ESTIMATION AND INFERENCE FOR DYNAMIC INTENSITY MODELS FOR RECURRENT EVENT DATA WITH APPLICATIONS TO A MALARIA TRIAL

by

Jing Xu

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Applied Mathematics

Charlotte

2024

Approved by:

_____
Dr. Yanqing Sun

_____
Dr. Qingning Zhou

_____
Dr. Yinghao Pan

_____
Dr. Ronald Sass

ABSTRACT

JING XU. ESTIMATION AND INFERENCE FOR DYNAMIC INTENSITY MODELS FOR RECURRENT EVENT DATA WITH APPLICATIONS TO A MALARIA TRIAL. (Under the direction of DR. YANQING SUN)

Recurrent events are commonly encountered in medical and epidemiological studies. It is often of interest what and how risk factors influence the occurrence of events. While many research of recurrent events address both time-independent and time-dependent effects, there is a possibility that these effects also vary with certain covariates. In this dissertation, we develop novel estimation and inference procedures of two intensity models for recurrent event data. Both models allow for the simultaneous measurement of time-varying and covariate-varying effects, with covariates potentially depend on event history.

We firstly consider a class of semiparametric models, the models feature unspecific time-varying effects, while covariate-varying and event history effects are modeled parametrically. The models offer much flexibility through the choice of different link functions and parametric functions. Estimation procedures involve local linear method and profile log-likelihood method, with asymptotic properties of estimators explored using martingale theory and empirical processes. Two hypothesis tests have been developed to assess the parametric functions of the covariate-varying effects, one being a supremum type test and the other a chi-square type test. Both tests are based on residual processes.

Secondly, we propose a nonparametric intensity model with frailty. The model includes unspecified time-varying and covariate-varying effects and we introduce a frailty term for each individual, following a Gamma distribution, which operates multiplicatively on the intensity function. Estimation of nonparametric functions and the parameters governing the Gamma frailty involves using an Expectation-Maximization (EM) algorithm and local linear estimation techniques. Variance estimators are ob-

tained through a weighted bootstrap procedure. Simulation studies demonstrate the satisfactory performance of this method.

Both of the proposed models have been applied to a malaria vaccine efficacy trial (MAL-094) to assess the efficacy of the RTS,S/AS01$_{\mathrm{E}}$ vaccine and examine how the efficacy varies over the gap time since the most recent infection or vaccination.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my advisor, Dr. Yanqing Sun, for her invaluable guidance, suggestions, encouragement, and patience throughout my research journey. I am truly grateful for her financial support, including the research assistant positions during both semesters and summers. Most importantly, I deeply appreciate her unwavering assistance and concern, not only in my academic endeavors but also in my job search and personal life.

I would like to express my gratitude to my committee members, Dr. Qingning Zhou, Dr. Yinghao Pan and Dr. Ronald Sass, for their valuable time, insightful comments, and helpful suggestions on my thesis. Special thanks go to Dr.Zhou for the abundant encouragement provided. Additionally, I extend my sincere thanks to Dr. Fei Heng for his assistance and engaging discussions on these projects. I appreciate the support from Dr. Shaozhong Deng and Dr. Mohammad A. Kazemi during my TA responsibilities.

I wish to extend my gratitude to the Graduate School for their financial support, including the fellowship provided during 2023 summer. This research was partially supported by the National Institutes of Health NIAID grant R37-AI054165, the National Science Foundation grant DMS-1915829, and a subaward from Fred Hutchinson Cancer Center, entitled "Longitudinal modeling of the dependency of the intensity of malaria infection on previous vaccinations and infection history". Many thanks to these institutions for their support.

Thanks to my friends, Yang Song, Xue Tan and Jian Wang. I would also like to thanks to my classmate, Xu Cao, with whom I have went through every milestone together since the Qualifying Exam.

Lastly, I would like to express my love and gratitude to my parents, Dongcheng Xu and Limei Liu, for their unconditional love and understanding.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Recurrent events refer to the events that can occur multiple times during a particular time period. They are often observed in medical studies where patients encounter events repetitively, such as hospital admissions, cancer recurrences, and infections of Covid-19 and many others. It is typically of interest to understand what and how the risk factors would influence the events. Evaluating the effects of risk factors and analyzing how these effects may change over time help us unravel the underlying mechanisms of the events.

1.1 Motivating Example - MAL-094 Trial Study

This dissertation is motivated by the complex challenges posed by the MAL-094 malaria vaccine efficacy trial. Malaria is a life-threatening disease with diverse genetic strains, adults and children can experience multiple malaria infections during their lifetime.

The MAL-094 trial is conducted by Glaxo SmithKline Biologicals (GSK) and PATH Malaria Vaccine Initiative, testing the RTS,S/AS01$_{\mathrm{E}}$ malaria vaccine. It took place in Sub-Saharan Africa from 2017 to 2022, randomly divided approximately 1500 children aged 5 to 17 months from two sites (Agogo in Ghana, and Siaya in Kenya) into five arms, each arm has around 300 participants. Four arms received vaccine versions administered at different doses and schedules and one arm served as control group receiving placebo. Children's vaccination and infections status over 20 months and 32 months follow-up periods have been recorded. The primary objectives of our work are to measure the efficacy of the RTS,S/AS01$_{\mathrm{E}}$ vaccine and investigate whether and how prior infections can confer protection against subsequent ones.

1.2 Literature Review

In this section, we provide an overview of relevant literature concerning the analysis of recurrent event data.

The most two commonly used approaches to model recurrent events are marginal methods and conditional methods. Both of them have been intensively studied, including statistical modelings and inference procedures.

Marginal methods model the population average behaviors of the recurrent event, focus on the overall effects and trends, rather than individual characteristics. Wei *et al.* (1989) analyzed multivariate failure time data, they used Cox proportional hazard models to model the marginal distribution of each failure time without imposing any structure of dependence among the failure times for each individual. Pepe and Cai (1993) proposed two rate functions to model the first infection and recurrent infection separately, providing likelihood-based estimating equations. Lawless (1997) proposed a semiparametric procedures to model the mean or rate function of the counting process, Lin *et al.* (2000) justified the inference procedure through empirical process theory and constructed confidence bands for the mean functions.

Rather than modeling the overall population behavior, conditional method can model the pattern based on event history for each individual. Conditional methods usually work on counting process, provide a flexible framework by modeling the intensity function of a counting process. The intensity of a counting process is the instantaneous rate of an event occurring at that time point, given the all the event and covariate history. Denote $\mathcal{F}_t$ as the filtration generated by the counting process $N_i(t)$ and possible covariate processes up to time t, the intensity function of counting process $N_i(t)$ is defined as

$$\lambda_i(t) = \lim_{\Delta t \to 0} \frac{Pr(\Delta N_i(t) = 1 | \mathcal{F}_t)}{\Delta t},$$

where $\Delta N_i(t) = N_i(t + \Delta t^-) - N_i(t^-)$ is the number of events in the time interval $[t, t + \Delta t]$.

A counting process is deemed to be of the Poisson type if, for non-overlapping time intervals, the number of events within these intervals is statistically independent. The recurrent event processes characterized by a constant intensity are referred to as homogeneous Poisson processes, while those with time-dependent intensity functions are termed inhomogeneous Poisson processes.

There are extensive work on modeling the intensity of the Poisson-type counting process. Andersen and Gill (1982) generalized the Cox model, modeling the intensity function of the recurrent events by the following function

$$\lambda(t) = \lambda_0(t) \exp\{\beta^\mathsf{T} X(t)\},$$

where $X(t)$ is a vector of possibly time-dependent covariates, $\lambda_0(t)$ is an unspecified baseline intensity function and β is a vector of unknown regression parameter.

Zeng and Lin (2006) proposed the following semiparametric transformation models

$$\Lambda_Z(t) = G\Big\{ \int_0^t Y^*(s) \exp^{\beta^\mathsf{T} Z(s)} d\Lambda(s) \Big\},$$

where $Z(\cdot)$ is a vector of possibly time-varying covariates, β is a vector of unknown parameters, $Y^*(\cdot)$ is the at risk indicator, $\Lambda(\cdot)$ is an unspecified increasing function. The transformation function $G(\cdot)$ provide much flexibility of the models and the estimated regression parameters β and cumulative intensity functions $\Lambda(\cdot)$ are obtained through non-parametric maximum likelihood method.

Gap time, also know as waiting time, refer to the time between two consecutive events for a particular subject. It is of natural interest to incorporating the gap times in the model to help us understand the intra-individual correlation. Prentice *et al.* (1981) proposed two classes of stratified proportional intensity function, one model

incorporated the baseline intensity as a function of time since enrollment, while the other model included the baseline function as a function of time since the most recent event. Chang (2004) considered an accelerated failure time (AFT) model, which assume the individual specific frailty, the covariate effects and the random errors act additively on the logarithm of gap time. Other works related with gap time includes Oakes and Cui (1994), Pena *et al.* (2001), Strawderman (2005) and some others.

In recurrent event models, covariates could be either fixed or time-varying. Martinussen and Scheike (1999) worked on a semiparametric additive model on longitudinal data, this model permits certain covariate effects to remain constant, while others vary nonparametrically with time, estimation are conducted by introducing marked point process and local estimating equations. Cai and Sun (2003) proposed a local linear method to estimate the time-varying coefficient in Cox regression model; Amorim *et al.* (2008) incorporated B splines method in a rates model for recurrent events to estimate the time-dependent coefficient. Sun *et al.* (2009) developed a marginal modeling approach on a multivariate recurrent event model. More local modeling methods could be referred to Fan and Gijbels (1996).

Another scenario arises where the effects are dependent on time-varying covariates. For instance, there might be a time delay for a treatment to take effect, or the impact may diminish after a certain exposure duration. Qi *et al.* (2017) studied a generalized semiparametric varying-coefficient model for longitudinal data, which can model time-independent effects, time-varying effects and covariate-varying effects at the same time, estimation procedure are based on profile weighted least squares.

Frailty models, also called random effect models, are able to account for unobserved heterogeneity in a population and induce dependence among the recurrent event times within subjects by introducing a random variable in the model. Lawless (1987) is an early work on the frailty models, it incorporated the random effects in the intensity function $\lambda_i(t) = \lambda_0(t)\exp\{\alpha_i + X_i'\beta\}$, where α_i are independent and identically dis-

tributed random variables. Gamma frailty is commonly used in frailty models by the conjugate properties, Nileson *et al.* (1992) introduced gamma frailty, which act multiplicatively on the intensity function, Klein (1992) specified the estimation procedures based on EM algorithm, Murphy proved the consistency and asymptotic properties of the gamma frailty model without covariates in Murphy (1994) and Murphy (1995), Parner (1998) extended the theories to the correlated gamma frailty models with covariates.

Zeng and Lin (2007) incorporated random effect denoted as b within a class of semiparametric transformation models. This extension is a generalization of the transformation models to multivariate failure times. In a separate work, Zeng *et al.* (2009) propose a different class of transformation models that incorporates the gamma frailty while also allowing the random effects to take the value of 0. There are also some work incorporated local smoothing methods into frailty frameworks, includes Yu *et al.* (2013), Chen *et al.* (2013), Mazroui *et al.* (2015) among others.

In this dissertation, we proposed two dynamic intensity models for recurrent event data, both of which can measure the time-varying and covariate-varying effects simultaneously. It is arranged as follows: In chapter 2, we discuss a generalized class of semiparametric dynamic intensity models, including the estimation procedure, testing the parametric functions, simulation studies and the application to the 20 months follow-up MAL-094 trial data. In chapter 3, we investigated the nonparametric dynamic intensity models with frailty, and apply the models to the 32 months MAL-094 trial data. Chapter 4 concludes the entire work and provides an overview of future work.

CHAPTER 2: GENERALIZED SEMIPARAMETRIC INTENSITY MODELS FOR RECURRENT EVENT DATA

2.1 Introduction

In this chapter, we introduce a class of generalized semiparametric intensity models. Semiparametric models offer several advantages. Firstly, in comparison to nonparametric models, they require less data for fitting. Additionally, if we possess some background information about the parametric forms, it can aid in enhancing our understanding of the data.

The proposed models feature unspecific time-varying effects and constant effects, while the effects that depend on time-varying covariates or event history are modeled parametrically. Section 2.2 describes the models and illustrates their estimation methods, including the computational algorithm and the selection of kernel functions and bandwidths. The asymptotic properties of estimators are developed in Section 2.3. Two hypothesis tests based on residual processes are developed in Section 2.4 to assess the parametric functions of the covariate-varying effects. Simulation studies in Section 2.5 show that the methods perform well in finite samples. Lastly, in Section 2.6, we apply the models on the 20 months follow-up data from MAL-094 malaria vaccine trial.

2.2 Model and Estimation

2.2.1 Model Description

Consider a random sample of n subjects, τ is the end of study time. Suppose for subject i, T_{ij} represents the occurrence time for jth event. If we denote n_i as the total event time for i subject during the study time, we have $T_{i1} < T_{i2} < ... < T_{in_i} \leq \tau$.

$X_i(t)$, $Z_i(t)$, $W_i(t)$ and $U_i(t)$ serve as subject-specific covariates, all of which could be time-dependent. Counting process $N_i^*(t) = \sum_{j=1}^{n_i} I(T_{ij} \leq t)$ is the number of events taken from ith subject by time t; denote $\Delta N_i^*(t) = N_i^*(t + \Delta t^-) - N_i^*(t)$ as the number of events occurring in the small time interval $[t, t + \Delta t)$. Modeling of recurrent events can be based on the intensity function of $N_i^*(t)$, which is defined as $\lambda_i(t) = \lim_{\Delta t \downarrow 0} \Pr(\Delta N_i^*(t) = 1 | \mathcal{F}_{it-}^*)/\Delta t$. $\lambda_i(t)dt$ is the instantaneous probability of an event occurring in $[t, t + \Delta t)$, i.e. $E(dN_i^*(t)|\mathcal{F}_{it-}^*) = \lambda_i(t)dt$, where $\mathcal{F}_{it-}^*$ is the filtration generated by $N_i^*(t)$ and the history of covariates for ith subject up to time t.

Let $\tau_i = \min\{\tau, C_i\}$, where C_i is the non-informative censoring time for subject i. Events for subject i can only be observed before τ_i. $Y_i(t) = I(\tau_i \geq t)$ is the at-risk process, indicates whether subject i is exposed to the event at time t. $N_i(t) = N_i^*(t \wedge \tau_i)$ is the observed counting process, $\mathcal{F}_{it-}$ is the filtration generated by the observed event history, covariate processes and censoring for subject i, thus we have $E(dN_i(t)|\mathcal{F}_{it-}) = Y_i(t)\lambda_i(t)dt$. Censorings are non-informative in the sense of $E\{dN_i(t)|\mathcal{F}_{it-}^*\} = E\{dN_i(t)|\mathcal{F}_{it-}\} = Y_i(t)\lambda_i(t)dt$.

We propose the following generalized semiparametric dynamic intensity models:

$$\lambda_i(t) = g^{-1}\{\alpha^{\mathsf{T}}(t)X_i(t) + \beta^{\mathsf{T}}Z_i(t) + \gamma^{\mathsf{T}}(U_i(t), \theta)W_i(t)\}, \tag{2.1}$$

for $0 \leq t \leq \tau$, where $\alpha(\cdot)$ is a p_1 dimensional vector, each element is an unspecified function; β is a p_2 dimensional vector of unknown time-independent parameters; $\gamma(U_i(t), \theta)$ is a p_3 dimensional vector, with each element is a parametric function defined on the range of $U_i(\cdot)$.

$g(\cdot)$ is a known link function and offers a lot of flexibility to the models. The logarithm link function yields a multiplicative intensity model, whereas choosing the identity link results in an additive intensity model.

Setting the first component of $X_i(t)$ equal 1 provides us with the nonparametric baseline function. $U_i(t)$ can be related with the event or treatment history. For example, $U_i(t) = t - T_{iN_i(t^-)}$ stands for the time since the most recent event; if we denote $V_i(t)$ as the most recent vaccination time, then $U_i(t) = t - V_i(t)$ signifies the time since the most recent vaccination.

For the sake of clarity in representation, we denote $\eta = (\beta^\mathsf{T}, \theta^\mathsf{T})^\mathsf{T}$, $\zeta(U_i(t), \eta) = (\beta^\mathsf{T}, \gamma^\mathsf{T}(U_i(t), \theta))^\mathsf{T}$ and $P_i(t) = (Z_i(t)^\mathsf{T}, W_i(t)^\mathsf{T})^\mathsf{T}$, then the intensity function (2.1) can be written as

$$\lambda_i(t) = g^{-1}\{\alpha^\mathsf{T}(t)X_i(t) + \zeta^\mathsf{T}(U_i(t), \eta)P_i(t)\}. \tag{2.2}$$

2.2.2 Estimation Procedure

Assume $\alpha(\cdot)$ is smooth enough on $t \in [0, \tau]$ and its first derivative $\dot{\alpha}(t)$ exists. Denote $\mathcal{N}_{t_0}$ as a neighbourhood of t_0, for $t \in \mathcal{N}_{t_0}$, by first order Taylor approximation, we have

$$\alpha(t) = \alpha(t_0) + \dot{\alpha}(t_0)(t - t_0) + O((t - t_0)^2).$$

The intensity function (2.2) can be approximated by

$$\lambda_i^*(t, \alpha^*, \eta | t_0) = g^{-1}\{\alpha^{*\mathsf{T}}(t_0)X_i^*(t|t_0) + \zeta^\mathsf{T}(U_i(t), \eta)P_i(t)\}, \tag{2.3}$$

where $\alpha^*(t_0) = (\alpha^\mathsf{T}(t_0), \dot{\alpha}^\mathsf{T}(t_0))^\mathsf{T}$ and $X_i^*(t|t_0) = (X_i^\mathsf{T}(t), X_i^\mathsf{T}(t)(t - t_0))^\mathsf{T}$.

By Cook and Lawless (2007), for fixed η, the likelihood function for the observed data can be constructed as follows:

$$\begin{aligned}\mathcal{L}_\alpha(\alpha, \eta) &= \prod_{0 \le t \le \tau} \Big[\Big\{ \prod_{i=1}^n \{Y_i(t)\lambda_i(t)\}^{dN_i(t)} \Big\} \Big\{ 1 - \sum_{i=1}^n Y_i(t)\lambda_i(t)dt \Big\}^{1 - dN_{i\cdot}(t)} \Big] \\ &= \Big\{ \prod_{0 \le t \le \tau} \prod_{i=1}^n \{Y_i(t)\lambda_i(t)\}^{dN_i(t)} \Big\} \exp\Big\{ -\sum_{i=1}^n \int_0^\tau Y_i(t)\lambda_i(t)dt \Big\},\end{aligned}$$

where $N_{i.}(t) = \sum_i^n N_i(t)$.

Take logarithm, we obtain the log-likelihood function for observed data:

$$\ell_\alpha(\alpha, \eta) = \sum_{i=1}^{n} \int_0^\tau \Big\{ \log\{Y_i(t)\lambda_i(t)\} dN_i(t) - Y_i(t)\lambda_i(t)dt \Big\} \tag{2.4}$$

Apply local smoothing method and plug in the approximated intensity function (2.3), the localized log-likelihood can be written as:

$$\begin{aligned}\ell_\alpha(\alpha^*; \eta, t_0) = \textstyle\sum_{i=1}^n \int_0^\tau K_h(t - t_0) \Big\{ & \log\{Y_i(t)\lambda_i^*(t, \alpha^*, \eta|t_0)\} dN_i(t) \\ & - Y_i(t)\lambda_i^*(t, \alpha^*, \eta|t_0)dt \Big\}\end{aligned} \tag{2.5}$$

where $K_h(.) = K(./h)/h$, $K(.)$ is a kernel function and h is the bandwidth parameter.

Take derivative of (2.5) with respect to $\alpha^*(t_0)$, the score function for $\alpha^*(t_0)$ for fixed η can be written as:

$$U_\alpha(\alpha^*; \eta, t_0) = \sum_{i=1}^{n} \int_0^\tau K_h(t - t_0) X_i^*(t|t_0) \Big\{ \frac{\dot{\lambda}_i^*(t, \alpha^*, \eta|t_0)}{\lambda_i^*(t, \alpha^*, \eta|t_0)} dN_i(t) - Y_i(t)\dot{\lambda}_i^*(t, \alpha^*, \eta|t_0)dt \Big\}. \tag{2.6}$$

Set $U_\alpha(\alpha^*; \eta, t_0) = \mathbf{0}$ and denote the solution as $\tilde{\alpha}^*(t_0, \eta)$. Let $\tilde{\alpha}(t.\eta)$ be the first p_1 components of $\tilde{\alpha}^*(t, \eta)$ and $\tilde{\lambda}_i(t, \eta)$ be the corresponding estimated intensity, i.e., $\tilde{\lambda}_i(t, \eta) = g^{-1}\{\tilde{\alpha}^\mathsf{T}(t, \eta)X_i(t) + \zeta^\mathsf{T}(U_i(t), \eta)P_i(t)\}$. The profile log-likelihood function for η can be written as:

$$\ell_\eta(\eta) = \textstyle\sum_{i=1}^n \int_{t_1}^{t_2} \Big\{ \log\{Y_i(t)\tilde{\lambda}_i(t, \eta)\} dN_i(t) - Y_i(t)\tilde{\lambda}_i(t, \eta)dt \Big\}, \tag{2.7}$$

where t_1 and t_2 are neighbour points of 0 and τ, we integrate on $[t_1, t_2] \subset [0, \tau]$ to avoid boundary effects.

Take derivative of (2.7) with respect to η, we get the profile score function for η,

$$U_\eta(\eta) = \sum_{i=1}^{n} \int_{t_1}^{t_2} \Big\{ \Big(\frac{\partial \tilde{\alpha}(t,\eta)}{\partial \eta} \Big)^{\mathsf{T}} X_i(t) + \Big(\frac{\partial \zeta(U_i(t),\eta)}{\partial \eta} \Big)^{\mathsf{T}} P_i(t) \Big\} \Big[\frac{\dot{\tilde{\lambda}}_i(t,\eta)}{\tilde{\lambda}_i(t,\eta)} dN_i(t) - Y_i(t) \dot{\tilde{\lambda}}_i(t,\eta) dt \Big], \tag{2.8}$$

where $\frac{\partial \tilde{\alpha}(t.\eta)}{\partial \eta}$ are the first p_1 rows of

$$\frac{\partial \tilde{\alpha}^*(t,\eta)}{\partial \eta} = -\Big\{ \frac{\partial U_\alpha(\alpha^*;\eta,t)}{\partial \alpha^*} \Big\}^{-1} \frac{\partial U_\alpha(\alpha^*;\eta,t)}{\partial \eta} \Big|_{\alpha^* = \tilde{\alpha}^*(t,\eta)}, \tag{2.9}$$

equation (2.9) is derived through taking derivative of $U_\alpha(\alpha^*;\eta,t) = \mathbf{0}$ with respect to η on both sides,

$$\frac{\partial U_\alpha(\alpha^*;\eta,t)}{\partial \alpha^*} \times \frac{\partial \alpha^*(t,\eta)}{\partial \eta} + \frac{\partial U_\alpha(\alpha^*;\eta,t)}{\partial \eta} = 0,$$

where

$$\frac{\partial U_\alpha(\alpha^*;\eta,t)}{\partial \alpha^*} = \sum_{i=1}^{n} \int_0^\tau K_h(s-t) X_i^*(s|t) X_i^{*\mathsf{T}}(s|t) \Big\{ \frac{\ddot{\lambda}_i^*(s,\tilde{\alpha}^*(s),\eta|t)\lambda_i^*(s,\tilde{\alpha}^*(s),\eta|t) - [\dot{\lambda}_i^*(s,\tilde{\alpha}^*(s),\eta|t)]^2}{[\lambda_i^*(s,\tilde{\alpha}^*(s),\eta|t)]^2} dN_i(s) - Y_i(s) \ddot{\lambda}_i^*(s,\tilde{\alpha}^*(s),\eta|t) ds \Big\},$$

and

$$\frac{\partial U_{\alpha^*}(\alpha^*;\eta,t)}{\partial \eta} = \sum_{i=1}^{n} \int_0^\tau K_h(s-t) X_i^*(s|t) P_i^{\mathsf{T}}(s) \Big(\frac{\partial \zeta(\eta,U_i(s))}{\partial \eta} \Big) \Big\{ \frac{\ddot{\lambda}_i^*(s,\tilde{\alpha}^*(s),\eta|t)\lambda_i^*(s,\tilde{\alpha}^*,\eta|t) - [\dot{\lambda}_i^*(s,\tilde{\alpha}^*(s),\eta|t)]^2}{[\lambda_i^*(s,\tilde{\alpha}^*(s),\eta|t)]^2} dN_i(s) - Y_i(s) \ddot{\lambda}_i^*(s,\tilde{\alpha}^*(s),\eta|t) ds \Big\}.$$

The parameter vector η can be updated by solving the profile estimating equation $U(\eta) = \mathbf{0}$ using Newton-Raphson method, where $U(\eta)$ is defined in equation (2.8).

Following this procedure, $\tilde{\alpha}^*(\eta, t_0)$ and η undergo iterative updates until convergences are reached for both, where the maximum likelihood is achieved.

2.2.3 Computational Algorithm

In this subsection, we summarize the computational algorithm to illustrate the profile maximum likelihood estimation procedure as we outlined in Section 2.2.2.

1. Generate the grid points over t.

2. Set initial values $\hat{\alpha}^{\{0\}}(t)$ and $\hat{\eta}^{\{0\}}$ for $\hat{\alpha}(t)$ and $\hat{\eta}$.

3. Let $\hat{\alpha}^{\{k-1\}}(t)$ and $\hat{\eta}^{\{k-1\}}$ be the estimates of $\alpha(t)$ and η in $(k-1)$th iteration. At each grid point, plug $\hat{\eta}^{\{k-1\}}$ into the localized score function (2.6), solve the equation $U_\alpha(\alpha^*; \eta^{\{k-1\}}, t_0) = \mathbf{0}$ and get $\hat{\alpha}^{*\{k\}}(t) = \hat{\alpha}^{*\{k\}}(t, \hat{\eta}^{\{k-1\}})$; take first p_1 components as the estimation of $\alpha(t)$ in kth iteration and denote it as $\hat{\alpha}^{\{k\}}(t)$;

4. Replace $\tilde{\alpha}(t, \eta)$ with $\hat{\alpha}^{\{k\}}(t)$ in estimating equation (2.8), solve the equation using Newton-Raphson method and get $\hat{\eta}^{\{k\}}$, which is the kth iteration estimate of η.

5. Repeat Step 3 and Step 4, $\hat{\alpha}^{\{k\}}(t)$ and $\hat{\eta}^{\{k\}}$ are updated at each iteration until both of them converge, the estimates $\hat{\alpha}(t)$ and $\hat{\eta}$ are $\hat{\alpha}^{\{k\}}(t)$ and $\hat{\eta}^{\{k\}}$ at convergence.

2.2.4 Selections of Kernel Functions and Bandwidths

The proposed estimation procedure integrates both local smoothing method and profile maximum log-likelihood estimation method. In this section, we discuss the selections of the kernel functions and bandwidths when estimate the nonparametric parameters using local smoothing method. For the kernel function, we choose the

Epanechnikov kernel function $K(x) = 3/4(1-x^2)I\{|x| \leq 1\}$, which has been showed many desirable properties (Epanechnikov (1969); Fan and Gijbels (1996)).

For the bandwidth selection, we employ Monte Carlo cross-validation method, also referred to as "leave-group-out" cross-validation (Cai *et al.* (2023)). This method helps reduce the randomness associated with data splitting during cross-validation.

Firstly, we bootstrap the original data without replacement, with a fixed proportion. The selected data is treated as the training data for model fitting, while the unselected data serves as the test data. In each bootstrap, we fit the model with training data across a range of bandwidth candidates, and record the one that maximizes the prediction accuracy on the test dataset. We repeat this procedure B times and take the average of B recorded bandwidths as the selected optimal bandwidth.

We sketch the bandwidth selection procedure as follows:

1. Create a set of candidate bandwidths H;

2. In bth bootstrap iteration, we do the following steps:

   2.1. Firstly, we randomly sample from the original dataset without replacement, utilizing a fixed proportion p. We obtain a dataset with size np (rounded to the nearest integer), which serves as the training dataset D_n^b and the subjects not selected into D_n^b form the test dataset D_t^b.

   2.2. For each h in H, we use h to fit the model on training dataset D_n^b, obtaining estimates $\hat{\alpha}^{(b,h)}(t)$, $\hat{\beta}^{(b,h)}$, $\hat{\theta}^{(b,h)}$ and the estimated intensity function for subject i takes as follows:

$$\hat{\lambda}_i^{(b,h)}(t) = g^{-1}\Big\{\hat{\alpha}^{(b,h)\mathsf{T}} X_i(t) + \hat{\beta}^{(b,h)\mathsf{T}} Z_i(t) + \hat{\gamma}^{(b,h)\mathsf{T}}(U_i(t),\theta) W_i(t)\Big\}.$$

   2.3. The prediction accuracy is defined as the estimated log-likelihood on test dataset, as proposed by Tian *et al.* (2005). The prediction accuracy in bth

bootstrap using bandwidth h, denoted as $ACC^{(b)}(h)$, is calculated by

$$ACC^{(b)}(h) = \sum_{i \in D_t^b} \int_{t_1}^{t_2} \Big\{ \log[\hat{\lambda}_i^{(b,h)}(t)] dN_i^b(t) - Y_i^{(b)}(t) \hat{\lambda}_i^{(b,h)}(t) dt \Big\}.$$

2.4. The recorded bandwidth in bth iteration h_b^* is the one that maximizes the prediction accuracy, i.e. $h_b^* = \underset{h}{\text{argmax}} ACC^{(b)}(h)$.

3. Repeat Step 2.1 to Step 2.4 B times, the optimal bandwidth, denoted as h_{opt}^*, is determined as the average of all the h_b^* we got in B repetitions of bootstrap.

2.3 Asymptotic Properties

In this section, we discuss the asymptotic properties of the estimator through, we introduce the notations which would be used as follows:

Let η_0 and $\alpha_0(t)$ be the true value of η and $\alpha(t)$, denote the first and second derivatives of $\alpha_0(t)$ by $\dot{\alpha}_0(t)$ and $\ddot{\alpha}_0(t)$. Let $\lambda_i(t) = g^{-1}\{\alpha_0^\mathsf{T}(t) X_i(t) + \zeta^\mathsf{T}(\eta_0, U_i(t)) P_i(t)\}$ and $\dot{\lambda}_i(t) = \dot{g}^{-1}\{\alpha_0^\mathsf{T}(t) X_i(t) + \zeta^\mathsf{T}(\eta_0, U_i(t)) P_i(t)\}$. Define $e_{11}(t) = E\{-Y_i(t) \frac{\dot{\lambda}_i(t)^2}{\lambda_i(t)} X_i(t)^{\otimes 2}\}$ and $e_{12}(t) = E\{-Y_i(t) \frac{\dot{\lambda}_i(t)^2}{\lambda_i(t)} X_i(t) P_i^\mathsf{T}(t) (\frac{\partial \zeta(\eta_0, U_i(t))}{\partial \eta})\}$.

Let $\hat{\lambda}_i(t) = g^{-1}\{\hat{\alpha}^\mathsf{T}(t) X_i(t) + \zeta^\mathsf{T}(\hat{\eta}, U_i(t)) P_i(t)\}$ and $\dot{\hat{\lambda}}_i(t) = \dot{g}^{-1}\{\hat{\alpha}^\mathsf{T}(t) X_i(t) + \zeta^\mathsf{T}(\hat{\eta}, U_i(t)) P_i(t)\}$ and $\ddot{\hat{\lambda}}_i(t) = \ddot{g}^{-1}\{\hat{\alpha}^\mathsf{T}(t) X_i(t) + \zeta^\mathsf{T}(\hat{\eta}, U_i(t)) P_i(t)\}$. Let $\hat{E}_{11}(t) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau Y_i(s) K_h(s-t) \{-\frac{\dot{\hat{\lambda}}_i(s)^2}{\hat{\lambda}_i(s)}\} X_i(s)^{\otimes 2} ds$ and $\hat{E}_{12}(t) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau Y_i(s) K_h(s-t)$ $\{-\frac{\dot{\hat{\lambda}}_i(s)^2}{\hat{\lambda}_i(s)}\} \{X_i(s) P_i^\mathsf{T}(s) (\frac{\partial \zeta(\hat{\eta}, U_i(s))}{\partial \eta})\} ds$.

Under condition A given in Appendix, we have the following theorems for the asymptotic properties of the estimators $\hat{\eta}$ and $\hat{\alpha}(t)$.

Theorem 1 *Under Condition A,* $\eta \xrightarrow{p} \eta_0$*, and* $\sqrt{n}(\hat{\eta} - \eta_0)$ *converges in distribution to a mean zero Gaussian random vector with covariance matrix* $A_\eta^{-1} \Sigma_\eta A_\eta^{-1}$*, with*

$$A_\eta = E\Big[\int_{t_1}^{t_2} \frac{\dot{\lambda}_i(t)^2}{\lambda_i(t)} \Big\{ \Big(\frac{\partial \zeta(U_i(t), \eta_0)}{\partial \eta} \Big)^\mathsf{T} P_i(t) - (e_{12}(t))^\mathsf{T} (e_{11}(t))^{-1} X_i(t) \Big\}^{\otimes 2} dt \Big]$$

and

$$\Sigma_\eta = E\Big[\int_{t_1}^{t_2} \frac{\dot\lambda_i(t)}{\lambda_i(t)}\Big\{\Big(\frac{\partial\zeta(U_i(t),\eta_0)}{\partial\eta}\Big)^{\mathsf T} P_i(t) - (e_{12}(t))^{\mathsf T}(e_{11}(t))^{-1}X_i(t)\Big\}dM_i(t)\Big]^{\otimes 2}$$

where $0 < t_1 < t_2 < \tau$, $\otimes$ *is the Kronecker product of matrices.*

A_η *can be estimated by*

$$\hat A_\eta = \frac{1}{n}\sum_{i=1}^{n}\int_{t_1}^{t_2} \frac{\hat{\dot\lambda}^2(t)}{\hat\lambda_i(t)}\Big\{\Big(\frac{\partial\zeta(U_i(t),\hat\eta)}{\partial\eta}\Big)^{\mathsf T} P_i(t) - (\hat E_{12}(t))^{\mathsf T}(\hat E_{11}(t))^{-1}X_i(t)\Big\}^{\otimes 2}dt$$

and Σ_η *can be estimated by*

$$\hat\Sigma_\eta = \frac{1}{n}\sum_{i=1}^{n}\Big[\int_{t_1}^{t_2}\frac{\hat{\dot\lambda}_i(t)}{\hat\lambda_i(t)}\Big\{\Big(\frac{\partial\zeta(U_i(t),\hat\eta)}{\partial\eta}\Big)^{\mathsf T} P_i(t) - (\hat E_{12}(t))^{\mathsf T}(\hat E_{11}(t))^{-1}X_i(t)\Big\} \Big\{dN_i(t) - Y_i(t)\hat\lambda_i(t)dt\Big\}\Big]^{\otimes 2}$$

Theorem 2 *Under Condition A,* $\hat\alpha(t) \xrightarrow{\mathcal{P}} \alpha_0(t)$*, uniformly in* $t\in[t_1,t_2]\subset[0,\tau]$*, and*

$$(nh)^{1/2}(\hat\alpha(t) - \alpha_0(t) - \frac{1}{2}\mu_2 h^2 \ddot\alpha_0{}^{\mathsf T}(t)) \xrightarrow{\mathcal{D}} N(0,\Sigma_\alpha(t))$$

where $\mu_2 = \int_{-1}^{1} t^2K(t)dt$.

$\Sigma_\alpha(t)$ *can be consistently estimated by* $(\hat E_{11}(t))^{-1}\hat\Sigma_e(t)(\hat E_{11}(t))^{-1}$*, with*

$$\begin{aligned}\hat\Sigma_e(t) = n^{-1}h\sum_{i=1}^{n}\Big[&\int_0^\tau \frac{\hat{\dot\lambda}_i(s)}{\hat\lambda_i(s)}\{dN_i(s) - Y_i(s)\hat\lambda_i(s)\}X_i(s)K_h(s-t)\\ &-\hat E_{12}(t)\hat A_\eta^{-1}\int_{t_1}^{t_2}\frac{\hat{\dot\lambda}_i(s)}{\hat\lambda_i(s)}\{dN_i(s) - Y_i(s)\hat\lambda_i(s)\}\\ &\Big\{\Big(\frac{\partial\zeta(U_i(s),\hat\eta)}{\partial\eta}\Big)^{\mathsf T} P_i(s) - (\hat E_{12}(s))^{\mathsf T}(\hat E_{11}(s))^{-1}X_i(s)\Big\}\Big]^{\otimes 2}\end{aligned}$$

2.4 Testing the Covariate-Varying Effects

In this section, we provide two hypothesis test procedures to test the adequacy of parametric form $\gamma(U_i(t), \theta)$. To test $H_0 : \gamma(u) = \gamma(u, \theta)$, we consider the following test process

$$R(u, \hat{\eta}) = n^{-\frac{1}{2}}(I_r \otimes \hat{A}_\eta^{-1}) \sum_{i=1}^{n} \Big\{ \int_{t_1}^{t_2} \frac{\hat{\dot{\lambda}}_i(t)}{\hat{\lambda}_i(t)} I\{U_i(t) \le u\} \otimes \hat{O}_i(t)\{dN_i(t) - Y_i(t)\hat{\lambda}_i(t)dt\} \Big\} \tag{2.10}$$

where

$$\hat{O}_i(t) = \Big(\frac{\partial \zeta(U_i(t), \hat{\eta})}{\partial \eta}\Big)^{\mathsf{T}} P_i(t) - (\hat{E}_{12}(t))^{\mathsf{T}} (\hat{E}_{11}(t))^{-1} X_i(t),$$

$u \in R^r$ is a grid of $U_i(t)$ and r is the dimension of $U_i(t)$, I_r is the $r \times r$ identity matrix, $\otimes$ is the Kronecker product of matrices. The test process is a stratified version of the score function based on the level of $U_i(t) \le u$.

Test statistics T_1 is based on the supreme norm. Define $T_1 = \sup_{u \in \triangle} \|R(u, \hat{\eta})\|$, where $\|\cdot\|$ represent the L_2 norm in R^r and $\triangle$ is a set of grid points in R^r.

Test statistics T_2 is based on the chi-square test. Take the grids of $U_i(t)$ as $\{u_1, ...u_K\}$, K is the number of the grid points. Let $L(\hat{\eta})$ be the difference of two consecutive $R(u, \hat{\eta})$,

$$L(\hat{\eta}) = \begin{bmatrix} R(u_2, \hat{\eta}) - R(u_1, \hat{\eta}) \\ R(u_3, \hat{\eta}) - R(u_2, \hat{\eta}) \\ ... \\ R(u_{K-1}, \hat{\eta}) - R(u_{K-2}, \hat{\eta}) \\ R(u_K, \hat{\eta}) - R(u_{K-1}, \hat{\eta}) \end{bmatrix}$$

define test statistic T_2 as the quadratic form of $L(\hat{\eta})$:

$$T_2 = L^{\mathsf{T}}(\hat{\eta})\{\text{cov}(L(\hat{\eta}), L(\hat{\eta}))\}^{-1}L(\hat{\eta}),$$

$\text{cov}\{L((\hat{\eta}), L(\hat{\eta}))\}$ is a $(K-1)\times(K-1)$ block matrix, for $1 \le q, s \le K-1$, the q, s th block equals

$$\begin{aligned}
&\text{cov}[R(u_{q+1},\hat{\eta}) - R(u_q,\hat{\eta}), R(u_{s+1},\hat{\eta}) - R(u_s,\hat{\eta})] \\
=&\text{cov}[R(u_{q+1},\hat{\eta}), R(u_{s+1},\hat{\eta})] - \text{cov}[R(u_{q+1},\hat{\eta}), R(u_s,\hat{\eta})] - \text{cov}[R(u_q,\hat{\eta}), R(u_{s+1},\hat{\eta})] \\
&+ \text{cov}[R(u_q,\hat{\eta}), R(u_s,\hat{\eta})]. \qquad (2.11)
\end{aligned}$$

Since the distribution of T_1 is unknown and complicated, we consider using Gaussian multiplier method to approximate its distribution (Lin *et al.* (1993)). The outline of this procedure is given in the following:

By first order approximation, we have

$$R(u,\hat{\eta}) = R(u,\eta_0) + \frac{\partial R(u,\eta_0)}{\partial \eta}(\hat{\eta} - \eta_0) + o_p(1), \qquad (2.12)$$

where

$$\begin{aligned}
R(u,\eta_0) =& n^{-\frac{1}{2}}(I_r \otimes A_\eta^{-1}) \sum_{i=1}^{n} \Big\{ \int_{t_1}^{t_2} \frac{\dot{\lambda}_i(t)}{\lambda_i(t)} I\{U_i(t) \le u\} \otimes O_i(t)[dN_i(t) - Y_i(t)\lambda_i(t)dt] \Big\} \\
&+ o_p(1), \qquad (2.13)
\end{aligned}$$

with

$$O_i(t) = \Big(\frac{\partial \zeta(U_i(t),\eta)}{\partial \eta}\Big)^{\mathsf{T}} P_i(t) - (e_{12}(t))^{\mathsf{T}}(e_{11}(t))^{-1}X_i(t).$$

As shown in Appendix A.1, we have

$$n^{\frac{1}{2}}(\hat{\eta}-\eta_0) = A_\eta^{-1} n^{-\frac{1}{2}} \sum_{i=1}^{n} \Big\{ \int_{t_1}^{t_2} \frac{\dot{\lambda}_i(t)}{\lambda_i(t)} O_i(t)\{dN_i(t) - Y_i(t)\lambda_i(t)dt\} \Big\} + o_p(1). \quad (2.14)$$

The derivative of $R(u,\eta)$ with respect of η can be shown that

$$n^{-\frac{1}{2}} \frac{\partial R(u,\eta)}{\partial \eta} \xrightarrow{p} -(I_r \otimes A_\eta^{-1}) A_u, \quad (2.15)$$

where

$$A_u = E\Big\{ \int_{t_1}^{t_2} [I\{U_i(t) \le u\} \otimes O_i(t)] O_i(t)^{\intercal} [\frac{\ddot{\lambda}_i(t)\lambda_i(t) - \dot{\lambda}_i^2(t)}{\lambda_i^2(t)} dN_i(t) - Y_i(t)\ddot{\lambda}_i(t)dt] \Big\}.$$

Combining equation (2.12)-(2.15), we have

$$R(u,\hat{\eta}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} D_i(u) + o_p(1)$$

where

$$D_i(u) = (I_r \otimes A_\eta^{-1}) \int_{t_1}^{t_2} \frac{\dot{\lambda}_i(t)}{\lambda_i(t)} \Big\{ I\{U_i(t) \le u\} \otimes O_i(t) - A_u A_\eta^{-1} O_i(t) \Big\} \Big\{ dN_i(t) - Y_i(t)\lambda_i(t)dt \Big\},$$

from Lin *et al.* (1993), we know $R(u,\hat{\eta})$ converges weakly to a mean zero Gaussian process $R(u)$, for $u \in R^r$.

Let

$$\hat{D}_i(u) = (I_r \otimes \hat{A}_\eta^{-1}) \int_{t_1}^{t_2} \frac{\hat{\dot{\lambda}}_i(t)}{\hat{\lambda}_i(t)} \Big[I\{U_i(t) \le u\} \otimes \hat{O}_i(t) - \hat{A}_u \hat{A}_\eta^{-1} \hat{O}_i(t) \Big] \{dN_i(t) - Y_i(t)\hat{\lambda}_i(t)dt\}$$

and ϕ_1, ... ,ϕ_n be n independent standard normal random variables, define the Gaus-

sian multiplier process

$$R^*(u) = n^{-1/2} \sum_{i=1}^{n} \hat{D}_i(u)\phi_i. \tag{2.16}$$

Given the observed data, the distribution of $R(u)$ can be approximated by the construction of $R^*(u)$. We hold the observed data sequence fixed and generate, we say, 500 sets of independent normal random variables, get 500 copies of $R^*(u)$, define $T_1^* = \sup_{u\in\triangle} \|R^*(u,\hat{\eta})\|$, then the critical value of T_1 can be determined by the percentile of the empirical distribution of T_1^*.

Based on the construction of $R^*(u)$, for $1 \le l \le m <= K$, $\mathrm{cov}(R(u_l,\hat{\eta}), R(u_m,\hat{\eta}))$ can be calculated by

$$\begin{aligned}
\mathrm{cov}(R(u_l,\hat{\eta}, R(u_m,\hat{\eta}) &= \mathrm{cov}(R^*(u_l), R^*(u_m)) \\
&= \mathrm{cov}\Big(n^{-1/2}\sum_{i=1}^{n} \hat{D}_i(u_l)\phi_i, n^{-1/2}\sum_{j=1}^{n} \hat{D}_j(u_m)\phi_j\Big) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \hat{D}_i(u_l)\hat{D}_j^{\mathsf{T}}(u_m)\mathrm{cov}(\phi_i,\phi_j) \\
&= \frac{1}{n}\sum_{i=1}^{n} \hat{D}_i(u_l)\hat{D}_i^{\mathsf{T}}(u_m)\mathrm{cov}(\phi_i,\phi_i) \\
&= \frac{1}{n}\sum_{i=1}^{n} \hat{D}_i(u_l)\hat{D}_i^{\mathsf{T}}(u_m).
\end{aligned}$$

Test statistic T_2 has asymptotic chi-square distribution. The critical value can be found based on the chi-square distribution. However, its critical value can also be approximated using the Gaussian multiplier approach which may perform better in small sample case.

Define

$$L^*(\hat{\eta}) = \begin{bmatrix} R^*(u_2, \hat{\eta}) - R^*(u_1, \hat{\eta}) \\ R^*(u_3, \hat{\eta}) - R^*(u_2, \hat{\eta}) \\ ... \\ R^*(u_{K-1}, \hat{\eta}) - R*(u_{K-2}, \hat{\eta}) \\ R^*(u_K, \hat{\eta}) - R*(u_{K-1}, \hat{\eta}) \end{bmatrix}$$

and let

$$T_2^* = L^*(\hat{\eta})^{\mathsf{T}} \{\mathrm{cov}(L(\hat{\eta}), L(\hat{\eta}))\}^{-1} L^*(\hat{\eta}),$$

the critical value of T_2 can be determined by the percentile of the empirical distribution of T_2^*.

2.5 Simulation Studies

In this section, we perform comprehensive simulations to assess the performance of the estimators in finite samples under two distinct link functions: the logarithm link function and the identity link function. Within the logarithm link function, we further examine three different scenarios, which are showed in section 2.5.1.

We use the following abbreviations in all the simulation studies hereafter. Bias = estimate- true value; SSE stands for the sample standard error of the estimates; ESE stands for the sample mean of the estimated standard errors; CP represents the 95% empirical coverage probability.

2.5.1 Simulation Studies under Logarithm Link Function

We consider the models with following intensity function

$$\lambda_i(t) = \exp\{\alpha_0(t) + \alpha_1(t)X_i + \beta Z_i + \gamma(U_i(t), \theta)W_i(t)\}, \tag{2.17}$$

for $0 \leq t \leq \tau$.

Scenario 1. $\gamma(U_i(t), \theta)$ is a linear function of $U_i(t)$, with the following settings:

- $\tau = 4$, subjects are censored up to a censoring time $C_i \sim U(3, 8)$;
- X_i is a uniform random variable on [-1, 1], Z_i is generated from truncated Normal distribution (0, 1, 0.5, 0.2);
- $U_i(t) = t - T_{i,N_i(t^-)}$ stands for the gap time since the most recent event; $W_i(t) = I(N_i(t^-) > 0)$ indicates whether there is an event occurred just before time t;
- $\gamma(U_i(t), \theta) = \theta_0 + \theta_1 U_i(t)$; with $\theta_0 = 0.5$ and $\theta_1 = 0.5$;
- $\beta = 0.1$, $\alpha_0(t) = -0.5 - 0.5\log(1+t)$, $\alpha_1(t) = -0.5\sin(1+0.4t)$.

Averagely, 2.7 events are observed per subject during the study time. Table 2.1 presents the estimation results for sample sizes of $n = 400, 600, 800$, with different bandwidth $h = 0.2, 0.3, 0.4$. Table 2.2 presents the sizes under null model $M_0 : \gamma(u) = 0.5 + 0.5u$, and the powers under alternative models $M_{11} : \gamma(u) = 1.5 - 0.4\sin(5u)$, $M_{12} : \gamma(u) = 1.5 - 0.5\sin(5u)$ and $M_{21} : \gamma(u) = 1.5 - u + 0.3u^2$ and $M_{22} : \gamma(u) = 1.5 - u + 0.4u^2$. Figure 2.1 shows the estimation results for $\alpha_0(t)$ and $\alpha_1(t)$.

Table 2.1: Bias, SEE, ESE and CP of β, θ_0, θ_1 under model (2.17) in Scenario 1, bandwidth $h = 0.2$, 0.3 and 0.4.

	$\beta = 0.1$				$\theta_0 = 0.5$				$\theta_1 = 0.5$			
n	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP
					h=0.2							
400	0.003	0.159	0.164	0.950	0.003	0.086	0.088	0.954	-0.003	0.064	0.065	0.950
600	-0.002	0.128	0.133	0.956	0.006	0.076	0.072	0.940	-0.006	0.056	0.053	0.936
800	-0.005	0.109	0.115	0.962	-0.001	0.062	0.062	0.948	-0.008	0.043	0.045	0.966
					h=0.3							
400	-0.009	0.161	0.165	0.946	-0.001	0.087	0.089	0.958	-0.004	0.065	0.065	0.956
600	-0.011	0.130	0.135	0.958	0.003	0.076	0.073	0.934	-0.006	0.057	0.053	0.934
800	-0.014	0.111	0.117	0.958	-0.003	0.063	0.063	0.954	-0.009	0.044	0.046	0.952
					h=0.4							
400	-0.019	0.161	0.167	0.950	-0.002	0.089	0.090	0.956	-0.004	0.065	0.066	0.964
600	-0.018	0.134	0.136	0.950	0.001	0.076	0.073	0.944	-0.006	0.057	0.054	0.932
800	-0.019	0.111	0.118	0.958	-0.005	0.063	0.063	0.960	-0.009	0.045	0.047	0.958

Table 2.2: Observed sizes and powers of test statistics T_1, T_{2g} and T_{2c} under model (2.17) in Scenario 1, bandwidth $h = 0.3$ and 0.4. T_1 is based on supremum test, T_{2g} is based on Gaussian multiplier distribution and T_{2c} is based on chi-square distribution.

		n=400			n=600			n=800		
Model	h	T_1	T_{2g}	T_{2c}	T_1	T_{2g}	T_{2c}	T_1	T_{2g}	T_{2c}
M_0	0.3	0.054	0.074	0.062	0.050	0.052	0.054	0.050	0.046	0.046
	0.4	0.052	0.076	0.064	0.044	0.050	0.050	0.052	0.048	0.046
M_{11}	0.3	0.234	0.850	0.830	0.518	0.960	0.964	0.800	0.998	0.998
	0.4	0.270	0.836	0.834	0.534	0.958	0.960	0.792	0.998	0.998
M_{12}	0.3	0.530	0.974	0.970	0.850	1.000	1.000	0.990	1.000	1.000
	0.4	0.532	0.968	0.964	0.878	1.000	1.000	0.982	1.000	1.000
M_{21}	0.3	0.810	0.368	0.366	0.946	0.544	0.546	0.988	0.770	0.774
	0.4	0.804	0.372	0.370	0.954	0.544	0.550	0.984	0.758	0.764
M_{22}	0.3	0.964	0.718	0.702	0.998	0.936	0.936	1.000	0.996	0.996
	0.4	0.958	0.704	0.704	0.996	0.928	0.926	1.000	0.996	0.992

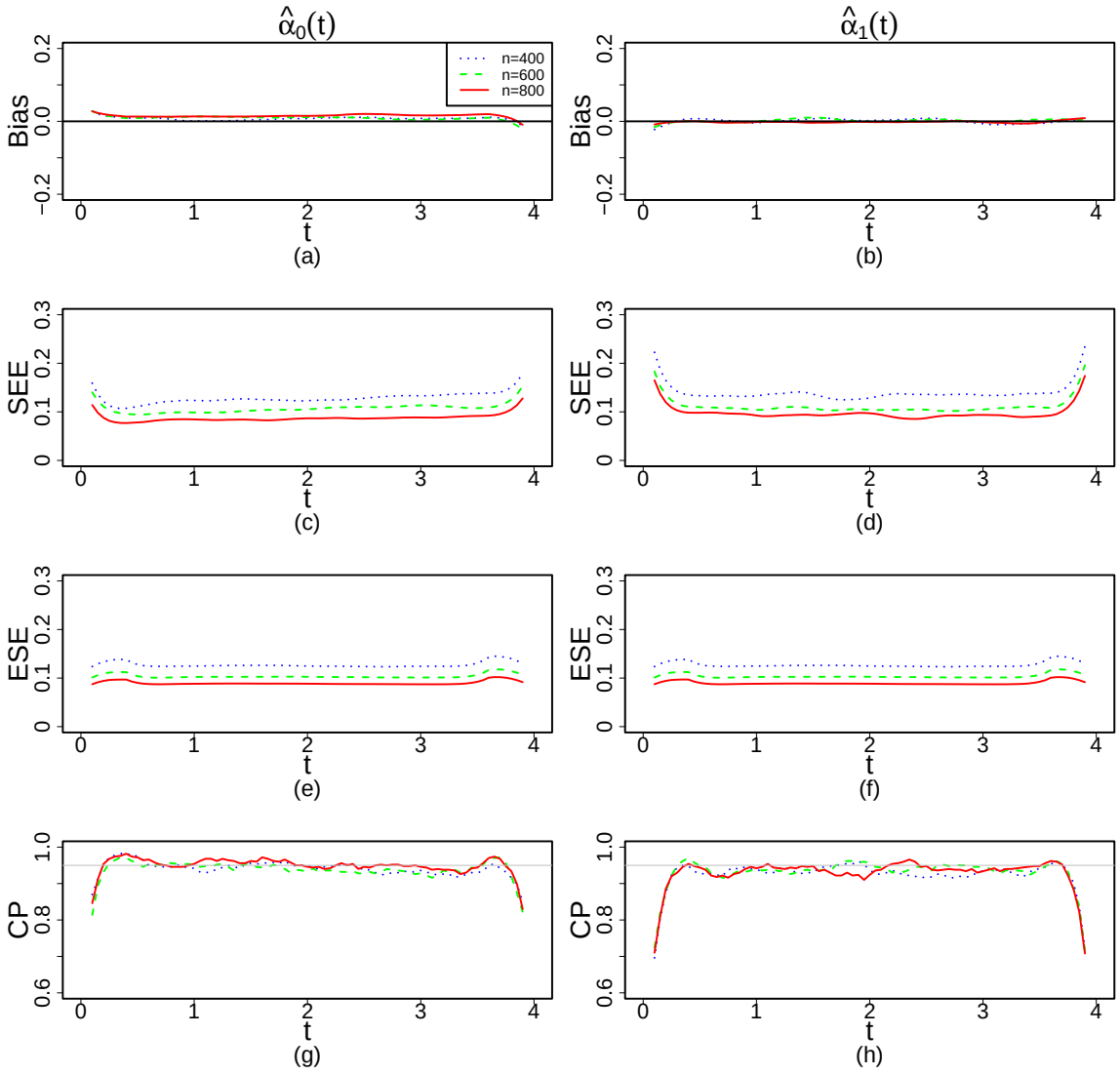

Figure 2.1: Bias, SEE, ESE and CP of $\alpha_0(t)$ and $\alpha_1(t)$ under model (2.17) in Scenario 1 with $h = 0.4$. The dotted, dashed and solid line represent $n = 400$, $n = 600$ and $n = 800$, respectively.

Scenario 2. $\gamma(U_i(t), \theta)$ is multidimensional, we consider model 2.17 under the following settings:

- $\tau = 4$, subjects are censored up to a censoring time $C_i \sim U(3, 8)$;
- X_i is a uniform random variable on [-1, 1], Z_i is generated from truncated Normal distribution (0, 1, 0.5, 0.2);
- $U_i(t) = t - T_{iN_i(t^-)}$ stands for the gap time since last event; $W_i = (W_{1i}, W_{2i})^\mathsf{T}$ with $W_{1i}(t) = I(N_i(t^-) > 0)$ and $W_{2i}(t) = I(N_i(t^-) > 0)B_i(t)$, where $B_i(t) \sim Ber(0.5)$;
- $\gamma(U_i(t), \theta) = (\gamma_1(U_i(t), \theta_1), \gamma_2(U_i(t), \theta_2))^\mathsf{T}$; $\gamma_1(U_i(t), \theta_1) = \theta_{10} + \theta_{11}U_i(t)$ and $\gamma_2(U_i(t), \theta_2) = \theta_{20} + \theta_{21}U_i(t)$ with $\theta_{10} = \theta_{11} = 0.3$, $\theta_{20} = \theta_{21} = 0.2$;
- $\beta = 0.1$, $\alpha_0(t) = -0.5 - 0.5\log(1 + t)$, $\alpha_1(t) = -0.5\sin(1 + 0.4t)$.

Averagely, 2.5 events are observed per subject during the study time. Table 2.3 presents the estimation results for sample sizes of $n = 400, 600, 800$ with $h = 0.4$ and 0.5. Table 2.4 presents the sizes under null model M_0: $\gamma_1(U_i(t), \theta_1) = 0.3 + 0.3U_i(t)$ and $\gamma_2(U_i(t), \theta_1) = 0.2 + 0.2U_i(t)$. Figure 2.2 shows the estimation results for $\alpha_0(t)$ and $\alpha_1(t)$.

Table 2.3: Bias, SEE, ESE and CP of β, θ_{10}, θ_{11}, θ_{20} and θ_{21} under model (2.17) in Scenario 2, bandwidth $h = 0.4$ and 0.5.

	$n = 400$				$n = 600$				$n = 800$			
	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP
						h=0.4						
β	-0.010	0.166	0.175	0.960	-0.008	0.142	0.144	0.950	-0.008	0.114	0.124	0.962
θ_{10}	-0.006	0.122	0.118	0.948	-0.001	0.095	0.097	0.946	0.003	0.083	0.084	0.948
θ_{11}	0.000	0.090	0.092	0.948	-0.002	0.077	0.075	0.948	-0.006	0.064	0.065	0.944
θ_{20}	-0.004	0.131	0.129	0.950	0.002	0.106	0.105	0.956	0.001	0.088	0.092	0.956
θ_{21}	0.012	0.126	0.124	0.950	0.000	0.103	0.101	0.950	0.000	0.084	0.088	0.946
						h=0.5						
β	-0.016	0.171	0.177	0.950	-0.014	0.143	0.145	0.954	-0.012	0.116	0.126	0.964
θ_{10}	-0.008	0.124	0.120	0.944	-0.003	0.095	0.098	0.950	0.003	0.084	0.085	0.948
θ_{11}	0.000	0.093	0.093	0.948	-0.002	0.078	0.076	0.948	-0.007	0.065	0.066	0.952
θ_{20}	-0.003	0.132	0.131	0.954	0.002	0.107	0.106	0.960	0.000	0.089	0.093	0.958
θ_{21}	0.011	0.128	0.127	0.950	0.001	0.105	0.103	0.952	0.001	0.086	0.089	0.950

Table 2.4: Observed sizes of test statistics T_1, T_{2g} and T_{2c} under model (2.17) in Scenario 2, bandwidth h=0.4 and 0.5. T_1 is based on the supremum test, T_{2g} is based on Gaussian multiplier distribution and T_{2c} is based on chi-square distribution.

		n=400			n=600			n=800		
Model	h	T_1	T_{2g}	T_{2c}	T_1	T_{2g}	T_{2c}	T_1	T_{2g}	T_{2c}
M_0	0.4	0.060	0.084	0.082	0.056	0.062	0.060	0.058	0.050	0.050
	0.5	0.062	0.098	0.096	0.062	0.062	0.044	0.044	0.066	0.068

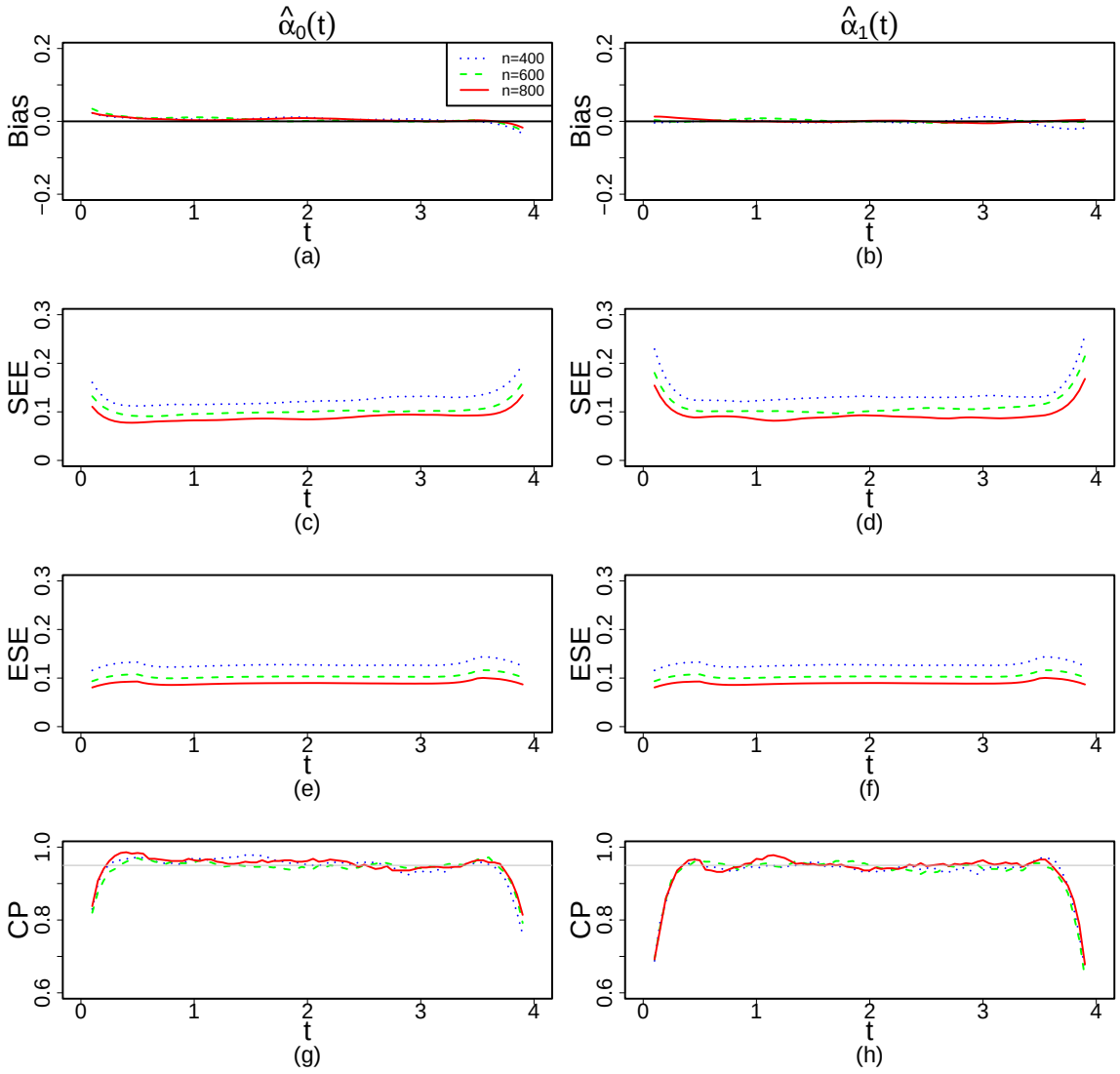

Figure 2.2: Bias, SEE, ESE and CP of $\alpha_0(t)$ and $\alpha_1(t)$ under model (2.17) in Scenario 2 with $h = 0.5$. The dotted, dashed and solid line represent $n = 400$, $n = 600$ and $n = 800$, respectively.

Scenario 3. $U_i(t)$ has multiple dimensions and $\gamma(U_i(t), \theta)$ is a nonlinear function of $U_i(t)$; we consider model 2.17 under following settings:

- $\tau = 4$, subjects are censored up to a censoring time $C_i \sim U(3, 8)$;
- X_i is a uniform random variable on [-1, 1], Z_i is generated from truncated Normal distribution (0, 1, 0.5, 0.2);
- $U_i(t) = (U_{1i}(t), U_{2i}(t))$, $U_{1i}(t) = t - T_{i,N_i(t^-)}$ stands for the gap time since last event, $U_{2i}(t) = \log(N_i(t^-) + 1)$ is derived from the current event number; $W_i(t) = I(N_i(t^-) > 0)$ indicates whether there is an event occurred just before time t;.
- $\gamma(U_i(t), \theta) = \log(\theta_0 + \theta_1 U_{1i}(t))^2 + \theta_2 U_{2i}(t)$; with $\theta_0 = 0.8$, $\theta_1 = 0.5$, $\theta_2 = 0.5$;
- $\beta = 0.1$, $\alpha_0(t) = -0.5 - 0.5 \log(1 + t)$, $\alpha_1(t) = -0.5 \sin(1 + 0.4t)$.

Averagely, 2.6 events are observed per subject during the study time. Table 2.5 present the estimation results for β, θ_0, θ_1 and θ_2 for sample sizes of $n = 400, 600, 800$ with different bandwidths. Figure 2.3 presents the estimation results for $\alpha_0(t)$ and $\alpha_1(t)$.

Table 2.5: Bias, SEE, ESE and CP of β, θ_0, θ_1 and θ_2 under model (2.17) in Scenario 3, bandwidth $h = 0.3, 0.4$ and 0.5.

	$\beta = 0.1$				$\theta_0 = 0.8$				$\theta_1 = 0.5$				$\theta_2 = 0.5$			
n	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP
								h=0.3								
400	0.023	0.170	0.177	0.956	0.014	0.082	0.075	0.930	0.007	0.054	0.054	0.956	-0.010	0.134	0.120	0.926
600	0.023	0.153	0.143	0.932	0.006	0.062	0.061	0.956	0.000	0.041	0.044	0.956	-0.003	0.100	0.097	0.946
800	0.020	0.125	0.124	0.944	0.006	0.054	0.053	0.944	0.002	0.037	0.038	0.942	-0.006	0.086	0.085	0.934
								h=0.4								
400	0.012	0.173	0.181	0.956	0.013	0.084	0.078	0.940	0.007	0.055	0.056	0.954	-0.010	0.138	0.123	0.922
600	0.020	0.157	0.148	0.936	0.006	0.063	0.065	0.948	0.000	0.042	0.045	0.952	-0.003	0.103	0.100	0.952
800	0.016	0.129	0.128	0.956	0.005	0.057	0.055	0.942	0.002	0.037	0.039	0.944	-0.006	0.089	0.088	0.934
								h=0.5								
400	0.007	0.178	0.186	0.960	0.014	0.088	0.081	0.926	0.006	0.056	0.057	0.946	-0.011	0.142	0.127	0.916
600	0.016	0.163	0.152	0.944	0.006	0.068	0.065	0.934	-0.002	0.043	0.046	0.944	-0.002	0.107	0.104	0.934
800	0.012	0.133	0.131	0.948	0.004	0.059	0.057	0.932	0.001	0.039	0.040	0.938	-0.005	0.092	0.090	0.928

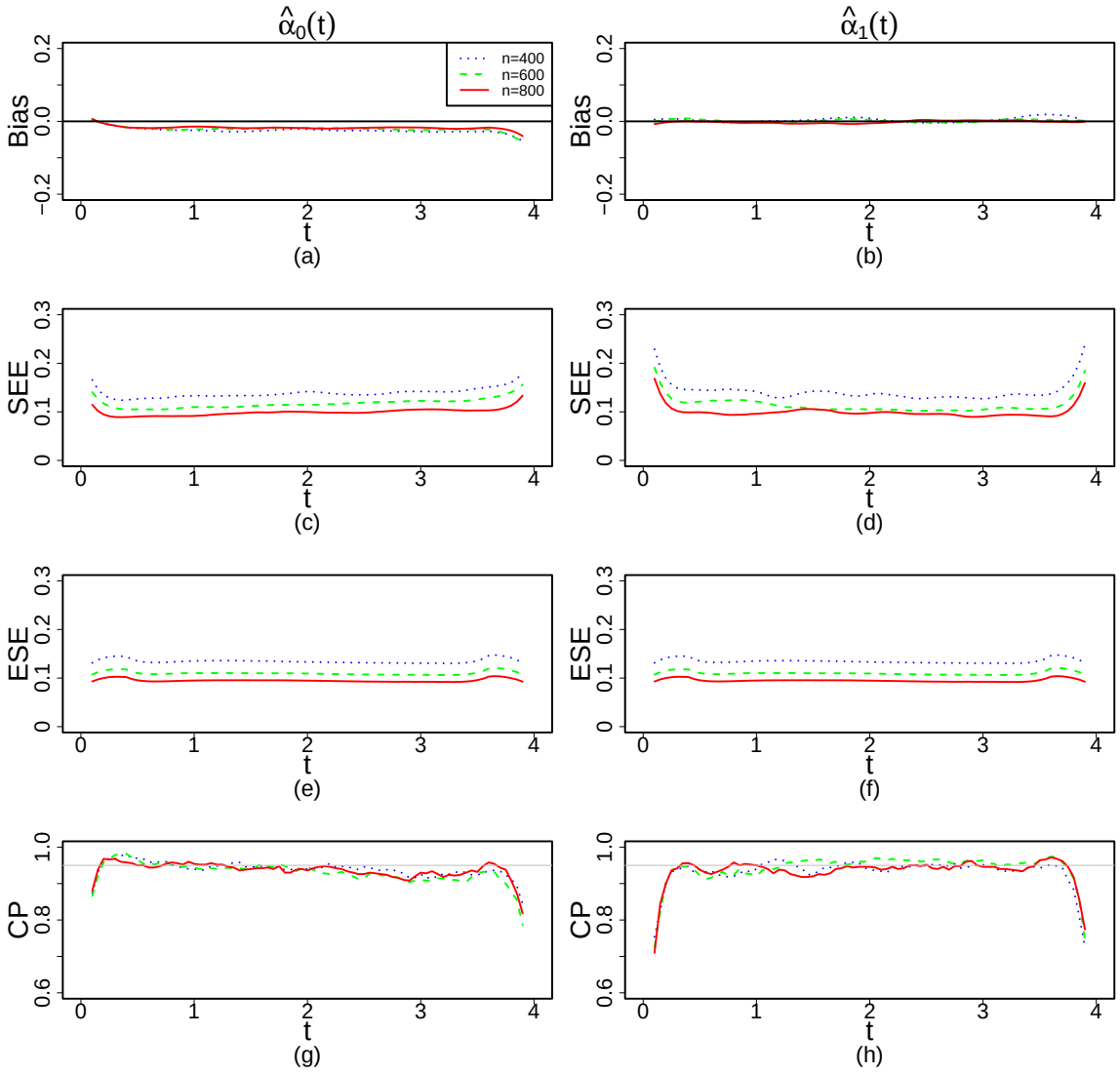

Figure 2.3: Bias, SEE, ESE and CP for $\alpha_0(t)$ and $\alpha_1(t)$ under model (2.17) in Scenario 3 with $h = 0.4$. The dotted, dashed and solid line represent $n = 400$, $n = 600$ and $n = 800$, respectively.

2.5.2 Simulation Studies under Identity Link Function

In this subsection, we conduct a simulation study when the link function being an identity link, we consider the models with intensity function as follows:

$$\lambda_i(t) = \alpha_0(t) + \alpha_1^{\mathsf{T}}(t)X_i + \beta Z_i + \gamma(U_i(t), \theta)W_i(t) \tag{2.18}$$

for $0 \leq t \leq \tau$. Covariates and parameters are set as follows:

- $\tau = 4$, subjects are censored up to a censoring time $C_i \sim U(3, 8)$;
- X_i is a uniform random variable on [-1, 1], Z_i is generated from truncated Normal distribution (0, 1, 0.5, 0.2);
- $U_i(t) = t - T_{i,N_i(t^-)}$ stands for the gap time since last event, $W_i(t) = I(N_i(t^-) > 0)$ indicates whether there is an event occurred just before time t;
- $\gamma(U_i(t), \theta) = \theta_0 + \theta_1 U_i(t)$; with $\theta_0 = 0.2$, $\theta_1 = 0.2$;
- $\beta = 0.1$, $\alpha_0(t) = 0.6 - 0.2\log(1 + t)$, $\alpha_1(t) = 0.2\sin(t)$.

Averagely, 2.5 events are observed per subject during the study time. Table 2.6 presents the estimation results for β, θ_0 and θ_1 for sample sizes of $n = 400, 600, 800$ with bandwidth $h = 0.7, 0.8, 0.9$; Table 2.7 presents the sizes under null model M_0 : $\gamma(u) = 0.2 + 0.2u$, and the powers under alternative models M_{11} : $\gamma(u) = 1.0 - 0.5\sin(5u)$, M_{12} : $\gamma(u) = 1.0 - 0.6\sin(5u)$, M_{21} : $\gamma(u) = 1.0 - u + 0.5u^2$ and M_{22} : $\gamma(u) = 1.0 - u + 0.6u^2$. Figure 2.4 presents the estimation results for $\alpha_0(t)$ and $\alpha_1(t)$.

Table 2.6: Bias, SEE, ESE and CP of β, θ_0, θ_1 under model (2.18), with $h = 0.7$, 0.8 and 0.9.

	$\beta = 0.1$				$\theta_0 = 0.2$				$\theta_1 = 0.2$			
n	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP
						h=0.7						
400	0.007	0.118	0.109	0.904	0.002	0.060	0.056	0.938	0.002	0.056	0.055	0.948
600	0.005	0.096	0.090	0.926	0.005	0.046	0.046	0.952	-0.004	0.045	0.045	0.934
800	0.003	0.080	0.078	0.942	0.003	0.040	0.040	0.942	-0.005	0.041	0.039	0.928
						h=0.8						
400	0.003	0.127	0.110	0.908	0.003	0.066	0.057	0.932	-0.003	0.058	0.056	0.938
600	0.004	0.098	0.091	0.934	0.005	0.047	0.047	0.942	-0.004	0.046	0.046	0.930
800	0.001	0.083	0.079	0.932	0.003	0.040	0.041	0.950	-0.005	0.042	0.040	0.926
						h=0.9						
400	-0.004	0.149	0.111	0.900	0.003	0.067	0.058	0.932	-0.002	0.059	0.057	0.934
600	0.001	0.099	0.093	0.936	0.005	0.048	0.048	0.946	-0.005	0.047	0.047	0.938
800	0.000	0.087	0.081	0.930	0.003	0.040	0.041	0.954	-0.005	0.043	0.040	0.922

Table 2.7: Observed sizes and powers of test statistics T_1, T_{2g} and T_{2c} under model (2.18), bandwidth $h = 0.8$ and 0.9. T_1 is based on supremum test, T_{2g} is based on Gaussian multiplier distribution and T_{2c} is based on chi-square distribution.

		n=400			n=600			n=800		
Model	h	T_1	T_{2g}	T_{2c}	T_1	T_{2g}	T_{2c}	T_1	T_{2g}	T_{2c}
M_0	0.8	0.052	0.060	0.060	0.042	0.040	0.036	0.038	0.044	0.046
	0.9	0.082	0.062	0.060	0.052	0.040	0.040	0.052	0.050	0.048
M_{11}	0.8	0.604	0.650	0.658	0.798	0.916	0.912	0.918	0.980	0.978
	0.9	0.604	0.650	0.640	0.788	0.914	0.904	0.920	0.970	0.970
M_{12}	0.8	0.820	0.842	0.840	0.966	0.988	0.988	0.988	1.000	1.000
	0.9	0.806	0.826	0.826	0.952	0.984	0.984	0.990	0.998	0.998
M_{21}	0.8	0.472	0.600	0.590	0.790	0.876	0.882	0.972	0.970	0.974
	0.9	0.482	0.586	0.578	0.760	0.886	0.878	0.956	0.964	0.962
M_{22}	0.8	0.642	0.698	0.694	0.902	0.948	0.948	0.998	0.992	0.994
	0.9	0.642	0.702	0.694	0.882	0.942	0.940	0.988	0.990	0.992

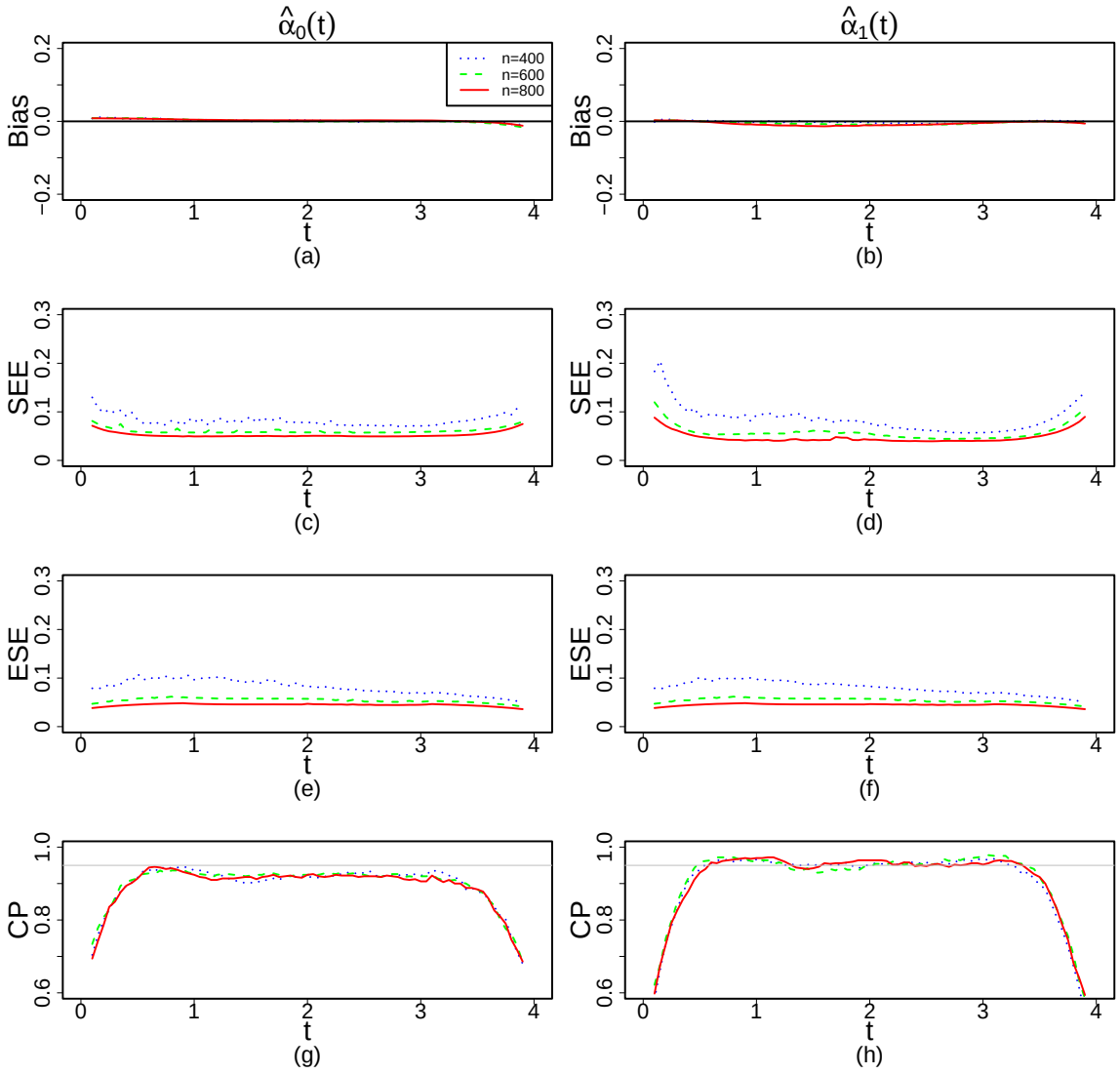

Figure 2.4: Bias, SEE, ESE and CP for $\alpha_0(t)$ and $\alpha_1(t)$ under model (2.18) with $h = 0.9$. The dotted, dashed and solid line represent $n = 400$, $n = 600$ and $n = 800$, respectively.

2.6 Data Application

We apply the proposed method to analyze the 20 months MAL-094 trial data. In this trial, approximately 1500 children aged 5 to 17 months from two sites (Agogo in Ghana and Siaya in Kenya) were randomly divided into five arms, four arms received the RTS,S/AS01$_{\mathrm{E}}$ vaccine with different doses and schedules, and one arm serve as a control arm, receiving the placebo. For all participants, there are records at scheduled and unscheduled visits for whether they get malaria infections detected molecularly. For each participant, visits after three consecutive missed scheduled visits and with no intervening unscheduled visits in-between are defined as censored.

For the 20-months follow up data, there are 3325 molecularly detected malaria infections (referred to as "infection" hereafter) were observed before censoring among 1464 participants. There are 975 participants have experienced at least one infection, with the largest number of infections being 25.

Our preliminary analysis indicates that hemoglobin and age of participants have significant effects on the risk of infection. Figure 2.5 displays histograms of these two variables.

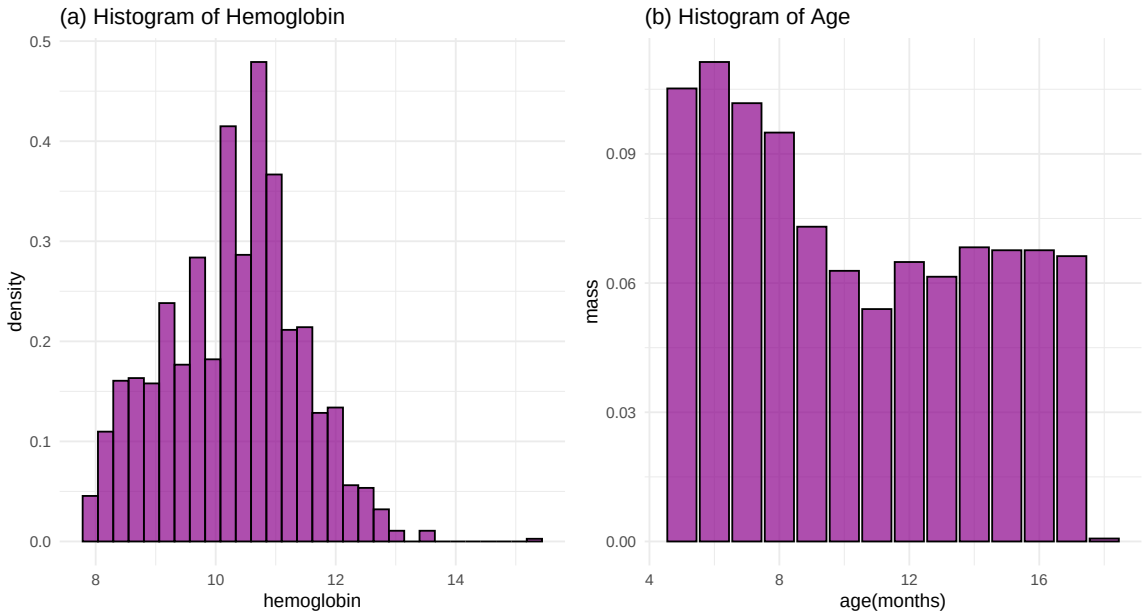

Figure 2.5: Histograms of hemoglobin and age for all participants.

Define T_{ij} as the jth infection time we observed for subject i, and denote n_i as the total event time for ith subject before the end of study or censoring, whichever comes first, we have $T_{i1} < T_{i2} < ... < T_{in_i}$. Define $N_i(t) = \sum_{j=1}^{n_i} I(T_{ij} \leq t)$ as the observed number of events taken from ith subject by time t; denote $\Delta N_i(t) = N_i(t+\Delta t^-) - N_i(t)$ as the number of events occurring in the small time interval $[t, t + \Delta t)$. The malaria infections can be modeled by the intensity function of $N_i(t)$. We combine the four vaccine arms as treatment group, and the vaccine efficacy is evaluated between the control and treatment groups.

2.6.1 Modeling Intensity as a Function of Calendar Time and Time Since the Most Recent Infection

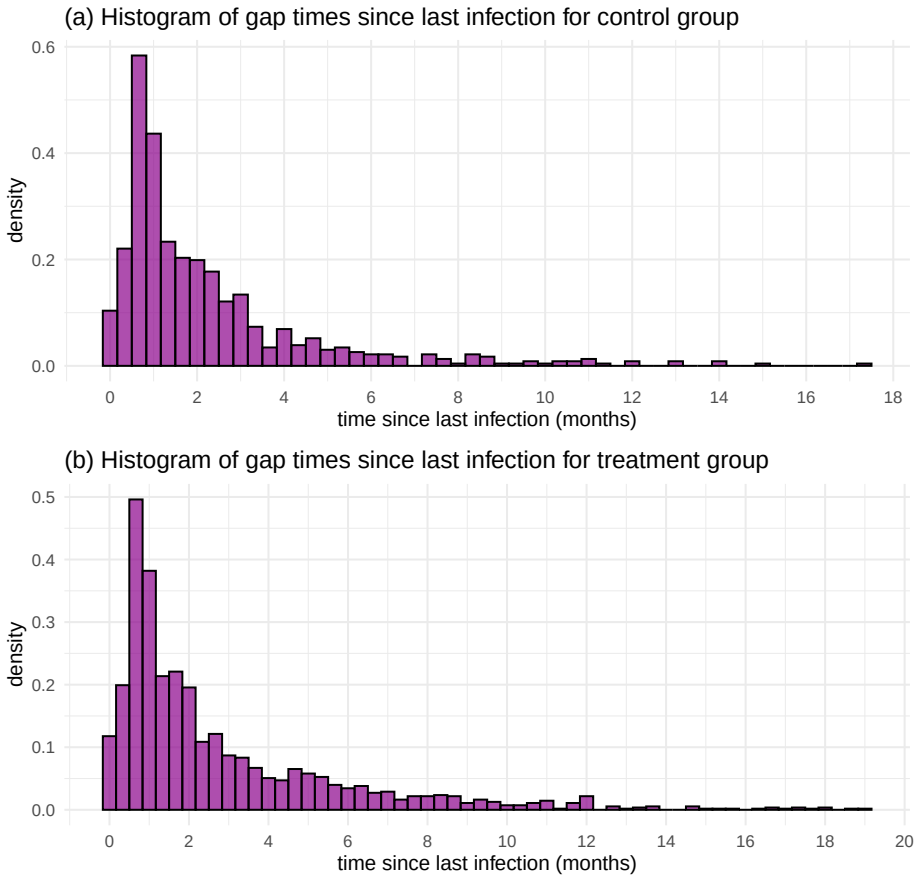

Figure 2.6: Histograms of gap times between infections for control and treatment groups in 20 months follow-up data.

Figure 2.6 is the histogram of gap times between infections for control group and treatment group. To assess the effectiveness of the RTS,S/AS01$_{\mathrm{E}}$ vaccine and explore how the vaccine effects are influenced by previous infections, as well as the effects of other risk factors, we consider the following multiplicative temporal intensity model

$$\lambda_i(t) = \exp\Big\{\alpha_0(t) + \alpha_1(t)\mathrm{Vacc}_i + \alpha_2(t)\mathrm{Agogo}_i + \alpha_3(t)\mathrm{Age}_i(year) + \beta\mathrm{Hemo}_i \\ +\gamma_0(t - T_{iN_i(t^-)})I(N_i(t^-) > 0) + \gamma_1(t - T_{iN_i(t^-)})I(N_i(t^-) > 0)Vacc_i\Big\} \quad (2.19)$$

with $\gamma_0(u,\theta) = \theta_0 + \theta_1 u$ and $\gamma_1(u,\theta) = \theta_2 + \theta_3 u$, where Vacc_i is the treatment group indicator (Vacc_i= 1 if assigned to one of the four RTS,S/AS01E vaccine arms, 0 if assigned to the control arm), and Agogo_i is the study site indicator (1= Agogo, 0 = Siaya), Age_i is the age in years at enrollment and Hemo_i is the standardized hemoglobin for ith subject.

Apply the Monte Carlo bootstrap cross-validation method as described in Section 2.2.4 to select the optimal bandwidth. 100 repetitions bootstrap cross-validation yield an optimal bandwidth of 1.35 months.

Table 2.8 presents the estimation results of parametric parameters β, θ_0, θ_1, θ_2 and θ_3, including their estimates, estimated standard errors and p values under the null hypotheses $H_0 : \beta = 0$, $H_0 : \theta_0 = 0$, $H_0 : \theta_1 = 0$, $H_0 : \theta_2 = 0$ and $H_0 : \theta_3 = 0$.

Table 2.9 shows the test statistics and p values under null hypotheses H_0: $\gamma_0(u,\theta) = \theta_0 + \theta_1 u$ and $\gamma_1(u,\theta) = \theta_2 + \theta_3 u$. At the 0.05 significant level, we draw same conclusions from the supremum and chi-square tests, which is the linearity hypothesis for $\gamma_0(u,\theta)$ and $\gamma_1(u,\theta)$ hold.

Figure 2.7 plots the test statistics and the Gaussian multiplier statistics. The gray lines represent 500 Gaussian multiplier approximations of test statistics under the null hypothesis H_0, while the red line depicts the test statistic obtained after fitting the model. Observing the plots, we notice that the test statistic for all components

falls within the gray region. This suggests no departure from the null hypothesis.

Table 2.8: Estimations, estimated standard error and p values of parametric parameters under model (2.19), using $h = 1.35$.

	β	θ_0	θ_1	θ_2	θ_3
EST	-0.148	0.814	-0.034	0.128	-0.039
ESE	0.024	0.112	0.018	0.131	0.020
p-value	$6.974e^{-10}$	$3.653e^{-13}$	0.059	0.329	0.051

Table 2.9: Test statistics and p values under model (2.19), T_1 is the test statistics of supreme test; T_2 is the test statistics of chi-square test; p_{2g} value is based on Gaussian multiplier distribution and p_{2c} value is based on chi-square distribution.

T_1	p_1	T_2	p_{2g}	p_{2c}
4.213	0.453	36.018	0.220	0.208

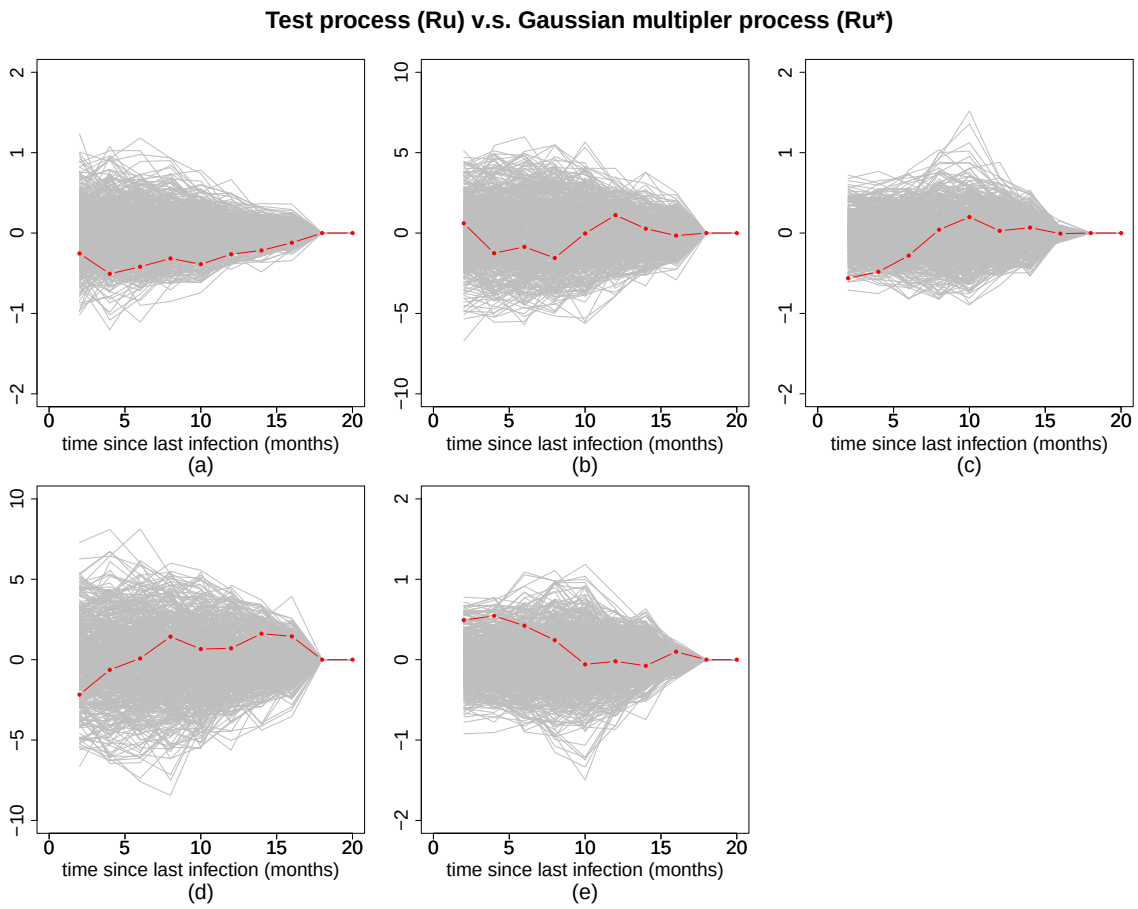

Figure 2.7: Test process R_u and Gaussian multiplier process R_u^* under model (2.19). The plots labeled (a) through (e) respectively represent the components β, θ_0, θ_1, θ_2 and θ_3.

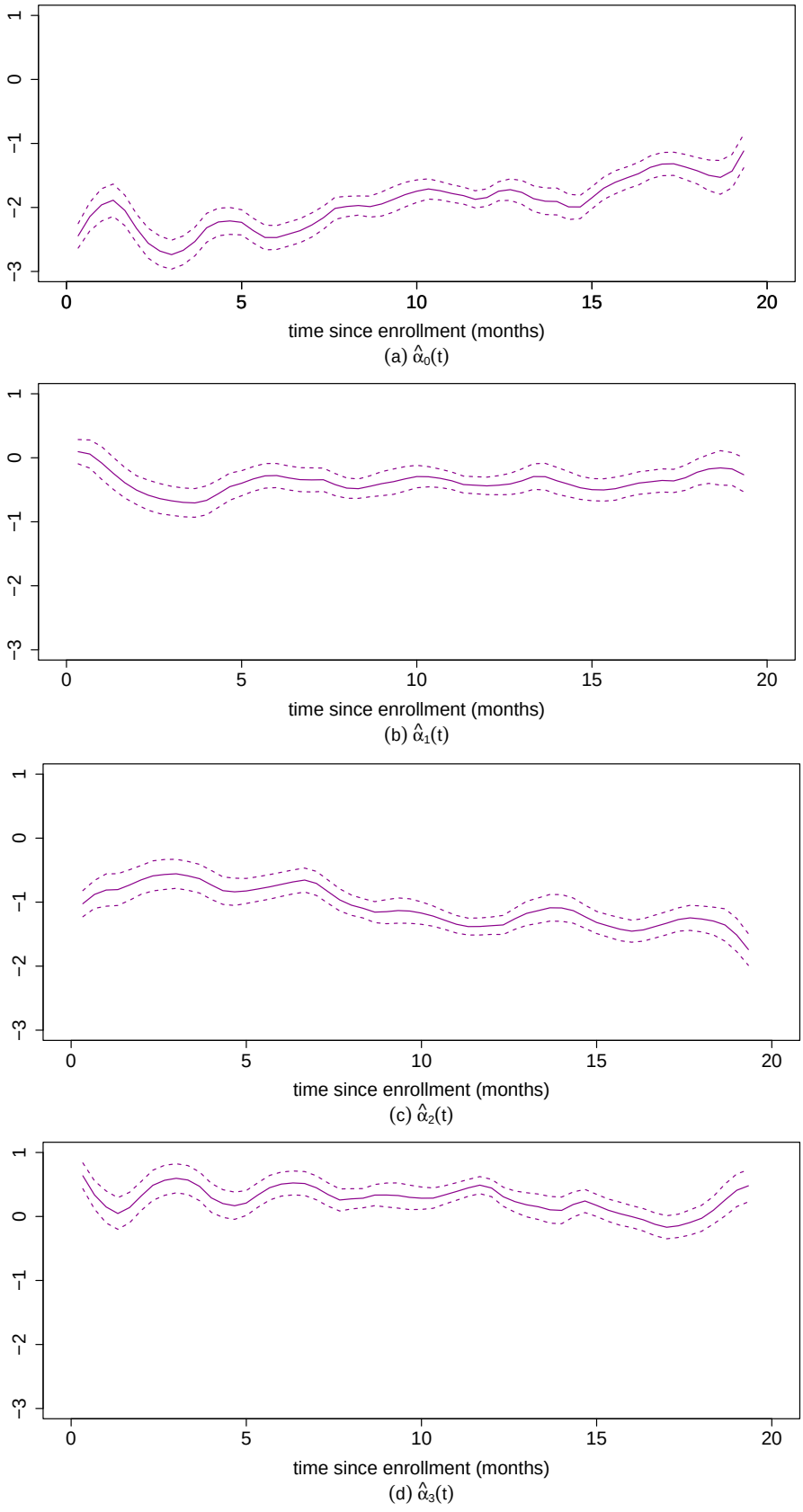

Figure 2.8: Estimation results of time-varying effects of covariates under model (2.19). The solid line represents the point estimate, while the shaded area signifies the 95% pointwise confidence interval.

Figure 2.8 presents the estimates and corresponding standard errors of $\alpha_0(t)$, $\alpha_1(t)$, $\alpha_2(t)$ and $\alpha_3(t)$. From the figure, we find that the baseline risk intensity increase with time. The risk of infection among participants residing in the Agogo site is lower compared to those residing in Siaya, and this difference becomes increasingly pronounced over time. The positive estimate of $\hat{\alpha}_3(t)$ also suggests that older children have higher risk of infections.

In the treatment group, participants consistently exhibit a lower infection risk compared to those in the control group.

Regarding the impact of prior infections, we observe an increase in the risk of subsequent infection within the control group, which gradually decreases over time. The infection risk associated with the most recent infection is initially higher in treatment group during about the first 3 months but subsequently diminishes, ultimately becoming lower than that of the control group.

To assess the quantity of protection against re-infection, we define vaccine efficacy as the percentage decrease in infection intensity among vaccinated individuals in comparison to those who are not vaccinated. Under model (2.19), the vaccine efficacy at time t equals

$$\mathrm{VE}(t) = 1 - \frac{\lambda_i(t|\mathrm{Vacc}_i = 1)}{\lambda_i(t|\mathrm{Vacc}_i = 0)} = 1 - \exp\left\{\alpha_1(t) + \gamma_1(t - T_{N_i(t^-)})I(N_i(t^-) > 0)\right\}.$$

When there are no prior infections, we have the vaccine efficacy against the first infection is represented by $\mathrm{VE}(t) = 1 - \exp\{\alpha_1(t)\}$. When prior infections exist, i.e. $I(N_i(t^-) > 0) = 1$, the vaccine efficacy for subsequent infections, including the second and beyond, is given by $\mathrm{VE}(t) = 1 - \exp\{\alpha_1(t) + \gamma_1(t - T_{N_i(t^-)})\}$. Figure 2.9 shows the estimated vaccine efficacy against both first infection and reinfections.

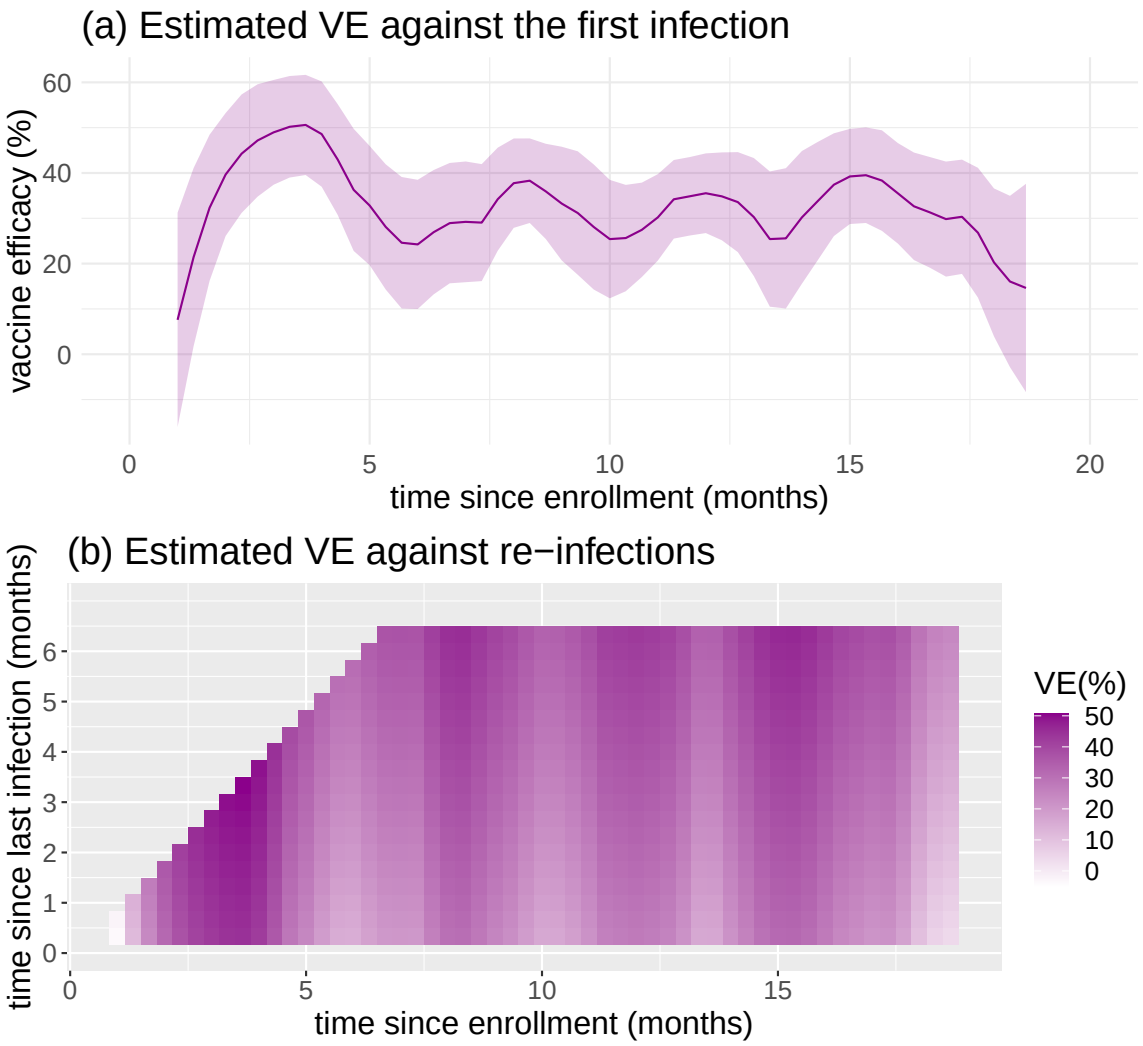

Figure 2.9: The estimated vaccine efficacy against the first infection (a) and re-infection (b) under model (2.19)

The vaccine efficacy for both the first infection and subsequent reinfections exhibits temporal variation. Strong efficacy effects are evident particularly within the first 5 months. However, after approximately 7 and 12 months, fluctuations occur, which could be partly attributed to the timing of vaccinations, as depicted in Figure 2.10.

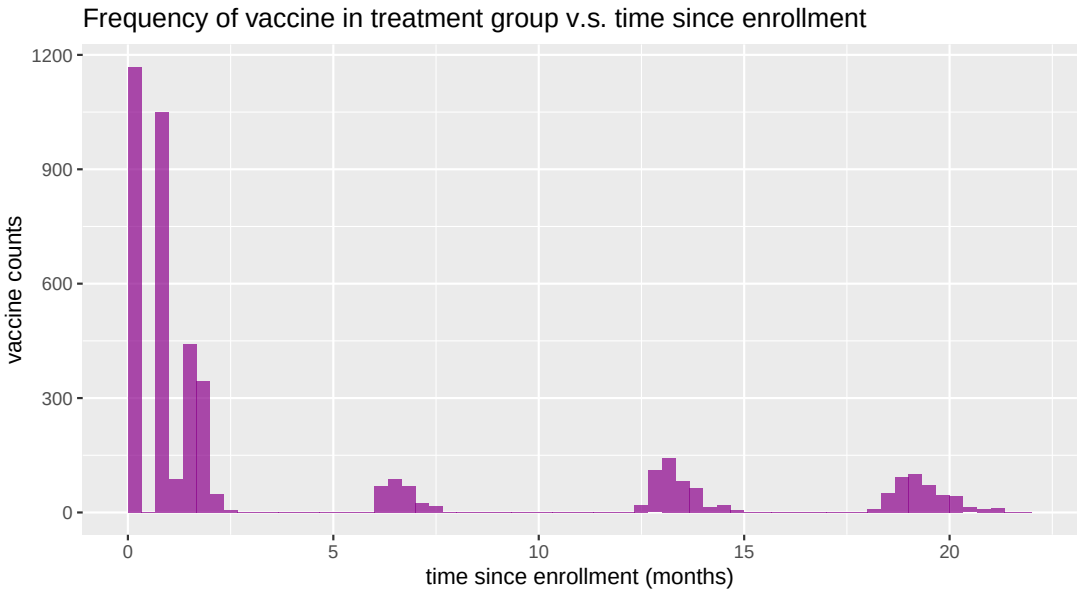

Figure 2.10: The frequency of vaccinations in treatment group over time since enrollment.

2.6.2 Modeling Intensity as a Function of Calendar Time and Time Since the Most Recent Vaccination

In this subsection, we model the effects of vaccinations parametrically and aim to investigate how these effects fluctuate over time since the last vaccination. Figure 2.11 shows the histogram of the gap times since last vaccinations when infections occur, distinguishing between control and treatment groups.

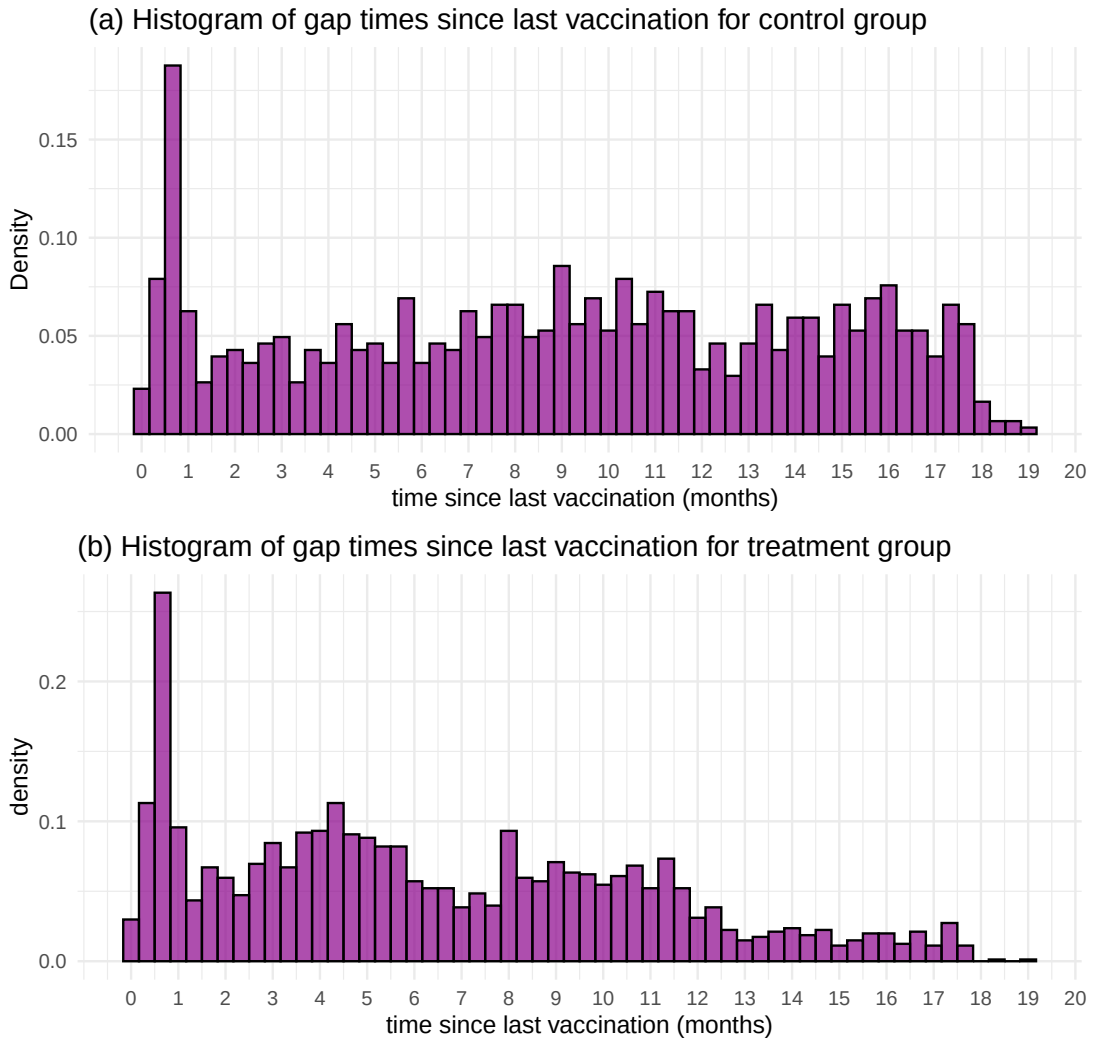

Figure 2.11: Histograms of gap time since last vaccination for control and treatment groups when infections occur in 20 months follow-up data.

Besides the definition of infection times and counting process of infections at the beginning of this section, we also need to define the vaccination times. Define T_{ik}^V as the kth vaccination time for subject i, and denote v_i as the total vaccine doses for i subject before the end of study or censoring, whichever comes first, we have $T_{i1}^V < T_{i2}^V < ... < T_{iv_i}^V$, we define counting process for the vaccination as $V_i(t) = \sum_{k=1}^{v_i} I(T_{ik}^V \leq t)$.

We model the intensity function of infections with two time scales as follows:

$$\lambda_i(t) = \exp\Big\{\alpha_0(t) + \alpha_1(t)\text{Agogo}_i + \alpha_2(t)\text{Age}_i(year) + \beta\text{Hemo}_i + \gamma(t - T^V_{iV_i(t^-)})I(V_i(t^-) > 0)\text{Vacc}_i\Big\}, \tag{2.20}$$

where $\gamma(U_i(t), \theta) = \theta_0 + \theta_1 U_{1i}(t)$. Vacc_i is the treatment group indicator, Agogo_i is the study site indicator (1= Agogo, 0 = Siaya), Age_i is the age in months at enrollment and Hemo_i is the standardized hemoglobin for ith subject.

Based on 100 repetitions of bootstrap as we discussed in Section 2.2.2, we use the selected bandwidth 1.42 months to fit model (2.20). Table 2.10 presents the estimation results of parametric parameters β, θ_0 and θ_1, including their estimates, estimated standard errors and p values under the null hypotheses $H_0 : \beta = 0$, $H_0 : \theta_0 = 0$ and $H_0 : \theta_1 = 0$. Table 2.11 shows the test statistics and p values under null hypotheses H_0: $\gamma(u, \theta) = \theta_0 + \theta_1 u$. At 0.05 significant level, both supremum and chi-square tests indicate acceptance of the linear hypothesise of $\gamma(u, \theta)$. Figure 2.12 are plots of test statistics and Gaussian multiplier statistics, the plots indicate that the test statistic for all components falls within the gray region. This suggests no departure from the null hypotheses.

Table 2.10: Estimations, estimated standard errors and p values of parametric parameters under model (2.20), using $h = 1.42$.

	β	θ_0	θ_1
EST	-0.195	-0.639	0.024
ESE	0.030	0.078	0.007
p value	8.032e^{-11}	2.220e^{-16}	6.068e^{-4}

Table 2.11: Test statistics and p values under model (2.20), T_1 is the test statistics of supreme test; T_2 is the test statistics of chi-square test; p_{2g} value is based on Gaussian multiplier distribution and p_{2c} value is based on chi-square distribution.

T_1	p_1-value	T_2	p_{2g}-value	p_{2c}-value
2.128	0.158	21.469	0.112	0.123

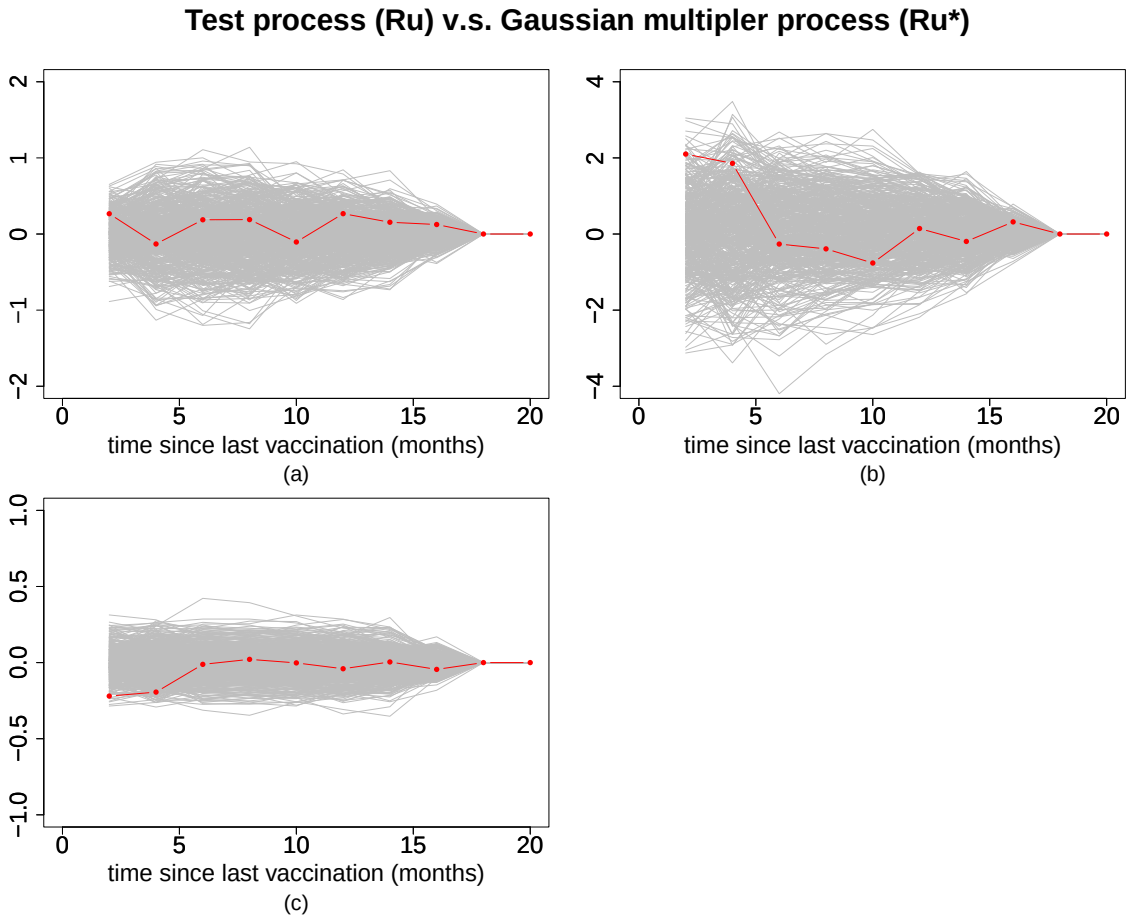

Figure 2.12: Test process $R(u)$ and Gaussian multiplier process $R(u)^*$ under model (2.20). The plots labeled (a) through (c) respectively represent the components β, θ_0 and θ_1.

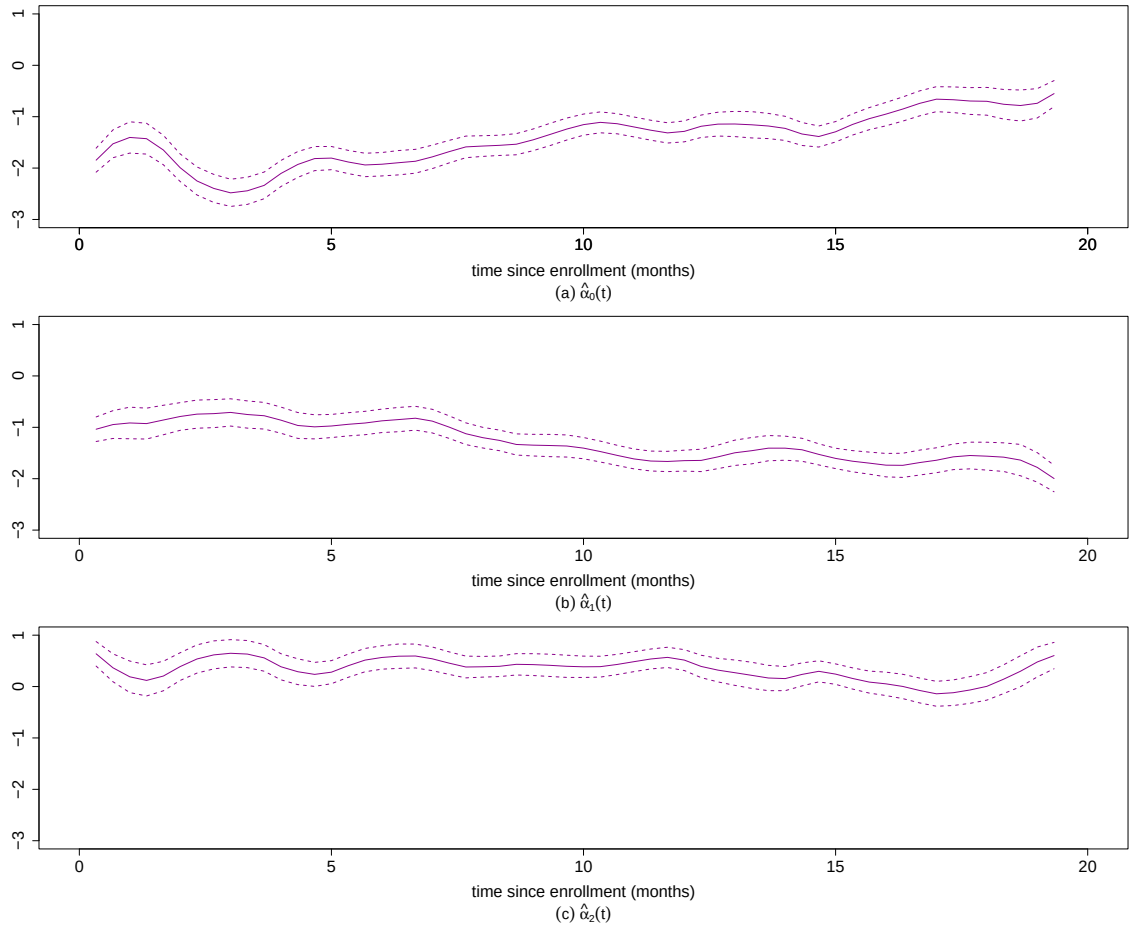

Figure 2.13: Estimation results of time-varying effects of covariates under model (2.20). The solid line represents the point estimate, while the shaded area signifies the 95% pointwise confidence interval.

From Figure 2.13, we get similar conclusions as we do in model (2.20). Firstly, as time passes, the baseline risk of infection rises. In comparing participants from the two study sites, the risk of infection is lower in Agogo. Older children are at a higher risk of infection, while those with higher levels of hemoglobin tend to have a relatively lower risk of infection.

The vaccine efficacy at time t can be defined as

$$\mathrm{VE}(t) = 1 - \frac{\lambda_i(t|\mathrm{Vacc}_i = 1)}{\lambda_i(t|\mathrm{Vacc}_i = 0)} = 1 - \exp\left\{\gamma(t - T_{iV_i(t^-)})I(V_i(t^-) > 0)\right\},$$

which can be demonstrated as the following plot. Basically, we can observe that as the time since last vaccination increases, the vaccine efficacy decrease.

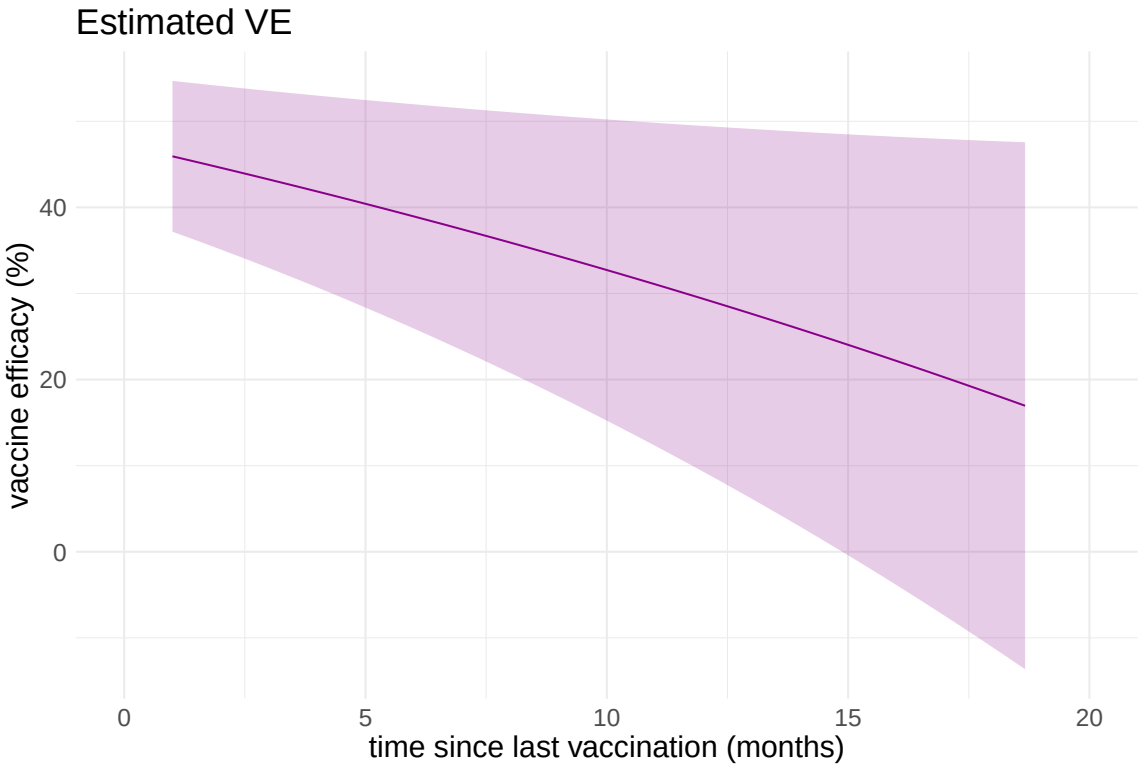

Figure 2.14: Estimated vaccine efficacy over time since last vaccination under model (2.20), the solid line represents the point estimate, while the shaded area signifies the 95% pointwise confidence interval.

CHAPTER 3: NONPARAMETRIC DYNAMIC INTENSITY MODELS WITH FRAILTY FOR RECURRENT EVENT DATA

3.1 Method and Estimation

3.1.1 Model Description

Suppose there is a random sample of n subjects, τ is the end of study time. T_{ij}, $i = 1, 2...n$, $j = 1, 2, ..., n_i$ are event times for subject i, where n_i is the total event number for subject i, we have $T_{i1} < T_{i2} < ... < T_{in_i} \leq \tau$. $X_i(t)$ $U_i(t)$ and $Z_i(t)$ are possible time-dependent covariates for subject i, and $U_i(t)$ could be derived from event history. Let counting process $N_i^*(t) = \sum_{j=1}^{n_i} I(T_{i,j} \leq t)$ be the number of events taken from ith subject by time t, where $I(\cdot)$ is the indicator function. $\Delta N_i^*(t) = N_i^*(t + \Delta t^-) - N_i^*(t)$ denotes the number of events occurring in the time interval $[t, t + \Delta t)$. Let $\mathcal{F}_{it}^*$ be the filtration generated by $N_i^*(t)$ and all covariates history up to time t for subject i, then the intensity of counting process $N_i^*(t)$ is defined as $\lambda_i^f(t) = \lim_{\Delta t \downarrow 0} \Pr(\Delta N_i^*(t) = 1 | \mathcal{F}_{it-}^*)/\Delta t$.

Let C_i be the censoring time, define $\tau_i = \min\{\tau, C_i\}$ as the end of follow-up time or censoring time whichever comes first, the events for subject i can only be observed before τ_i. $Y_i(t) = I(\tau_i \geq t)$ is the at-risk process. For subject i, $N_i(t) = N_i^*(t \wedge \tau_i)$ is the counting process for observed events; $\mathcal{F}_{it} = \mathcal{F}_{it}^* \vee C_i$ is the filtration of observed events, covariates history up to time t and censoring. Censoring is assumed to be non-informative in the sense that $E(dN_i(t)|\mathcal{F}_{t^-}) = E(dN_i(t)|\mathcal{F}_{t^-}^*) = Y_i(t)\lambda_i(t)dt$, where $\mathcal{F}_t^* = \vee_i^n \mathcal{F}_{it}^*$ and $\mathcal{F}_t = \vee_i^n \mathcal{F}_{it}$. The observed data consists of $\mathcal{D} = \Big\{N_i(t), Y_i(t), X_i(t), Z_i(t), U_i(t), t \in [0, C_i]\Big\}, (i = 1, ..., n)$.

The nonparametric dynamic intensity models incorporating frailty are proposed as

follows:

$$\lambda_i^f(t) = \xi_i \exp\{\alpha^{\mathsf{T}}(t)X_i(t) + \gamma^{\mathsf{T}}(U_i(t))Z_i(t)\}, \tag{3.1}$$

for $0 \leq t \leq \tau$, where $\alpha(\cdot)$ and $\gamma(\cdot)$ are p_1 and p_2 dimensional vectors of unspecified functions.

ξ_i is a frailty, also termed as a random effect, which is independent with both the subjects' covariates and stochastic processes. ξ_i are independent and identically distributed, following $Gamma(\theta, \theta)$, with probability density function

$$f(\xi) = \frac{\theta^\theta \xi^{\theta-1} e^{-\theta\xi}}{\Gamma(\theta)},$$

for $\xi > 0$, $\theta > 0$, where $\Gamma(\theta)$ is the Gamma function. Here, the two parameters of Gamma distribution are constrained to be identical in order to achieve scaling such that the resulting distribution has mean 1 and variance of $\frac{1}{\theta}$ to avoid identifiability problem. Here, the identifiability problem refers that if we multiply ξ_i and divide the whole exponential term by the same constant, ξ_i maintains its gamma distribution, but with a new scale parameter (Nileson *et al.* (1992)).

The possibly time-dependent covariate $U_i(t)$ offer flexibility in capturing the temporal patterns on an additional time scale. $U_i(t)$ could be derived from event or treatment history, enables the intensity function to change with these history. For example, by setting $U_i(t) = t - T_{i,N_i(t^-)}$ and $Z_i(t) = W_i(t)I(N_i(t^-) > 0)$, $U_i(t)$ stands the time since most recent event, there is an indicator $I(N_i(t^-) > 0)$ multiplying on $W_i(t)$ because the gap time $U_i(t)$ is only meaningful after the subject has experienced at least one event. By setting $U_i(t) = t - V_i(t)$ and $Z_i(t) = I(t > V_i(t))$, where $V_i(t)$ represents the time of vaccination or intervention, model (3.1) captures the alteration intensity following the vaccination or intervention.

The frailty term ξ_i, operating multiplicatively on the intensity function, seeks to

model unobserved heterogeneity among the subjects and induce the dependence structure among recurrent events with subjects. When ξ_i is greater than 1, it indicates that the subject is more likely to experience an event; conversely, when ξ_i is less than 1, it signifies that the subject is less likely to have an event.

3.1.2 Nonparametric Maximum Likelihood Estimation

For ease of representation, we denote $\lambda_i(t) = \exp\{\alpha(t)^\mathsf{T} X_i(t) + \gamma^\mathsf{T}(U_i(t))Z_i(t)\}$, then the intensity function can be written as

$$\lambda_i^f(t) = \xi_i \lambda_i(t) = \xi_i \exp\{\alpha^\mathsf{T}(t)X_i(t) + \gamma^\mathsf{T}(U_i(t))Z_i(t)\}.$$

By Cook and Lawless (2007), for given ξ_i, the likelihood function takes the form

$$\begin{aligned}\mathcal{L}_n(\alpha(\cdot),\gamma(\cdot),\theta) &= \prod_{0\le t\le\tau}\Big[\Big\{\prod_{i=1}^{n}\{Y_i(t)\xi_i\lambda_i(t)\}^{dN_i(t)}\Big\}\Big\{\{1-\sum_{i}^{n}Y_i(t)\xi_i\lambda_i(t)dt\}^{1-dN_{i\cdot}(t)}\Big\}\Big] \\ &= \Big\{\prod_{0\le t\le\tau}\prod_{i=1}^{n}\{Y_i(t)\xi_i\lambda_i(t)\}^{dN_i(t)}\Big\}\exp\Big\{-\sum_{i=1}^{n}\xi_i\int_0^\tau Y_i(t)\lambda_i(t)dt\Big\},\end{aligned} \tag{3.2}$$

where $N_{i\cdot}(t) = \sum_i^n N_i(t)$.

Integrate over ξ_i and get the observed likelihood function

$$\begin{aligned}&\prod_{i=1}^{n}\Big[\int_0^\infty\Big\{\prod_{0\le t\le\tau}\{Y_i(t)\xi_i\lambda_i(t)\}^{dN_i(t)}\Big\}\exp\Big\{-\xi_i\int_0^\tau Y_i(t)\lambda_i(t)dt\Big\}f(\xi_i)d\xi_i\Big] \\ =&\prod_{i=1}^{n}\Big[\Big\{\prod_{0\le t\le\tau}\{Y_i(t)\lambda_i(t)\}^{dN_i(t)}\Big\}\frac{\theta^\theta}{\Gamma(\theta)}\int_0^\infty\xi_i^{\theta+N_i(\tau)-1}\exp\Big\{-\xi_i\{\theta+\int_0^\tau Y_i(t)\lambda_i(t)dt\}\Big\}d\xi_i\Big] \\ =&\prod_{i=1}^{n}\Big[\Big\{\prod_{0\le t\le\tau}\{Y_i(t)\lambda_i(t)\}^{dN_i(t)}\Big\}\frac{\theta^\theta}{\Gamma(\theta)}\frac{\Gamma\{\theta+N_i(\tau)\}}{\{\theta+\int_0^\tau Y_i(t)\lambda_i(t)dt\}^{\theta+N_i(\tau)}}\Big],\end{aligned}$$

take logarithm, the log-likelihood function for observed data is obtained by

$$\log l_n^{\circ}(\alpha(\cdot),\gamma(\cdot),\theta) = \sum_{i=1}^{n}\Big[\int_0^{\tau}\log\{Y_i(t)\lambda_i(t)\}dN_i(t) + \theta\log\theta - \log\Gamma(\theta) + \log\Gamma\{\theta+N_i(\tau)\} - \{\theta+N_i(\tau)\}\log\Big\{\theta+\int_0^{\tau}Y_i(t)\lambda_i(t)dt\Big\}\Big]. \tag{3.3}$$

We use local linear smoothing method to estimate the nonparametric estimators $\alpha(\cdot)$ and $\gamma(\cdot)$. Assume that $X_i(t)$ and $Z_i(t)$ do not have common covariates, and assume $\alpha(t)$ and $\gamma(u)$ are smooth enough and their first derivatives exist. We do Taylor expansions for $\alpha(t)$ and $\gamma(u)$ at $\mathcal{N}_{t0}$ and $\mathcal{N}_{u0}$, the neighbourhoods of t_0 and u_0, getting

$$\alpha(t) = \alpha(t_0) + \dot{\alpha}(t_0)(t-t_0) + O((t-t_0)^2)$$

and

$$\gamma(u) = \gamma(u_0) + \dot{\gamma}(u_0)(u-u_0) + O((u-u_0)^2).$$

Denote $\lambda_i^{f*}(t,\vartheta^*|t_0,u_0)$ as the approximated intensity function localized at (t_0,u_0), we have

$$\lambda_i^{f*}(t,\vartheta^*|t_0,u_0) = \xi_i\exp\{\vartheta^{*T}(t_0,u_0)\tilde{X_i}^*(t|t_0,u_0)\} \tag{3.4}$$

where

$$\vartheta^*(t_0,u_0) = (\alpha^{\mathsf{T}}(t_0),\dot{\alpha}^{\mathsf{T}}(t_0),\gamma^{\mathsf{T}}(u_0),\dot{\gamma}^{\mathsf{T}}(u_0))^{\mathsf{T}}$$

and

$$\tilde{X_i}^*(t|t_0,u_0) = (X_i^\mathsf{T}(t), X_i^\mathsf{T}(t)(t-t_0), Z_i^\mathsf{T}(t), Z_i^\mathsf{T}(t)(U_i(t)-u_0))^\mathsf{T}.$$

Define

$$\lambda_i^*(t,\vartheta^*|t_0,u_0) = \exp\{\vartheta^{*T}(t_0,u_0)\tilde{X}_i^*(t|t_0,u_0)\}, \tag{3.5}$$

equation (3.4) can be written as

$$\lambda_i^{f*}(t,\vartheta^*|t_0,u_0) = \xi_i\lambda_i^*(t,\vartheta^*|t_0,u_0).$$

We employ the Expectation-Maximization (EM) algorithm to achieve maximum likelihood estimators. During this process, we treat ξ_i as a latent variable, obtaining its conditional expectation in E-step, and maximizing conditional expectation of the complete log-likelihood in M-step.

E Step. From Nileson *et al.* (1992), we know that given the observed data and current estimate $\hat{\alpha}(t)$, $\hat{\gamma}(u)$ and $\hat{\theta}$, ξ_i is conditional Gamma distributed and follows Gamma$(\hat{\theta}+N_i(\tau), \hat{\theta}+\int_0^\tau Y_i(t)\hat{\lambda}_i(t)dt)$, where $\hat{\lambda}_i(t)$ take the form:

$$\hat{\lambda}_i(t) = \exp\{\hat{\alpha}^\mathsf{T}(t)X_i(t) + \hat{\gamma}^\mathsf{T}(U_i(t))Z_i(t)\}. \tag{3.6}$$

Denote $\hat{E}(\cdot|\mathcal{D})$ as the conditional expectation given the observed data, we have closed forms of $\hat{E}(\xi_i|\mathcal{D})$ and $\hat{E}(\log\xi_i|\mathcal{D})$.

$$\hat{E}(\xi_i|\mathcal{D}) = \frac{\theta+N_i(\tau)}{\theta+\int_0^\tau Y_i(t)\hat{\lambda}_i(t))dt} \tag{3.7}$$

and

$$\hat{E}(\log \xi_i|\mathcal{D}) = \frac{\Gamma'(\theta + N_i(\tau))}{\Gamma(\theta + N_i(\tau))} - \log\Big\{\theta + \int_0^\tau Y_i(t)\hat{\lambda}_i(t))dt\Big\} \tag{3.8}$$

M-step. We maximize the conditional expectation of localized complete log-likelihood with respect to $\mathcal{D}$:

$$\begin{aligned} E[l(\vartheta^*|t_0,u_0)] = &\sum_{i=1}^n \int_0^\tau K_h(t-t_0)K_b(U_i(t)-u_0)\Big[\Big\{\hat{E}(\log\xi_i|\mathcal{D}) \\ &+ \log\{Y_i(t)\lambda_i^*(t,\vartheta^*|t_0,u_0)\}\Big\}dN_i(t) - Y_i(t)\hat{E}(\xi_i|\mathcal{D})\lambda_i^*(t,\vartheta^*|t_0,u_0)dt\Big] \\ &+ \sum_{i=1}^n \hat{E}\{\log f(\xi_i)|\mathcal{D}\} \end{aligned} \tag{3.9}$$

where $K_h(.) = K_1(./h)/h$, $K_b(.) = K_2(./b)/b$, $K_1(.)$, $K_2(.)$ are kernel functions, h and b are bandwidths, $\lambda_i^*(t,\vartheta^*|t_0,u_0)$ has been defined in (3.5).

Equation (3.9) can be written as $E[l(\vartheta^*,\theta|t_0,u_0)] = L_1(\vartheta^*|t_0,u_0) + L_2(\theta)$, with

$$\begin{aligned} L_1(\vartheta^*|t_0,u_0) = &\sum_{i=1}^n \int_0^\tau K_h(t-t_0)K_b(U_i(t)-u_0)\Big[\Big\{\hat{E}(\log\xi_i|\mathcal{D}) + \\ &\log\{Y_i(t)\lambda_i^*(t,\vartheta^*|t_0,u_0)\}\Big\}dN_i(t) - Y_i(t)\hat{E}(\xi_i|\mathcal{D})\lambda_i^*(t,\vartheta^*|t_0,u_0)dt\Big], \end{aligned} \tag{3.10}$$

and

$$\begin{aligned} L_2(\theta) &= \sum_i^n \hat{E}(\log f(\xi_i)|\mathcal{D}) \\ &= n\theta\log\theta - n\log\Gamma(\theta) + (\theta-1)\sum_{i=1}^n \hat{E}(\log\xi_i|\mathcal{D}) - \theta\sum_{i=1}^n \hat{E}(\xi_i|\mathcal{D}). \end{aligned} \tag{3.11}$$

Since $L_1(\vartheta^*|t_0,u_0)$ only contains ϑ^* and $L_2(\theta)$ only contains θ, we can maximize each of them individually. $\hat{E}(\xi_i|\mathcal{D})$ and $\hat{E}(\log\xi_i|\mathcal{D})$ can be calculated in the E-step based on the observed data and estimates at previous iteration.

For $L_1(\vartheta^*|t_0,u_0)$, we take derivative with respect to ϑ^*, get the local score function

$$U(\vartheta^*|t_0,u_0)=\sum_{i=1}^{n}\int_0^{\tau}K_h(t-t_0)K_b(U_i(t)-u_0)\Big\{dN_i(t) - Y_i(t)\hat{E}(\xi_i|\mathcal{D})\lambda_i^*(t,\vartheta^*|t_0,u_0)dt\Big\}\tilde{X}_i^*(t|t_0,u_0), \tag{3.12}$$

setting $U(\vartheta^*|t_0,u_0)=\mathbf{0}$, the bivariate estimate $\hat{\vartheta}^*(t_0,u_0)$ can be obtained through Newton-Rapson method.

Let $\hat{\alpha}(t_0,u_0)$ be the first p_1 elements and $\hat{\gamma}(t_0,u_0)$ be the $2p_1+1$ to $2p_1+p_2$ elements of $\hat{\vartheta}^*(t_0,u_0)$, aggregate them along each direction to get $\hat{\alpha}(t_0)$ and $\hat{\gamma}(u_0)$:

$$\hat{\alpha}(t_0)=n^{-1}\sum_{1=1}^{n}\hat{\alpha}(t_0,U_i(t_0)) \tag{3.13}$$

and

$$\hat{\gamma}(u_0)=n_{u_0}^{-1}\sum_{t_{u_0}\in\mathcal{V}_{u_0}}\hat{\gamma}(t_{u_0},u_0) \tag{3.14}$$

where $\mathcal{V}_{u_0}=\bigcup_{i=1}^{n}U_i^{-1}(u_0)$, $U_i^{-1}(u_0)=\{t:U_i(t)=u_0\}$, and $n_{u_0}=|\mathcal{V}_{u_0}|$ is the cardinality of $\mathcal{V}_{u_0}$.

For $L_2(\theta)$, we take derivative with respect to θ and set it to zero, getting

$$\log(\theta)-\frac{\Gamma'(\theta)}{\Gamma(\theta)}+\frac{1}{n}\sum_{i=1}^{n}\hat{E}(\log\xi_i|\mathcal{D})-\frac{1}{n}\sum_{i=1}^{n}\hat{E}(\xi_i|\mathcal{D})+1=0, \tag{3.15}$$

$\hat{\theta}$ is obtained by solving equation (3.15).

This process iterates between E-step and M-step until convergence, yielding the estimates $\hat{\alpha}(t)$, $\hat{\gamma}(u)$ and $\hat{\theta}$.

3.1.3 Computational Algorithm

In this section, we elaborate on the algorithm for the estimation procedure:

1. Generate the grid points over t and u.

2. Set initial values $\hat{\alpha}^{\{0\}}(t)$, $\hat{\gamma}^{\{0\}}(u)$ and $\hat{\theta}^{\{0\}}$.

3. Let $\hat{\alpha}^{\{k-1\}}(t)$, $\hat{\gamma}^{\{k-1\}}(u)$ and $\hat{\theta}^{\{k-1\}}$ be the estimates of $\alpha(t)$, $\gamma(u)$ and θ in $(k-1)$th iteration. At kth iteration, update the conditional expectation $\hat{E}^{\{k\}}(\xi_i|\mathcal{D})$ and $\hat{E}^{\{k\}}(\log\xi_i|\mathcal{D})$, by plugging $\hat{\alpha}^{\{k-1\}}(t)$, $\hat{\gamma}^{\{k-1\}}(u)$ and $\hat{\theta}^{\{k-1\}}$ in equation (3.7) and (3.8), more specifically,

$$\hat{E}^{\{k\}}(\xi_i) = \frac{\hat{\theta}^{\{k-1\}} + N_i(\tau)}{\hat{\theta}^{\{k-1\}} + \int_0^\tau Y_i(t)\lambda_i\{t, \hat{\alpha}^{\{k-1\}}(t), \hat{\gamma}^{\{k-1\}}(u_i(t))\}dt}$$

and

$$\hat{E}^{\{k\}}(\log\xi_i) = \frac{\Gamma'(\hat{\theta}^{\{k-1\}} + N_i(\tau))}{\Gamma(\hat{\theta}^{\{k-1\}} + N_i(\tau))} - \log\left\{\hat{\theta}^{\{k-1\}} + \int_0^\tau Y_i(t)\lambda_i\{t, \hat{\alpha}^{\{k-1\}}(t), \hat{\gamma}^{\{k-1\}}(u_i(t))\}dt\right\}$$

where

$$\lambda_i\{t, \hat{\alpha}^{\{k-1\}}(t), \hat{\gamma}^{\{k-1\}}(u_i(t))\} = \exp\{\hat{\alpha}^{\{k-1\}\mathsf{T}}(t)X_i(t) + \hat{\gamma}^{\{k-1\}\mathsf{T}}(U_i(t))Z_i(t)\}.$$

4. Update $\hat{\alpha}^{\{k\}}(t)$, $\hat{\gamma}^{\{k\}}(u)$ by solving $U(\vartheta^*|t_0, u_0) = \mathbf{0}$, where $U(\vartheta^*|t_0, u_0)$ takes the form of (3.12) with $\hat{E}(\xi_i|\mathcal{D})$ replaced by $\hat{E}^{\{k\}}(\xi_i|D)$, then we take corresponding elements and aggregate through equation (3.13) and (3.14). Update $\hat{\theta}^{\{k\}}$ by solving equation (3.15).

5. Repeat Step 3 and Step 4 iteratively until converge, estimates $\hat{\alpha}(t)$, $\hat{\gamma}(u)$ and $\hat{\theta}$ are $\hat{\alpha}^{\{k\}}(t)$, $\hat{\gamma}^{\{k\}}(u)$ and $\hat{\theta}^{\{k\}}$ at convergence.

3.1.4 Adaptive Estimation Algorithm

In previous algorithm, we have an implicit assumption $X_i(t) \neq Z_i(t)$, but if certain covariates are shared between $X_i(t)$ and $Z_i(t)$, $\alpha(t)$ and $\gamma(u)$ may not be distinguishable in a local area of using the local linear estimation method. For example, we consider model $\lambda_i^f(t) = \xi_i \exp\{\alpha(t) + \gamma(U_i(t))I(N_i(t^-) > 0)\}$, which is model (3.1) with $X_i(t) = 1$, $U_i(t) = t - t_{N_i(t^-)}$ and $Z_i(t) = I(N_i(t^-) > 0)$. If we consider a neighborhood that all subjects have experienced events, i.e. $P(N_i(t^-) > 0) = 1$ for $t \in \mathcal{N}_h(t_0)) = (t_0 - h, t_0 + h)$, $\alpha(t)$ and $\gamma(u)$ will have the identifiable problems. In this scenario, we develop an adaptive estimation

algorithm. At an early time stage, there must have some subjects that have not experienced events yet, we find a maximum of this time point t^*, to ensure when $t \in \mathcal{N}_h(t^*) = (t^*-h, t^*+h)$, $0 < P(N_i(t^-) = 0) < 1$ and $P(U_i(t) \in \mathcal{N}_b(u_0)) > 0$, where $\mathcal{N}_b(u_0)) = (u_0 - b, u_0 + b)$ is one bandwidth neighborhood of u_0. For $(t_0, u_0) \in \Delta = \{0 \le u \le t \le t^*\}$, where $t^* \le h + b$, $\alpha(t_0)$ and $\gamma(u_0)$ can be locally identified since $\tilde{X}_i^*(t|t_0, u_0)$ is full rank now. For later times $t > t^*$, we use integration method to estimate $\alpha(t)$, and then get estimate $\gamma(u)$ separately.

The overall estimation procedure follows a similar outline as sketched in Section 3.1.3, with the primary difference occurring when we update $\hat{\alpha}(t)$ and $\hat{\gamma}(u)$ in M-step of each iteration, instead of solving equation $U(\vartheta^*|t_0, u_0) = \mathbf{0}$ using double kernel method, we need to the following procedure to estimate $\alpha(t)$ and $\gamma(u)$ separately.

1. For $(t_0, u_0) \in \Delta = \{0 \le u \le t \le t^*\}$, where $t^* \le h + b$, we can estimate $\vartheta^*(t_0, u_0)$ by solving $U(\vartheta^*|t_0, u_0) = 0$ as we described in equation (3.12), then do aggregation by $\hat{\alpha}(t_0) = n^{-1} \sum_{i=1}^n \hat{\alpha}(t_0, U_i(t_0))$ and $\hat{\dot{\alpha}}(t_0) = n^{-1} \sum_{i=1}^n \hat{\dot{\alpha}}(t_0, U_i(t_0))$ to get $\hat{\alpha}(t_0)$ and $\hat{\dot{\alpha}}(t_0)$;

2. Suppose $\hat{\alpha}(t_{l_0})$ and $\hat{\dot{\alpha}}(t_{l_0})$ are the last points we can estimate by Step 1, consider the recursive formula $\hat{\alpha}(t_{l+1}) = \hat{\alpha}(t_l) + \Delta t \hat{\dot{\alpha}}(t_l)$, it hep us to get $\hat{\alpha}(t_{l_0+1}) = \hat{\alpha}(t_{l_0}) + \Delta t \hat{\dot{\alpha}}(t_{l_0})$. For $l = l_0 + 1, l_0 + 2$, and so on, the recursive formula is used to estimate $\alpha(t_{l+1})$ with the current estimate $\hat{\alpha}(t_l)$ and by estimating $\dot{\alpha}(t)$ at the grid points t_l using the following profile procedure with the plugged-in $\hat{\alpha}(t_l)$.

3. For $t_0 = t_l$ and u_0 be one of the grid points in $\mathcal{U}$, we firstly separate $\alpha(t_0)$ from $\vartheta^*(t_0, u_0)$ in notations. Let $\vartheta^*(t_0, u_0) = (\alpha^\mathsf{T}(t_0), \vartheta^{**\mathsf{T}}(t_0, u_0))^\mathsf{T}$ where $\vartheta^{**}(t_0, u_0) = (\dot{\alpha}^\mathsf{T}(t_0), \gamma^\mathsf{T}(u_0), \dot{\gamma}^\mathsf{T}(u_0))^\mathsf{T}$; let $\tilde{X_i}^*(t|t_0, u_0) = (X_i^\mathsf{T}(t), \tilde{X_i}^{**\mathsf{T}}(t|t_0, u_0))^\mathsf{T}$, where$\tilde{X_i}^{**}(t|t_0, u_0) = (X_i^\mathsf{T}(t)(t - t_0), Z_i^\mathsf{T}(t), Z_i^\mathsf{T}(t)(U_i(t) - u_0))^\mathsf{T}$. Then the localized intensity can be written as $\lambda_i^{f*}(t) = \xi_i \exp\{\alpha^\mathsf{T}(t) X_i(t) + \vartheta^{**\mathsf{T}}(t_0, u_0) \tilde{X_i}^{**}(t|t_0, u_0)\}$, and the conditional

complete log-likelihood that contains $\alpha(t_0)$ and $\vartheta^{**}(t_0,u_0)$ can be expressed as

$$L_1(\alpha,\vartheta^{**}|t_0,u_0)=\sum_{i=1}^{n}\int_0^\tau K_h(t-t_0)K_b(U_i(t)-u_0)\Big[\Big\{\hat{E}(\log\xi_i|\mathcal{D})+ \log\{Y_i(t)\lambda_i^{**}(t,\alpha,\vartheta^{**}|t_0,u_0)\}\Big\}dN_i(t)-Y_i(t)\hat{E}(\xi_i|\mathcal{D})\lambda_i^{**}(t,\alpha,\vartheta^{**}|t_0,u_0)dt\Big], \tag{3.16}$$

plug $\hat{\alpha}(t_0)$ for $\alpha(t_0)$ in $L_1(\alpha,\vartheta^{**}|t_0,u_0)$, we maximize the likelihood with respect to $\vartheta^{**}(t_0,u_0)$ and get $\hat{\vartheta}{**}(t_0,u_0)=(\hat{\alpha}^{\mathsf{T}}(t_0),\hat{\gamma}^{\mathsf{T}}(u_0),\hat{\dot{\gamma}}^{\mathsf{T}}(u_0))^{\mathsf{T}}$ for all grid points u_0. Aggregate them and get $\hat{\dot{\alpha}}(t_0)=n^{-1}\sum_{i=1}^{n}\hat{\dot{\alpha}}(t_0,U_i(t_0))$. Then, obtain $\hat{\gamma}(u_0)$ by aggregating all $\hat{\gamma}^{\mathsf{T}}(t_0,u_0)$ through equation (3.14).

4. Repeat Step 2 and Step 3 one point after another until estimate all the points.

3.2 Variance Estimator

In this session, we employed the weighted bootstrap procedure (Ma and R.Kosorok (2005)) to get the variance estimators of $\alpha(t)$, $\gamma(u)$ and θ. Let $\{\omega_1,\omega_2,...\omega_n\}$ be n independent realization of random variable Ω, which following exponential distribution with mean 1. $\{\omega_1,\omega_2,...\omega_n\}$ are independent with the observed data $\mathcal{D}$.

We aim to get the weighted bootstrap estimators through maximizing the weighted log-likelihood function of the observed data

$$\log l_n^{\omega}(\alpha(\cdot),\gamma(\cdot),\theta)=\sum_{i=1}^{n}\omega_i\Big[\int_0^\tau \log(Y_i(t)\lambda_i(t))dN_i(t)+\theta\log\theta-\log\Gamma(\theta)+\log\Gamma\{\theta+N_i(\tau)\} -\{\theta+N_i(\tau)\}\log\Big\{\theta+\int_0^\tau Y_i(t)\lambda_i(t)dt\Big\}\Big]. \tag{3.17}$$

To achieve the maximum weighted log-likelihood, we still implement the EM algorithm. The weighted version of conditional localized complete log-likelihood takes the form

$$E^{\omega}[l(\vartheta^*,\theta|t_0,u_0)]=L_1^{\omega}(\vartheta^*|t_0,u_0)+L_2^{\omega}(\theta),$$

where

$$L_1^{\omega}(\vartheta^*|t_0,u_0)=\sum_{i=1}^{n}\omega_i\int_0^{\tau}K_h(t-t_0)K_b(U_i(t)-u_0)\Big[\Big\{\hat{E}(\log\xi_i|\mathcal{D})+ \log\{Y_i(t)\lambda_i^*(t,\vartheta^*|t_0,u_0)\}\Big\}dN_i(t)-Y_i(t)\hat{E}(\xi_i|\mathcal{D})\lambda_i^*(t,\vartheta^*|t_0,u_0)dt\Big], \tag{3.18}$$

and

$$L_2^{\omega}(\theta)=\sum_{i=1}^{n}\omega_i\theta\log\theta-\sum_{i=1}^{n}\omega_i\log\Gamma(\theta)+(\theta-1)\sum_{i=1}^{n}\omega_i\hat{E}(\log\xi_i|\mathcal{D})-\theta\sum_{i=1}^{n}\omega_i\hat{E}(\xi_i|\mathcal{D}).$$

The estimation procedures follow the same approach as outlined in Section 3.1.2, but in a weighted version. In bth bootstrap, we generate a set of bootstrap weights $\{\omega_1,\omega_2,...\omega_n\}$, where $\omega_i \overset{\text{iid}}{\sim} Exp(1)$, then use the revised EM algorithm to get the weighted estimators $\hat{\alpha}^{\{b\}}(t)$, $\hat{\gamma}^{\{b\}}(u)$ and $\hat{\theta}^{\{b\}}$. Suppose we undertake 100 repetitions of weighted bootstrap, resulting in 100 weighted estimators, the sample variance of these weighted estimators are the estimated variance of $\hat{\alpha}(t)$, $\hat{\gamma}(u)$ and $\hat{\theta}$.

3.3 Simulation Studies

In this section, we perform simulations to demonstrate the effectiveness of the proposed methods. Section 3.3.1 focuses on the double kernel estimation method, and Section 3.3.2 delves into the adaptive method. All the variance estimators are obtained through weighted bootstrap that we illustrated in Section 3.2.

3.3.1 Simulations Using Double Kernel Algorithm

We consider the following intensity model

$$\lambda_i^f(t)=\xi_i\exp\{\alpha_0(t)+\alpha_1(t)X_i+\gamma(U_i(t))I(N_i(t^-)>0)W_i\}, \tag{3.19}$$

$t\in[0,\tau]$, with the following settings,

- $\tau=4$, For each subject i, we generate censoring time $C_i \overset{\text{iid}}{\sim} U(3,8)$, the study time

for each subject is the minimum of C_i and τ.

- $X_i \overset{\text{iid}}{\sim} Ber(0.5)$, $W_i \overset{\text{iid}}{\sim} U(0,1)$, $U_i(t) = t - T_{iN_i(t^-)}$ stands for the time since last event.
- $\alpha_0(t) = 1 - \log(1 + 0.2\log(1+t))$, $\alpha_1(t) = -0.5 + 0.1t$ and $\gamma(u) = -\frac{0.3}{1+u}$;
- random effect $\xi_i \overset{\text{iid}}{\sim} Gamma(\theta, \theta)$,

Under these settings, the average event numbers per subject is approximately 7. A smaller θ in gamma distribution implies a larger variability of the ξ_i. In this subsection, we simulate on three scenarios: $\theta = 1$, $\theta = 2$, and $\theta = 5$.

Table 3.1 summarizes the estimation results for parameter θ, and Figure 3.1, 3.2 and 3.3 illustrate the estimation results for the nonparameters $\alpha_0(t)$, $\alpha_1(t)$ and $\gamma(u)$.

Table 3.1: Bias, ESE, SEE and CP of θ under model (3.19) for $\theta = 1, 2, 5$ when n= 800, 1000, 1200, bandwidths are taken as $h = b = 0.9$, $h = b = 0.5$ and $h = b = 0.3$ for $\theta = 1$, $\theta = 2$, and $\theta = 5$, respectively.

θ	n	Bias	SEE	ESE	CP
1	800	0.005	0.067	0.072	0.968
	1000	0.008	0.063	0.064	0.940
	1200	0.008	0.057	0.058	0.956
2	800	0.001	0.223	0.225	0.962
	1000	0.003	0.204	0.200	0.948
	1200	0.011	0.165	0.178	0.964
5	800	-0.192	1.106	1.105	0.946
	1000	-0.111	0.968	1.007	0.952
	1200	-0.142	0.901	0.930	0.950

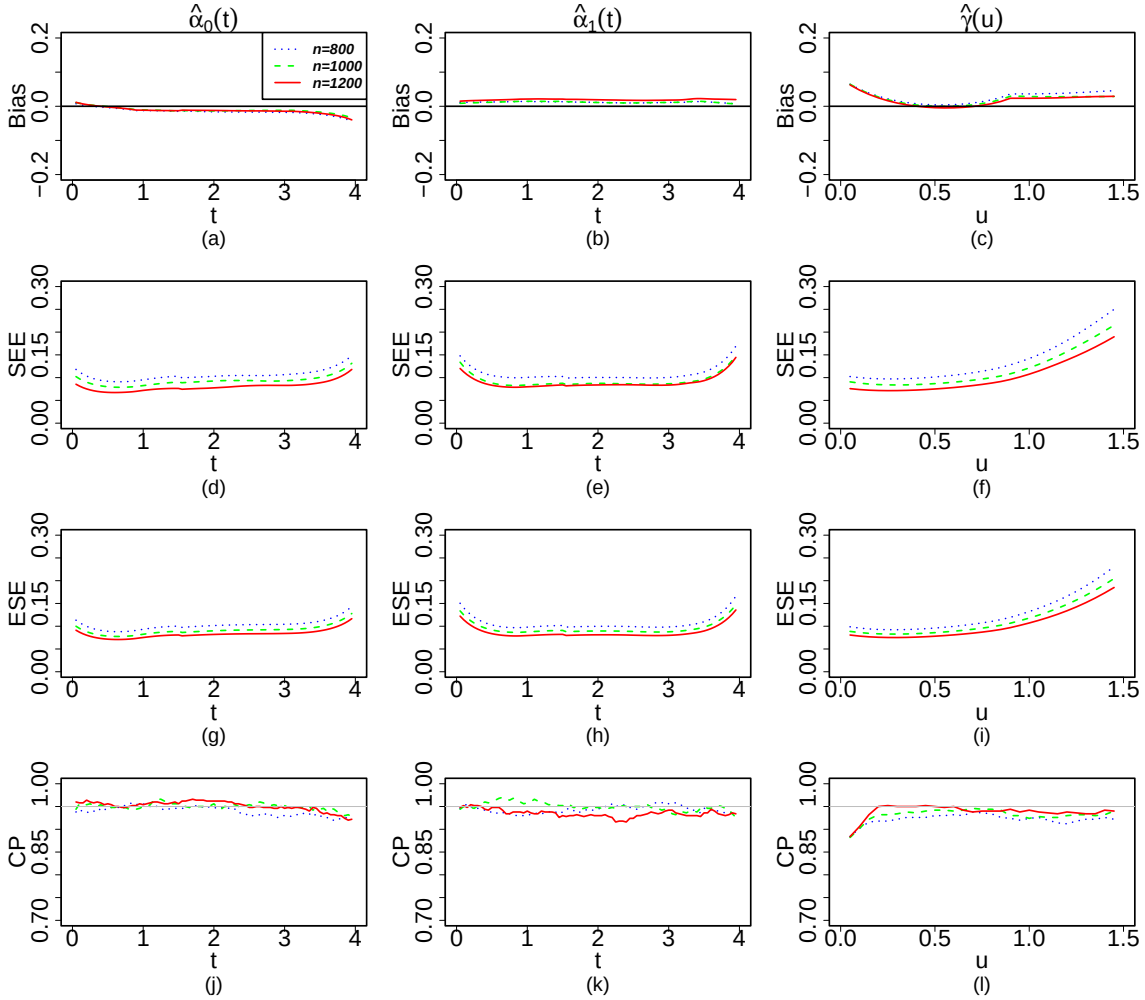

Figure 3.1: Bias, SEE, ESE and CP of $\alpha_0(t)$, $\alpha_1(t)$ and $\gamma(u)$ under model (3.19) for $\theta = 1$ with $h = b = 0.9$. The dotted, dashed and solid lines represent $n = 800$, $n = 1000$ and $n = 1200$, respectively.

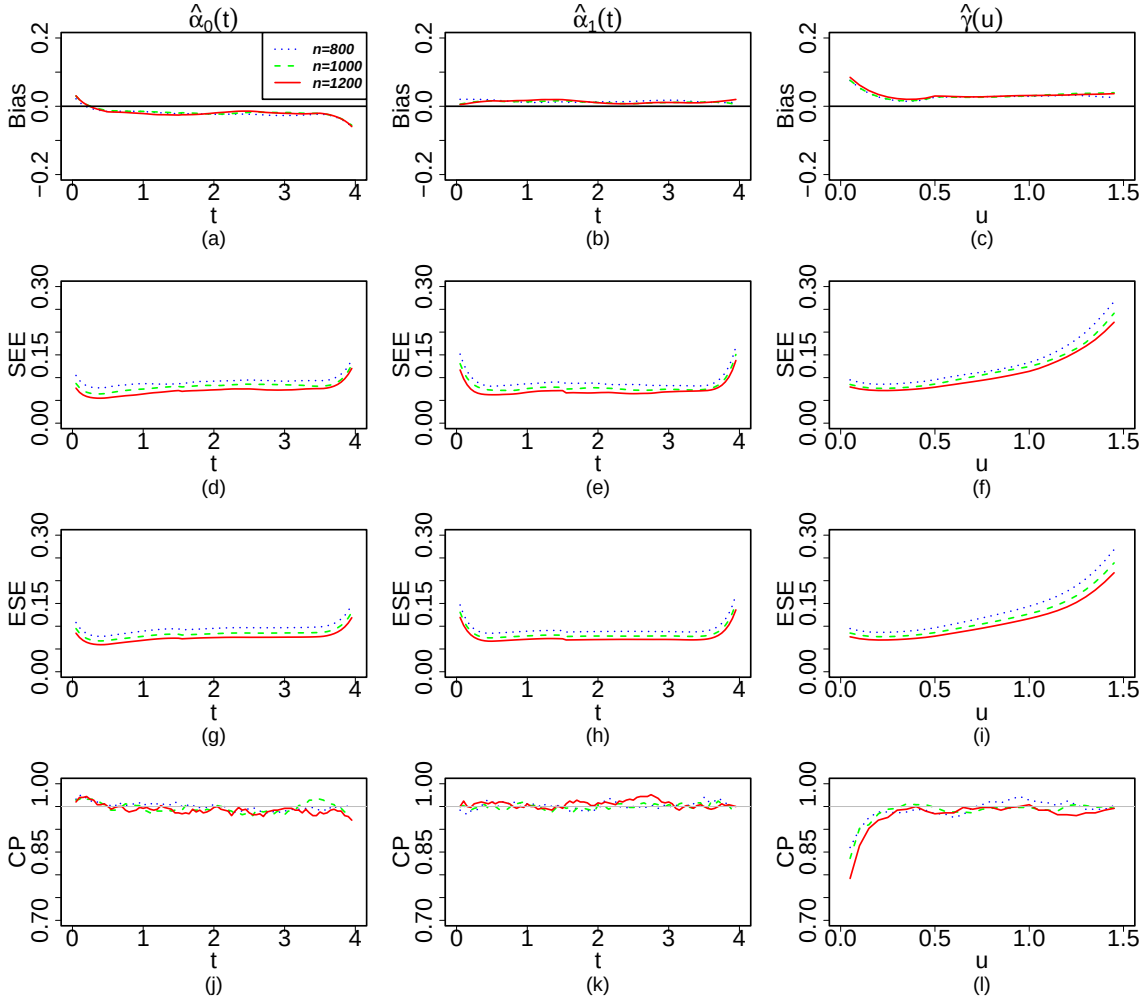

Figure 3.2: Bias, SEE, ESE and CP of $\alpha_0(t)$, $\alpha_1(t)$ and $\gamma(u)$ under model (3.19) for $\theta = 2$ with $h = b = 0.5$. The dotted, dashed and solid lines represent $n = 800$, $n = 1000$ and $n = 1200$, respectively.

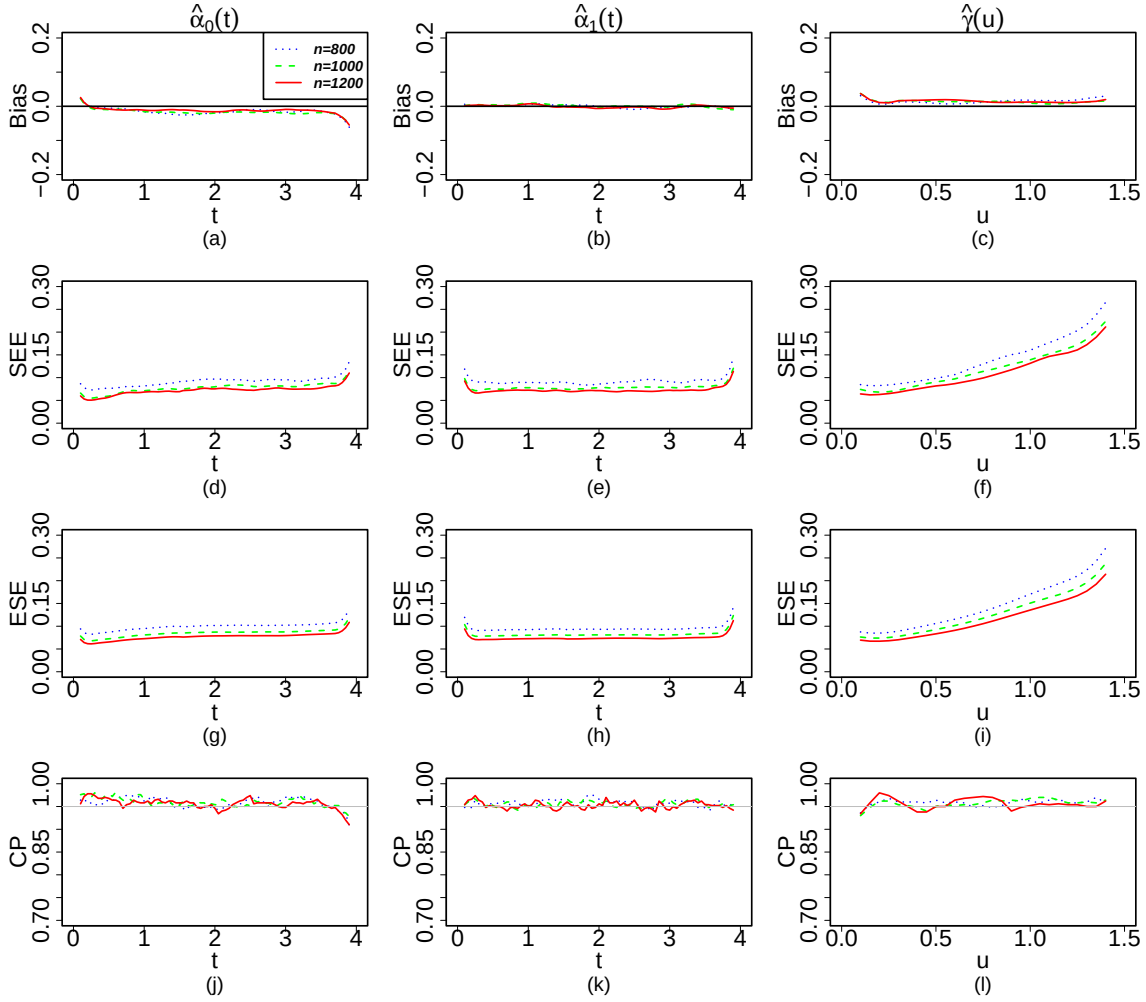

Figure 3.3: Bias, SEE, ESE and CP of $\alpha_0(t)$, $\alpha_1(t)$ and $\gamma(u)$ under model (3.19) for $\theta = 5$ with $h = b = 0.3$. The dotted, dashed and solid lines represent $n = 800$, $n = 1000$ and $n = 1200$, respectively.

Table 3.2 Figure 3.4 compares the estimation results using different bandwidths for $\theta = 2$ and $n = 800$. We can observe that all three pairs of bandwidths work well, the one with larger bandwidths ($h = b = 0.5$) have smaller biases, sample and estimated standard errors for both θ, $\alpha_0(t)$, $\alpha_1(t)$ and $\gamma(u)$. In practice, we can use the Monte Carlo cross validation method we described in Section 2.2 to choose the optimal bandwidth as well, h and b are not necessarily required to be the same.

Table 3.2: Bias, ESE, SEE and CP of θ under model (3.19) for $\theta = 2$ and $n = 800$, with bandwidths $h = b = 0.3$, $h = b = 0.4$ and $h = b = 0.5$.

$[h, b]$	Bias	SEE	ESE	CP
$[0.3, 0.3]$	-0.108	0.395	0.370	0.940
$[0.4, 0.4]$	-0.034	0.279	0.280	0.942
$[0.5, 0.5]$	0.001	0.225	0.223	0.962

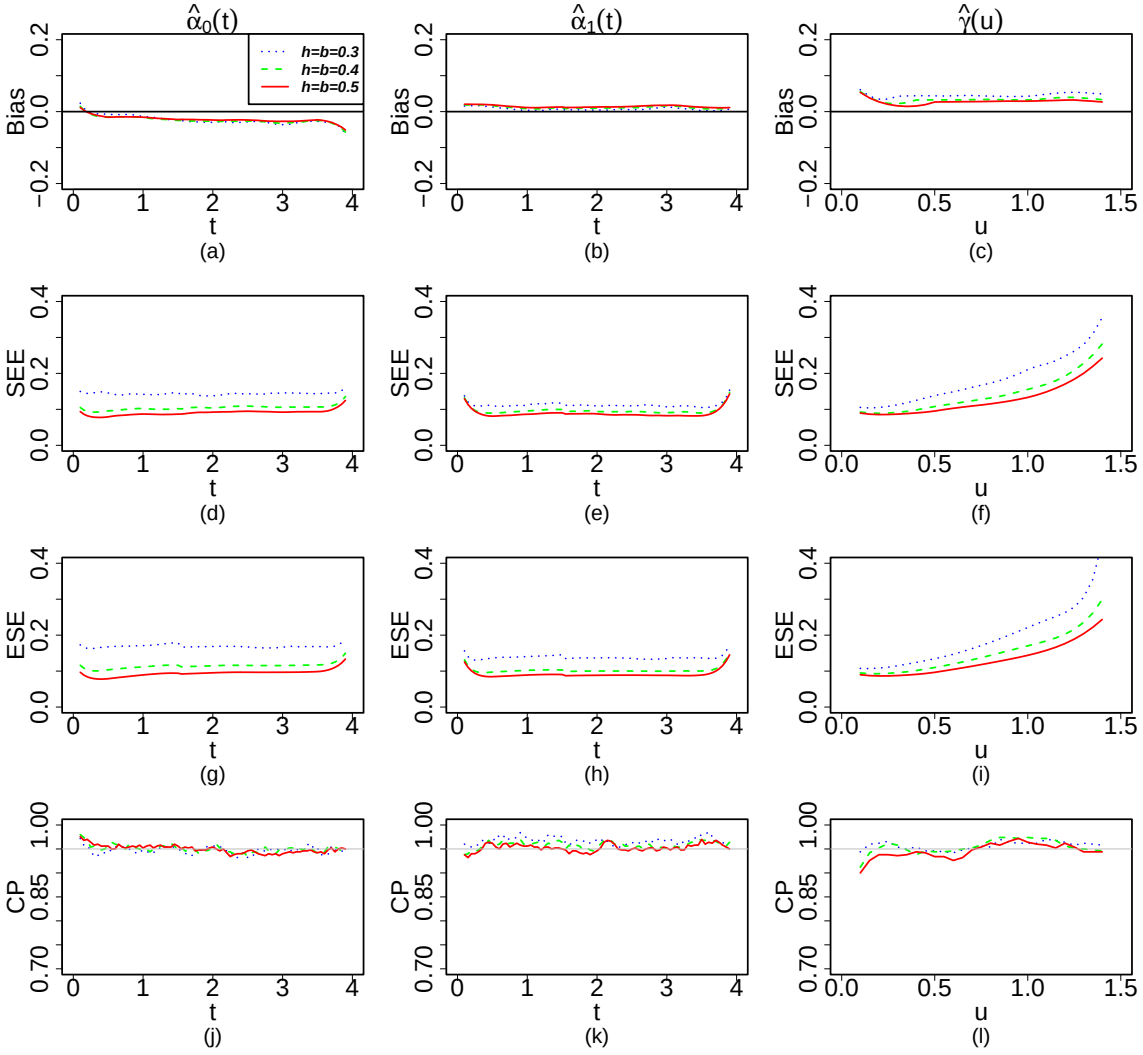

Figure 3.4: Bias, ESE, SEE and CP of $\alpha_0(t)$, $\alpha_1(t)$ and $\gamma(u)$ under model (3.19) for $\theta = 2$ and $n = 800$. The dotted, dashed and solid lines represent $h = b = 0.3$, $h = b = 0.4$ and $h = b = 0.5$, respectively.

3.3.2 Simulations Using Adaptive Algorithm

In this section, we conduct simulations using adaptive algorithm described in Section 3.1.4.

We generate data from the following intensity model

$$\lambda_i^f(t) = \xi_i \exp\{\alpha_0(t) + \gamma(U_i(t))I(N_i(t^- > 0)\}, \tag{3.20}$$

$t \in [0, \tau]$, with the following settings:

- Study time is on $[0, \tau]$, with $\tau = 4$. For each subject i, we generate censoring time $C_i \overset{\text{iid}}{\sim} U(3, 8)$, the study time for each individual is the minimum of C_i and τ;
- $U_i(t) = t - T_{iN_i(t^-)}$ represents the time since the most recent event;
- $\alpha_0(t) = 2 - \log(1 + 0.2\log(1 + t))$ and $\gamma(u) = -\frac{1-u}{\exp(1-u)^2}$;
- random effect $\xi_i \overset{\text{iid}}{\sim} Gamma(\theta, \theta)$ with $\theta = 2$;

As we discussed in Section 3.1.4, it may be challenging to identify $\alpha(t)$ and $\gamma(u)$ in certain local regions. It is necessary to employ the adaptive method to estimate the parameters, all variance estimators come from 100 repetitions of weighted bootstrap as we outlined in section 3.2.

Table 3.3 shows the estimation results for θ and Figure 3.5 shows the estimation results for $\alpha_0(t)$ and $\gamma(u)$.

Table 3.3: Bias, ESE, SEE and CP of θ under model (3.20) for $\theta = 2$ when $n = 800, 1000, 1200$, bandwidths $h = b = 0.3$.

n	Bias	SEE	ESE	CP
800	-0.035	0.162	0.167	0.944
1000	-0.039	0.148	0.148	0.932
1200	-0.037	0.135	0.136	0.942

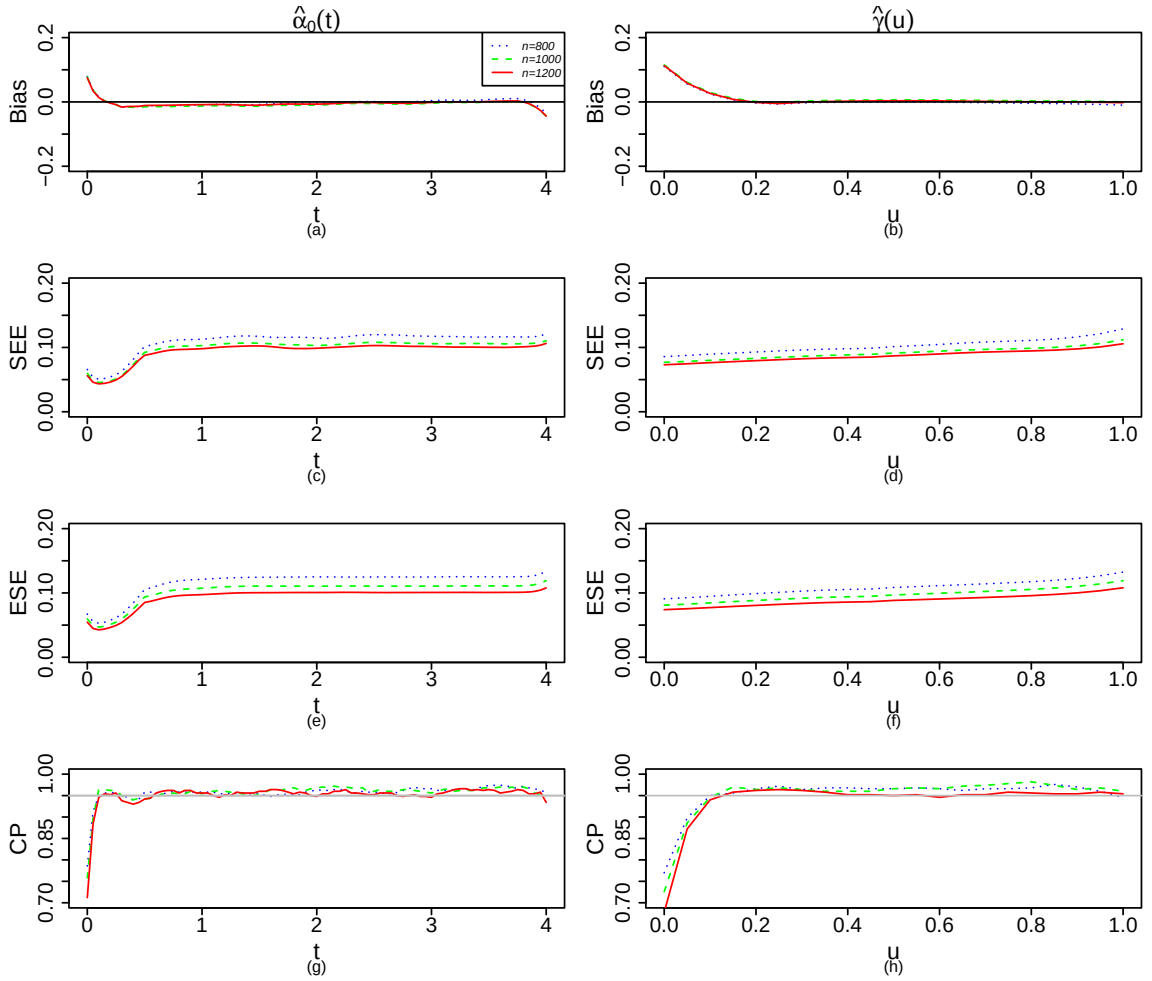

Figure 3.5: Bias, ESE, SEE and CP of $\alpha_0(t)$, $\alpha_1(t)$ and $\gamma(u)$ under model (3.20) for $\theta = 2$, bandwidths $h = b = 0.3$. The dotted, dashed and solid lines represent $n = 800$, $n = 1000$ and $n = 1200$, respectively.

3.4 Data Application

In this section, we apply the proposed nonparametric frailty model to the 32 months follow-up MAL-094 malaria vaccine trial data. Censoring is defined as: for each participant, visits after three consecutive missed scheduled visits and with no intervening unscheduled visits in-between are defined as censored. During the 32 months study time, 4633 malaria infections were observed before censored among 1461 participants. There are 1065 participants have experienced at least one infection with the largest infection number being 34.

We define the observed infection time and counting process the same way as we do in Section 2.6. Define T_{ij} as the jth infection time we observed for subject i, and denote n_i

as the total event time for i subject before the end of study or censoring, whichever comes first, we have $T_{i1} < T_{i2} < ... < T_{in_i}$. Define $N_i(t) = \sum_{j=1}^{n_i} I(T_{ij} \leq t)$ as the observed number of events taken from ith subject by time t; denote $\Delta N_i(t) = N_i(t + \Delta t^-) - N_i(t)$ as the number of events occurring in the small time interval $[t, t + \Delta t)$. The malaria infections are modeled by the conditional intensity function of $N_i(t)$. We combine the four vaccine arms as the the treatment group, and the vaccine efficacy is evaluated between the control group (arm) and treatment group.

To explore the effects of the RTS,S/AS01$_{\mathrm{E}}$ vaccine over time and their relations with prior infections, as well as to assess heterogeneity among participants, we consider the following multiplicative dynamic intensity model with frailty

$$\lambda_i(t) = \xi_i \exp\Big\{\alpha_0(t) + \alpha_1(t)\mathrm{Vacc}_i + \alpha_2(t)\mathrm{Agogo}_i + \alpha_3(t)\mathrm{Age}_i$$
$$+\gamma_0(t - T_{iN_i(t^-)})I(N_i(t^-) > 0) + \gamma_1(t - T_{iN_i(t^-)})I(N_i(t^-) > 0)Vacc_i\Big\} \tag{3.21}$$

where $\alpha_0(t)$, $\alpha_1(t)$, $\alpha_2(t)$ and $\alpha_3(t)$, $\gamma_0(u)$ and $\gamma_1(u)$ are unspecified functions, ξ_i follows $Gamma(\theta, \theta)$ with unknown parameter θ. Vacc_i is the treatment group indicator (Vacc_i= 1 if assigned to one of the four RTS,S/AS01E vaccine arms, 0 if assigned to the control arm), and Agogo_i is the study site indicator (1= Agogo, 0 = Siaya) and Age_i is the age in years at enrollment. The model is fitted on $u \in [0, 8]$, where is about $90th$ percentile of the gap times between infections.

Figure 3.6 presents the estimation results of the nonparametric parameters. We can see the baseline has slightly increasing trend. The participants living in Agogo have a lower risk while the older children tend to have a larger infection risk. Regarding the comparison between treatment and control groups, individuals in the treatment groups show a lower infection risk during the initial 7 months. However, this difference diminishes after 7 months. Following previous infections, the risk increases for the control group, maintaining a relatively consistent increase over time. For the treatment group, the risk is lower than the participants in control group.

The estimate of θ is 1.863, with estimated standard error we got in 100 times bootstraps is

0.511. The estimated θ indicates there is a spectrum of frailty among individuals. For each participant, we have the conditional expectation $E(\xi_i)$ at convergence. $E(\xi_i) > 1$ indicates the individual is more susceptible to the malaria infections while $E(\xi_i) < 1$ indicates the individual is relatively less likely to get infected, given all covariates are the same. When predicting the infection intensity for a particular individual, along with the covariates, we should also take this susceptibility or resistance into consideration.

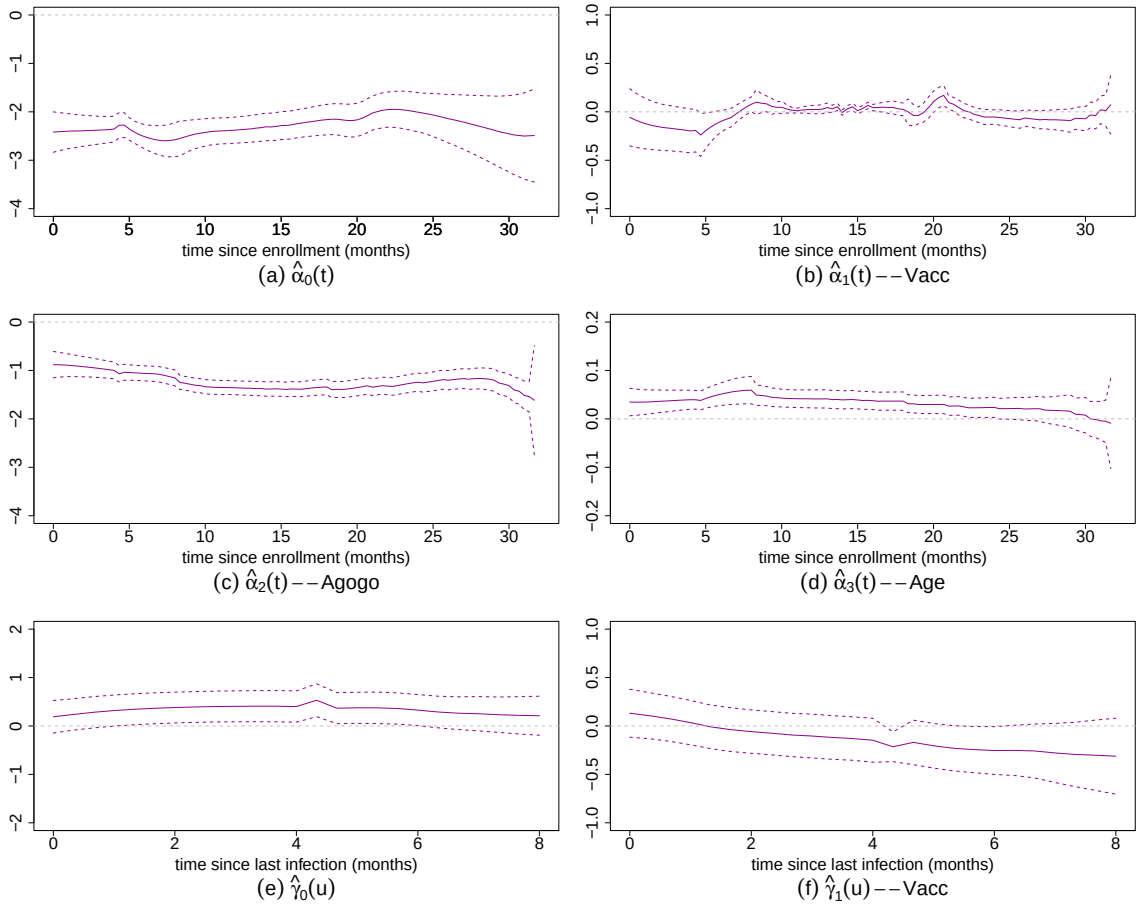

Figure 3.6: Estimation results of time-varying effects of covariates under model (3.21). The solid line represents the point estimate, while the shaded area signifies the 95% pointwise confidence interval.

CHAPTER 4: CONCLUSIONS AND FUTURE WORK

In conclusion, we propose two dynamic intensity models for recurrent event data, our simulation and application results demonstrate their effectiveness. The proposed models have the potential to address numerous clinical challenges, particular those involving the time-varying and covariate-varying effects, as well as the underlying structure of recurrent events is of interest.

In chapter 2, we developed a class of semiparametric intensity models, which are quite flexible through the choices of different link functions and covariates functions. We conducted simulations under logarithm link function and identity link function, and in the 20 months follow-up MAL-094 malaria data application, we fit the model with the logarithm link function under two scenarios, one models the intensity function with calendar time and time since last infection, the other models the intensity function with calendar time and time since last vaccination.

The semiparametric models could more robust if we consider the Box-Cox transformations $G(x) = \{(1+x)^{\rho} - 1\}/\rho(\rho \geq 0)$ as discussed in Zeng and Lin (2006), different value of ρ yield different link functions, with $\rho = 0$ corresponding to $G(x) = \log(1+x)$ and $\rho = 1$ corresponding to identity link function $G(x) = x$. We can select the value of ρ based on likelihood through Akaike (1985) information criterion (AIC) model selection method.

An additional extension of the semiparametric intensity models involves incorporating random effects. This enables us to parametrically assess covariate-varying effects and concurrently evaluate population homogeneity. We have conducted simulations on incorporating the Gamma frailty into the intensity function with a logarithm link function, the simulation results suggest the feasibility of this approach.

The nonparametric frailty models in Chapter 3 help us gain deeper insights into the underlying factors and vaccine efficacy of the MAL-094 malarial trial data. In future research,

we would like to develop hypothesis tests to assess the statistical significance of frailty models. These tests will help determine whether frailty is a significant factor, ensuring that accurate conclusions are drawn.

In malarial endemic areas, individuals may be exposed to diverse parasite genotypes. We also intend to propose a multivariate frailty model in the future to explore correlations between multiple types of events. This research aims to illuminate the impact of prior infections with specific parasite genotypes on the likelihood of subsequent infection with a different genotype.

REFERENCES

Akaike, H. (1985). *Prediction and entropy. In A Celebration of Statistics.* Springer-Verlag, New York.

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

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

Cai, B., Pellegrini, F., Pang, M., de Moor, C., Shen, C., Charu, V., and Tian, L. (2023). Bootstrapping the cross-validation estimate.

Cai, Z. and Sun, Y. (2003). A semiparametric additive regression model for longitudinal data. *Scandinavian Journal of Statistics*, **30**, 93–111.

Chang, H.-H. (2004). Estimating marginal effects in accelerated failure time models for serial sojourn times among repeated events. *Lifetime Data Analysis*, **10**, 175–190.

Chen, Q., Zeng, D., Ibrahim, J. G., Akacha, M., and Schmidli, H. (2013). Estimating time-varying effects for overdispersed recurrent events data with treatment switching. *Biometrika*, **100**(2), 339–354.

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

Epanechnikov, V. A. (1969). Non-parametric estimation of a multivariate probability density. *Theory of Probability Its Applications*, **14**(1), 153–158.

Fan, J. and Gijbels, I. . (1996). *Local polynomial modelling and its applications.* Chapman and Hall, London.

Klein, J. P. (1992). Semiparametric estimation of random effects using the cox model based on the em algorithm. *Biometrics*, **48**(3), 795–806.

Lawless, J. F.and Nadeau, C. C. R. J. (1997). Analysis of mean and rate functions for recurrent events. In D. Y. Lin and T. R. Fleming, editors, *Proceedings of the First Seattle Symposium in Biostatistics*, pages 37–49, New York, NY. Springer US.

Lawless, J. F. (1987). Regression methods for poisson process data. *Journal of the American Statistical Association*, **82**, 805–815.

Lin, D. Y., Wei, L. J., and Ying, Z. (1993). Checking the cox model with cumulative sums of martingale-based residuals. *Biometrika*, **80**(3), 557–572.

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

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

Ma, S. and R.Kosorok, M. (2005). Robust semiparametric m-estimation and weighted bootstrap. *Journal of Multivariate Analysis*, **97**, 190–217.

Martinussen, T. and Scheike, T. H. (1999). A semiparametric additive regression model for longitudinal data. *Biometrika*, **86**, 691–702.

Mazroui, Y., Mauguen, A., Mathoulin-Pelissier, S., MacGrogan, G., Brouste, V., and Rondeau, V. (2015). Time-varying coefficients in a multivariate frailty model: Application to breast cancer recurrences of several types and death. *Lifetime Data Anal*, **22**, 191–215.

Murphy, S. (1994). Consistency in a proportional hazard model incorporating a random effect. *The Annals of Statistics*, **22**(2), 712–731.

Murphy, S. (1995). Asymptotic theory for the frailty model. *The Annals of Statistics*, **23**(1), 182–198.

Nileson, G. G., Gill, R. D., Andersen, P. K., and Sorensen, T. I. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics*, **19**, 25–43.

Oakes, D. and Cui, L. (1994). On semiparametric inference for modulated renewal processes. *Biometrika*, **81**, 83–90.

Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *The Annals of Statistics*, **26**(1), 183–214.

Pena, E. A., Strawderman, R. L., and Hollander, M. (2001). Nonparametric estimation with recurrent event data. *Journal of the American Statistical Association*, **96**, 1299–1315.

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

Prentice, R. L., Williams, B., and Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika*, **68,2**, 373–379.

Qi, L., Sun, Y., and B.Gilbert, P. (2017). Generalized semiparametric varying-coefficient model for longitudinal data with applications to adaptive treatment randomizations. *Biometrics*, **73**, 441–451.

Strawderman, R. L. (2005). The accelerated gap times model. *Biometrika*, **92**, 647–666.

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

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

Vaart, A. V. D. (1998). *Asymptotic Statistics*. Cambridege University Press.

Wei, L., Lin, D., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American statistical association*, **84**, 1065–1073.

Yu, Z., Liu, L., Bravata, D. M., Williamsc, L. S., and Teppere, R. S. (2013). A semiparametric recurrent events model with time-varying coefficients. *statistics in Medicine*.

Zeng, D. and Lin, D. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika*, **93**, 627–640.

Zeng, D. and Lin, D. (2007). Semiparametric transformation models with random effects for recurrent events. *Journal of the American Statistical Association*, **102**, 167–180.

Zeng, D., Chen, Q., and IBRAHIM, J. G. (2009). Gamma frailty transformation models for multivariate survival times. *Biometrika*, **96**(2), 277–291.

APPENDIX A: PROOF OF THEOREMS IN CHAPTER 2

In this section, we approve Theorem 2.1 and Theorem 2.2 in Chapter 2. The following regularity conditions are assumed through proving.

Condition A.

(1) Censoring times are non-informative for the model in the sense of $E\{dN_i(t)|X_i(t), Z_i(t), U_i(t), C_i \geq t\} = E\{dN_i(t)|X_i(t), Z_i(t), U_i(t)\}$;

(2) The inverse function of the link function $g^{-1}(\cdot)$ is twice differentiable;

(3) The processes $X_i(t)$, $Z_i(t)$, $U_i(t)$ and $\lambda_i(t)$, $0 \leq t \leq \tau$, are left-continuous, bounded and their total variations are bounded by a constant;

(4) The kernel function $K(\cdot)$ is symmetric with compact support on $[-1, 1]$ and Lipschitz continuous; Bandwidths $h \asymp b$; $h \to 0$; $nh^2 \to \infty$ and nh^5 is bounded;

(5) $\alpha_0(t)$, $e_{11}(t)$ and $e_{12}(t)$ are twice differentiable on $t \in [0, \tau]$, $(e_{11}(t))^{-1}$ is bounded over $0 \leq t \leq \tau$;

(6) The matrices A_η and Σ_η are positive definite;

(6) The density $f_U(t, u)$ is twice continuously differentiable with respect to u and satisfies $\inf_{t\in[0,\tau],u\in\mathcal{U}} f_U(t, u) > 0$.

A.1 Proof of Theorem 2.1

We consider the left side of profile estimating equation (6),

$$
\begin{aligned}
&\frac{1}{n}U_\eta(\eta)\\
=&\frac{1}{n}\sum_{i=1}^{n}\int_{t_1}^{t_2}\Big\{\frac{\dot{\tilde{\lambda}}_i(t,\eta)}{\tilde{\lambda}_i(t,\eta)}dN_i(t)-Y_i(t)\dot{\tilde{\lambda}}_i(t,\eta)dt\Big\}\Big\{\Big(\frac{\partial\tilde{\alpha}(t,\eta)}{\partial\eta}\Big)^{\mathsf{T}}X_i(t)+\Big(\frac{\partial\zeta(U_i(t),\eta)}{\partial\eta}\Big)^{\mathsf{T}}P_i(t)\Big\}\\
=&\frac{1}{n}\sum_{i=1}^{n}\int_{t_1}^{t_2}\frac{\dot{\tilde{\lambda}}_i(t,\eta)}{\tilde{\lambda}_i(t,\eta)}\Big\{dN_i(t)-Y_i(t)\tilde{\lambda}_i(t,\eta)dt\Big\}\Big\{\Big(\frac{\partial\tilde{\alpha}(t,\eta)}{\partial\eta}\Big)^{\mathsf{T}}X_i(t)+\Big(\frac{\partial\zeta(U_i(t),\eta)}{\partial\eta}\Big)^{\mathsf{T}}P_i(t)\Big\}\\
\xrightarrow{p}&E\int_{t_1}^{t_2}\frac{\dot{\tilde{\lambda}}_i(t,\eta)}{\tilde{\lambda}_i(t,\eta)}\Big\{dN_i(t)-Y_i(t)\tilde{\lambda}_i(t,\eta)dt\Big\}\Big\{\Big(\frac{\partial\zeta(U_i(t),\eta)}{\partial\eta}\Big)^{\mathsf{T}}P_i(t)\\
&-e_{12}(t,U_i(t))^{\mathsf{T}}e_{11}(t,U_i(t))^{-1}X_i(t)\Big\}\\
=&E\int_{t_1}^{t_2}\frac{\dot{\tilde{\lambda}}_i(t,\eta)}{\tilde{\lambda}_i(t,\eta)}\Big\{\lambda_i(t)dt-Y_i(t)\tilde{\lambda}_i(t,\eta)dt\Big\}\Big\{\Big(\frac{\partial\zeta(U_i(t),\eta)}{\partial\eta}\Big)^{\mathsf{T}}P_i(t)\\
&-e_{12}(t,U_i(t))^{\mathsf{T}}e_{11}(t,U_i(t))^{-1}X_i(t)\Big\}\\
=&u(\eta)
\end{aligned}
$$

Consider the derivative of $U_\eta(\eta)$ with respect to η at η_0, we have

$$
\begin{aligned}
&-\frac{1}{n}\frac{\partial U_\eta(\eta)}{\partial\eta}\Big|_{\eta=\eta_0}\\
=&-\frac{1}{n}\sum_{i=1}^{n}\int_{t_1}^{t_2}\Big\{\frac{\ddot{\tilde{\lambda}}_i(t,\eta_0)\tilde{\lambda}_i(t,\eta_0)-[\dot{\tilde{\lambda}}_i(t,\eta_0)]^2}{[\tilde{\lambda}_i(t,\eta_0)]^2}dN_i(t)-Y_i(t)\ddot{\tilde{\lambda}}_i(t,\eta_0)dt\Big\}\\
&\Big[\Big(\frac{\partial\tilde{\alpha}(t,\eta_0)}{\partial\eta}\Big)^{\mathsf{T}}X_i(t)+\Big(\frac{\partial\zeta(U_i(t),\eta_0)}{\partial\eta}\Big)^{\mathsf{T}}P_i(t)\Big]^{\otimes 2}\\
&-\frac{1}{n}\sum_{i=1}^{n}\int_{t_1}^{t_2}[\frac{\dot{\tilde{\lambda}}_i(t,\eta_0)}{\tilde{\lambda}_i(t,\eta_0)}dN_i(t)-Y_i(t)\dot{\tilde{\lambda}}_i(t,\eta_0)dt]\Big[\Big(\frac{\partial^2\tilde{\alpha}(t,\eta_0)}{\partial\eta^2}\Big)^{\mathsf{T}}X_i(t)+\Big(\frac{\partial\zeta(U_i(t),\eta_0)}{\partial\eta^2}\Big)^{\mathsf{T}}P_i(t)\Big]
\end{aligned}
$$

The second term converges to zero as n goes to infinity, and the first term

$$
\begin{aligned}
&-\frac{1}{n}\sum_{i=1}^{n}\int_{t_1}^{t_2}\Big\{\frac{\ddot{\tilde{\lambda}}_i(t,\eta_0)\tilde{\lambda}_i(t,\eta_0)-[\dot{\tilde{\lambda}}_i(t,\eta_0)]^2}{[\tilde{\lambda}_i(t,\eta_0)]^2}dN_i(t)-Y_i(t)\ddot{\tilde{\lambda}}_i(t,\eta_0)dt\Big\}\\
&\times\Big[\Big(\frac{\partial\tilde{\alpha}(t,\eta_0)}{\partial\eta}\Big)^{\mathsf{T}}X_i(t)+\Big(\frac{\partial\zeta(U_i(t),\eta_0)}{\partial\eta}\Big)^{\mathsf{T}}P_i(t)\Big]^{\otimes 2}\\
\xrightarrow{p}\ &-E\int_{t_1}^{t_2}\Big\{\frac{\ddot{\tilde{\lambda}}_i(t,\eta_0)\tilde{\lambda}_i(t,\eta_0)-[\dot{\tilde{\lambda}}_i(t,\eta_0)]^2}{[\tilde{\lambda}_i(t,\eta_0)]^2}dN_i(t)\\
&-Y_i(t)\ddot{\tilde{\lambda}}_i(t,\eta_0)dt\Big\}\Big[\Big(\frac{\partial\zeta(U_i(t),\eta_0)}{\partial\eta}\Big)^{\mathsf{T}}P_i(t)-e_{12}(t,U_i(t))^{\mathsf{T}}e_{11}(t,U_i(t))^{-1}X_i(t)\Big]^{\otimes 2}\\
&=E\int_{t_1}^{t_2}Y_i(t)\frac{\dot{\lambda}_i(t)^2}{\lambda_i(t)}\Big[\Big(\frac{\partial\zeta(U_i(t),\eta_0)}{\partial\eta}\Big)^{\mathsf{T}}P_i(t)-e_{12}(t,U_i(t))^{\mathsf{T}}e_{11}(t,U_i(t))^{-1}X_i(t)\Big]^{\otimes 2}dt\\
&\equiv A_\eta
\end{aligned}
\tag{A.1}
$$

Since A_η is positive definite, η_0 is the unique root of $u(\eta)=0$ in a neighborhood of η_0. By theorem 5.9 of Vaart (1998), we have

$$\hat{\eta}\xrightarrow{p}\eta_0$$

For asymptotic normality of $\hat{\eta}$, we start with Taylor expansion,

$$U_\eta(\hat{\eta})=U_\eta(\eta_0)+\frac{\partial U_\eta(\eta)}{\partial\eta}\Big|_{\eta=\eta_0}(\hat{\eta}-\eta_0)+O_p(||\hat{\eta}-\eta_0||^2)$$

we know $U_\eta(\hat{\eta})=0$, so we have

$$\hat{\eta}-\eta_0=-\left(\frac{\partial U_\eta(\eta)}{\partial\eta}\Big|_{\eta=\eta_0}\right)^{-1}U_\eta(\eta_0)$$

$$\sqrt{n}(\hat{\eta}-\eta_0)=-\left(\frac{1}{n}\frac{\partial U_\eta(\eta)}{\partial\eta}\Big|_{\eta=\eta_0}\right)^{-1}\frac{1}{\sqrt{n}}U_\eta(\eta_0)\tag{A.2}$$

then consider

$$\begin{aligned}\frac{1}{\sqrt{n}}U_\eta(\eta_0) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{t_1}^{t_2}\Big\{\frac{\dot{\tilde{\lambda}}_i(t,\eta_0)}{\tilde{\lambda}_i(t,\eta_0)}dN_i(t) - Y_i(t)\dot{\tilde{\lambda}}_i(t,\eta_0)dt\Big\}\\ &\times\Big\{\Big(\frac{\partial\tilde{\alpha}(t,\eta_0)}{\partial\eta}\Big)^{\mathsf{T}}X_i(t) + \Big(\frac{\partial\zeta(\eta_0,U_i(t))}{\partial\eta}\Big)^{\mathsf{T}}P_i(t)\Big\}\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{t_1}^{t_2}\frac{\dot{\tilde{\lambda}}_i(t,\eta_0)}{\tilde{\lambda}_i(t,\eta_0)}\Big\{dN_i(t) - Y_i(t)g^{-1}\{\alpha_0^{\mathsf{T}}(t)X_i(t) + \zeta^{\mathsf{T}}(U_i(t),\eta_0)P_i(t)\}dt\\ &+ Y_i(t)g^{-1}\{\alpha_0^{\mathsf{T}}(t)X_i(t) + \zeta^{\mathsf{T}}(U_i(t),\eta_0)P_i(t)\}dt - Y_i(t)g^{-1}\{\tilde{\alpha}(t,\eta_0)X_i(t) + \zeta(U_i(t),\eta_0)^{\mathsf{T}}P_i(t)\}dt\Big\}\\ &\times\Big\{\Big(\frac{\partial\tilde{\alpha}(t,\eta_0)}{\partial\eta}\Big)^{\mathsf{T}}X_i(t) + \Big(\frac{\partial\zeta(U_i(t),\eta_0)}{\partial\eta}\Big)^{\mathsf{T}}P_i(t)\Big\}\end{aligned} \tag{A.3}$$

By Lemma 1 in Lin *et al.* (2001), the last two terms equal

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{t_1}^{t_2}Y_i(t)\dot{g}^{-1}\{\alpha_0^{\mathsf{T}}(t)X_i(t) + \zeta^{\mathsf{T}}(U_i(t),\eta_0)P_i(t)\}[\alpha_0^{\mathsf{T}}(t) - \tilde{\alpha}^{\mathsf{T}}(t,\eta_0)]X_i(t)\\ &\times\Big\{\Big(\frac{\partial\tilde{\alpha}(t,\eta_0)}{\partial\eta}\Big)^{\mathsf{T}}X_i(t) + \Big(\frac{\partial\zeta(U_i(t),\eta_0)}{\partial\eta}\Big)^{\mathsf{T}}P_i(t)\Big\} = op(1)\end{aligned}$$

so

$$\begin{aligned}\frac{1}{\sqrt{n}}U_\eta(\eta_0) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{t_1}^{t_2}\frac{\dot{\tilde{\lambda}}_i(t,\eta_0)}{\tilde{\lambda}_i(t,\eta_0)}\Big\{\Big(\frac{\partial\tilde{\alpha}(t,\eta_0)}{\partial\eta}\Big)^{\mathsf{T}}X_i(t) + \Big(\frac{\partial\zeta(U_i(t),\eta_0)}{\partial\eta}\Big)^{\mathsf{T}}P_i(t)\Big\}dM_i(t) + op(1)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{t_1}^{t_2}\frac{\dot{\tilde{\lambda}}_i(t,\eta_0)}{\tilde{\lambda}_i(t,\eta_0)}\Big\{\Big(\frac{\partial\zeta(\eta_0,U_i(t))}{\partial\eta}\Big)^{\mathsf{T}}P_i(t) - (e_{12}(t))^{\mathsf{T}}(e_{11}(t))^{-1}X_i(t)\Big\}dM_i(t)\\ &+ op(1)\end{aligned} \tag{A.4}$$

By martingale central limit theorem, $\frac{1}{\sqrt{n}}U_\eta(\eta_0) \sim N(0,\Sigma_\eta)$, where

$$\Sigma_\eta = E\Big\{\int_{t_1}^{t_2}\frac{\dot{\lambda}_i(t,\eta_0)}{\lambda_i(t,\eta_0)}\Big[\Big(\frac{\partial\zeta(U_i(t),\eta_0)}{\partial\eta}\Big)^{\mathsf{T}}P_i(t) - e_{12}(t)^{\mathsf{T}}(e_{11}(t))^{-1}X_i(t)\Big]dM_i(t)\Big\}^{\otimes 2}$$

By slustky theorem and combine with $-\frac{1}{n}\frac{\partial U_\eta(\eta)}{\partial \eta}\Big|_{\eta=\eta_0} \xrightarrow{p} A_\eta$,

$$\sqrt{n}(\hat{\eta}-\eta_0) \xrightarrow{p} N(0, A_\eta^{-1}\Sigma_\eta A_\eta^{-1})$$

Σ_η can be estimated by

$$\hat{\Sigma}_\eta = \frac{1}{n}\sum_{i=1}^{n}\Big[\int_{t_1}^{t_2}\frac{\hat{\dot{\lambda}}_i(t)}{\hat{\lambda}_i(t)}\Big\{\Big(\frac{\partial\zeta(U_i(t),\hat{\eta})}{\partial\eta}\Big)^{\mathsf{T}}P_i(t)-\hat{E}_{12}(t)^{\mathsf{T}}\hat{E}_{11}(t)^{-1}X_i(t)\Big\}\Big\{dN_i(t)-Y_i(t)\hat{\lambda}_i(t)dt\Big\}\Big]^{\otimes 2}$$

where $0<t_1<t_2<\tau$. A_η can be estimated by

$$\hat{A}_\eta = \frac{1}{n}\sum_{i=1}^{n}\int_{t_1}^{t_2}\frac{\hat{\dot{\lambda}}^2(t)}{\hat{\lambda}_i(t)}\Big\{\Big(\frac{\partial\zeta(U_i(t),\hat{\eta})}{\partial\eta}\Big)^{\mathsf{T}}P_i(t)-\hat{E}_{12}(t)^{\mathsf{T}}\hat{E}_{11}(t)^{-1}X_i(t)\Big\}^{\otimes 2}dt$$

A.2 Proof of Theorem 2.2

Now, we derive the asymptotic property for $\hat{\alpha}(t)=\tilde{\alpha}(t,\hat{\eta})$,

$$\begin{aligned}&(nh)^{1/2}[\hat{\alpha}(t)-\alpha_0(t)-\frac{1}{2}\mu_2 h^2\ddot{\alpha_0}^{\mathsf{T}}(t)]\\ =&(nh)^{1/2}[\tilde{\alpha}(t,\hat{\eta})-\tilde{\alpha}(t,\eta_0)+\tilde{\alpha}(t,\eta_0)-\alpha_0(t)-\frac{1}{2}\mu_2 h^2\ddot{\alpha_0}^{\mathsf{T}}(t)]\end{aligned} \tag{A.5}$$

By the mean value theorem,

$$(nh)^{1/2}[\tilde{\alpha}(t,\hat{\eta})-\tilde{\alpha}(t,\eta_0)] = (nh)^{1/2}\frac{\partial\tilde{\alpha}(t,\eta_m)}{\partial\eta}(\hat{\eta}-\eta_0) \tag{A.6}$$

where η_m is on the segment between η_0 and $\hat{\eta}$, and $\frac{\partial\tilde{\alpha}(t,\eta_m)}{\partial\eta}\xrightarrow{p}-e_{11}^{-1}(t)e_{12}(t)$. So combining equation (A.2) with equation (A.5)(A.6), we can continue writing the first two terms of

equation (A.5) as

$$
\begin{aligned}
(nh)^{1/2}[\tilde{\alpha}(t,\hat{\eta})-\tilde{\alpha}(t,\eta_0)] &= (nh)^{1/2}\frac{\partial\tilde{\alpha}(t,\eta_m)}{\partial\eta}(\hat{\eta}-\eta_0) \\
&= h^{1/2}e_{11}^{-1}(t)e_{12}(t)A_{\eta_0}^{-1}\frac{1}{\sqrt{n}}U_\eta(\eta_0)+\mathrm{op}(h^{1/2}) \\
&= h^{1/2}e_{11}^{-1}(t)e_{12}(t)A_{\eta_0}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{t_1}^{t_2}\frac{\dot{\tilde{\lambda}}_i(s,\eta_0)}{\tilde{\lambda}_i(s,\eta_0)}\Big\{\Big(\frac{\partial\zeta(U_i(s)),\eta_0}{\partial\eta}\Big)^{\mathsf{T}}P_i(s) \\
&\quad -(e_{12}(s))^{\mathsf{T}}(e_{11}(s))^{-1}X_i(s)\Big\}dM_i(s)+\mathrm{op}(h^{1/2}) \qquad \text{(A.7)}
\end{aligned}
$$

Then, we consider the last three terms of (A.5), for given t, we know $\tilde{\alpha}(t,\eta_0)$ is the first p_1 elements of $\tilde{\alpha}^*(t,\eta_0)$, which is solved from equation (2.6) when $\eta=\eta_0$.

From equation (2.6),

$$
U_{\alpha^*}(\alpha^*,\eta_0,t)=\sum_{i=1}^{n}\int_0^\tau\frac{\dot{\lambda}_i^*(s,\alpha^*,\eta_0|t)}{\lambda_i^*(s,\alpha^*,\eta_0|t)}\Big\{dN_i(s)-Y_i(s)\lambda_i^*(s,\alpha^*,\eta|t)ds\Big\}X_i^*(s|t)K_h(s-t)
$$

By Taylor expansion, we have

$$
U_{\alpha^*}(\tilde{\alpha}^*(t,\eta_0),\eta_0,t)=U_{\alpha^*}(\alpha_0^*(t),\eta_0,t)+\frac{\partial U_{\alpha^*}}{\partial\alpha^*}\Big|_{\alpha^*=\alpha_0^*(t)}\Big(\tilde{\alpha}^*(t,\eta_0)-\alpha_0^*(t)\Big)
$$

Since $U_{\alpha^*}(\tilde{\alpha}^*(t,\eta_0),\eta_0,t)=0$, we have

$$
\tilde{\alpha}^*(t,\eta_0)-\alpha_0^*(t)=-\left(\frac{\partial U_{\alpha^*}}{\partial\alpha^*}\Big|_{\alpha^*=\alpha_0^*(t)}\right)^{-1}U_{\alpha^*}(\alpha_0^*(t),\eta_0,t)
$$

Since $\tilde{\alpha}(t,\eta_0)-\alpha_0(t)$ is the first p_1 components of $\tilde{\alpha}^*(t,\eta_0)-\alpha_0^*(t)$, we have

$$
\begin{aligned}
\tilde{\alpha}(t,\eta_0)-\alpha_0(t) &= -(ne_{11}(t))^{-1}\sum_{i=1}^{n}\int_0^\tau\frac{\dot{\lambda}_i^*(s,\alpha_0^*(s),\eta_0|t)}{\lambda_i^*(s,\alpha_0^*(s),\eta_0|t)}\Big\{dN_i(s)-Y_i(s)\lambda_i^*(s,\alpha_0^*(s),\eta_0|t)ds\Big\} \\
&\quad X_i(s)K_h(s-t) \\
&= -(ne_{11}(t))^{-1}\sum_{i=1}^{n}\int_0^\tau\frac{\dot{\lambda}_i^*(s,\alpha_0^*(s),\eta_0|t)}{\lambda_i^*(s,\alpha_0^*(s),\eta_0|t)}\Big\{dN_i(s)-Y_i(s)\lambda_i(s)ds+Y_i(s)\lambda_i(s)ds \\
&\quad -Y_i(s)\lambda_i^*(s,\alpha_0^*(s),\eta_0|t)ds\Big\}X_i(s)K_h(s-t)
\end{aligned}
$$

Note that

$$\lambda_i(s) = g^{-1}\{\alpha_0(s)X_i(s) + \zeta(\eta_0, U_i(s))P_i(s)\}$$

and

$$\lambda_i^*(s, \alpha_0^*(s), \eta_0|t) = g^{-1}\{\alpha_0^*(t)X_i^*(s|t) + \zeta(\eta_0, U_i(s))\}$$

The last two terms can be written as

$$\begin{aligned}
&\lambda_i(s) - \lambda_i^*(s, \alpha_0^*(s), \eta_0|t) \\
=&\dot{g}^{-1}\{\alpha_0^{*\mathsf{T}}(t)X_i^*(s|t) + \zeta^{\mathsf{T}}(\eta_0, U_i(s))P_i(s)\}(\alpha_0(s)X_i(s) - \alpha_0^*(t)X_i^*(s|t)) \\
=&\dot{g}^{-1}\{\alpha_0^{*\mathsf{T}}(t)X_i^*(s|t) + \zeta^{\mathsf{T}}(\eta_0, U_i(s))P_i(s)\}[\frac{1}{2}\ddot{\alpha}_0(t)(s-t)^2X_i(s)] \\
=&\dot{\lambda}_i(s, \alpha_0^*(s), \eta_0|t)[\frac{1}{2}\ddot{\alpha}_0(t)(s-t)^2X_i(s)]
\end{aligned}$$

By the definition of $e_{11}(t)$,

$$-n^{-1}\sum_{i=1}^{n}\int_0^\tau K_h(s-t)\frac{\dot{\lambda}_i(s, \alpha_0^*(s), \eta_0|t)^2}{\lambda_i(s, \alpha_0^*(s), \eta_0|t)}X_i(s)^{\otimes 2}ds \xrightarrow{\mathcal{P}} e_{11}(t)$$

Let $dM_i(s) = dN_i(s) - Y_i(s)\lambda_i(s)ds$, we have

$$\begin{aligned}
&(nh)^{1/2}[\tilde{\alpha}(t, \eta_0) - \alpha_0(t) - \frac{1}{2}\mu_2 h^2\ddot{\alpha}_0(t)] \\
=&-n^{-1/2}h^{1/2}e_{11}(t)^{-1}\sum_{i=1}^{n}\int_0^\tau \frac{\dot{\lambda}^*(s, \alpha_0^*(s), \eta_0|t)}{\lambda^*(s, \alpha_0^*(s), \eta_0|t)}X_i(s)K_h(s-t)dM_i(s) \qquad \text{(A.8)}
\end{aligned}$$

where $\mu_2 = \int_{-1}^{1} t^2 K(t)dt$.

Combine equation (A.7) and (A.8),

$$
\begin{aligned}
&(nh)^{1/2}[\hat{\alpha}(t) - \alpha_0(t) - \frac{1}{2}\mu_2 h^2 \ddot{\alpha_0}^{\mathsf{T}}(t)] \\
=&(nh)^{1/2}[\tilde{\alpha}(t,\hat{\eta}) - \tilde{\alpha}(t,\eta_0) + \tilde{\alpha}(t,\eta_0) - \alpha_0(t) - \frac{1}{2}\mu_2 h^2 \ddot{\alpha_0}^{\mathsf{T}}(t)] \\
=&(n^{-1}h)^{1/2} e_{11}^{-1}(t) \Bigg[e_{12}(t) A_{\eta_0}^{-1} \sum_{i=1}^{n} \int_{t_1}^{t_2} \frac{\dot{\tilde{\lambda}}_i(s,\eta_0)}{\tilde{\lambda}_i(s,\eta_0)} \Big\{ \Big(\frac{\partial \zeta(U_i(s),\eta_0)}{\partial \eta} \Big)^{\mathsf{T}} P_i(s) \\
&- (e_{12}(s))^{\mathsf{T}} (e_{11}(s))^{-1} X_i(s) \Big\} dM_i(s) - \sum_{i=1}^{n} \int_0^{\tau} \frac{\dot{\lambda}^*(s,\alpha_0^*,\eta_0|t)}{\lambda^*(s,\alpha_0^*,\eta_0|t)} X_i(s) K_h(s-t) dM_i(s) \Bigg]
\end{aligned}
$$

By CLT for martingale, we have

$$(nh)^{1/2}[\hat{\alpha}(t) - \alpha_0(t) - \frac{1}{2}\mu_2 h^2 \ddot{\alpha_0}^{\mathsf{T}}(t)] \xrightarrow{D} N(0, \Sigma_\alpha)$$

$\Sigma_\alpha(t)$ can be estimated by$\hat{E}_{11}(t)^{-1}\hat{\Sigma}_e(t)\hat{E}_{11}(t)^{-1}$, with

$$
\begin{aligned}
&\hat{\Sigma}_e(t) = n^{-1}h \sum_{i=1}^{n} \Big[\int_0^{\tau} \frac{\dot{\hat{\lambda}}_i(s)}{\hat{\lambda}_i(s)} \{dN_i(s) - Y_i(s)\hat{\lambda}_i(s)\} X_i(s) K_h(s-t) \\
&-\hat{E}_{12}(t)\hat{A}_\eta^{-1} \int_{t_1}^{t_2} \frac{\dot{\hat{\lambda}}_i(s)}{\hat{\lambda}_i(s)} \{dN_i(s) - Y_i(s)\hat{\lambda}_i(s)\} \Big\{ \Big(\frac{\partial \zeta(U_i(s),\hat{\eta})}{\partial \eta} \Big)^{\mathsf{T}} P_i(s) - \hat{E}_{12}(s)^{\mathsf{T}} \hat{E}_{11}(s)^{-1} X_i(s) \Big\} \Big]^{\otimes 2}
\end{aligned}
$$