi)\beta_0$ and $\Lambda_\xi = \xi\Lambda + (1-\xi)\Lambda_0$. Then

$$
\begin{aligned}
& \int I(U_1 < t < U_2) \exp\left(\gamma_0^\top W(t) + \beta_0^\top(t) X(t)\right) d\Lambda_0(t) - \\
& \qquad \int I(U_1 < t < U_2) \exp\left(\gamma_0^\top W(t) + \beta_0^\top(t) X(t)\right) d\Lambda(t) = g(1) - g(0).
\end{aligned}
$$

By the mean value theorem, there exist $\xi^* \in [0,1]$ such that $g(1) - g(0) = g'(\xi^*)$.

Since

$$g'(\xi) = \int_{U_1}^{U_2} \Big(1 + \xi(\gamma - \gamma_0)^\top W(t) + \xi(\beta - \beta_0)^\top(t)\, X(t)\Big)$$
$$\exp\Big((\xi(\gamma - \gamma_0) + \gamma_0)^\top W(t) + (\xi(\beta - \beta_0) + \beta_0)^\top(t)\, X(t)\Big)\, d(\Lambda(t) - \Lambda_0(t))$$
$$+ \int_{U_1}^{U_2} \Big((\gamma - \gamma_0)^\top W(t) + (\beta - \beta_0)^\top(t)\, X(t)\Big)$$
$$\exp\Big((\xi(\gamma - \gamma_0) + \gamma_0)^\top W(t) + (\xi(\beta - \beta_0) + \beta_0)^\top(t)\, X(t)\Big)\, d\Lambda_0(t).$$

we have

$$\begin{aligned}
&\mathbf{M}(\vartheta_0) - \mathbf{M}(\vartheta) \\
\gtrsim&\nu_1 \Bigg(\Bigg\{\int_{U_1}^{U_2} \Big(1 + (\gamma - \gamma_0)^\top W(t) + \xi(\beta - \beta_0)^\top(t)\, X(t)\Big)\, d(\Lambda(t) - \Lambda_0(t)) \\
+& \int_{U_1}^{U_2} \Big((\gamma - \gamma_0)^\top W(t) + (\beta - \beta_0)^\top(t)\, X(t)\Big)\, d\Lambda_0(t)\Bigg\}^2\Bigg) \\
=&\nu_1 \{g_1 h_1 + g_2\}^2
\end{aligned}$$

where $g_1(U_1, U_2, V) = \int_{U_1}^{U_2} \Big((\gamma - \gamma_0)^\top W(t) + (\beta - \beta_0)^\top(t)\, X(t)\Big)\, d\Lambda_0(t)$,

$$g_2(U_1, U_2) = \Lambda(U_2) - \Lambda_0(U_2) - (\Lambda(U_1) - \Lambda_0(U_1))$$

and

$$h_1(U_1, U_2, V) = \xi \left\{ \frac{\int_{U_1}^{U_2} \Big((\gamma - \gamma_0)^\top W(t) + (\beta - \beta_0)^\top(t)\, X(t)\Big)\, d\Lambda(t)}{\int_{U_1}^{U_2} \Big((\gamma - \gamma_0)^\top W(t) + (\beta - \beta_0)^\top(t)\, X(t)\Big)\, d\Lambda_0(t)} \right\} + 1 - \xi,$$

in the notation of Lemma 8.8 in van der Vaart (2002). Our goal is to show

$$\nu_1 \left\{g_1 h_1 + g_2\right\}^2 \gtrsim \nu_1 \left(g_1^2\right) + \nu_1 \left(g_2^2\right)$$

by using Lemma 8.8 in van der Vaart (2002). For convenience, we write the expectation under ν_1 as E_1 Then, for $\eta \in (0, 1)$, we need to verify

$$\left\{E_1 \left[g_1 g_2\right]\right\}^2 \leq (1 - \eta)\, E_1 \left[g_1^2\right] E_1 \left[g_2^2\right]$$

to apply Lemma 8.8 in van der Vaart (2002). By the Cauchy–Schwarz inequality, we have

$$
\begin{aligned}
& \left\{E_1\left[g_1 g_2\right]\right\}^2 \\
= & \left\{E_1\left[g_2 E_1\left[g_1 | U_1, U_2\right]\right]\right\}^2 \\
\leq & E_1\left[g_2^2\right] E_1\left[\left\{E_1\left[g_1 | U_1, U_2\right]\right\}^2\right] \\
= & E_1\left[g_2^2\right] E_1\left[\left\{\int_{U_1}^{U_2} E_1\left[\left(\left(\gamma-\gamma_0\right)^{\top} W(t)+\left(\beta-\beta_0\right)^{\top}(t) X(t)\right) | U_1, U_2\right] d \Lambda_0(t)\right\}^2\right] \\
= & E_1\left[g_2^2\right] E_1\left[\int \int I\left(U_1<t<U_2\right) I\left(U_1<s<U_2\right)\right. \\
& \left\{E_1\left[\left(\beta-\beta_0\right)^{\top}(t) X(t)+\left(\gamma-\gamma_0\right)^{\top} W(t) | U_1, U_2\right]\right. \\
& \left.\times E_1\left[X^{\top}(s)\left(\beta-\beta_0\right)(s)+\left(\gamma-\gamma_0\right)^{\top} W(s) | U_1, U_2\right]\right\} d \Lambda_0(t) d \Lambda_0(s) \\
\leq & (1-\eta) E_1\left[g_2^2\right] E_1\left[\int \int I\left(U_1<t<U_2\right) I\left(U_1<s<U_2\right)\right. \\
& \times\left\{\left(\beta-\beta_0\right)^{\top}(t) E_1\left[X(t) X^{\top}(s) | U_1, U_2\right]\left(\beta-\beta_0\right)(s)\right. \\
& \left.+\left(\gamma-\gamma_0\right)^{\top} E_1\left[W(t) W^{\top}(s) | U_1, U_2\right]\left(\gamma-\gamma_0\right)\right\} d \Lambda_0(t) d \Lambda_0(s) \\
= & (1-\eta) E_1\left[g_2^2\right] E_1\left[g_1^2\right] .
\end{aligned}
$$

The last inequality is due to Condition (C7). Consequently, Lemma 8.8 in van der Vaart (2002) yields $\nu_1\left\{g_1 h_1+g_2\right\}^2 \gtrsim \nu_1\left(g_1^2\right)+\nu_1\left(g_2^2\right)$. For $\nu_1\left(g_1^2\right)$, by Jensen's inequality,

$$
\begin{aligned}
&\nu_1\left(\left\{\int_{U_1}^{U_2}\left(\beta-\beta_0\right)^{\top}(t)\, X(t)\, d\Lambda_0(t)\right\}^2\right)\\
=&\nu_1\left(\int_{U_1}^{U_2}\int_{U_1}^{U_2}\left\{\left(\beta-\beta_0\right)^{\top}(t)\, X(t)\, X(s)\left(\beta-\beta_0\right)(s)\right.\right.\\
&\left.\left.+\ \left(\gamma-\gamma_0\right)^{\top} W(t)\, W^{\top}(s)\left(\gamma-\gamma_0\right)\right\} d\Lambda_0(t)\, d\Lambda_0(s)\right)\\
&\gtrsim \left\|\beta-\beta_0\right\|_{L(\tilde{\mu}_0)}^2+\left\|\gamma-\gamma_0\right\|_2^2
\end{aligned}
$$

by $\nu_1\left(\int_{U_1}^{U_2}\int_{U_1}^{U_2} d\Lambda_0(t)\, d\Lambda_0(s)\right)>0$ because of the positive Λ_0' under (C1). Therefore, $\mathbf{M}(\vartheta_0)-\mathbf{M}(\vartheta)\gtrsim d(\vartheta,\vartheta_0)$.

Next, we need to derive $\phi_n(\eta)$ such that

$$
E\sup_{d(\vartheta,\vartheta_0)<\eta}\left|\mathbb{G}_n\left(m_{\vartheta}(O)-m_{\vartheta_0}(O)\right)\right|\lesssim\phi_n(\eta)\,.
$$

Define classes $\mathcal{B}_\eta=\{m_\vartheta(O)-m_{\vartheta_0}(O): d(\vartheta,\vartheta_0)<\eta, \vartheta\in\mathcal{A}\times\mathcal{B}_n\times\mathcal{F}\}$. By the similar argument in (Lu et al., 2009), this shows that ϵ-bracketing number for $\mathcal{B}_\eta$ under the Bernstein norm $\|f\|_{P,B}=\left\{2\mathbf{P}\left(e^{|f|}-1-|f|\right)\right\}^{1/2}$ (Wellner and Zhang, 2007) will be of the order

$$
\exp\left\{c\left(\frac{1}{\epsilon}+(p_1+p_2q_n)\log\left(\frac{\delta}{\epsilon}\right)\right)\right\}
$$

for some constant $\eta>\epsilon$ and c. Hence

$$
\log N_{[]}\left(\epsilon,\mathcal{B}_\eta,\|\cdot\|_{P,B}\right)\lesssim c\left(\frac{1}{\epsilon}+(p_1+p_2q_n)\log\left(\frac{\delta}{\epsilon}\right)\right)
$$

and then by Lemma 3.4.3 of van der Vaart and Wellner (1996),

$$E\left\|\sqrt{n}\left(\mathbb{P}_n-\mathbf{P}\right)\right\|_{\mathcal{F}_\delta}\leq cJ_{[]}\left(\delta,\mathcal{F}_\delta,\|\cdot\|_{\mathbf{P},B}\right)\left\{1+\frac{J_{[]}\left(\delta,\mathcal{F}_\delta,\|\cdot\|_{\mathbf{P},B}\right)}{\delta^2 n^{1/2}}\right\}$$

where

$$\begin{aligned}\tilde{J}_{[]}\left(\eta,\mathcal{B}_\eta,\|\cdot\|_{P,B}\right)&=\int_0^\eta\sqrt{1+\log N_{[]}\left(\epsilon,\mathcal{B}_\eta,\|\cdot\|_{P,B}\right)}d\epsilon\\&\lesssim q_n^{1/2}\delta^{1/2}.\end{aligned}$$

Thus,

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

Obviously, $\varphi_n(\delta)/\delta$ is decreasing in δ as the leading term is $\delta^{-1/2}$. Therefore

$$\rho_n^2\varphi_n\left(\frac{1}{\rho_n}\right)=\rho_n^{3/2}q_n^{1/2}+\rho_n^3 q_n n^{-1/2}\lesssim n^{1/2}$$

if $\rho_n=\min\left\{n^{\frac{1-\nu}{3}},n^{r\nu}\right\}$ and $0<\nu<1/2$.

Moreover, using similar argument in Lu et al. (2009), we can show $\mathbb{M}_n\left(\hat{\vartheta}_n\right)-\mathbb{M}_n\left(\vartheta_0\right)>-O_p\left(n^{-2r\nu}\right)\geq O_p\left(r_n^2\right)$. Then, by Theorem 3.2.5 of Wellner and Zhang (2007), we have $\rho_n d\left(\hat{\vartheta}_n,\vartheta_0\right)=O_p(1)$. If ν is chosen as $1/(3r+1)$, we obtain the optimal rate $n^{r/(3r+1)}$ because $(1-\nu)/3=r\nu$.

A.3.3 Asymptotic Normality

We mainly use the method in He et al. (2017). We define a sequence of maps S_n mapping a neighborhood of ϑ_0, denoted by $\mathcal{U}$, in the parameter space for ϑ into

$l^\infty \left(\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3\right)$ where $\mathcal{H}_1 = \{h_1 : h_1 \in \mathcal{A}\}$ and

$$\mathcal{H}_2 = \{h_2 (t) = (h_{21} (t), \ldots, h_{2p_2} (t)) : \\ h_{2k} \text{ is a function with bounded total variation in } [0, \tau], k = 1, \ldots p_2\}$$

as

$$\begin{aligned}
S_n (\vartheta) [h_1, h_2, h_3] =& \frac{d}{d\epsilon} \, l_n (\gamma + \epsilon h_1, \beta + \epsilon h_2, \Lambda + \epsilon h_3)|_{\epsilon=0} \\
=& \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{J_i} \Delta N_{ij} \, \frac{\int_{T_{i(j-1)}}^{T_{ij}} \exp\left(\gamma^\top W_i (t) + \beta^\top (t) Z_i (t)\right)}{\int_{T_{i(j-1)}}^{T_{ij}} \exp\left(\gamma^\top W_i (t) + \beta^\top (t) Z_i (t)\right) d\Lambda (t)} \\
& \left\{\left(h_1^\top W_i (t) + h_2^\top (t) Z_i (t)\right) d\Lambda (t) + dh_3 (t)\right\} \\
& - \left\{ \int_{T_{i(j-1)}}^{T_{ij}} \exp\left(\gamma^\top W_i (t) + \beta^\top (t) Z_i (t)\right) \right. \\
& \left. \left\{\left(h_1^\top W_i (t) + h_2^\top (t) Z_i (t)\right) d\Lambda (t) + dh_3 (t)\right\} \right\} \\
=& A_{n1} (\vartheta) [h_1] + A_{n2} (\vartheta) [h_2] + A_{n3} (\vartheta) [h_3] \\
=& \mathbb{P}_n \psi (\vartheta) [h_1, h_2, h_3],
\end{aligned}$$

where $A_{nk} (\vartheta) [h_k] = \mathbb{P}_n a_k (\vartheta) [h_k]$ for $k = 1, \ldots, 3$, and

$$\psi (\vartheta) [h_1, h_2, h_3] = \sum_{k=1}^{3} a_k (\vartheta) [h_k]$$

. Here

$$a_1(\vartheta)[h_1] = \sum_{j=1}^{J} \Delta N_{\cdot j} \frac{\int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\gamma^\top W(t) + \beta^\top(t) Z(t)\right) h_1^\top W(t)\, d\Lambda(t)}{\int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\gamma^\top W(t) + \beta^\top(t) Z(t)\right) d\Lambda(t)} - \left\{\int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\gamma^\top W(t) + \beta^\top(t) Z(t)\right) h_1^\top W(t)\, d\Lambda(t)\right\}$$

$$a_2(\vartheta)[h_2] = \sum_{j=1}^{J} \Delta N_{\cdot j} \frac{\int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\gamma^\top W(t) + \beta^\top(t) Z(t)\right) h_2^\top(t)(t) Z(t)\, d\Lambda(t)}{\int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\gamma^\top W(t) + \beta^\top(t) Z(t)\right) d\Lambda(t)} - \left\{\int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\gamma^\top W(t) + \beta^\top(t) Z(t)\right) h_2^\top(t) Z(t)\, d\Lambda(t)\right\}$$

and

$$a_3(\vartheta)[h_3] = \sum_{j=1}^{J} \Delta N_{\cdot j} \frac{\int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\gamma^\top W_i(t) + \beta^\top(t) Z_i(t)\right) dh_3(t)}{\int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\gamma^\top W_i(t) + \beta^\top(t) Z_i(t)\right) d\Lambda(t)} - \left\{\int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\gamma^\top W_i(t) + \beta^\top(t) Z_i(t)\right) dh_3(t)\right\}$$

Correspondingly, we define the limit map $S : \mathcal{U} \to l^\infty(\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3)$ as

$$S(\vartheta)[h_1, h_2, h_3] = A_1(\vartheta)[h_1] + A_2(\vartheta)[h_2] + A_3(\vartheta)[h_3]$$

where $A_k(\vartheta)[h_k] = \mathbf{P}a_k(\vartheta)[h_k]$.

To derive the asymptotic normality of $\hat{\vartheta}_n$, we need to verify the following five conditions in He et al. (2017).

(a1) $\sqrt{n}(S_n - S)\left(\hat{\vartheta}_n\right) - \sqrt{n}(S_n - S)(\vartheta_0) = o_p(1)$.

(a2) $S(\vartheta_0) = 0$ and $S_n\left(\hat{\vartheta}_n\right) = o_p\left(n^{-1/2}\right)$.

(a3) $\sqrt{n}\,(S_n - S)(\theta_0)$ converges in distribution to a tight Gaussian process on $l^\infty(\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3)$.

(a4) $S(\vartheta)$ is Frï¿œchet-differentiable at ϑ_0 denoted by $\dot{S}(\vartheta_0)$.

(a5) $S\left(\hat{\vartheta}_n\right) - S(\vartheta_0) - \dot{S}(\vartheta_0)\left(\hat{\vartheta}_n - \vartheta_0\right) = o_p\left(n^{-1/2}\right)$.

Using similar argument in Lu et al. (2009), it is not hard to show

$$\{\psi(\vartheta)[h_1, h_2, h_3] - \psi(\vartheta_0)[h_1, h_2, h_3] : d(\vartheta, \vartheta_0) < \delta, (h_1, h_2, h_3) \in \mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3\}$$

is a Donkser class for some δ. Therefore,

$$\sup_{(h_1,h_2,h_3)\in\mathcal{A}\times\mathcal{B}\times\mathcal{F}} \mathbf{P}\left\{\psi\left(\hat{\vartheta}_n\right)[h_1, h_2, h_3] - \psi(\vartheta_0)[h_1, h_2, h_3]\right\}^2 \to 0$$

as $d\left(\hat{\vartheta}_n, \vartheta_0\right) \to 0$ in probability and thus (a1) holds.

For (a2), clearly, $S(\vartheta_0) = 0$. For $h_2 \in \mathcal{H}_2$, let h_{2n} be the B-spline function approximation of h_2 with $\max_{j=1,\ldots,p_2} \|h_{2j} - h_{2nj}\|_\infty = O\left(n^{-\nu r}\right)$ (Schumaker, 2007). Then we have $S_n\left(\hat{\vartheta}_n\right)[h_1, \boldsymbol{h}_{2n}, h_3] = 0$. Thus,

$$\begin{aligned} S_n\left(\hat{\vartheta}_n\right)[h_1, h_2, h_3] &= \sqrt{n}\mathbb{P}_n\psi\left(\hat{\vartheta}_n\right)[h_1, h_2, h_3] - \sqrt{n}\mathbf{P}\psi\left(\hat{\vartheta}_n\right)[h_1, \boldsymbol{h}_{2n}, h_3] \\ &= I_{n1} - I_{n2} + I_{n3} + I_{n4} \end{aligned}$$

where

$$I_{n1} = \sqrt{n}\,(\mathbb{P}_n - \mathbf{P})\left\{\psi\left(\hat{\vartheta}_n\right)[h_1, h_2, h_3] - \psi(\vartheta_0)[h_1, h_2, h_3]\right\}$$

$$I_{n2} = \sqrt{n}\left(\mathbb{P}_n - \mathbf{P}\right)\left\{\psi\left(\hat{\vartheta}_n\right)[h_1, \boldsymbol{h}_{2n}, h_3] - \psi\left(\vartheta_0\right)[h_1, \boldsymbol{h}_{2n}, h_3]\right\}$$

$$I_{n3} = \sqrt{n}\mathbb{P}_n\left\{\psi\left(\vartheta_0\right)[h_1, h_2, h_3] - \psi\left(\vartheta_0\right)[h_1, \boldsymbol{h}_{2n}, h_3]\right\}$$

and

$$I_{n4} = \sqrt{n}\mathbf{P}\left\{\psi\left(\hat{\vartheta}_n\right)[h_1, h_2, h_3] - \psi\left(\hat{\vartheta}_n\right)[h_1, \boldsymbol{h}_{2n}, h_3]\right\}.$$

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

$$\begin{aligned} I_{n3} =& \sqrt{n}\left(\mathbb{P}_n - \mathbf{P}\right)\left\{\psi\left(\vartheta_0\right)[h_1, h_2, h_3] - \psi\left(\vartheta_0\right)[h_1, \boldsymbol{h}_{2n}, h_3]\right\} \\ & + \sqrt{n}\mathbf{P}\left\{\psi\left(\vartheta_0\right)[h_1, h_2, h_3] - \psi\left(\vartheta_0\right)[h_1, \boldsymbol{h}_{2n}, h_3]\right\} \\ =& I_{n31} + I_{n32}. \end{aligned}$$

Similarly to proving (a1), $I_{n31} = o_p(1)$ and $I_{n32} = 0$ since $S(\vartheta_0) = 0$ for any $h_{2,} h_{2n} \in \mathcal{H}_2$. For I_{n4},

$$\begin{aligned} |I_{n4}| &\le \sqrt{n} d\left(\hat{\vartheta}_n, \theta_0\right)\left(\max_{j=1,\ldots,p_2}\left\|h_{2j} - h_{2nj}\right\|_\infty\right) \\ &= O_p\left(\max\left\{n^{-(1-\nu)/3}, n^{-r\nu}\right\} n^{-rv+1/2}\right) \\ &= o_p(1) \end{aligned}$$

if $1/(4r) < \nu < 1/2$. Thus (a2) holds.

Condition (a3) holds because $\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$ is a Donsker class and the functionals $A_1(\vartheta)[h_1]$, $A_2(\vartheta)[h_2]$ and $A_3(\vartheta)[h_3]$ are bounded Lipschitz functions with respect to $\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$ due the compactness of $\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$.

For (a4), by the smoothness of $S(\vartheta)$ the Frï¿œchet differentiability holds and the derivative of $S(\vartheta)$ at ϑ_0, denoted by $\dot{S}(\vartheta_0)$ is a map from the space $\{\vartheta - \vartheta_0 : \vartheta \in \mathcal{U}\}$ to $l^\infty(\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3)$. Now we calculate $\dot{S}(\vartheta_0)$ as

$$
\begin{aligned}
&\dot{S}(\vartheta_0)(\vartheta - \vartheta_0)[h_1, h_2, h_3] \\
=&\frac{d}{d\epsilon}\left\{A_1(\vartheta_0 + \epsilon(\vartheta - \vartheta_0))[h_1]\right\}\big|_{\epsilon=0} \\
&+ \frac{d}{d\epsilon}\left\{A_2(\vartheta_0 + \epsilon(\vartheta - \vartheta_0))[h_2]\right\}\big|_{\epsilon=0} \\
&+ \frac{d}{d\epsilon}\left\{A_3(\vartheta_0 + \epsilon(\vartheta - \vartheta_0))[h_3]\right\}\big|_{\epsilon=0} \\
=& - \mathbf{P}\sum_{j=1}^{J}\int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\gamma_0^\top W(t) + \beta_0^\top(t) Z(t)\right) \\
&\times \left(\left(h_1^\top W(t) + h_2^\top(t) Z(t)\right) d\Lambda_0 + dh_3(t)\right) \\
&/\left\{\int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\gamma_0^\top W(t) + \beta_0^\top(t) Z(t)\right) d\Lambda_0\right\} \\
&\left\{\int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\gamma_0^\top W(t) + \beta_0^\top(t) Z(t)\right)\right. \\
&\left.\times\left[\left((\gamma - \gamma_0)^\top(t) W(t) + (\beta - \beta_0)^\top(t) Z(t)\right) d\Lambda_0 + d(\Lambda - \Lambda_0)\right]\right\} \\
=&(\gamma - \gamma_0)^\top Q_1(h_1, h_2, h_3) + \int_0^\tau (\beta - \beta_0)^\top(u)\, dQ_2(h_1, h_2, h_3, u) \\
&+ \int_0^\tau Q_3(h_1, h_2, h_3, u)\, d(\Lambda - \Lambda_0)
\end{aligned}
$$

where

$$
\begin{aligned}
Q_1(h_1, h_2, h_3) = -\mathbf{P}\sum_{j=1}^{J} &\frac{\int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\gamma_0^\top W(t) + \beta_0^\top(t) Z(t)\right)}{\int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\gamma_0^\top W(t) + \beta_0^\top(t) Z(t)\right) d\Lambda_0} \\
&\left(\left(h_1^\top W(t) + h_2^\top(t) Z(t)\right) d\Lambda_0 + dh_3(t)\right) \\
&\qquad\left\{\int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\gamma_0^\top W(t) + \beta_0^\top(t) Z(t)\right) W(t)\, d\Lambda_0\right\},
\end{aligned}
$$

$$dQ_2(h_1,h_2,h_3,t) = -\mathbf{P}\sum_{j=1}^{J} \frac{\int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\gamma_0^\top W(t) + \beta_0^\top(t) Z(t)\right)}{\int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\gamma_0^\top W(t) + \beta_0^\top(t) Z(t)\right) d\Lambda_0}$$

$$\left(\left(h_1^\top W(t) + h_2^\top(t) Z(t)\right) d\Lambda_0 + dh_3(t)\right)$$

$$I\left(T_{\cdot(j-1)} < u < T_{\cdot j}\right) \exp\left(\gamma_0^\top W(u) + \beta_0^\top(u) Z(u)\right) Z(u)\, d\Lambda_0(u),$$

and

$$Q_3(h_1,h_2,h_3,t) = -\mathbf{P}\sum_{j=1}^{J} \frac{\int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\gamma_0^\top W(t) + \beta_0^\top(t) Z(t)\right)}{\int_{T_{\cdot(j-1)}}^{T_{\cdot j}} \exp\left(\gamma_0^\top W(t) + \beta_0^\top(t) Z(t)\right) d\Lambda_0}$$

$$\left(\left(h_1^\top W(t) + h_2^\top(t) Z(t)\right) d\Lambda_0 + dh_3(t)\right)$$

$$I\left(T_{\cdot(j-1)} < u < T_{\cdot j}\right) \exp\left(\gamma_0^\top W(u) + \beta_0^\top(u) Z(u)\right)$$

We can also show $Q = (Q_1, Q_2, Q_3)$ is one-to-one by the similar method in He et al. (2017).

Last for (a5), by using the mean value theorem twice, we can show under conditions (C1), (C2), (C4) and (C8),

$$\left|S\left(\hat{\theta}\right) - S(\theta_0) - \dot{S}(\theta_0)\left(\hat{\theta} - \theta_0\right)\right| \lesssim O\left(d^2\left(\hat{\theta}, \theta\right)\right).$$

Since $n^{1/2} d^2\left(\hat{\vartheta}_n, \vartheta\right) = O_p\left(n^{1/2}\max\left\{n^{-2(1-\nu)/3}, n^{-2r\nu}\right\}\right) = o_p(1)$ if $1/(4r) < \nu < 1/4$, we can conclude that $S\left(\hat{\vartheta}_n\right) - S(\vartheta_0) - \dot{S}(\vartheta_0)\left(\hat{\vartheta}_n - \vartheta_0\right) = o_p\left(n^{-1/2}\right)$ and (a5) holds.

If (a1)-(a5) hold, according to He et al. (2017), we have

$$-\sqrt{n}\dot{S}(\vartheta_0)\left(\hat{\vartheta}_n - \vartheta_0\right)[h_1,h_2,h_3] = \sqrt{n}(S_n - S)(\vartheta_0)[h_1,h_2,h_3] + o_p(1),$$

uniformly in h_1, h_2, h_3. For each $(h_1, h_2, h_3) \in \mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$, Q is invertible by the similar arguement in He et al. (2017). Then there exists $(h_1, h_2, h_3) \in \mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$ such that

$$Q_1 (h_1, h_2, h_3) = h_1, Q_2 (h_1, h_2, h_3) = h_2, Q_{3k} (h_1, h_2, h_3) = h_{3k}.$$

Therefore, we have

$$\begin{aligned}
&\sqrt{n} (\hat{\gamma} - \gamma_0)^T h_1 + \sqrt{n} \int_0^{\tau} \left(\hat{\beta}_n (t) - \beta_0 (t)\right)^T dh_2 (t) \\
&+ \sqrt{n} \int_0^{\tau} \left(\hat{\Lambda} (t) - \Lambda_0 (t)\right) dh_3 (t) \\
=&\sqrt{n} (S_n - S) (\vartheta_0) [h_1, h_2, h_3] + o_p (1) \to_d N \left(0, \sigma^2\right)
\end{aligned}$$

where $\sigma^2 = E [\psi^2 (\vartheta_0) [h_1, h_2, h_3]]$ because of (a3). To find the asymptotic distribution of γ only, we can find h_1, h_2 and h_3 as a solution of $Q_2 = 0$ and $Q_3 = 0$. Unfortunately, we cannot find the explicit forms of h_1, h_2 and h_3 as He et al. (2017). Hence, we adopt the variance estimation method in the main body.

Bibliography

Amorim, L. D., Cai, J., Zeng, D., & Barreto, M. L. (2008). Regression splines in the time-dependent coefficient rates model for recurrent event data. *Statistics in Medicine*, *27*(28), 5890–5906.

Cai, J., Fan, J., Zhou, H., & Zhou, Y. (2007). Hazard models with varying coefficients for multivariate failure time data. *The Annals of Statistics*, *35*(1), 324–354.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. In *Handbook of Econometrics* (pp. 5549–5632).

Cook, R., & Lawless, J. (2007). *The Statistical Analysis of Recurrent Events*.

Dempster, A. P., Laird, N. M., & work(s): D. B. R. R. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*(1), 1–38.

He, X., Tong, X., Sun, J., & Cook, R. J. (2007). Regression analysis of multivariate panel count data. *Biostatistics*, *9*(2), 234–248.

He, X., Tong, X., & Sun, J. (2008). Semiparametric analysis of panel count data with correlated observation and follow-up times. *Lifetime Data Analysis*, *15*(2), 177–196.

He, X., Feng, X., Tong, X., & Zhao, X. (2017). Semiparametric partially linear varying coefficient models with panel count data. *Lifetime Data Analysis*, *23*(3), 439–466.

Hu, X. J., Sun, J., & Wei, L.-J. (2003). Regression parameter estimation from panel counts. *Scandinavian Journal of Statistics*, *30*(1), 25–43.

Hua, L., Zhang, Y., & Tu, W. (2014). A spline-based semiparametric sieve likelihood method for over-dispersed panel count data. *Canadian Journal of Statistics*, *42*(2), 217–245.

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

Li, N., Park, D.-H., Sun, J., & Kim, K. (2011). Semiparametric transformation models for multivariate panel count data with dependent observation process. *Canadian Journal of Statistics*, *39(3)*, 458–474.

Li, N., Zhao, H., & Sun, J. (2013). Semiparametric transformation models for panel count data with correlated observation and follow-up times. *Statistics in Medicine*, *32*(17), 3039–3054.

Lin, D. Y., Wei, L. J., & Ying, Z. (2001). Semiparametric transformation models for point processes. *Journal of the American Statistical Association*, *96*(454), 620–628.

Lin, H., Fei, Z., & Li, Y. (2015). A semiparametrically efficient estimator of the time-varying effects for survival data with time-dependent treatment. *Scandinavian Journal of Statistics*, *43*(3), 649–663.

Lu, M., Zhang, Y., & Huang, J. (2007). Estimation of the mean function with panel count data using monotone polynomial splines. *Biometrika*, *94*(3), 705–718.

Lu, M., Zhang, Y., & Huang, J. (2009). Semiparametric estimation methods for panel count data using monotone B-splines. *Journal of the American Statistical Association*, *104*(487), 1060–1070.

Murphy, S. A., & Vaart, A. W. V. D. (2000). On profile likelihood. *Journal of the American Statistical Association*, *95*(450), 449–465.

Oliveira, J. (2016). The value of children: inter-generational support, fertility, and human capital. *Journal of Development Economics*, *120*, 1–16.

Perperoglou, A. (2012). Reduced rank hazard regression with fixed and time-varying effects of the covariates. *Biometrical Journal*, *55*(1), 38–51.

Perperoglou, A. (2013). Cox models with dynamic ridge penalties on time-varying effects of the covariates. *Statistics in Medicine*, *33*(1), 170–180.

Schumaker, L. L. (2007). *Spline Functions: Basic Theory.* New York: Cambridge University Press.

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

Sun, J., & Zhao, X. (2013). *Statistical Analysis of Panel Count Data.* Springer Science + Business Inc.

Sun, L., Guo, S., & Chen, M. (2009). Marginal regression model with time-varying coefficients for panel data. *Communications in Statistics - Theory and Methods*, *38*(8), 1241–1261.

Tian, L. (2018). Less is more: the effects of children on parents' income, assets and intergenerational transfers. *Innovation in Aging*, *2*, 468–468.

Tian, L., Zucker, D., & Wei, L. J. (2005). On the Cox model with time-varying regression coefficients. *Journal of the American Statistical Association*, *100*(469), 172–183.

van der Vaart, A. W., & Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer Series in Statistics. New York: Springer.

van der Vaart, W. A. (2002). Semiparametric statistics. In *Lectures on probability theory andstatistics. école d'etè de probabilitès de saint-flour xxix-1999. lecture notes in math* (Vol. 1781, pp. 331–457). Springer, New York.

Wang, Y., & Yu, Z. (2021a). A kernel regression model for panel count data with nonparametric covariate functions. *Biometrics*, Accepted.

Wang, Y., & Yu, Z. (2021b). A kernel regression model for panel count data with time-varying coefficients. *Statistica Sinica*, Accepted. arXiv: http://arxiv.org/abs/1903.10233v1 [math.ST]

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

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

Xie, Y., He, Z., Tu, W., & Yu, Z. (2019). Variable selection for joint models with time-varying coefficients. *Statistical Methods in Medical Research*, *29*(1), 309–322.

Yan, J., & Huang, J. (2012). Model selection for cox models with time-varying coefficients. *Biometrics*, *68*(2), 419–428.

Yu, Z., & Lin, X. (2010). Semiparametric regression with time-dependent coefficients for failure time data analysis. *Statistica Sinica*, *20*(2), 853.

Zeng, D., Gao, F., & Lin, D. Y. (2017). Maximum likelihood estimation for semiparametric regression models with multivariate interval-censored data. *Biometrika*, *104*(3), 505–525.

Zhang, H., Sun, J., & Wang, D. (2013). Variable selection and estimation for multivariate panel count data via the seamless-L_0 penalty. *Canadian Journal of Statistics*, *41*(2), 368–385.

Zhang, Y. (2002). A semiparametric pseudolikelihood estimation method for panel count data. *Biometrika*, *89*(1), 39–48.

Zhang, Y., & Jamshidian, M. (2004). On algorithms for the nonparametric maximum likelihood estimator of the failure function with censored data. *Journal of Computational and Graphical Statistics*, *13*(1), 123–140.

Zhao, H., Tu, W., & Yu, Z. (2018). A nonparametric time-varying coefficient model for panel count data. *Journal of Nonparametric Statistics*, *30*(3), 640–661.

Zhao, H., Li, Y., & Sun, J. (2012). Analyzing panel count data with a dependent observation process and a terminal event. *Canadian Journal of Statistics*, *41*(1), 174–191.

Zhao, H., Li, Y., & Sun, J. (2013). Semiparametric analysis of multivariate panel count data with dependent observation processes and a terminal event. *Journal of Nonparametric Statistics*, *25*(2), 379–394.

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

Zhu, L., Zhang, Y., Li, Y., Sun, J., & Robison, L. L. (2017). A semiparametric likelihood-based method for regression analysis of mixed panel-count data. *Biometrics*, *74*(2), 488–497.

VITA

Yuanyuan Guo received her Bachelor of Science in 2016 from The Central South University. Then she joined the Ph.D. program in the Department of Statistics at the University of Missouri-Columbia in August 2016. She will start her postdoctoral training at Duke University in July 2021.