A PREDICTIVE TIME-TO-EVENT MODELING APPROACH WITH LONGITUDINAL MEASUREMENTS AND MISSING DATA

A Dissertation
Submitted to
the Temple University Graduate Board

---

in Partial Fulfillment
of the Requirements for the Degree of
DOCTOR OF PHILOSOPHY

---

by
Lili Zhu
December, 2019

ABSTRACT

A PREDICTIVE TIME-TO-EVENT MODELING APPROACH WITH LONGITUDINAL MEASUREMENTS AND MISSING DATA

Lili Zhu
DOCTOR OF PHILOSOPHY

Temple University, December, 2019

Dr. Cheng Yong Tang, Chair

An important practical problem in survival analysis is predicting the time to a future event such as the death or failure of a subject. It is of great importance for medical decision making to investigate how the predictor variables including repeated measurements of the same subjects are affecting future time-to-event. Such a prediction problem is particularly more challenging due to the fact that the future values of predictor variables are unknown, and they may vary dynamically over time.

In this thesis, we consider a predictive approach based on modeling the forward intensity function. To handle the practical difficulty due to missing data in longitudinal measurements, and to accommodate observations at irregularly spaced time points, we propose a smoothed composite likelihood approach for estimations. The forward intensity function approach intrinsically incorporates the future dynamics in the predictor variables that affect the stochastic occurrence of the future event. Thus the proposed framework is advantageous and parsimonious from requiring no separated modeling step for the stochastic mechanism of the predictor variables. Our theoretical analysis establishes the validity of the forward intensity modeling approach and the smoothed composite likelihood method. To model the parameters as continuous functions of time, we introduce the penalized B-spline into the proposed approach.

Extensive simulations and real-data analyses demonstrate the promising performance of the proposed predictive approach.

ACKNOWLEDGEMENTS

I want to express my sincere gratitude to my PhD advisor Dr. Cheng Yong Tang for giving me an opportunity to do research and providing invaluable support and guidance. Doing my PhD research under Dr. Tang's guidance has been the most rewarding experience, as it really shaped my capability of logical thinking, perseverance, attention to details and problems solving, which not only I benefit for my research but also for my life.

I would like to thank my committee members Dr. William W.S. Wei and Dr. Xu Han from Temple University, and Dr. Yong Chen from University of Pennsylvania for their insightful questions, comments, and suggestions.

I would also like to thank my many colleagues at Bristol-Myers Squibb for their support and encouragement during my PhD studies. Special thanks to Dr. Kay Tatsuoka who supported me to apply graduate school. I am very grateful to Dr. Kay Tatsuoka, Dr. Li-An Xu, and Dr. Jingli Song who provide me invaluable guidance and opportunities for growth and career development, and meanwhile encourage and support me to have work, life and school balance. I am also very appreciative of my colleague Dr. Ming Zhou who gave me many excellent suggestions over years on my research. I greatly value the comradeship with Dr. Tina Young as we both pursued PhD education while working and her emotional support and encouragement through the challenging time. I would like to extend my sincere thanks to many colleagues and friends whose names are not mentioned here for being so supportive and keen to know how I was proceeding over the past 5 years. Last but not the least I want to thank Bristol-Myers Squibb for providing me the financial support for my PhD research.

DEDICATION

I dedicate my dissertation work to my family. This dissertation would not have been possible without the love, caring, sacrifices and support of my families.

A deep gratitude to my parents who taught me perseverance and audacity. In these five years, Sophie and Isabelle arrived and joined our growing family. By staying with us and taking care of the whole family for one and a half years after Sophie and Isabelle were born, my loving mother and father made it possible for me to enjoy being a mother of two newborn children, and in the meantime to be able to work full time and do research part-time.

I also dedicate this dissertation to my beloved mother in-law and father in-law who always emphasize the importance of education in life and support me to pursue my PhD education. Before my PhD screening examination, my mother in-law and father in-law took care of Sophie in their city while Sophie was one year old such that I could better focus on my examination preparation.

Last but most importantly, I dedicate my dissertation to the love of my life Amir Zabet. I am so blessed of having Amir's love, sacrifices and emotional support. I don't recall how many nights in the past five years that I left Sophie and Isabelle to Amir alone while I went to school for classes and by the time I was home, both children were asleep after he fed them, entertained them, and put them into bed (this was after his full day of work); how many weekends I asked Amir to plan play dates for the children so I could focus on my research; and how many times when stress got my nerves, and Amir told me it would be ok. Without Amir, I wouldn't be the person I am today.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Survival analysis predictively concerning the future time-to-event is of importance in practice. Longitudinal measurements are commonly available in those problems. For example, clinical studies assessing the time-to-event outcomes (such as progression free survival or overall survival) as the primary objectives often monitor patients' key longitudinal data over time. These longitudinal measurements may include, for example, the virologic status such as the CD4 count in human immune virus (HIV) patients or the prostate specific antigen (PSA) in prostate cancer patients. Physicians need prediction tools to assess the future time-to-event outcomes based on these longitudinal measures up to the current time, ultimately making individualized medical practice possible and improving clinical outcomes.

A class of the modeling device to characterize the association between the longitudinal measurements and the survival process is the joint modeling approach. It is generally comprised of two linked components through shared random effects, for example, a linear mixed effects model for the longitudinal process, and a Cox (Cox, 1972) model for the survival process. We refer to Tsiatis and Davidian (2004) and Yu et al. (2004) for a comprehensive overview on this methodology. Recently, there are growing interests in utilizing the joint modeling method as a prediction tool; see, for example, Taylor et al. (2005), Yu et al. (2008), Taylor et al. (2013), Proust-Lima and Taylor (2009), Rizopoulos

(2011), and Rizopoulos (2016).

Upon building the joint model, a two-step procedure is needed for predicting the future time-to-events. That is, the future longitudinal measures need to be predicted first, before quantifying the survival probabilities and other relevant quantitative measures. Clearly, the accuracy of the predictions using the joint modeling approach depends on the correctness of the model specification for both the survival process and longitudinal measurements. The misspecification of the dynamics of the longitudinal measurements could lead to biased projection of the future longitudinal measurements and subsequently impacting the predicted survival probabilities. Therefore, one focus of the joint modeling research is to develop more adaptive and flexible models, either to characterize the hazard functions (Tseng et al., 2005, 2015) when the proportionality assumption fails, or to model longitudinal process when the linear mixed model was inadequate to represent the trajectories (Ding and Wang, 2008; Yu et al., 2004; Taylor et al., 2005; Yu et al., 2008; Garre et al., 2008).

Another potential practical limitation of the joint modeling approach as a prognostic tool is its high computation complexity; see, among others, Pauler and Finkelstein (2002). For example, the expectation-maximization (EM) algorithm proposed by Wulfsohn and Tsiatis (1997) to estimate the parameters in the joint modeling requires multidimensional numerical integrations that can be computationally prohibitive for a model with many parameters. The computation burden increases exponentially with the dimensionality of the integration especially given that the shared random effects are individual-specific and increase with the number of individuals. One area of joint modeling research has been to find numerical integration techniques that ease the computation burden, such as the Monte Carlo approach proposed by Henderson et al. (2000), Gauss-Hermite quadrature by Song et al. (2002), and fully exponential Laplace approximations by Rizopoulos et al. (2009). Faucett and Thomas. (1996) took a Bayesian approach using Markov Chain Monte Carlo (MCMC) technique of Gibbs sampling to estimate the unknown parameters of the joint posterior distribution in the model. Ibrahim et al. (2001) and

Wang and Taylor. (2001) extended the Bayesian joint modeling approach by proposing a semiparametric cure rate model for the event process and more flexible longitudinal model respectively.

In this thesis, we consider a new framework for the predictive time-to-event analysis. Our framework is built on a new foundation that is different from the existing joint modeling approach for solving the predictive survival analysis problems. In particular, our framework is constructed based on the forward intensity function defined as the conditional future event intensity function given the information up to the current time t. The forward intensity function approach has been considered in corporate default predictions in finance; see Chen (2007) and Duan et al. (2012). The key advantage of the forward intensity function approach is that it intrinsically incorporates the future dynamics in the predictor variables that affect the stochastic occurrence of the future event. By its definition, the forward intensity function is predictive so that our framework is a one-step synthetic modeling approach that requires no separated layers of models for the future longitudinal measurements and event occurrence.

Another related class of approaches in survival analysis is the landmark analysis; see Houwelingen (2007), Parast et al. (2012), Parast et al. (2014), among others. The similarity is that the landmarking methods concern a set of survivors at a given time point, and is targeting at the future events after that time point. As noted in Houwelingen (2007), the landmark analysis typically works with the original time scale, and existing methods generally specify a "global" semiparametric model that covers the entire time-span of the study. For predictive analysis, working in the original time scale may encounter difficulty when concerns are beyond the time-span in which the data are collected. A major difference of our approach on this aspect is that the forward intensity model can be viewed as a bivariate function of t and u – a given time point and the unit of time after that. This can be viewed as a reset of the time origin at t as the starting point, and it is advantageous for building a predictive model. There is few study carrying the interpretation as a reset of the time origin, e.g., Zheng and Heagerty (2005); but the development therein is also a "global" semiparametric model. Our model development in this study is different; it can be viewed as "local" for each given pair of t and u, and then appropriate aggregations are developed for general multiple-period predictions.

In clinical settings, it is a common issue to encounter missing data and irregularly spaced longitudinal measurements. In existing studies, the last observation carried forward (LOCF) method is commonly applied, replacing the missing datum with the nearest observation. One potential issue with the LOCF is that it may introduce substantial bias in the model parameter estimation (Prentice, 1982). To properly handle the missing data, we develop a smoothed composite likelihood approach with some kernel function that appropriately adjusts the weight of contribution to the likelihood from multiple observations near the missing one. Our smoothed composite likelihood approach is related to the local likelihood method of Tibshirani and Hastie (1987) to capture nonlinear dependencies. The method has also been applied to both the generalized linear model and the Cox proportional hazard model; see, for example, Fan et al. (1997), Staniswalis (1989), Hunsberger (1994), Tibshirani and Hastie (1987), Cai and Sun (2003) and Tian et al. (2005). Recently, Cao et al. (2015) extended the work of Cai and Sun (2003) and Tian et al. (2005) and proposed smoothing approach for the Cox model. While Cao et al. (2015) used the smoothing method to estimate time-invariant parameters in the proportional hazard Cox model, our primary goal is to improve the prediction performance by the smoothed composite likelihood approach under the forward intensity model. Under proper choice of bandwidth, our proposed smoothed estimators are proven to be consistent and asymptotically normal even when the predictor variables have missing values or are not regularly observed.

The parameter estimators by either the LOCF or the smoothed composite likelihood approach are obtained at each prediction horizon time point (TP) separately. To model the parameter estimators as a continuous function of time, we extend the research and introduce B-splines in the forward intensity model. We refer to (Boor, 1978) for comprehensive overview of B-spline basics. The composite likelihood would include contributions from all prediction horizons examined and those time points with longitudinal measurements available. The composite likelihood approach using B-splines enables the estimation to borrow information across multiple prediction horizons without LOCF or kernel smoothing. To optimize the parameter estimation, we adopt the penalized B-spline method (Eilers and Marx, 1996) by introducing a penalty term of differences between coefficients of adjacent B-splines in the composite likelihood function. The penalized B-spline method was first developed by O'Sullivan (1988) which includes a penalty term on the squared second order derivative to restrict the flexibility of the fitted curve and prevent overfitting. The computation of derivatives in the penalty term can be quite tedious when the third or higher order of derivatives are used in the penalty term. Wand and Ormerod (2008) extended O'Sullivan's idea to higher orders of derivatives using a computer algebra system. Eilers and Marx (1996) proposed a simplified penalty term, also called P-splines, using differences of B-spline coefficients and showed strong connections for the second order differences to the spline penalty by O'Sullivan (1988). Eilers et al. (2015) gave a comprehensive review of the development of P-splines in the last two decades. In our study, the penalty coefficient is chosen by the n-fold cross-validation. The advantage of this approach compared to the estimation by each prediction horizon time point is demonstrated in simulations with less variability in the parameter estimation especially when the sample size is small. The real data analysis shows that the estimators by the composite likelihood approach using the penalized B-spline is a smooth function of time and could well capture the curvature of the estimators by the estimation at each prediction horizon time point.

Our investigation makes several contributions to the field. Foremost for the predictive survival analysis, we propose and analyze a modeling approach with a fresh concept of the forward intensity function. Our work is among the first investigating the forward intensity function broadly as a predictive modeling device for survival analysis using longitudinal measurements with missing data. We demonstrate promising performance with this new framework. Our new modeling framework is advantageous for solving the problems by its convenience in both the model building and the prediction computations. Second, we reveal new insights via our technical analysis. In existing study of Duan et al. (2012), the truth of the model parameter is assumed under the discrete data model in the technical development. Since the sampling frequency may vary in different scenarios, ambiguity exists in the definition of the model parameters at the population level. How to broadly characterize the statistical properties of the estimators remains less explored. Our analysis connects the important practical reality of discretized longitudinal measurements with the ideal setting of some continuous-time processes. We establish conditions under which the estimators of our approach are consistent with discretized data. Third, missing data in the longitudinal measurements are handled by the smoothed composite likelihood in our proposed framework, and its validity is established through the examination of the asymptotic properties. Such a contribution more broadly extends the scope of the forward intensity modeling approach in survival analysis.

The rest of this thesis is organized as follow. Section 2 provides literature review in related areas. We outline our methodological framework in Section 3 including the forward intensity function modeling, its estimation with discrete-time observations, the smoothed composite likelihood method to deal with missing data and observations at irregularly spaced time points, and the composite likelihood method using the penalized B-spline to model the parameters as continuous functions of time. Numerical examples with real data analyses and simulation studies are presented in Section 4. Theoretical investigations justifying the validity of the smoothed composite likelihood approach are given in Section 5, followed by some discussions in Section 6. The technical proofs and additional simulation studies are given in Supplementary Material of this thesis.

CHAPTER 2

LITERATURE REVIEW

This section reviews the survival analysis, competing risks, joint modeling, landmarking analyses, forward intensity model, missing longitudinal data, and B-spline.

2.1 Survival analysis

Survival analysis is unique in the sense that the underlying distribution is often not normal and data can be 'censored'. Two basic quantities to describe time-to-event data are survival function and hazard rate. The survival function is defined as probability of an individual surviving beyond time t.

$$S(t) = \mathbb{P}(T > t)$$

The hazard rate is the instantaneous rate of occurrence of an event.

$$\lambda(t) = lim_{\triangle t \to 0} \frac{\mathbb{P}(t \leq T < t + \triangle t | T \geq t)}{\triangle t}$$

There are two fundamental mathematical connections between the survival function and the hazard rate. First we have,

$$\lambda(t) = lim_{\triangle t \to 0} \frac{1}{\triangle t} \frac{S(t) - S(t + \triangle t)}{S(t)} = -\frac{S'(t)}{S(t)}$$

By integration we have,

$$S(t) = \exp\{-\int_0^t \lambda(s)ds\}$$

A commonly used method to characterize relationship between time-to-event data and covariates is Cox (1972) proportional hazard model. It is assumed that the hazard rate of an individual with covariates $\boldsymbol{Z}_i = (Z_{i1}, ..., Z_{ip})$ takes the form,

$$\lambda_i(t) = \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i)\lambda_0(t)$$

Here $\lambda_0(t)$ is a baseline hazard that describes the shape of the hazard rate as a function of time, and $\exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i)$ is a hazard ratio that describes how the size of the hazard rate depends on covariates and $\boldsymbol{\beta}$ is the parameter vector. The survival function for an individual i can be expressed as

$$S_i(t) = \exp\{-\int_0^t \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i)\lambda_0(s)ds\}$$

The covariates are assumed to be fixed over time. It can be generalized to allow time-varying covariates.

$$\lambda_i(t) = \exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)\}\lambda_0(t)$$

The Cox model is considered as a semi-parametric model because the covariate parts assume a functional form whereas the baseline hazard has non-parametric functional form. Although there is no assumption on the distribution of the hazard function, it assumes that the hazard ratio is independent of time, or in other words, the effect of the predictor variables is assumed to be constant over time. This is because for two individuals with covariates $\boldsymbol{Z}_k$ and $\boldsymbol{Z}_{k'}$ respectively, the baseline hazard as a function of time is canceled out.

$$\frac{\lambda_k(t)}{\lambda_{k'}(t)} = \frac{\exp(\boldsymbol{\beta}^T \boldsymbol{Z}_k)\lambda_0(t)}{\exp(\boldsymbol{\beta}^T \boldsymbol{Z}_{k'})\lambda_0(t)} = \frac{\exp(\boldsymbol{\beta}^T \boldsymbol{Z}_k)}{\exp(\boldsymbol{\beta}^T \boldsymbol{Z}_{k'})}$$

The proportionality assumption on the hazard ratio in the Cox model may not always hold. An alternative method is the accelerated failure time (AFT) model which models the effect of the predictor variables on the survival time X_i instead of hazard function $\lambda_i(t)$ by the Cox model.

$$log(X_i) = \boldsymbol{\beta}^T \boldsymbol{Z}_i + \sigma\epsilon_i$$

where σ is the scale parameter and ϵ_i is the random effect, assumed to follow a certain distribution. Hence the AFT model is a parametric model. The survival time X_i can be expressed as a function of ϵ_i.

$$\begin{aligned} S_i(t) &= P(X_i \geq t) \\ &= P\{log(X_i) \geq log(t)\} \\ &= P\{\boldsymbol{\beta}^T \boldsymbol{Z}_i + \sigma\epsilon_i \geq log(t)\} \\ &= P\{\epsilon_i \geq \frac{log(t) - \boldsymbol{\beta}^T \boldsymbol{Z}_i}{\sigma}) \end{aligned}$$

The effect of the covariate (parameter $\boldsymbol{\beta}$) is to accelerate or delay the survival time X_i by a constant amount. The AFT model does require the specification of the distribution of the survival time, and an appropriate distribution may be difficult to identify.

While Cox model is popularly used to capture the association between the time-to-event data and predictor covariates, it is not often used for the event prediction. The non-parametric nature of the baseline hazard function in the Cox model makes the predictions at the end of the follow-up quite unstable. In contrast, the AFT model as a parametric survival model, yields concise, parsimonious formulation, and smoothing of the underlying hazard. This makes the extrapolation readily possible. Hence AFT model can be considered for prediction purpose with time invariant covariates.

Below we review the likelihood function under counting process notation, and how it is related to the Cox partial likelihood.

Denote T_i the true event time for the ith subject, and C_i the censoring time. The observed data is $(X_i, \triangle_i), i = 1, ..., n$, where

$$X_i = \min(T_i, C_i)$$

$$\triangle_i = I(T_i \leq C_i) = \begin{cases} 1 & \text{if } X_i = T_i \leq C_i \\ 0 & \text{if } X_i = C_i < T_i \end{cases}$$

We follow the similar convention in section 3.6 of Klein and Moeschberger (2005) and chapter 5 of Aalen et al. (2008) to construct the likelihood using counting processes. Assume counting processes $N_1(t), N_2(t), ..., N_n(t)$ where $N_i(t) = I(X_i \leq t, \triangle_i = 1)$ with corresponding intensity processes of event $\lambda_{E_1}(t), \lambda_{E_2}(t), ..., \lambda_{E_n}(t)$. The time interval $[0, \tau]$ is partitioned into a number of small intervals $0 = t_0 < t_1 < t_2 < ... < t_K = \tau$ with length $\triangle t$ being small enough that at most one of the counting processes will have a jump in an interval. Let $dN_i(t)$ be the increment of $N_i(t)$,

$$dN_i(t) = \begin{cases} 1 & \text{if } N_i(t+) - N_i(t) = 1 \\ 0 & \text{otherwise} \end{cases}$$

While $\lambda_{T_i}(t)$ is the intensity process of the event at time t, $dN_i(t)$ follows a Bernoulli distribution with probability of $\lambda_{T_i}(t)\triangle t$ of having $dN_i(t) = 1$. The event contribution to the likelihood at a given time t is proportional to

$$\lambda_{T_i}(t)^{dN_i(t)}\{1 - \lambda_{T_i}(t)dt\}^{1-dN_i(t)}$$

In the first part $\triangle t$ is neglected as it would be canceled out by forming likelihood ratios. Given the time interval $(0, \tau]$, the likelihood is

$$\{\prod_{0<t\leq\tau} \lambda_{T_i}(t)^{dN_i(t)}\} \exp\{-\int_0^\tau \lambda_{T_i}(u)du\}$$

Considering all n observations, the likelihood becomes

$$L_E = \{\prod_{i=1}^{n} \prod_{0<t\leq\tau} \lambda_{T_i}(t)^{dN_i(t)}\} \exp\{-\sum_{i=1}^{n} \int_0^\tau \lambda_{T_i}(u)du\}$$

This is only a partial likelihood considering the intensity process of the event but not that of the censor. If the intensity process of event is modeled parametrically as,

$$\begin{aligned} \lambda_{T_i}(t) &= lim_{\triangle t \to 0} \frac{P(T_i < t + \triangle t | T_i \geq t)}{\triangle t} \\ &= I(X_i \geq t) \exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)\} \lambda_0(t) \end{aligned}$$

where $\lambda_0(x)$ is the baseline hazard. The partial likelihood can be rewritten as

$$\begin{aligned}
L_E &= (\prod_{i=1}^{n} \prod_{0<t\leq\tau} [I(X_i \geq t)\exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)\}\lambda_0(t)]^{dN_i(t)}) \\
&\times \exp[-\sum_{i=1}^{n}\int_0^{\tau} I(X_i \geq u)\exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_i(u)\}\lambda_0(u)du] \\
&= (\prod_{i=1}^{n} \prod_{0<t\leq\tau} [\frac{I(X_i \geq t)\exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)\}}{\sum_{j=1}^{n} I(X_j \geq t)\exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_j(t)\}}]^{dN_i(t)}) \\
&\times [\prod_{0<t\leq\tau} \sum_{j=1}^{n} I(X_j \geq t)\exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_j(t)\}]^{dN_i(t)} \\
&\times \exp[-\sum_{i=1}^{n}\int_0^{\tau} I(X_i \geq u)\exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_i(u)\}\lambda_0(u)du]
\end{aligned}$$

The Cox partial likelihood is defined as the leading factor of the above partial likelihood by ignoring the likelihood for the aggregated process.

$$\prod_{i=1}^{n} \prod_{0<t\leq\tau} [\frac{I(X_i \geq t)\exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)\}}{\sum_{j=1}^{n} I(X_j \geq t)\exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_j(t)\}}]^{dN_i(t)}$$

The inference is made by maximizing this partial likelihood. This is predicated on the availability of $\boldsymbol{Z}_i(t)$ for all individuals at each observed time. In reality $\boldsymbol{Z}_i(t)$ is only available intermittently. The missing values can be imputed by most recent value for the individual i, so called the LOCF approach. However Prentice (1982) showed that the LOCF approach could introduce substantial bias into the parameter estimation.

2.2 Competing risks

The concept of competing risks arises when the failure could be due to one of K ($K > 1$) mutually exclusive causes of failure. The partial likelihood derived from section 2.1 only considers the contribution from one cause of failure - the event. Similar as the notation convention in chapter 8 of Kalbfleisch and Prentice (2011), the full likelihood under Cox (1972) model is a product of K

components for each cause of failure.

$$L = \{\prod_{i=1}^{n}\prod_{k=1}^{K}\prod_{0<t\leq\tau}\lambda_{ik}(t)^{dN_{ik}(t)}\}\exp\{-\sum_{i=1}^{n}\int_0^{\tau}\sum_{k=1}^{K}\lambda_{ik}(u)du\}$$

where $\lambda_{ik}(t)$ is the cause-specific intensity function for the k^{th} cause.

$$\lambda_{ik}(t) = lim_{\triangle t\to 0}\frac{P\{t \leq X_i < t + \triangle t, dN_{ik}(t) = 1|X_i \geq t\}}{\triangle t}$$

And $dN_{ik}(t) = 1$ if the failure occurs at time t due to the k^{th} cause. If the cause-specific intensity process is modeled parametrically,

$$\lambda_{ik}(t) = I(X_i \geq t)\exp(\boldsymbol{\beta}_k^T\boldsymbol{Z}_i)\lambda_{0k}(t)$$

where $\lambda_{0k}(t)$ is the k^{th} cause-specific baseline hazard. The full likelihood can be rewritten as

$$L = (\prod_{i=1}^{n}\prod_{k=1}^{K}\prod_{0<t\leq\tau}[I(X_i \geq t)\exp\{\boldsymbol{\beta}_k^T\boldsymbol{Z}_i(t)\}\lambda_{0k}(t)]^{dN_{ik}(t)})$$
$$\times\exp[-\sum_{i=1}^{n}\int_0^{\tau}\sum_{k=1}^{K}I(X_i \geq u)\exp\{\boldsymbol{\beta}_k^T\boldsymbol{Z}_i(u)\}\lambda_{0k}(u)du]$$

The challenges posed by the intermittent longitudinal covariates and interest to characterize the relationship between the longitudinal process and survival process have led to research interest in joint modeling method.

2.3 Joint modeling

There has been growing interest in characterizing the relationship between the time-to-event outcome and the trajectory of the longitudinal observations in diverse fields such as medicine, epidemiology, engineering, and economics. Joint modeling provides a framework to characterize the association between longitudinal and survival processes. Tsiatis and Davidian (2004) and Yu et al. (2004) provided comprehensive overview on this methodology.

The model is generally comprised of two linked submodels through shared random effects, a random effect model for the longitudinal process, and a Cox (1972) proportional hazard model for the survival process. We follow similar notation in Tsiatis and Davidian (2004). The hazard for individual i at time t $\lambda_{T_i}(t)$ is a function of longitudinal data $Z_i(t)$ at time t.

$$\begin{aligned}\lambda_{T_i}(t) &= \lim_{\triangle t \to 0} \frac{P(t < T_i < t + \triangle t | T_i \geq t)}{\triangle t} \\ &= \exp\{\beta Z_i(t)\}\lambda_0(t)\end{aligned}$$

where β is the parameter and $\lambda_0(t)$ is the baseline hazard function. For the longitudinal response process, a standard approach is to characterize $Z_i(t)$ by the subject-specific random effect. For example, the "true" value of the longitudinal data $Z_i(t)$ can be a simple linear function of time t with parameter $\boldsymbol{\theta}_i = (\theta_{0i}, \theta_{1i})$ as the subject-specific intercept and slope.

$$Z_i(t) = \theta_{0i} + \theta_{1i} t$$

Tsiatis and Wulfsohn (1995) and Wulfsohn and Tsiatis (1997) used this simple linear function to model the 'true' log-transformed CD4 trajectories.

One focus of the joint modeling research is to develop more adaptive and flexible models, either to characterize hazard relationship (Tseng et al. (2005) and Tseng et al. (2015)) when the proportionality assumption failed or to model the longitudinal process when the linear mixed model was inadequate to represent the trajectories (Ding and Wang (2008), Yu et al. (2004), Taylor et al. (2005), Yu et al. (2008), and Garre et al. (2008)).

For the event process, to monitor the disease recurrence in prostate cancer patients by measuring prostate-specific antigen (PSA) periodically, both Yu et al. (2004) and Yu et al. (2008) used mixture cure model to account for the proportion of individuals who are cured by the treatment and will never expect events. Tseng et al. (2005) introduced the use of accelerated failure time model under the joint modeling framework when the proportional hazard assumption fails to capture the association between survival time and longitudinal covariates. The hazard rate for an individual takes the form,

$$\lambda_{T_i}(t) = \lambda_0 \int_0^t \exp\{\beta Z_i(s)\} ds \exp\{\beta Z_i(t)\}$$

Tseng et al. (2015) proposed to replace parameter β with β_1 and β_2 to make the model more flexible.

$$\lambda_{T_i}(t) = \lambda_0 \int_0^t \exp\{\beta_1 Z_i(s)\} ds \exp\{\beta_2 Z_i(t)\}$$

When $\beta_1 = 0$, it's a Cox model. When $\beta_1 = \beta_2$, it becomes accelerated failure time model. The baseline hazard function λ_0 is left unspecified, hence it is a semi-parametric model.

For the longitudinal process, more flexible models may also be considered such as the polynomial function of time t, $Z_i(t) = \beta_{0i} + \beta_{1i} t + \cdots + \beta_{pi} t^p$, or more generally,

$$Z_i(t) = \boldsymbol{f}(t)^T \boldsymbol{\theta}_i$$

where the vector $\boldsymbol{f}(t)$ is generalized as a function of time t (including splines). Brown et al. (2005) and Ding and Wang (2008) argued that the proposed non-parametric B-splines to specify the longitudinal model is more flexible. Specifically in Ding and Wang (2008), the longitudinal data $Z_i(t)$ is connected with the population mean function $\mu(t) = E\{Z_i(t)\}$ through subject-specific multiplicative factor θ_i.

$$Z_i(t) = \theta_i \mu(t)$$

where $E(\theta_i) = 1$. The mean function $\mu(t)$ is approximated by the B-spline basis function $B_l(.)$, $1 \leq l \leq L$.

$$\mu(t) \approx \sum_{l=1}^{L} \gamma_l B_l(t) = \boldsymbol{B}^T(t)\gamma$$

Henderson et al. (2000) considered longitudinal process of the form,

$$Z_i(t) = \boldsymbol{f}(t)^T \boldsymbol{\theta}_i + U_i(t)$$

where $U_i(t)$ is a mean-zero stochastic process, specially stationary Gaussian process. This allows the consideration of within-subject serial autocorrelation. Given the specification for $Z_i(t)$, the observed longitudinal data W_i is

$$W_i(t_{ij}) = Z_i(t_{ij}) + e_i(t_{ij})$$

where $e_i(t_{ij})$ is assumed to follow $N(0, \sigma^2)$. The $e_i(t_{ij})$ represents the measurement error and biological variations. The distribution of parameters $\boldsymbol{\theta}_i$ is with density of $p(\boldsymbol{\theta}_i|\boldsymbol{\delta})$ and $\boldsymbol{\delta}$ being the parameters of the assumed distribution. Wulfsohn and Tsiatis (1997) assumed the shared random effect parameters $\boldsymbol{\theta}_i$ to be bivariate normal.

$$\begin{pmatrix} \theta_{0i} \\ \theta_{1i} \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix}, \begin{pmatrix} \sigma_{00} & \sigma_{01} \\ \sigma_{01} & \sigma_{11} \end{pmatrix}\right)$$

The normality assumption on the random effect parameters $\boldsymbol{\theta}_i$ may be considered too strict, so Song et al. (2002) relaxed this assumption and assumed the random effects having a distribution with a smooth density.

Parameter estimation and inference is one area of research (Wulfsohn and Tsiatis (1997), Henderson et al. (2000), Song et al. (2002), and Rizopoulos et al. (2009)). The likelihood under the joint modeling for the full set of parameters $\Omega = \{\lambda_0(.), \beta, \boldsymbol{\theta}, \sigma^2, \boldsymbol{\delta}\}$ takes the form,

$$\begin{aligned}
&\prod_{i=1}^{n} \int \left[\prod_{0<t\leq X_i} f\{X_i, dN_i(t)|\boldsymbol{\theta}_i, \beta, \lambda_0\} \right] \prod_{j=1}^{m_i} f\{W_i(t_{ij})|\boldsymbol{\theta}_i, \sigma^2\} p(\boldsymbol{\theta}_i|\boldsymbol{\delta}) d\boldsymbol{\theta}_i \\
&= \prod_{i=1}^{n} \int \{ \prod_{0<t\leq X_i} [I(X_i \geq t) \exp\{\beta Z_i(t)\}\lambda_0(t)]^{dN_i(t)} \} \\
&\times \exp\left[- \int_0^{X_i} I(X_i \geq u) \exp\{\beta Z_i(u)\}\lambda_0(u) du \right] \\
&\times \frac{1}{(2\pi\sigma^2)^{m_i/2}} \exp\left[- \sum_{j=1}^{m_i} \frac{\{W_i(t_{ij}) - Z_i(t_{ij})\}^2}{2\sigma^2} \right] p(\boldsymbol{\theta}_i|\boldsymbol{\delta}) d\boldsymbol{\theta}_i
\end{aligned}$$

The parameters are estimated by jointly maximizing the likelihood from both processes. Wulfsohn and Tsiatis (1997) proposed to use expectation-maximization algorithm to estimate the parameters. This is done by iterating between an E-step, in which the expected log-likelihood of the complete data is calculated conditioning on the observed data and the current estimate of the parameters Ω, and a M-step in which the parameter estimates are updated by maximizing the log-likelihood. The parameters are estimated in the maximization step by solving the score function of the log-likelihood. There exists close-form estimators for all parameters using maximum likelihood method except β which is updated by Newton-Raphson algorithm in each iteration. Maximizing this likelihood function is computationally challenging. The expectation-maximization (EM) algorithm proposed by Wulfsohn and Tsiatis (1997) to estimate parameters in the joint modeling requires the numerical multidimensional integration. The computation burden increases exponentially with not only the dimensionality of the integration but also the number of individuals as the shared random effects are individual-specific. The numerical complexity of joint models has so far limited their application as a prognostic tool (Pauler and Finkelstein, 2002). One area of the joint modeling research has been to find numerical integration techniques that ease the computation burden, such as Monte Carlo proposed by Henderson et al. (2000), Gauss-Hermite quadrature by Song et al. (2002), and fully exponential Laplace approximations by Rizopoulos et al. (2009).

Wulfsohn and Tsiatis (1997) took the frequentist approach to fit the model, whereas Faucett and Thomas. (1996) took the Bayesian approach using Markov chain Monte Carlo technique of Gibbs sampling to estimate the joint posterior distribution of the unknown parameters of the model. Ibrahim et al. (2001) extended the Bayesian joint modeling approach by proposing a semiparametric cure rate model for the event process. Under the Bayesian framework, Wang and Taylor. (2001) proposed a more flexible longitudinal model that incorporated a mean structure dependent on covariates, a random intercept, a stochastic process, and measurement error. The Bayesian joint modeling approach may be more straightforward as it avoids complicated approximation required by the fequentist approach.

There is also growing interest to utilize joint modeling method as a prediction tool (Taylor et al. (2005),Yu et al. (2008), Taylor et al. (2013), Proust-Lima and Taylor (2009), Rizopoulos (2011), Rizopoulos (2016)). Taylor et al. (2005), Taylor et al. (2013), Yu et al. (2008), Proust-Lima and Taylor (2009) used the joint modeling to predict future serial prostate specific antigen (PSA) values in patients with prostate cancer and to predict the risk of disease progression. Rizopoulos (2011) predicted the time-to-death using longitudinal CD4 cell count in human immunodeficiency virus infected patients.

For the purpose of event prediction, Rizopoulos (2011) estimated the predicted survival probabilities under the Bayesian framework,

$$\begin{aligned}\pi_i(u|t) &= \mathbb{P}(T_i \geq u | T_i > t, \boldsymbol{Z_i}(t), \Omega) \\ &= \int \mathbb{P}(T_i \geq u | T_i > t, \boldsymbol{Z_i}(t), \Omega) p(\Omega | D_n) d\Omega\end{aligned}$$

where $u > t$ and D_n denotes the data on which the joint model was fitted $D_n = \{X_i, \triangle_i, \boldsymbol{Z_i}; i = 1, ..., n\}$. The first part of the integrand is given by

$$\begin{aligned}&\mathbb{P}(T_i \geq u | T_i > t, \boldsymbol{Z_i}(t), \Omega) \\ &\quad = \int \mathbb{P}(T_i \geq u | T_i > t, \boldsymbol{Z_i}(t), \boldsymbol{\theta}_i; \Omega) p(\boldsymbol{\theta}_i | T_i > t, \boldsymbol{Z_i}(t), \Omega) d\boldsymbol{\theta}_i \\ &\quad = \int \frac{S_i[u|\{\boldsymbol{Z_i}(u), 0 \leq u < t\}]}{S_i[t|\{\boldsymbol{Z_i}(u), 0 \leq u < t\}]} p(\boldsymbol{\theta}_i | T_i > t, \boldsymbol{Z_i}(t), \Omega) d\boldsymbol{\theta}_i\end{aligned}$$

The conditional survival probability

$$S_i[t|\{\boldsymbol{Z_i}(u), 0 \leq u < t\}] = \exp\left(- \int_0^t \lambda_{T_i}[s|\{\boldsymbol{Z_i}(u), 0 \leq u < s\}; \beta, \lambda_0(s)] ds \right)$$

For the second part it is assumed that Ω can be approximated by a multivariate normal distribution with mean $\hat{\Omega}$ and covariance matrix $v\hat{a}r(\hat{\Omega})$. Then $\pi_i(u|t)$ is estimated by Monte Carlo method.

Both Rizopoulos (2011) and Rizopoulos (2016) derived accuracy measures under the joint modeling framework to discriminate patients who will have high risk of events of interests within a time interval since the last assessment from those who will not. However the definition of sensitivity and specificity were different between two papers. Rizopoulos (2011) defined the sensitivity as the probability that the longitudinal measure is less than some threshold value $S_i(t,k,c) = \{Z_i(R_{il}) \leq c_l; k \leq l \leq t\}$ given that the timing of the event is within $(t, t+s]$.

$$\text{Sensitivity}(c, t+s) : \mathbb{P}\{S_i(t,k,c)|T_i > t, T_i \in (t, t+s]\}$$

$$\text{Specificity}(c, t+s) : \mathbb{P}\{F_i(t,k,c)|T_i > t, T_i > t+s\}$$

where $F_i(t,k,c) = \{Z_i(R_{il}) > c_l; k \leq l \leq t\}$. Per Rizopoulos (2011) the overall discrimination capability of the longitudinal measures can be assessed using the ROC curve

$$ROC_t^s(p) = TP_t^s\{[FP_t^s]^{-1}(p)\}$$

where p is in [0, 1], $TP_t^s(c)$ denotes the true positive rate, FP_t^s the false positive rate, and $[FP_t^s]^{-1}(p) = \inf_c\{c : FP_t^s(c) \leq p\}$. The corresponding area under the ROC curve (AUC) is obtained by

$$AUC_t^s = \int_0^1 ROC_t^s(p)dp$$

and can be estimated for each subject using rocJM() function in the R package JM.

Rizopoulos (2016) defined the sensitivity and specificity as

$$\text{Sensitivity}(c, t+s) : \mathbb{P}\{\pi_i(t+s|t) \leq c|T_i > t, T_i \in (t, t+s]\}.$$

$$\text{Specificity}(c, t+s) : \mathbb{P}\{\pi_i(t+s|t) > c|T_i > t, T_i > t+s\}.$$

For a randomly chosen pair of subjects i and j both of which had longitudinal measures up to time t, the AUC is defined as

$$AUC_t^s = \mathbb{P}[\pi_i(t+s|t) < \pi_j(t+s|t)|\{T_i \in (t, t+s] \cap \{T_j > t+s\}].$$

2.4 Landmarking analyses

Anderson et al. (1983) proposed to evaluate survival selecting some fixed time after the initiation of therapy as a landmark for conducting the analysis.

Those patients still on study at the landmark time are analyzed based on the probability estimates and statistical tests conditional on the response status of patients at the landmark time. It is called partly conditional modeling by Zheng and Heagerty (2005). This is because rather than conditioning on the entire covariate history, the model conditions on the partial history measured at the landmark time point.

Houwelingen (2007) developed the landmark methodology for dynamic prediction of event. Unlike the joint modeling method, the landmark method does not require separate modeling of the event process and the longitudinal process. The landmark method computes the predictive probability of any event of interest up to a certain horizon t_{hor} at a landmark time point t_{LM}. For a dataset with observations $(t_i, \delta_{T_i}, \mathbf{Z}_i)$ (observation time, event indicator, vector of covariates respectively) for $i = 1, \ldots, n$, the basic landmark model in Houwelingen (2007) is

$$h(t|\mathbf{Z}, t_{LM}) = h_0(t|t_{LM}) \exp\{\mathbf{Z}^T \boldsymbol{\beta}_{LM}\} \text{ for } s = t_{LM} \leq t \leq t_{hor}.$$

This can be used to obtain simple approximations for the probabilities

$$S_{LM}(t_{hor}|\mathbf{Z}, t_{LM}) = \hat{\mathbb{P}}(T > t_{hor}|\mathbf{Z}, T > t_{LM}) = \frac{S(t_{hor}|\mathbf{Z})}{S(t_{LM}|\mathbf{Z})}.$$

The parameter $\boldsymbol{\beta}_{LM}(s)$ can be a linear model,

$$\boldsymbol{\beta}_{LM}(s) = f(s)\theta. \tag{2.4.1}$$

with a suitable set of basis functions $f(s)$ and a vector θ of parameters.

A dataset for landmarking at s can be created by selecting all individuals with $t_i \geq s$. The corresponding partial log-likelihood is given by

$$pl_s\{\boldsymbol{\beta}_{LM}(s)\} = \sum_{t_i \geq s} d_i \left[\left\{ \mathbf{Z}_i^T \boldsymbol{\beta}_{LM}(s) - \ln\{\sum_{t_j \geq t_i} e^{\mathbf{Z}_j^T \boldsymbol{\beta}_{LM}(s)}\} \right\} \right].$$

For different values of s, multiple landmarking datasets can be created. Individual i will have a record in all data sets with $s \leq t_i$. These datasets for different values of s can be merged into one big 'stacked' dataset. The dependence of $\boldsymbol{\beta}_{LM}(s)$ on s can be investigated by fitting a Cox model using a pseudo-partial log-likelihood

$$\sum_s pl_s\{\boldsymbol{\beta}_{LM}(s)\}.$$

Fitting this model is equivalent of maximizing the integrated partial log-likelihood

$$\begin{aligned} ipl\{\boldsymbol{\beta}_{LM}\} &= \int_0^{t_{hor}} pl_s\{\boldsymbol{\beta}_{LM}(s)\}\psi(s)ds \\ &= \sum_{i=1}^{n} d_i\Big[\mathbf{Z}_i^T \int_0^{t_i} \boldsymbol{\beta}_{LM}(s)\psi(s)ds - \int_0^{t_i} \ln\{\sum_{t_j \geq t_i} e^{\mathbf{Z}_j^T \boldsymbol{\beta}_{LM}(s)}\}\psi(s)ds\Big]. \end{aligned} \tag{2.4.2}$$

where $\psi(s)$ is a non-negative function that is introduced to weight possible landmark points in fitting models for $\boldsymbol{\beta}_{LM}(s)$ and to simplify the notation. The parameter θ in (2.4.1) is estimated by maximizing $ipl\{\boldsymbol{\beta}_{LM}\}$ in (2.4.2).

The baseline hazard at the event time t_i can be estimated by a Breslow-type estimator

$$\hat{h}_0(t_i|t_{LM}=s) = \frac{1}{\sum_{t_j \geq t_i} \exp\{\mathbf{Z}_j^T \hat{\boldsymbol{\beta}}_{LM}(s)\}}.$$

The hazard is modeled as

$$h(t|\mathbf{Z}, s) = h_0(t)\exp\{\mathbf{Z}^T \boldsymbol{\beta}_{LM}(s) + \gamma(s)\}.$$

Maximizing with respect to the baseline hazard at the event times leads to a slightly different version of the integrated partial log-likelihood,

$$\begin{aligned} ipl^*\{\boldsymbol{\beta}_{LM}, \gamma\} = \sum_{t_i} d_i \int_0^{t_i} \Big[&\{\mathbf{Z}_i^T \boldsymbol{\beta}_{LM}(s) + \gamma(s)\} \\ &- \ln\{\sum_{t_j \geq t_i} \int_0^{t_i} e^{\mathbf{Z}_j^T \boldsymbol{\beta}_{LM}(s) + \gamma(s)} \psi(s)ds\}\Big]\psi(s)ds. \end{aligned}$$

The corresponding baseline hazard can be estimated by

$$\hat{h}_0^*(t_i) = \frac{\int_0^{t_i}\psi(s)ds}{\sum_{t_j \geq t_i} \int_0^{t_i} \exp\{\mathbf{Z}_j^T \hat{\boldsymbol{\beta}}_{LM}(s) + \hat{\gamma}(s)\}\psi(s)ds}.$$

The landmark survival probabilities can be estimated as,

$$\hat{\mathbb{P}}\{T > t | T \geq s, \mathbf{Z}\} = \exp\left[-\exp\{\mathbf{Z}^T \hat{\boldsymbol{\beta}}_{LM}(s) + \hat{\gamma}(s)\}\{\hat{H}_0^*(t) - \hat{H}_0^*(s-)\}\right].$$

where the cumulative baseline hazard $\hat{H}_0^*(t) = \sum_{t_i \leq t} \hat{h}_0^*(t_i)$.

Houwelingen (2007) argued that fitting a simple Cox model to the data on the interval from the landmark time point t_{LM} to the prediction horizon t_{hor} is an easy and valid way of making prediction models. The regression coefficients in such a model can be seen as some weighted average of the time-varying covariate effects over the interval (t_{LM}, t_{hor}).

Van Houwelingen and Putter (2008) proposed a landmark prediction procedure to incorporate the status of the short-term event by a landmark point t_0. This method was applied in an acute lymphoid leukemia data. Parast et al. (2012) extended the landmark Cox model of Van Houwelingen and Putter (2008) and proposed to incorporate the short-term event time information along with multiple covariates up to a landmark time point via a flexible varying-coefficient model. The event time for the short-term and long-term events are denoted as T_S and T_L. The probability $P(T_L \leq t_0 + \tau | T_L > t_0, T_S, \mathbf{Z}) = P^{(2)}_{t_0,\tau}(T_S, \mathbf{Z})$ for a predefined landmark point t_0 for any given τ given covariates $\mathbf{Z}$ and the information on T_S is estimated as

$$\hat{P}^{(2)}_{t_0,\tau}(T_S, \mathbf{Z}) = g(\hat{\boldsymbol{\beta}}^T_{T_S}\mathbf{Z}) I(T_S \leq t_0) + g(\hat{\boldsymbol{\beta}}^T_{t_0}\mathbf{Z}) I(T_S > t_0).$$

where g is a known, strictly increasing link function and $\boldsymbol{\beta}_{t_0}$ and $\boldsymbol{\beta}_{T_S}$ are parameters allowing the effect of $\mathbf{Z}$ on the risk to vary over the landmark time point t_0 and the short-term outcome respectively. Parast et al. (2012) stated that this varying-coefficient model can potentially approximate the underlying disease process better compared with commonly used semiparametric models. Parast et al. (2014) proposed a two-stage procedure by first using a semiparametric approach to incorporate baseline covariates and intermediate event information observed before some landmark time; and then estimating the marginal survival nonparametrically by smoothing over risk scores derived from the model in the first stage.

As noted in Houwelingen (2007), the landmark analysis typically works with the original time scale, and existing methods generally specify a "global" semiparametric model that covers the entire time-span of the study. Zheng and Heagerty (2005) decoupled the time scale for modeling the hazard from the time scale for accrual of available longitudinal covariate information. Specifically the hazard is converted to the derived time scale, $t^* = t - s_{ik}$, which measures the follow-up time since the measurement of the marker s_{ik}. The regression model for the partly conditional hazard function can take the general form $\lambda_{ik}(t^*|\mathbf{Z}_{ik}, 0 \leq s_{ik} \leq T_i) = g\{\lambda_0(t^*, s), \beta(t^*, s)^T \mathbf{Z}_{ik}\}$, where $g(\lambda, \eta)$ is a link function. The baseline hazard $\lambda_0(t^*, s)$ and the regression coefficient $\beta(t^*, s)$ may be functions of both the time since measurement t^*, and measurement time s. Zheng and Heagerty (2005) stated that their proposed methods allow a flexible characterization of the association between a longitudinal covariate process and a survival time, and facilitate the direct prediction of survival probabilities in the time-varying covariate setting.

2.5 Forward intensity model

Duan et al. (2012) proposed a different and computationally efficient framework that enables direct prediction of time-to-event outcomes using longitudinal data. Although the data that motivated research in Duan et al. (2012) was the Credit Research Initiative (CRI) database which is different from clinical data, a shared interest is to predict events using longitudinal measures. Duan et al. (2012) used monthly financial indexes and company attributes to forecast the chance of default or bankruptcy, whereas clinical longitudinal data are used to predict probabilities of survival or disease progression.

The hazard function definition in Duan et al. (2012) is different from Cox (1972). The hazard function $\lambda_i(t)$ at time t by Cox (1972) is a function of time t and takes the form

$$\lambda_i(t) = \exp(\boldsymbol{\beta}^T \boldsymbol{Z}_i)\lambda_0(t)$$

where $\boldsymbol{Z}_i$ is a vector of measurements for individual i and may be a function of time and $\boldsymbol{\beta}$ is the parameter. In comparison, Duan et al. (2012) defined the forward intensity functions as,

$$\begin{aligned} f_{it}(u) &= \exp\{\varphi_{it}(\tau)(\tau)\} \lim_{\triangle t \to 0} \frac{1}{\triangle t} P_t(t + \tau < X_i = T_i \leq t + \tau + \triangle t) \\ &= \exp[-\ln E_t \exp\{-\int_t^{t+\tau} (\lambda_{is} + \varphi_{is}) ds\}] \\ &\times \lim_{\triangle t \to 0} \frac{1}{\triangle t} \int_{t+\tau}^{t+\tau+\triangle t} \exp\{-\int_t^s (\lambda_{iu} + \phi_{iu}) du\} \lambda_{is} ds \end{aligned}$$

where λ_{iu} and ϕ_{iu} are the intensity of event and exit due to other reasons respectively. Duan et al. (2012) proposed to model the forward intensity functions, $f_{it}(u)$ at a future prediction horizon u from time t for individual i, as exponential functions of longitudinal measures $\boldsymbol{Z}_i(t)$ at time t and horizon-specific parameters $\boldsymbol{\alpha}(u)$.

$$f_{it}\{\boldsymbol{\alpha}(u)\} = \exp\{\boldsymbol{\alpha}^T(u)\boldsymbol{Z}_i(t)\} = \exp\{\alpha_0(u) + \alpha_1(u)Z_{i1}(t) + ... + \alpha_k(u)Z_{ik}(t)\}$$

The forward intensity of exit due to other reason for prediction horizon u from time t for individual i is modeled as,

$$k_{it}\{\boldsymbol{\beta}(u)\} = \exp\{\boldsymbol{\beta}^T(u)\boldsymbol{Z}_i(t)\} = \exp\{\beta_0(u) + \beta_1(u)Z_i(t) + ... + \beta_k(u)Z_{ik}(t)\}$$

The forward intensity of event $f_{it}(u)$ and exit due to other reasons $k_{it}(u)$ were considered as two forms of competing risks and modeled as two independent doubly stochastic processes. The forward intensity functions determine the future event probability.

T_i is denoted as default time which is equivalent of event time. The individual could exit the study due to other reasons. X_i is the combined exit time. If an individual exits the study due to event (default), then $X_i = T_i$, otherwise $X_i < T_i$. The pseudo-likelihood function for future u-unit is expressed as

$$L_\tau(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{i=1}^{N} \prod_{t=0}^{T-1} L_{\tau,i,t}(\boldsymbol{\alpha}, \boldsymbol{\beta})$$

where

$$
\begin{aligned}
L_{\tau,i,t}(\boldsymbol{\alpha},\boldsymbol{\beta}) = {} & 1_{(t_{0i}\le t; X_i > t+\tau)} P_t(X_i > t+\tau) \\
& + 1_{(t_{0i}\le t; T_i = X_i \le t+\tau)} P_t(X_i; T_i = X_i \le t+\tau) \\
& + 1_{(t_{0i}\le t; T_i \ne X_i, X_i \le t+\tau)} P_t(X_i; X_i \ne T_i \text{ and } X_i \le t+\tau) \\
& + 1_{(t_{0i} > t)} + 1_{(X_i < t)}
\end{aligned}
$$

The first term in the pseudo-likelihood is the probability of surviving beyond $t+\tau$. The second term is the probability of having an event by time $t+\tau$. The third term is the probability of exiting the study due to reasons other than having an event by time $t+\tau$. If the individual does not appear in the sample at time t, then the contribution to the likelihood function is 1 which is transformed to 0 in the log-likelihood function. Duan et al. (2012) proposed

$$
P_t(X_i > t+\tau) = \exp\left(-\sum_{u=0}^{\tau-1} [f_{it}\{\boldsymbol{\alpha}(u)\} + k_{it}\{\boldsymbol{\alpha}(u)\}]\triangle t \right)
$$

$$
P_t(X_i; X_i = T_i \le t+\tau)
$$
$$
= \begin{cases}
1 - \exp\left[- f_{it}\{\boldsymbol{\alpha}(0)\}\triangle t \right] & \text{if } X_i = t+1 \\
\exp\left(- \sum_{u=0}^{X_i - t - 2} [f_{it}\{\boldsymbol{\alpha}(u)\} + k_{it}\{\boldsymbol{\alpha}(u)\}]\triangle t \right) & \\
\quad \times \left(1 - \exp[-f_{it}\{\boldsymbol{\alpha}(X_i - t - 1)\}\triangle t] \right) & \text{if } t+1 < X_i \le t+\tau
\end{cases}
$$

$$P_t(X_i; X_i \neq T_i, \text{ and } X_i \leq t+\tau)$$

$$= \begin{cases} \exp\Big[-f_{it}\{\boldsymbol{\alpha}(0)\}\triangle t\Big] \\ -\exp\Big(-[f_{it}\{\boldsymbol{\alpha}(0)\}+k_{it}\{\boldsymbol{\alpha}(0)\}]\triangle t\Big) & \text{if } X_i = t+1 \\ \exp\Big(-\sum_{u=0}^{X_i-t-2}[f_{it}\{\boldsymbol{\alpha}(u)\}+k_{it}\{\boldsymbol{\alpha}(u)\}]\triangle t\Big) \\ \quad\times\Big\{\exp[-f_{it}\{\boldsymbol{\alpha}((X_i-t-1)\}\triangle t] \\ \quad -\exp(-[f_{it}\{\boldsymbol{\alpha}((X_i-t-1)\} \\ \quad +k_{it}\{\boldsymbol{\alpha}((X_i-t-1)\}]\triangle t)\Big\} & \text{if } t+1 < X_i \leq t+\tau \end{cases}$$

The $\triangle t$ is the increment of time.

The parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ will be estimated by maximizing the likelihood function. As the likelihood is the product of two separate terms involving $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, the parameter estimates $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ can be achieved by maximizing two separate likelihood functions independently. Furthermore, the pseudo-likelihood function can be further decomposed to separate terms for $\boldsymbol{\alpha}(u)$ and $\boldsymbol{\beta}(u)$ for different values of u. The future u-unit specific likelihood functions are

$$L\{\boldsymbol{\alpha}(u)\} = \prod_{i=1}^{N}\prod_{t=0}^{T-u-1} L_{i,t}\{\boldsymbol{\alpha}(u)\}$$

and

$$L\{\boldsymbol{\beta}(u)\} = \prod_{i=1}^{N}\prod_{t=0}^{T-u-1} L_{i,t}\{\boldsymbol{\beta}(u)\}$$

where $u = 0, 1, ..., \tau - 1$.

Here

$$
\begin{aligned}
L_{i,t}\{\boldsymbol{\alpha}(u)\} = {} & 1_{(t_{0i}\le t; X_i > t+u+1)} \exp\left[-f_{it}\{\boldsymbol{\alpha}(u)\}\triangle t\right] \\
& + 1_{(t_{0i}\le t; T_i = X_i = t+u+1)}\Big(1-\exp\left[-f_{it}\{\boldsymbol{\alpha}(u)\}\triangle t\right]\Big) \\
& + 1_{(t_{0i}\le t; T_i \neq X_i, X_i = t+u+1)} \exp\left[-f_{it}\{\boldsymbol{\alpha}(u)\}\triangle t\right] \\
& + 1_{(t_{0i} > t)} + 1_{(X_i < t+u+1)}
\end{aligned}
$$

and

$$
\begin{aligned}
L_{i,t}\{\boldsymbol{\beta}(u)\} = {} & 1_{(t_{0i}\le t; X_i > t+u+1)} \exp\left[-k_{it}\{\boldsymbol{\beta}(u)\}\triangle t\right] \\
& + 1_{(t_{0i}\le t; T_i = X_i = t+u+1)} \\
& + 1_{(t_{0i}\le t; T_i \neq X_i, X_i = t+u+1)}\Big(1-\exp\left[-k_{it}\{\boldsymbol{\beta}(u)\}\triangle t\right]\Big) \\
& + 1_{(t_{0i} > t)} + 1_{(X_i < t+u+1)}
\end{aligned}
$$

The decomposition of the likelihood function makes the parameter estimation less computationally intensive.

Once the parameters are estimated from maximizing the likelihood functions, Duan et al. (2012) proposed to use the following quantities to evaluate the risk and predict events.

The forward event probability at time t for the period of $[t+\tau, t+\tau+1]$ is defined as,

$$
\begin{aligned}
& P_t(t+\tau < T_i = X_i \le t+\tau+1) \\
& = \exp\Big(-\sum_{u=0}^{\tau-1}[f_{it}\{\boldsymbol{\alpha}(u)\} + k_{it}\{\boldsymbol{\alpha}(u)\}]\triangle t\Big)\Big(1-\exp\left[-f_{it}\{\boldsymbol{\alpha}(\tau)\}\triangle t\right]\Big)
\end{aligned}
$$

The cumulative event probability at time t for the period $[t, t+\tau]$ is defined as

$$
\begin{aligned}
& P_t(t < T_i = X_i \le t+\tau) \\
& = \sum_{u=0}^{\tau-1}\exp\Big(-\sum_{u=0}^{u-1}[f_{it}\{\boldsymbol{\alpha}(u)\} + k_{it}\{\boldsymbol{\alpha}(u)\}]\triangle t\Big)\Big(1-\exp\left[-f_{it}\{\boldsymbol{\alpha}(s)\}\triangle t\right]\Big)
\end{aligned}
$$

It is necessary to have complete knowledge of the entire trajectory of the longitudinal measures for each individual in order to maximize the composite likelihood and consequently estimate the parameter estimates. Duan et al. (2012) used the LOCF approach to impute any missing values.

2.6 Missing longitudinal data

In clinical setting, it's a common issue to have missing and irregularly scheduled longitudinal measurements and the last observation carried forward approach (LOCF) approach could introduce substantial bias in parameter estimation (Prentice, 1982). The LOCF approach which naively imputed the missing longitudinal value by the most recent observed data.

To properly handle the missing data, one can use kernel smoothed approach that appropriately adjusts the contribution to the likelihood from multiple observations near the missing one. The kernel smoothed approach is related the local likelihood method proposed by Tibshirani and Hastie (1987) to capture the nonlinear dependencies. For random variables X and Y with $Y|X = x \sim f(Y, \theta)$, the local likelihood is defined

$$\begin{aligned} L(\theta_1, ..., \theta_n) &= \prod_{i=1}^{n} f(y_i; \theta_i) \\ &= \prod_{i=1}^{n} f\{y_i; s(x_i)\} \\ &= \prod_{i=1}^{n} f\{y_i; \beta_{0i} + \beta_{1i} x_i\} \end{aligned}$$

The θ is a parametric function of x in the likelihood function. Instead of assuming $\theta_i = \beta_0 + \beta_1 x_i$ in a standard linear regression, θ_i is assumed to be $\theta_i = s(x_i) = \beta_{0i} + \beta_{1i} x_i$ where parameters β_{0i} and β_{1i} are estimated by maximizing the local likelihood $L_i(\beta_{0i}, \beta_{1i})$

$$L_i(\beta_{0i}, \beta_{1i}) = \prod_{j \in N_i} f(y_j, \beta_{0i} + \beta_{1i} x_j)$$

Given the value of i, the local likelihood $L_i(\beta_{0i}, \beta_{1i})$ considers contributions from the nearest neighborhood $j \in N_i$. The local likelihood method produces a smooth estimate of the curve $s(.)$ at the points $x_1, ..., x_n$. This method was applied to both the generalized linear models and the Cox proportional hazard model by Tibshirani and Hastie (1987).

The local partial likelihood introduced by Tibshirani and Hastie (1987) considers the nearest neighbor of uniform windows, and Fan et al. (1997) generalized it by introducing nonuniform kernels into the local partial likelihood. It considersed two scenarios when the baseline hazard function was either parametric or non-parametric. The former used the knowledge of the baseline hazard function while the latter one ignored it. The asymptotic normality of the estimators was established.

Staniswalis (1989) and Hunsberger (1994) generalized the local likelihood theory and extended the kernel-weighted likelihood approach in regression analysis. The log-likelihood of the sample (x_i, Y_i) for $i = 1, ..., n$ is $\sum_{i=1}^{n} logf(Y_i, \lambda_i)$ where $x_i \in [0,1]^d$ are lattice points and Y_i are the independent random variables from a family of distributions with parameter $\lambda_t = g(x_i)$. Staniswalis (1989) proposed the estimator $\hat{\lambda}_0$ that maximizes the weighted likelihood function

$$\sum_{i=1}^{n} W[(x_0 - x_i)/b] \log f(Y_i, \lambda)$$

with f the density of Y_i, W a symmetric kernel with compact support, and b the bandwidth controls the size of the neighborhood. The kernel estimator is developed for censored survival times under the assumption of the Cox proportional hazards model.

Cai and Sun (2003) and Tian et al. (2005) adopted kernel weighting approach in the localized Cox regression likelihood. The smoothing occurs at the population level where the same weights are applied to all individuals. The coefficients were assumed to vary over time $\boldsymbol{\beta}(t)$.

$$\lambda_i(t) = \exp\{\boldsymbol{\beta}^T(t)\boldsymbol{Z_i}\}\lambda_0(t)$$

Recall that the log-likelihood of the partial likelihood function under Cox model in section 2.1 can be written as

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{\tau} \left\{ I(X_i \geq t)\boldsymbol{\beta}^T \boldsymbol{Z}_i(t) - \log[\sum_{j=1}^{n} I(X_j \geq t) \exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_j(t)\}] \right\} dN_i(t)$$

Denote $\tilde{\boldsymbol{Z}}_i(u, u-t) = \boldsymbol{Z}_i(u)(1, u-t)^T$. The weighted local linear partial likelihood function becomes,

$$\begin{aligned} l^*(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{\tau} K_h(u-t) \Big\{ & I(X_i \geq u)\boldsymbol{\beta}^T \tilde{\boldsymbol{Z}}_i(u, u-t) \\ & - \log[\sum_{j=1}^{n} I(X_j \geq u) \exp\{\boldsymbol{\beta}^T \tilde{\boldsymbol{Z}}_j(u, u-t)\}] \Big\} dN_i(u) \end{aligned}$$

Cao et al. (2015) extended the work done by Cai and Sun (2003) and Tian et al. (2005) and the smoothing occurs at the individual level. It is assumed that the covariates change over time $\boldsymbol{Z}_i(t)$ and the $\boldsymbol{\beta}$ are fixed.

$$\lambda_i(t) = \exp\{\boldsymbol{\beta}^T \boldsymbol{Z_i}(t)\}\lambda_0$$

Unlike the LOCF method, the contribution of a covariate observation to the likelihood was weighted by the distance to the event time by the kernel smoothing method in Cao et al. (2015). With potential irregular observed longitudinal covariates, Cao et al. (2015) considered those covariates with observation time before time u and contribution of covariates to the partial likelihood is weighted by the distance to the time of interest. The smoothed log-likelihood becomes,

$$\begin{aligned} & l^*(\boldsymbol{\beta}) \\ & = \sum_{i=1}^{n} \int_0^{\tau} \sum_{k=1}^{M_i} K_{h_n}(u - R_{ik}) I(R_{ik} \leq u) \times \Big\{ I(X_i \geq t)\boldsymbol{\beta}^T \boldsymbol{Z}_i(t) \\ & - log[\sum_{j=1}^{n} \sum_{l=1}^{M_j} K_{h_n}(u - R_{jl}) I(R_{jl} \leq u) I(X_j \geq t) \exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_j(t)\}] \Big\} dN_i(t) \end{aligned}$$

where $K_{h_n}(t) = K(t/h_n)/h_n$ is the kernel function with $K(t)$ being a symmetric probability density. The h_n is the bandwidth. The weighting approach reduces the impact of the covariates that are distant in time from time of interest.

The kernel smoothed estimators in Cao et al. (2015) under proper choice of bandwidth were proven to be consistent and asymptotic normal. The simulations demonstrated that the kernel smoothing method had smaller bias than LOCF.

2.7 B-spline

B-spline function is applied extensively in curve and shape optimization methods. Boor (1978) gave comprehensive review of the B-spline basics. Here we review some basics of the B-spline.

B-splines are constructed from polynomial pieces, joined at a certain values of u called the knots. Let $\boldsymbol{u} = \{u_i\}$ be a nondecreasing real sequence. Given a non-decreasing knot sequence u_i and u_{i+1}, the B-splines of order 1 are defined by

$$B_{i,1}(u) = X_i(u) = \begin{cases} 1, & \text{if } u_i \leq u < u_{i+1} \\ 0, & \text{otherwise} \end{cases}.$$

Note that the function is right-continuous. The only constraint is that the sum of these B-splines should add up to 1, i.e., $\sum_i B_{i,1}(u) = 1$, for all u. The ith B-spline of order q denoted by $B_{i,q}(u)$ can be obtained by recursion formula.

$$B_{i,q}(u) = w_{i,q}B_{i,q-1}(u) + (1 - w_{i+1,q})B_{i+1,q-1}(u). \tag{2.7.1}$$

in which

$$w_{i,q}(u) = \begin{cases} \frac{u-u_i}{u_{i+q-1}-u_i} & \text{if } u_i \neq u_{i+q-1} \\ 0 & \text{otherwise} \end{cases}.$$

It is common to call a B-spline with order q as a B-spline with degree $q-1$. The positions at which the B-spline segments join are called the knots. Once the knots are chosen, B-spline can be calculated recursively for any degree of polynomials. Hence the second-order B-spline is,

$$\begin{aligned}B_{i,2} &= w_{i2}B_{i,1} + (1 - w_{i+1,2})B_{i+1,2} \\ &= w_{i,2}X_i + (1 - w_{i+1,q})X_{i+1}.\end{aligned}$$

which consists of two linear pieces that join continuously to form a piecewise linear function that vanishes outside the interval $[t_i, \dots, t_{i+1})$. The third-order B-spline is,

$$\begin{aligned}B_{i,3} &= w_{i3}B_{i2} + (1 - w_{i+1,3})B_{i+1,2} \\ &= w_{i3}w_{i,2}B_{i,1} + \{w_{i3}(1 - w_{i+1,2}) + (1 - w_{i+1,3})w_{i+1,2}\}B_{i+1,1} \\ &+ (1 - w_{i+1,3})(1 - w_{i+2,2})B_{i+2,1} \\ &= w_{i3}w_{i2}X_i + \{w_{i3}(1 - w_{i+1,2}) + (1 - w_{i+1,3})w_{i+1,2}\}X_{i+1} \\ &+ (1 - w_{i+1,3})(1 - w_{i+2,2})X_{i+2}.\end{aligned}$$

which consists of three quadratic pieces that vanishes outside $[t_i, \dots, t_{i+2})$. After $q - 1$ steps of recurrence, $B_{i,q}$ is in the form

$$B_{i,q} = \sum_{j=i}^{i+q-1} b_{j,q}X_j$$

with each $b_{j,q}$ a polynomial of degree $< q$ since it is the sum of products of $q-1$ linear polynomials. Hence a B-spline of order q $B_{i,q}$ is a piecewise polynomial function of degree $q - 1$ which vanishes outside the interval $[u_i, u_{i+q})$ and has possible $q + 1$ knots $u_i, \dots, u_{i+q}$.

If each knot is separated by the same distance h $(= u_{i+1} - u_i)$ i.e. $\boldsymbol{u} = c(\dots, -2, -1, 0, 1, 2, \dots)$, the corresponding B-splines are called 'uniform' (or cardinal B-spline).

A spline of order q with knot sequence $\boldsymbol{u}$ is a linear combination of the B-splines $B_{i,q}$. $S_{q,\boldsymbol{u}}$ is denoted the collection of all such splines.

$$S_{q,\boldsymbol{u}} = \sum_{i=1}^{I} a_iB_{i,q}$$

where $a_i \in \mathbb{R}$ and is also called the control point. Let p be the number of internal knots within a closed interval. The total number of B-spline basis is $I = q+p$. The estimation of the function $S_{q,\boldsymbol{u}}$ is then reduced to the estimation of a set of coefficients a_i which can be estimated using the least square method.

$$\hat{\boldsymbol{a}} = \mathrm{argmin}_{\boldsymbol{a}} \left[\sum_{j=1}^{m} \{y_j - \sum_{i=1}^{I} a_i B_{i,q}(u_j)\}^2 \right] = \mathrm{argmin}_{\boldsymbol{a}} ||\boldsymbol{y} - \boldsymbol{B}\boldsymbol{a}||^2$$

where y_j is the observed data and $\boldsymbol{B} = [B_{i,q}]$ is the so called basis matrix. This is a standard linear regression problem and the solution is well known: $\hat{\boldsymbol{a}} = (\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}\boldsymbol{y}$.

Important properties of B-spline basis function of order q are summarized below,

- The B-splines of order q with a given knot sequence that do not vanish over some knot interval are linearly indpendent with each other over this interval;

- The derivative of a single B-spline of order q is a function of B-splines of order $q-1$ and is given by.
$$\frac{dB_{i,q}(u)}{du} = (q-1)\{\frac{-B_{i+1,q-1}(u)}{u_{i+q} - u_{i+1}} + \frac{B_{i,q-1}(u)}{u_{i+q-1} - u_i}\}$$

- At the knots (joining points), derivatives of the polynomial pieces up to order $q-1$ are continuous;

- The B-splines of order q (degree $q-1$) consists of q polynomial pieces, each of degree $q-1$;

- Each B-spline basis function is non-zero on only a few adjacent subintervals, therefore the B-spline functions are a local approximation method.

If the number of internal knots and location of the knots are given, the B-spline functions are determined. How to choose the optimal number and position of knots remain a question. The flexibility of spline modeling comes at the price of a number of tuning parameters. One area of research was to optimize the number and location of knots. The higher the number of internal knots, the more flexible is the spline function but it may also results in overfitting the data. On the contrary, the number of internal knots being too low could bring in bias in the estimation.

In case of non-homogeneous curves or curves with non-trivial points, the least square method with uniform knots could result in overfitting the data. In order to identify the number of internal knots and their respective locations in the non-uniform B-spline curve fitting, Tjahjowidodo (2017) proposed a two-step method via local algorithm. In the first step, the data is split using a bisecting method and the subset data are fitted by a single-piece B-spline with the pre-defined error bound criterion. Secondly, the knots are optimized by employing a local non-linear least square technique in order to identify the knots positions and continuity levels.

O'Sullivan (1988) proposed the penalized splines to facilitate the selection of the knots. The penalized spline method includes a penalty term on the squared second order derivative to restrict the flexibility of the fitted curve.

$$\begin{aligned}\hat{\boldsymbol{a}} &= \operatorname{argmin}_{\boldsymbol{a}}\left[\sum_{j=1}^{m}\{y_j - \sum_{i=1}^{I} a_i B_{i,q}(u_j)\}^2 + \lambda \int_{u_{min}}^{u_{max}} \{\sum_{i=1}^{I} a_i B''_{i,q}(u)\}^2 du\right] \\ &= \operatorname{argmin}_{\boldsymbol{a}}\left[||\boldsymbol{y} - \boldsymbol{B}\boldsymbol{a}||^2 + \lambda \boldsymbol{a}^T \boldsymbol{P}\boldsymbol{a}\right]\end{aligned}$$

The matrix $\boldsymbol{P}$ can be derived as integrals of products of second derivatives of neighboring B-splines and $\boldsymbol{a}^T\boldsymbol{P}\boldsymbol{a} = \int_{u_{min}}^{u_{max}}\{\sum_{i=1}^{I} a_i B''_{i,q}(u)\}^2 du$. The parameter λ sets the influence of the penalty. The parameter estimate $\hat{\boldsymbol{a}}$ is optimized by minimizing this function which implies a trade-off between the model fit and smoothness of the curve. Hence the larger the λ, the smoother the curve. The parameter vector $\boldsymbol{a}$ can be solved from the equation.

$$(\boldsymbol{B}^T\boldsymbol{B} + \lambda\boldsymbol{P})\hat{\boldsymbol{a}} = \boldsymbol{B}^T\boldsymbol{y}$$

The idea is to use large number of internal knots and let the tuning parameter λ to control the amount of smoothness. Lower or higher orders of derivatives can be used as well.

The computation of $\boldsymbol{P}$ can be quite tedious when the third or higher order of derivatives are used in the penalty term. Wand and Ormerod (2008) extended O'Sullivan's idea to higher orders of derivatives using a computer algebra system.

Eilers and Marx (1996) simplified the penalized B-spline method proposed by O'Sullivan (1988) and based the penalty on the sum of squares of differences between coefficients of adjacent B-splines.

$$\begin{aligned}\hat{\boldsymbol{a}} &= \mathrm{argmin}_{\boldsymbol{a}}\left[\sum_{j=1}^{m}\{y_j - \sum_{i=1}^{I} a_i B_{i,q}(u_j)\}^2 + \lambda \sum_{i=q+1}^{I}\{\Delta^q a_i\}^2\right] \\ &= \mathrm{argmin}_{\boldsymbol{a}}\left[||\boldsymbol{y} - \boldsymbol{B}\boldsymbol{a}||^2 + \lambda||\boldsymbol{D}_q\boldsymbol{a}||^2\right]\end{aligned}$$

in which $\Delta a_i = a_i - a_{i-1}$ and $\boldsymbol{D}_q$ is a matrix such that $\boldsymbol{D}_q\boldsymbol{a} = \Delta^q\boldsymbol{a}$. The second-order difference operator is defined as $\Delta^2 a_i = (a_i - a_{i-1}) - (a_{i-1} - a_{i-2})$ and so on for higher order q. Mostly q of 2 or 3 is used. The approach Eilers and Marx (1996) proposed is also called P-splines. It circumvents the issue by dropping derivatives and integrals completely. When λ equals to 0, it reduces again to a linear regression on the B-spline basis functions. Eilers and Marx (1996) stated that it is straightforward to extend the difference penalty to higher orders differences Δ^q with $q > 2$ whereas the penalty proposed by O'Sullivan (1988) requires complex equations for higher order B-splines. Also Eilers and Marx (1996) showed very strong connection between a penalty on second-order differences of the B-spline coefficients and O'Sullivan's choice of a penalty on the second derivative of the fitted function. To optimize the weight of penalty λ, Eilers and Marx (1996) used cross-validation and the Akaike information criterion (AIC). Eilers et al. (2015) gave a comprehensive review of the development of P-splines in the last two decades.

CHAPTER 3

METHODOLOGY

3.1 Forward intensity

We are interested in the time, denoted by T, when the event of interest happens. As a prediction problem, it is most interesting to consider that T occurs after some specified time t. Here t, for example, can naturally be taken as the current time, so that the future time after t until the event happens draws the key attention. To address this prediction problem, we intend to investigate some models for the conditional distribution of T given that the event time is no earlier than the time t, i.e., $\mathbb{P}(T > t + \tau | T > t)$. Clearly, evaluating this quantity is the key for predicting the future survival probabilities.

We first introduce the concept of the forward intensity function. Denote by $(\mathcal{F}_t)_{t \geq 0}$ the information observed up to time t is the sub-sigma-algebra satisfying $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $0 \leq s < t$ and $\mathcal{F}_t = \mathcal{F}_{t+}$. Let $N_T(u)$ $(u \geq 0)$ be the event counting process such that $dN_T(u) = 1$ if $N_T(u+) - N_T(u) = 1$. For any t $(t \geq 0)$, and given $T > t$, we consider the forward looking counting process originated at t as $N_T(t + u)$ concerning any u $(u \geq 0)$ unit time after t:

$$dN_T(t+u) = \begin{cases} 1 & \text{if } N_T\{(t+u)+\} - N_T(t+u) = 1 \\ 0 & \text{otherwise} \end{cases}. \tag{3.1.1}$$

For any specific t, let

$$\begin{aligned} d\Lambda_{T,t}(u) &= \mathbb{E}[\{N_T(t+u+du) - N_T(t+u)\}|\mathcal{F}_t, T > t] \\ &= \mathbb{P}\{dN_T(t+u) = 1|\mathcal{F}_t, T > t\}. \end{aligned} \tag{3.1.2}$$

We define

$$M_T(t+\tau) = N_T(t+\tau) - \Lambda_{T,t}(\tau) \tag{3.1.3}$$

where $\Lambda_{T,t}(\tau) = \int_0^\tau d\Lambda_{T,t}(u)$ is defined as the cumulative forward intensity function for the time interval $(t, t+\tau)$. Then, conditioning on $T > t$ and $\mathcal{F}_t$,

$$\mathbb{E}[dM_T(t+u)|\mathcal{F}_t, T > t] = \mathbb{E}[\{dN_T(t+u) - d\Lambda_{T,t}(u)\}|\mathcal{F}_t, T > t] = 0.$$

That is, $M_T(t+u)$ is a martingale given $T > t$ and $\mathcal{F}_t$. If we assume that $\Lambda_{T,t}(u)$ is a differentiable function, then the u-time future forward intensity function after time t exists and is defined as

$$\lambda_{T,t}(u) = \frac{d\Lambda_{T,t}(u)}{du}. \tag{3.1.4}$$

the expected event rate at time $T = t+u$ conditioning on $T > t$ and $\mathcal{F}_t$. Furthermore, it is seen that the time t-specific conditional future event probability can be expressed as a function of the forward intensity, for example:

$$S_{T,t}(\tau) = \mathbb{P}(T > t+\tau|\mathcal{F}_t, T > t) = \exp\{-\Lambda_{T,t}(\tau)\} = \exp\left\{-\int_0^\tau \lambda_{T,t}(u)du\right\}. \tag{3.1.5}$$

3.2 Modeling based on the forward intensity function

A crucial implication from (3.1.5) is that the forward intensity function determines the future event probability. Most remarkably, by its definition in (3.1.2), the forward intensity function at t, $\lambda_{T,t}(u)$ $(u \geq 0)$ is a $\mathcal{F}_t$-measurable function. Hence if $\lambda_{T,t}(\cdot)$ is modeled as a function involving only information up to time t, and estimated with data, the predictive distribution of the future event time can be analytically evaluated using (3.1.5). That is, the forward intensity function is a convenient device for assessing the future event probability after t, but requiring only information up to time t.

To model the forward intensity function, we denote by $\boldsymbol{Z}(t) = \{1, Z_1(t), \ldots, Z_k(t)\}^{\mathrm{T}}$ the vector containing predictor variables at time t. The specification of $\mathbf{Z}(t)$ is flexible, allowing its dependence on t. Candidates of $\boldsymbol{Z}(t)$ may include subject specific longitudinal measurements, time t itself, and broad functions of t. In our framework, we consider a parametric model for the forward intensity function $\lambda_{T,t}(u)$:

$$\lambda_{T,t}(u) = \exp\{\boldsymbol{\alpha}(u)^{\mathrm{T}}\boldsymbol{Z}(t)\} = \exp\{\alpha_0(u) + \alpha_1(u)Z_1(t) + \cdots + \alpha_k(u)Z_k(t)\}. \tag{3.2.1}$$

The implication from (3.2.1) is that given the predictor variables $\boldsymbol{Z}(t)$ at time t, the u-time future event intensity depends on u only through $\boldsymbol{\alpha}(u)$ – the model parameter. The predictor variables can also be generalized to include baseline covariates.

The exponential form (3.2.1), ensuring positive $\lambda_{T,t}(u)$, is a conventional device in survival analysis, and it is related to the notion of proportional hazard; and it was applied in Duan et al. (2012). In the context of the forward intensity function, the form (3.2.1) for $\lambda_{T,t}(u)$ is also reasonable. As a concrete example, we illustrate it with a case that T follows the Gompertz distribution. Then by the definition, it is straightforward to show that the conditional survival function is

$$S_{T,t}(\tau) = \mathbb{P}(T > t + \tau | T > t) = \exp\left[\frac{b}{a}\{e^{at} - e^{a(t+\tau)}\}\right],$$

where $a > 0$ and $b > 0$ are the shape and rate parameters. Given the relationship between the forward intensity function and conditional survival function in (3.1.5), the corresponding forward intensity function at t is $\lambda_{T,t}(u) = \exp\{\log(b) + au + at\}$, which can be expressed as $\lambda_{T,t}(u) = \exp\{\alpha_0(u) + \alpha_1(u)t\}$ with $\alpha_0(u) = \log(b) + au$ and $\alpha_1(u) = a$. The intercept $\alpha_0(u) = \log(b) + au$ indicates that the baseline future event intensity increases when u gets larger.

Furthermore, $\lambda_{T,t}(u)$ is an increasing function of t. Both implications are quite sensible, reflecting the expected increment of the risk associated with later time (u) in the future, and longer survival time (t).

When more predictor variables are available at time t, they can be conveniently included in (3.2.1) by appropriately using the form $\boldsymbol{\alpha}(u)^{\mathrm{T}}\boldsymbol{Z}(t)$. Through the exponential additive form, it can be interpreted as some scaling effects from the predictor variables on the forward intensity function $\lambda_{T,t}(u)$. We note that besides the exponential form, other functions can also be broadly investigated for parametrically modeling the forward intensity function; and related problems are fit for separate investigations of their own interests.

An implication from (3.2.1) is that the model is developed "locally" for each pair of t and u via $\boldsymbol{\alpha}(u)$. We note that the parameter $\boldsymbol{\alpha}(u)$ ($u > 0$) is a function of u on its entire support, if we consider the problem in an ideal continuous time setting. In reality, the estimation of the forward intensity function on its entire support is done approximately and so is the evaluation of the conditional future survival probability by (3.1.5); see our development in later sections for more details.

3.3 Competing risks

Other than the event time T of interests, there are typically other competing event time(s), so that only the first event time is observable. We consider the competing risk setting under the forwarding intensity framework. For ease in presentation, we consider one type of competing event with O denoting the time of the other event. There is no fundamental difficulty generalizing the setting to multiple categories of competing events.

Analogous to (3.1.1), we define the counting process $N_O(t+u)$ conditioning on $O > t$ originated at t for the competing event as

$$dN_O(t+u) = \begin{cases} 1 & \text{if } N_O\{(t+u)+\} - N_O(t+u) = 1 \\ 0 & \text{otherwise} \end{cases}. \tag{3.3.1}$$

Additionally, the forward intensity function $\lambda_{O,t}(u)$ is similarly defined to that in (3.1.2)-(3.1.4). Similar to (3.2.1), we parametrically model $\lambda_{O,t}(u)$ as

$$\lambda_{O,t}(u) = \exp\{\boldsymbol{\beta}(u)^{\mathrm{T}}\boldsymbol{Z}(t)\} = \exp\{\beta_0(u) + \beta_1(u)Z_1(t) + \cdots + \beta_k(u)Z_k(t)\}. \tag{3.3.2}$$

We note that $\lambda_{T,t}(u)$ and $\lambda_{O,t}(u)$ may include different predictor variables. For simplicity in notations, we do not differentiate them in our presentation.

Let $X = \min(T, O)$ be the observed event time. Conditioning on $X > t$ and $\mathcal{F}_t$, we have the conditional survival probability beyond another τ unit ahead of time t:

$$S_{X,t}(\tau) = \mathbb{P}(X > t + \tau | \mathcal{F}_t, X > t) = \exp\left[-\int_0^\tau \{\lambda_{T,t}(u) + \lambda_{O,t}(u)\}du\right]. \tag{3.3.3}$$

Then, we have that at u unit of time after t and conditioning on $X > t$ and $\mathcal{F}_t$,

$$\begin{aligned} \mathbb{P}\{dN_T(t+u) = 1 | \mathcal{F}_t, X > t\} &= \lambda_{T,t}(u)S_{X,t}(u)du, \\ \mathbb{P}\{dN_O(t+u) = 1 | \mathcal{F}_t, X > t\} &= \lambda_{O,t}(u)S_{X,t}(u)du. \end{aligned} \tag{3.3.4}$$

3.4 Model estimation with a composite likelihood

The forward intensity model implies the unknown functional-valued model parameters $\boldsymbol{\alpha}(\cdot)$ and $\boldsymbol{\beta}(\cdot)$. Now we discuss their estimations. For a realistic time-to-event setting, there is an end time of the study. We denote by E the end time so that the observable time $X = \min(T, O, E)$. Conceptually, when E is observed, it means that the subject survived at the end of the study.

We consider the study contains a total of n subjects with observed time X_i and predictor process $\boldsymbol{Z}_i(t)$ $(i = 1, \ldots, n)$. For the ith subject, ideally, at any given time $t < X_i$, the data for estimating $\boldsymbol{\alpha}(u)$ and $\boldsymbol{\beta}(u)$ are contained in

$\{dN_{T_i}(s), dN_{O_i}(s), \boldsymbol{Z}_i(t)\}_{t\le s\le X_i}$, where $N_{T_i}(s)$ and $N_{O_i}(s)$ are respectively the counting process of the competing events associated with the ith subject. By letting $V_{it} = X_i - t$, then the contribution from the ith subject to the t-specific forward intensity function is

$$L_i(t) = \{\lambda_{T_i,t}(V_{it})\}^{dN_{T_i}(X_i)}\{\lambda_{O_i,t}(V_{it})\}^{dN_{O_i}(X_i)} S_{X_i,t}(V_{it}). \tag{3.4.1}$$

Since the contribution (3.4.1) is well defined for any $t < X_i$, for the ith subject, we then construct the total contribution as $\prod_{t<X_i} L_i(t)$ so that the composite likelihood function is

$$L\{\boldsymbol{\alpha}(\cdot), \boldsymbol{\beta}(\cdot)\} = \prod_{i=1}^{n}\prod_{t<X_i}\{\lambda_{T_i,t}(V_{it})\}^{dN_{T_i}(X_i)}\{\lambda_{O_i,t}(V_{it})\}^{dN_{O_i}(X_i)} S_{X_i,t}(V_{it}), \tag{3.4.2}$$

where the product is taken over the support of the time. We note that the composite likelihood function (3.4.2) contains information for the continuous-time functionals $\boldsymbol{\alpha}(u)$ and $\boldsymbol{\beta}(u)$ ($u \ge 0$). We also remark that (3.4.2) is not the full likelihood, but a product of conditional likelihood. Thus, we refer to it as a composite likelihood function; for an overview of the composite likelihood, see Varin and Firth (2011). By maximizing (3.4.2), one then obtains the estimators for the model parameters denoted by $\hat{\boldsymbol{\alpha}}^{(1)}(u)$ and $\hat{\boldsymbol{\beta}}^{(1)}(u)$.

Clearly, continuous-time observation is required for computing the composite likelihood (3.4.2), which is practically not realistic. Thus, practically evaluating (3.4.2) for estimating the model parameters requires approximations with discrete-time observations.

3.5 Handling data observed at discrete time points

Practically, one only knows that an event happens in a small time interval, say $(t + t_1, t + t_1 + \Delta)$ in the future with $t_1 > 0$, without knowing the exact time when it happens. To incorporate this feature, we propose a discrete-time counter part of (3.3.4):

$$\begin{aligned}&\mathbb{P}(t+t_1<X=T<t+t_1+\Delta|\mathcal{F}_t,X>t)\\&\quad=S_{X,t}(t_1)\exp\left\{-\int_{t_1}^{t_1+\Delta}\lambda_{O,t}(u)du\right\}\left[1-\exp\left\{-\int_{t_1}^{t_1+\Delta}\lambda_{T,t}(u)du\right\}\right].\end{aligned} \tag{3.5.1}$$

The first part of (3.5.1) reflects that the subject survives at least up to t_1 after t. The second part incorporates the fact that the competing event does not happen before the event of interest, and the third part is the probability that the event happens in the interval $(t+t_1, t+t_1+\Delta)$. Clearly, as $\Delta \to 0$, (3.5.1) becomes (3.3.4) in the limiting case. More generally, we treat (3.5.1) as a discrete-time approximation of (3.3.4).

When Δ is small, or the forward intensity function is constant over the interval $(t_1, t_1+\Delta)$ the conditional probability in (3.5.1) is well approximated by

$$Q_{X,t}^{(1)}(t_1) = S_{X,t}(t_1)\exp\{-\lambda_{O,t}(t_1)\Delta\}[1-\exp\{-\lambda_{T,t}(t_1)\Delta\}].$$

Similarly, we have the following approximation for the competing event probability in the Δ interval after $t+t_1$:

$$Q_{X,t}^{(2)}(t_1) = S_{X,t}(t_1)[1-\exp\{-\lambda_{O,t}(t_1)\Delta\}]\exp\{-\lambda_{T,t}(t_1)\Delta\},$$

which approximates $\mathbb{P}(t+t_1<X=O<t+t_1+\Delta|\mathcal{F}_t, X>t)$. Additionally, when no event happens,

$$Q_{X,t}^{(3)}(t_1) = S_{X,t}(t_1)\exp\{-\lambda_{T,t}(t_1)\Delta\}\exp\{-\lambda_{O,t}(t_1)\Delta\}]$$

approximates $\mathbb{P}(X>t+t_1+\Delta|\mathcal{F}_t, X>t) = S_{X,t}(t_1+\Delta)$. Finally, the survival function $S_{X,t}(\tau)$ defined by (3.3.3) also needs to be approximately evaluated with data observed at discrete time points. For ease in presentation, we assume hereinafter that data is observed at equally spaced time points with interval

Δ. From the definition (3.4.2), equally space data are not essential, but it helps simplifying the notations. Then $S_{X,t}(\tau)$ is approximated by

$$S_{X,t}(\tau) \approx \exp\left[-\Delta \sum_{u<\tau}\{\lambda_{T,t}(u) + \lambda_{O,t}(u)\}\right]. \tag{3.5.2}$$

Concretely, for the ith ($i = 1, \dots, n$) subject, let $\mathcal{T}_i = \{t_{ij}\}_{j=1}^{m_i}$ be the observation times with m_i of them in total. We also observe the connections that $t_{im_i} = X_i$, and $t_{ij} = t_{i(j+1)} - \Delta$, $(j = 1, \dots, m_i - 1)$. Let $V_{it,\Delta} = V_{it} - \Delta = X_i - t - \Delta$, $\delta_T(X_i) = N_{T_i}(X_i) - N_{T_i}(X_i - \Delta)$ indicate the event, and $\delta_O(X_i) = N_{O_i}(X_i) - N_{O_i}(X_i - \Delta)$ indicate the competing event. Then by incorporating the discrete nature of the observations, for the ith subject, for any $t \in \mathcal{T}_i$, the counter part of (3.4.1) is

$$L_i(t) = \{Q^{(1)}_{X_i,t}(V_{it,\Delta})\}^{\delta_T(X_i)}\{Q^{(2)}_{X_i,t}(V_{it,\Delta})\}^{\delta_O(X_i)}\{Q^{(3)}_{X_i,t}(V_{it,\Delta})\}^{\{1-\delta_T(X_i)\}\{1-\delta_O(X_i)\}} \tag{3.5.3}$$

Then the composite likelihood function with discrete-time observations is

$$L\{\boldsymbol{\alpha}(\cdot), \boldsymbol{\beta}(\cdot)\} = \prod_{i=1}^{n}\prod_{t\in\mathcal{T}_i} L_i(t). \tag{3.5.4}$$

Furthermore, we can predictively evaluate the cumulative probability of event over the period $(t_i, t_i + \tau]$ by

$$\begin{aligned}\mathbb{P}(t_i < X_i = T_i \le t_i + \tau | \mathcal{F}_{t_i}, X_i > t_i) &= \sum_{k=1}^{[\tau/\Delta]} Q^{(1)}_{X_i,t_i}(k\Delta) \qquad (3.5.5)\\ &= \sum_{k=1}^{[\tau/\Delta]}\Bigg(\exp\Big[-\Delta\sum_{j=1}^{k-1}\{\lambda_{T_i,t_i}(j\Delta) + \lambda_{O_i,t_i}(j\Delta)\}\Big]\\ &\quad \times \exp\{-\lambda_{O_i,t_i}(k\Delta)\Delta\}[1 - \exp\{-\lambda_{T_i,t_i}(k\Delta)\Delta\}]\Bigg).\end{aligned}$$

in which t_i could be the last observation time for predictor variables for individual i. The prediction performance can be evaluated using time-dependent receiver operating characteristic (ROC) curve with the sensitivity and specificity defined based on (3.5.5) for $c \in [0,1]$:

$$\begin{aligned}
&\text{Sensitivity}(c, t+\tau):\\
&\mathbb{P}\{\mathbb{P}(t_i < X_i = T_i \leq t_i + \tau | \mathcal{F}_{t_i}, X_i > t_i) \geq c | X_i > t_i, T_i \in (t_i, t_i+\tau]\}.\\
&\text{Specificity}(c, t+\tau):\\
&\mathbb{P}\{\mathbb{P}(t_i < X_i = T_i \leq t_i + \tau | \mathcal{F}_{t_i}, X_i > t_i) < c | X_i > t_i, T_i > t_i + \tau\}.
\end{aligned}$$

3.6 Estimating $\boldsymbol{\alpha}(\cdot)$ and $\boldsymbol{\beta}(\cdot)$

A key observation from (3.5.3) and (3.5.4) is that the composite likelihood can be separated into two parts: one part only contains information for $\boldsymbol{\alpha}(\cdot)$ and the other only contains information for $\boldsymbol{\beta}(\cdot)$; see also Duan et al. (2012). More specifically, thanks to the exponential additive form,

$$L\{\boldsymbol{\alpha}(\cdot), \boldsymbol{\beta}(\cdot)\} = L\{\boldsymbol{\alpha}(\cdot)\} L\{\boldsymbol{\beta}(\cdot)\} \tag{3.6.1}$$

where

$$\begin{aligned}
L\{\boldsymbol{\alpha}(\cdot)\} &= \prod_{i=1}^{n} \prod_{t \in \mathcal{T}_i} S_{T_i,t}(V_{it,\Delta}) [1 - \exp\{-\lambda_{T_i,t}(V_{it,\Delta})\Delta\}]^{\delta_{T_i}(X_i)}\\
&\quad \times [\exp\{-\lambda_{T_i,t}(V_{it,\Delta})\Delta\}]^{\{1-\delta_{T_i}(X_i)\}},\\
L\{\boldsymbol{\beta}(\cdot)\} &= \prod_{i=1}^{n} \prod_{t \in \mathcal{T}_i} S_{O_i,t}(V_{it,\Delta}) [1 - \exp\{-\lambda_{O_i,t}(V_{it,\Delta})\Delta\}]^{\delta_{O_i}(X_i)}\\
&\quad \times [\exp\{-\lambda_{O_i,t}(V_{it,\Delta})\Delta\}]^{\{1-\delta_{O_i}(X_i)\}}.
\end{aligned}$$

From (3.5.2)

$$\begin{aligned}
S_{T_i,t}(V_{it,\Delta}) = \exp\left\{-\Delta \sum_{u \in \mathcal{T}_i, u > t} \lambda_{T_i,t}(u-t)\right\}, S_{O_i,t}(V_{it,\Delta})\\
= \exp\left\{-\Delta \sum_{u \in \mathcal{T}_i, u > t} \lambda_{O_i,t}(u-t)\right\}
\end{aligned}$$

are respectively the conditional survival functions corresponding to the competing events. Conceptually, $\boldsymbol{\alpha}(\cdot)$ and $\boldsymbol{\beta}(\cdot)$ are functional-valued. Typically, estimating the functions on its entire support is challenging, and would require additional structural assumptions. In our setting with discrete time points and the aforementioned approximations, we note that estimating the entire function is not necessary. Instead, one only needs to estimate $\boldsymbol{\alpha}(u)$ and $\boldsymbol{\beta}(u)$ for $u = \Delta, 2\Delta, \ldots, [\tau/\Delta]$ where τ is the future time after t at which the event probability is of interest.

For $u = k\Delta$ $(k = 1, \ldots, [\tau/\Delta])$, and let $\mathcal{T}_{i,k} = \{t_{ij}\}_{j=1}^{m_i-k-1}$. Then, the relevant composite likelihood functions can be further simplified as

$$L\{\boldsymbol{\alpha}(k\Delta)\} = \prod_{i=1}^{n} \left[\prod_{t \in \mathcal{T}_{i,k}} \exp\left\{-\Delta\lambda_{T_i,t}(k\Delta)\right\} \right] \times \left[\frac{1-\exp\left\{-\Delta\lambda_{T_i,(X_i-k\Delta-\Delta)}(k\Delta)\right\}}{\exp\left\{-\Delta\lambda_{T_i,(X_i-k\Delta-\Delta)}(k\Delta)\right\}} \right]^{\delta_{T_i}(X_i)}. \tag{3.6.2}$$

$$L\{\boldsymbol{\beta}(k\Delta)\} = \prod_{i=1}^{n} \left[\prod_{t \in \mathcal{T}_{i,k}} \exp\left\{-\Delta\lambda_{O_i,t}(k\Delta)\right\} \right] \times \left[\frac{1-\exp\left\{-\Delta\lambda_{O_i,(X_i-k\Delta-\Delta)}(k\Delta)\right\}}{\exp\left\{-\Delta\lambda_{O_i,(X_i-k\Delta-\Delta)}(k\Delta)\right\}} \right]^{\delta_{O_i}(X_i)}. \tag{3.6.3}$$

Then by maximizing (3.6.2) and (3.6.3), we obtain the estimators denoted by $\hat{\boldsymbol{\alpha}}^{(2)}(u)$ and $\hat{\boldsymbol{\beta}}^{(2)}(u)$ for $u = k\Delta$ $(k = 1, \ldots, [\tau/\Delta])$.

3.7 Handling missing data and irregularly spaced time points

To evaluate the likelihood functions (3.6.2) and (3.6.3), longitudinal observations of the predictor variables $\boldsymbol{Z}(t)$ at $t = \Delta, 2\Delta, \ldots$ are required. In practice, however, predictor variables could be missing at some time points, or observed at irregularly spaced time. A heuristic approach is to replace the missing one with the last observed value, which is the so-called the last observation carried forward (LOCF) approach. As discussed earlier, the LOCF approach may lead to biased estimations. Indeed, such bias is evidential as seen in our simulation studies.

To address this problem, we propose a smoothed composite likelihood approach to appropriately handle the missing measurements. Recall that $\mathcal{T}_{i,k}$ is the equally spaced time points at which $\boldsymbol{Z}(t)$ is observed for estimating $\boldsymbol{\alpha}(k\Delta)$. We then denote by $\mathcal{S}_{i,k}$ the actual set of time points at which $\boldsymbol{Z}(t)$ is observed. When there is missing data, $\mathcal{S}_{i,k}$ is a subset of $\mathcal{T}_{i,k}$. More broadly, $\mathcal{S}_{i,k}$ can be a general set of un-equally spaced time points when observations are taken.

The smoothed composite likelihood approach is constructed based on the log-composite likelihood. From (3.6.2) and (3.6.3), the log-likelihood is given by

$$\ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\} = \sum_{i=1}^{n}\Bigg(\sum_{t\in\mathcal{T}_{i,k}}\{-\Delta\lambda_{T_i,t}(k\Delta)\} \\ + \delta_{T_i}(X_i)\log\left[\frac{1-\exp\left\{-\Delta\lambda_{T_i,(X_i-k\Delta-\Delta)}(k\Delta)\right\}}{\exp\left\{-\Delta\lambda_{T_i,(X_i-k\Delta-\Delta)}(k\Delta)\right\}}\right]\Bigg). \tag{3.7.1}$$

$$\ell^{(2)}\{\boldsymbol{\beta}(k\Delta)\} = \sum_{i=1}^{n}\Bigg(\sum_{t\in\mathcal{T}_{i,k}}\{-\Delta\lambda_{O_i,t}(k\Delta)\} \\ + \delta_{O_i}(X_i)\log\left[\frac{1-\exp\left\{-\Delta\lambda_{O_i,(X_i-k\Delta-\Delta)}(k\Delta)\right\}}{\exp\left\{-\Delta\lambda_{O_i,(X_i-k\Delta-\Delta)}(k\Delta)\right\}}\right]\Bigg). \tag{3.7.2}$$

Denote by $K_{h_n}(t) = h_n^{-1}K(t/h_n)$ a kernel function with bandwidth h_n, where $K(\cdot)$ is a probability density function. For a comprehensive overview of the kernel smoothing methods, we refer to Fan (1996). Then the smoothed log-likelihood functions can be written as

$$\ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\} = \sum_{i=1}^{n}\Bigg\{-\sum_{t\in\mathcal{T}_{i,k}}\sum_{s\in\mathcal{S}_{i,k}} K_{h_n}(s-t)I(s\le t)\Delta\lambda_{T_i,s}(k\Delta) \\ + \sum_{s\in\mathcal{S}_{i,k}} K_{h_n}\{s-(X_i-k\Delta-\Delta)\}I(s\le X_i-k\Delta-\Delta) \\ \times\log\left[\frac{1-\exp\{-\Delta\lambda_{T_i,s}(k\Delta)\}}{\exp\{-\Delta\lambda_{T_i,s}(k\Delta)\}}\right]\delta_{T_i}(X_i)\Bigg\}, \tag{3.7.3}$$

$$\ell^{(3)}\{\boldsymbol{\beta}(k\Delta)\} = \sum_{i=1}^{n}\Bigg\{-\sum_{t\in\mathcal{T}_{i,k}}\sum_{s\in\mathcal{S}_{i,k}} K_{h_n}(s-t)I(s\leq t)\Delta\lambda_{O_i,s}(k\Delta)$$
$$+\sum_{s\in\mathcal{S}_{i,k}} K_{h_n}\{s-(X_i-k\Delta-\Delta)\}I(s\leq X_i-k\Delta-\Delta)$$
$$\times\log\left[\frac{1-\exp\{-\Delta\lambda_{O_i,s}(k\Delta)\}}{\exp\{-\Delta\lambda_{O_i,s}(k\Delta)\}}\right]\delta_{O_i}(X_i)\Bigg\}. \tag{3.7.4}$$

The smoothed log-likelihood considers contribution from predictor variables prior to the event of interest and the weight of the contribution is based on the distance between the time of predictor covariates and the event time of interest.

Then by maximizing (3.7.3) and (3.7.4), we obtain the maximum smoothed composite likelihood estimator denoted by $\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta)$ and $\hat{\boldsymbol{\beta}}^{(3)}(k\Delta)$ $(k = 1, 2, \ldots)$.

The properties of the estimators $\hat{\boldsymbol{\alpha}}^{(i)}(\cdot)$ and $\hat{\boldsymbol{\beta}}^{(i)}(\cdot)$ $(i = 1, 2, 3)$ will be presented in Section 5.

3.8 Estimating $\boldsymbol{\alpha}(\cdot)$ and $\boldsymbol{\beta}(\cdot)$ as continuous functions of time using the penalized B-spline method

Using the LOCF or the kernel smooth approach in Section 3.7 under the proposed framework, parameters are estimated at each prediction horizon time point (TP) of u. To model the estimators as a continuous function of time, we introduce B-splines (Boor, 1978) in the forward intensity model.

Let $B_{j,q}(u)$ denote the value of the jth B-spline basis function of order q at $u = k\Delta$ $(k = 1, 2, \ldots)$. Let p the number of internal knots. The rth parameter $\alpha_r(k\Delta)$ in the parameter vector $\boldsymbol{\alpha}(k\Delta)$ for the forward intensity function of event can be modeled by $\hat{\alpha}_r(k\Delta) = \sum_{j=1}^{q+p}\hat{a}_{rj}B_{j,q}(k\Delta) = \hat{\boldsymbol{a}}_r^T\boldsymbol{B}_q(k\Delta)$. The vector $\boldsymbol{B}_q(k\Delta) = \{B_{1,q}(k\Delta), \ldots, B_{q+p,q}(k\Delta)\}^T$ and $\hat{\boldsymbol{a}}_r = \{\hat{a}_{r1}, \hat{a}_{r(q+p)}\}^T$. Denote $\hat{\boldsymbol{a}} = \{\hat{\boldsymbol{a}}_0^T, \ldots, \hat{\boldsymbol{a}}_r^T, \ldots\}^T$ the parameter vector used in the B-spline functions to model all parameters in the parameter vector $\boldsymbol{\alpha}(k\Delta)$. Similarly the rth parameter $\beta_r(k\Delta)$ in the parameter vector $\boldsymbol{\beta}(k\Delta)$ for the forward intensity function of the competing event can be modeled by $\hat{\beta}_r(k\Delta) = \sum_{j=1}^{q+p} \hat{b}_{rj} B_{j,q}(k\Delta) = \hat{\boldsymbol{b}}_r^T \boldsymbol{B}_q(k\Delta)$. Denote $\hat{\boldsymbol{b}} = \{\hat{\boldsymbol{b}}_0^T, \ldots, \hat{\boldsymbol{b}}_r^T, \ldots\}^T$ the parameter vector used in the B-spline functions to model all parameters in the parameter vector $\boldsymbol{\beta}(k\Delta)$.

The log-likelihood using the B-spines can be expressed as

$$\begin{aligned}\ell^{(4)}[\boldsymbol{\alpha}\{\boldsymbol{B}_q(k\Delta)\}] = \sum_{i=1}^{n} \Bigg(& \sum_{s \in \mathcal{S}_{i,k}} \{-\Delta\lambda_{T_i,s}(k\Delta)\} \\ & + \delta_{T_i}(X_i) \log \left[\frac{1 - \exp\left\{-\Delta\lambda_{T_i,(X_i - k\Delta - \Delta)}(k\Delta)\right\}}{\exp\left\{-\Delta\lambda_{T_i,(X_i - k\Delta - \Delta)}(k\Delta)\right\}} \right] \Bigg). \end{aligned} \tag{3.8.1}$$

$$\begin{aligned}\ell^{(4)}[\boldsymbol{\beta}\{\boldsymbol{B}_q(k\Delta)\}] = \sum_{i=1}^{n} \Bigg(& \sum_{s \in \mathcal{S}_{i,k}} \{-\Delta\lambda_{O_i,s}(k\Delta)\} \\ & + \delta_{O_i}(X_i) \log \left[\frac{1 - \exp\left\{-\Delta\lambda_{O_i,(X_i - k\Delta - \Delta)}(k\Delta)\right\}}{\exp\left\{-\Delta\lambda_{O_i,(X_i - k\Delta - \Delta)}(k\Delta)\right\}} \right] \Bigg). \end{aligned} \tag{3.8.2}$$

which only include the contribution from the actual set of time points $s \in \mathcal{S}_{i,k}$ at which the $\boldsymbol{Z}(s)$ is observed. Then by maximizing $\sum_{k=1}^{[\tau/\Delta]} \ell^{(4)}[\boldsymbol{\alpha}\{\boldsymbol{B}_q(k\Delta)\}]$ and $\sum_{k=1}^{[\tau/\Delta]} \ell^{(4)}[\boldsymbol{\beta}\{\boldsymbol{B}_q(k\Delta)\}]$ considering all prediction horizons examined together, we obtain the maximum composite likelihood estimator vectors of $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{b}}$ for the B-spline basis functions.

In order to optimize the parameter estimation, we consider the penalized B-spline approach by Eilers and Marx (1996) and propose the penalized log-likelihood function under the forward intensity functions as,

$$\sum_{k=1}^{[\tau/\Delta]} \ell^{(4)}[\boldsymbol{\alpha}\{\boldsymbol{B}_q(k\Delta)\}] - \lambda \sum_{r} \sum_{j=3}^{q+p} (\Delta^2 a_{rj})^2. \tag{3.8.3}$$

where $\Delta^2 a_{rj} = a_{rj} - 2a_{r(j-1)} + a_{r(j-2)}$. Similarly,

$$\sum_{k=1}^{[\tau/\Delta]} \ell^{(4)}[\boldsymbol{\beta}\{\boldsymbol{B}_q(k\Delta)\}] - \lambda \sum_{r} \sum_{j=3}^{q+p} (\Delta^2 b_{rj})^2. \tag{3.8.4}$$

in which $\Delta^2 b_{rj} = b_{rj} - 2b_{r(j-1)} + b_{r(j-2)}$. For example, when the order q of the B-spline function is 3 (equivalent of degree 2) and the number of internal knots p is 0, the penalty term in function (3.8.3) is $\sum_r (\Delta^2 a_{r3})^2 = \sum_r (a_{r3} - 2a_{r2} + a_{r1})^2$ multiplied by λ. When the order of the B-spline function is 4 (equivalent of degree 3), the penalty term in function (3.8.3) is $\sum_r \sum_{j=3}^{4} (\Delta^2 a_{rj})^2 = \sum_r \{(a_{r3} - 2a_{r2} + a_{r1})^2 + (a_{r4} - 2a_{r3} + a_{r2})^2\}$ multiplied by λ. The parameter λ controls the smoothness of the fit. The parameter vectors $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{b}}$ can be estimated by maximizing the penalized log-likelihood functions (3.8.3) and (3.8.4).

Once the parameter vectors $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{b}}$ are estimated, the B-spline maximum composite likelihood estimators $\hat{\boldsymbol{\alpha}}^{(4)}\{\boldsymbol{B}_q(k\Delta)\}$ and $\hat{\boldsymbol{\beta}}^{(4)}\{\boldsymbol{B}_q(k\Delta)\}$ can be obtained by $\hat{\alpha}_r^{(4)}(k\Delta) = \sum_{j=1}^{q+p} \hat{a}_{rj} B_j(k\Delta; q) = \hat{\boldsymbol{a}}_r^T \boldsymbol{B}_q(k\Delta)$ and $\hat{\beta}_r^{(4)}(k\Delta) = \sum_{j=1}^{q+p} \hat{b}_{rj} B_{j,q}(k\Delta) = \hat{\boldsymbol{b}}_r^T \boldsymbol{B}_q(k\Delta)$ for each prediction horizon $u = k\Delta$ ($k = 1, \ldots, [\tau/\Delta]$) and each parameter $r = 0, 1, 2, \ldots$. Compared to the discrete estimation by each prediction horizon, the parameter estimators by maximizing the penalized log-likelihood functions (3.8.3) and (3.8.4) are continuous over prediction horizons $u = k\Delta$ ($k = 1, \ldots, [\tau/\Delta]$).

In order to optimize the penalty coefficient λ under the penalized B-spline method, the 3-fold cross validation can be performed. First the data is randomly split into 3 groups with equal number of individuals in each group. The train set 1 would include data from 2 out of the 3 groups, whereas the test set 1 would include data from the 3rd group that is not included in the train set 1. The train and test sets 2 and 3 can be prepared similarly considering different combinations of the 3 groups. The B-spline maximum composite likelihood estimators are obtained on the train set 1 first, and the prediction performance is examined in the test set 1 by calculating AUC for each prediction horizon time point $u = k\Delta(k = 1, \ldots, [\tau/\Delta])$. The overall prediction performance in the test set 1 across all prediction horizon $u = k\Delta(k = 1, \ldots, [\tau/\Delta])$ is then evaluated by taking the average of the AUC across k values ($k = 1, \ldots, [\tau/\Delta]$). The same process are repeated to evaluate the average of the AUC across k values ($k = 1, 2, \ldots$) in test sets 2 and 3 using the parameter estimates from train sets 2 and 3 respectively. The penalty coefficient λ can be optimized by maximizing the average AUC across all k values among the 3 test sets.

CHAPTER 4

NUMERICAL EXAMPLES

4.1 CD4 data analysis

We now illustrate the application and performance of the proposed approach with a data set from an immunodeficiency virus infection (HIV) clinical trial; see, among others, Abrams et al. (1994) and Goldman et al. (1996) for the conduct and findings of this study in the literature. The longitudinal data CD4 lymphocyte counts were collected for 467 patients at the study entry time, and 2, 6, 12 and 18 months afterward if available. CD4 Lymphocytes count has been extensively used as a prognostic factor or surrogate marker for disease progression or death of HIV patients. We are interested in using the CD4 count profile over time to predict patients' survival outcome.

We adopt the model (3.2.1) for the forward intensity function of death at the prediction horizon τ from time t. Predictor variables include the baseline covariates: the treatment (Drug: didanosine (ddI) or zalcitabine (ddC)), the indicator (prevOI) for patients with previous opportunistic infection, and the time-dependent covariates CD4 counts and its observation time. Both CD4 counts (cells/mm^3) and square root of CD4 counts are known to be right-skewed, and hence a transformed CD4$^{1/4}$ is used. The forward intensity function is specified as

$$\lambda_{T,t}(u) = \exp[\alpha_0(u) + \alpha_1(u)\{\text{CD4}(t)\}^{1/4} + \alpha_2(u)\text{Drug} + \alpha_3(u)t + \alpha_4(u)\text{prevOI}]$$

Out of 279 subjects who survived in the study, only 33 subjects had the CD4 assessment at month 18. The forward intensity function for the 246 subjects who exited the study due to reason other than death is modeled separately as a competing risk:

$$\lambda_{O,t}(u) = \exp[\beta_0(u) + \beta_1(u)\{\text{CD4}(t)\}^{1/4} + \beta_2(u)\text{Drug} + \beta_3(u)t + \beta_4(u)\text{prevOI}].$$

The prediction performance is examined both in-sample and out-of-sample. The in-sample evaluation includes all subjects for parameter estimations, and the same sample is used in calculating the area under the ROC curve (AUC) for the prediction evaluation. While for the out-sample evaluation, we first randomly partition the CD4 data set into a training set and a testing set with the same number of patients, then the parameters are estimated from the training set, and the AUCs are then calculated based on the testing set.

We implement our smoothed composite likelihood approach with multiple bandwidths in the real data analysis; see also the bandwidth selection approach of Cao et al. (2015). In particular, we set $h_n = 3 \times (Q_3 - Q_1) \times n^{-\gamma}$, where $Q_1 = 0$ month and $Q_3 = 6$ months are the 25% and 75% quantile of CD4 assessment time, and γ is chosen as $0.2, 0.5$, and 0.7 respectively. The total number of events n is 188 in the full data set and 94 in the train set. This results in bandwidths of 6.32, 1.31, and 0.46 respectively for the full data set, and 7.26, 1.86, and 0.75 respectively for the training set.

We compare the smoothed composite likelihood approach with the LOCF method. Considering the full sample CD4 data, the point estimates and the standard errors of $\boldsymbol{\alpha}(\tau)$ and $\boldsymbol{\beta}(\tau)$ in the forward intensity function are reported in Tables A.1 and A.2 in the Supplementary Material for both the smoothed composite likelihood approach and the LOCF method. We also plot in Figures 6.1 and 6.2 the estimates and point-wise 95% confidence interval (CI) for each predictor variable used in the forward intensity functions. The results presented are associated with the bandwidth $h = 6.32$, and we note that the results are similar for different choices of the bandwidth examined.

From the point estimates and the standard errors, we can see that the impact from the transformed CD4 is significant and it contributes negatively to the forward intensity function of event for all prediction time horizons ($\tau = 1, \ldots, 12$), which suggests that a subject with higher CD4 measurement has lower forward event intensity; thus he or she is subject to a lower risk of a future death. Additionally, the standard errors associated with the estimates for the transformed CD4 are also getting larger for longer time horizons, reflecting an increasing level of uncertainties for predictions associated with longer times. On the other hand, we observe that no significant association is found between the transformed CD4 and the forward intensity of exit due to other reasons for most of the prediction horizons. In an existing study, Goldman et al. (1996) concluded that a CD4 response did not necessarily correlate with improved survival outcome and therefore not a useful surrogate outcome. However, their analysis was static and based on landmark analyses at 2 months using proportional hazard model. Due to the limitation of the landmark analysis, Goldman et al. (1996) only concluded that the a classified CD4 response at 2 months was not predictive of death. In contrast, as illustrated by our proposed method, the CD4 count is statistically significant in the forward looking survival predictions, which could be due to the unaccounted impact from the entire trajectories of the CD4 profile.

The coefficients for treatment (ddC) are negative which suggests that subjects receiving ddC has lower forward intensity of death. However the treatment effect is not statistically significant at 5% level in our forward intensity approach. The assignment of treatment arm did not contribute to the forward intensity of exit due to other reasons.

The coefficients of the CD4 assessment time are positive and significant in the forward intensity function for other exits. This suggests that the longer a subject was followed in the study, the higher the forward intensity of exit due to other reasons and hence more likely this subject would drop out. But no such significant association is observed between the CD4 observation time and the forward intensity function of event. Moreover, the forward intensity of event is estimated to be higher if a subject was previously diagnosed with AIDS, but no association is observed for the forward intensity of exit due to other reasons.

To show how informative the forward looking event probability is in predicting the risk of death, we examine closely the two subjects 312 and 313. Subject 312 died at 13.9 months and subject 313 survived till the end of the study and the last CD4 observation time was 18.97 months. The transformed CD4$^{1/4}$ values over time for these two subjects are presented in Figure 6.3. Both subjects had similar CD4 value at 0 month (4.15 vs. 4.16). CD4 counts were under control during the study for subject 313, but decreased over time for subject 312. Both subjects received ddC treatment and were not previously diagnosed with AIDS.

The risk of death can be monitored by two subject-specific probabilities, the forward event probability over the period $(t_i+\tau-\Delta, t_i+\tau]$, $P(t_i+\tau-\Delta < X_i = T_i \leq t_i+\tau | X_i > t_i) = Q^{(1)}_{X_i,t_i}(\tau)$ and the cumulative event probability over the period $(t_i, t_i+\tau]$, $P(t_i < X_i = T_i \leq t_i+\tau | X_i > t_i) = \sum_{k=1}^{\tau/\Delta} Q^{(1)}_{X_i,t_i}(k\Delta)$ for prediction horizons $\tau = 1, \ldots, 12$ months and $\Delta = 1$ month. The probabilities plotted in Figure 6.4 are calculated using these two subjects' CD4 data up to $t_i = 0$, 2 and 6 months respectively and the smoothed parameters presented in Tables A.1 and A.2 in the Supplementary Material. As these two subjects shared the same baseline characteristics (treatment, AID diagnosis and CD4 count at study entry), the event probabilities considering only 0 month CD4 data nearly overlap for these two subjects. At 2 month, in contrast to a slight increase for subject 313, the CD4 count dropped significantly for subject 312. This is manifested in significant increase in the forward event probability up to 8-month prediction horizon for subject 312 and slight decrease for subject 313 considering CD4 data up to 2 months. The cumulative event probability for subject 312 at prediction horizon 12 month was also twice as high as that of subject 313 (0.06 vs. 0.03). At the 6 month, the CD4 count continued to drop for subject 312. The curve of the forward event probability peaked 3 month after the CD4 assessment at 6 month, indicating a high risk of death in coming months. This example demonstrates that the proposed approach incorporates the stochastic mechanism of the predictor variables in the forward event probabilities and discriminate the subjects with high risk from those with low risk.

To assess the predictive performance of the forward intensity function approach using LOCF and the smoothed composite likelihood, we utilize the ROC measure based on the conditional probability of event over the period $(t_i, t_i + \tau]$ $P(t_i < X_i = T_i \leq t_i + \tau | X_i > t_i) = \sum_{k=1}^{\tau/\Delta} Q^{(1)}_{X_i,t_i}(k\Delta)$. For a given subject, t_i is the last available CD4 assessment time and τ is a prediction horizon of interest.

We also compare the prediction performance of the proposed approach with the joint modeling approach. As documented in Rizopoulos (2011), the definitions of the sensitivity and specificity under the joint modeling approach are slightly different. In particular, the prediction rule of Rizopoulos (2011) is that for an HIV infected patient, he or she is considered as an event if the CD4 count is lower than a specific threshold. This is implemented by the 'rocJM' function in the JM package of R (Rizopoulos, 2010). For a direct comparison of the predictive performance using consistent definitions, we also evaluate the ROC considering the conditional survival probability $\pi_i(t_i + \tau | t_i) = \mathbb{P}(T_i \geq t_i + \tau | T_i > t_i)$ under the joint modeling framework of Rizopoulos (2011).

We implemented the joint modeling approach described in Rizopoulos (2011). The survival model includes treatment, gender, the indicator variable for azidothymidine (AZT) failure, and the indicator variable for the previous opportunistic infection. In addition to these covariates included in the survival submodel, fixed-effects in the longitudinal submodel also include the CD4 assessment time and the interaction effects of time with all the other covariates. The model also includes random effects; see Rizopoulos (2011).

We compute the ROC curve under different methods and then evaluate the AUC in order to compare the prediction performances. The AUC over prediction horizon 2-12 months under different methods are presented in Table A.3 in the Supplementary Material and plotted in Figure 6.6 for both in-sample and out-of-sample CD4 data. Due to very limited number of events (only 8) occurred by 1 month after the last available CD4 assessment, prediction horizon τ=1 month is not evaluated. The ROC curves for different values of τ examined using smoothed composite likelihood are also plotted in Figure 6.5 in which $\gamma = 0.2$ for both in-sample and out-of-sample data.

Using the definitions of sensitivity and specificity per Rizopoulos (2011) under the joint modeling framework, the mean AUC across subjects at each prediction horizon τ in the in-sample data is in the range of 66-70% as presented in Table A.3 in the Supplementary Material. They are in the similar scale as the AUC calculated in Table 1 of Rizopoulos (2011). This is consistent with the conclusion in Rizopoulos (2011) that CD4 cell count as a marker for death did not exhibit great discrimination power for advanced HIV-infected patients in this clinical trial. The AUC using predicted survival probabilities $\pi_i(t_i + \tau|t_i)$ under the joint model framework also doesn't show a good prediction power with AUC falling below 50% in the out-of-sample data.

Overall, we observe that the proposed methods with forward intensity function modeling yield substantially higher AUCs than the ones by Rizopoulos (2011). The advantage of our methods is more prominent in longer prediction horizons with AUCs increased to $\sim$ 88% which demonstrate that strong discriminative and predictive capability of CD4 counts for time to death. Compared to the LOCF method, the smoothed composite likelihood approach improves AUC by $\sim$5-10% up to 6 months prediction horizons and $\sim$ 2-5% beyond 7 months. Different choices of bandwidth within the range examined using the smoothed method have minimum impact on AUC. Moreover, the robustness of the prediction is demonstrated by similar AUCs in both in-sample and out-of-sample data studies.

4.2 PBC data analysis

We now illustrate the application of the forward intensity method in a data set. The data is from two clinical trials evaluating the use of D-penicillamine for treating patients with primary biliary cirrhosis (PBC); see Dickson et al. (1985) and Murtaugh et al. (1994) for the conduct and findings of this study in the literature. Primary biliary cirrhosis (PBC) is a chronic liver disease for which the liver transplantation is considered as an effective treatment. The serum bilirubin level has been known to be a prognostic factor associated with the disease progression of PBC and survival. There were 312 patients in the study with the PBC disease who were followed between 1974 and 1988 by the Mayo clinic. They were randomized to receive either D-penicillamine or placebo (158 versus 154). Most of the 312 patients in this study made annual visits to the Mayo Clinic after their initial referral which generated 1,945 patient visits. This enables us to use the serum bilirubin profile over time to predict a composite event, liver transplant or death in these patients. Rizopoulos (2016) predicted this composite event considering the repeated measurements of the serum bilirubin level under the joint modeling framework. Figure 6.7 plotted the log-transformed serum Bilirubin over years in this study. The 169 patients who had liver transplant or death show an increase in the serum bilirubin over years. This is not observed in the majority of the 143 patients who were alive.

We adopt the model (3.2.1) for the forward intensity function of event at prediction horizon τ from time t. Predictor variables include the baseline covariates: the treatment (Drug: D-penicillamine or placebo (reference)), and log-transformed serum bilirubin and its observation time. The forward intensity function is specified as

$$\lambda_{T,t}(u) = \exp[\alpha_0(u) + \alpha_1(u)\log\{\text{serBilir}(t)\} + \alpha_2(u)\text{Drug} + \alpha_3(u)t]$$

Out of 143 patients who survived in the study, only 25 subjects had the serum bilirubin assessment beyond year 10. Hence there were 118 patients who exited the study due to reason other than transplant or death. The forward intensity function for the competing risk is modeled separately:

$$\lambda_{O,t}(u) = \exp[\beta_0(u) + \beta_1(u)\log\{\text{serBilir}(t)\} + \beta_2(u)\text{Drug} + \beta_3(u)t].$$

The prediction performance is examined through out-of-sample evaluation. We first randomly partition the PBC data set into an in-sample set and an out-of-sample set by a ratio of 2:1 for the number of patients. The in-sample set includes 208 patients and the out-of-sample set includes 104 patients. The parameters are estimated from the in-sample set, and the prediction performance is evaluated on the out-of-sample set using ROC curves based on the cumulative probability of event $\mathbb{P}(t_i < X_i = T_i \leq t_i + \tau | \mathcal{F}_{t_i}, X_i > t_i)$ in (3.5.5). Among 208 individuals in the in-sample set, 114 (55%) patients had the event of interest (transplant or death), 81 (39%) exited due to other reason, and 13 (6%) survived with the last available serum Bilirubin measurement beyond 10 years. Among 104 individuals in the out-of-sample set, 55 (53%) patients had the event , 37 (36%) exited due to other reason, and 12 (12%) survived with the last available serum Bilirubin measurement beyond 10 years.

We are interested in prediction performance in prediction horizon $\tau = 1, \ldots, 9$ years from the last available serum bilirubin assessment time t_i. Parameters are estimated for $k = 1, \ldots, 9$ $(\Delta = 1)$ by maximizing

1. $l^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}$ and $l^{(2)}\{\boldsymbol{\beta}(k\Delta)\}$ in (3.7.1) and (3.7.2) using LOCF for imputation for each k value

2. $l^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}$ and $l^{(3)}\{\boldsymbol{\beta}(k\Delta)\}$ the smooth composite likelihood approach in (3.7.3) and (3.7.4) for each k value

3. $\sum_{k=1}^{9} \ell^{(4)}[\boldsymbol{\alpha}\{\boldsymbol{B}_q(k\Delta)\}] - \lambda \sum_r \sum_{j=3}^{q} (\Delta^2 a_{rj})^2$ and $\sum_{k=1}^{9} \ell^{(4)}[\boldsymbol{\beta}\{\boldsymbol{B}_q(k\Delta)\}] - \lambda \sum_r \sum_{j=3}^{q} (\Delta^2 b_{rj})^2$ using the penalized B-spline method with $\Delta^2 a_{rj} = \{a_{rj} - a_{r(j-1)}\} - \{a_{r(j-1)} - a_{r(j-2)}\}$.

The rth parameter in the parameter vector in the 3rd approach is modeled by

$$\hat{\alpha}_r(k\Delta) = \sum_j \hat{a}_{rj} B_{j,q}(k\Delta)$$
$$\hat{\beta}_r(k\Delta) = \sum_j \hat{b}_{rj} B_{j,q}(k\Delta)$$

in which $r = 0, 1, 2, 3$. The B-spline functions with the order of 3, 4 and 5 (corresponding to the degree of 2, 3 and 4 respectively) are considered for the penalized B-spline method and plotted over time u (-2 to 12) in Figure 6.9. An intercept is included and there is 0 inner knot. The penalized composite log-likelihood functions only include the contribution from the time points with serum Bilirubin available $s \in \mathcal{S}_{i,k}$.

The first two approaches give parameter estimators $\boldsymbol{\alpha}(k\Delta)$ and $\boldsymbol{\beta}(k\Delta)$ for each prediction horizon time point ($u = k\Delta = 1, \ldots, 9$) separately. We refer them as the time point (TP) estimation thereafter. The bandwidth selection in the smoothed composite likelihood approach follows $h_n = 3 \times (Q_3 - Q_1) \times n^{-\gamma}$ in (Cao et al., 2015) which yields bandwidth ranging between 1.80 and 4.65.

The penalty coefficient λ is optimized in the in-sample set by 3-fold cross validation which are explained in Section 3.8. The penalty coefficient λ examined is in the range of $\lambda = 0, \ldots, 1$ in an increment of 0.1. The λ value chosen is the one with the highest average AUC across all 9 prediction horizon years among all 3 test sets. The λ value with the highest average AUC is 0.

Once the penalty coefficient λ is optimized by the 3-fold cross-validation, the parameters are then re-estimated considering all 208 patients in the in-sample set. The parameter estimates by 3 different method are plotted in Figure 6.10 along with the 95% CI for the smoothed composite likelihood estimator in the in-sample set.

The difference in the $\boldsymbol{\alpha}$ parameter estimation is minimal comparing three different methods. This difference is larger in the $\boldsymbol{\beta}$ estimation. This is due to the lower rate of missing longitudinal observations in patients who had event than those who left the study due to other reasons (8% versus 36% considering all time points up to the observable time in the study). This indicates that the patients who had event were more closely monitored annually up to the event occurrence.

The parameter estimate of log-transformed serum Bilirubin in the forward intensity function of event is positive and much higher than that of the competing event, and it contributes positively to the forward intensity function of event for all prediction time horizon ($k\Delta = 1, \ldots, 9$ years), which suggests that a subject with higher serum bilirubin levels has higher forward event intensity; thus he or she is subject to a higher risk of a future transplant or death. This is consistent with the trend observed in Figure 6.7.

The estimates by the smoothed composite likelihood approach and the penalized B-spline method (PBS) are smoother over time than those by the LOCF, especially in longer prediction horizons. This is because kernel smoothing approach and PBS method allow estimation to borrow strength from neighboring data points and across multiple prediction horizons respectively.

To show how informative the forward looking event probability is in predicting the risk of transplant or death, we examine closely the two subjects 72 and 183. Subject 183 had transplant at 6.1 years and subject 72 survived till the end of the study (13.3 years) with the last serum bilirubin assessed at 4 years. The log-transformed serum bilirubin values over time for these two subjects are presented in Figure 6.11. Both subjects had the same serum bilirubin value ($\log(0.5 \text{ mg/dl}) = -0.693$) at study entry. The serum bilirubin level was under control during the study for subject 72, but increased over years for subject 183. Both subjects received placebo.

The risk of death can be monitored by the subject-specific forward event probability, $P(t+\tau-\Delta < X_i = T_i \leq t+\tau | X_i > t) = Q^{(1)}_{X_i,t}(\tau)$ between 4 years and 5 years. The probabilities plotted in Figure 6.12 are calculated using these two subjects' serum bilirubin levels at the time $t = 0, 1 \ldots,$ up to 4 years for $\tau = 5, 4, \ldots, 1$ years respectively (such that $t + \tau = 5$ years). The parameter estimates are based on the B-spline method with the degree of 4 B-splines and plotted in Figure 6.10. As both of these two subjects received placebo and had the same serum bilirubin level at study entry, the event probabilities at 5 years considering the serum bilirubin level at study entry are the same for these two subjects. At year 2, the serum bilirubin level was elevated for subject 183, but still under control for subject 72. This is manifested in significant increase in the forward event probability at 5 years for subject 183 indicating a much higher risk of transplant or death at 5 years. This example demonstrates that the proposed approach incorporates the stochastic mechanism of the predictor variables in the forward event probabilities and discriminate the subjects with high risk from those with low risk.

To assess the predictive performance of the forward intensity function approach by all 3 methods, we utilize the ROC measure based on the conditional probability of event over the period $(t_i, t_i + \tau]$ $P(t_i < X_i = T_i \leq t_i + \tau | X_i > t_i) = \sum_{k=1}^{\tau/\Delta} Q^{(1)}_{X_i,t_i}(k\Delta)$. For a given subject, t_i is the last available serum bilirubin assessment time and τ is a prediction horizon of interest. The prediction horizon years examined are from 1 year up to 9 years. Using the parameter estimates plotted in Figure 6.10 which are estimated from the in-sample set, the prediction performance is evaluated using AUCs in the out-of-sample set. Figure 6.13 plots the AUCs up to 9 years prediction horizon from the last available serum bilirubin assessment time in the out-of-sample PBC data for all three methods. AUC is the highest at $\tau = 5$ years prediction horizon, and above 90% for all prediction horizon examined ($\tau = 1, \ldots, 9$ years) from the last available assessment time t_i. The high AUCs in the out-of-sample set indicating strong discriminative and predictive capability of the serum bilirubin for time to transplant or death using the forward intensity method. Figure 6.14 plots the ROC curve at τ =5 years prediction horizon in the out-of-sample PBC data. The smoothed composite likelihood method and penalized B-spline method yield marginally higher AUCs than the time point estimation with LOCF.

4.3 Simulation studies

We conduct extensive simulations to examine the performance of the forward intensity modeling approach. Study E-1 examines the time point estimation by maximizing log-likelihood functions $l^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}$ and $l^{(2)}\{\boldsymbol{\beta}(k\Delta)\}$ in (3.7.1) and (3.7.2) for each prediction horizon time point ($u = k\Delta = 1, \ldots$) separately. The accuracy of the parameter estimations are assessed in terms of their biases and variances. In Study E-2 with missing longitudinal data, the accuracy of the parameter estimators are compared between the time point estimation with LOCF and the smoothed composite likelihood method which maximizes log-likelihood functions $l^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}$ and $l^{(3)}\{\boldsymbol{\beta}(k\Delta)\}$ in (3.7.3) and (3.7.4). In Study E-3, the accuracy of the parameter estimations and prediction performance are compared between the time point estimation with LOCF and the penalized B-spline approach by maximizing the penalized log-likelihood functions (3.8.3) and (3.8.4). In all settings, the simulations are repeated for 100 times. Study E-1 and E-2 consider two sample sizes of 100 and 300 respectively. Study E-3 considers two sample sizes of 45 and 90 respectively. Additional simulation results are included in the Supplementary Material.

Study E-1

In this study, we generate the event time considering a time-variant predictor in the hazard function. The hazard function for $t \geq 0$ takes the form $h_i(t) = \exp\{\log(\lambda) + z_i(t)\}$ where $\log(\lambda) = -8$. The predictor is generated as a function of time t such that $z_i(t) = a^t z_i(0)$ with $a = 1.15$ for time points $t = 0, 1, 2 \ldots$ up to 25. The baseline value $z_i(0)$ is subject-specific and generated by normal distribution $z_i(0) \sim N(0.8, 0.2^2)$. This implies that the conditional survival function is $S_{T,t}(u) = \mathbb{P}(T > t + u | T > t) = \exp[-\int_t^{t+u} \lambda \exp\{a^x z_i(0)\} dx]$. Hence the conditional forward intensity function can be modeled as $\lambda_{T,t}(u) = \exp\{\alpha_0(u) + \alpha_1(u) z_i(t)\}$ where

$\alpha_0(u) = \log(\lambda) = -8$ and $\alpha_1(u) = a^u = 1.15^u$. The intercept parameter $\alpha_0(u)$ is a constant over different prediction horizons u and the slope parameter $\alpha_1(u)$ changes over different u values. The event time T is generated by using the implied survival function $S(T) = \exp\{-\int_0^T h_i(u)du\} = \exp[-\int_0^T \exp\{\log(\lambda)+a^u z_i(0)\}du]$ in which the survival function $S(T)$ is generated from Unif$(0,1)$. A simulation scenario without considering a time-variant predictor in the hazard function is evaluated in Study E-0 provided in the Supplementary Material. Censor time C follows the exponential distribution with mean of 40. For a given individual, the observation time $X = \min(T, C)$ with event indicator $\delta_T = 1$ if $X = T$ and $\delta_T = 0$ if otherwise.

Parameter estimators are obtained using the time point estimation by maximizing log-likelihood functions $l^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}$ and $l^{(2)}\{\boldsymbol{\beta}(k\Delta)\}$ in (3.7.1) and (3.7.2) for each prediction horizon time point ($u = k\Delta = 1, \ldots, 10$) separately.

Table 6.1 summarizes the accuracy of parameter estimates in terms of their bias, relative bias (RB), standard deviation (SD), and square root of mean square error ($\sqrt{MSE}$). The RB is small and stable over different prediction horizons $u = 1, 2, \ldots, 10$. The standard deviation increases slightly with increasing prediction horizon u, but overall still stable. The SD improves with higher number of subjects in the simulations.

Study E-2

The predictor variable is available for every time point in Study E-1. In this study, we consider a scenario with missing predictor data and compare the performance between the LOCF and the smoothed composite likelihood approach. The predictor z_i, the event time T and the observation time X are the same as those generated in Study E-1.

The probability of being missing for the predictor at time t follows a Bernoulli distribution with the probability being a logistic function of time $P(t) = \frac{1}{1+\exp(-\frac{t-15}{8})}$. $P(t)$ ranges from 0.133 to 0.777 from t = 0 to 25. For example, for the 1st simulated data set, 31.6% of the predictor values are considered as missing.

The composite likelihood component corresponding to the missing data $z_i(t)$ is imputed by either the LOCF method or the smoothed composite likelihood approach. The parameter estimators are obtained using the time point estimation by maximizing either the log-likelihood functions $l^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}$ and $l^{(2)}\{\boldsymbol{\beta}(k\Delta)\}$ in (3.7.1) and (3.7.2) with LOCF or the smoothed composite log-likelihood functions $l^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}$ and $l^{(3)}\{\boldsymbol{\beta}(k\Delta)\}$ in (3.7.3) and (3.7.4) for each prediction horizon time point ($u = k\Delta = 1, \ldots, 10$) separately.

We use the bandwidth $h_n = 3 \times (Q_3 - Q_1) \times n^{-\gamma}$; see also the study of Cao et al. (2015). We use the 1st simulated data to calculate Q_1, Q_3 and n as the difference among simulated data sets on the bandwidth calculation is minimal. In particular, we set $Q_1 = 2$ month and $Q_3 = 9$ months which are the 25% and 75% quantile of the predictor time, and γ is in the range of 0.7 - 0.9. The total number of events n is 51 and 163 for sample size of 100 and 300 respectively. The bandwidth in the smoothed composite likelihood method is optimized by minimizing the MSE with an increment of 0.05 in each setting.

Table 6.2 and Table 6.3 summarize the parameter estimates using both the LOCF method and the smoothed composite likelihood approach when the number of subjects is 100 and 300. It presents the best smoothed results with the minimum $\sqrt{MSE}$ and the corresponding bandwidth is 0.75 for sample size of 100 and 0.6 for sample size of 300. The variability and RB improve with larger number of subjects. The smoothed method yields smaller bias than the LOCF method, and the SD is comparable between two methods. The smoothed method has lower $\sqrt{MSE}$ than LOCF approach, and this advantage becomes more prominent with higher number of subjects.

Study E-3

In the previous simulation studies, the parameter estimators are obtained using the time point estimation by maximizing either the composite log-likelihood functions $\ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}$ and $\ell^{(2)}\{\boldsymbol{\beta}(k\Delta)\}$ in (3.7.1) and (3.7.2) or the smoothed composite log-likelihood functions $\ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}$ and $\ell^{(3)}\{\boldsymbol{\beta}(k\Delta)\}$ in (3.7.3) and (3.7.4) at each forward-looking time point $u = k\Delta$ $(k = 1, 2, \ldots)$. In this study, we also estimate the parameters by maximizing the penalized log-likelihood functions (3.8.3) and (3.8.4) which consider all $u = k\Delta$ $(k = 1, 2, \ldots)$ values together. We compare its parameter estimators and prediction performance with the time point estimation.

In this study, the event time T and the competing event time O are assumed to follow Gompertz and Exponential distributions respectively. A simulation scenario with both the event time T and the competing event time O following the Gompertz distribution is evaluated in Study G-3 provided in the Supplementary Material.

The event time T is generated from the Gompertz distribution with shape parameter $a = 0.5$ and rate parameter $b = 0.005$, so that the conditional survival function is $S_{T,t}(u) = \mathbb{P}(T > t + u|T > t) = \exp[-\int_t^{t+u} be^{ax}dx] = \exp[\frac{b}{a}\{e^{at} - e^{a(t+u)}\}]$. This implies that the forward event intensity function at t is $\lambda_{T,t}(u) = \exp\{\log(b) + au + at\}$, which can be expressed as $\lambda_{T,t}(u) = \exp\{\alpha_0(u) + \alpha_1(u)t\}$, where $\alpha_0(u) = \log(b) + au = \log(0.005) + 0.5u$ and $\alpha_1(u) = a$=0.5.

The competing event time O is generated from the exponential distribution with the hazard function taking the form $h_i(t) = \exp\{\log(\lambda) + z_i(t)\}$ where $\log(\lambda) = -2$ and $t \geq 0$. The predictor is generated as a function of time t such that $z_i(t) = 0.4^t z_i(0)$ for time points $t = 0, 1, 2 \ldots$ up to 25. The baseline value $z_i(0)$ is subject-specific and generated by normal distribution $z_i(0) \sim N(1.1, 0.2^2)$. This implies that the conditional survival function of the competing event is $S_{O,t}(u) = \mathbb{P}(O > t + u|O > t) = \exp[-\int_t^{t+u} \lambda \exp\{0.4^x z_i(0)\}dx]$. Hence the conditional forward intensity function of the competing event can be modeled as $\lambda_{O,t}(u) = \exp\{\beta_0(u) + \beta_1(u)z_i(t)\}$ where $\beta_0(u) = \log(\lambda) = -2$ and $\beta_1(u) = 0.4^u$.

The end time of the study E is assumed to be 10. For a given individual, the observation time $X = \min(T, O, 10)$. The event indicator $\delta_T = 1$ if $X = T$. The competing event indicator $\delta_O = 1$ if $X = O$.

We estimate the parameters by three methods that maximizes

1. $l^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}$ and $l^{(2)}\{\boldsymbol{\beta}(k\Delta)\}$ in (3.7.1) and (3.7.2) for each k value

2. $l^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}$ and $l^{(3)}\{\boldsymbol{\beta}(k\Delta)\}$ the smooth composite likelihood approach in (3.7.3) and (3.7.4) for each k value

3. $\sum_{k=1}^{4} \ell^{(4)}[\boldsymbol{\alpha}\{\boldsymbol{B}_q(k\Delta)\}] - \lambda \sum_r \sum_{j=3}^{q} (\Delta^2 a_{rj})^2$ and $\sum_{k=1}^{4} \ell^{(4)}[\boldsymbol{\beta}\{\boldsymbol{B}_q(k\Delta)\}] - \lambda \sum_r \sum_{j=3}^{q} (\Delta^2 b_{rj})^2$ using the penalized B-spline method with $\Delta^2 a_{rj} = \{a_{rj} - a_{r(j-1)}\} - \{a_{r(j-1)} - a_{r(j-2)}\}$.

Under the 3rd approach the penalized B-spline method, the intercept $\alpha_0(u)$ and $\beta_0(u)$ and the slope $\alpha_1(u)$ and $\beta_1(u)$ in the forward intensity function of event and competing event are modeled as a function of the B-spline basis function $B_{j,q}(k\Delta)$.

$$\hat{\alpha}_r(k\Delta) = \sum_j \hat{a}_{rj} B_{j,q}(k\Delta)$$
$$\hat{\beta}_r(k\Delta) = \sum_j \hat{b}_{rj} B_{j,q}(k\Delta)$$

in which $r = 0, 1$ and $u = k\Delta = 1, 2, 3, 4$. The B-spline functions with the order of 3 and 4 (corresponding to the degree of 2, and 3 respectively) are considered for the penalized B-spline method and plotted over time u (0 to 5) in Figure 6.15. An intercept is included and there is 0 inner knot.

The first two approaches give parameter estimators $\boldsymbol{\alpha}(k\Delta)$ and $\boldsymbol{\beta}(k\Delta)$ for each prediction horizon time point ($u = k\Delta = 1, \ldots, 4$) separately. We refer them as the time point (TP) estimation thereafter. The bandwidth selection in the smoothed composite likelihood approach follows $h_n = 3 \times (Q_3 - Q_1) \times n^{-\gamma}$ in (Cao et al., 2015).

The penalty coefficient of λ examined for the penalized B-spline method is in the range of 0 to 5 with an increment of 1. It is optimized using the 3-fold cross validation by choosing the highest average AUC across all 4 prediction horizon time points considering all three test sets together. Details were explained in Section 3.8. The mean and SD of the average AUC across all prediction horizon time points ($u = k\Delta = 1, 2, 3, 4$) in each of the three test sets of the 100 simulated data are summarized in Table 6.4. The average AUCs across three test sets are also summarized in Table 6.4. When the sample size is 45, the highest average AUC is 83.08% and 83.09% by the penalized B-spline method using degree of 2 and 3 B-splines respectively. The λ value corresponding to the best average AUC by the penalized B-spline method ranges from 2, $\ldots$, 5 for degree of 2 B-splines and is 3 for degree of 3 B-splines. When the sample size is 90, the highest average AUC is 82.75% by the penalized B-spline method using both degree of 2 and 3 B-splines. The λ value corresponding to the best average AUC ranges from 2, $\ldots$, 5 for degree of 2 B-splines and is 4 for degree of 3 B-splines. Table 6.5 summarizes the mean and SD of the AUC at each prediction time point $u = k\Delta = 1, 2, 3,$ and 4 in each of the 3 test sets of the 100 simulated data using the optimized λ value presented in Table 6.4.

In Table 6.6, the parameter estimators by 3 methods are summarized using three measures $\sum_{k,r} |\text{bias}_{\alpha_r(k\Delta),\beta_r(k\Delta)}|$, $\sum_{k,r} SD_{\alpha_r(k\Delta),\beta_r(k\Delta)}$ and $\sum_{k,r} \sqrt{MSE_{\alpha_r(k\Delta),\beta_r(k\Delta)}}$ which consider all 4 parameters $\alpha_r(k\Delta)$ and $\beta_r(k\Delta)$ $(r = 0, 1)$ and prediction horizons $u = k\Delta = 1, 2, \ldots, 4$. The smoothed composite likelihood approach and the penalized B-spline method (PBS) yield better parameter estimates than the time point estimation without smoothing, especially when sample size is small. The improvement mainly comes from smaller SD. With proper choice of λ value, the bias and SD are similar using different degree of the B-splines. The $\sqrt{MSE}$ improves with larger number of subjects in the simulations.

Table 6.7 summarizes the accuracy of parameter estimates of $\alpha_r(k\Delta)$ and $\beta_r(k\Delta)$ $(r = 0, 1)$ in terms of their bias, relative bias (RB), standard deviation (SD), and square root of mean square error ($\sqrt{MSE}$). These measures are summarized for each parameter at each prediction horizon $u = k\Delta = 1, 2, 3, 4$ for all 3 methods. When sample size is 45, Table 6.7 presents the results for λ of 5 and 3 for degree of 2 and 3 B-splines respectively which correspond to the highest average AUC and the lowest $\sum_{k,r} \sqrt{MSE_{\alpha_r(k\Delta),\beta_r(k\Delta)}}$. When the sample size is 90, Table 6.7 presents the results for λ of 5 and 4 for degree of 2 and 3 B-splines respectively. The improvement in bias, SD and $\sqrt{MSE}$ by the smoothed composite likelihood approach and the penalized B-spline method is more prominent in longer prediction horizons when there are fewer data points available for the composite likelihood function for the estimation by each forward-looking time point without smoothing. This is more so for the parameter $\beta_1(u) = 0.4^u$ (a power function of u) at time $u = 4$ when the sample size $n = 45$. Table 6.8 summarizes the AUC at each prediction horizon for sample size of 45 and 90 by each of three methods.

In summary, the advantage of the smoothed composite likelihood approach and the penalized B-spline method compared to the estimation by each forward-looking time point without smoothing is more prominent when the sample size is smaller. This is because the kernel smoothing approach and PBS method allow estimation to borrow strength from neighboring data points and across multiple prediction horizons respectively.

CHAPTER 5

THEORETICAL PROPERTIES

We now present the properties of $\hat{\boldsymbol{\alpha}}^{(i)}(\cdot)$ $(i = 1, 2, 3)$. The index $i = 1, 2, 3$ corresponds to three scenarios, 1) continuous observation time, 2) discrete observation time, and 3) the smoothed composite likelihood approach when there are missing observations. The counterparts for $\hat{\boldsymbol{\beta}}^{(i)}(\cdot)$ $(i = 1, 2, 3)$ can be established following the same development. The proofs are provided in the Supplementary Material. Our setting is assuming the forward intensity model with the form (3.2.1), and targeting at estimating $\boldsymbol{\alpha}(\cdot)$. To achive that, we assume dense longitudinal measurements with $\Delta \to 0$. Alternative development is feasible with different settings of the underlying forward intensity model and sampling scheme.

We denote by $\boldsymbol{\alpha}_0(u)$ the truth, and assume the following regularity conditions.

Condition 1

(1A) There exists a neighborhood $\mathcal{A}_0^{(1)}(u)$ of $\boldsymbol{\alpha}_0(u)$ such that for $\boldsymbol{\alpha}(u) \in \mathcal{A}_0^{(1)}(u)$ and $t \in \mathcal{T}$, (i) $\lambda_{T,(t-u)}$ is strictly bounded away from 0, and (ii) the first, second and third order derivatives of $\lambda_{T,(t-u)}$ w.r.t. $\boldsymbol{\alpha}(u)$ are continuous and finite for $\boldsymbol{\alpha}(u) \in \mathcal{A}_0^{(1)}(u)$.

(1B) As $n \to \infty$, there exist a finite positive definite function supported on $\mathcal{A}_0^{(1)}(u)$, such that the probability limit exists and is positive definite for the matrix valued function,

$$\boldsymbol{\Omega}_{1,n}\{\boldsymbol{\alpha}(u)\} = \frac{1}{n}\sum_{i=1}^{n}\int_{t>u}^{\tau} \boldsymbol{Z}_i(t-u)\boldsymbol{Z}_i^T(t-u)\lambda_{T_i,(t-u)}\{\boldsymbol{\alpha}(u)\}dt.$$

(1C) For all $\epsilon > 0$ as $n \to \infty$

$$\frac{1}{n}\sum_{i=1}^{n}\left[\int_{t>u}^{\tau} \boldsymbol{Z}_i(t-u)\boldsymbol{Z}_i^T(t-u)I\{\frac{1}{\sqrt{n}}|\boldsymbol{Z}_i(t-u)|>\epsilon\}\lambda_{T_i,t-u}\{\boldsymbol{\alpha}_0^{(1)}(u)\}dt\right] \xrightarrow{p} 0.$$

Theorem 5.1 *Under Condition 1,* $\hat{\boldsymbol{\alpha}}^{(1)}(u)$ *is consistent and asymptotically normal:*

$$\sqrt{n}\boldsymbol{\Omega}_{1,n}^{1/2}\{\hat{\boldsymbol{\alpha}}^{(1)}(u)\}\{\hat{\boldsymbol{\alpha}}^{(1)}(u) - \boldsymbol{\alpha}_0(u)\} \xrightarrow{d} N(0, \mathbf{I}).$$

We assume the following condition for studying the properties of $\hat{\boldsymbol{\alpha}}^{(2)}(u)$.

Condition 2

(2A) There exists a neighborhood $\mathcal{A}_0^{(2)}(k\Delta)$ of $\boldsymbol{\alpha}_0(k\Delta)$ such that for all n, $\boldsymbol{\alpha}(k\Delta) \in \mathcal{A}_0^{(2)}(k\Delta)$ and $t \in \mathcal{T}$, the first, second and third order derivatives of $\ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}$ w.r.t $\boldsymbol{\alpha}(k\Delta)$ are continuous and finite for $\boldsymbol{\alpha}(k\Delta) \in \mathcal{A}_0^{(2)}(k\Delta)$.

(2B) As $n \to \infty$, the limits exist and are positive definite for the matrix valued functions

$$\boldsymbol{\Sigma}_{2,n}\{\boldsymbol{\alpha}_0(k\Delta)\} = \mathrm{var}_{\boldsymbol{\alpha}_0(k\Delta)}[\frac{1}{\sqrt{n}}\frac{\partial \ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T}]$$
$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}} \boldsymbol{Z}_i(t)\boldsymbol{Z}_i^T(t)\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}\Delta + o(\Delta);$$
$$\mathbf{W}_{2,n}\{\boldsymbol{\alpha}(k\Delta)\} = -\frac{1}{n}\frac{\partial^2 \ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial\{\boldsymbol{\alpha}(k\Delta)\}^T\partial\{\boldsymbol{\alpha}(k\Delta)\}}$$
$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}} \lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{Z}_i(t)\boldsymbol{Z}_i^T(t)\Delta\left[1 - \delta_{T_i}(t + k\Delta + \Delta)\right] + o(\Delta).$$

(2C) $\Delta \to 0$, $n \to \infty$, $\sqrt{n}\Delta \to 0$.

Theorem 5.2 *Let* $\boldsymbol{\Omega}_{2,n}\{\boldsymbol{\alpha}(k\Delta)\} = \mathbf{W}_{2,n}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{\Sigma}_{2,n}^{-1}\{\boldsymbol{\alpha}(k\Delta)\}[\mathbf{W}_{2,n}\{\boldsymbol{\alpha}(k\Delta)\}]^T$. *Under condition 2,* $\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta)$ *is consistent and asymptotically normal:*

$$\sqrt{n}\boldsymbol{\Omega}_{2,n}^{1/2}\{\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta)\}\{\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta) - \boldsymbol{\alpha}_0(u)\} \xrightarrow{d} N(0, \mathbf{I}).$$

We assume the following condition for studying the property of $\hat{\boldsymbol{\alpha}}^{(3)}(u)$.

Condition 3

(3A) There exists a neighborhood $\mathcal{A}_0^{(3)}(k\Delta)$ of $\boldsymbol{\alpha}_0(k\Delta)$ such that for all n, $\boldsymbol{\alpha}(k\Delta) \in \mathcal{A}_0^{(3)}(k\Delta)$ and $t \in \mathcal{T}$, the first, second and third order derivatives of $\ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}$ w.r.t $\boldsymbol{\alpha}(k\Delta)$ are continuous and finite for $\boldsymbol{\alpha}(k\Delta) \in \mathcal{A}_0^{(3)}(k\Delta)$.

(3B) As $n \to \infty$, $\Delta \to 0$, $h_n \to 0$, and $\frac{\Delta}{h_n} \to 0$, the limits exist and are positive definite for below matrix valued functions,

$$\begin{aligned}
&\boldsymbol{\Sigma}_{3,n}\{\boldsymbol{\alpha}_0(k\Delta)\} = \mathrm{var}_{\boldsymbol{\alpha}_0(k\Delta)}[\sqrt{n}\frac{1}{n}\frac{\partial \ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial \boldsymbol{\alpha}^T(k\Delta)}]\Delta^2 \\
&= \frac{1}{4n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}} \boldsymbol{Z}_i(t)\boldsymbol{Z}_i(t)^T\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}\Delta + o(\Delta); \\
&\mathbf{W}_{3,n}\{\boldsymbol{\alpha}(k\Delta)\} = -\frac{\Delta}{n}\frac{\partial^2 \ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial\boldsymbol{\alpha}(k\Delta)^T\partial\boldsymbol{\alpha}(k\Delta)} \\
&= \frac{1}{2n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}} \lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{Z}_i(t)\boldsymbol{Z}_i^T(t)\Delta\Big[1-\delta_{T_i}(t+k\Delta+\Delta)\Big] + o(\Delta).
\end{aligned}$$

(3C) $K(z)$ is a symmetric density function that satisfies $\int K(z)dz = 1$, $\int K^2(z)dz < \infty$, $h_n \to 0$, and $\sqrt{n}\Delta \to 0$.

Theorem 5.3 *Let* $\boldsymbol{\Omega}_{3,n}\{\boldsymbol{\alpha}(k\Delta)\} = \mathbf{W}_{3,n}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{\Sigma}_{3,n}^{-1}\{\boldsymbol{\alpha}(k\Delta)\}[\mathbf{W}_{3,n}\{\boldsymbol{\alpha}(k\Delta)\}]^T$. *Under condition 3,* $\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta)$ *is consistent and asymptotically normal:*

$$\sqrt{n}\boldsymbol{\Omega}_{3,n}^{1/2}\{\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta)\}\{\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta) - \boldsymbol{\alpha}_0(u)\} \xrightarrow{d} N(0,\mathbf{I}).$$

CHAPTER 6

DISCUSSION

We develop in this thesis a new framework for predictive survival analysis with longitudinal measurements where missing data and observations at irregularly spaced time points may occur. Our framework is parametric, modeling the forward intensity function with exponential parametric functions. Our theoretical analysis establishes the validity of the forward intensity modeling approach and the smoothed composite likelihood method. We introduce the penalized B-spline into the proposed forward intensity method in order to model the parameters as continuous functions of time. Simulations and real-data analyses demonstrate the promising performance of the proposed predictive approach.

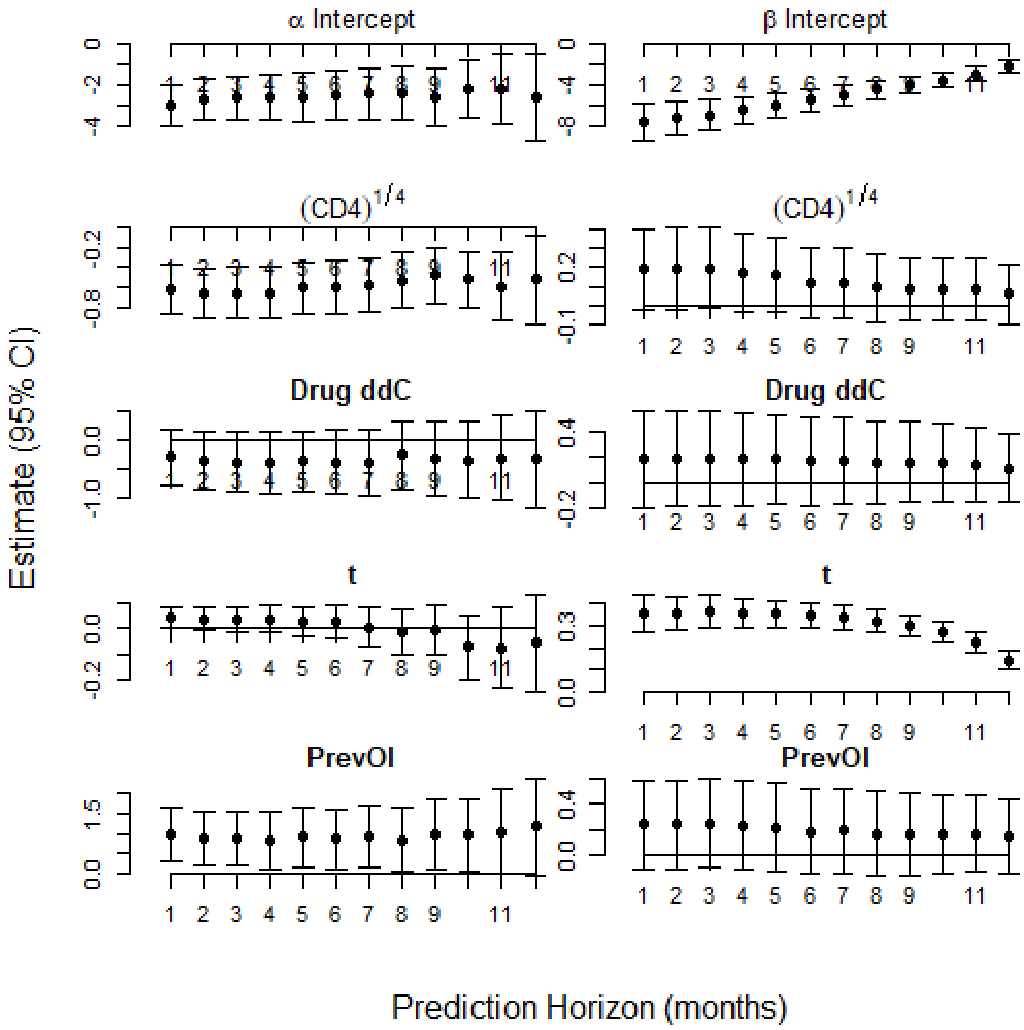

Figure 6.1: Parameter estimates (95% CI) of the forward intensity function using the LOCF approach in the CD4 data; the left panel presents parameter estimate $\hat{\boldsymbol{\alpha}}(\tau)$ over the prediction horizon τ=1-12 months and the right panel presents $\hat{\boldsymbol{\beta}}(\tau)$ over τ; predictor variables include drug ddC, the indicator (prevOI) for patients with previous opportunistic infection, transformed CD4 counts (CD4$^{1/4}$) and its observation time t; x-axis is plotted at y=0.

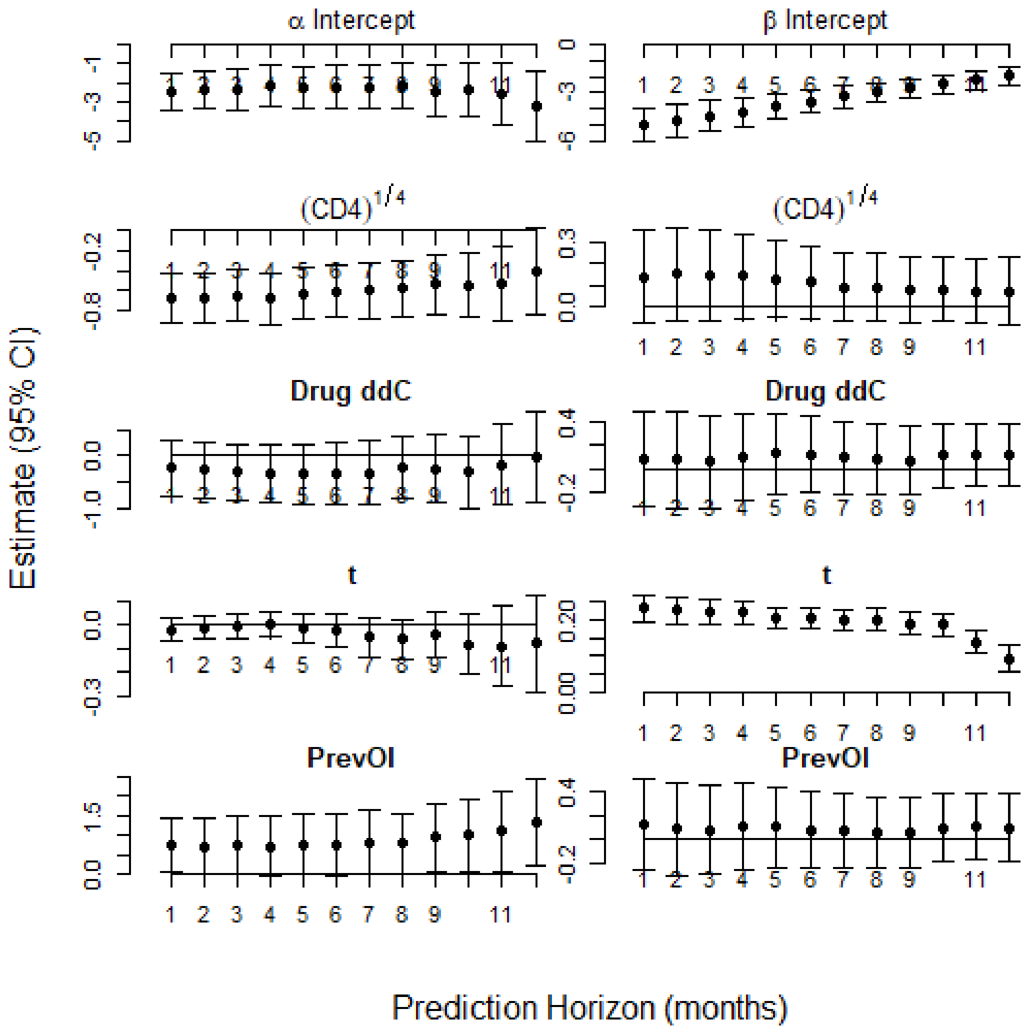

Figure 6.2: Parameter estimates (95% CI) of the forward intensity function using the smoothed composite likelihood approach in the CD4 data; the left panel presents parameter estimates $\hat{\boldsymbol{\alpha}}(\tau)$ over the prediction horizon τ=1-12 months and the right panel presents $\hat{\boldsymbol{\beta}}(\tau)$ over τ; predictor variables include drug ddC, the indicator (prevOI) for patients with previous opportunistic infection, transformed CD4 counts ($CD4^{1/4}$) and its observation time t; x-axis is plotted at y=0.

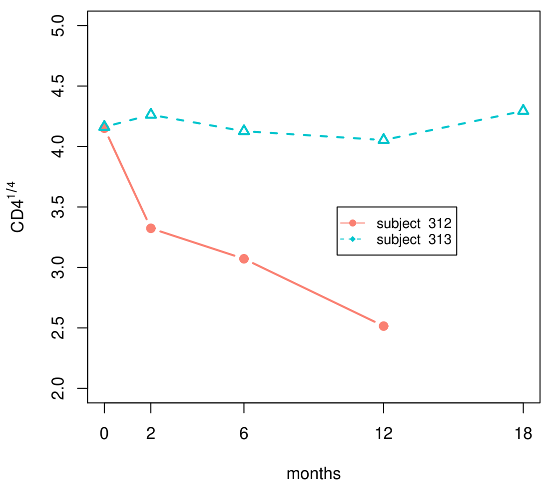

Figure 6.3: CD4 profiles over time for subjects 312 and 313.

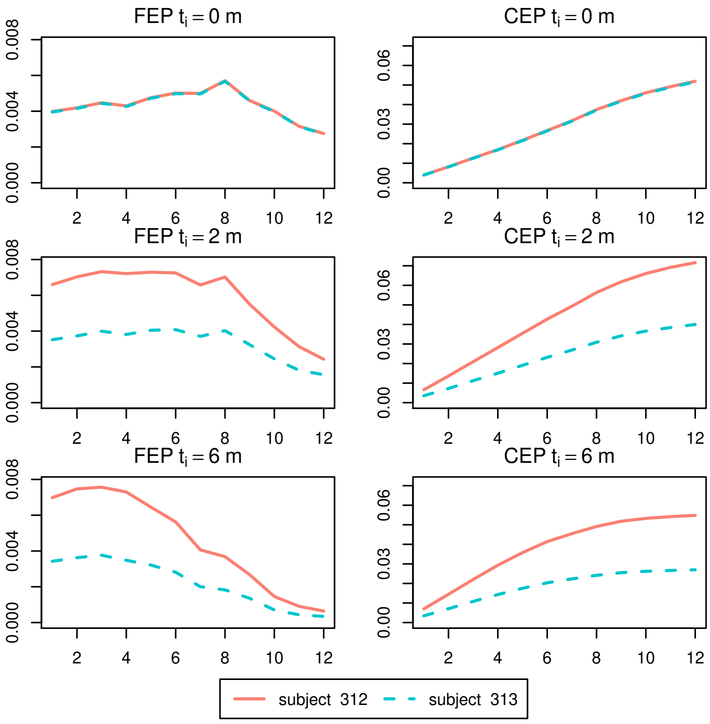

Figure 6.4: Predicted forward event probability (FEP) and cumulative event probability (CEP) over future prediction horizon 1-12 months for subjects 312 and 313 using CD4 data up to 2, 6 and 12 months; the origin 0 month in the plots is the last CD4 assessment time point t_i considered in the modeling (e.g. for the scenario above considering up to 2m data, the origin 0 month corresponds to $t_i = 2$ month from the study entry).

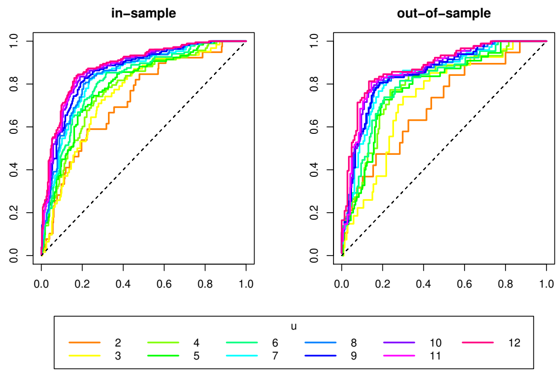

Figure 6.5: ROC curve using the smoothed composite likelihood method in both in-sample and out-of sample CD4 data for prediction horizons 2-12 months.

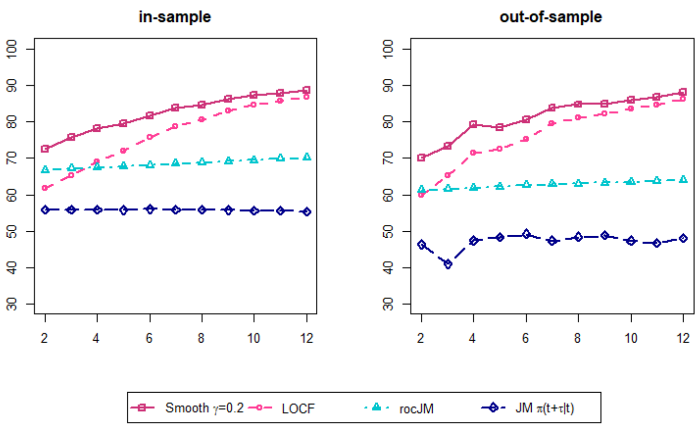

Figure 6.6: Plot of AUC (%) over prediction horizons 2-12 months in CD4 data comparing the joint modeling approach and the forward intensity method; both LOCF and smoothed composite likelihood approach are considered for the forward intensity method; both in-sample (left) and out-of-sample (right) CD4 data are used; rocJM is the mean AUC under the joint modeling method using the definitions of sensitivity and specificity per Rizopoulos (2011); $\pi(t+\tau|t)$ is the predicted survival probability under joint model framework; γ is the parameter value for bandwidth calculation in the proposed smoothed composite likelihood method.

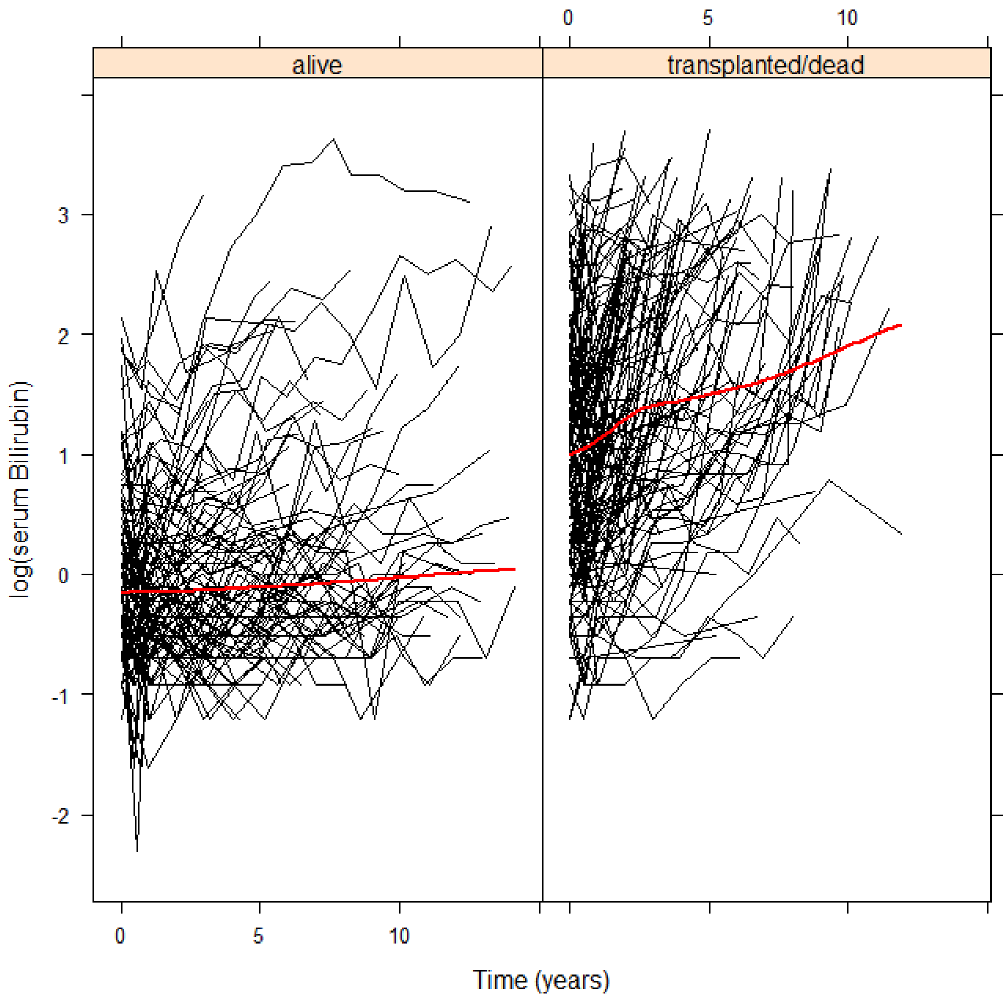

Figure 6.7: Plot of log-transformed serum Bilirubin over years in the PBC data; it's presented by patients who were alive versus those who had transplant or died in the PBC data.

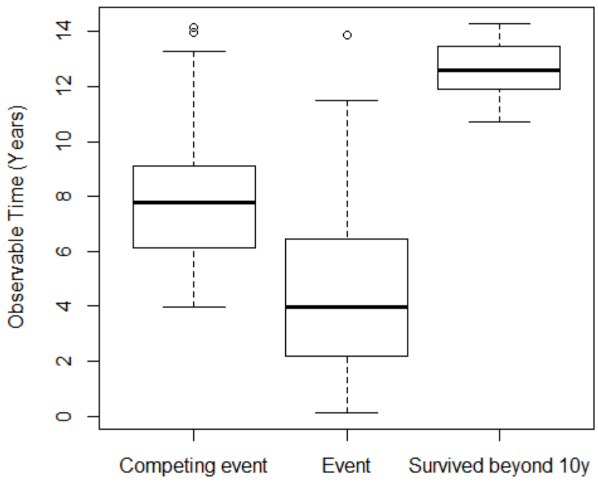

Figure 6.8: Box plot of the observable time (years) in the PBC data.

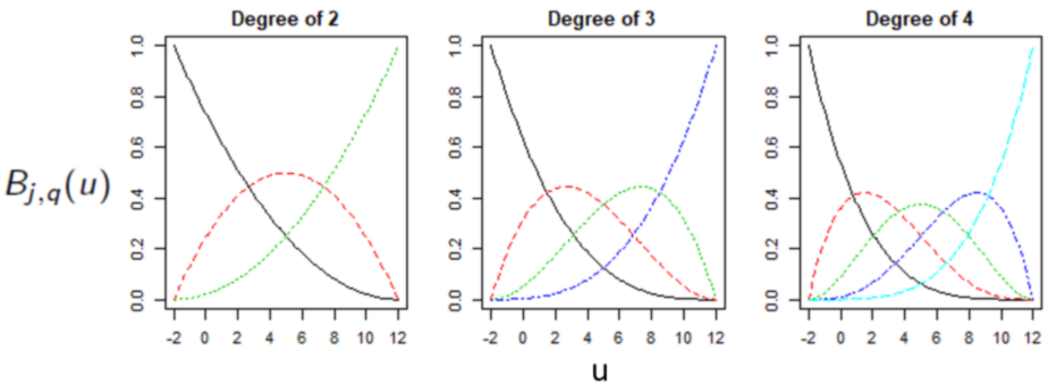

Figure 6.9: Plot of B-spline basis function used in the PBC data analysis with degree of 2, 3, and 4 respectively; intercept is included; the number of inner knots is 0.

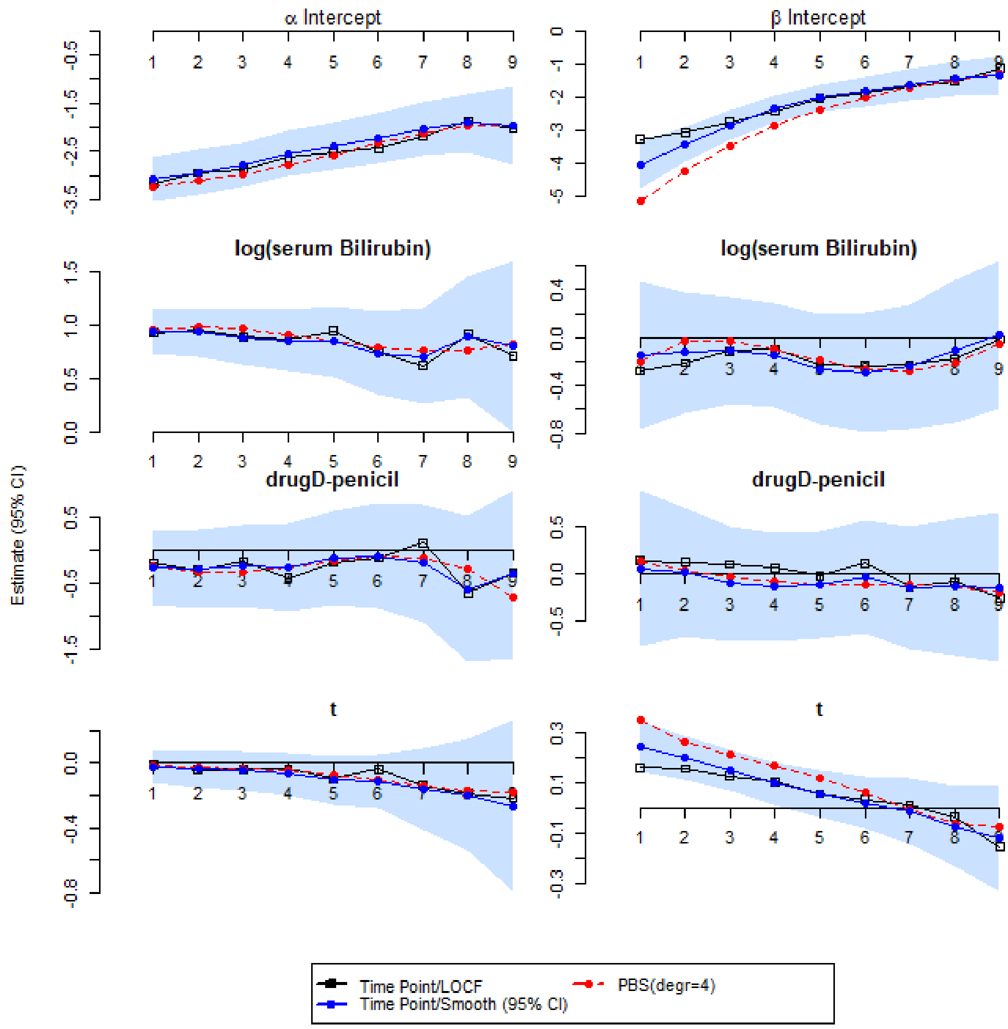

Figure 6.10: Parameter estimates of the forward intensity function in the in-sample PBC data by estimation at each time point with LOCF, the smoothed composite likelihood approach and the penalized B-spline method; the left panel presents parameter estimate $\hat{\boldsymbol{\alpha}}(\tau)$ over the prediction horizon τ=1-9 years and the right panel presents $\hat{\boldsymbol{\beta}}(\tau)$ over τ; predictor variables include drug, and log-transformed serum bilirubin level and its observation time t; x-axis is plotted at y=0.

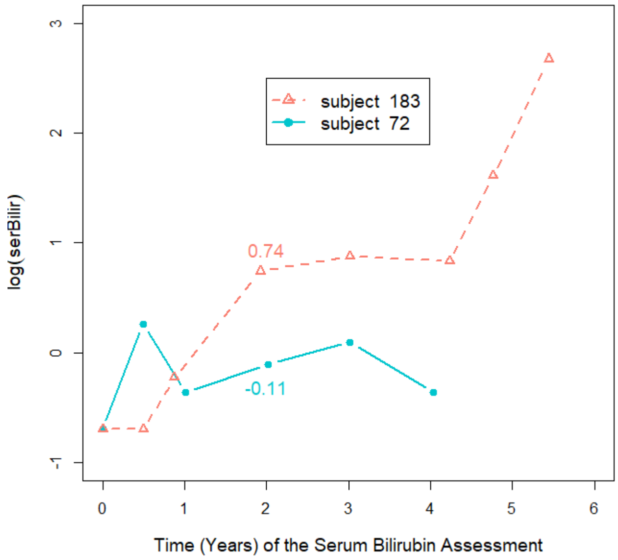

Figure 6.11: Serum Bilirubin profiles over time for subjects 72 and 183 in the out-of-sample PBC data.

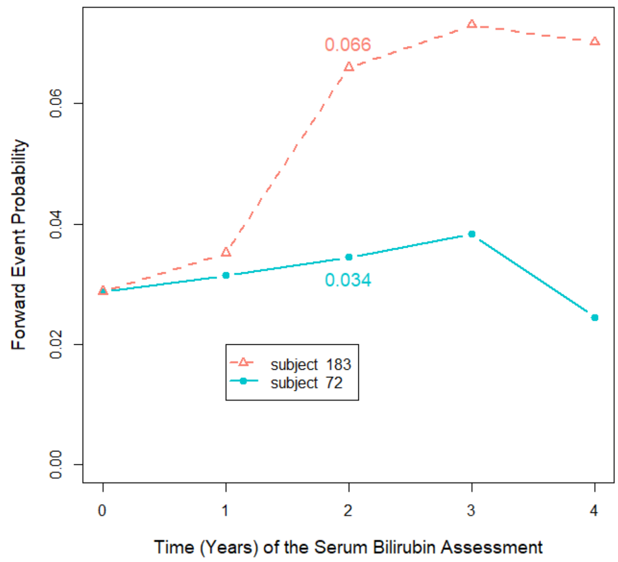

Figure 6.12: Predicted forward event probability (FEP) at 5 years for subjects 72 and 183 considering the serum bilirubin data at the study entry, year 1, . . . , up to year 4 in the out-of-sample PBC data.

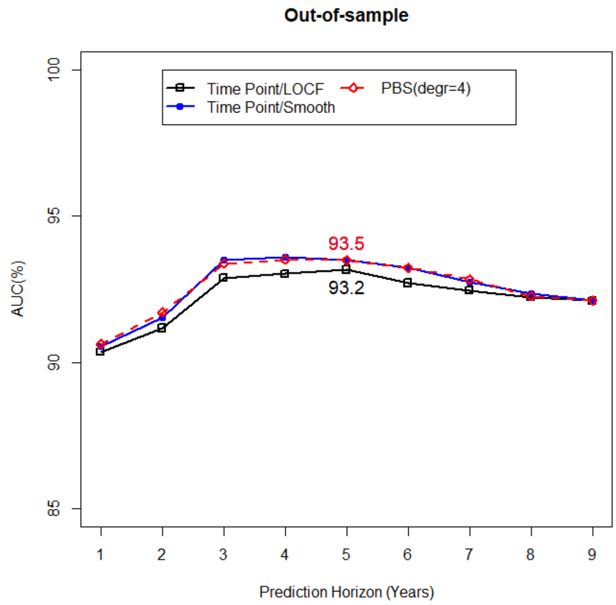

Figure 6.13: AUC over prediction horizons τ $(1, \ldots, 9$ years) in the out-of-sample PBC data.

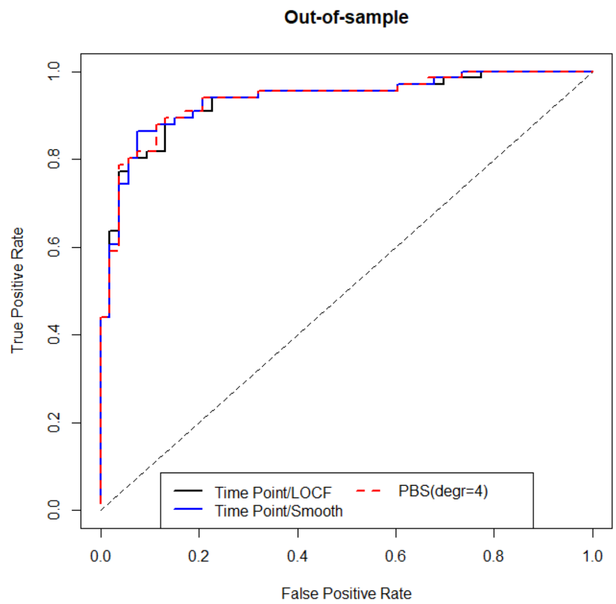

Figure 6.14: ROC curve at τ =5 years prediction horizon in the out-of-sample PBC data.

Table 6.1: Simulation results of the forward intensity method in Study E-1.

	u	True	n = 100 Bias	RB	SD	$\sqrt{MSE}$	n = 300 Bias	RB	SD	$\sqrt{MSE}$
$\alpha_0(u)$	1	-8	-0.019	0.002	0.618	0.619	-0.091	0.011	0.372	0.383
	2	-8	-0.001	<0.001	0.622	0.622	-0.082	0.01	0.375	0.384
	3	-8	0.001	<0.001	0.629	0.629	-0.081	0.01	0.382	0.39
	4	-8	-0.027	0.003	0.672	0.673	-0.08	0.01	0.404	0.412
	5	-8	-0.014	0.002	0.674	0.674	-0.085	0.011	0.41	0.418
	6	-8	-0.036	0.004	0.718	0.719	-0.084	0.01	0.421	0.429
	7	-8	-0.061	0.008	0.756	0.759	-0.085	0.011	0.437	0.445
	8	-8	-0.059	0.007	0.761	0.764	-0.092	0.011	0.447	0.457
	9	-8	-0.045	0.006	0.796	0.797	-0.091	0.011	0.468	0.477
	10	-8	-0.078	0.01	0.854	0.857	-0.056	0.007	0.496	0.5
$\alpha_1(u)$	1	1.15	0.004	0.004	0.103	0.103	0.02	0.018	0.062	0.066
	2	1.322	0.001	0.001	0.119	0.119	0.022	0.016	0.073	0.076
	3	1.521	0.001	0.001	0.137	0.137	0.025	0.016	0.085	0.088
	4	1.749	0.008	0.005	0.17	0.17	0.028	0.016	0.103	0.106
	5	2.011	0.005	0.003	0.194	0.195	0.034	0.017	0.119	0.124
	6	2.313	0.013	0.006	0.235	0.236	0.038	0.017	0.14	0.145
	7	2.66	0.024	0.009	0.282	0.283	0.044	0.017	0.167	0.173
	8	3.059	0.026	0.009	0.325	0.326	0.054	0.018	0.196	0.203
	9	3.518	0.023	0.007	0.388	0.389	0.062	0.018	0.237	0.245
	10	4.046	0.045	0.011	0.477	0.48	0.052	0.013	0.285	0.29

NOTE: The n is the number of subjects considered in the simulation; True is the true parameter value for $\alpha_0(u)$ and $\alpha_1(u)$ for different prediction horizons u; Bias is the difference between the parameter estimate and the true parameter value; RB is the bias divided by the true parameter value; SD is the standard deviation; $\sqrt{MSE}$ is the square root of the mean square error.

Table 6.2: Simulation results of the forward intensity method in Study E-2 comparing LOCF and the smoothed composite likelihood method when sample size n is 100.

			LOCF				Smooth			
	u	True	Bias	RB	SD	$\sqrt{MSE}$	Bias	RB	SD	$\sqrt{MSE}$
$\alpha_0(u)$	1	-8	1.007	-0.126	0.407	1.086	0.287	-0.036	0.641	0.702
	2	-8	0.925	-0.116	0.427	1.019	0.301	-0.038	0.684	0.748
	3	-8	0.827	-0.103	0.45	0.942	0.364	-0.046	0.693	0.782
	4	-8	0.771	-0.096	0.509	0.924	0.367	-0.046	0.655	0.75
	5	-8	0.75	-0.094	0.528	0.917	0.39	-0.049	0.622	0.734
	6	-8	0.679	-0.085	0.548	0.873	0.325	-0.041	0.65	0.727
	7	-8	0.66	-0.083	0.556	0.863	0.406	-0.051	0.636	0.755
	8	-8	0.562	-0.07	0.612	0.831	0.376	-0.047	0.691	0.787
	9	-8	0.593	-0.074	0.663	0.89	0.466	-0.058	0.719	0.857
	10	-8	0.466	-0.058	0.76	0.892	0.506	-0.063	0.701	0.865
$\alpha_1(u)$	1	1.15	-0.1	-0.087	0.076	0.125	-0.013	-0.011	0.112	0.113
	2	1.322	-0.104	-0.079	0.092	0.139	-0.015	-0.012	0.135	0.136
	3	1.521	-0.104	-0.069	0.108	0.15	-0.029	-0.019	0.157	0.16
	4	1.749	-0.113	-0.064	0.14	0.179	-0.036	-0.021	0.173	0.177
	5	2.011	-0.132	-0.066	0.162	0.209	-0.051	-0.025	0.186	0.193
	6	2.313	-0.136	-0.059	0.192	0.235	-0.036	-0.016	0.225	0.228
	7	2.66	-0.162	-0.061	0.221	0.274	-0.078	-0.029	0.249	0.261
	8	3.059	-0.152	-0.05	0.277	0.315	-0.072	-0.024	0.311	0.32
	9	3.518	-0.2	-0.057	0.343	0.397	-0.132	-0.037	0.367	0.39
	10	4.046	-0.164	-0.041	0.447	0.477	-0.162	-0.04	0.412	0.443

NOTE: The bandwidth used in the smoothed composite likelihood method is h=0.75.

Table 6.3: Simulation results of the forward intensity method in Study E-2 comparing LOCF and the smoothed composite likelihood method when sample size n is 300.

			LOCF				Smooth			
	u	True	Bias	RB	SD	$\sqrt{MSE}$	Bias	RB	SD	$\sqrt{MSE}$
$\alpha_0(u)$	1	-8	0.964	-0.12	0.246	0.995	0.109	-0.014	0.385	0.4
	2	-8	0.865	-0.108	0.259	0.903	0.153	-0.019	0.368	0.398
	3	-8	0.808	-0.101	0.278	0.854	0.179	-0.022	0.377	0.417
	4	-8	0.754	-0.094	0.304	0.813	0.235	-0.029	0.365	0.434
	5	-8	0.69	-0.086	0.313	0.758	0.19	-0.024	0.374	0.419
	6	-8	0.641	-0.08	0.353	0.732	0.171	-0.021	0.431	0.464
	7	-8	0.628	-0.079	0.38	0.734	0.247	-0.031	0.388	0.46
	8	-8	0.599	-0.075	0.393	0.717	0.226	-0.028	0.432	0.487
	9	-8	0.559	-0.07	0.376	0.674	0.248	-0.031	0.421	0.488
	10	-8	0.522	-0.065	0.442	0.684	0.299	-0.037	0.484	0.569
$\alpha_1(u)$	1	1.15	-0.089	-0.077	0.049	0.101	0.008	0.007	0.068	0.069
	2	1.322	-0.088	-0.067	0.057	0.105	0.004	0.003	0.074	0.074
	3	1.521	-0.095	-0.062	0.07	0.118	-0.001	-0.001	0.089	0.089
	4	1.749	-0.102	-0.058	0.086	0.133	-0.015	-0.008	0.097	0.098
	5	2.011	-0.104	-0.052	0.099	0.144	-0.003	-0.002	0.117	0.117
	6	2.313	-0.113	-0.049	0.127	0.17	<0.001	<0.001	0.151	0.151
	7	2.66	-0.139	-0.052	0.152	0.206	-0.033	-0.013	0.154	0.157
	8	3.059	-0.155	-0.051	0.183	0.24	-0.027	-0.009	0.203	0.205
	9	3.518	-0.164	-0.047	0.199	0.258	-0.041	-0.012	0.215	0.219
	10	4.046	-0.175	-0.043	0.262	0.315	-0.074	-0.018	0.284	0.293

NOTE: The bandwidth used in the smoothed composite likelihood method is h=0.6.

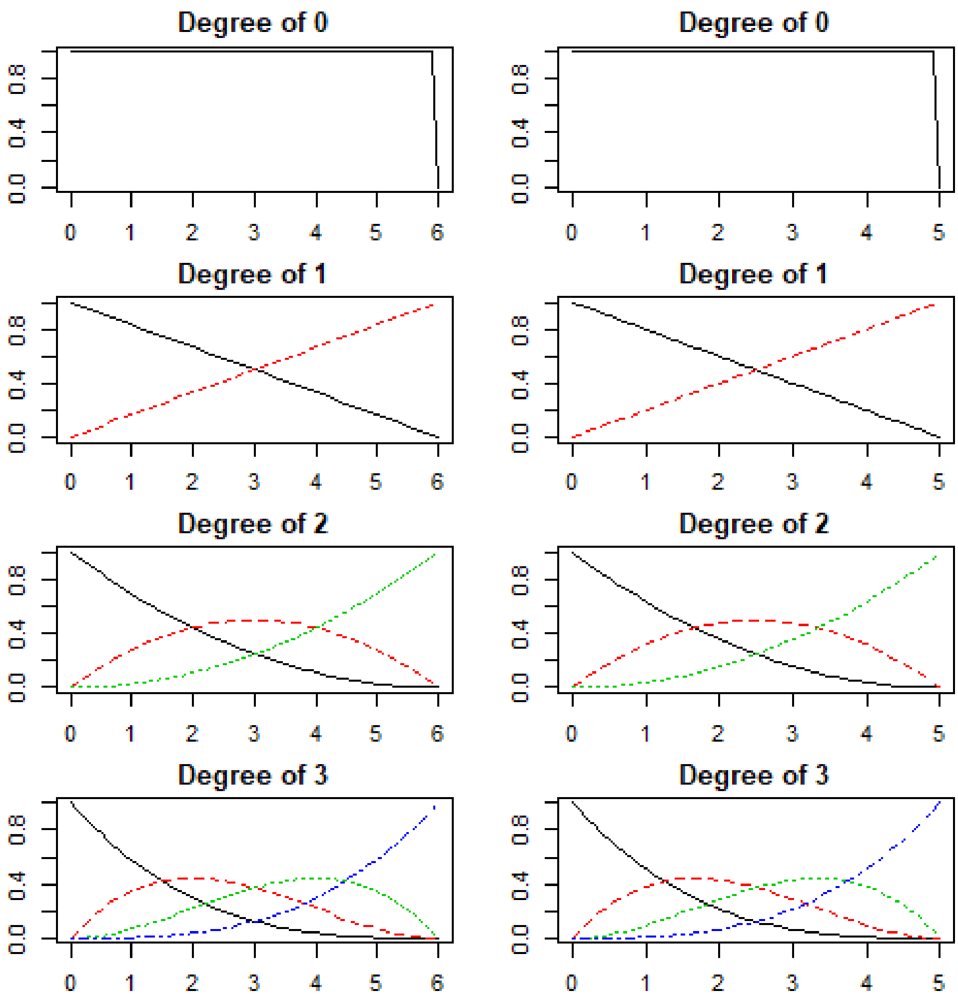

Figure 6.15: Plot of B-spline basis function with degree of 0, 1, 2, and 3 respectively; the Study G-3 use B-spline basis functions with the degree of 2 and 3 in the 1st column; Study E-3 uses B-spline basis functions with the degree of 2 and 3 in the 2nd column; intercept is included; the number of inner knots is 0.

Table 6.4: Average AUC (% Mean/SD) across prediction horizons examined in Study E-3 using the penalized B-spline (PBS/Degree) under the forward intensity method.

			Test set 1		Test set 2		Test set 3		Average	
n	Method	λ	Mean	SD	Mean	SD	Mean	SD	Mean	SD
45	PBS/2	0	84.01	13.01	83.83	12.51	81.18	14.36	83.01	13.29
		1	84.07	12.93	83.96	12.26	81.2	14.96	83.08	13.38
		2	84.08	12.92	83.97	12.27	81.2	14.96	83.08	13.38
		3	84.08	12.92	83.97	12.26	81.2	14.96	83.08	13.38
		4	84.08	12.92	83.97	12.26	81.2	14.96	83.08	13.38
		5	84.08	12.92	83.97	12.26	81.2	14.96	83.08	13.38
	PBS/3	0	83.98	13.08	83.74	12.43	81.45	14.09	83.06	13.2
		1	84.08	12.93	83.96	12.26	81.22	14.92	83.08	13.37
		2	84.06	12.92	83.96	12.26	81.23	14.92	83.08	13.37
		3	84.07	12.93	83.96	12.26	81.23	14.92	83.09	13.37
		4	84.07	12.93	83.96	12.26	81.2	14.96	83.08	13.38
		5	84.07	12.93	83.96	12.26	81.2	14.96	83.08	13.38
90	PBS/2	0	82.48	9.8	82.34	8.49	83.37	8.59	82.73	8.96
		1	82.49	9.76	82.34	8.43	83.4	8.52	82.74	8.9
		2	82.49	9.75	82.34	8.43	83.41	8.52	82.75	8.9
		3	82.49	9.75	82.34	8.43	83.41	8.52	82.75	8.9
		4	82.49	9.75	82.34	8.43	83.41	8.52	82.75	8.9
		5	82.49	9.75	82.34	8.43	83.41	8.52	82.75	8.9

Table 6.4: Average AUC (% Mean/SD) across prediction horizons examined in Study E-3 using the penalized B-spline (PBS/Degree) under the forward intensity method.

			Test set 1		Test set 2		Test set 3		Average	
n	Method	λ	Mean	SD	Mean	SD	Mean	SD	Mean	SD
	PBS(3)	0	82.5	9.8	82.34	8.49	83.37	8.6	82.74	8.96
		1	82.5	9.76	82.34	8.48	83.39	8.53	82.74	8.92
		2	82.49	9.75	82.35	8.46	83.39	8.54	82.74	8.91
		3	82.48	9.75	82.34	8.45	83.41	8.52	82.74	8.91
		4	82.48	9.75	82.35	8.43	83.41	8.52	82.75	8.9
		5	82.49	9.76	82.34	8.43	83.4	8.52	82.74	8.9

Table 6.5: AUC (% Mean/SD) at each prediction horizon (u) in the Test Sets in Study E-3 by the penalized B-spline (PBS/Degree) under the forward intensity method.

				Test Set 1 (u)				Test Set 2 (u)				Test Set 3 (u)			
n	Method	λ	Statistics	1	2	3	4	1	2	3	4	1	2	3	4
45	PBS/2	5	Mean	83.44	84.17	84.52	84.20	84.02	84.22	84.30	83.36	82.30	81.51	81.16	79.83
			SD	12.96	12.77	12.77	13.20	12.10	12.46	12.06	12.42	14.56	15.17	14.98	15.12
	PBS/3	3	Mean	83.39	84.17	84.52	84.20	83.99	84.22	84.27	83.36	82.30	81.64	81.16	79.83
			SD	12.97	12.77	12.77	13.20	12.08	12.46	12.07	12.42	14.56	15.04	14.98	15.12
90	PBS/2	5	Mean	82.00	82.57	82.86	82.53	81.67	82.46	82.59	82.62	83.14	83.57	83.58	83.37
			SD	9.62	9.50	9.89	9.99	8.18	8.31	8.48	8.76	8.77	8.56	8.29	8.44
	PBS/3	4	Mean	82.00	82.57	82.85	82.51	81.67	82.46	82.62	82.63	83.15	83.54	83.58	83.36
			SD	9.62	9.50	9.90	9.98	8.18	8.31	8.49	8.76	8.78	8.55	8.29	8.46

Table 6.6: Parameter estimation summary in Study E-3 comparing the estimation by time point (TP) with and without kernel smoothing, and penalized B-spline (PBS/Degree) under the forward intensity method.

n	Method/Degree	λ/h_n	$\sum_{k,r} \lvert \text{bias}_{\alpha_r(k\Delta),\beta_r(k\Delta)} \rvert$	$\sum_{k,r} SD_{\alpha_r(k\Delta),\beta_r(k\Delta)}$	$\sum_{k,r} \sqrt{MSE_{\alpha_r(k\Delta),\beta_r(k\Delta)}}$
45	Time point without smooth	NA	1.229	10.174	10.261
	Time point with smooth	2.26	0.669	6.727	6.806
	PBS/2	5	0.682	6.105	6.155
	PBS/3	3	0.685	6.111	6.161
90	Time point without smooth	NA	1.02	5.413	5.563
	Time point with smooth	1.16	1.209	5.156	5.365
	PBS/2	5	0.932	4.885	5.029
	PBS/3	4	0.931	4.890	5.034

Table 6.7: Parameter summary in Study E-3 comparing estimation by time point (TP) with and without smoothing and penalized B-spline (PBS/Degree).

				TP without smoothing				TP with smoothing				PBS/2			
n	Parameter	u	True	Bias	RB	SD	$\sqrt{MSE}$	Bias	RB	SD	$\sqrt{MSE}$	Bias	RB	SD	$\sqrt{MSE}$
45	$\alpha_0(u)$	1	-4.798	-0.137	0.028	0.701	0.714	-0.023	0.005	0.713	0.714	-0.101	0.021	0.701	0.708
		2	-4.298	-0.062	0.015	0.635	0.638	0.028	-0.006	0.669	0.669	-0.063	0.015	0.596	0.599
		3	-3.798	-0.004	0.001	0.579	0.579	0.052	-0.014	0.646	0.648	-0.025	0.006	0.539	0.54
		4	-3.298	-0.011	0.003	0.58	0.58	0.039	-0.012	0.669	0.67	0.014	-0.004	0.547	0.547
	$\alpha_1(u)$	1	0.5	-0.004	-0.009	0.132	0.132	-0.021	-0.042	0.136	0.137	-0.008	-0.017	0.13	0.131
		2	0.5	-0.019	-0.037	0.136	0.137	-0.033	-0.067	0.148	0.152	-0.02	-0.04	0.137	0.138
		3	0.5	-0.035	-0.069	0.156	0.16	-0.046	-0.091	0.172	0.178	-0.032	-0.064	0.151	0.155
		4	0.5	-0.042	-0.084	0.188	0.193	-0.055	-0.109	0.209	0.216	-0.044	-0.088	0.172	0.178
	$\beta_0(u)$	1	-2	0.053	-0.026	0.244	0.25	0.038	-0.019	0.279	0.281	0.046	-0.023	0.244	0.248
		2	-2	0.036	-0.018	0.305	0.307	0.027	-0.014	0.34	0.341	0.041	-0.021	0.283	0.286
		3	-2	0.037	-0.019	0.364	0.366	0.017	-0.008	0.41	0.411	0.036	-0.018	0.354	0.355
		4	-2	0.063	-0.031	0.483	0.487	-0.011	0.005	0.524	0.524	0.031	-0.016	0.441	0.442
	$\beta_1(u)$	1	0.4	-0.073	-0.184	0.421	0.427	-0.182	-0.456	0.332	0.379	-0.08	-0.2	0.412	0.419
		2	0.16	-0.041	-0.253	0.538	0.54	-0.064	-0.402	0.38	0.385	0.027	0.171	0.338	0.339
		3	0.064	-0.082	-1.285	0.623	0.628	-0.027	-0.424	0.456	0.456	-0.009	-0.144	0.434	0.434
		4	0.026	-0.53	-20.699	4.088	4.122	-0.007	-0.267	0.644	0.644	-0.103	-4.035	0.627	0.635
90	$\alpha_0(u)$	1	-4.798	0.013	-0.003	0.587	0.587	0.046	-0.01	0.605	0.607	0.024	-0.005	0.603	0.603
		2	-4.298	0.022	-0.005	0.564	0.564	0.051	-0.012	0.589	0.592	0.021	-0.005	0.536	0.536
		3	-3.798	0.005	-0.001	0.571	0.571	0.052	-0.014	0.571	0.574	0.019	-0.005	0.501	0.502
		4	-3.298	0.021	-0.006	0.496	0.497	0.068	-0.021	0.507	0.512	0.017	-0.005	0.506	0.506
	$\alpha_1(u)$	1	0.5	-0.037	-0.073	0.095	0.101	-0.041	-0.083	0.099	0.107	-0.037	-0.075	0.097	0.104
		2	0.5	-0.045	-0.09	0.107	0.116	-0.05	-0.099	0.114	0.124	-0.046	-0.092	0.105	0.115

Table 6.7: Parameter summary in Study E-3 comparing estimation by time point (TP) with and without smoothing and penalized B-spline (PBS/Degree) under the forward intensity method.

				TP without smoothing				TP with smoothing				PBS/2			
n	Parameter	u	True	Bias	RB	SD	$\sqrt{MSE}$	Bias	RB	SD	$\sqrt{MSE}$	Bias	RB	SD	$\sqrt{MSE}$
		3	0.5	-0.051	-0.103	0.132	0.142	-0.06	-0.121	0.134	0.146	-0.055	-0.11	0.121	0.133
		4	0.5	-0.068	-0.136	0.144	0.159	-0.079	-0.158	0.147	0.167	-0.064	-0.129	0.141	0.155
	$\beta_0(u)$	1	-2	0.057	-0.029	0.186	0.194	0.063	-0.032	0.19	0.2	0.058	-0.029	0.19	0.198
		2	-2	0.079	-0.04	0.211	0.226	0.08	-0.04	0.221	0.235	0.07	-0.035	0.2	0.212
		3	-2	0.086	-0.043	0.267	0.28	0.083	-0.042	0.269	0.282	0.083	-0.041	0.243	0.256
		4	-2	0.092	-0.046	0.323	0.336	0.077	-0.039	0.329	0.338	0.095	-0.048	0.305	0.32
	$\beta_1(u)$	1	0.4	-0.075	-0.188	0.295	0.305	-0.177	-0.443	0.243	0.301	-0.115	-0.286	0.292	0.314
		2	0.16	-0.134	-0.835	0.367	0.39	-0.139	-0.866	0.303	0.333	-0.014	-0.086	0.229	0.23
		3	0.064	-0.11	-1.719	0.475	0.487	-0.087	-1.354	0.378	0.388	-0.057	-0.886	0.324	0.329
		4	0.026	-0.124	-4.832	0.593	0.606	-0.057	-2.221	0.458	0.461	-0.157	-6.136	0.493	0.517

Table 6.8: AUC summary in Study E-3 comparing the estimation by time point (TP) with and without kernel smoothing, and penalized B-spline (PBS/Degree) under the forward intensity method.

n	Method/Degree	λ/h_n	Statistics	1	2	3	4
45	Time point without smooth	NA	Mean	82.81	83.03	83.07	82.61
			SD	7.23	7.46	7.58	7.46
	Time point with smooth	2.26	Mean	82.69	82.92	82.92	82.37
			SD	7.26	7.42	7.66	7.56
	PBS/2	5	Mean	82.73	83.10	83.14	82.52
			SD	7.18	7.40	7.50	7.54
	PBS/3	3	Mean	82.71	83.09	83.15	82.54
			SD	7.18	7.40	7.50	7.55
90	Time point without smooth	NA	Mean	81.59	82.45	82.87	82.69
			SD	5.46	5.54	5.34	5.36
	Time point with smooth	1.16	Mean	81.67	82.34	82.86	82.70
			SD	5.35	5.44	5.37	5.33
	PBS/2	5	Mean	81.69	82.39	82.89	82.71
			SD	5.45	5.55	5.40	5.40
	PBS/3	4	Mean	81.67	82.39	82.87	82.71
			SD	5.44	5.55	5.38	5.39

Supplementary Material to "A Predictive Time-to-Event Modeling Approach with Longitudinal Measurements and Missing Data"

This Supplementary Material contains more numerical results and proofs of the theorems.

REFERENCES

Aalen, O., O. Borgan, and H. Gjessing (2008). *Survival and event history analysis: a process point of view.* Springer Science Business Media.

Abrams, D. I., A. I. Goldman, C. Launer, J. A. Korvick, J. D. Neaton, L. R. Crane, M. Grodesky, S. Wakefield, K. Muth, S. Kornegay, et al. (1994). A comparative trial of didanosine or zalcitabine after treatment with zidovudine in patients with human immunodeficiency virus infection. *New England Journal of Medicine 330*(10), 657–662.

Anderson, J. R., K. C. Cain, and R. D. Gelber (1983). Analysis of survival by tumor response. *J Clin Oncol 1*(11), 710–719.

Boor, D. (1978). *A practical guide to splines.* New York: springer-verlag.

Brown, E. R., J. G. Ibrahim, and V. DeGruttola (2005). A flexible b-spline model for multiple longitudinal biomarkers and survival. *Biometrics 61*(1), 64–73.

Cai, Z. and Y. Sun (2003). Local linear estimation for time-dependent coefficients in cox's regression models. *Scandinavian Journal of Statistics 30*(1), 93–111.

Cao, H., M. M. Churpek, D. Zeng, and J. P. Fine (2015). Analysis of the proportional hazards model with sparse longitudinal covariates. *Journal of the American Statistical Association 110*(511), 1187–1196.

Chen, C.-J. (2007). The instantaneous and forward default intensity of structural models. *Working paper*.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society 34*(2), 187–220.

Dickson, E., T. Fleming, R. Wiesner, W. Baldus, C. Fleming, J. Ludwig, and J. McCall (1985). Trial of penicillamine in advanced primary biliary cirrhosis. *New England Journal of Medicine 312*(16), 1011–1015.

Ding, J. and J.-L. Wang (2008). Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. *Biometrics 64*(2), 546–556.

Duan, J.-C., J. Sun, and T. Wang (2012). Multiperiod corporate default prediction—a forward intensity approach. *Journal of Econometrics 170*(1), 191–209.

Eilers, P. H. and B. D. Marx (1996). Flexible smoothing with b-splines and penalties. *Statistical science*, 89–102.

Eilers, P. H., B. D. Marx, and M. Durbán (2015). Twenty years of p-splines. *SORT: statistics and operations research transactions 39*(2), 0149–186.

Fan, J. (1996). On local smoothing of nonparametric curve estimators. *Journal of the American Statistical Association 91*(433), 258–266.

Fan, J., I. Gijbels, and M. King (1997). Local likelihood and local partial likelihood in hazard regression. *The Annals of Statistics*, 1661–1690.

Faucett, C. L. and D. C. Thomas. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: a gibbs sampling approach. *Statistics in medicine 15*(15), 1663–1685.

Garre, F. G., A. H. Zwinderman, R. B. Geskus, and Y. W. Sijpkens (2008). A joint latent class changepoint model to improve the prediction of time to graft failure. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 171*(1), 299–308.

Goldman, A. I., B. P. Carlin, L. R. Crane, C. Launer, J. A. Korvick, L. Deyton, and D. I. Abrams (1996). Response of cd4 lymphocytes and clinical consequences of treatment using ddi or ddc in patients with advanced hiv infection. *JAIDS Journal of Acquired Immune Deficiency Syndromes 11*(2), 161–169.

Henderson, R., P. Diggle, and A. Dobson (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics 1*(4), 465–480.

Houwelingen, H. C. V. (2007, mar). Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics 34*(1), 70–85.

Hunsberger, S. (1994). Semiparametric regression in likelihood-based models. *Journal of the American Statistical Association 89*(428), 1354–1365.

Ibrahim, J. G., M. Chen, and D. Sinha (2001). Bayesian semiparametric models for survival data with a cure fraction. *Biometrics 57*(2), 383–388.

Kalbfleisch, J. D. and R. L. Prentice (2011). *The statistical analysis of failure time data*, Volume 360. John Wiley Sons.

Klein, J. P. and M. L. Moeschberger (2005). *Survival analysis: techniques for censored and truncated data*. Springer Science Business Media.

Lenglart, E. (1977). Relation de domination entre deux processus. *Ann. Inst. H. Poincaré Sect. B (NS) 13*(2), 171–179.

Murtaugh, P. A., E. R. Dickson, G. M. Van Dam, M. Malinchoc, P. M. Grambsch, A. L. Langworthy, and C. H. Gips (1994). Primary biliary cirrhosis: prediction of short-term survival based on repeated patient visits. *Hepatology 20*(1), 126–134.

O'Sullivan, F. (1988). Fast computation of fully automated log-density and log-hazard estimators. *SIAM Journal on scientific and statistical computing 9*(2), 363–379.

Parast, L., S.-C. Cheng, and T. Cai (2012, aug). Landmark prediction of long-term survival incorporating short-term event time information. *Journal of the American Statistical Association 107*(500), 1492–1501.

Parast, L., L. Tian, and T. Cai (2014, jan). Landmark estimation of survival and treatment effect in a randomized clinical trial. *Journal of the American Statistical Association 109*(505), 384–394.

Pauler, D. K. and D. M. Finkelstein (2002). Predicting time to prostate cancer recurrence based on joint models for non-linear longitudinal biomarkers and event time outcomes. *Statistics in medicine 21*(24), 3897–3911.

Prentice, R. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika 69*(2), 331–342.

Proust-Lima, C. and J. M. Taylor (2009). Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment psa: a joint modeling approach. *Biostatistics 10*(3), 535–549.

Rizopoulos, D. (2010). Jm: An r package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software (Online) 35*(9), 1–33.

Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics 67*(3), 819–829.

Rizopoulos, D. (2016). The r package jmbayes for fitting joint models for longitudinal and time-to-event data using mcmc. *Journal of Statistical Software 72*(7).

Rizopoulos, D., G. Verbeke, and E. Lesaffre (2009). Fully exponential laplace approximations for the joint modelling of survival and longitudinal data.

Journal of the Royal Statistical Society: Series B (Statistical Methodology) 71(3), 637–654.

Shao, J. (1999). *Mathematical Statistics*. Springer.

Song, X., M. Davidian, and A. A. Tsiatis (2002). A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics 58*(4), 742–753.

Staniswalis, J. G. (1989). The kernel estimate of a regression function in likelihood-based models. *Journal of the American Statistical Association 84*(405), 276–283.

Taylor, J. M., Y. Park, D. P. Ankerst, C. Proust-Lima, S. Williams, L. Kestin, K. Bae, T. Pickles, and H. Sandler (2013). Real-time individual predictions of prostate cancer recurrence using joint models. *Biometrics 69*(1), 206–213.

Taylor, J. M., M. Yu, and H. M. Sandler (2005). Individualized predictions of disease progression following radiation therapy for prostate cancer. *Journal of clinical oncology 23*(4), 816–825.

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

Tibshirani, R. and T. Hastie (1987). Local likelihood estimation. *Journal of the American Statistical Association 82*(398), 559–567.

Tjahjowidodo, T. (2017). A direct method to solve optimal knots of b-spline curves: An application for non-uniform b-spline curves fitting. *PloS one*.

Tseng, Y., Y. Su, M. Mao, and J. Wang (2015). An extended hazard model with longitudinal covariates. *Biometrika 102*(1), 135–150.

Tseng, Y.-K., F. Hsieh, and J.-L. Wang (2005). Joint modelling of accelerated failure time and longitudinal data. *Biometrika 92*(3), 587–603.

Tsiatis, A. A. and M. Davidian (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, 809–834.

Tsiatis, A. A., V. D. and M. S. Wulfsohn (1995). Modeling the relationship of survival to longitudinal data measured with error. applications to survival and cd4 counts in patients with aids. *Journal of the American Statistical Association*, 27–37.

Van Houwelingen, H. C. and H. Putter (2008). Dynamic predicting by landmarking as an alternative for multi-state modeling: an application to acute lymphoid leukemia data. *Lifetime data analysis 14*(4), 447.

Varin, Cristiano, N. R. and D. Firth (2011). An overview of composite likelihood methods. *Statistica Sinica*, 5–42.

Wand, M. P. and J. T. Ormerod (2008). On semiparametric regression with o'sullivan penalized splines. *Australian and New Zealand Journal of Statistics 50*(2), 179–198.

Wang, Y. and J. M. G. Taylor. (2001). Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association 96*(455), 895–905.

Wulfsohn, M. S. and A. A. Tsiatis (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 330–339.

Yu, M., N. J. Law, J. M. Taylor, and H. M. Sandler (2004). Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica*, 835–862.

Yu, M., J. M. G. Taylor, and H. M. Sandler (2008). Individual prediction in prostate cancer studies using a joint longitudinal survival–cure model. *Journal of the American Statistical Association 103*(481), 178–187.

Zheng, Y. and P. J. Heagerty (2005, jun). Partly conditional survival models for longitudinal data. *Biometrics 61*(2), 379–391.

APPENDIX A

MORE NUMERICAL RESULTS

A.1 More results from the data analysis

Tables A.1 and A.2 contains the estimates of the parameters plotted in Figures 6.1 and 6.2.

Table A.1: Parameter estimates of the forward intensity function of the event in the CD4 data.

	$\hat{\boldsymbol{\alpha}}(\tau)$ (SD)									
	LOCF					Smoothed				
τ	Int.	$(\mathrm{CD4})^{\frac{1}{4}}$	Drug	t	PrevOI	Int.	$(\mathrm{CD4})^{\frac{1}{4}}$	Drug	t	PrevOI
1	-2.985 (0.502)	-0.616 (0.122)	-0.285 (0.247)	0.041 (0.021)	0.964 (0.332)	-2.474 (0.495)	-0.673 (0.125)	-0.238 (0.266)	-0.021 (0.025)	0.754 (0.357)
2	-2.689 (0.513)	-0.665 (0.126)	-0.338 (0.252)	0.035 (0.023)	0.883 (0.335)	-2.351 (0.509)	-0.678 (0.128)	-0.269 (0.270)	-0.011 (0.025)	0.716 (0.363)
3	-2.611 (0.538)	-0.652 (0.133)	-0.368 (0.258)	0.032 (0.025)	0.875 (0.345)	-2.344 (0.533)	-0.647 (0.132)	-0.308 (0.271)	-0.005 (0.027)	0.741 (0.372)
4	-2.572 (0.564)	-0.649 (0.135)	-0.385 (0.267)	0.037 (0.026)	0.832 (0.356)	-2.196 (0.548)	-0.679 (0.131)	-0.332 (0.280)	0.002 (0.027)	0.713 (0.382)
5	-2.587 (0.587)	-0.600 (0.133)	-0.362 (0.271)	0.025 (0.029)	0.904 (0.373)	-2.259 (0.565)	-0.623 (0.128)	-0.353 (0.289)	-0.011 (0.031)	0.763 (0.398)
6	-2.499 (0.606)	-0.598 (0.134)	-0.370 (0.283)	0.023 (0.032)	0.861 (0.378)	-2.240 (0.575)	-0.603 (0.132)	-0.340 (0.297)	-0.021 (0.036)	0.760 (0.407)
7	-2.427 (0.617)	-0.578 (0.138)	-0.367 (0.293)	0.005 (0.037)	0.912 (0.385)	-2.213 (0.592)	-0.597 (0.142)	-0.324 (0.305)	-0.053 (0.043)	0.814 (0.416)
8	-2.397 (0.637)	-0.529 (0.139)	-0.251 (0.303)	-0.015 (0.043)	0.846 (0.390)	-2.194 (0.596)	-0.574 (0.141)	-0.221 (0.308)	-0.060 (0.043)	0.779 (0.397)
9	-2.606 (0.706)	-0.482 (0.141)	-0.317 (0.328)	-0.007 (0.049)	0.974 (0.439)	-2.428 (0.685)	-0.539 (0.152)	-0.244 (0.336)	-0.039 (0.049)	0.932 (0.453)

Table A.1: Parameter estimates of the forward intensity function of the event in the CD4 data.

	$\hat{\boldsymbol{\alpha}}(\tau)$ (SD)									
	LOCF					Smoothed				
τ	Int.	$(\mathrm{CD4})^{\frac{1}{4}}$	Drug	t	PrevOI	Int.	$(\mathrm{CD4})^{\frac{1}{4}}$	Drug	t	PrevOI
10	-2.200 (0.715)	-0.522 (0.140)	-0.333 (0.334)	-0.072 (0.063)	0.955 (0.448)	-2.334 (0.714)	-0.544 (0.160)	-0.318 (0.349)	-0.082 (0.064)	0.991 (0.477)
11	-2.188 (0.865)	-0.588 (0.172)	-0.302 (0.374)	-0.075 (0.078)	1.034 (0.529)	-2.593 (0.814)	-0.534 (0.191)	-0.169 (0.394)	-0.089 (0.088)	1.095 (0.534)
12	-2.544 (1.057)	-0.522 (0.223)	-0.321 (0.428)	-0.057 (0.095)	1.160 (0.606)	-3.198 (0.924)	-0.404 (0.222)	-0.023 (0.446)	-0.078 (0.105)	1.319 (0.570)

NOTE: Both LOCF and smoothed composite likelihood approach are used to estimate the parameters; SD is the standard deviation; τ is the prediction horizon examined; Int. is the intercept in the forward intensity function; predictor variables include drug ddC, the indicator (prevOI) for patients with previous opportunistic infection, transformed CD4 counts ($\mathrm{CD4}^{1/4}$) and its observation time t.

Table A.2: Parameter estimates of the forward intensity function of the competing event in the CD4 data.

	$\hat{\boldsymbol{\beta}}(\tau)$ (SD)									
	LOCF					Smoothed				
τ	Int.	$(\mathrm{CD4})^{\frac{1}{4}}$	Drug	t	PrevOI	Int.	$(\mathrm{CD4})^{\frac{1}{4}}$	Drug	t	PrevOI
	-7.608	0.191	0.184	0.359	0.237	-4.948	0.139	0.080	0.232	0.125
1	(0.892)	(0.107)	(0.194)	(0.043)	(0.177)	(0.513)	(0.109)	(0.207)	(0.019)	(0.193)
	-7.238	0.194	0.185	0.357	0.239	-4.710	0.152	0.074	0.228	0.084
2	(0.824)	(0.108)	(0.192)	(0.040)	(0.177)	(0.504)	(0.110)	(0.208)	(0.019)	(0.196)
	-6.979	0.198	0.191	0.363	0.250	-4.436	0.147	0.057	0.222	0.082
3	(0.774)	(0.107)	(0.191)	(0.038)	(0.177)	(0.477)	(0.107)	(0.200)	(0.018)	(0.192)
	-6.506	0.176	0.186	0.360	0.233	-4.188	0.142	0.104	0.221	0.105
4	(0.687)	(0.103)	(0.186)	(0.034)	(0.175)	(0.442)	(0.100)	(0.188)	(0.016)	(0.183)
	-6.036	0.159	0.182	0.354	0.219	-3.868	0.130	0.126	0.209	0.100
5	(0.623)	(0.099)	(0.182)	(0.032)	(0.172)	(0.399)	(0.092)	(0.178)	(0.015)	(0.174)
	-5.476	0.117	0.170	0.346	0.188	-3.539	0.113	0.121	0.207	0.082
6	(0.557)	(0.095)	(0.178)	(0.030)	(0.169)	(0.355)	(0.084)	(0.165)	(0.014)	(0.164)
	-5.047	0.119	0.168	0.336	0.190	-3.240	0.091	0.096	0.202	0.080
7	(0.503)	(0.092)	(0.173)	(0.029)	(0.165)	(0.334)	(0.082)	(0.157)	(0.015)	(0.157)
	-4.508	0.093	0.159	0.321	0.171	-3.004	0.091	0.085	0.203	0.055
8	(0.447)	(0.088)	(0.168)	(0.027)	(0.164)	(0.322)	(0.080)	(0.151)	(0.016)	(0.152)
	-4.042	0.087	0.162	0.304	0.162	-2.719	0.079	0.070	0.190	0.060
9	(0.402)	(0.084)	(0.162)	(0.024)	(0.161)	(0.304)	(0.077)	(0.144)	(0.016)	(0.147)

Table A.2: Parameter estimates of the forward intensity function of the competing event in the CD4 data.

	$\hat{\boldsymbol{\beta}}(\tau)$ (SD)									
	LOCF					Smoothed				
τ	Int.	$(\mathrm{CD4})^{\frac{1}{4}}$	Drug	t	PrevOI	Int.	$(\mathrm{CD4})^{\frac{1}{4}}$	Drug	t	PrevOI
10	-3.545 (0.378)	0.089 (0.083)	0.152 (0.157)	0.273 (0.024)	0.164 (0.158)	-2.487 (0.289)	0.081 (0.075)	0.108 (0.137)	0.186 (0.016)	0.100 (0.142)
11	-3.009 (0.351)	0.091 (0.082)	0.139 (0.148)	0.226 (0.022)	0.168 (0.151)	-2.231 (0.284)	0.071 (0.076)	0.120 (0.135)	0.139 (0.017)	0.106 (0.141)
12	-2.306 (0.315)	0.063 (0.079)	0.115 (0.138)	0.142 (0.021)	0.155 (0.144)	-1.982 (0.287)	0.073 (0.079)	0.118 (0.136)	0.094 (0.019)	0.099 (0.143)

Table A.3 contains the AUC corresponding to the CD4 data analysis.

A.2 Additional simulations

This section provides additional simulation scenarios considered.

Study G-1

The event time T is generated from the Gompertz distribution with shape parameter $a = 0.1$ and rate parameter $b = 0.005$, so that the conditional survival function is $S_{T,t}(u) = \mathbb{P}(T > t+u|T > t) = \exp[\frac{b}{a}\{e^{at} - e^{a(t+u)}\}]$. This implies that the forward event intensity function at t is $\lambda_{T,t}(u) = \exp\{\log(b) + au + at\}$, which can be expressed as $\lambda_{T,t}(u) = \exp\{\alpha_0(u) + \alpha_1(u)t\}$, where $\alpha_0(u) = \log(b) + au = \log(0.005) + 0.1u$ and $\alpha_1(u) = a$=0.1. We randomly generate censoring time C from an exponential distribution with the mean parameter 40. For a given individual, the observation time $X = \min(T, C)$ with event indicator $\delta_T = 1$ if $X = T$ and $\delta_T = 0$ if otherwise.

This is a benchmarking case with no missing data and using the time alone as the explanatory variable, based on which we evaluate the validity and accuracy of the composite likelihood approach for estimating the forward intensity function. Parameter estimators are obtained using the time point estimation by maximizing log-likelihood functions $l^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}$ and $l^{(2)}\{\boldsymbol{\beta}(k\Delta)\}$ in (3.7.1) and (3.7.2) for each prediction horizon time point ($u = k\Delta = 1, \ldots, 10$) separately.

Table A.8 summarizes the accuracy of parameter estimates in terms of their biases and variances. The standard deviation is small and stable over different prediction horizons $u = 1, 2, \ldots, 10$. The relative bias (RB) for $\alpha_0(u)$ is also small and stable over different prediction horizons $u = 1, 2, \ldots, 10$. The RB for $\alpha_1(u)$ increases slightly with increasing prediction horizon u, but overall still small and stable. Both RB and SD improve with larger number of subjects in the simulations.

Study G-2

This study incorporates additionally a time-invariant predictor and a categorical grouping variable in the event time generation. We generate T with the hazard function $h(t) = b \times \exp(I \times a z_i t)$, in which $a = 0.3$ and $b = 0.02$. The time-invariant predictor $z_i \sim N(0.4, 0.2^2)$ and $I \sim \text{Bernoulli}(0.5)$. The event time T is generated by using the implied survival function $S(T) = \exp\{-\int_0^T h(u)du\}$. When $I = 1$, the conditional forward survival function for $t \geq 0$ is $S_{T,t}(u) = \mathbb{P}(T > t + u | T > t) = \exp[-\frac{b}{az_i}\{e^{az_i(t+u)} - e^{az_i t}\}]$. When $I = 0$, the conditional forward survival function for $t \geq 0$ is $S_{T,t}(u) = \mathbb{P}(T > t + u | T > t) = \exp(-bu)$. The forward intensity function at time t is $\lambda_{T,t}(u) = \exp\{\alpha_0(u) + \alpha_1(u) I \times z_i + \alpha_2(u) I \times t z_i\}$, where $\alpha_0(u) = \log b = \log(0.02) = -3.912$, $\alpha_1(u) = au = 0.3u$ and $\alpha_2(u) = a = 0.3$. We have included an extra explanatory variable in this setting. We also generate the random censoring time C from the exponential distribution with mean of 40. For a given individual, the observation time $X = \min(T, C)$ with event indicator $\delta_T = 1$ if $X = T$ and $\delta_T = 0$ if otherwise.

Parameter estimators are obtained using the time point estimation by maximizing log-likelihood functions $l^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}$ and $l^{(2)}\{\boldsymbol{\beta}(k\Delta)\}$ in (3.7.1) and (3.7.2) for each prediction horizon time point ($u = k\Delta = 1, \ldots, 10$) separately.

Table A.9 summarizes the accuracy of parameter estimates in terms of their biases and variances. The standard deviation is small and stable over different prediction horizons $u = 1, 2, \ldots, 10$. The RB for $\alpha_0(u)$ is small and stable over different prediction horizons. The RB for $\alpha_2(u)$ increase slightly with increasing prediction horizon u, but overall still small and stable. As a function of u, $\alpha_1(u)$ has higher RB for some early prediction horizons while the other two parameters with constant true values have low RB. Both RB and SD improve with larger number of subjects in the simulations.

Study G-3

In the study, both event time T and the competing event time O are assumed to follow Gompertz distributions but with different rate and shape parameters. In the forward intensity function of both the event and the competing event, the intercept is a linear function of u, and the slope is a constant. They are generated as below.

The event time T is generated from the Gompertz distribution with shape parameter $a = 0.02$ and rate parameter $b = 0.01$, so that the conditional survival function of the event is $S_{T,t}(u) = \mathbb{P}(T > t + u | T > t) = \exp[\frac{b}{a}\{e^{at} - e^{a(t+u)}\}]$. This implies that the forward event intensity function at t is $\lambda_{T,t}(u) = \exp\{\log(b) + au + at\}$, which can be expressed as $\lambda_{T,t}(u) = \exp\{\alpha_0(u) + \alpha_1(u)t\}$, where $\alpha_0(u) = \log(b) + au = \log(0.02) + 0.01u$ and $\alpha_1(u) = a$=0.01.

The competing event time O is generated from the Gompertz distribution with shape parameter $a = 0.005$ and rate parameter $b = 0.33$. Hence the conditional forward intensity function of the competing event can be modeled as $\lambda_{O,t}(u) = \exp\{\beta_0(u) + \beta_1(u)t\}$, where $\beta_0(u) = \log(0.005) + 0.33u$ and $\beta_1(u)$=0.33.

The end time of the study E is assumed to be 15. For a given individual, the observation time $X = \min(T, O, 15)$. The event indicator $\delta_T = 1$ if $X = T$, and 0 otherwise. The competing event indicator $\delta_O = 1$ if $X = O$, and 0 otherwise.

Under the penalized B-spline method, the intercept $\alpha_0(u)$ and $\beta_0(u)$ and the slope $\alpha_1(u)$ and $\beta_1(u)$ in the forward intensity function of event and competing event are modeled as a function of the B-spline function $B_{j,q}(k\Delta)$.

$$\hat{\alpha}_r(k\Delta) = \sum_j \hat{a}_{rj} B_{j,q}(k\Delta);$$
$$\hat{\beta}_r(k\Delta) = \sum_j \hat{b}_{rj} B_{j,q}(k\Delta).$$

in which $r = 0, 1$ and $u = k\Delta = 1, 2, 3, 4, 5$. The degree $(q - 1)$ of 2 and 3 (corresponding to the order of 3 and 4 respectively) for the B-spline basis functions are considered for each of the 4 parameters. With 0 internal knot considered, the number of B-spline basis functions is the same as the degree of the B-spline function plus 1 (or the order of the B-spline function q). Hence the number of B-spline basis functions is 3 and 4 for degree of 2 and 3 B-splines respectively. The lower boundary knot is set at 0 and upper boundary knot is set at 6. The first column in Figure 6.15 shows the B-spline basis functions with degree ranging from 0 to 3 and no inner knot chosen.

The penalty coefficient of λ examined for the penalized B-spline method is in the range of 0 to 5 with an increment of 1. It is optimized using the 3-fold cross validation which was explained in Section 3.8.

In Table A.4, the parameter estimators by each forward-looking time point and the penalized B-spline method (PBS) are summarized using three measures $\sum_{k,r} |\text{bias}_{\alpha_r(k\Delta),\beta_r(k\Delta)}|$, $\sum_{k,r} SD_{\alpha_r(k\Delta),\beta_r(k\Delta)}$ and $\sum_{k,r} \sqrt{MSE_{\alpha_r(k\Delta),\beta_r(k\Delta)}}$ considering all 4 parameters $\alpha_r(k\Delta)$ and $\beta_r(k\Delta)$ ($r = 0, 1$) and $u = k\Delta = 1, 2, \ldots, 5$ in each of the 3 train sets among the 100 simulated data. The average of each of these three measures among three train train sets are also summarized in the Table A.4.

The prediction performance is examined by AUC. The mean and SD of the average AUC across all prediction horizon time points ($u = k\Delta = 1, 2, \ldots, 5$) in each of the three test sets of the 100 simulated data are summarized in Table A.6. The average AUC across three test sets are also summarized in Table A.6. The λ value is optimized by choosing the highest average AUC across all prediction horizon time points considering all three test sets together.

When the sample size is 45, the highest average AUC is 70.5% and 70.49% by the penalized B-spline method using degree of 2 and 3 B-spline functions respectively compared to 70.09% by time point estimation. The λ value corresponding to the best average AUC by the penalized B-spline method ranges from 3 to 5 for both degree of 2 and 3 B-splines. When $\lambda \geq 1$, the parameter estimates measure $\sum_{k,r} \sqrt{MSE_{\alpha_r(k\Delta),\beta_r(k\Delta)}}$ by the penalized B-spline method is less than 41% of that by the time point estimation. The

$\sum_{k,r} \sqrt{MSE_{\alpha_r(k\Delta),\beta_r(k\Delta)}}$ is similar for different $\lambda \geq 1$ values regardless of the degree of the B-splines, and is the smallest when $\lambda = 5$. The improvement in $\sum_{k,r} \sqrt{MSE_{\alpha_r(k\Delta),\beta_r(k\Delta)}}$ by the penalized B-spline method comes from both smaller bias and less variability especially when $\lambda \geq 1$.

When the sample size is 90, the highest average AUC is 71.62% and 71.64% by the penalized B-spline method using degree of 2 and 3 B-splines respectively compared to 71.43% by time point estimation. The λ value corresponding to the best average AUC ranges between 3 and 5 for degree of 2 B-splines, and between 1 and 3 for degree of 3. When $\lambda \geq 1$, the parameter estimates measure $\sum_{k,r} \sqrt{MSE_{\alpha_r(k\Delta),\beta_r(k\Delta)}}$ by the penalized B-spline method is less than 95% of that by the time point estimation. The $\sum_{k,r} \sqrt{MSE_{\alpha_r(k\Delta),\beta_r(k\Delta)}}$ is similar for different $\lambda \geq 1$ values regardless of the degree of the B-splines, and is the smallest when $\lambda = 5$. The improvement in $\sum_{k,r} \sqrt{MSE_{\alpha_r(k\Delta),\beta_r(k\Delta)}}$ by the penalized B-spline method comes from both smaller bias and less variability especially when $\lambda \geq 1$.

Table A.5 summarizes the accuracy of parameter estimates of $\alpha_r(k\Delta)$ and $\beta_r(k\Delta)$ $(r = 0, 1)$ in terms of their bias, relative bias (RB), standard deviation (SD), and square root of mean square error ($\sqrt{MSE}$). These measures are summarized for each parameter at each prediction horizon $u = k\Delta = 1, 2, 3, 4$ and 5 within each train set by both methods. When sample size is 45, Table A.5 presents the results for λ of 5 for both degree of 2 and 3 B-splines which corresponds to the highest average AUC and the lowest $\sum_{k,r} \sqrt{MSE_{\alpha_r(k\Delta),\beta_r(k\Delta)}}$. When the sample size is 90, Table A.5 presents the results for λ of 5 and 3 for degree of 2 and 3 B-splines respectively. The improvement in bias, SD and $\sqrt{MSE}$ by the penalized B-spline method is more prominent in longer prediction horizon when there are fewer data points available for the composite likelihood function for the estimation by each forward-looking time point. This is more so for the parameters $\alpha_0(u)$ and $\alpha_1(u)$ at time $u = 4, 5$ when the sample size $n = 45$. The $\sqrt{MSE}$ by the penalized B-spline method is less than 25% of the $\sqrt{MSE}$ by time point estimation for these two parameters at time $u = 5$. The improvement in $\sqrt{MSE}$ at time $u = 4, 5$ becomes less significant when $n = 90$. The $\sqrt{MSE}$ by the penalized B-spline method is less than 89% of the $\sqrt{MSE}$ by time point estimation. Table A.7 summarizes the mean and SD of the AUC at each prediction time point $u = k\Delta = 1, 2, 3, 4$ and 5 in each of the 3 test sets of the 100 simulated data using the optimized λ value presented in Table A.5.

In summary, the results observed in Study G-3 is very similar as that in Study E-3. The advantage of the penalized B-spline method compared to the estimation by each forward-looking time point is more prominent when sample size is smaller measured by both parameter estimates measure $\sum_{k,r} \sqrt{MSE_{\alpha_r(k\Delta),\beta_r(k\Delta)}}$ and average AUC. This is due to the fact that the penalized B-spline method allows the composite log-likelihood function to incorporate contributions from multiple prediction horizons examined. With more data taken into account, it enables the parameter estimation to leverage strength across prediction horizons examined. The parameter estimates measure $\sum_{k,r} SD_{\alpha_r(k\Delta),\beta_r(k\Delta)}$ and $\sum_{k,r} \sqrt{MSE_{\alpha_r(k\Delta),\beta_r(k\Delta)}}$ become smaller with larger sample size. $\sum_{k,r} |\text{bias}_{\alpha_r(k\Delta),\beta_r(k\Delta)}|$ and $\sum_{k,r} SD_{\alpha_r(k\Delta),\beta_r(k\Delta)}$ are not susceptible to the choice of the degree of the B-splines with proper choice of λ value and the results are fairly consistent when the penalty coefficient $\lambda \geq 1$. This demonstrates the robustness and validity of the penalized B-spline method.

Table A.3: Prediction performance evaluated by AUC (%) over prediction horizons 2-12 months in CD4 data comparing the joint modeling approach and the forward intensity method.

		JM		LOCF	Smooth		
	τ	rocJM	$\pi_i(t_i+\tau\|t)$		γ=0.7	γ=0.5	γ=0.2
in-sample	2	66.8	55.8	61.7	73.8	72.9	72.6
	3	67.1	55.8	65.2	76.5	74.3	75.7
	4	67.4	55.8	69.1	76.5	75.8	78.0
	5	67.7	55.7	71.9	78.5	77.7	79.6
	6	68.1	56.0	75.6	80.6	80.3	81.7
	7	68.5	55.9	78.8	82.6	82.8	83.8
	8	68.8	55.9	80.7	84.1	84.1	84.5
	9	69.2	55.7	83.0	86.1	86.1	86.2
	10	69.5	55.6	84.6	87.4	87.3	87.2
	11	69.8	55.5	85.6	88.3	88.2	87.9
	12	70.1	55.3	86.7	89.1	88.9	88.6
out-of-sample	2	61.3	46.3	60.0	74.0	71.9	70.2
	3	61.6	40.9	65.1	76.1	73.2	73.3
	4	61.9	47.4	71.5	80.1	78.7	79.2
	5	62.2	48.3	72.4	79.6	78.6	78.5
	6	62.5	49.2	75.2	81.8	80.8	80.7
	7	62.8	47.3	79.5	84.6	84.0	83.8
	8	63.0	48.4	81.0	85.4	85.1	84.8
	9	63.2	48.9	82.1	85.9	85.4	84.9
	10	63.5	47.3	83.6	87.1	86.7	86.0
	11	63.8	46.7	84.6	87.8	87.4	86.7
	12	64.0	48.1	86.3	89.3	88.8	88.1

NOTE: Both in-sample (top) and out-of-sample (bottom) CD4 data are used; τ is the prediction horizon examined; JM is the joint modeling method; rocJM is the mean AUC under the joint modeling method using the definitions of sensitivity and specificity per Rizopoulos (2011); $\pi_i(t_i+\tau|t)$ is the predicted survival probability under the joint model framework; both LOCF and smoothed composite likelihood approach are evaluated under the forward intensity method; γ is the parameter value for bandwidth calculation in the proposed smoothed composite likelihood.

Table A.4: Simulation results in Study G-3 comparing estimation by time point (TP) and penalized B-spline (PBS/Degree) under the forward intensity method.

			Train set 1			Train set 2			Train set 3			Average		
n	Method	λ	$\sum_{k,r}\lvert Bias\rvert$	$\sum_{k,r} SD$	$\sum_{k,r}\sqrt{MSE}$	$\sum_{k,r}\lvert Bias\rvert$	$\sum_{k,r} SD$	$\sum_{k,r}\sqrt{MSE}$	$\sum_{k,r}\lvert Bias\rvert$	$\sum_{k,r} SD$	$\sum_{k,r}\sqrt{MSE}$	$\sum_{k,r}\lvert Bias\rvert$	$\sum_{k,r} SD$	$\sum_{k,r}\sqrt{MSE}$
45	TP	NA	3.134	19.431	19.689	2.903	21.881	22.092	4.051	28.245	28.564	3.362	23.186	23.448
	PBS/2	0	1.43	8.654	8.779	2.24	18.254	18.413	2.964	22.91	23.127	2.211	16.606	16.773
		1	1.279	7.989	8.099	1.109	7.962	8.063	0.742	7.984	8.044	1.043	7.978	8.069
		2	1.279	7.983	8.094	1.108	7.958	8.058	0.741	7.979	8.039	1.043	7.973	8.064
		3	1.279	7.98	8.092	1.108	7.956	8.056	0.74	7.976	8.037	1.043	7.971	8.061
		4	1.279	7.979	8.09	1.108	7.954	8.054	0.74	7.975	8.036	1.042	7.969	8.06
		5	1.28	7.978	8.089	1.108	7.953	8.054	0.74	7.974	8.035	1.042	7.968	8.059
	PBS/3	0	2.342	14.832	15.023	1.428	9.7	9.826	1.461	11.872	11.989	1.744	12.135	12.279
		1	1.279	8.003	8.113	1.109	7.974	8.074	0.746	7.997	8.058	1.044	7.991	8.082
		2	1.279	7.996	8.107	1.109	7.968	8.069	0.744	7.991	8.051	1.044	7.985	8.076
		3	1.278	7.992	8.103	1.109	7.965	8.065	0.743	7.987	8.048	1.044	7.981	8.072
		4	1.278	7.99	8.1	1.109	7.963	8.063	0.743	7.984	8.045	1.043	7.979	8.07
		5	1.279	7.988	8.099	1.109	7.962	8.062	0.742	7.983	8.043	1.043	7.977	8.068
90	TP	NA	1.308	6.529	6.713	1.118	5.513	5.662	1.16	5.977	6.134	1.195	6.006	6.17
	PBS/2	0	1.276	6.368	6.551	1.089	5.356	5.504	1.131	5.814	5.972	1.166	5.846	6.009
		1	1.239	6.192	6.373	1.052	5.136	5.284	1.106	5.658	5.817	1.133	5.662	5.825
		2	1.239	6.188	6.37	1.053	5.133	5.281	1.106	5.654	5.813	1.133	5.659	5.821
		3	1.239	6.187	6.369	1.053	5.132	5.28	1.106	5.652	5.811	1.133	5.657	5.82
		4	1.239	6.186	6.368	1.053	5.131	5.279	1.106	5.65	5.81	1.133	5.656	5.819
		5	1.24	6.185	6.367	1.053	5.13	5.278	1.107	5.649	5.809	1.133	5.655	5.818
	PBS/3	0	1.296	6.464	6.648	1.099	5.409	5.557	1.148	5.912	6.069	1.181	5.929	6.091
		1	1.24	6.2	6.381	1.053	5.145	5.292	1.109	5.669	5.828	1.134	5.671	5.834
		2	1.24	6.196	6.377	1.053	5.141	5.288	1.107	5.663	5.823	1.133	5.667	5.829
		3	1.239	6.194	6.375	1.053	5.138	5.286	1.106	5.661	5.82	1.133	5.664	5.827

4	1.239	6.192	6.374	1.052	5.137	5.285	1.106	5.659	5.818	1.133	5.663	5.825
5	1.239	6.191	6.373	1.052	5.136	5.284	1.106	5.657	5.817	1.133	5.661	5.824

Table A.5: Parameter summary in Study G-3 comparing estimation by time point (TP) and penalized B-spline (PBS/Degree) under the forward intensity method.

						Train set 1				Train set 2				Train set 3			
n	Method	λ	Parameter	u	True	Bias	RB	SD	$\sqrt{MSE}$	Bias	RB	SD	$\sqrt{MSE}$	Bias	RB	SD	$\sqrt{MSE}$
45	TP	NA	$\alpha_0(u)$	1	-3.902	-0.078	0.02	0.688	0.692	0.021	-0.005	0.779	0.779	-0.019	0.005	0.63	0.63
				2	-3.892	-0.085	0.022	0.722	0.727	-0.066	0.017	0.794	0.796	-0.057	0.015	0.621	0.624
				3	-3.882	-0.069	0.018	0.702	0.705	-0.173	0.045	0.825	0.843	-0.073	0.019	0.61	0.615
				4	-3.872	-0.183	0.047	1.075	1.091	-0.491	0.127	3.502	3.536	-0.805	0.208	4.87	4.936
				5	-3.862	-0.903	0.234	4.958	5.04	-0.63	0.163	3.643	3.697	-0.791	0.205	4.892	4.956
			$\alpha_1(u)$	1	0.01	-0.022	-2.205	0.115	0.117	-0.048	-4.769	0.143	0.151	-0.035	-3.508	0.13	0.135
				2	0.01	-0.029	-2.869	0.144	0.146	-0.046	-4.641	0.172	0.178	-0.039	-3.884	0.152	0.157
				3	0.01	-0.041	-4.133	0.172	0.177	-0.174	-17.395	1.496	1.506	-0.342	-34.219	2.231	2.257
				4	0.01	-0.372	-37.178	2.499	2.527	-0.404	-40.376	3.614	3.637	-0.761	-76.102	5.084	5.141
				5	0.01	-0.764	-76.421	5.087	5.144	-0.507	-50.743	3.78	3.814	-0.965	-96.5	5.402	5.488
			$\beta_0(u)$	1	-4.968	0.105	-0.021	0.626	0.635	-0.066	0.013	0.624	0.627	-0.068	0.014	0.741	0.744
				2	-4.638	0.132	-0.029	0.583	0.598	-0.046	0.01	0.587	0.588	-0.037	0.008	0.703	0.704
				3	-4.308	0.105	-0.024	0.597	0.606	-0.075	0.017	0.553	0.558	-0.011	0.003	0.64	0.64
				4	-3.978	0.088	-0.022	0.556	0.563	-0.08	0.02	0.52	0.526	-0.013	0.003	0.562	0.563
				5	-3.648	0.084	-0.023	0.522	0.529	-0.061	0.017	0.506	0.509	-0.024	0.007	0.564	0.565
			$\beta_1(u)$	1	0.33	-0.012	-0.036	0.068	0.069	0.002	0.007	0.062	0.062	0.006	0.017	0.075	0.076
				2	0.33	-0.016	-0.048	0.07	0.072	0	0.001	0.064	0.064	0.003	0.008	0.08	0.08
				3	0.33	-0.015	-0.045	0.079	0.08	0.004	0.011	0.067	0.067	0	0	0.082	0.082
				4	0.33	-0.015	-0.044	0.082	0.08	0.005	0.014	0.071	0.071	0	0.001	0.081	0.081
				5	0.33	-0.016	-0.05	0.087	0.088	0.003	0.008	0.08	0.08	0.002	0.005	0.094	0.094
	PBS/2	5	$\alpha_0(u)$	1	-3.902	-0.039	0.01	0.672	0.673	0.026	-0.007	0.78	0.78	-0.008	0.002	0.606	0.606
				2	-3.892	-0.064	0.016	0.627	0.631	-0.032	0.008	0.684	0.685	-0.031	0.008	0.545	0.546
				3	-3.882	-0.089	0.023	0.686	0.692	-0.089	0.023	0.674	0.679	-0.054	0.014	0.57	0.573
				4	-3.872	-0.113	0.029	0.827	0.835	-0.146	0.038	0.751	0.766	-0.076	0.02	0.672	0.676
				5	-3.862	-0.138	0.036	1.016	1.025	-0.203	0.053	0.895	0.918	-0.099	0.026	0.822	0.828
			$\alpha_1(u)$	1	0.01	-0.025	-2.47	0.114	0.117	-0.047	-4.686	0.142	0.15	-0.034	-3.43	0.129	0.134
				2	0.01	-0.035	-3.481	0.141	0.145	-0.053	-5.263	0.168	0.176	-0.049	-4.95	0.175	0.182
				3	0.01	-0.045	-4.499	0.184	0.19	-0.058	-5.844	0.213	0.221	-0.065	-6.475	0.232	0.241
				4	0.01	-0.055	-5.523	0.236	0.242	-0.064	-6.429	0.267	0.275	-0.08	-8.006	0.295	0.305
				5	0.01	-0.066	-6.554	0.291	0.298	-0.07	-7.018	0.326	0.334	-0.095	-9.542	0.359	0.372
			$\beta_0(u)$	1	-4.968	0.125	-0.025	0.635	0.647	-0.059	0.012	0.629	0.631	-0.048	0.01	0.756	0.757
				2	-4.638	0.116	-0.025	0.581	0.592	-0.06	0.013	0.569	0.572	-0.038	0.008	0.679	0.68
				3	-4.308	0.107	-0.025	0.542	0.553	-0.062	0.014	0.525	0.528	-0.028	0.006	0.616	0.616
				4	-3.978	0.098	-0.025	0.523	0.532	-0.063	0.016	0.499	0.503	-0.017	0.004	0.569	0.569

Table A.5: Parameter summary in Study G-3 comparing estimation by time point (TP) and penalized B-spline (PBS/Degree) under the forward intensity method.

						Train set 1				Train set 2				Train set 3			
n	Method	λ	Parameter	u	True	Bias	RB	SD	$\sqrt{MSE}$	Bias	RB	SD	$\sqrt{MSE}$	Bias	RB	SD	$\sqrt{MSE}$
				5	-3.648	0.089	-0.024	0.526	0.534	-0.064	0.018	0.495	0.499	-0.007	0.002	0.543	0.543
			$\beta_1(u)$	1	0.33	-0.014	-0.041	0.069	0.07	0.002	0.005	0.062	0.062	0.004	0.011	0.077	0.077
				2	0.33	-0.014	-0.043	0.07	0.071	0.002	0.006	0.063	0.063	0.003	0.008	0.077	0.077
				3	0.33	-0.015	-0.046	0.073	0.075	0.002	0.007	0.065	0.065	0.002	0.005	0.079	0.079
				4	0.33	-0.016	-0.048	0.079	0.08	0.003	0.008	0.07	0.07	0.001	0.002	0.083	0.083
				5	0.33	-0.017	-0.051	0.086	0.087	0.003	0.009	0.076	0.076	0	-0.001	0.089	0.089
	PBS/3	5	$\alpha_0(u)$	1	-3.902	-0.04	0.01	0.672	0.673	0.025	-0.006	0.78	0.78	-0.009	0.002	0.607	0.607
				2	-3.892	-0.064	0.017	0.627	0.631	-0.032	0.008	0.684	0.685	-0.031	0.008	0.546	0.546
				3	-3.882	-0.089	0.023	0.686	0.692	-0.089	0.023	0.674	0.68	-0.054	0.014	0.571	0.573
				4	-3.872	-0.113	0.029	0.827	0.835	-0.146	0.038	0.752	0.766	-0.076	0.02	0.672	0.677
				5	-3.862	-0.138	0.036	1.017	1.026	-0.203	0.053	0.896	0.918	-0.099	0.026	0.823	0.829
			$\alpha_1(u)$	1	0.01	-0.025	-2.468	0.114	0.117	-0.047	-4.682	0.142	0.15	-0.034	-3.429	0.129	0.134
				2	0.01	-0.035	-3.46	0.141	0.145	-0.052	-5.25	0.168	0.176	-0.049	-4.932	0.175	0.182
				3	0.01	-0.045	-4.48	0.184	0.189	-0.058	-5.834	0.213	0.221	-0.065	-6.459	0.232	0.241
				4	0.01	-0.055	-5.529	0.236	0.242	-0.064	-6.433	0.267	0.275	-0.08	-8.01	0.295	0.305
				5	0.01	-0.066	-6.605	0.292	0.299	-0.07	-7.048	0.327	0.334	-0.096	-9.584	0.36	0.372
			$\beta_0(u)$	1	-4.968	0.125	-0.025	0.635	0.647	-0.059	0.012	0.629	0.632	-0.049	0.01	0.756	0.758
				2	-4.638	0.116	-0.025	0.58	0.592	-0.061	0.013	0.569	0.572	-0.038	0.008	0.679	0.68
				3	-4.308	0.107	-0.025	0.542	0.553	-0.062	0.014	0.525	0.528	-0.028	0.006	0.616	0.616
				4	-3.978	0.098	-0.025	0.525	0.534	-0.063	0.016	0.5	0.504	-0.017	0.004	0.57	0.57
				5	-3.648	0.089	-0.024	0.531	0.538	-0.065	0.018	0.498	0.502	-0.007	0.002	0.546	0.546
			$\beta_1(u)$	1	0.33	-0.014	-0.041	0.069	0.07	0.002	0.005	0.062	0.062	0.004	0.011	0.077	0.077
				2	0.33	-0.014	-0.043	0.07	0.071	0.002	0.006	0.063	0.063	0.003	0.009	0.077	0.077
				3	0.33	-0.015	-0.045	0.073	0.075	0.002	0.007	0.065	0.065	0.002	0.006	0.079	0.079
				4	0.33	-0.016	-0.048	0.079	0.08	0.003	0.008	0.07	0.07	0.001	0.002	0.083	0.083
				5	0.33	-0.017	-0.051	0.087	0.089	0.003	0.01	0.077	0.077	0	-0.001	0.09	0.09
90	TP	NA	$\alpha_0(u)$	1	-3.902	-0.008	0.002	0.515	0.515	-0.026	0.007	0.476	0.476	-0.076	0.019	0.528	0.533
				2	-3.892	-0.065	0.017	0.708	0.711	-0.023	0.006	0.477	0.478	-0.032	0.008	0.541	0.542
				3	-3.882	-0.118	0.03	0.791	0.8	-0.078	0.02	0.535	0.54	-0.06	0.015	0.531	0.534
				4	-3.872	-0.113	0.029	0.858	0.865	-0.099	0.026	0.65	0.657	-0.019	0.005	0.585	0.585
				5	-3.862	-0.085	0.022	0.839	0.843	-0.061	0.016	0.662	0.665	-0.026	0.007	0.721	0.721
			$\alpha_1(u)$	1	0.01	-0.018	-1.775	0.083	0.085	-0.019	-1.946	0.079	0.082	-0.014	-1.371	0.09	0.091
				2	0.01	-0.014	-1.392	0.097	0.098	-0.025	-2.456	0.094	0.097	-0.024	-2.357	0.101	0.103
				3	0.01	-0.011	-1.129	0.123	0.123	-0.022	-2.19	0.112	0.114	-0.025	-2.513	0.117	0.12

Table A.5: Parameter summary in Study G-3 comparing estimation by time point (TP) and penalized B-spline (PBS/Degree) under the forward intensity method.

						Train set 1				Train set 2				Train set 3			
n	Method	λ	Parameter	u	True	Bias	RB	SD	$\sqrt{MSE}$	Bias	RB	SD	$\sqrt{MSE}$	Bias	RB	SD	$\sqrt{MSE}$
				4	0.01	-0.02	-2.004	0.156	0.157	-0.026	-2.555	0.13	0.133	-0.04	-3.964	0.133	0.139
				5	0.01	-0.034	-3.358	0.173	0.177	-0.046	-4.599	0.171	0.177	-0.058	-5.821	0.196	0.204
			$\beta_0(u)$	1	-4.968	0.131	-0.026	0.432	0.452	0.104	-0.021	0.431	0.443	0.126	-0.025	0.498	0.513
				2	-4.638	0.146	-0.031	0.413	0.438	0.121	-0.026	0.414	0.431	0.145	-0.031	0.466	0.487
				3	-4.308	0.136	-0.032	0.387	0.411	0.105	-0.024	0.368	0.383	0.134	-0.031	0.417	0.438
				4	-3.978	0.147	-0.037	0.362	0.391	0.121	-0.03	0.344	0.365	0.141	-0.035	0.401	0.425
				5	-3.648	0.147	-0.04	0.345	0.375	0.128	-0.035	0.338	0.362	0.132	-0.036	0.385	0.407
			$\beta_1(u)$	1	0.33	-0.017	-0.052	0.045	0.048	-0.017	-0.051	0.043	0.046	-0.016	-0.049	0.05	0.053
				2	0.33	-0.021	-0.062	0.047	0.052	-0.021	-0.062	0.046	0.05	-0.02	-0.06	0.052	0.056
				3	0.33	-0.022	-0.066	0.049	0.054	-0.021	-0.064	0.045	0.05	-0.021	-0.064	0.052	0.056
				4	0.33	-0.026	-0.079	0.051	0.058	-0.026	-0.079	0.047	0.054	-0.025	-0.075	0.055	0.06
				5	0.33	-0.03	-0.091	0.054	0.062	-0.031	-0.094	0.051	0.06	-0.027	-0.082	0.059	0.065
	PBS/2	5	$\alpha_0(u)$	1	-3.902	-0.016	0.004	0.563	0.563	-0.018	0.005	0.487	0.487	-0.06	0.015	0.544	0.547
				2	-3.892	-0.039	0.01	0.574	0.576	-0.03	0.008	0.432	0.433	-0.045	0.012	0.491	0.493
				3	-3.882	-0.062	0.016	0.645	0.648	-0.042	0.011	0.442	0.444	-0.03	0.008	0.486	0.487
				4	-3.872	-0.084	0.022	0.759	0.764	-0.054	0.014	0.513	0.515	-0.015	0.004	0.53	0.53
				5	-3.862	-0.107	0.028	0.9	0.907	-0.066	0.017	0.623	0.627	-0.001	0	0.611	0.611
			$\alpha_1(u)$	1	0.01	-0.016	-1.593	0.083	0.085	-0.02	-1.973	0.083	0.085	-0.014	-1.43	0.091	0.092
				2	0.01	-0.018	-1.804	0.092	0.094	-0.024	-2.4	0.087	0.091	-0.023	-2.315	0.098	0.101
				3	0.01	-0.02	-2.024	0.112	0.114	-0.028	-2.834	0.102	0.106	-0.032	-3.214	0.113	0.118
				4	0.01	-0.023	-2.251	0.139	0.141	-0.033	-3.275	0.124	0.128	-0.041	-4.124	0.133	0.14
				5	0.01	-0.025	-2.487	0.169	0.171	-0.037	-3.723	0.15	0.154	-0.05	-5.047	0.157	0.165
			$\beta_0(u)$	1	-4.968	0.138	-0.028	0.443	0.464	0.109	-0.022	0.438	0.452	0.137	-0.028	0.505	0.523
				2	-4.638	0.141	-0.03	0.401	0.425	0.113	-0.024	0.395	0.411	0.137	-0.03	0.453	0.473
				3	-4.308	0.143	-0.033	0.369	0.395	0.117	-0.027	0.361	0.379	0.137	-0.032	0.412	0.434
				4	-3.978	0.145	-0.036	0.349	0.378	0.121	-0.03	0.337	0.358	0.137	-0.034	0.385	0.408
				5	-3.648	0.147	-0.04	0.343	0.373	0.125	-0.034	0.328	0.35	0.136	-0.037	0.375	0.399
			$\beta_1(u)$	1	0.33	-0.018	-0.053	0.046	0.049	-0.017	-0.052	0.044	0.047	-0.017	-0.052	0.051	0.054
				2	0.33	-0.02	-0.061	0.046	0.05	-0.02	-0.061	0.044	0.048	-0.019	-0.058	0.051	0.054
				3	0.33	-0.023	-0.07	0.047	0.052	-0.023	-0.07	0.044	0.05	-0.022	-0.066	0.051	0.056
				4	0.33	-0.026	-0.079	0.05	0.056	-0.026	-0.079	0.046	0.053	-0.024	-0.074	0.053	0.059
				5	0.33	-0.029	-0.089	0.053	0.061	-0.03	-0.09	0.049	0.057	-0.027	-0.082	0.057	0.063
	PBS/3	3	$\alpha_0(u)$	1	-3.902	-0.017	0.004	0.563	0.563	-0.019	0.005	0.487	0.487	-0.061	0.016	0.544	0.548
				2	-3.892	-0.039	0.01	0.574	0.576	-0.03	0.008	0.432	0.433	-0.046	0.012	0.491	0.494

Table A.5: Parameter summary in Study G-3 comparing estimation by time point (TP) and penalized B-spline (PBS/Degree) under the forward intensity method.

n	Method	λ	Parameter	u	True	Train set 1				Train set 2				Train set 3			
						Bias	RB	SD	$\sqrt{MSE}$	Bias	RB	SD	$\sqrt{MSE}$	Bias	RB	SD	$\sqrt{MSE}$
				3	-3.882	-0.062	0.016	0.646	0.649	-0.042	0.011	0.442	0.444	-0.03	0.008	0.486	0.487
				4	-3.872	-0.084	0.022	0.761	0.765	-0.053	0.014	0.513	0.516	-0.015	0.004	0.531	0.531
				5	-3.862	-0.107	0.028	0.902	0.908	-0.065	0.017	0.625	0.628	0	0	0.613	0.613
			$\alpha_1(u)$	1	0.01	-0.016	-1.596	0.083	0.085	-0.02	-1.976	0.083	0.085	-0.014	-1.442	0.091	0.092
				2	0.01	-0.018	-1.767	0.092	0.093	-0.024	-2.366	0.087	0.09	-0.023	-2.264	0.098	0.101
				3	0.01	-0.02	-1.989	0.111	0.113	-0.028	-2.802	0.101	0.105	-0.032	-3.161	0.113	0.117
				4	0.01	-0.023	-2.26	0.139	0.141	-0.033	-3.283	0.124	0.128	-0.041	-4.131	0.134	0.14
			$\beta_0(u)$	5	0.01	-0.026	-2.58	0.171	0.173	-0.038	-3.809	0.152	0.157	-0.052	-5.174	0.159	0.167
				1	-4.968	0.136	-0.027	0.444	0.464	0.107	-0.022	0.439	0.452	0.135	-0.027	0.506	0.524
				2	-4.638	0.139	-0.03	0.401	0.425	0.112	-0.024	0.396	0.411	0.136	-0.029	0.454	0.474
				3	-4.308	0.143	-0.033	0.369	0.395	0.117	-0.027	0.361	0.379	0.137	-0.032	0.412	0.434
				4	-3.978	0.146	-0.037	0.349	0.379	0.122	-0.031	0.338	0.359	0.137	-0.035	0.386	0.41
				5	-3.648	0.149	-0.041	0.345	0.376	0.126	-0.035	0.33	0.353	0.138	-0.038	0.377	0.402
			$\beta_1(u)$	1	0.33	-0.018	-0.053	0.046	0.049	-0.017	-0.052	0.044	0.047	-0.017	-0.052	0.051	0.054
				2	0.33	-0.02	-0.06	0.046	0.05	-0.02	-0.06	0.044	0.048	-0.019	-0.058	0.051	0.054
				3	0.33	-0.023	-0.069	0.047	0.052	-0.023	-0.069	0.044	0.05	-0.021	-0.065	0.051	0.055
				4	0.33	-0.026	-0.079	0.05	0.056	-0.026	-0.08	0.046	0.053	-0.024	-0.074	0.053	0.059
				5	0.33	-0.03	-0.091	0.054	0.062	-0.03	-0.092	0.05	0.058	-0.028	-0.084	0.058	0.064

Table A.6: Average AUC (% Mean/SD) in Study G-3 comparing estimation by time point (TP) and penalized B-spline (PBS/Degree) under the forward intensity method.

			Test set 1		Test set 2		Test set 3		Average	
n	Method	λ	Mean	SD	Mean	SD	Mean	SD	Mean	SD
45	TP	NA	69.06	27.38	68.29	25.82	72.93	21.52	70.09	24.9
	PBS/2	0	69.38	27.12	68.51	25.7	73.07	21.34	70.32	24.72
		1	69.61	26.91	68.62	25.54	73.24	21.48	70.49	24.64
		2	69.6	26.92	68.63	25.54	73.25	21.44	70.49	24.63
		3	69.63	26.88	68.63	25.55	73.26	21.44	70.5	24.62
		4	69.63	26.88	68.63	25.55	73.26	21.44	70.5	24.62
		5	69.63	26.88	68.63	25.55	73.26	21.44	70.5	24.62
	PBS/3	0	69.2	27.21	68.33	25.77	72.94	21.43	70.16	24.8
		1	69.6	26.91	68.62	25.54	73.23	21.47	70.48	24.64
		2	69.61	26.93	68.62	25.54	73.22	21.48	70.48	24.65
		3	69.6	26.92	68.64	25.53	73.24	21.49	70.49	24.64
		4	69.6	26.92	68.62	25.53	73.24	21.48	70.49	24.64
		5	69.61	26.91	68.62	25.54	73.25	21.48	70.49	24.64
90	TP	NA	71.1	19.85	73.57	18.69	69.63	21.64	71.43	20.06
	PBS/2	0	71.1	19.78	73.71	18.63	69.82	21.47	71.54	19.96
		1	71.18	19.49	73.89	18.4	69.79	21.38	71.62	19.76
		2	71.21	19.48	73.87	18.4	69.8	21.39	71.62	19.76
		3	71.21	19.48	73.87	18.41	69.79	21.38	71.62	19.76
		4	71.21	19.48	73.87	18.41	69.79	21.38	71.62	19.75
		5	71.21	19.48	73.87	18.41	69.8	21.38	71.62	19.75
	PBS/3	0	71.15	19.83	73.57	18.67	69.7	21.55	71.47	20.02
		1	71.19	19.48	73.92	18.38	69.82	21.37	71.64	19.74
		2	71.18	19.49	73.92	18.4	69.81	21.37	71.64	19.75
		3	71.18	19.49	73.92	18.38	69.8	21.37	71.64	19.75
		4	71.18	19.49	73.9	18.39	69.8	21.38	71.62	19.75
		5	71.19	19.49	73.89	18.4	69.79	21.38	71.62	19.76

Table A.7: AUC (% Mean/SD) at each prediction horizon (u) in Study G-3 comparing estimation by time point (TP) and penalized B-spline (PBS/Degree) under the forward intensity method.

				Test Set 1 (u)					Test Set 2 (u)					Test Set 3 (u)				
n	Method	λ	Statistics	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
45	TP	NA	mean	59.29	65.96	70.54	73.41	76.10	60.10	64.68	68.53	72.59	75.53	64.41	69.89	73.50	77.21	79.62
			sd	32.00	29.13	27.23	25.13	23.38	32.12	28.53	24.88	22.53	21.03	27.89	24.42	19.68	18.64	16.96
	PBS/2	5	Mean	59.84	67.24	70.79	73.78	76.49	60.20	65.06	69.38	73.06	75.45	65.28	70.03	73.48	77.69	79.80
			SD	31.42	28.56	27.30	24.56	22.56	31.24	28.76	25.03	21.78	20.93	27.77	23.69	20.49	18.67	16.59
	PBS/3	5	Mean	59.86	67.24	70.79	73.78	76.37	60.17	64.95	69.51	73.02	75.45	65.25	70.03	73.53	77.69	79.74
			SD	31.37	28.56	27.30	24.56	22.77	31.22	28.93	24.83	21.77	20.93	27.83	23.69	20.50	18.67	16.71
90	TP	NA	mean	63.32	68.03	72.62	74.88	76.63	66.37	72.05	74.58	76.53	78.31	60.77	66.39	70.43	74.09	76.48
			SD	27.02	22.10	18.61	16.64	14.90	24.52	21.00	18.51	16.00	13.41	28.89	24.43	20.67	18.06	16.14
	PBS/2	5	Mean	63.07	68.16	72.65	75.25	76.90	67.15	72.40	74.78	76.83	78.20	60.65	66.60	70.91	74.25	76.56
			SD	26.62	22.00	18.35	16.05	14.37	23.88	20.99	18.10	15.40	13.67	28.49	24.22	20.39	17.87	15.92
	PBS/3	3	Mean	62.93	68.16	72.65	75.24	76.92	67.32	72.38	74.86	76.87	78.20	60.70	66.59	70.93	74.24	76.55
			SD	26.67	22.00	18.35	16.05	14.38	23.88	20.96	18.09	15.33	13.66	28.51	24.22	20.34	17.88	15.92

Study E-0

We generate event time T which follows the exponential distribution with mean of 30. The conditional forward survival function is $S_{T,t}(u) = \mathbb{P}(T > t + u|T > t) = \exp(-\lambda u)$ and the conditional forward intensity function is constant $\lambda_{T,t}(u) = \exp\{\alpha_0(u)\} = \lambda$ with $\alpha_0(u) = \log(\lambda) = \log(1/30) = -3.401$. Censor time C follows the exponential distribution with mean of 40. For a given individual, the observation time $X = \min(T, C)$ with event indicator $\delta_T = 1$ if $X = T$ and $\delta_T = 0$ if otherwise.

Table A.10 summarizes the accuracy of parameter estimates in terms of their biases and variances. The RB and SD are small and stable over different prediction horizons $u = 1, 2, \ldots, 10$. Both RB and SD improve with larger number of subjects in the simulations.

Table A.8: Simulation results in Study G-1 when the sample size $n = 100$ and 300.

			$n=100$				$n=300$			
	u	True	Bias	RB	SD	$\sqrt{MSE}$	Bias	RB	SD	$\sqrt{MSE}$
$\alpha_0(u)$	1	-5.198	-0.062	0.012	0.287	0.294	-0.031	0.006	0.19	0.193
	2	-5.098	-0.044	0.009	0.284	0.287	-0.029	0.006	0.195	0.197
	3	-4.998	-0.042	0.008	0.283	0.286	-0.033	0.007	0.201	0.204
	4	-4.898	-0.054	0.011	0.267	0.272	-0.032	0.006	0.196	0.199
	5	-4.798	-0.064	0.013	0.269	0.276	-0.028	0.006	0.192	0.194
	6	-4.698	-0.070	0.015	0.262	0.272	-0.032	0.007	0.196	0.199
	7	-4.598	-0.055	0.012	0.274	0.279	-0.025	0.005	0.186	0.188
	8	-4.498	-0.065	0.014	0.265	0.273	-0.028	0.006	0.179	0.181
	9	-4.398	-0.068	0.016	0.268	0.277	-0.036	0.008	0.165	0.169
	10	-4.298	-0.079	0.018	0.274	0.285	-0.037	0.009	0.167	0.171
$\alpha_1(u)$	1	0.1	0.003	0.030	0.013	0.013	0.001	0.012	0.008	0.008
	2	0.1	0.002	0.024	0.013	0.013	0.001	0.012	0.008	0.009
	3	0.1	0.002	0.024	0.013	0.013	0.001	0.014	0.009	0.009
	4	0.1	0.003	0.030	0.013	0.014	0.001	0.014	0.009	0.009
	5	0.1	0.004	0.036	0.014	0.014	0.001	0.013	0.009	0.009
	6	0.1	0.004	0.041	0.015	0.015	0.002	0.015	0.01	0.01
	7	0.1	0.004	0.036	0.015	0.016	0.001	0.013	0.01	0.01
	8	0.1	0.004	0.042	0.015	0.016	0.001	0.015	0.01	0.01
	9	0.1	0.005	0.046	0.016	0.016	0.002	0.019	0.009	0.009
	10	0.1	0.005	0.053	0.017	0.018	0.002	0.021	0.009	0.01

NOTE: "True" is the true parameter value for $\alpha_0(u)$ and $\alpha_1(u)$ for different prediction horizons u; "Bias" is the difference between the parameter estimate and the true parameter value; RB is the bias divided by the true parameter value; "SD" is the standard deviation; "$\sqrt{MSE}$" is the square root of the mean square error.

Table A.9: Simulation results in Study G-2 when the sample size n = 100 and 300.

	u	True	n = 100 Bias	RB	SD	$\sqrt{MSE}$	n = 300 Bias	RB	SD	$\sqrt{MSE}$
$\alpha_0(u)$	1	-3.912	-0.054	0.014	0.2	0.207	-0.018	0.005	0.128	0.129
	2	-3.912	-0.056	0.014	0.209	0.217	-0.018	0.005	0.13	0.131
	3	-3.912	-0.054	0.014	0.21	0.217	-0.023	0.006	0.137	0.139
	4	-3.912	-0.065	0.017	0.222	0.231	-0.024	0.006	0.139	0.141
	5	-3.912	-0.057	0.015	0.236	0.243	-0.027	0.007	0.143	0.146
	6	-3.912	-0.053	0.014	0.237	0.243	-0.026	0.007	0.146	0.148
	7	-3.912	-0.059	0.015	0.244	0.251	-0.023	0.006	0.148	0.15
	8	-3.912	-0.066	0.017	0.243	0.252	-0.022	0.006	0.151	0.153
	9	-3.912	-0.068	0.017	0.241	0.251	-0.016	0.004	0.15	0.151
	10	-3.912	-0.06	0.015	0.245	0.253	-0.01	0.002	0.157	0.158
$\alpha_1(u)$	1	0.3	-0.191	-0.638	0.935	0.955	-0.062	-0.207	0.468	0.472
	2	0.6	-0.123	-0.204	0.94	0.948	-0.049	-0.082	0.457	0.46
	3	0.9	-0.14	-0.155	0.882	0.893	-0.059	-0.065	0.44	0.444
	4	1.2	-0.078	-0.065	0.91	0.913	-0.052	-0.043	0.43	0.433
	5	1.5	-0.073	-0.048	0.915	0.918	-0.083	-0.055	0.439	0.447
	6	1.8	-0.058	-0.032	0.966	0.967	-0.053	-0.029	0.448	0.451
	7	2.1	-0.024	-0.012	0.998	0.998	-0.056	-0.027	0.468	0.471
	8	2.4	-0.048	-0.02	0.981	0.982	-0.06	-0.025	0.486	0.49
	9	2.7	0.016	0.006	0.879	0.879	-0.037	-0.014	0.515	0.516
	10	3	0.079	0.026	0.915	0.918	-0.038	-0.013	0.576	0.578
$\alpha_2(u)$	1	0.3	0.033	0.11	0.076	0.083	0.008	0.026	0.035	0.036
	2	0.3	0.03	0.101	0.082	0.087	0.007	0.025	0.034	0.035
	3	0.3	0.034	0.114	0.08	0.087	0.01	0.032	0.035	0.036
	4	0.3	0.034	0.113	0.085	0.091	0.01	0.034	0.036	0.038
	5	0.3	0.035	0.115	0.087	0.094	0.015	0.049	0.037	0.04
	6	0.3	0.036	0.121	0.1	0.107	0.013	0.044	0.041	0.043
	7	0.3	0.038	0.125	0.111	0.117	0.014	0.047	0.046	0.048
	8	0.3	0.047	0.155	0.123	0.131	0.016	0.055	0.054	0.056
	9	0.3	0.045	0.15	0.123	0.131	0.014	0.047	0.062	0.064
	10	0.3	0.04	0.135	0.132	0.138	0.014	0.046	0.072	0.073

NOTE: "True" is the true parameter value for $\alpha_0(u)$ and $\alpha_1(u)$ for different prediction horizons u; "Bias" is the difference between the parameter estimate and the true parameter value; RB is the bias divided by the true parameter value; "SD" is the standard deviation; "$\sqrt{MSE}$" is the square root of the mean square error.

Table A.10: Simulation results in Study Study E-0 when the sample size n = 100 and 300.

			n = 100				n = 300			
	u	True	Bias	RB	SD	$\sqrt{MSE}$	Bias	RB	SD	$\sqrt{MSE}$
$\alpha(u)$	1	-3.401	-0.03	0.009	0.149	0.152	-0.021	0.006	0.073	0.076
	2	-3.401	-0.028	0.008	0.146	0.149	-0.022	0.006	0.075	0.078
	3	-3.401	-0.024	0.007	0.152	0.154	-0.018	0.005	0.079	0.081
	4	-3.401	-0.026	0.008	0.164	0.166	-0.016	0.005	0.082	0.084
	5	-3.401	-0.029	0.009	0.164	0.166	-0.012	0.003	0.086	0.087
	6	-3.401	-0.03	0.009	0.166	0.169	-0.012	0.003	0.086	0.087
	7	-3.401	-0.027	0.008	0.165	0.167	-0.012	0.004	0.087	0.088
	8	-3.401	-0.03	0.009	0.162	0.165	-0.012	0.003	0.084	0.084
	9	-3.401	-0.027	0.008	0.164	0.166	-0.013	0.004	0.089	0.09
	10	-3.401	-0.029	0.008	0.162	0.165	-0.016	0.005	0.088	0.089

NOTE: "True" is the true parameter value for $\alpha_0(u)$ and $\alpha_1(u)$ for different prediction horizons u; "Bias" is the difference between the parameter estimate and the true parameter value; RB is the bias divided by the true parameter value; "SD" is the standard deviation; "$\sqrt{MSE}$" is the square root of the mean square error.

APPENDIX B

PROOFS

B.1 Proof of Theorem 1

Proof: When $\Delta \to 0$, the continuous counterpart of the discrete-time composite likelihood (3.6.2) is

$$\begin{aligned}\frac{1}{n}\ell^{(1)}\{\boldsymbol{\alpha}(u)\} &= \frac{1}{n}\sum_{i=1}^{n}\left[-\int_{t>u}\lambda_{T_i,t-u}(u)dt + \delta_{T_i}(X_i)\log\{\lambda_{T_i,(X_i-u)}(u)\}\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[-\int_{t>u}\lambda_{T_i,t-u}(u)dt + \int_{t>u}dN_{T_i}(t)\log\{\lambda_{T_i,(t-u)}(u)\}\right]\end{aligned}$$

By the Taylor expansion of the score function around the true value $\boldsymbol{\alpha}_0^{(1)}(u) \in \mathcal{A}_0(u)$, we have

$$0 = \frac{1}{n}\frac{\partial\ell^{(1)}\{\hat{\boldsymbol{\alpha}}^{(1)}(u)\}}{\partial\boldsymbol{\alpha}^T(u)} = \frac{1}{n}\frac{\partial\ell^{(1)}\{\boldsymbol{\alpha}_0^{(1)}(u)\}}{\partial\boldsymbol{\alpha}^T(u)} + \frac{1}{n}\frac{\partial^2\ell^{(1)}\{\boldsymbol{\alpha}(u)\}}{\partial\boldsymbol{\alpha}^T(u)\boldsymbol{\alpha}(u)}|_{\boldsymbol{\alpha}(u)=\tilde{\boldsymbol{\alpha}}^{(1)}(u)}\{\hat{\boldsymbol{\alpha}}^{(1)}(u) - \boldsymbol{\alpha}_0^{(1)}(u)\}$$

where $\tilde{\boldsymbol{\alpha}}^{(1)}(u) \in (\hat{\boldsymbol{\alpha}}^{(1)}(u), \boldsymbol{\alpha}_0^{(1)}(u))$. Therefore in order to prove the existence of a consistent solution of the score function $\frac{1}{n}\frac{\partial\ell^{(1)}\{\boldsymbol{\alpha}(u)\}}{\partial\boldsymbol{\alpha}^T(u)} = 0$, we need to show both

$$\frac{1}{n}\frac{\partial\ell^{(1)}\{\boldsymbol{\alpha}_0^{(1)}(u)\}}{\partial\boldsymbol{\alpha}(u)^T} \xrightarrow{p} 0 \tag{B.1.1}$$

and

$$\boldsymbol{\Omega}_{1,n}\{\boldsymbol{\alpha}_0^{(1)}(u)\} = -\frac{1}{n}\frac{\partial^2\ell^{(1)}\{\boldsymbol{\alpha}_0^{(1)}(u)\}}{\partial\boldsymbol{\alpha}^T(u)\partial\boldsymbol{\alpha}(u)} \xrightarrow{p} \int_{t>u}^{\tau}\boldsymbol{Z}_1(t-u)\boldsymbol{Z}_1^T(t-u)E_{\boldsymbol{\alpha}_0^{(1)}(u)}[\lambda_{T_1,t-u}\{\boldsymbol{\alpha}(u)\}]dt \tag{B.1.2}$$

The score function at the true parameter $\boldsymbol{\alpha}_0^{(1)}(u)$ takes the form,

$$\begin{aligned}\frac{\partial \ell^{(1)}\{\boldsymbol{\alpha}_0^{(1)}(u)\}}{\partial \boldsymbol{\alpha}(u)^T} &= \sum_{i=1}^{n}\left[-\int_{t>u}^{\tau}\frac{\partial \lambda_{T_i,t-u}\{\boldsymbol{\alpha}_0^{(1)}(u)\}}{\partial \boldsymbol{\alpha}(u)^T}dt + \delta_{T_i}(X_i)\frac{\partial \log[\lambda_{T_i,(X_i-u)}\{\boldsymbol{\alpha}_0^{(1)}(u)\}]}{\partial \boldsymbol{\alpha}(u)^T}\right] \\ &= \sum_{i=1}^{n}\left[\int_{t>u}^{\tau}\boldsymbol{Z}_i(t-u)[dN_{T_i}(t)-\lambda_{T_i,t-u}\{\boldsymbol{\alpha}_0^{(1)}(u)\}dt]\right] \\ &= \sum_{i=1}^{n}\left[\int_{t>u}^{\tau}\boldsymbol{Z}_i(t-u)\{dM_{T_i}(t)\}\right]\end{aligned}$$

where $dM_{T_i}(t) = dN_{T_i}(t) - \lambda_{T_i,t-u}\{\boldsymbol{\alpha}_0^{(1)}(u)\}dt$ is the orthogonal local square integrable martingale. We have $E_{\boldsymbol{\alpha}_0^{(1)}(u)}[\frac{\partial \ell^{(1)}\{\boldsymbol{\alpha}(u)\}}{\partial \boldsymbol{\alpha}(u)^T}] = 0$ because $E_{\boldsymbol{\alpha}_0^{(1)}(u)}\{dN_{T_i}(t) - \lambda_{T_i,t-u}(u)dt\} = E_{\boldsymbol{\alpha}_0^{(1)}(u)}\{dM_{T_i}(t)\} = 0$ according to the martingale property. In addition the predictable variation process is

$$\boldsymbol{\Omega}_{1,n}\{\boldsymbol{\alpha}_0^{(1)}(u)\} = \langle\frac{1}{\sqrt{n}}\frac{\partial \ell^{(1)}\{\boldsymbol{\alpha}_0^{(1)}(u)\}}{\partial \boldsymbol{\alpha}(u)^T}\rangle = \frac{1}{n}\sum_{i=1}^{n}\left[\int_{t>u}^{\tau}\boldsymbol{Z}_i(t-u)\boldsymbol{Z}_i^T(t-u)\lambda_{T_i,t-u}\{\boldsymbol{\alpha}_0^{(1)}(u)\}dt\right]$$

Hence by condition (1B) as $n\to\infty$

$$\boldsymbol{\Omega}_{1,n}\{\boldsymbol{\alpha}_0^{(1)}(u)\} = \langle\frac{1}{\sqrt{n}}\frac{\partial \ell^{(1)}\{\boldsymbol{\alpha}_0^{(1)}(u)\}}{\partial \boldsymbol{\alpha}(u)^T}\rangle \xrightarrow{p} \int_{t>u}^{\tau}\boldsymbol{Z}_1(t-u)\boldsymbol{Z}_1^T(t-u)E_{\boldsymbol{\alpha}_0^{(1)}(u)}[\lambda_{T_1,t-u}\{\boldsymbol{\alpha}(u)\}]dt$$

By the inequality of Lenglart (1977), for any $\eta > 0$ and $\delta > 0$, $P(\sup_{\Gamma}|M| > \eta) \le \frac{\delta}{\eta^2} + P(\langle M\rangle(\tau) > \delta)$. That is if $\langle M\rangle(\tau)$ the prediction variation process for a integrable martingale M at τ is small, M is small in absolute value throughout Γ.

Therefore

$$P(\sup_{t\in\mathcal{T}}|\frac{1}{n}\frac{\partial \ell^{(1)}\{\boldsymbol{\alpha}_0^{(1)}(u)\}}{\partial \boldsymbol{\alpha}(u)^T}| > \eta) \le \frac{\delta}{\eta^2} + P(\langle\frac{1}{n}\frac{\partial \ell^{(1)}\{\boldsymbol{\alpha}_0^{(1)}(u)\}}{\partial \boldsymbol{\alpha}(u)^T}\rangle > \delta)$$

which proves (B.1.1).

Next is to prove (B.1.2). First note that

$$\begin{aligned}&\frac{1}{n}\frac{\partial^2 \ell^{(1)}\{\boldsymbol{\alpha}_0^{(1)}(u)\}}{\partial \boldsymbol{\alpha}(u)^T\partial \boldsymbol{\alpha}(u)} \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[-\int_{t>u}^{\tau}\frac{\partial^2 \lambda_{T_i,t-u}\{\boldsymbol{\alpha}_0^{(1)}(u)\}}{\partial \boldsymbol{\alpha}(u)^T\partial \boldsymbol{\alpha}(u)}dt + \int_{t>u}^{\tau}dN_{T_i}(t+u)\frac{\partial^2 \log[\lambda_{T_i,(t-u)}\{\boldsymbol{\alpha}_0^{(1)}(u)\}]}{\partial \boldsymbol{\alpha}(u)^T\partial \boldsymbol{\alpha}(u)}\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[-\int_{t>u}^{\tau}\boldsymbol{Z}_i(t-u)\boldsymbol{Z}_i^T(t-u)\lambda_{T_i,t-u}\{\boldsymbol{\alpha}_0^{(1)}(u)\}dt\right] \\ &= -\boldsymbol{\Omega}_{1,n}\{\boldsymbol{\alpha}_0^{(1)}(u)\}\end{aligned}$$

in which $\frac{\partial^2 \log[\lambda_{T_i,(t-u)}\{\boldsymbol{\alpha}_0^{(1)}(u)\}]}{\partial\boldsymbol{\alpha}(u)^T\partial\boldsymbol{\alpha}(u)} = 0$. By condition (1B) and the law of large number, (B.1.2) is proved. Hence there exists a consistent solution $\hat{\boldsymbol{\alpha}}^{(1)}(u)$ for the equation $\frac{1}{n}\frac{\partial\ell^{(1)}\{\boldsymbol{\alpha}(u)\}}{\partial\boldsymbol{\alpha}^T(u)} = 0$.

Since the second order derivative of the log-likelihood $\frac{\partial^2\ell^{(1)}\{\boldsymbol{\alpha}(u)\}}{\partial\boldsymbol{\alpha}(u)^T\partial\boldsymbol{\alpha}(u)}$ is negative, $\hat{\boldsymbol{\alpha}}^{(1)}(u)$ is a local maximum. Hence we have the consistency of $\hat{\boldsymbol{\alpha}}^{(1)}(u)$.

Next we will show the asymptotic normality of $\hat{\boldsymbol{\alpha}}^{(1)}(u)$. Note that for any random $\tilde{\boldsymbol{\alpha}}^{(1)}(u)$ that $\tilde{\boldsymbol{\alpha}}^{(1)}(u) \xrightarrow{p} \boldsymbol{\alpha}_0(u)$

$$\sqrt{n}\{\hat{\boldsymbol{\alpha}}^{(1)}(u) - \boldsymbol{\alpha}_0^{(1)}(u)\} = -[\frac{1}{n}\frac{\partial^2\ell^{(1)}\{\tilde{\boldsymbol{\alpha}}^{(1)}(u)\}}{\partial\boldsymbol{\alpha}^T(u)\boldsymbol{\alpha}(u)}]^{-1}\sqrt{n}\frac{1}{n}\frac{\partial\ell^{(1)}\{\boldsymbol{\alpha}_0^{(1)}(u)\}}{\partial\boldsymbol{\alpha}^T(u)}$$

To prove the asymptotic distribution of $\hat{\boldsymbol{\alpha}}^{(1)}(u)$, we need to show both

$$[\boldsymbol{\Omega}_{1,n}\{\hat{\boldsymbol{\alpha}}^{(1)}(u)\}]^{-1/2}\sqrt{n}\frac{1}{n}\frac{\partial\ell^{(1)}\{\boldsymbol{\alpha}_0^{(1)}(u)\}}{\partial\boldsymbol{\alpha}^T(u)} \xrightarrow{d} N(0,\mathbf{I}) \tag{B.1.3}$$

and

$$-\frac{1}{n}\frac{\partial^2\ell^{(1)}\{\tilde{\boldsymbol{\alpha}}^{(1)}(u)\}}{\partial\boldsymbol{\alpha}(u)^T\partial\boldsymbol{\alpha}(u)} = \boldsymbol{\Omega}_{1,n}\{\tilde{\boldsymbol{\alpha}}^{(1)}(u)\} \xrightarrow{p} \int_{t>u}^{\tau}\boldsymbol{Z}_1(t-u)\boldsymbol{Z}_1^T(t-u)E_{\boldsymbol{\alpha}_0^{(1)}(u)}[\lambda_{T_1,t-u}\{\boldsymbol{\alpha}(u)\}]dt \tag{B.1.4}$$

as $n \to \infty$.

As shown before $E_{\boldsymbol{\alpha}_0^{(1)}(u)}[\frac{\partial\ell^{(1)}\{\boldsymbol{\alpha}(u)\}}{\partial\boldsymbol{\alpha}(u)^T}] = 0$. Also as shown before we have $\langle\frac{1}{\sqrt{n}}\frac{\partial\ell^{(1)}\{\boldsymbol{\alpha}_0^{(1)}(u)\}}{\partial\boldsymbol{\alpha}(u)^T}\rangle \xrightarrow{p} \int_{t>u}^{\tau}\boldsymbol{Z}_1(t-u)\boldsymbol{Z}_1^T(t-u)E_{\boldsymbol{\alpha}_0^{(1)}(u)}[\lambda_{T_1,t-u}\{\boldsymbol{\alpha}(u)\}]dt$ as $n \to \infty$.

By the martingale central limit theorem and the condition (1C), we have (B.1.3) proved.

For the proof of (B.1.4), by Taylor Expansion

$$-\frac{1}{n}\frac{\partial^2\ell^{(1)}\{\tilde{\boldsymbol{\alpha}}^{(1)}(u)\}}{\partial\boldsymbol{\alpha}(u)^T\partial\boldsymbol{\alpha}(u)} = -\frac{1}{n}\frac{\partial^2\ell^{(1)}\{\boldsymbol{\alpha}_0^{(1)}(u)\}}{\partial\boldsymbol{\alpha}(u)^T\partial\boldsymbol{\alpha}(u)} - \frac{1}{n}\frac{\partial^3\ell^{(1)}\{\check{\boldsymbol{\alpha}}(u)\}}{\partial\boldsymbol{\alpha}(u)^T\partial\boldsymbol{\alpha}(u)\partial\boldsymbol{\alpha}(u)^T}\{\tilde{\boldsymbol{\alpha}}^{(1)}(u) - \boldsymbol{\alpha}_0^{(1)}(u)\} \tag{B.1.5}$$

where $\check{\boldsymbol{\alpha}}(u) \in (\tilde{\boldsymbol{\alpha}}^{(1)}(u), \boldsymbol{\alpha}_0^{(1)}(u))$. By condition (1A), the second term in (B.1.5) is bounded in probability by $M||\tilde{\boldsymbol{\alpha}}^{(1)}(u) - \boldsymbol{\alpha}_0^{(1)}(u)||$ for some finite constant M not depending on $\check{\boldsymbol{\alpha}}(u)$. We have proved (B.1.2) earlier. Hence we have (B.1.4).

By the Slutsky Theorem, we prove the asymptotic normality of $\hat{\boldsymbol{\alpha}}^{(1)}(u)$.

B.2 Proof of Theorem 2

Proof:

We follow the method in section 4.5.2 in Shao (1999) to prove the consistency of $\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta) \xrightarrow{p} \boldsymbol{\alpha}_0^{(2)}(k\Delta)$ by showing that for the estimator $\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta)$,

$$\mathbb{P}[\frac{1}{n}\frac{\partial\ell^{(2)}\{\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta)\}}{\partial\boldsymbol{\alpha}(k\Delta)^T} = 0] \to 1 \text{ and } \hat{\boldsymbol{\alpha}}^{(2)}(k\Delta) \xrightarrow{p} \boldsymbol{\alpha}_0^{(2)}(k\Delta) \tag{B.2.1}$$

First denote

$$\boldsymbol{\Sigma}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\} = \text{var}_{\boldsymbol{\alpha}_0^{(2)}(k\Delta)}[\frac{1}{\sqrt{n}}\frac{\partial \ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T}]$$
$$= \frac{1}{n}E_{\boldsymbol{\alpha}_0^{(2)}(k\Delta)}([\frac{\partial \ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T}][\frac{\partial \ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T}]^T)$$

and

$$\mathcal{A}_n^{(2)}(c;k\Delta) = \{\boldsymbol{\alpha}(k\Delta) : ||\{\boldsymbol{\alpha}(k\Delta) - \boldsymbol{\alpha}_0^{(2)}(k\Delta)\}[\boldsymbol{\Sigma}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}]^{1/2}|| \leq c\} \text{ for } c > 0$$

Since the parameter space $\mathcal{A}_0^{(2)}(k\Delta)$ is open, for each $c > 0$, $\mathcal{A}_n^{(2)}(c;k\Delta) \subset \mathcal{A}_0^{(2)}(k\Delta)$ for sufficiently large n. Since $\mathcal{A}_n^{(2)}(c;k\Delta)$ shrinks to $\boldsymbol{\alpha}_0^{(2)}(k\Delta)$ as $n \to \infty$, the existence of $\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta)$ satisfying (B.2.1) is implied by that for any $\epsilon > 0$, there exists $c > 0$ and $n_0 > 1$ such that

$$P[\frac{1}{n}\ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\} - \frac{1}{n}\ell^{(2)}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\} < 0 \text{ for all } \boldsymbol{\alpha}(k\Delta) \in \partial\mathcal{A}_n^{(2)}(c;k\Delta)] \geq 1 - \epsilon, n \geq n_0 \tag{B.2.2}$$

where $\partial\mathcal{A}_n^{(2)}(c;k\Delta)$ is the boundary of $\mathcal{A}_n^{(2)}(c;k\Delta)$. For $\boldsymbol{\alpha}(k\Delta) \in \mathcal{A}_n^{(2)}(c;k\Delta)$, by the Taylor expansion we have,

$$\frac{1}{n}\ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\} - \frac{1}{n}\ell^{(2)}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\} = c\frac{1}{n}\frac{\partial \ell^{(2)}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T}[\boldsymbol{\Sigma}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}]^{-1/2}\lambda^T$$
$$+ \frac{c^2}{2}\lambda[\boldsymbol{\Sigma}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}]^{-1/2}[-\mathbf{W}_{2,n}\{\tilde{\boldsymbol{\alpha}}^{(2)}(k\Delta)\}][\boldsymbol{\Sigma}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}]^{-1/2}\lambda^T$$

where $\lambda = \{\boldsymbol{\alpha}(k\Delta) - \boldsymbol{\alpha}_0^{(2)}(k\Delta)\}[\boldsymbol{\Sigma}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}]^{1/2}/c$ satisfying $||\lambda|| = 1$, $-\mathbf{W}_{2,n}\{\boldsymbol{\alpha}(k\Delta)\} = \frac{1}{n}\frac{\partial^2 \ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial\{\boldsymbol{\alpha}(k\Delta)\}^T\partial\{\boldsymbol{\alpha}(k\Delta)\}}$ and $\tilde{\boldsymbol{\alpha}}^{(2)}(k\Delta) \in (\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta), \boldsymbol{\alpha}_0^{(2)}(k\Delta))$.
Note that

$$E||\mathbf{W}_{2,n}\{\tilde{\boldsymbol{\alpha}}^{(2)}(k\Delta)\} - \mathbf{W}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}||$$
$$\leq \frac{1}{n}E\max_{\boldsymbol{\alpha}(k\Delta)\in\mathcal{A}_n^{(2)}(c;k\Delta)}||\frac{\partial^2 \ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T\partial \boldsymbol{\alpha}(k\Delta)} - \frac{\partial^2 \ell^{(2)}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T\partial \boldsymbol{\alpha}(k\Delta)}||$$
$$\leq E\max_{\boldsymbol{\alpha}(k\Delta)\in\mathcal{A}_n^{(2)}(c;k\Delta)}||\frac{\partial^2 Q_1\{\boldsymbol{\alpha}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T\partial \boldsymbol{\alpha}(k\Delta)} - \frac{\partial^2 Q_1\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T\partial \boldsymbol{\alpha}(k\Delta)}||$$
$$\to 0$$

This follows from three conditions, (1) condition (2A) that the second order derivative of $\ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}$ is continuous for $\boldsymbol{\alpha}(k\Delta) \in \mathcal{A}_0^{(2)}(k\Delta)$; (2) $\mathcal{A}_n^{(2)}(c;k\Delta)$ shrinks to $\boldsymbol{\alpha}_0^{(2)}(k\Delta)$ as $n \to \infty$; and (3) when $n \to \infty$, $E\max_{\boldsymbol{\alpha}(k\Delta)\in\mathcal{A}_n^{(2)}(c;k\Delta)}||\frac{\partial^2 Q_1\{\boldsymbol{\alpha}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T\partial \boldsymbol{\alpha}(k\Delta)} - \frac{\partial^2 Q_1\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T\partial \boldsymbol{\alpha}(k\Delta)}||$ is finite.

As shown in Results 4 and 5 in the section of technical proofs when $\Delta \to 0$,

$$\begin{aligned}\boldsymbol{\Sigma}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\} &= \text{var}_{\boldsymbol{\alpha}_0^{(2)}(k\Delta)}[\frac{1}{\sqrt{n}}\frac{\partial \ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T}] \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}} \boldsymbol{Z}_i(t)\boldsymbol{Z}_i^T(t)\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}\Delta + o(\Delta)\end{aligned}$$

and

$$\begin{aligned}\mathbf{W}_{2,n}\{\boldsymbol{\alpha}^{(2)}(k\Delta)\} &= -\frac{1}{n}\frac{\partial^2 \ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial\{\boldsymbol{\alpha}(k\Delta)\}^T\partial\{\boldsymbol{\alpha}(k\Delta)\}} \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}} \lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{Z}_i(t)\boldsymbol{Z}_i^T(t)\Delta\Big[1-\delta_{T_i}(t+k\Delta+\Delta)\Big] + o(\Delta)\end{aligned}$$

Hence given $\tilde{\boldsymbol{\alpha}}^{(2)}(k\Delta) \in (\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta), \boldsymbol{\alpha}_0^{(2)}(k\Delta))$,

$$[\boldsymbol{\Sigma}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}]^{-1/2}[-\mathbf{W}_{2,n}\{\tilde{\boldsymbol{\alpha}}^{(2)}(k\Delta)\}][\boldsymbol{\Sigma}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}]^{-1/2} \approx -1$$

Hence we have

$$\frac{1}{n}\ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\} - \frac{1}{n}\ell^{(2)}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\} = c\frac{1}{n}\frac{\partial \ell^{(2)}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T}[\boldsymbol{\Sigma}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}]^{-1/2}\lambda^T - \frac{c^2}{2} + o_p(1)$$

Note that $\max_\lambda(\frac{1}{n}\frac{\partial \ell^{(2)}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T}[\boldsymbol{\Sigma}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}]^{-1/2}\lambda^T) = ||\frac{1}{n}\frac{\partial \ell^{(2)}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T}[\boldsymbol{\Sigma}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}]^{-1/2}||$ given $||\lambda||= 1$. By choosing c sufficiently large we have

$$\begin{aligned}&\mathbb{P}(||\frac{1}{n}\frac{\partial \ell^{(2)}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T}[\boldsymbol{\Sigma}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}]^{-1/2}||< \frac{c}{4}) \\ &\geq 1 - (\frac{c}{4})^2 E||\frac{1}{n}\frac{\partial \ell^{(2)}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T}[\boldsymbol{\Sigma}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}]^{-1/2}||^2 \\ &= 1 - k(\frac{c}{4})^2 \\ &\geq 1-\epsilon\end{aligned}$$

Hence we have (B.2.2) and complete the proof of (B.2.1) that $\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta) \xrightarrow{p} \boldsymbol{\alpha}_0^{(2)}(k\Delta)$. To complete the proof of the consistency $\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta) \xrightarrow{p} \boldsymbol{\alpha}_0^{(1)}(u)$, next we need to show that

$$\frac{1}{n}\frac{\partial \ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)} - \frac{1}{n}\frac{\partial \ell^{(1)}\{\boldsymbol{\alpha}(u)\}}{\partial \boldsymbol{\alpha}(u)} = o(\Delta) \tag{B.2.3}$$

This was proven in Result 3 (B.4.2) in the section of technical proofs when $\Delta \to 0$, $k\Delta = u$ and $\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}$ being constant within each increment Δ of $t \in \mathcal{T}_{i,k}$.

Also the second order derivative of the discrete composite log-likelihood takes the form,

$$
\begin{aligned}
&-\mathbf{W}_{2,n}\{\boldsymbol{\alpha}(k\Delta)\}\\
&=\frac{1}{n}\frac{\partial^2\ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial\boldsymbol{\alpha}(k\Delta)^T\partial\boldsymbol{\alpha}(k\Delta)}\\
&=-\frac{1}{n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}}\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{Z}_i(t)\boldsymbol{Z}_i^T(t)\Delta\Big[1-\delta_{T_i}(t+k\Delta+\Delta)\Big]+o(\Delta)
\end{aligned}
$$

and it is negative. Hence $\ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}$ is a global convex function for $\boldsymbol{\alpha}(k\Delta)\in\mathcal{A}_0^{(2)}(k\Delta)$.

This completes the proof for the consistency $\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta)\xrightarrow{p}\boldsymbol{\alpha}_0^{(1)}(u)$ as $n\to\infty$, $\Delta\to 0$, $k\Delta=u$, and $\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}$ being constant within each increment Δ of $t\in\mathcal{T}_{i,k}$.

Next we will show the asymptotic normality of $\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta)$. First note that

$$
\sqrt{n}\{\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta)-\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}=-[\frac{1}{n}\frac{\partial^2\ell^{(2)}\{\tilde{\boldsymbol{\alpha}}^{(2)}(k\Delta)\}}{\partial\boldsymbol{\alpha}(k\Delta)^T\partial\boldsymbol{\alpha}(k\Delta)}]^{-1}\sqrt{n}\frac{1}{n}\frac{\partial\ell^{(2)}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}}{\partial\boldsymbol{\alpha}(k\Delta)^T}
$$

where $\tilde{\boldsymbol{\alpha}}^{(2)}(k\Delta)\in(\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta),\boldsymbol{\alpha}_0^{(2)}(k\Delta))$. In order to prove the asymptotic normality of $\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta)$, we need to show both

$$
[\boldsymbol{\Sigma}_{2,n}\{\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta)\}]^{-1/2}\sqrt{n}\frac{1}{n}\frac{\partial\ell^{(2)}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}}{\partial\boldsymbol{\alpha}^T(k\Delta)}\xrightarrow{d}N(0,\mathbf{I}) \tag{B.2.4}
$$

where

$$
\begin{aligned}
\boldsymbol{\Sigma}_{2,n}\{\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta)\}&=\mathrm{var}_{\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta)}[\frac{1}{\sqrt{n}}\frac{\partial\ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial\boldsymbol{\alpha}(k\Delta)^T}]\\
&=\frac{1}{n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}}\boldsymbol{Z}_i(t)\boldsymbol{Z}_i^T(t)\lambda_{T_i,t}\{\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta)\}\Delta+o(\Delta)
\end{aligned}
$$

and

$$
\mathbf{W}_{2,n}\{\tilde{\boldsymbol{\alpha}}^{(2)}(k\Delta)\}=-\frac{1}{n}\frac{\partial^2\ell^{(2)}\{\tilde{\boldsymbol{\alpha}}^{(2)}(k\Delta)\}}{\partial\boldsymbol{\alpha}(k\Delta)^T\partial\boldsymbol{\alpha}(k\Delta)}\xrightarrow{p}\mathbf{W}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\} \tag{B.2.5}
$$

for $\tilde{\boldsymbol{\alpha}}^{(2)}(k\Delta)\in(\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta),\boldsymbol{\alpha}_0^{(2)}(k\Delta))$ as $n\to\infty$.

By definition $E_{\boldsymbol{\alpha}_0^{(2)}(k\Delta)}[\frac{\partial\ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial\boldsymbol{\alpha}(k\Delta)^T}]=0$.

As shown in the proof of Result 3 in the section of technical proofs, the score function of $\ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}$ takes the form,

$$
\begin{aligned}
&\frac{\partial\ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial\{\boldsymbol{\alpha}(k\Delta)^T\}}\\
&=\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}}\boldsymbol{Z}_i(t)\Big[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}+\Big(\frac{\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp\left[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}\right]}\Big)\delta_{T_i}(t+k\Delta+\Delta)\Big]
\end{aligned}
$$

The second term $\frac{\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp\left[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}\right]} \approx 1$ by the Taylor expansion of the exponential function. Hence at the true value $\boldsymbol{\alpha}_0^{(2)}(k\Delta)$, as shown in the proof of the Result 4 in the section of technical proofs, the variance $\boldsymbol{\Sigma}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}$ can be expressed as,

$$\begin{aligned}\boldsymbol{\Sigma}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\} &= \mathrm{var}_{\boldsymbol{\alpha}_0^{(2)}(k\Delta)}[\frac{1}{\sqrt{n}}\frac{\partial\ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial\boldsymbol{\alpha}(k\Delta)^T}]\\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}} \boldsymbol{Z}_i(t)\boldsymbol{Z}_i^T(t)\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}\Delta + o(\Delta)\end{aligned}$$

By central limit theorem, we have

$$[\boldsymbol{\Sigma}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}]^{-\frac{1}{2}}\sqrt{n}\frac{1}{n}\frac{\partial\ell^{(2)}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}}{\partial\boldsymbol{\alpha}^T(k\Delta)} \xrightarrow{d} N(0,\mathbf{I})$$

As $\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta) \xrightarrow{p} \boldsymbol{\alpha}_0^{(2)}(k\Delta)$, we also have $\boldsymbol{\Sigma}_{2,n}\{\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta)\} \xrightarrow{p} \boldsymbol{\Sigma}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}$ by continuous mapping theorem. Hence we have (B.2.4).

Next we will show (B.2.5). It was shown in the proof of Result 5 in the section of technical proofs,

$$\begin{aligned}\mathbf{W}_{2,n}\{\boldsymbol{\alpha}(k\Delta)\} &= -\frac{1}{n}\frac{\partial^2\ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial\{\boldsymbol{\alpha}(k\Delta)\}^T\partial\{\boldsymbol{\alpha}(k\Delta)\}}\\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}} \lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{Z}_i(t)\boldsymbol{Z}_i^T(t)\Delta\Big[1-\delta_{T_i}(t+k\Delta+\Delta)\Big] + o(\Delta)\end{aligned}$$

By continuous mapping theorem, $\tilde{\boldsymbol{\alpha}}^{(2)}(k\Delta) \in (\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta), \boldsymbol{\alpha}_0^{(2)}(k\Delta))$, and $\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta) \xrightarrow{p} \boldsymbol{\alpha}_0^{(2)}(k\Delta)$, we have $\mathbf{W}_{2,n}\{\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta)\} \xrightarrow{p} \mathbf{W}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}$ and (B.2.5).

Let $\boldsymbol{\Omega}_{2,n}\{\boldsymbol{\alpha}(k\Delta)\} = \mathbf{W}_{2,n}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{\Sigma}_{2,n}^{-1}\{\boldsymbol{\alpha}(k\Delta)\}[\mathbf{W}_{2,n}\{\boldsymbol{\alpha}(k\Delta)\}]^T$. By Slutsky Theorem we have,

$$\sqrt{n}\boldsymbol{\Omega}_{2,n}^{1/2}\{\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta)\}\{\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta) - \boldsymbol{\alpha}_0^{(2)}(k\Delta)\} \xrightarrow{d} N(0,\mathbf{I})$$

We have shown before when $\boldsymbol{\alpha}_0^{(2)}(k\Delta) - \boldsymbol{\alpha}_0^{(1)}(u) = o(\Delta)$, $k\Delta = u$ and $\Delta \to 0$, we have $\frac{1}{n}\ell^{(2)}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\} - \frac{1}{n}\ell^{(1)}\{\boldsymbol{\alpha}_0^{(1)}(u)\} = o(\Delta)$. When $\Delta \to 0$ sufficiently fast, $k\Delta = u$ and $\sqrt{n}\Delta \to 0$ we have $\sqrt{n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta) - \boldsymbol{\alpha}_0^{(1)}(u)\} \xrightarrow{p} o_p(1)$. Given

$$\sqrt{n}\{\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta) - \boldsymbol{\alpha}_0^{(1)}(u)\} = \sqrt{n}[\{\hat{\boldsymbol{\alpha}}^{(2)}(k\Delta) - \boldsymbol{\alpha}_0^{(2)}(k\Delta)\} + \{\boldsymbol{\alpha}_0^{(2)}(k\Delta) - \boldsymbol{\alpha}_0^{(1)}(u)\}]$$

By Slutsky's theorem, we complete the proof of asymptotic normality in theorem 2.

Moreover $\boldsymbol{\Sigma}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\} - \boldsymbol{\Omega}_{1,n}\{\boldsymbol{\alpha}_0^{(1)}(u)\} = o(\Delta)$ and $\mathbf{W}_{2,n}\{\boldsymbol{\alpha}^{(2)}(k\Delta)\} - \boldsymbol{\Omega}_{1,n}\{\boldsymbol{\alpha}^{(1)}(u)\} = o(\Delta)$ as shown in Results 4 and 5 in the section of technical proofs. Hence we have

$$\boldsymbol{\Omega}_{2,n}\{\boldsymbol{\alpha}^{(2)}(k\Delta)\} - \boldsymbol{\Omega}_{1,n}\{\boldsymbol{\alpha}^{(1)}(u)\} = o(\Delta)$$

when $\Delta \to 0$, $k\Delta = u$ and $\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}$ being constant within each increment Δ of $t \in \mathcal{T}_{i,k}$.

B.3 Proof of Theorem 3

Proof:

The proof of consistency $\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta) \xrightarrow{p} \boldsymbol{\alpha}_0^{(3)}(k\Delta)$ is based on Shao (1999) and similar as that in theorem 2. We need to show that for the estimator $\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta)$,

$$P[\frac{1}{n}\frac{\partial \ell^{(3)}\{\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T} = 0] \to 1 \text{ for } \hat{\boldsymbol{\alpha}}^{(3)}(k\Delta) \xrightarrow{p} \boldsymbol{\alpha}_0^{(3)}(k\Delta) \tag{B.3.1}$$

First denote

$$\begin{aligned}\boldsymbol{\Sigma}_{3,n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\} &= \mathrm{var}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)}[\frac{1}{\sqrt{n}}\frac{\partial \ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T}] \\ &= \frac{1}{n}E_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)}([\frac{\partial \ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T}][\frac{\partial \ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T}]^T)\end{aligned}$$

and

$$\mathcal{A}_n^{(3)}(c;k\Delta) = \{\boldsymbol{\alpha}(k\Delta) : ||\{\boldsymbol{\alpha}(k\Delta) - \boldsymbol{\alpha}_0^{(3)}(k\Delta)\}[\boldsymbol{\Sigma}_{3,n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}]^{1/2}|| \leq c\} \text{ and } c > 0$$

Since the parameter space $A_0^{(3)}(k\Delta)$ is open, for each $c > 0$ $\mathcal{A}_n^{(3)}(c;k\Delta) \subset A_0^{(3)}(k\Delta)$ for sufficiently large n. Since $\mathcal{A}_n^{(3)}(c;k\Delta)$ shrinks to the true value $\boldsymbol{\alpha}_0^{(3)}(k\Delta)$ as $n \to \infty$, the existence of $\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta)$ satisfying (B.3.1) is implied by that for any $\epsilon > 0$, there exists $c > 0$ and $n_0 > 1$ such that

$$\mathbb{P}[\frac{1}{n}\ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\} - \frac{1}{n}\ell^{(3)}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\} < 0 \text{ for all } \boldsymbol{\alpha}(k\Delta) \in \partial\mathcal{A}_n^{(3)}(c;k\Delta)] \geq 1-\epsilon, n \geq n_0 \tag{B.3.2}$$

where $\partial\mathcal{A}_n^{(3)}(c;k\Delta)$ is the boundary of $\mathcal{A}_n^{(3)}(c;k\Delta)$. For $\boldsymbol{\alpha}(k\Delta) \in \mathcal{A}_n^{(3)}(c;k\Delta)$, by Taylor expansion we have,

$$\begin{aligned}&\frac{1}{n}\ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\} - \frac{1}{n}\ell^{(3)}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\} = c\frac{1}{n}\frac{\partial \ell^{(3)}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T}[\boldsymbol{\Sigma}_{3,n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}]^{-1/2}\lambda^T \\ &+ \frac{c^2}{2}\lambda[\boldsymbol{\Sigma}_{3,n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}]^{-1/2}\frac{1}{n}\frac{\partial^2 \ell^{(3)}\{\tilde{\boldsymbol{\alpha}}^{(3)}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T\partial \boldsymbol{\alpha}(k\Delta)}[\boldsymbol{\Sigma}_{3,n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}]^{-1/2}\lambda^T\end{aligned}$$

where $\lambda = \{\boldsymbol{\alpha}(k\Delta) - \boldsymbol{\alpha}_0^{(3)}(k\Delta)\}[\boldsymbol{\Sigma}_{3,n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}]^{1/2}/c$ satisfying $||\lambda|| = 1$, $-\mathbf{W}_{3,n}\{\boldsymbol{\alpha}^{(3)}(k\Delta)\} = \frac{1}{n}\frac{\partial^2 \ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial\{\boldsymbol{\alpha}(k\Delta)\}^T\partial\{\boldsymbol{\alpha}(k\Delta)\}}$ and $\tilde{\boldsymbol{\alpha}}^{(3)}(k\Delta) \in (\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta), \boldsymbol{\alpha}_0^{(3)}(k\Delta))$. Note that

$$\begin{aligned}&E||\mathbf{W}_{3,n}\{\tilde{\boldsymbol{\alpha}}^{(3)}(k\Delta)\} - \mathbf{W}_{3,n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}|| \\ &\leq \frac{1}{n}E\max_{\boldsymbol{\alpha}(k\Delta)\in\mathcal{A}_n^{(3)}(c;k\Delta)}||\frac{\partial^2 \ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T\partial \boldsymbol{\alpha}(k\Delta)} - \frac{\partial^2 \ell^{(3)}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T\partial \boldsymbol{\alpha}(k\Delta)}|| \\ &\to 0\end{aligned}$$

which follows from three conditions, (1) condition (3A) that the second order derivative of $\ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}$ is continuous in a neighborhood of $\boldsymbol{\alpha}_0^{(3)}(k\Delta) \in \mathcal{A}_0^{(3)}(k\Delta)$; (2) $\mathcal{A}_n^{(3)}(c;k\Delta)$ shrinks to $\boldsymbol{\alpha}_0^{(3)}(k\Delta)$ as $n \to \infty$; and (3) when $n \to \infty$, $E\max_{\boldsymbol{\alpha}(k\Delta)\in\mathcal{A}_n^{(3)}(c;k\Delta)}||\frac{\partial^2\ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial\boldsymbol{\alpha}(k\Delta)^T\partial\boldsymbol{\alpha}(k\Delta)} - \frac{\partial^2\ell^{(3)}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}}{\partial\boldsymbol{\alpha}(k\Delta)^T\partial\boldsymbol{\alpha}(k\Delta)}||$ is finite.

As shown in Results 4 and 5 in the section of technical proofs when $\Delta \to 0$,

$$\begin{aligned}
\boldsymbol{\Sigma}_{3,n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\} &= \mathrm{var}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)}[\sqrt{n}\frac{1}{n}\frac{\partial\ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial\boldsymbol{\alpha}^T(k\Delta)}]\Delta^2 \\
&= \frac{1}{4n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}} \boldsymbol{Z}_i(t)\boldsymbol{Z}_i(t)^T\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}\Delta + o(\Delta)
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{W}_{3,n}\{\boldsymbol{\alpha}(k\Delta)\} &= -\frac{\Delta}{n}\frac{\partial^2\ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial\boldsymbol{\alpha}(k\Delta)^T\partial\boldsymbol{\alpha}(k\Delta)} \\
&= \frac{1}{2n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}} \lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{Z}_i(t)\boldsymbol{Z}_i^T(t)\Delta\Big[1-\delta_{T_i}(t+k\Delta+\Delta)\Big] + o(\Delta)
\end{aligned}$$

Hence

$$[\boldsymbol{\Sigma}_{3,n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}]^{-1/2}[-\mathbf{W}_{3,n}\{\tilde{\boldsymbol{\alpha}}^{(3)}(k\Delta)\}][\boldsymbol{\Sigma}_{3,n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}]^{-1/2} \approx -\frac{1}{2}$$

Hence we have

$$\begin{aligned}
&\frac{1}{n}\ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\} - \frac{1}{n}\ell^{(3)}\{\boldsymbol{\alpha}_0(k\Delta)\} = c\frac{1}{n}\frac{\partial\ell^{(3)}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}}{\partial\boldsymbol{\alpha}(k\Delta)^T}[\boldsymbol{\Sigma}_{3,n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}]^{-1/2}\lambda^T \\
&-\frac{c^2}{4} + o_p(1)
\end{aligned}$$

Note that $\max_\lambda(\frac{1}{n}\frac{\partial\ell^{(3)}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}}{\partial\boldsymbol{\alpha}(k\Delta)^T}[\boldsymbol{\Sigma}_{3,n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}]^{-1/2}\lambda^T) = ||\frac{1}{n}\frac{\partial\ell^{(3)}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}}{\partial\boldsymbol{\alpha}(k\Delta)^T}[\boldsymbol{\Sigma}_{3,n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}]^{-1/2}||$. By choosing c sufficiently large we have,

$$\begin{aligned}
&\mathbb{P}(||\frac{1}{n}\frac{\partial\ell^{(3)}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}}{\partial\boldsymbol{\alpha}(k\Delta)^T}[\boldsymbol{\Sigma}_{3,n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}]^{-1/2}||< \frac{c}{4}) \\
&\geq 1 - (\frac{c}{4})^2 E||\frac{1}{n}\frac{\partial\ell^{(3)}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}}{\partial\boldsymbol{\alpha}(k\Delta)^T}[\boldsymbol{\Sigma}_{3,n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}]^{-1/2}||^2 \\
&= 1 - k(\frac{c}{4})^2 \\
&\geq 1 - \epsilon
\end{aligned}$$

Hence we have (B.3.2) and complete the proof of (B.3.1) that $\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta) \xrightarrow{p} \boldsymbol{\alpha}_0^{(3)}(k\Delta)$.

To complete the proof of the consistency $\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta) \xrightarrow{p} \boldsymbol{\alpha}_0^{(1)}(u)$, next we need to show that

$$\frac{\Delta}{n}\frac{\partial \ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)} - \frac{1}{2n}\frac{\partial \ell^{(1)}\{\boldsymbol{\alpha}(u)\}}{\partial \boldsymbol{\alpha}(u)} = o(\Delta) \tag{B.3.3}$$

This was proven in the section of technical proofs in Result 3 (B.4.2) and (B.4.3) when $\Delta \to 0$, $k\Delta = u$, $h_n \to 0$, $\frac{\Delta}{h_n}$ and $\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}$ being constant within each increment Δ of $t \in \mathcal{T}_{i,k}$.

Also the second order derivative of the smoothed composite log-likelihood as shown in Result 5 in the section of technical proofs takes the form,

$$\begin{aligned}
-\mathbf{W}_{3,n}\{\boldsymbol{\alpha}(k\Delta)\} &= \frac{\Delta}{n}\frac{\partial^2 \ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T \partial \boldsymbol{\alpha}(k\Delta)} \\
&= -\frac{1}{2n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}} \lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{Z}_i(t)\boldsymbol{Z}_i^T(t)\Delta\Big[1-\delta_{T_i}(t+k\Delta+\Delta)\Big] + o(\Delta)
\end{aligned}$$

which is negative. Hence $\ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}$ is a global convex function for $\boldsymbol{\alpha}(k\Delta) \in \mathcal{A}_0^{(3)}(k\Delta)$.

This completes the proof for the consistency $\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta) \xrightarrow{p} \boldsymbol{\alpha}_0^{(1)}(u)$ as $\Delta \to 0$, $k\Delta = u$, $h_n \to 0$, $\frac{\Delta}{h_n}$ and $\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}$ being constant within each increment Δ of $t \in \mathcal{T}_{i,k}$.

Next we will show the asymptotic normality of $\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta)$. First note that

$$\begin{aligned}
&\sqrt{n}\{\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta) - \boldsymbol{\alpha}_0^{(1)}(u)\} \\
&= \sqrt{n}[\{\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta) - \boldsymbol{\alpha}_0^{(3)}(k\Delta)\} + \{\boldsymbol{\alpha}_0^{(3)}(k\Delta) - \boldsymbol{\alpha}_0^{(1)}(u)\}]
\end{aligned}$$

Hence to show the asymptotic normality, we will first find the distribution of $\sqrt{n}[\{\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta) - \boldsymbol{\alpha}_0^{(3)}(k\Delta)\}]$. Note that

$$\sqrt{n}\{\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta) - \boldsymbol{\alpha}_0^{(3)}(k\Delta)\} = -[\frac{1}{n}\frac{\partial^2 \ell^{(3)}\{\tilde{\boldsymbol{\alpha}}^{(3)}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T \partial \boldsymbol{\alpha}(k\Delta)}\Delta]^{-1}\sqrt{n}\frac{1}{n}\frac{\partial \ell^{(3)}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T}\Delta$$

where $\tilde{\boldsymbol{\alpha}}^{(3)}(k\Delta) \in (\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta), \boldsymbol{\alpha}_0^{(3)}(k\Delta))$. Therefore in order to prove the asymptotic normality of the estimated parameter $\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta)$, we need to show both

$$\boldsymbol{\Sigma}_{3,n}^{-1/2}\{\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta)\}\sqrt{n}\frac{1}{n}\frac{\partial \ell\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T}\Delta \xrightarrow{d} N(0, \mathbf{I}). \tag{B.3.4}$$

where

$$
\begin{aligned}
&\mathbf{\Sigma}_{3,n}\{\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta)\} = \mathrm{var}_{\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta)}[\sqrt{n}\frac{1}{n}\frac{\partial \ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial \boldsymbol{\alpha}^T(k\Delta)}]\Delta^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathrm{var}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)}\Big(\sum_{t\in\mathcal{T}_{i,k}}\sum_{s\in\mathcal{S}_{i,k}}\frac{1}{h_n}K(\frac{s-t}{h_n})I(s\leq t)\boldsymbol{Z}_i(s)[-\Delta\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\} \\
&+ \Big(\frac{\Delta\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp\left[-\Delta\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\}\right]}\Big)\delta_{T_i}(s+k\Delta+\Delta)]\Big) \\
&= \frac{1}{4n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}}\boldsymbol{Z}_i(t)\boldsymbol{Z}_i(t)^T\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}\Delta + o(\Delta) \qquad \text{(B.3.5)}
\end{aligned}
$$

and

$$
\mathbf{W}_{3,n}\{\tilde{\boldsymbol{\alpha}}^{(3)}(k\Delta)\} = -\frac{1}{n}\frac{\partial^2\ell^{(2)}\{\tilde{\boldsymbol{\alpha}}^{(3)}(k\Delta)\}}{\partial\boldsymbol{\alpha}(k\Delta)^T\partial\boldsymbol{\alpha}(k\Delta)} \xrightarrow{p} \mathbf{W}_{3,n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\} \qquad \text{(B.3.6)}
$$

First we will prove (B.3.4). By definition $\mathrm{E}[\frac{\partial\ell^{(3)}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}}{\partial\boldsymbol{\alpha}(k\Delta)^T}] = 0$. As shown in Result 4 in the section of technical proofs, under conditions $h_n \to 0$, $\Delta \to 0$, $\frac{\Delta}{h_n} \to 0$; $k\Delta = u$ and $\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}$ being constant within each increment Δ of $t \in \mathcal{T}_{i,k}$ we have (B.3.5).

By central limit theorem,

$$
\mathbf{\Sigma}_{3,n}^{-1/2}\{\boldsymbol{\alpha}_0(k\Delta)\}\sqrt{n}\frac{1}{n}\frac{\partial\ell\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}}{\partial\boldsymbol{\alpha}(k\Delta)^T}\Delta \xrightarrow{d} N(0,\mathbf{I}).
$$

By continuous mapping theorem as $\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta) \xrightarrow{p} \boldsymbol{\alpha}_0^{(3)}(k\Delta)$, we have $\mathbf{\Sigma}_{3,n}\{\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta)\} \xrightarrow{p} \mathbf{\Sigma}_{3,n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}$. Therefore we have (B.3.4).

Next we will show (B.3.6). It was shown in the proof of Result 5 in the section of technical proofs,

$$
\begin{aligned}
\mathbf{W}_{3,n}\{\boldsymbol{\alpha}(k\Delta)\} &= -\frac{\Delta}{n}\frac{\partial^2\ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial\boldsymbol{\alpha}(k\Delta)^T\partial\boldsymbol{\alpha}(k\Delta)} \\
&= \frac{1}{2n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}}\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{Z}_i(t)\boldsymbol{Z}_i^T(t)\Delta\Big[1-\delta_{T_i}(t+k\Delta+\Delta)\Big] + o(\Delta)
\end{aligned}
$$

By continuous mapping theorem, $\tilde{\boldsymbol{\alpha}}^{(3)}(k\Delta) \in (\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta), \boldsymbol{\alpha}_0^{(3)}(k\Delta))$, and $\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta) \xrightarrow{p} \boldsymbol{\alpha}_0^{(3)}(k\Delta)$, we have $\mathbf{W}_{3,n}\{\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta)\} \xrightarrow{p} \mathbf{W}_{3,n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}$ and (B.3.6).

Let $\mathbf{\Omega}_{3,n}\{\boldsymbol{\alpha}(k\Delta)\} = \mathbf{W}_{3,n}\{\boldsymbol{\alpha}(k\Delta)\}\mathbf{\Sigma}_{3,n}^{-1}\{\boldsymbol{\alpha}(k\Delta)\}[\mathbf{W}_{3,n}\{\boldsymbol{\alpha}(k\Delta)\}]^T$. By continuous mapping theorem $\mathbf{W}_{3,n}\{\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta)\} \xrightarrow{p} \mathbf{W}_3\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}$. We now have,

$$
\sqrt{n}\mathbf{\Omega}_{3,n}^{1/2}\{\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta)\}\{\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta) - \boldsymbol{\alpha}_0^{(3)}(k\Delta)\} \xrightarrow{d} N(0,\mathbf{I}).
$$

When $\boldsymbol{\alpha}_0^{(3)}(k\Delta) - \boldsymbol{\alpha}_0^{(1)}(u) = o(\Delta)$, $k\Delta = u$ and $\Delta \to 0$, we have $\frac{\Delta}{n}\ell^{(3)}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\} - \frac{1}{2n}\ell^{(1)}\{\boldsymbol{\alpha}_0^{(1)}(u)\} = o(\Delta)$. When $\Delta \to 0$ sufficiently fast, $k\Delta = u$ and $\sqrt{n}\Delta \to 0$ we have

$\sqrt{n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta) - \boldsymbol{\alpha}_0^{(1)}(u)\} \xrightarrow{p} o_p(1)$. Given

$$\sqrt{n}\{\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta) - \boldsymbol{\alpha}_0^{(1)}(u)\} = \sqrt{n}[\{\hat{\boldsymbol{\alpha}}^{(3)}(k\Delta) - \boldsymbol{\alpha}_0^{(3)}(k\Delta)\} + \{\boldsymbol{\alpha}_0^{(3)}(k\Delta) - \boldsymbol{\alpha}_0^{(1)}(u)\}]$$

By Slutsky's theorem, we complete the proof of asymptotic normality in theorem 3.

Moreover as shown in Results 4 and 5 in the section of technical proofs we have $\boldsymbol{\Sigma}_{3,n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\} - \frac{1}{4}\boldsymbol{\Sigma}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\} = o(\Delta)$ and $\mathbf{W}_{3,n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\} - \frac{1}{2}\mathbf{W}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\} = o(\Delta)$ when $h_n \to 0$, $\Delta \to 0$, $\frac{\Delta}{h_n} \to 0$, $k\Delta = u$ and $\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}$ being constant within each increment Δ of $t \in \mathcal{T}_{i,k}$.

Given $\boldsymbol{\Omega}_{3,n}\{\boldsymbol{\alpha}(k\Delta)\} = \mathbf{W}_{3,n}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{\Sigma}_{3,n}^{-1}\{\boldsymbol{\alpha}(k\Delta)\}[\mathbf{W}_3\{\boldsymbol{\alpha}(k\Delta)\}]^T$ and $\boldsymbol{\Omega}_{2,n}\{\boldsymbol{\alpha}(k\Delta)\} = \mathbf{W}_{2,n}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{\Sigma}_{2,n}^{-1}\{\boldsymbol{\alpha}(k\Delta)\}[\mathbf{W}_{2,n}\{\boldsymbol{\alpha}(k\Delta)\}]^T$, we have $\boldsymbol{\Omega}_{3,n}\{\boldsymbol{\alpha}(k\Delta)\} - \boldsymbol{\Omega}_{2,n}\{\boldsymbol{\alpha}(k\Delta)\} = o(\Delta)$.

B.4 Technical proofs

B.4.1 Result 1:

$$\begin{aligned}
\mathrm{cov}_i =& \mathrm{cov}_{\boldsymbol{\alpha}_0^{(2)}(k\Delta)}\Big[- \Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\} + \Big(\frac{\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}]}\Big)\delta_{T_i}(t+k\Delta+\Delta), \qquad \text{(B.4.1)}\\
& - \Delta\lambda_{T_i,t'}\{\boldsymbol{\alpha}(k\Delta)\} + \Big(\frac{\Delta\lambda_{T_i,t'}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,t'}\{\boldsymbol{\alpha}(k\Delta)\}]}\Big)\delta_{T_i}(t'+k\Delta+\Delta)\Big]\\
& = \Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\} + o(\Delta)
\end{aligned}$$

$$\mathrm{cov}_i = \begin{cases} \Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\} + o(\Delta) & \text{if } t = t' \\ -2\Delta^2\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}\lambda_{T_i,t'}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\} & \text{if } t \neq t' \end{cases}.$$

Proof:

$$\begin{aligned}
& \mathrm{var}_{\boldsymbol{\alpha}_0^{(2)}(k\Delta)}\Big[- \Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\} + \Big(\frac{\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}]}\Big)\delta_{T_i}(t+k\Delta+\Delta)\Big]\\
&= \mathrm{var}_{\boldsymbol{\alpha}_0^{(2)}(k\Delta)}[\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}]\\
&+ \mathrm{var}_{\boldsymbol{\alpha}_0^{(2)}(k\Delta)}\{\Big(\frac{\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}]}\Big)\delta_{T_i}(t+k\Delta+\Delta)\}\\
&- \mathrm{cov}_{\boldsymbol{\alpha}_0^{(2)}(k\Delta)}\Big[\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}, \Big(\frac{\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}]}\Big)\delta_{T_i}(t+k\Delta+\Delta)\Big]
\end{aligned}$$

in which

$$\mathrm{var}_{\boldsymbol{\alpha}_0^{(2)}(k\Delta)}[\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}] = 0$$

$$\begin{aligned}
&\mathrm{var}_{\boldsymbol{\alpha}_0^{(2)}(k\Delta)}\{\Big(\frac{\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}]}\Big)\delta_{T_i}(t+k\Delta+\Delta)\}\\
&= \Big(\frac{\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}}{1-\exp\Big[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}\Big]}\Big)^2 \exp[-\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}\Delta]\Big(1-\exp[-\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}\Delta]\Big)\\
&= \frac{[\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}]^2}{1-\exp\Big[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}\Big]}\exp[-\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}\Delta]\\
&= \Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\} + o(\Delta)
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{cov}_{\boldsymbol{\alpha}_0^{(2)}(k\Delta)}\Big[\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\},\Big(\frac{\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}]}\Big)\delta_{T_i}(t+k\Delta+\Delta)\Big]\\
&= \Big(\frac{[\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}]^2}{1-\exp\Big[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}\Big]}\Big)E_{\boldsymbol{\alpha}_0^{(2)}(k\Delta)}\delta_{T_i}(t+k\Delta+\Delta)\\
&= [\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}]^2 = o(\Delta)
\end{aligned}$$

Hence

$$\begin{aligned}
&\mathrm{var}_{\boldsymbol{\alpha}_0^{(2)}(k\Delta)}\Big[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}+\Big(\frac{\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}]}\Big)\delta_{T_i}(t+k\Delta+\Delta)\Big]\\
&= \Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\} + o(\Delta)
\end{aligned}$$

and for $t \neq t'$,

$$\begin{aligned}
&\mathrm{cov}_{\boldsymbol{\alpha}_0^{(2)}(k\Delta)}\Big[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}+\Big(\frac{\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}]}\Big)\delta_{T_i}(t+k\Delta+\Delta),\\
&-\Delta\lambda_{T_i,t'}\{\boldsymbol{\alpha}(k\Delta)\}+\Big(\frac{\Delta\lambda_{T_i,t'}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,t'}\{\boldsymbol{\alpha}(k\Delta)\}]}\Big)\delta_{T_i}(t'+k\Delta+\Delta)\Big]\\
&= -\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}\Big(\frac{\Delta\lambda_{T_i,t'}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}}{1-\exp\Big[-\Delta\lambda_{T_i,t'}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}\Big]}\Big)E_{\boldsymbol{\alpha}_0^{(2)}(k\Delta)}\delta_{T_i}(t'+k\Delta+\Delta)\\
&-\Delta\lambda_{T_i,t'}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}\Big(\frac{\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}}{1-\exp\Big[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}\Big]}\Big)E_{\boldsymbol{\alpha}_0^{(2)}(k\Delta)}\delta_{T_i}(t+k\Delta+\Delta)\\
&+\Big(\frac{\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}}{1-\exp\Big[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}\Big]}\Big)\Big(\frac{\Delta\lambda_{T_i,t'}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}}{1-\exp\Big[-\Delta\lambda_{T_i,t'}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}\Big]}\Big)\\
&\times \mathrm{cov}_{\boldsymbol{\alpha}_0^{(2)}(k\Delta)}\{\delta_{T_i}(t+k\Delta+\Delta),\delta_{T_i}(t'+k\Delta+\Delta)\}\\
&= -2\Delta^2\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}\lambda_{T_i,t'}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}
\end{aligned}$$

Under condition 3 of Theorem 3, we have Results 2 and 3.

B.4.2 Result 2:

$$\begin{aligned}
&\boldsymbol{\Sigma}_{3,n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\} = \mathrm{var}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)}[\sqrt{n}\frac{1}{n}\frac{\partial \ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial \boldsymbol{\alpha}^T(k\Delta)}]\Delta^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathrm{var}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)}\Big(\sum_{t\in\mathcal{T}_{i,k}}\sum_{s\in\mathcal{S}_{i,k}}\frac{1}{h_n}K(\frac{s-t}{h_n})I(s\le t)\boldsymbol{Z}_i(s)[-\Delta\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\} \\
&+ \Big(\frac{\Delta\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\}]}\Big)\delta_{T_i}(s+k\Delta+\Delta)]\Big) \\
&= \frac{1}{n}\sum_{i=1}^{n}V_i \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{\Delta}{4}\sum_{t\in\mathcal{T}_{i,k}}\boldsymbol{Z}_i(t)\boldsymbol{Z}_i(t)^T\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\} + o(\Delta)
\end{aligned}$$

in which,

$$V_i = \begin{cases}
\frac{\Delta}{4}\sum_{t\in\mathcal{T}_{i,k}}\boldsymbol{Z}_i(t)\boldsymbol{Z}_i(t)^T\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}+o(\Delta) & \text{if } s_1=s_2 \\
\frac{\Delta^2}{h_n}\sum_{t\in\mathcal{T}_{i,k}}\int_{-\infty}^{0}K^2(z)dz\boldsymbol{Z}_i(t)\boldsymbol{Z}_i(t)^T\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}+o(\frac{\Delta^2}{h_n}) & \text{if } t_1=t_2 \text{ and } s_1=s_2 \\
\frac{\Delta^2}{2h_n}\sum_{t\in\mathcal{T}_{i,k}}K(0)\boldsymbol{Z}_i(t)\boldsymbol{Z}_i(t)^T\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}+o(\frac{\Delta^2}{h_n}) & \text{if } s_1=s_2=t_1 \\
 & \text{or } s_1=s_2=t_2 \\
 & \\
-\frac{\Delta^2}{2}\sum_{t_1\in\mathcal{T}_{i,k}}\sum_{t_2\in\mathcal{T}_{i,k}}\boldsymbol{Z}_i(t_1)\boldsymbol{Z}_i(t_2)^T & \\
\times\lambda_{T_i,t_1}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}\lambda_{T_i,t_2}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}+o(\Delta^2) & \text{if } t_1\ne t_2\ne s_1\ne s_2 \\
-\frac{\Delta^2}{2}\sum_{t\in\mathcal{T}_{i,k}}\boldsymbol{Z}_i(t)\boldsymbol{Z}_i(t)^T\lambda^2_{T_i,t}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}+o(\Delta^2) & \text{if } s_1=t_2 \text{ or } s_2=t_1 \\
 & \text{or } t_1=t_2 \\
\frac{\Delta^3}{h_n^2}\sum_{s\in\mathcal{S}_{i,k}}K^2(0)\boldsymbol{Z}_i(s)\boldsymbol{Z}_i(s)^T\lambda_{T_i,s}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}+o(\frac{\Delta^3}{h_n^2}) & \text{if } s_1=t_1=s_2=t_2 \\
-\frac{\Delta^3}{h_n}\sum_{t\in\mathcal{T}_{i,k}}\sum_{s\in\mathcal{S}_{i,k}}K(0)\boldsymbol{Z}_i(s)\boldsymbol{Z}_i(t)^T & \\
\times\lambda_{T_i,s}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}+o(\frac{\Delta^3}{h_n}) & \text{if } s_1=t_1 \text{ or } s_2=t_2 \\
-\frac{\Delta^3}{h_n}\sum_{s\in\mathcal{S}_{i,k}}\boldsymbol{Z}_i(s)\boldsymbol{Z}_i(s)^T\lambda^2_{T_i,s}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}+o(\frac{\Delta^3}{h_n}) & \text{if } t_1=t_2=s_1 \\
 & \text{or } t_1=t_2=s_2 \\
 & \\
-\frac{2\Delta^4}{h_n^2}\sum_{s_1\in\mathcal{S}_{i,k}}\sum_{s_2\in\mathcal{S}_{i,k}}K^2(0)\boldsymbol{Z}_i(s_1)\boldsymbol{Z}_i(s_2)^T & \\
\times\lambda_{T_i,s_1}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}\lambda_{T_i,s_2}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\} & \text{if } s_1=t_1 \text{ and } s_2=t_2
\end{cases}.$$

When $s_1 = s_2 = s$,

$$\Delta^2 \sum_{t_1\in\mathcal{T}_{i,k}} \sum_{t_2\in\mathcal{T}_{i,k}} \sum_{s\in\mathcal{S}_{i,k}} \mathrm{cov}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)}\Bigg\{\Bigg(\frac{1}{h_n}K(\frac{s-t_1}{h_n})I(s\le t_1)\boldsymbol{Z}_i(s)[-\Delta\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\}$$

$$+\Big(\frac{\Delta\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\}]}\Big)\delta_{T_i}(s+k\Delta+\Delta)]\Bigg)$$

$$,\Bigg(\frac{1}{h_n}K(\frac{s-t_2}{h_n})I(s\le t_2)\boldsymbol{Z}_i(s)[-\Delta\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\}$$

$$+\Big(\frac{\Delta\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\}]}\Big)\delta_{T_i}(s+k\Delta+\Delta)]\Bigg)\Bigg\}$$

$$=\Delta^2 \sum_{t_1\in\mathcal{T}_{i,k}} \sum_{t_2\in\mathcal{T}_{i,k}} \sum_{s\in\mathcal{S}_{i,k}} \frac{1}{h_n^2}K(\frac{s-t_1}{h_n})K(\frac{s-t_2}{h_n})I(s\le t_1)I(s\le t_2)\boldsymbol{Z}_i(s)\boldsymbol{Z}_i(s)^T$$

$$\times\,\mathrm{var}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)}\Bigg([-\Delta\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\}+\Big(\frac{\Delta\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\}]}\Big)\delta_{T_i}(s+k\Delta+\Delta)]\Bigg)$$

$$=\sum_{t_1\in\mathcal{T}_{i,k}} \int_{-\infty}^{0}\int_{-\infty}^{0} K(z_1)K(z_2)dz_1dz_2\boldsymbol{Z}_i(t_1+z_1h_n)\boldsymbol{Z}_i(t_1+z_1h_n)^T\Delta\lambda_{T_i,t_1+z_1h_n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}$$

$$+\,o(\Delta)$$

$$=\frac{\Delta}{4}\sum_{t\in\mathcal{T}_{i,k}} \boldsymbol{Z}_i(t)\boldsymbol{Z}_i(t)^T\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}+o(\Delta)$$

When $t_1 = t_2$ and $s_1 = s_2$,

$$\Delta^2 \sum_{t_1\in\mathcal{T}_{i,k}} \sum_{s_1\in\mathcal{S}_{i,k}} \mathrm{var}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)}\Bigg\{\Bigg(\frac{1}{h_n}K(\frac{s_1-t_1}{h_n})I(s_1\le t_1)\boldsymbol{Z}_i(s_1)[-\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}$$

$$+\Big(\frac{\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}]}\Big)\delta_{T_i}(s_1+k\Delta+\Delta)]\Bigg)\Bigg\}$$

$$=\Delta^2 \sum_{t_1\in\mathcal{T}_{i,k}} \sum_{s_1\in\mathcal{S}_{i,k}} \frac{1}{h_n^2}K^2(\frac{s_1-t_1}{h_n})I(s_1\le t_1)\boldsymbol{Z}_i(s_1)\boldsymbol{Z}_i(s_1)^T\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}$$

$$=\frac{\Delta^2}{h_n}\sum_{t_1\in\mathcal{T}_{i,k}} \int_{-\infty}^{0} K^2(z_1)\boldsymbol{Z}_i(t_1+z_1h_n)\boldsymbol{Z}_i(t_1+z_1h_n)^T\lambda_{T_i,t_1+z_1h_n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}dz_1+o(\frac{\Delta^2}{h_n})$$

$$=\frac{\Delta^2}{h_n}\sum_{t\in\mathcal{T}_{i,k}} \int_{-\infty}^{0} K^2(z)dz\boldsymbol{Z}_i(t)\boldsymbol{Z}_i(t)^T\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}+o(\frac{\Delta^2}{h_n})$$

When $s_1 = s_2 = t_1$ or $s_1 = s_2 = t_2$

$$\Delta^2 \sum_{t_2 \in \mathcal{T}_{i,k}} \sum_{s_1 \in \mathcal{S}_{i,k}} \mathrm{cov}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)} \Big\{ \Big(\frac{1}{h_n} K(0) \boldsymbol{Z}_i(s_1) [-\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}$$
$$+ \Big(\frac{\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}]} \Big) \delta_{T_i}(s_1 + k\Delta + \Delta)] \Big)$$
$$, \Big(\frac{1}{h_n} K(\frac{s_1 - t_2}{h_n}) I(s_1 \le t_2) \boldsymbol{Z}_i(s_1) [-\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}$$
$$+ \Big(\frac{\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}]} \Big) \delta_{T_i}(s_1 + k\Delta + \Delta)] \Big) \Big\}$$
$$= \Delta^2 \sum_{t_2 \in \mathcal{T}_{i,k}} \sum_{s_1 \in \mathcal{S}_{i,k}} \frac{1}{h_n^2} K(0) K(\frac{s_1 - t_2}{h_n}) I(s_1 \le t_2) \boldsymbol{Z}_i(s_1) \boldsymbol{Z}_i(s_1)^T$$
$$\times \mathrm{var}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)} \Big\{ [-\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\} + \Big(\frac{\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}]} \Big) \delta_{T_i}(s_1 + k\Delta + \Delta)] \Big\}$$
$$= \frac{\Delta^2}{h_n} \sum_{t_2 \in \mathcal{T}_{i,k}} \int_{-\infty}^{0} K(0) K(z_1) dz_1 \boldsymbol{Z}_i(t_2 + z_1 h_n) \boldsymbol{Z}_i(t_2 + z_1 h_n)^T \lambda_{T_i, t_2 + z_1 h_n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\} + o(\frac{\Delta^2}{h_n})$$
$$= \frac{\Delta^2}{2h_n} \sum_{t \in \mathcal{T}_{i,k}} K(0) \boldsymbol{Z}_i(t) \boldsymbol{Z}_i(t)^T \lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\} + o(\frac{\Delta^2}{h_n})$$

When $t_1 \neq t_2 \neq s_1 \neq s_2$,

$$\Delta^2 \sum_{t_1 \in \mathcal{T}_{i,k}} \sum_{t_2 \in \mathcal{T}_{i,k}} \sum_{s_1 \in \mathcal{S}_{i,k}} \sum_{s_2 \in \mathcal{S}_{i,k}} \mathrm{cov}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)} \Big\{ \Big(\frac{1}{h_n} K(\frac{s_1 - t_1}{h_n}) I(s_1 \le t_1) \boldsymbol{Z}_i(s_1) [-\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}$$
$$+ \Big(\frac{\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}]} \Big) \delta_{T_i}(s_1 + k\Delta + \Delta)] \Big)$$
$$, \Big(\frac{1}{h_n} K(\frac{s_2 - t_2}{h_n}) I(s_2 \le t_2) \boldsymbol{Z}_i(s_2) [-\Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\}$$
$$+ \Big(\frac{\Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\}]} \Big) \delta_{T_i}(s_2 + k\Delta + \Delta)] \Big) \Big\}$$
$$= \Delta^2 \sum_{t_1 \in \mathcal{T}_{i,k}} \sum_{t_2 \in \mathcal{T}_{i,k}} \sum_{s_1 \in \mathcal{S}_{i,k}} \sum_{s_2 \in \mathcal{S}_{i,k}} \frac{1}{h_n^2} K(\frac{s_1 - t_1}{h_n}) I(s_1 \le t_1) K(\frac{s_2 - t_2}{h_n}) I(s_2 \le t_2) \boldsymbol{Z}_i(s_1) \boldsymbol{Z}_i(s_2)^T$$
$$\times \mathrm{cov}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)} \Big\{ [-\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\} + \Big(\frac{\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}]} \Big) \delta_{T_i}(s_1 + k\Delta + \Delta)]$$
$$, [-\Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\} + \Big(\frac{\Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\}]} \Big) \delta_{T_i}(s_2 + k\Delta + \Delta)] \Big\}$$
$$= \sum_{t_1 \in \mathcal{T}_{i,k}} \sum_{t_2 \in \mathcal{T}_{i,k}} \int_{-\infty}^{0} \int_{-\infty}^{0} K(z_1) K(z_2) dz_1 dz_2 \boldsymbol{Z}_i(t_1 + z_1 h_n) \boldsymbol{Z}_i(t_2 + z_2 h_n)^T$$
$$\times [-2\Delta^2 \lambda_{T_i, t_1 + z_1 h_n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\} \lambda_{T_i, t_2 + z_2 h_n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}] + o(\Delta^2)$$
$$= -\frac{\Delta^2}{2} \sum_{t_1 \in \mathcal{T}_{i,k}} \sum_{t_2 \in \mathcal{T}_{i,k}} \boldsymbol{Z}_i(t_1) \boldsymbol{Z}_i(t_2)^T \lambda_{T_i,t_1}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\} \lambda_{T_i,t_2}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\} + o(\Delta^2)$$

When $s_1 = t_2$ or $s_2 = t_1$,

$$
\begin{aligned}
&\Delta^2 \sum_{t_1 \in \mathcal{T}_{i,k}} \sum_{s_1 \in \mathcal{S}_{i,k}} \sum_{s_2 \in \mathcal{S}_{i,k}} \mathrm{cov}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)} \Bigg\{ \bigg(\frac{1}{h_n} K(\frac{s_1 - t_1}{h_n}) I(s_1 \le t_1) \boldsymbol{Z}_i(s_1) [- \Delta \lambda_{T_i, s_1} \{\boldsymbol{\alpha}(k\Delta)\} \\
&+ \Big(\frac{\Delta \lambda_{T_i, s_1}\{\boldsymbol{\alpha}(k\Delta)\}}{1 - \exp[-\Delta \lambda_{T_i, s_1}\{\boldsymbol{\alpha}(k\Delta)\}]} \Big) \delta_{T_i}(s_1 + k\Delta + \Delta)] \bigg) \\
&, \bigg(\frac{1}{h_n} K(\frac{s_2 - s_1}{h_n}) I(s_2 \le s_1) \boldsymbol{Z}_i(s_2) [- \Delta \lambda_{T_i, s_2} \{\boldsymbol{\alpha}(k\Delta)\} \\
&+ \Big(\frac{\Delta \lambda_{T_i, s_2}\{\boldsymbol{\alpha}(k\Delta)\}}{1 - \exp[-\Delta \lambda_{T_i, s_2}\{\boldsymbol{\alpha}(k\Delta)\}]} \Big) \delta_{T_i}(s_2 + k\Delta + \Delta)] \bigg) \Bigg\} \\
&= \Delta^2 \sum_{t_1 \in \mathcal{T}_{i,k}} \sum_{s_1 \in \mathcal{S}_{i,k}} \sum_{s_2 \in \mathcal{S}_{i,k}} \frac{1}{h_n^2} K(\frac{s_1 - t_1}{h_n}) I(s_1 \le t_1) K(\frac{s_2 - s_1}{h_n}) I(s_2 \le s_1) \boldsymbol{Z}_i(s_1) \boldsymbol{Z}_i(s_2)^T \\
&\times \mathrm{cov}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)} \Bigg\{ [- \Delta \lambda_{T_i, s_1}\{\boldsymbol{\alpha}(k\Delta)\} + \Big(\frac{\Delta \lambda_{T_i, s_1}\{\boldsymbol{\alpha}(k\Delta)\}}{1 - \exp[-\Delta \lambda_{T_i, s_1}\{\boldsymbol{\alpha}(k\Delta)\}]} \Big) \delta_{T_i}(s_1 + k\Delta + \Delta)] \\
&, [- \Delta \lambda_{T_i, s_2}\{\boldsymbol{\alpha}(k\Delta)\} + \Big(\frac{\Delta \lambda_{T_i, s_2}\{\boldsymbol{\alpha}(k\Delta)\}}{1 - \exp[-\Delta \lambda_{T_i, s_2}\{\boldsymbol{\alpha}(k\Delta)\}]} \Big) \delta_{T_i}(s_2 + k\Delta + \Delta)] \Bigg\} \\
&= \sum_{t_1 \in \mathcal{T}_{i,k}} \int_{-\infty}^{0} \int_{-\infty}^{0} K(z_1) K(z_2) dz_1 dz_2 \boldsymbol{Z}_i(t_1 + z_1 h_n) \boldsymbol{Z}_i\{t_1 + (z_1 + z_2) h_n\}^T \\
&\times [-2\Delta^2 \lambda_{T_i, t_1 + z_1 h_n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\} \lambda_{T_i, \{t_1 + (z_1 + z_2) h_n\}}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}] \\
&= -\frac{\Delta^2}{2} \sum_{t \in \mathcal{T}_{i,k}} \boldsymbol{Z}_i(t) \boldsymbol{Z}_i(t)^T \lambda_{T_i, t}^2 \{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\} + o(\Delta^2)
\end{aligned}
$$

When $t_1 = t_2$,

$$\Delta^2 \sum_{t_1\in\mathcal{T}_{i,k}} \sum_{s_1\in\mathcal{S}_{i,k}} \sum_{s_2\in\mathcal{S}_{i,k}} \mathrm{cov}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)}\Big\{\Big(\frac{1}{h_n}K(\frac{s_1-t_1}{h_n})I(s_1\leq t_1)\boldsymbol{Z}_i(s_1)[-\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}$$
$$+\Big(\frac{\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}]}\Big)\delta_{T_i}(s_1+k\Delta+\Delta)]\Big)$$
$$,\Big(\frac{1}{h_n}K(\frac{s_2-t_1}{h_n})I(s_2\leq t_1)\boldsymbol{Z}_i(s_2)[-\Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\}$$
$$+\Big(\frac{\Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\}]}\Big)\delta_{T_i}(s_2+k\Delta+\Delta)]\Big)\Big\}$$
$$=\Delta^2 \sum_{t_1\in\mathcal{T}_{i,k}} \sum_{s_1\in\mathcal{S}_{i,k}} \sum_{s_2\in\mathcal{S}_{i,k}} \frac{1}{h_n^2}K(\frac{s_1-t_1}{h_n})I(s_1\leq t_1)K(\frac{s_2-t_1}{h_n})I(s_2\leq t_1)\boldsymbol{Z}_i(s_1)\boldsymbol{Z}_i(s_2)^T$$
$$\times\, \mathrm{cov}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)}\Big\{[-\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}+\Big(\frac{\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}]}\Big)\delta_{T_i}(s_1+k\Delta+\Delta)]$$
$$,[-\Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\}+\Big(\frac{\Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\}]}\Big)\delta_{T_i}(s_2+k\Delta+\Delta)]\Big\}$$
$$=\sum_{t_1\in\mathcal{T}_{i,k}} \int_{-\infty}^0\int_{-\infty}^0 K(z_1)K(z_2)dz_1dz_2\boldsymbol{Z}_i(t_1+z_1h_n)\boldsymbol{Z}_i(t_1+z_2h_n)^T$$
$$\times\,[-2\Delta^2\lambda_{T_i,t_1+z_1h_n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}\lambda_{T_i,t_1+z_2h_n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}]+o(\Delta^2)$$
$$=-\frac{\Delta^2}{2}\sum_{t\in\mathcal{T}_{i,k}}\boldsymbol{Z}_i(t)\boldsymbol{Z}_i(t)^T\lambda^2_{T_i,t}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}+o(\Delta^2)$$

When $t_1 = t_2 = s_1 = s_2$

$$\Delta^2\sum_{s_1\in\mathcal{S}_{i,k}}\mathrm{var}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)}\Big\{\Big(\frac{1}{h_n}K(0)\boldsymbol{Z}_i(s_1)[-\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}$$
$$+\Big(\frac{\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}]}\Big)\delta_{T_i}(s_1+k\Delta+\Delta)]\Big)\Big\}$$
$$=\frac{\Delta^3}{h_n^2}\sum_{s\in\mathcal{S}_{i,k}}K^2(0)\boldsymbol{Z}_i(s)\boldsymbol{Z}_i(s)^T\lambda_{T_i,s}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}+o(\frac{\Delta^3}{h_n^2})$$

When $s_1 = t_1$ or $s_2 = t_2$,

$$\Delta^2 \sum_{t_2 \in \mathcal{T}_{i,k}} \sum_{s_1 \in \mathcal{S}_{i,k}} \sum_{s_2 \in \mathcal{S}_{i,k}} \mathrm{cov}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)} \Bigg\{ \Bigg(\frac{1}{h_n} K(0) \boldsymbol{Z}_i(s_1) [-\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}$$

$$+ \Big(\frac{\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}]} \Big) \delta_{T_i}(s_1 + k\Delta + \Delta)] \Bigg)$$

$$, \Bigg(\frac{1}{h_n} K(\frac{s_2 - t_2}{h_n}) I(s_2 \le t_2) \boldsymbol{Z}_i(s_2) [-\Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\}$$

$$+ \Big(\frac{\Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\}]} \Big) \delta_{T_i}(s_2 + k\Delta + \Delta)] \Bigg) \Bigg\}$$

$$= \Delta^2 \sum_{t_2 \in \mathcal{T}_{i,k}} \sum_{s_1 \in \mathcal{S}_{i,k}} \sum_{s_2 \in \mathcal{S}_{i,k}} \frac{1}{h_n^2} K(0) K(\frac{s_2 - t_2}{h_n}) I(s_2 \le t_2) \boldsymbol{Z}_i(s_1) \boldsymbol{Z}_i(s_2)^T$$

$$\times \mathrm{cov}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)} \Bigg\{ [-\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\} + \Big(\frac{\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}]} \Big) \delta_{T_i}(s_1 + k\Delta + \Delta)]$$

$$, [-\Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\} + \Big(\frac{\Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\}]} \Big) \delta_{T_i}(s_2 + k\Delta + \Delta)] \Bigg\}$$

$$= \frac{\Delta}{h_n} \sum_{t_2 \in \mathcal{T}_{i,k}} \sum_{s_1 \in \mathcal{S}_{i,k}} \int_{-\infty}^{0} K(0) K(z_2) dz_2 \boldsymbol{Z}_i(s_1) \boldsymbol{Z}_i(t_2 + z_2 h_n)^T$$

$$\times [-2\Delta^2 \lambda_{T_i,s_1}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\} \lambda_{T_i,t_2+z_2h_n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}]$$

$$= -\frac{\Delta^3}{h_n} \sum_{t \in \mathcal{T}_{i,k}} \sum_{s \in \mathcal{S}_{i,k}} K(0) \boldsymbol{Z}_i(s) \boldsymbol{Z}_i(t)^T \lambda_{T_i,s}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\} \lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}] + o(\frac{\Delta^3}{h_n})$$

When $t_1 = t_2 = s_1$ or $t_1 = t_2 = s_2$,

$$
\begin{aligned}
&\Delta^2 \sum_{s_1 \in \mathcal{S}_{i,k}} \sum_{s_2 \in \mathcal{S}_{i,k}} \mathrm{cov}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)} \Big\{ \Big(\frac{1}{h_n} K(0) \boldsymbol{Z}_i(s_1) [- \Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\} \\
&+ \Big(\frac{\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}]} \Big) \delta_{T_i}(s_1 + k\Delta + \Delta)] \Big) \\
&, \Big(\frac{1}{h_n} K(\frac{s_2 - s_1}{h_n}) I(s_2 \le s_1) \boldsymbol{Z}_i(s_2) [- \Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\} \\
&+ \Big(\frac{\Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\}]} \Big) \delta_{T_i}(s_2 + k\Delta + \Delta)] \Big) \Big\} \\
&= \Delta^2 \sum_{s_1 \in \mathcal{S}_{i,k}} \sum_{s_2 \in \mathcal{S}_{i,k}} \frac{1}{h_n^2} K(0) K(\frac{s_2 - s_1}{h_n}) I(s_2 \le s_1) \boldsymbol{Z}_i(s_1) \boldsymbol{Z}_i(s_2)^T \\
&\times \mathrm{cov}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)} \Big\{ [- \Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\} + \Big(\frac{\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}]} \Big) \delta_{T_i}(s_1 + k\Delta + \Delta)] \\
&, [- \Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\} + \Big(\frac{\Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\}]} \Big) \delta_{T_i}(s_2 + k\Delta + \Delta)] \Big\} \\
&= \frac{\Delta}{h_n} \sum_{s_1 \in \mathcal{S}_{i,k}} \int_{-\infty}^{0} K(0) K(z_1) dz_1 \boldsymbol{Z}_i(s_1) \boldsymbol{Z}_i(s_1 + z_1 h_n)^T \\
&\times [-2\Delta^2 \lambda_{T_i,s_1}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\} \Delta\lambda_{T_i,s_1+z_1h_n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}] + o(\frac{\Delta^3}{h_n}) \\
&= -\frac{\Delta^3}{h_n} \sum_{s \in \mathcal{S}_{i,k}} \boldsymbol{Z}_i(s) \boldsymbol{Z}_i(s)^T \lambda_{T_i,s}^2 \{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\} + o(\frac{\Delta^3}{h_n})
\end{aligned}
$$

When $s_1 = t_1$ and $s_2 = t_2$

$$
\begin{aligned}
&\Delta^2 \sum_{s_1 \in \mathcal{S}_{i,k}} \sum_{s_2 \in \mathcal{S}_{i,k}} \mathrm{cov}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)} \Big\{ \Big(\frac{1}{h_n} K(0) \boldsymbol{Z}_i(s_1) [- \Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\} \\
&+ \Big(\frac{\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}]} \Big) \delta_{T_i}(s_1 + k\Delta + \Delta)] \Big) \\
&, \Big(\frac{1}{h_n} K(0) \boldsymbol{Z}_i(s_2) [- \Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\} \\
&+ \Big(\frac{\Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\}]} \Big) \delta_{T_i}(s_2 + k\Delta + \Delta)] \Big) \Big\} \\
&= \Delta^2 \sum_{s_1 \in \mathcal{S}_{i,k}} \sum_{s_2 \in \mathcal{S}_{i,k}} \frac{1}{h_n^2} K^2(0) \boldsymbol{Z}_i(s_1) \boldsymbol{Z}_i(s_2)^T \\
&\times \mathrm{cov}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)} \Big\{ [- \Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\} + \Big(\frac{\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s_1}\{\boldsymbol{\alpha}(k\Delta)\}]} \Big) \delta_{T_i}(s_1 + k\Delta + \Delta)] \\
&, [- \Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\} + \Big(\frac{\Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s_2}\{\boldsymbol{\alpha}(k\Delta)\}]} \Big) \delta_{T_i}(s_2 + k\Delta + \Delta)] \Big\} \\
&= -\frac{2\Delta^4}{h_n^2} \sum_{s_1 \in \mathcal{S}_{i,k}} \sum_{s_2 \in \mathcal{S}_{i,k}} K^2(0) \boldsymbol{Z}_i(s_1) \boldsymbol{Z}_i(s_2)^T \lambda_{T_i,s_1}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\} \lambda_{T_i,s_2}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}
\end{aligned}
$$

B.4.3 Result 3:

When $\Delta \to 0$, $k\Delta = u$ and $\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}$ being constant within each increment Δ of $t \in \mathcal{T}_{i,k}$,

$$\frac{\partial \ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial\{\boldsymbol{\alpha}(k\Delta)^T\}} - \frac{\partial \ell^{(1)}\{\boldsymbol{\alpha}(u)\}}{\partial \boldsymbol{\alpha}(u)^T} = o(\Delta) \tag{B.4.2}$$

In addition to the above condition, when $h_n \to 0$, $\frac{\Delta}{h_n} \to 0$,

$$\frac{\Delta}{n}\frac{\partial \ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T} - \frac{1}{2n}\frac{\partial \ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T} = o(\Delta) \tag{B.4.3}$$

Proof: Given $\lambda_{T,t}(u) = \exp\{\boldsymbol{\alpha}(u)^{\mathrm{T}}\boldsymbol{Z}(t)\}$ as defined in (3.2.1), the score function of $\ell^{(1)}\{\boldsymbol{\alpha}(u)\}$ takes the form,

$$\begin{aligned}
\frac{\partial \ell^{(1)}\{\boldsymbol{\alpha}(u)\}}{\partial \boldsymbol{\alpha}(u)^T} &= \sum_{i=1}^{n}\left[-\int_{t>u}^{\tau}\frac{\partial \lambda_{T_i,t-u}\{\boldsymbol{\alpha}(u)\}}{\partial \boldsymbol{\alpha}(u)^T}dt + \delta_{T_i}(X_i)\frac{\partial \log[\lambda_{T_i,(X_i-u)}\{\boldsymbol{\alpha}(u)\}]}{\partial \boldsymbol{\alpha}(u)^T}\right] \\
&= \sum_{i=1}^{n}\left[-\int_{t>u}^{\tau}\boldsymbol{Z}_i(t-u)\lambda_{T_i,t-u}dt + \delta_{T_i}(X_i)\boldsymbol{Z}_i(X_i-u)\right] \\
&= \sum_{i=1}^{n}\left[\int_{t>u}^{\tau}\boldsymbol{Z}_i(t-u)[-\lambda_{T_i,t-u}\{\boldsymbol{\alpha}(u)\}dt + dN_{T_i}(t)]\right]
\end{aligned}$$

At the true value $\boldsymbol{\alpha}_0^{(1)}(u)$, we have $-\lambda_{T_i,t-u}\{\boldsymbol{\alpha}_0^{(1)}(u)\}dt+dN_{T_i}(t) = dM_{T_i}(t)$ as a martingale. The score function of $\ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}$ takes the form,

$$\begin{aligned}
&\frac{\partial \ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial\{\boldsymbol{\alpha}(k\Delta)^T\}} \\
&= \sum_{i=1}^{n}\left\{-\sum_{t\in\mathcal{T}_{i,k}}\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{Z}_i(t) + \left(\frac{\Delta\lambda_{T_i,(X_i-k\Delta-\Delta)}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{Z}_i(X_i-k\Delta-\Delta)}{1-\exp\left[-\Delta\lambda_{T_i,(X_i-k\Delta-\Delta)}\{\boldsymbol{\alpha}(k\Delta)\}\right]}\right)\delta_{T_i}(X_i)\right\} \\
&= \sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}}\boldsymbol{Z}_i(t)\left[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\} + \left(\frac{\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp\left[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}\right]}\right)\delta_{T_i}(t+k\Delta+\Delta)\right]
\end{aligned}$$

The second term $\frac{\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp\left[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}\right]} \approx 1$ by the Taylor expansion of the exponential function. Hence we have (B.4.2) when $\Delta \to 0$, $k\Delta = u$ and $\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}$ being constant within each increment Δ of $t \in \mathcal{T}_{i,k}$.

$$
\begin{aligned}
&\frac{\Delta}{n}\frac{\partial \ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T} \\
&= \frac{\Delta}{n}\sum_{i=1}^{n}\Big\{-\sum_{t\in\mathcal{T}_{i,k}}\sum_{s\in\mathcal{S}_{i,k}}\frac{1}{h_n}K(\frac{s-t}{h_n})I(s\le t)\Delta\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{Z}_i(s) \\
&+\sum_{s\in\mathcal{S}_{i,k}}\frac{1}{h_n}K\{\frac{s-(X_i-k\Delta-\Delta)}{h_n}\}I(s\le X_i-k\Delta-\Delta) \\
&\times\Big(\frac{\Delta\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{Z}_i(s)}{1-\exp\big[-\Delta\lambda_{T_i,(s)}\{\boldsymbol{\alpha}(k\Delta)\}\big]}\Big)\delta_{T_i}(X_i)\Big\} \\
&= \frac{\Delta}{n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}}\sum_{s\in\mathcal{S}_{i,k}}\frac{1}{h_n}K(\frac{s-t}{h_n})I(s\le t)\Big\{-\Delta\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{Z}_i(s) \\
&+\Big(\frac{\Delta\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{Z}_i(s)}{1-\exp\big[-\Delta\lambda_{T_i,(s)}\{\boldsymbol{\alpha}(k\Delta)\}\big]}\Big)\delta_{T_i}(s+k\Delta+\Delta)\Big\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}}\int_{-\infty}^{0}K(z)dz\Big[-\boldsymbol{Z}_i(t+h_nz)\Delta\lambda_{T_i,t+h_nz}\{\boldsymbol{\alpha}(k\Delta)\} \\
&+\Big(\frac{\Delta\lambda_{T_i,t+h_nz}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{Z}_i(t+h_nz)}{1-\exp\big[-\Delta\lambda_{T_i,(t+h_nz)}\{\boldsymbol{\alpha}(k\Delta)\}\big]}\Big)\delta_{T_i}(t+h_nz+k\Delta+\Delta)\Big]+o(\Delta) \\
&= \frac{1}{2n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}}\Big[-\boldsymbol{Z}_i(t)\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\} \\
&+\Big(\frac{\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{Z}_i(t)}{1-\exp\big[-\Delta\lambda_{T_i,(t)}\{\boldsymbol{\alpha}(k\Delta)\}\big]}\Big)\delta_{T_i}(t+k\Delta+\Delta)\Big]+o(\Delta)
\end{aligned}
$$

Hence we have (B.4.3) when $h_n \to 0$, $\Delta \to 0$, $\frac{\Delta}{h_n} \to 0$, and $k\Delta = u$.

B.4.4 Result 4:

When $\Delta \to 0$, $k\Delta = u$ and $\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}$ being constant within each increment Δ of $t \in \mathcal{T}_{i,k}$,

$$
\begin{aligned}
\boldsymbol{\Sigma}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\} &= \mathrm{var}_{\boldsymbol{\alpha}_0^{(2)}(k\Delta)}[\frac{1}{\sqrt{n}}\frac{\partial \ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T}] \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}}\boldsymbol{Z}_i(t)\boldsymbol{Z}_i^T(t)\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}\Delta+o(\Delta) \qquad \text{(B.4.4)}
\end{aligned}
$$

$$
\begin{aligned}
&\boldsymbol{\Sigma}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}-\boldsymbol{\Omega}_{1,n}\{\boldsymbol{\alpha}_0^{(1)}(u)\} \qquad \text{(B.4.5)} \\
&= \mathrm{var}_{\boldsymbol{\alpha}_0^{(2)}(k\Delta)}[\frac{1}{\sqrt{n}}\frac{\partial \ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T}]-\langle\frac{1}{\sqrt{n}}\frac{\partial \ell^{(1)}\{\boldsymbol{\alpha}_0^{(1)}(u)\}}{\partial \boldsymbol{\alpha}(u)^T}\rangle = o(\Delta)
\end{aligned}
$$

In addition to above conditions, when $h_n \to 0$, $\frac{\Delta}{h_n} \to 0$,

$$\begin{aligned}\boldsymbol{\Sigma}_{3,n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\} &= \mathrm{var}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)}[\frac{1}{\sqrt{n}}\frac{\partial \ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial \boldsymbol{\alpha}^T(k\Delta)}]\Delta^2 \\ &= \frac{1}{4n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}} \boldsymbol{Z}_i(t)\boldsymbol{Z}_i(t)^T\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}\Delta + o(\Delta) \end{aligned} \tag{B.4.6}$$

$$\begin{aligned}&\boldsymbol{\Sigma}_{3,n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\} - \frac{1}{4}\boldsymbol{\Sigma}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\} \\ &= \mathrm{var}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)}[\frac{1}{\sqrt{n}}\frac{\partial \ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial \boldsymbol{\alpha}^T(k\Delta)}]\Delta^2 - \frac{1}{4}\mathrm{var}_{\boldsymbol{\alpha}_0^{(2)}(k\Delta)}[\frac{1}{\sqrt{n}}\frac{\partial \ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial \boldsymbol{\alpha}(k\Delta)^T}] = o(\Delta)\end{aligned} \tag{B.4.7}$$

Proof: Given $\lambda_{T,t}(u) = \exp\{\boldsymbol{\alpha}(u)^{\mathrm{T}}\boldsymbol{Z}(t)\}$ as defined in (3.2.1),

$$\boldsymbol{\Omega}_{1,n}\{\boldsymbol{\alpha}_0^{(1)}(u)\} = \langle\frac{1}{\sqrt{n}}\frac{\partial \ell^{(1)}\{\boldsymbol{\alpha}_0^{(1)}(u)\}}{\partial \boldsymbol{\alpha}(u)^T}\rangle = \frac{1}{n}\sum_{i=1}^{n}\left[\int_{t>u}^{\tau}\boldsymbol{Z}_i(t-u)\boldsymbol{Z}_i^T(t-u)\lambda_{T_i,t-u}\{\boldsymbol{\alpha}_0^{(1)}(u)\}dt\right]$$

Based on Result 1 in the section of the technical proofs, the variance $\boldsymbol{\Sigma}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}$ can be expressed as,

$$\begin{aligned}&\boldsymbol{\Sigma}_{2,n}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{var}_{\boldsymbol{\alpha}_0^{(2)}(k\Delta)}\Bigg(\sum_{t\in\mathcal{T}_{i,k}}\boldsymbol{Z}_i(t)\Big[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\} \\
&+\Big(\frac{\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp\left[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}\right]}\Big)\delta_{T_i}(t+k\Delta+\Delta)\Big]\Bigg) \\
&= \frac{1}{n}\sum_{i=1}^{n}\Bigg(\sum_{t\in\mathcal{T}_{i,k}}\boldsymbol{Z}_i(t)\boldsymbol{Z}_i(t)^T\mathrm{var}_{\boldsymbol{\alpha}_0^{(2)}(k\Delta)}\Big[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}+ \\
&\Big(\frac{\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp\left[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}\right]}\Big)\delta_{T_i}(t+k\Delta+\Delta)\Big] \\
&+2\sum_{t\neq t',t,t'\in\mathcal{T}_{i,k}}\boldsymbol{Z}_i(t)\boldsymbol{Z}_i(t')^T\mathrm{cov}_{\boldsymbol{\alpha}_0^{(2)}(k\Delta)}\Big[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}+ \\
&\Big(\frac{\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp\left[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}\right]}\Big)\delta_{T_i}(t+k\Delta+\Delta), \\
&-\Delta\lambda_{T_i,t'}\{\boldsymbol{\alpha}(k\Delta)\}+\Big(\frac{\Delta\lambda_{T_i,t'}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp\left[-\Delta\lambda_{T_i,t'}\{\boldsymbol{\alpha}(k\Delta)\}\right]}\Big)\delta_{T_i}(t'+k\Delta+\Delta)\Big]\Bigg) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}}\boldsymbol{Z}_i(t)\boldsymbol{Z}_i^T(t)\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(2)}(k\Delta)\}\Delta + o(\Delta)\end{aligned}$$

Hence we have (B.4.4) and (B.4.5) when $\Delta \to 0$, $k\Delta = u$ and $\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}$ being constant within each increment Δ of $t \in \mathcal{T}_{i,k}$. When $h_n \to 0$, $\Delta \to 0$, $\frac{\Delta}{h_n} \to 0$ and $k\Delta = u$, based on Result 2 and 3 in the section of the technical proofs we have (B.4.6),

$$\boldsymbol{\Sigma}_{3,n}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\} = \text{var}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)}[\sqrt{n}\frac{1}{n}\frac{\partial\ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial\boldsymbol{\alpha}^T(k\Delta)}]\Delta^2$$
$$= \frac{1}{n}\sum_{i=1}^{n}\text{var}_{\boldsymbol{\alpha}_0^{(3)}(k\Delta)}\Bigg(\sum_{t\in\mathcal{T}_{i,k}}\sum_{s\in\mathcal{S}_{i,k}}\frac{1}{h_n}K(\frac{s-t}{h_n})I(s\leq t)\boldsymbol{Z}_i(s)[-\Delta\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\}$$
$$+\Big(\frac{\Delta\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\}}{1-\exp[-\Delta\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\}]}\Big)\delta_{T_i}(s+k\Delta+\Delta)]\Bigg)$$
$$= \frac{1}{4n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}}\boldsymbol{Z}_i(t)\boldsymbol{Z}_i(t)^T\lambda_{T_i,t}\{\boldsymbol{\alpha}_0^{(3)}(k\Delta)\}\Delta + o(\Delta)$$

and (B.4.7).

B.4.5 Result 5:

When $\Delta \to 0$, $k\Delta = u$ and $\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}$ being constant within each increment Δ of $t \in \mathcal{T}_{i,k}$,

$$\mathbf{W}_{2,n}\{\boldsymbol{\alpha}(k\Delta)\} = -\frac{1}{n}\frac{\partial\ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial\{\boldsymbol{\alpha}(k\Delta)\}^T\partial\{\boldsymbol{\alpha}(k\Delta)\}}$$
$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}}\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{Z}_i(t)\boldsymbol{Z}_i^T(t)\Delta\Big[1-\delta_{T_i}(t+k\Delta+\Delta)\Big] + o(\Delta) \quad \text{(B.4.8)}$$

$$\mathbf{W}_{2,n}\{\boldsymbol{\alpha}(k\Delta)\} - \boldsymbol{\Omega}_{1,n}\{\boldsymbol{\alpha}(u)\}$$
$$= -\frac{1}{n}\frac{\partial\ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)^T\boldsymbol{\alpha}(k\Delta)\}}{\partial\{\boldsymbol{\alpha}(k\Delta)\}^T\partial\{\boldsymbol{\alpha}(k\Delta)\}} + \frac{1}{n}\frac{\partial^2\ell^{(1)}\{\boldsymbol{\alpha}(u)\}}{\partial\boldsymbol{\alpha}(u)^T\partial\boldsymbol{\alpha}(u)} = o(\Delta). \quad \text{(B.4.9)}$$

In addition to the above conditions, when $h_n \to 0$,

$$\mathbf{W}_{3,n}\{\boldsymbol{\alpha}(k\Delta)\} = -\frac{\Delta}{n}\frac{\partial^2\ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial\boldsymbol{\alpha}(k\Delta)^T\partial\boldsymbol{\alpha}(k\Delta)} \quad \text{(B.4.10)}$$
$$= \frac{1}{2n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}}\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{Z}_i(t)\boldsymbol{Z}_i^T(t)\Delta\Big[1-\delta_{T_i}(t+k\Delta+\Delta)\Big] + o(\Delta)$$

$$\mathbf{W}_{3,n}\{\boldsymbol{\alpha}(k\Delta)\} - \frac{1}{2}\mathbf{W}_{2,n}\{\boldsymbol{\alpha}(k\Delta)\} \quad \text{(B.4.11)}$$
$$= -\frac{\Delta}{n}\frac{\partial^2\ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial\boldsymbol{\alpha}(k\Delta)^T\partial\boldsymbol{\alpha}(k\Delta)} + \frac{1}{2n}\frac{\partial\ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)^T\boldsymbol{\alpha}(k\Delta)\}}{\partial\{\boldsymbol{\alpha}(k\Delta)\}^T\partial\{\boldsymbol{\alpha}(k\Delta)\}} = o(\Delta).$$

Proof: Given $\lambda_{T,t}(u) = \exp\{\boldsymbol{\alpha}(u)^{\mathrm{T}}\boldsymbol{Z}(t)\}$ as defined in (3.2.1),

$$\boldsymbol{\Omega}_{1,n}\{\boldsymbol{\alpha}_0^{(1)}(u)\} = -\frac{1}{n}\frac{\partial^2 \ell^{(1)}\{\boldsymbol{\alpha}(u)\}}{\partial\boldsymbol{\alpha}(u)^T\partial\boldsymbol{\alpha}(u)}$$
$$= -\frac{1}{n}\sum_{i=1}^{n}\left[-\int_{t>u}^{\tau}\frac{\partial^2\lambda_{T_i,t-u}\{\boldsymbol{\alpha}(u)\}}{\partial\boldsymbol{\alpha}(u)^T\partial\boldsymbol{\alpha}(u)}dt + \int_{t>u}^{\tau}dN_{T_i}(t+u)\frac{\partial^2\log[\lambda_{T_i,(t-u)}\{\boldsymbol{\alpha}(u)\}]}{\partial\boldsymbol{\alpha}(u)^T\partial\boldsymbol{\alpha}(u)}\right]$$
$$= \frac{1}{n}\sum_{i=1}^{n}\int_{t>u}^{\tau}\boldsymbol{Z}_i(t-u)\boldsymbol{Z}_i^T(t-u)\lambda_{T_i,t-u}\{\boldsymbol{\alpha}(u)\}dt$$

When $\Delta \to 0$,

$$\mathbf{W}_{2,n}\{\boldsymbol{\alpha}^{(2)}(k\Delta)\} = -\frac{1}{n}\frac{\partial^2\ell^{(2)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial\{\boldsymbol{\alpha}(k\Delta)\}^T\partial\{\boldsymbol{\alpha}(k\Delta)\}}$$
$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}}\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{Z}_i(t)\boldsymbol{Z}_i^T(t)\Delta\times\Bigg[1-\delta_{T_i}(t+k\Delta+\Delta)$$
$$\times\frac{(1-\exp[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}])-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}\exp[-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}]}{[1-\exp\{-\Delta\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}\}]^2}\Bigg]$$
$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}}\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{Z}_i(t)\boldsymbol{Z}_i^T(t)\Delta\Big[1-\delta_{T_i}(t+k\Delta+\Delta)\Big]+o(\Delta)$$

Hence we have (B.4.9) when $\Delta \to 0$, $k\Delta = u$ and $\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}$ being constant within each increment Δ of $t \in \mathcal{T}_{i,k}$.

$$\mathbf{W}_{3,n}\{\boldsymbol{\alpha}(k\Delta)\} = -\frac{\Delta}{n}\frac{\partial^2\ell^{(3)}\{\boldsymbol{\alpha}(k\Delta)\}}{\partial\boldsymbol{\alpha}(k\Delta)^T\partial\boldsymbol{\alpha}(k\Delta)}$$
$$= -\frac{\Delta}{n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}}\sum_{s\in\mathcal{S}_{i,k}}\frac{1}{h_n}K(\frac{s-t}{h_n})I(s\le t)$$
$$\times\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{Z}_i(s)\boldsymbol{Z}_i^T(s)\Delta\Bigg[-1+\delta_{T_i}(s+k\Delta+\Delta)$$

$$\times\frac{(1-\exp[-\Delta\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\}])-\Delta\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\}\exp[-\Delta\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\}]}{[1-\exp\{-\Delta\lambda_{T_i,s}\{\boldsymbol{\alpha}(k\Delta)\}\}]^2}\Bigg]$$
$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}}\int_{-\infty}^{0}K(z)dz\lambda_{T_i,t+zh_n}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{Z}_i(t+zh_n)\boldsymbol{Z}_i^T(t+zh_n)\Delta$$
$$\times\Big[1-\delta_{T_i}(t+zh_n+k\Delta+\Delta)\Big]+o(\Delta)$$
$$= \frac{1}{2n}\sum_{i=1}^{n}\sum_{t\in\mathcal{T}_{i,k}}\lambda_{T_i,t}\{\boldsymbol{\alpha}(k\Delta)\}\boldsymbol{Z}_i(t)\boldsymbol{Z}_i^T(t)\Delta\Big[1-\delta_{T_i}(t+k\Delta+\Delta)\Big]+o(\Delta)$$

Hence when $\Delta \to 0$, $h_n \to 0$, and $k\Delta = u$, we have (B.4.11).