ESTIMATION AND INFERENCE OF 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 of 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 most existing work on recurrent events address both time-independent and time-dependent effects, our challenges in analyzing a real-world vaccine trial data emphasize the importance of considering scenarios where these effects vary with specific 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.

In the first project, we consider a generalized class of semiparametric intensity 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 are investigated through local linear approximation and profile log-likelihood method. A cross-validation bandwidth selection method is discussed. Asymptotic properties of estimators are explored using martingale theory and empirical processes. Two hypothesis tests based on the martingale residual have been developed to assess the parametric functions of the covariate-varying effects. A Gaussian multiplier method has been derived to approximate the underlying distribution of test statistics.

In the second project, we propose a nonparametric intensity model with frailty that captures unspecified time-varying and covariate-varying effects. Each individual is associated with a frailty term following a Gamma distribution, which acts mul-

tiplicatively on the intensity function. We develop maximum likelihood estimation procedure using local linear approximation method with double kernels. The maximization is achieved through an EM algorithm. Variance estimators are obtained using a weighted bootstrap procedure.

The simulation studies reveal the satisfactory performance of both models, which have subsequently been employed to analyze the MAL-094 malaria vaccine efficacy trial data. Our data applications demonstrate that these proposed models successfully address the questions raised by the MAL-094 malaria vaccine efficacy trial data.

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 throughout 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 of interest that can occur repeatedly over time. They are often observed in medical studies, such as hospital admissions, cancer recurrences, infections of Covid-19, malaria 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.

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. It is transmitted through the bite of infected female Anopheles mosquitoes. Malaria can cause flu-like symptoms, and sometimes even be life-threatening. 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, with each arm containing 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 infection statuses have been recorded.

The primary objectives of our work are to measure the effects of the RTS,S/AS01$_{\mathrm{E}}$ malaria vaccine. Our research questions focus on: (1) Whether and how the risks of malaria infections vary over time? (2) How do previous events, such as prior infections and/or vaccinations, correlate with subsequent infections? Additionally, how do the

vaccine effects evolve over time following the most recent infection or vaccination?

Being able to answer these questions with mathematical models is an enormous benefit to the malaria vaccine research and development. To most of our knowledge, we can not find existing models that can perfectly addressing these questions. There remains a need to explore new models. Subsequently, we undertake a literature review and introduce conditional intensity models for 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 *et al.* (1997) proposed a semiparametric procedures to model the mean or rate function for recurrent events. The models involve a baseline mean or rate function which can be arbitrary, multiplied by a parametrically specified function of covariates. Lin *et al.* (2000) justified the inference procedure through empirical process theory and constructed confidence bands for the mean functions. 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.

Rather than modeling the overall population behavior, conditional methods can model the pattern based on event history. Conditional methods provide a flexible framework by modeling the intensity function of counting process of events over time.

For a given time t, let $N_i(t)$ be the counting process, which registers the number of events of subject i that have experienced up to, and including, time t. Let $\mathcal{F}_{it}$ be the history of the events up to time t. Mathematically, it can be a σ-algebra generated by the counting process $N_i(t)$ and possible covariate processes. The intensity function of counting process $N_i(t)$ is defined as

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

where $\Delta N_i(t) = N_i(t + \Delta t^-) - N_i(t^-)$ is the number of events in the time interval $[t, t+\Delta t]$. By definition, we have $E(dN_i(t)|\mathcal{F}_{it^-}) = \lambda_i(t)dt$. The intensity of a counting process at time t is the instantaneous risk rate of an event occurrence at the time point, given the event and covariate history.

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) studied the proportional intensity model for recurrent events:

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

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 $G(\cdot)$ provides 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 assumed the individual specific frailty, the covariate effects and the random errors acted additively on the logarithm of gap time. Other works related with gap time include Oakes and Cui (1994), Pena *et al.* (2001), Strawderman (2005) and some others.

Much of the existing literature on recurrent event data present certain limitations for our analysis of malaria trial data. Their models either focus solely on constant effects or time-varying effects and can not address our specific question: how does the vaccine effect change over time since the most recent infection or vaccination?

Qi *et al.* (2017) studied a generalized class of semiparametric varying-coefficient models for longitudinal data, which can model time-independent effects, time-varying effects and covariate-varying effects. In the first project, we study a similar class of models for the intensity of recurrent events. These models can be used to study effects across two time scales: time-varying effects in calendar time and effects based on time since a treatment exposure or event exposure. They are valuable for analyzing the malaria vaccine trial data and understanding how infection risk depends on covariates

and prior infections.

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 distributed random variables. Gamma frailty is commonly used in frailty models by its conjugate properties. Nileson *et al.* (1992) introduced Gamma frailty, which act multiplicatively on the intensity function as follows:

$$\lambda_i(t) = Z_i Y_i(t)\exp\{\beta^{\mathsf{T}} X_i(t)\}\alpha(t), \tag{1.1}$$

where the frailty variable Z_i is drawn from a Gamma distribution, with unknown parameter. $Y_i(t)$ is an observable non-negative predictable process, $X_i(t)$ is a possibly time-dependent covariates, β is a vector of unknown regression parameters and $\alpha(t)$ is an unspecified function.

Klein (1992) specified the estimation procedures of Nelson's model 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 within a class of semiparametric transformation models:

$$\Lambda(t|X, Z; b) = G\Big(\int_0^t \lambda(s)e^{\beta^{\mathsf{T}} X(s)+b^{\mathsf{T}} Z(s)}ds\Big),$$

where $G(\cdot)$ is a transformation function, $X(s)$ and $Z(s)$ are possibly time-dependent covariates, $\lambda(.)$ is an arbitrary positive function, β is a set of unknown parameters and b is a set of random effects.

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 of intensity models with frailty for recurrent event data, includes Yu *et al.* (2013), Chen *et al.* (2013), Mazroui *et al.* (2015) among others.

In the second project, we study a nonparametric dynamic intensity model that incorporates frailty to account for unobserved heterogeneity. Our goal is to explore the vaccine effects and examine how these effects vary across two time scales, after taking the unobserved heterogeneity into consideration.

This dissertation presents the research of two projects. In Chapter 2, we investigate a generalized class of semiparametric dynamic intensity models. Estimation and hypothesis testing procedures are developed. Simulation studies are conducted to evaluate the validity of the proposed procedures. We derive the asymptotic properties of the estimators based on martingale theory and empirical processes. The methods are applied to analyzing the MAL-094 malaria vaccine trial data. In Chapter 3, we investigate a nonparametric dynamic intensity model with frailty. We provide estimation procedure and obtain variance estimators using weighted bootstraps. Simulation results demonstrate that the procedures perform well in finite samples. We also apply the methods to the MAL-094 malaria vaccine trial data. Chapter 4 discusses some concluding remarks and outlines future work.

CHAPTER 2: GENERALIZED SEMIPARAMETRIC INTENSITY MODELS FOR RECURRENT EVENT DATA

2.1 Introduction

In this chapter, we introduce a generalized class of semiparametric intensity models. 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.

Semiparametric models present multiple advantages. First, they necessitate less data for fitting compared to nonparametric models, making them particularly efficient in situations where data is limited. Additionally, if we have some prior knowledge about the parametric forms, it can help enhance our understanding of the data.

Section 2.2 details the models and the estimation procedures for unknown parameters, covering the computational algorithm and the selection of bandwidth. The asymptotic properties of the estimators are derived in Section 2.3. In order to evaluate whether the parametric functions are proper, two hypothesis tests are developed in Section 2.4. Simulation studies in Section 2.5 show that the methods perform well in finite samples under different link functions and different parametric functions.

We apply the methods on the MAL-094 malaria vaccine trial data. In Section 2.6.1, we model the intensity as a function of Calendar time and time since the most recent infection, to explore how the vaccine effects changes over these two time scales. In Section 2.6.2, we model the intensity as a function of Calendar time and time since the most recent vaccination, to explore how the vaccine effects change with the time since the most recent vaccination.

2.2 Model and Estimation

2.2.1 Model Descriptions

Consider a random sample of n subjects, τ is duration of study. 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 $N_i^*(t)$. It is defined as $\lambda_i(t) = \lim_{\Delta t \downarrow 0} \Pr(\Delta N_i^*(t) = 1|\mathcal{F}_{it^-}^*)/\Delta t$, where $\mathcal{F}_{it^-}^*$ is the filtration generated by $N_i^*(t)$ and the history of covariates for ith subject up to time t. By definition, we have $E(dN_i^*(t)|\mathcal{F}_{it^-}^*) = \lambda_i(t)dt$. Therefore, $\lambda_i(t)dt$ is the instantaneous probability of an event occurring in $[t, t+\Delta t)$.

Let C_i be the non-informative censoring time for subject i. Let $\tau_i = \min\{\tau, C_i\}$, 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. By definition, we have $E(dN_i(t)|\mathcal{F}_{it^-}) = Y_i(t)\lambda_i(t)dt$. Censoring 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 $0 \leq t \leq \tau$.

In model 2.1, $\alpha(\cdot)$ represents a p_1 dimensional vector, with each element denoting

an unspecified function. β is a p_2 dimensional vector comprising unknown time-independent parameters. Let $\mathcal{U} \in R^r$ be the range of r dimensional process $U_i(t)$, $\gamma(u, \theta)$ is a p_3 dimensional vector of parametric functions defined on $\mathcal{U}$ for $\theta \in \Theta$, where Θ is a q dimension compact set.

The known function $g(\cdot)$ offers a lot of modeling flexibility. 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, the time since last vaccination can be written as $U_i(t) = t - V_i(t)$, where $V_i(t)$ be the most recent vaccination time. In another example, $U_i(t) = t - T_{iN_i(t^-)}$ stands for the time since the most recent event.

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 $P_i(t) = (Z_i(t)^\mathsf{T}, W_i(t)^\mathsf{T})^\mathsf{T}$. The intensity function 2.1 can be written as

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

for $0 \leq t \leq \tau$. The parameters to be estimated are $\alpha(\cdot)$ and η.

2.2.2 Estimation Procedure

We consider profile approach and use local linear approximation in the estimation. Assume $\alpha(\cdot)$ is smooth enough on $t \in [0, \tau]$ and its first and second derivatives $\dot{\alpha}(t)$ and $\ddot{\alpha}(t)$ exists. Denote $\mathcal{N}_{t_0}$ as a neighbourhood of t_0. For $t \in \mathcal{N}_{t_0}$, we have

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

The approximated intensity function for t in the neighborhood of t_0 is:

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

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

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

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

For fixed η, apply local linear method (Cai and Sun (2003)) and plug in the approximated intensity function 2.3, the localized log-likelihood for $\alpha(\cdot)$ at each t_0 is:

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

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:

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

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 $\tilde{\alpha}^*(t,\eta)$. Let $\tilde{\lambda}_i(t,\eta)$ 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:

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

where $[t_1, t_2] \subset (0,\tau)$. Here we integrate over the interval $[t_1, t_2]$ to avoid boundary effects.

Taking derivative of 2.7 with respect to η, the profile maximum likelihood estimate $\hat{\eta}$ is obtained by solving the following estimating equation:

$$\begin{aligned} U_\eta(\eta) = \sum_{i=1}^{n}\int_{t_1}^{t_2}\Big\{\Big(\frac{\partial\tilde{\alpha}(t,\eta)}{\partial\eta}\Big)^{\mathsf{T}}X_i(t) + \Big(\frac{\partial\zeta(U_i(t),\eta)}{\partial\eta}\Big)^{\mathsf{T}}P_i(t)\Big\} \\ \times\Big\{\frac{\dot{\tilde{\lambda}}_i(t,\eta)}{\tilde{\lambda}_i(t,\eta)}dN_i(t) - Y_i(t)\dot{\tilde{\lambda}}_i(t,\eta)dt\Big\} = \mathbf{0}. \end{aligned} \tag{2.8}$$

In the estimation equation 2.8, $\frac{\partial\tilde{\alpha}(t.\eta)}{\partial\eta}$ 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 $U_\alpha(\tilde{\alpha}^*(t,\eta);\eta,t) = \mathbf{0}$ with respect to η on both sides,

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

The estimate $\hat{\eta}$ can be updated by solving the profile estimating equation 2.8 using Newton-Raphson method. Subsequently, $\hat{\eta}$ is plugged into $\tilde{\alpha}(t,\eta)$ to obtain $\hat{\alpha}(t) = \tilde{\alpha}(t,\hat{\eta})$. Estimators $\hat{\alpha}(t)$ and $\hat{\eta}$ can be obtained by iteratively updating the estimates of $\tilde{\alpha}^*(t,\eta)$ and η until convergence is achieved for both, thereby maximizing the likelihood.

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 Bandwidth Selection

The proposed estimation procedure integrates a local linear approximation approach, which involves the selection of kernel function and bandwidth. Choice of kernel function has little impact on the model performance (Silverman (1986)). We use 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)). However, the selection of bandwidth can influence the estimation results. It is important to select an appropriate bandwidth.

We employ Monte Carlo cross-validation (also referred to as "leave-group-out" cross-validation) method using bootstraps to select bandwidth (Cai *et al.* (2023)).

This method helps reduce the randomness associated with data splitting during cross-validation.

Create a set of candidate bandwidths H. In jth bootstrap iteration, we do the following procedures:

1. Randomly sample from the original dataset without replacement with a fixed proportion to obtain a training dataset D_n^j. The subjects not selected into D_n^j form the test dataset D_t^j.

2. For each h in H, using h to fit the model on training dataset D_n^j to obtain estimates $\hat{\alpha}^{(j,h)}(t)$, $\hat{\beta}^{(j,h)}$, $\hat{\theta}^{(j,h)}$. The estimated intensity function for subject i takes as

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

3. The prediction accuracy for jth bootstrap using bandwidth h, denoted as $ACC^{(j)}(h)$, is defined as

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

   where $[t_1, t_2] \subset (0, \tau)$. The prediction accuracy is the log-likelihood on test dataset, a similar criterion was proposed by Tian *et al.* (2005) for survival data.

4. The recorded bandwidth in jth iteration h_j^* is the one that maximizes the prediction accuracy, i.e. $h_j^* = \underset{h}{\text{argmax}} ACC^{(j)}(h)$.

Repeat Step 1 to Step 4 B times, the optimal bandwidth, denoted as h_{opt}^*, is determined by the average of all the h_j^*.

2.3 Asymptotic Properties

In this section, we discuss the asymptotic properties of the estimators. We firstly introduce some notations that would be used in the theorems.

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 $\lambda_i(t) = g^{-1}\{\alpha_0^{\mathsf{T}}(t)X_i(t)+\zeta^{\mathsf{T}}(U_i(t),\eta_0)P_i(t)\}$ and $\dot{\lambda}_i(t) = \dot{g}^{-1}\{\alpha_0^{\mathsf{T}}(t)X_i(t)+\zeta^{\mathsf{T}}(U_i(t),\eta_0,)P_i(t)\}$. Define $e_{11}(t) = E\{-Y_i(t)\frac{\dot{\lambda}_i^2(t)}{\lambda_i(t)}[X_i(t)]^{\otimes 2}\}$ and $e_{12}(t) = E\{-Y_i(t)\frac{\dot{\lambda}_i^2(t)}{\lambda_i(t)}X_i(t)P_i^{\mathsf{T}}(t)(\frac{\partial\zeta(U_i(t),\eta_0)}{\partial\eta})\}$.

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

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,* $\eta \xrightarrow{\mathcal{P}} \eta_0$*, 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 vectors, for a vector* a*,* $a^{\otimes 2} = aa^{\mathsf{T}}$*.*

A_η can be estimated by

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

and Σ_η can be estimated by

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

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 $\mu_2=\int_{-1}^{1}t^2K(t)dt$ *and* $\Sigma_\alpha(t)=e_{11}(t)\Sigma_e(t)(e_{11}(t))^{-1}$*, with*

$$\Sigma_e(t) = \lim_{n\to\infty} hE\Big\{\int_0^\tau K_h^2(s-t)\frac{\dot{\lambda}^2(s)}{\lambda(s)}[X_i(s)]^{\otimes 2}ds\Big\}.$$

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

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

2.4 Testing the Covariate-Varying Effects

The parametric functions of the covariate-varying effects can be based on some prior knowledge or understanding behind the process. Otherwise, we can start from polynomials or linear combinations of basis functions. In this section, we provide two hypothesis test procedures to test the adequacy of parametric form $\gamma(U_i(t), \theta)$.

To test $H_0 : \gamma(u) = \gamma(u, \theta)$, $\theta \in \Theta$, 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

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

In 2.10, $u \in R^r$ 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 weighted martingale residual stratified by $U_i(t)$.

Defined supremum type test statistics $T_1 = \sup_{u \in \triangle} \|R(u, \hat{\eta})\|$, where $\|\cdot\|$ represent the L_2 norm in R^r and $\triangle$ is a set of grid points in $\mathcal{U}$.

Let $\{u_1, ...u_K\}$ be the grid points for $U_i(t)$. Let

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

Define quadratic type test statistic

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

where $\widehat{\mathrm{Cov}}(L(\hat{\eta}), L(\hat{\eta}))$ is the estimated covariance matrix of $\mathrm{Cov}(L(\hat{\eta}), L(\hat{\eta}))$. $\widehat{\mathrm{Cov}}(L(\hat{\eta}), L(\hat{\eta}))$ has $(K-1)\times(K-1)$ blocks, for $1 \le q, s \le K-1$, the (q, s) th block equals

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

The estimation of $\mathrm{cov}[R(u_l, \hat{\eta}), R(u_m, \hat{\eta})]$, $1 \le l, m \le K$ will be given later in 2.17.

T_2 has a chi-square distribution, but 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 as follows.

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

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

It can be shown that

$$n^{-\frac{1}{2}} \frac{\partial R(u, \eta_0)}{\partial \eta} \xrightarrow{p} -(I_r \otimes A_\eta^{-1}) A_u, \quad (2.15)$$

where

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

Combining equation 2.12,2.14 and 2.15, we have

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

where

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

It follows by the theorems of empirical process, $R(u, \hat{\eta})$ converges weakly to a mean zero Gaussian process $R(u)$, for $u \in \mathcal{U}$.

Denote

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

Let ϕ_1, ... ,ϕ_n be n independent standard normal random variables. Define the Gaussian 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 $R^*(u)$(Lin *et al.* (1993)). The critical values of T_1 and T_2 can be obtained using Gaussian multiplier approach. Hold the observed data sequence fixed and generate, say 500, sets of $\{\phi_1, \ ... \ ,\phi_n\}$ to get 500 realizations of $R^*(u)$. The critical value of T_1 can be determined by the percentile of the empirical distribution of $T_1^* = \sup_{u\in\triangle} \|R^*(u)\|$.

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 estimated by

$$\begin{aligned}
\widehat{\mathrm{Cov}}(R(u_l, \hat{\eta}), R(u_m, \hat{\eta})) =& \mathrm{Cov}(R^*(u_l), R^*(u_m)|\mathcal{D}) \\
=& \mathrm{Cov}\Big(n^{-1/2} \sum_{i=1}^{n} \hat{D}_i(u_l)\phi_i, n^{-1/2} \sum_{j=1}^{n} \hat{D}_j(u_m)\phi_j \Big| \mathcal{D} \Big) \\
=& \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \hat{D}_i(u_l) \hat{D}_j^{\mathsf{T}}(u_m) \mathrm{Cov}(\phi_i, \phi_j) \\
=& \frac{1}{n} \sum_{i=1}^{n} \hat{D}_i(u_l) \hat{D}_i^{\mathsf{T}}(u_m) \mathrm{Cov}(\phi_i, \phi_i) \\
=& \frac{1}{n} \sum_{i=1}^{n} \hat{D}_i(u_l) \hat{D}_i^{\mathsf{T}}(u_m).
\end{aligned} \tag{2.17}$$

Define

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

and

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

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

Test statistic T_2 has an asymptotic chi-square distribution with degrees of freedom $(K-1) \times \dim(\eta)$, where $\dim(\eta)$ represents the dimension of η. The critical value of T_2 can also be determined from the chi-square distribution.

2.5 Simulation Studies

In this section, we conduct simulations to evaluate the performance of the estimators in finite samples. Simulation are conducted under two distinct link functions: the logarithm link function and the identity link function. Under the logarithm link function, we further examine the performance of estimators in 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. In all simulations, we use Epanechnikov kernel function $K(x) = 3/4(1-x^2)I\{|x| \leq 1\}$.

2.5.1 Simulation Studies with Logarithm Link Function

Using the logarithm link function $g(x) = \log(x)$ results in a multiplicative intensity function. Consider the models with following intensity function

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

for $0 \leq t \leq \tau$ in three different scenarios.

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).
- $U_i(t) = t - T_{iN_i(t^-)}$, $W_i(t) = I(N_i(t^-) > 0)$ 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. The integral interval $[t_1, t_2]$ in equation 2.8 is taken as $[0.5h, \tau - 0.5h]$, where h is the bandwidth. Table 2.1 presents the estimation results for sample sizes of $n = 400, 600, 800$, with different bandwidth $h = 0.2, 0.3, 0.4$. Figure 2.1 shows the estimation results for $\alpha_0(t)$ and $\alpha_1(t)$. 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 $M_{22} : \gamma(u) = 1.5 - u + 0.4u^2$.

Table 2.1: Bias, SEE, ESE and CP for estimators of β, θ_0, θ_1 under model 2.18 in Scenario 1 for sample size $n = 400$, 600 and 800, using bandwidth $h = 0.2$, 0.3 and 0.4. The results are based on 500 simulations.

	$\beta = 0.1$				$\theta_0 = 0.5$				$\theta_1 = 0.5$			
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

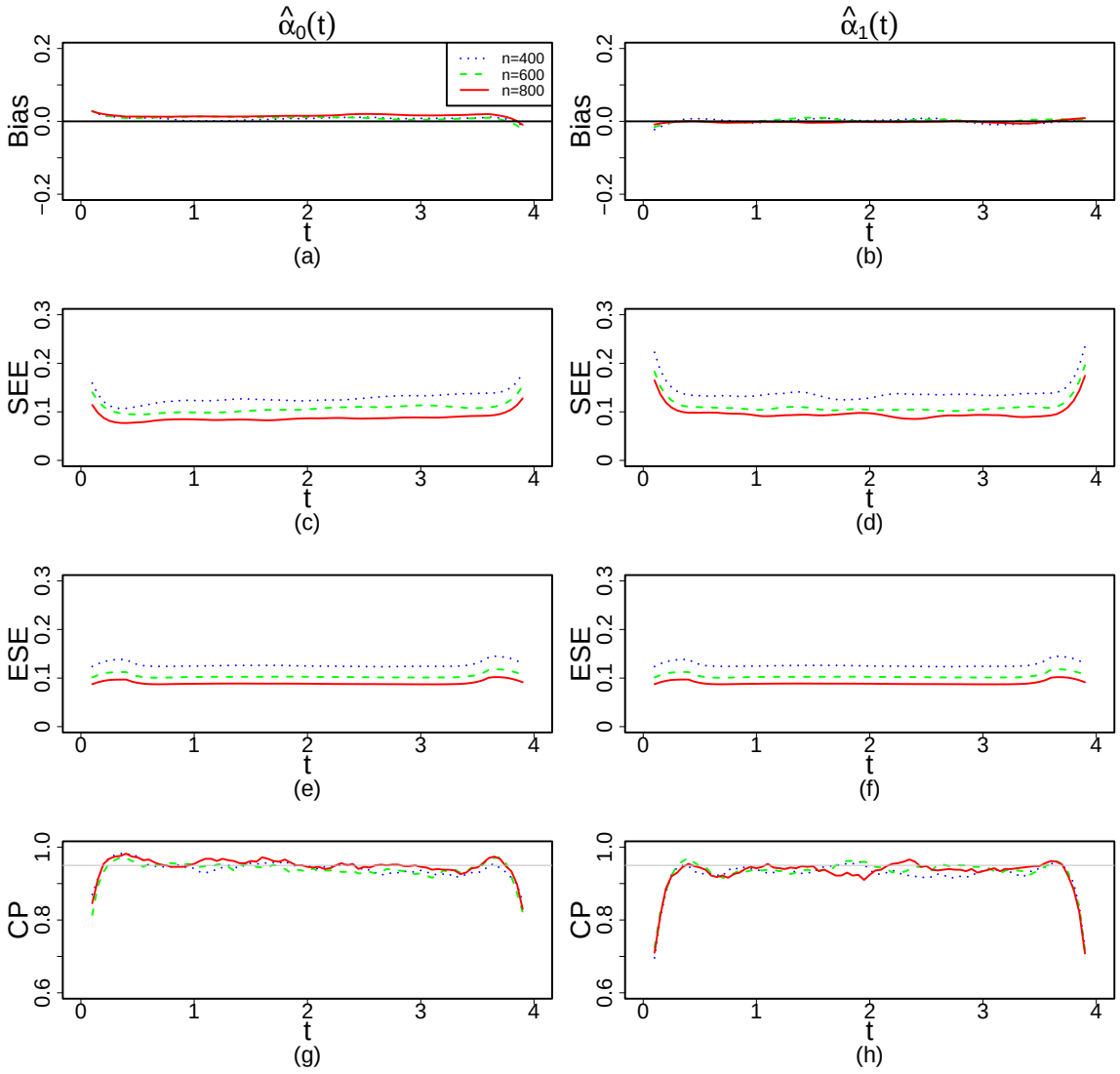

Figure 2.1: Bias, SEE, ESE and CP for estimators of $\alpha_0(t)$ and $\alpha_1(t)$ under model 2.18 in Scenario 1 using bandwidth $h = 0.4$. The dotted, dashed and solid lines represent sample size $n = 400$, 600 and 800, respectively. The results are based on 500 simulations.

Table 2.2: Observed sizes and powers of test T_1, T_{2g} and T_{2c} under model 2.18 in Scenario 1 for sample size $n = 400$, 600 and 800, using bandwidth $h = 0.3$ and 0.4. T_1 is the supremum type test. T_{2g} and T_{2c} are chi-square type tests. T_{2g} is the test based on test statistic T_2 and the critical value is based on Gaussian multiplier distribution. T_{2c} is the test based on test statistic T_2 and the critical value is based on chi-square distribution. Each entry is based on 500 simulations with 500 Gaussian multiplier samples.

		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

Scenario 2. $\gamma(U_i(t), \theta)$ is multidimensional, we consider model 2.18 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).
- $U_i(t) = t - T_{iN_i(t^-)}$; $W_i(t) = (W_{1i}(t), W_{2i}(t))^{\mathsf{T}}$ with $W_{1i}(t) = I(N_i(t^-) > 0)$ and $W_{2i}(t) = I(N_i(t^-) > 0)B_i$, where $B_i \sim Ber(0.5)$.
- $\gamma(U_i(t), \theta) = (\gamma_1(U_i(t), \theta_1), \gamma_2(U_i(t), \theta_2))^{\mathsf{T}}$; $\gamma_1(U_i(t), \theta_1) = \theta_{10} + \theta_{11}U_i(t)$ 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. The integral interval $[t_1, t_2]$ in equation 2.8 is taken as $[0.5h, \tau - 0.5h]$, where h is the bandwidth.

Table 2.3 presents the estimation results for sample sizes of $n = 400, 600, 800$ with bandwidth $h = 0.4$ and 0.5. Figure 2.2 shows the estimation results for $\alpha_0(t)$ and $\alpha_1(t)$. Table 2.4 presents the sizes under null models M_0: $\gamma_1(u, \theta_1) = \theta_{10} + \theta_{11}u$ and $\gamma_2(u, \theta_2) = \theta_{20} + \theta_{21}u$, with $\theta_{10} = \theta_{11} = 0.3$ and $\theta_{20} = \theta_{21} = 0.2$.

Table 2.3: Bias, SEE, ESE and CP for estimators of β, θ_{10}, θ_{11}, θ_{20} and θ_{21} under model 2.18 in Scenario 2 for sample size $n = 400$, 600 and 800, using bandwidth $h = 0.4$ and 0.5. The results are based on 500 simulations.

	$n = 400$				$n = 600$				$n = 800$			
	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

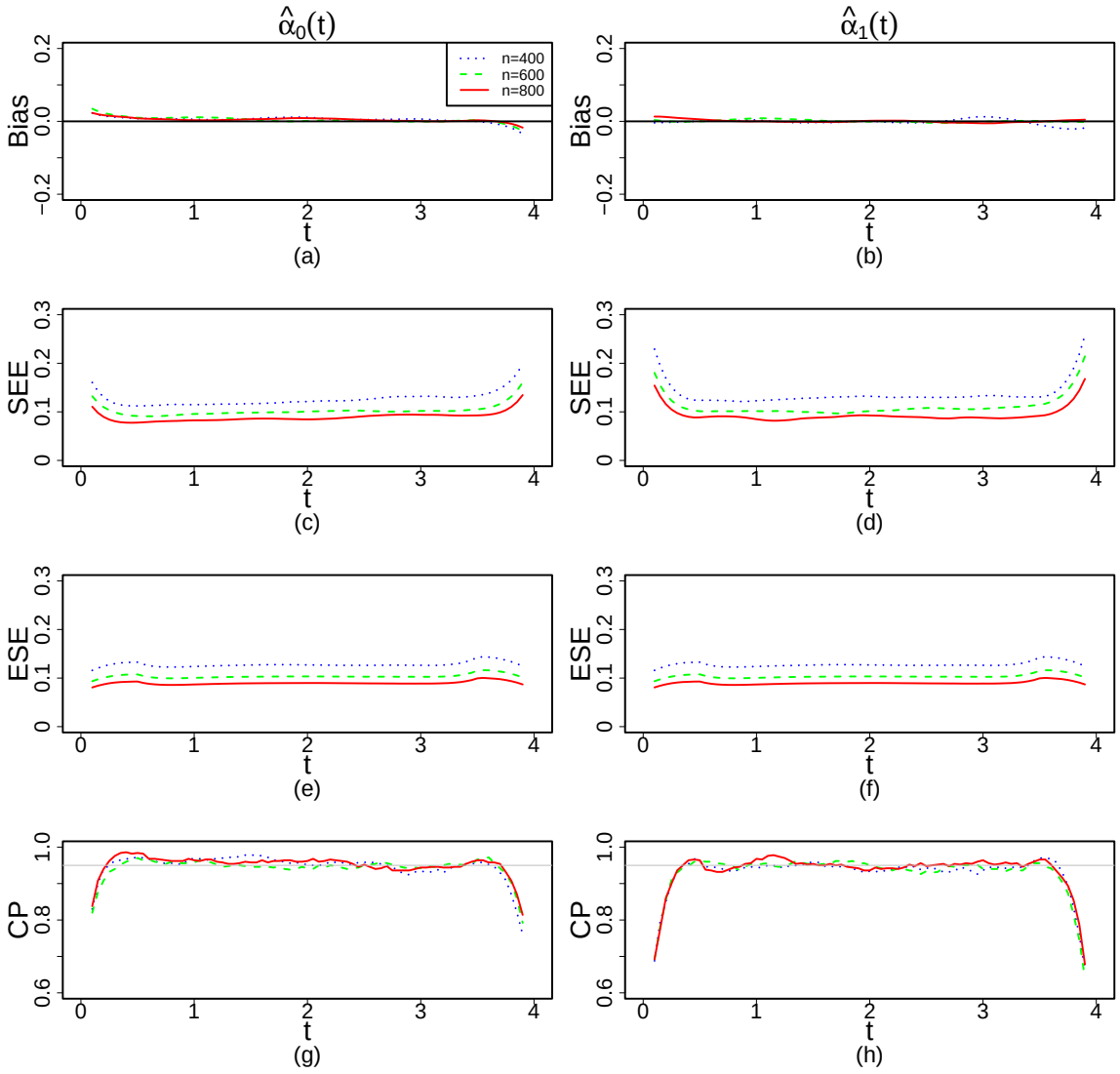

Figure 2.2: Bias, SEE, ESE and CP for estimators of $\alpha_0(t)$ and $\alpha_1(t)$ under model 2.18 in Scenario 2 using bandwidth $h = 0.5$. The dotted, dashed and solid lines represent sample size $n = 400$, 600 and 800, respectively. The results are based on 500 simulations.

Table 2.4: Observed sizes of test T_1, T_{2g} and T_{2c} under model 2.18 in Scenario 2 for sample size $n = 400$, 600 and 800, using bandwidth h=0.4 and 0.5. T_1 is the supremum type test. T_{2g} and T_{2c} are chi-square type tests. T_{2g} is the test based on test statistic T_2 and the critical value is based on Gaussian multiplier distribution. T_{2c} is the test based on test statistic T_2 and the critical value is based on chi-square distribution. Each entry is based on 500 simulations with 500 Gaussian multiplier samples.

		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

Scenario 3. Covariate $U_i(t)$ has multiple dimensions and $\gamma(U_i(t), \theta)$ is a nonlinear function of $U_i(t)$, we consider model 2.18 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).
- $U_i(t) = (U_{1i}(t), U_{2i}(t))$, $U_{1i}(t) = t - T_{iN_i(t^-)}$, $U_{2i}(t) = \log(N_i(t^-) + 1)$. $W_i(t) = I(N_i(t^-) > 0)$ 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. The integral interval $[t_1, t_2]$ in equation 2.8 is taken as $[h, \tau - h]$, where h is the bandwidth. 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 for estimators of β, θ_0, θ_1 and θ_2 under model 2.18 in Scenario 3 for sample size $n = 400$, $n = 600$ and $n = 800$, using bandwidth $h = 0.3, 0.4$ and 0.5. The results are based on 500 simulations.

	$\beta = 0.1$				$\theta_0 = 0.8$				$\theta_1 = 0.5$				$\theta_2 = 0.5$			
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

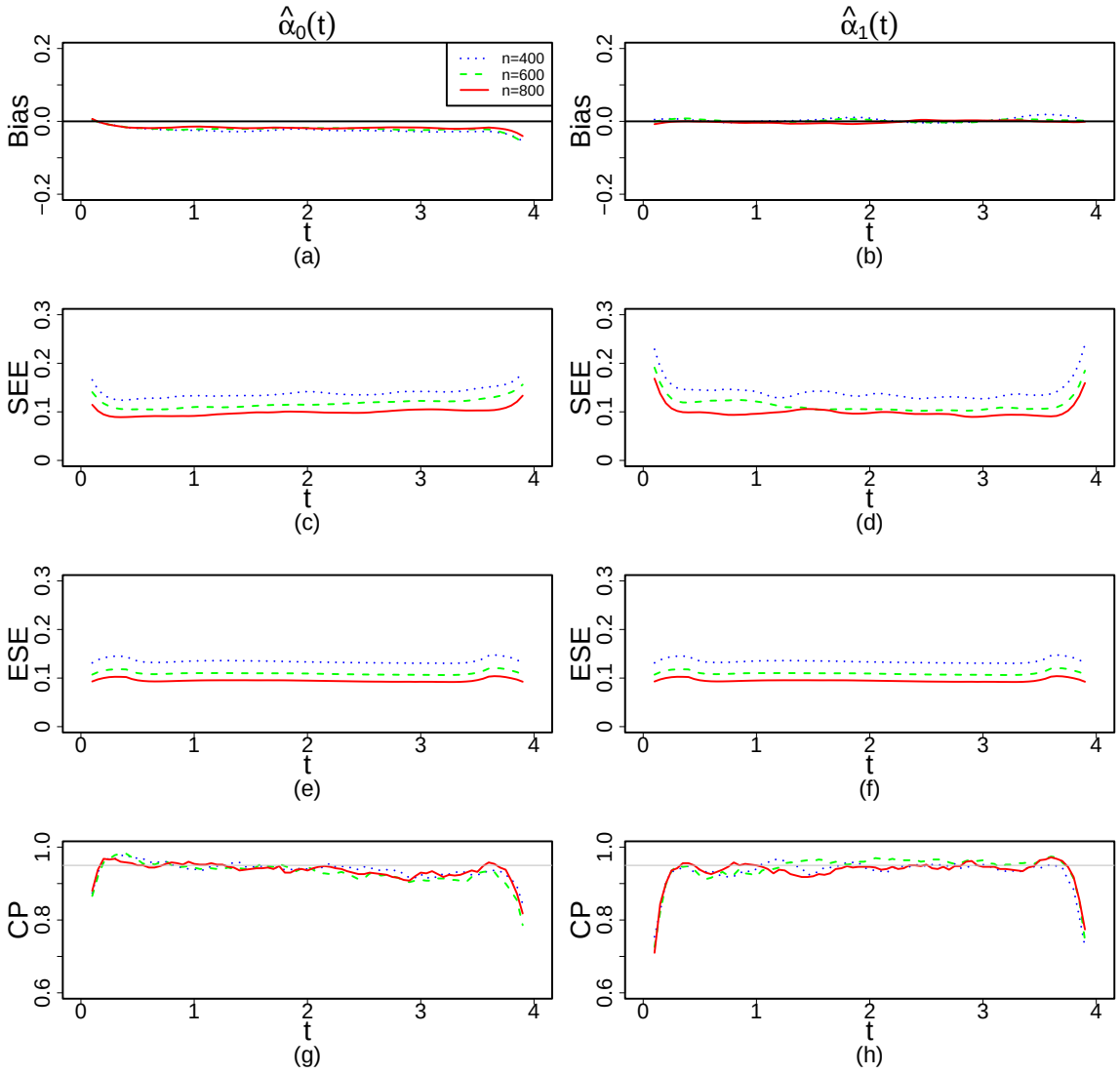

Figure 2.3: Bias, SEE, ESE and CP for estimators of $\alpha_0(t)$ and $\alpha_1(t)$ under model 2.18 in Scenario 3 using bandwidth $h = 0.4$. The dotted, dashed and solid lines represent sample size $n = 400$, 600 and 800, respectively. The results are based on 500 simulations.

2.5.2 Simulation Studies with Identity Link Function

In this subsection, we conduct a simulation study using the identity link function $g(x) = x$, which leads to an additive intensity model. We generated data from the following model:

$$\lambda_i(t) = \alpha_0(t) + \alpha_1(t)X_i + \beta Z_i + \gamma(U_i(t), \theta)W_i(t) \tag{2.19}$$

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).
- $U_i(t) = t - T_{iN_i(t^-)}$, $W_i(t) = I(N_i(t^-) > 0)$ 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. The integral interval $[t_1, t_2]$ in equation 2.8 is taken as $[0.5h, \tau - 0.5h]$, where h is the bandwidth. Table 2.6 presents the estimation results for β, θ_0 and θ_1 for sample sizes of $n = 400, 600, 800$ with bandwidth $h = 0.7, 0.8, 0.9$. Figure 2.4 presents the estimation results for $\alpha_0(t)$ and $\alpha_1(t)$. 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 $M_{22} : \gamma(u) = 1.0 - u + 0.6u^2$.

Table 2.6: Bias, SEE, ESE and CP for estimators of β, θ_0, θ_1 under model 2.19 for sample size $n = 400$, 600 and 800, using bandwidth $h = 0.7$, 0.8 and 0.9. The results are based on 500 simulations.

	$\beta = 0.1$				$\theta_0 = 0.2$				$\theta_1 = 0.2$			
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

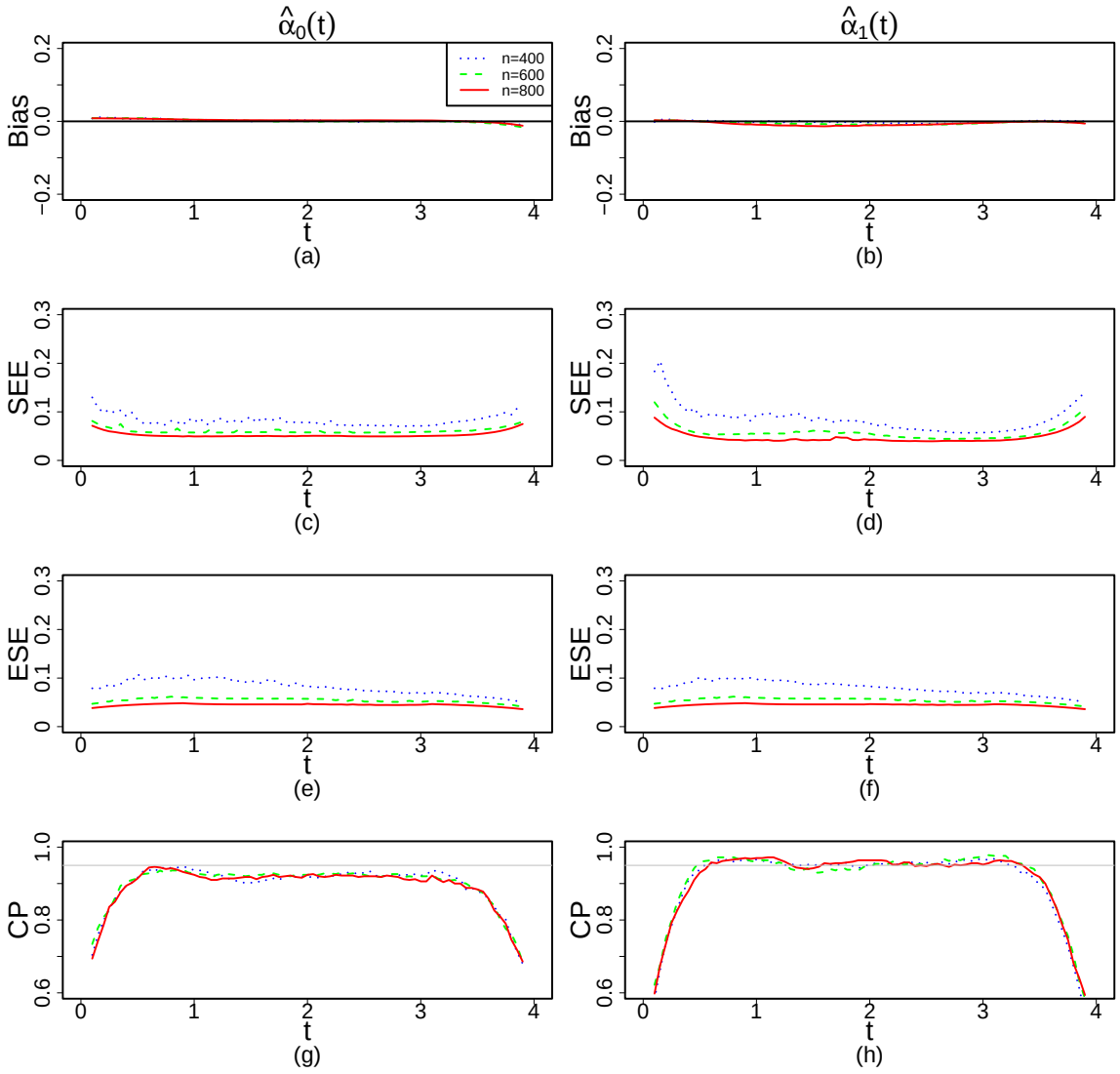

Figure 2.4: Bias, SEE, ESE and CP for estimators of $\alpha_0(t)$ and $\alpha_1(t)$ under model 2.19 using bandwidth $h = 0.9$. The dotted, dashed and solid lines represent sample size $n = 400$, 600 and 800, respectively. The results are based on 500 simulations.

Table 2.7: Observed sizes and powers of test T_1, T_{2g} and T_{2c} under model 2.19 for sample size $n = 400$, 600 and 800, using bandwidth $h = 0.8$ and 0.9. T_1 is supremum type test. T_{2g} and T_{2c} are chi-square type tests. T_{2g} is the test based on test statistic T_2 and the critical value is based on Gaussian multiplier distribution. T_{2c} is the test based on test statistic T_2 and the critical value is based on chi-square distribution. Each entry is based on 500 simulations with 500 Gaussian multiplier samples.

		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

The simulation results in this section demonstrate that our estimation and testing procedures perform well under both logarithmic and identity link functions. We observe that all biases are close to 0, and the coverage percentages are nearly aligned with the nominal value of 95%. Referring to Table 2.2 and Table 2.7, it's evident that the sizes of all three tests T_1, T_{2g} and T_{2c} are closely aligned with the significance level of 0.05. As anticipated, the powers of the tests increase with sample size. Overall, these results confirm the good performance of the tests we developed.

2.6 Data Application

In this section, we apply the proposed models to analyze the MAL-094 malaria vaccine trial. 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 occurring after three consecutive missed scheduled visits with no intervening unscheduled visits in-between are defined as censored. We analyze 20 months follow-up data from the MAL-094 trial. In the data analysis, the four RTS,S/AS01$_{\mathrm{E}}$ vaccine arms are combined as the treatment group, while the placebo arm is regarded as the control group.

Before censoring, a total of 3325 molecularly detected malaria infections (referred to as "infections" hereafter) are observed among 1464 participants. Out of the participants, 975 have experienced at least one infection, with the highest number of infections recorded being 25.

The available covariates in the data includes site, hemoglogbin and age at enrollment. There are 740 participants in Agogo and 724 participants in Siaya. Figure 2.5 displays histograms of the hemoglobin and age.

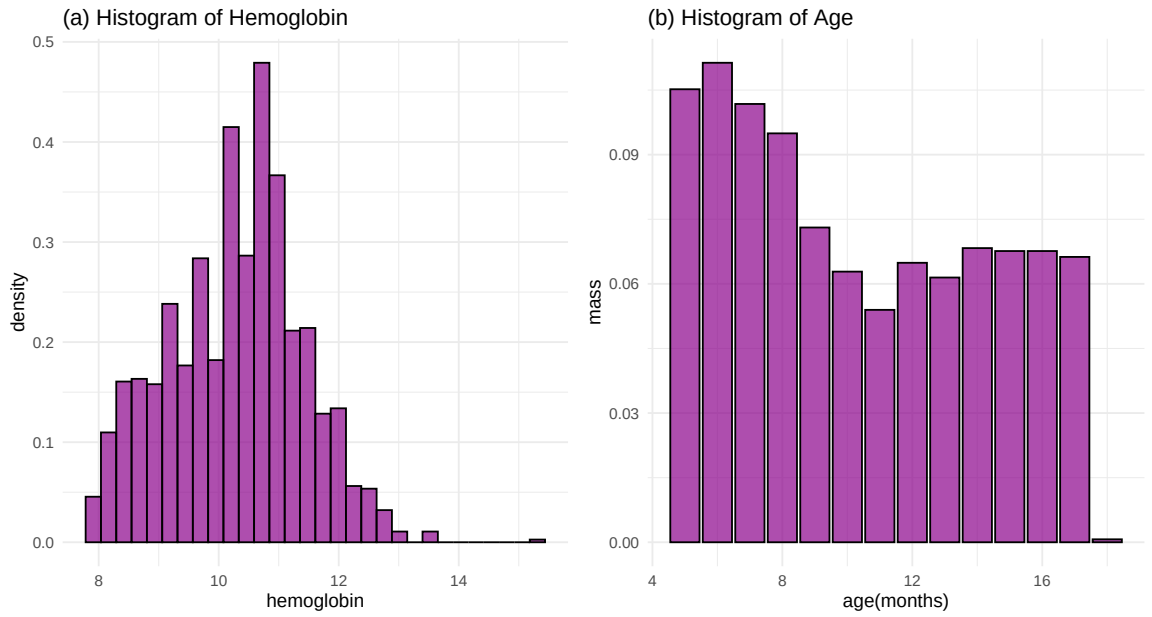

Figure 2.5: Histograms of hemoglobin and age for all participants.

Define T_{ij} as the jth infection time we observed for subject i. Denote n_i as the total event number 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 $N_i(t)$.

To investigate how the risk of malaria infection varies over time and understand the factors influencing this risk, we model the intensity function of malaria infections using covariates such as hemoglobin levels, age, and site. To explore the vaccine effects, we also include the treatment group indicator as a covariate in the model to determine the specific impact of the vaccines on malaria infection risk.

To explore the influence of previous infections and vaccinations on subsequent infections, we derive a covariate for each individual representing the time since their most recent infection or vaccination. This allows us to examine how the effects of malaria vaccines change over these time scales. In Section 2.6.1, we model the in-

tensity of malaria infections using two time scales: calendar time and time since the most recent infection. In Section 2.6.2, we model the intensity of malaria infections using two time scales: calendar time and time since the most recent vaccination. In both models, we initially consider linear forms for the covariate-varying effects due to the simplicity. Tests are conducted to confirm the acceptability of these linear forms at the end of each subsection.

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

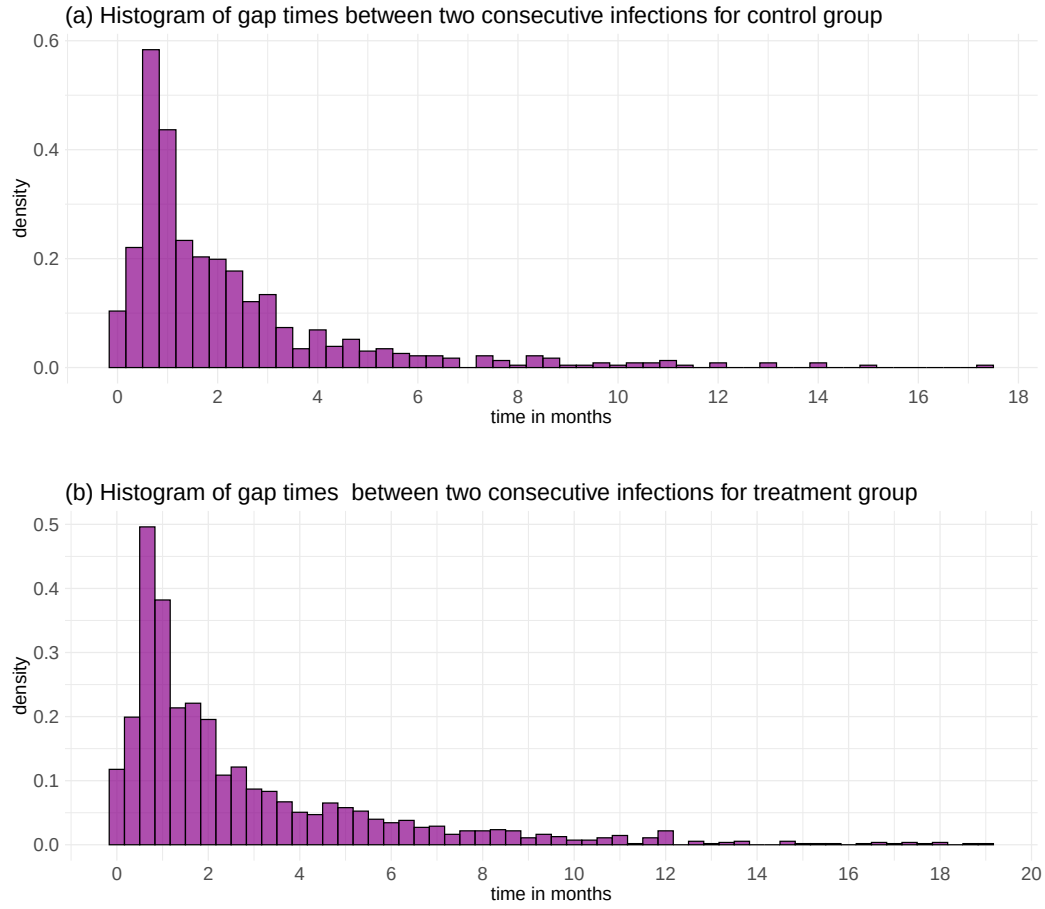

Figure 2.6: Histograms of gap times between consecutive infections for control and treatment group in the 20 months follow-up data.

Figure 2.6 is the histogram of gap times between consecutive infections for control group and treatment group.

To assess the effects of the RTS,S/AS01$_{\mathrm{E}}$ vaccine in two time scales, as well as get the effects of other risk factors, we consider the following multiplicative dynamic intensity model

$$\lambda_i(t) = \exp\Big\{\alpha_0(t) + \alpha_1(t)\mathrm{Vacc}_i + \alpha_2(t)\mathrm{Agogo}_i + \alpha_3(t)\mathrm{Age}_i(year) + \beta\mathrm{Hemo}_i$$
$$+\gamma_0(t - T_{iN_i(t^-)}, \theta)I(N_i(t^-) > 0) + \gamma_1(t - T_{iN_i(t^-)}, \theta)I(N_i(t^-) > 0)Vacc_i\Big\}, \quad (2.20)$$

for $0 \leq t \leq 20$.

In model 2.20, $\alpha_0(t)$, $\alpha_1(t)$, $\alpha_2(t)$ and $\alpha_3(t)$ are time-varying unspecified functions. Assume $\gamma_0(u, \theta) = \theta_0 + \theta_1 u$ and $\gamma_1(u, \tilde{\theta}) = \theta_2 + \theta_3 u$, with unknown parameters θ_0, θ_1, θ_2 and θ_3. 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). 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.

We use Epanechnikov kernel function $K(x) = 3/4(1 - x^2)I\{|x| \leq 1\}$ in the estimation and apply the Monte Carlo cross-validation method as described in Section 2.2.4 to select the optimal bandwidth. 100 repetitions of bootstrap cross-validation yield an optimal bandwidth of 1.35 months. Figure 2.7 displays the averaged accuracy per individual on the test dataset, with proportion $\frac{K-1}{K}$ of individuals are sampled as training dataset when $K = 3$ and $K = 5$.

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 $H_{01} : \beta = 0$, $H_{02} : \theta_0 = 0$, $H_{03} : \theta_1 = 0$, $H_{04} : \theta_2 = 0$ and $H_{05} : \theta_3 = 0$. At 0.05 significance level, we can see that $\hat{\beta}$ and $\hat{\theta}_0$ are significant and $\hat{\theta}_3$ is at the boundary. The hemoglobin level is negatively associated with the risk of malaria infections, meaning that a higher level of hemoglobin corresponds to a lower infection risk. Following prior infections, the risk of subsequent infection increases for the

control group ($\hat{\gamma_0}(u,\theta) > 0$), and this increment in risk appears to remain relatively consistent over time. However, for participants in the treatment group, the risk becomes lower than those in the control group after about 3 months since last infection ($\hat{\gamma_1}(u,\theta) < 0$).

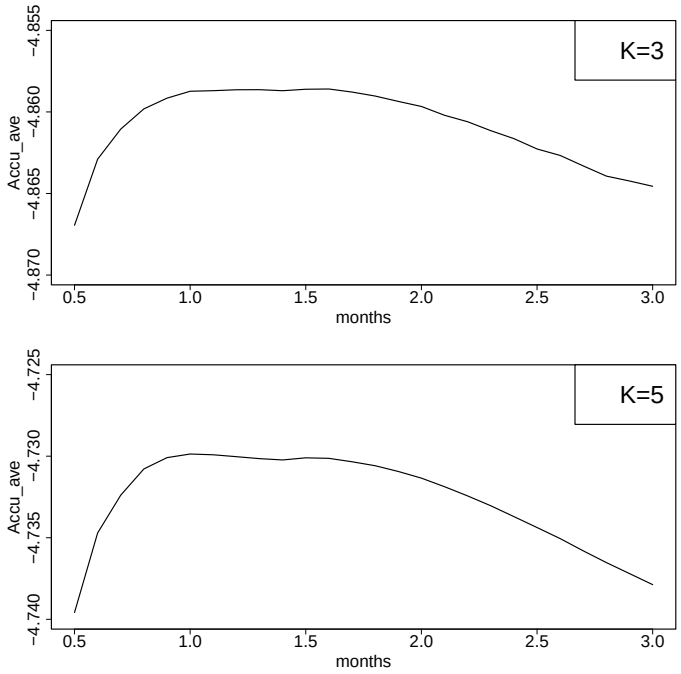

Figure 2.7: Average accuracy v.s. bandwidths when K=3 and K=5 under model 2.20.

Table 2.8: Estimates of parameters, their standard errors and p values under model 2.20, using bandwidth $h = 1.35$.

	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	$\ll 0.001$	$\ll 0.001$	0.059	0.329	0.051

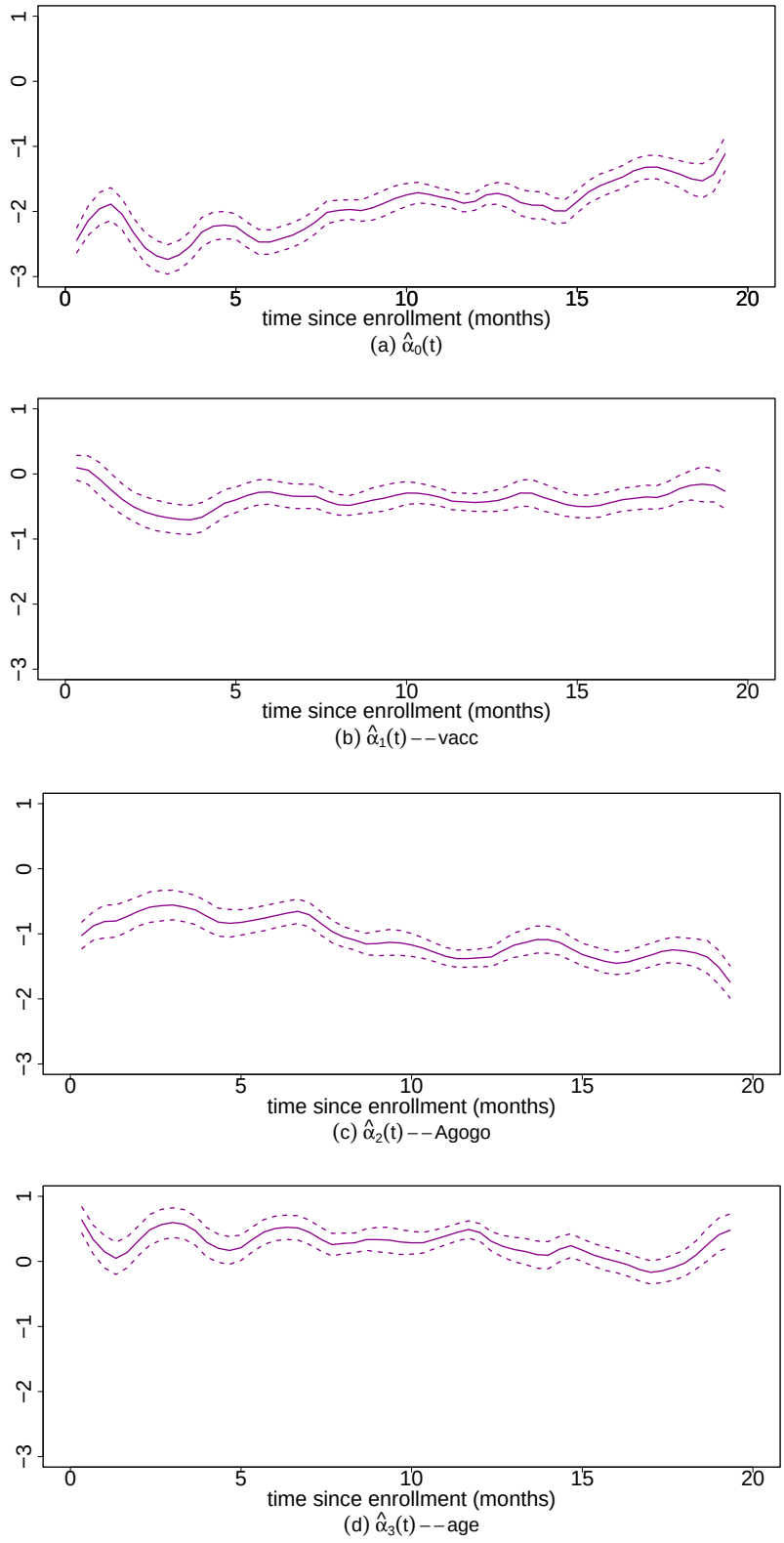

Figure 2.8: Estimation results of time-varying effects of covariates under model 2.20. The solid lines represent the point estimates, while the dashed lines signify the 95% pointwise confidence intervals.

Figure 2.8 presents the estimates and 95% pointwise confidence intervals of $\alpha_0(t)$, $\alpha_1(t)$, $\alpha_2(t)$ and $\alpha_3(t)$. From the figure, we find that the baseline intensity increases with time. The risk of infections among participants residing in Agogo is lower compared to those residing in Siaya, and this difference becomes increasingly pronounced over time. The positive estimate of $\alpha_3(t)$ suggests that older children have higher risk of infections.

In treatment group, participants consistently exhibit a lower infection risk against the first infections compared to those in the control group. The risk of reinfections becomes lower than those of control group after three months since the most recent infection.

To assess the level of protection against infections, we define vaccine efficacy as the percentage reduction in intensity among vaccinated individuals compared to those in the control group.

Under model 2.20, the vaccine efficacy at time t is defined as

$$\begin{aligned}\text{VE}(t) =& 1 - \frac{\lambda_i(t|\text{Vacc}_i = 1)}{\lambda_i(t|\text{Vacc}_i = 0)} \\ =& 1 - \exp\left\{\alpha_1(t) + \gamma_1(t - T_{iN_i(t^-)}, \theta) I(N_i(t^-) > 0)\right\}.\end{aligned}$$

When there are no prior infections, we have the vaccine efficacy against the first infection is represented by $\text{VE}(t) = 1 - \exp\{\alpha_1(t)\}$. 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 $\text{VE}(t) = 1 - \exp\{\alpha_1(t) + \gamma_1(t - T_{iN_i(t^-)}, \theta)\}$.

Figure 2.9 (a) shows the estimated vaccine efficacy against the first infections, the shaded area represents 95% confidence interval. The standard deviation is calculated by delta method, which equals $\exp\{\alpha_1(t)\}\text{ESE}(\alpha_1(t))$, where $\text{ESE}(\alpha_1(t))$ is the estimated standard error of $\alpha_1(t)$. From the plot, we can tell that the vaccine efficacy against first infections is approximately 40-50% within the first 5 months, declines to about 30% thereafter.

Figure 2.9 (b) shows the estimated vaccine efficacy against reinfections. This is a heat map, deeper color represents higher vaccine efficacy. It appears that the vaccine efficacy against reinfections is strongest within the first five months, ranging from 40-50%. Subsequently, the efficacy declines to around 10% at approximately 7 months after enrollment and then varies between 10% and 40%. This variability may be influenced by the timing of vaccinations, as illustrated in Figure 2.10.

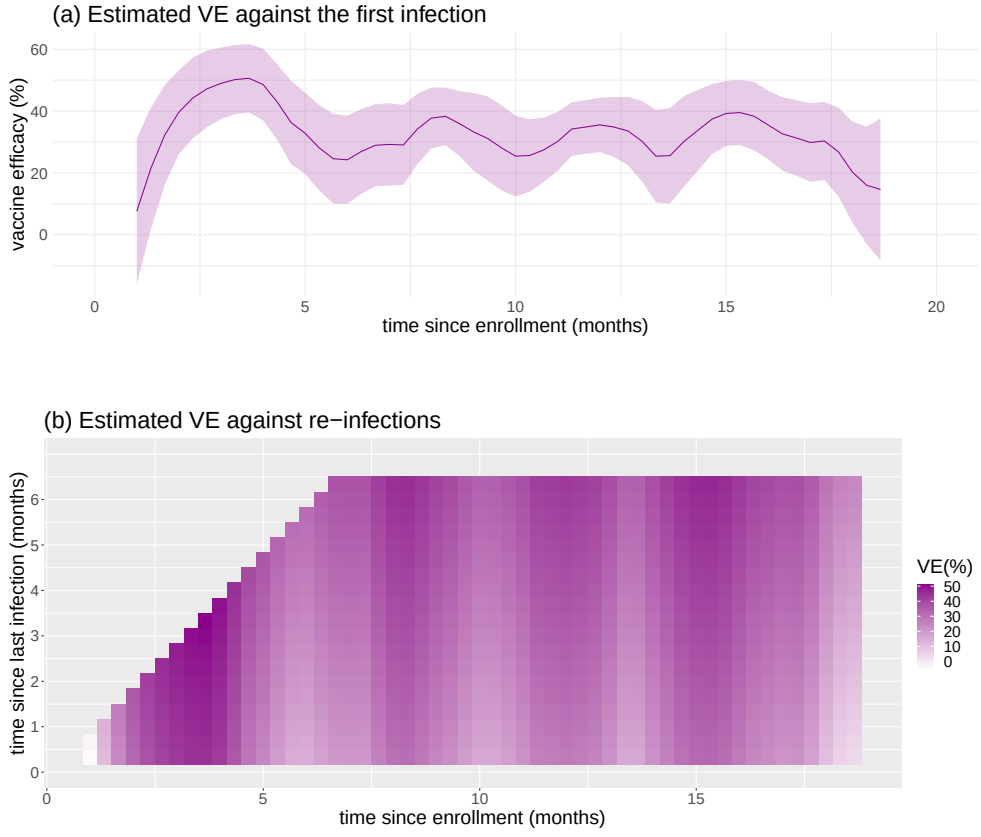

Figure 2.9: Estimated vaccine efficacy against the first infection (a) and re-infection (b) under model 2.20.

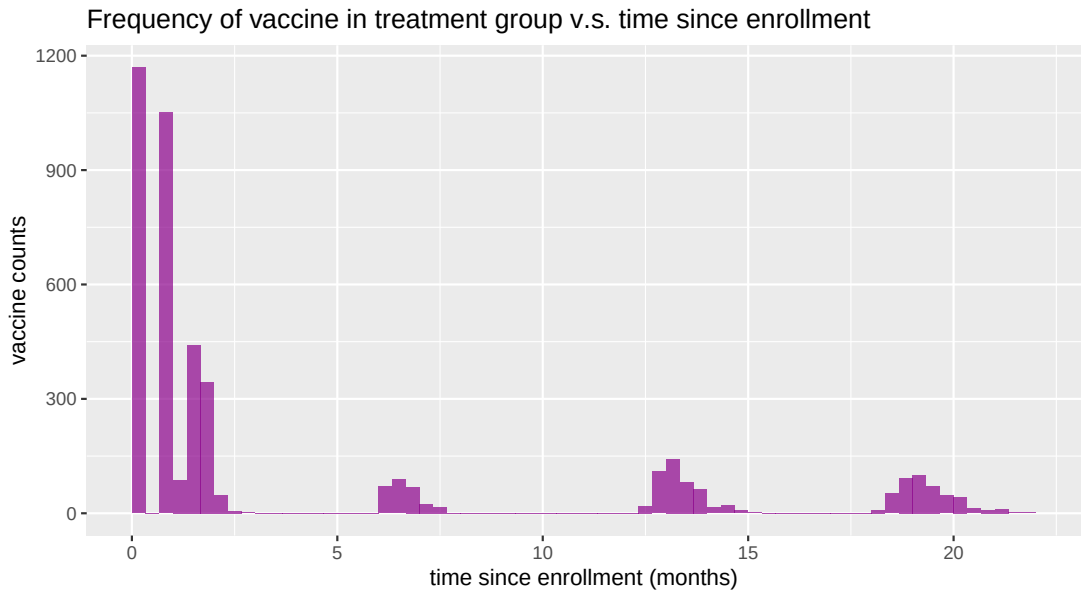

Figure 2.10: The frequency of vaccinations in treatment group over time since enrollment.

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 $\gamma_1(u,\tilde{\theta}) = \theta_2 + \theta_3 u$. At 0.05 significance level, we draw the same conclusions from both the supremum and chi-square tests. The linearity hypotheses for $\gamma_0(u,\theta)$ and $\gamma_1(u,\tilde{\theta})$ are found to hold.

Figure 2.11 plots the test process and the Gaussian multiplier processes. The red lines depict the test process $R(u,\hat{\eta})$, while the gray lines represent 500 realizations of Gaussian multiplier processes $R^*(u)$. Observing the plots, we notice that the test processes for all parameters fall within the gray regions. This suggests no evidence of departure from the null hypotheses, which is consistent with the p values given in Table 2.9.

Table 2.9: Test statistics and p values under model 2.20. T_1 is supremum type test. T_{2g} and T_{2c} are chi-square type tests. The critical value of T_{2g} is based on Gaussian multiplier distribution and the critical value of T_{2c} is based on chi-square distribution. The results are based on 500 Gaussian multiplier samples.

	T_1	T_{2g}	T_{2c}
Test statistic	2.391	27.002	27.002
p value	0.852	0.126	0.135

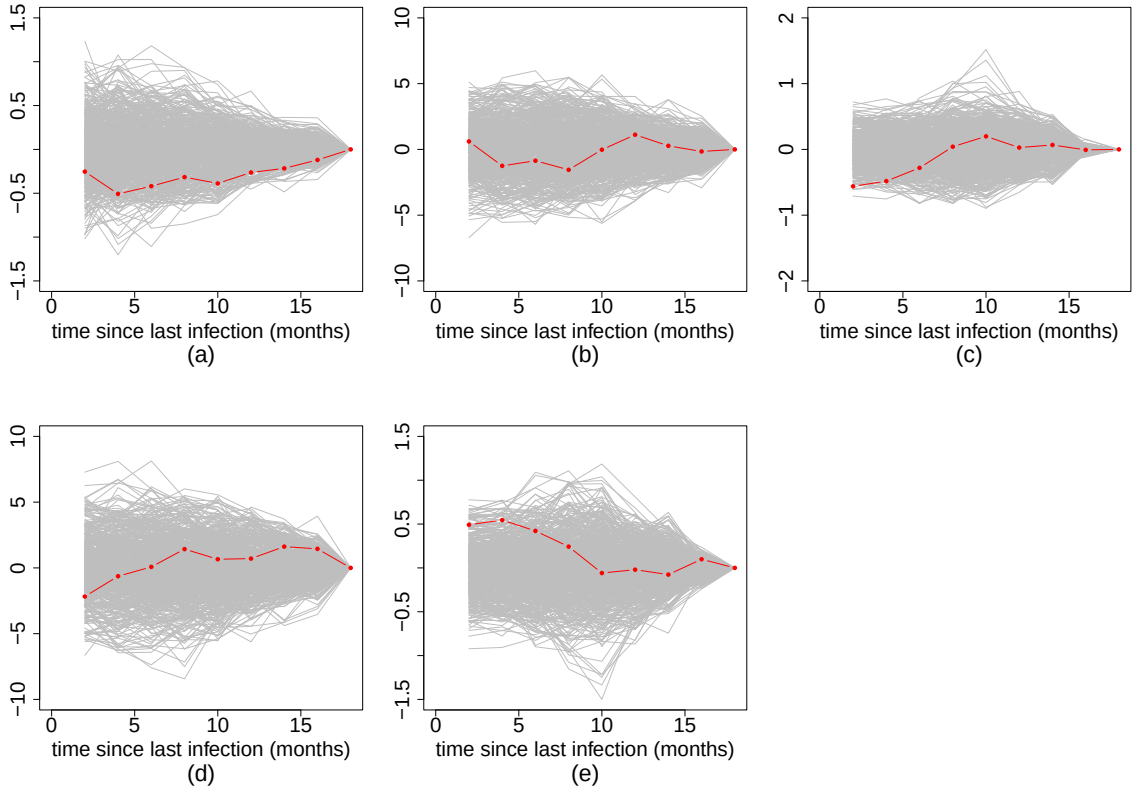

Figure 2.11: Test process $R(u, \hat{\eta})$ and 500 realizations of the Gaussian multiplier processes $R^*(u)$ under model 2.20. The plots in (a) to (e) correspond to parameters β, θ_0, θ_1, θ_2 and θ_3, respectively.

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

In this subsection, we model the vaccine effects parametrically and aim to investigate how these effects fluctuate over time since the most recent vaccination. Figure

2.12 shows the histogram of the gap times since last vaccination when infections occur, distinguishing between control and treatment group.

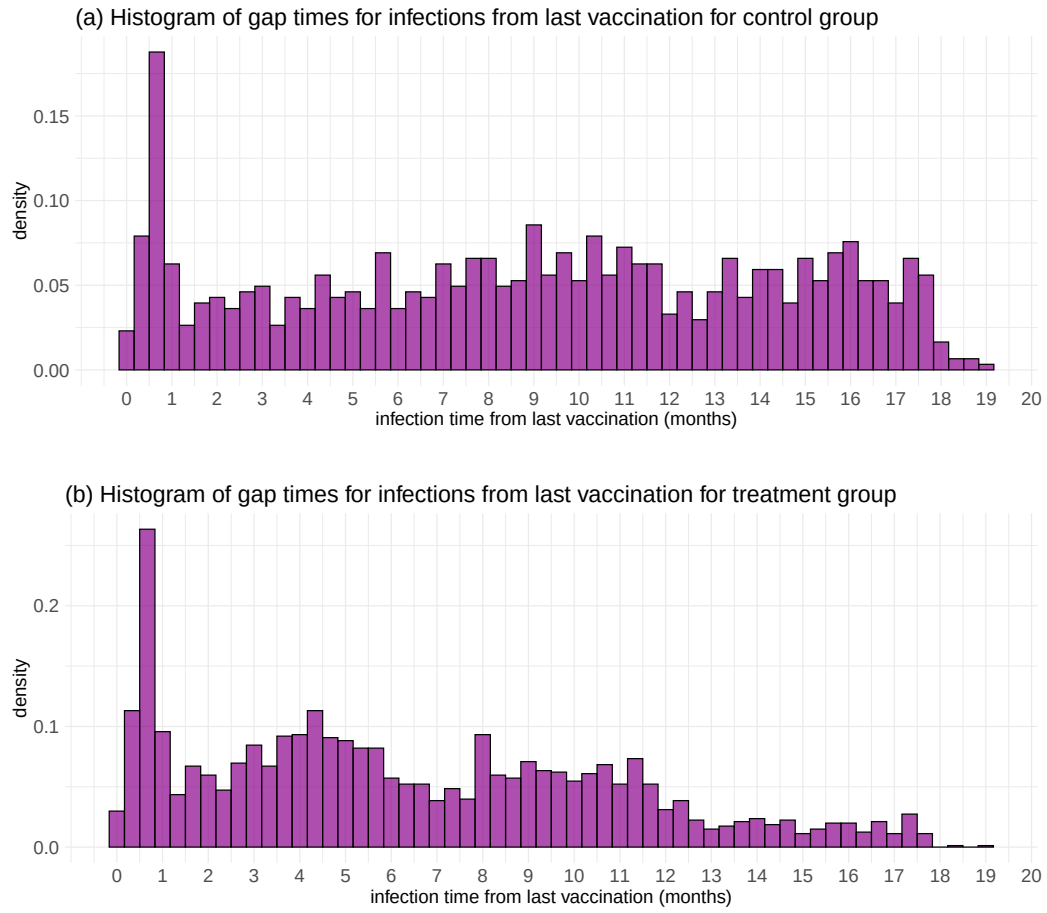

Figure 2.12: Histograms of gap times for infections from last vaccination for control and treatment group in the 20 months follow-up data.

Besides the definitions of infection times and counting process of infections, we also need to define the vaccination times. Define T_{ik}^V as the kth vaccination time for subject i. Denote v_i as the total vaccine doses for subject i before the end of study or censoring, whichever comes first. We have $T_{i1}^V < T_{i2}^V < ... < T_{iv_i}^V$. The counting process of vaccination is defined 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:

$$\lambda_i(t) = \exp\Big\{\alpha_0(t) + \alpha_1(t)\text{Agogo}_i + \alpha_2(t)\text{Age}_i(year) + \beta\text{Hemo}_i + \gamma(t - T^V_{iV_i(t^-)}, \theta)I(V_i(t^-) > 0)\text{Vacc}_i\Big\}, \tag{2.21}$$

for $0 \leq t \leq 20$(months).

In model 2.21, $\alpha_0(t)$, $\alpha_1(t)$, and $\alpha_2(t)$ are all unspecified functions that vary with calendar time t. $\gamma(u, \theta) = \theta_0 + \theta_1 u$, with unknown parameters θ_0 and θ_1. 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). Agogo_i is the study site indicator (1= Agogo, 0 = Siaya). Age_i represents the age in years at enrollment, and Hemo_i is the standardized hemoglobin level for the ith subject.

We use Epanechnikov kernel function $K(x) = 3/4(1 - x^2)I\{|x| \leq 1\}$ in the estimation and apply the Monte Carlo cross-validation method as described in Section 2.2.4 to select the optimal bandwidth. 100 repetitions of bootstrap yields an optimal bandwidth 1.42 months. Figure 2.13 displays the averaged accuracy per individual on the test dataset, with $\frac{K-1}{K}$ proportion of individuals are sampled as training dataset when $K = 3$ and $K = 5$.

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 $H_{01} : \beta = 0$, $H_{02} : \theta_0 = 0$ and $H_{03} : \theta_1 = 0$. At 0.05 significance level, the p values suggest us to conclude that hemoglobin and the vaccine effects are statistically significant. Hemoglobin level is negatively associated with the risk of infections. A higher level of hemoglobin corresponds to a lower malaria infection risk. The individuals in treatment group have lower infection risk than those in the control group ($\hat{\theta}_0 < 0$), but the effects slightly diminish over time since the most recent vaccination ($\hat{\theta}_1 > 0$).

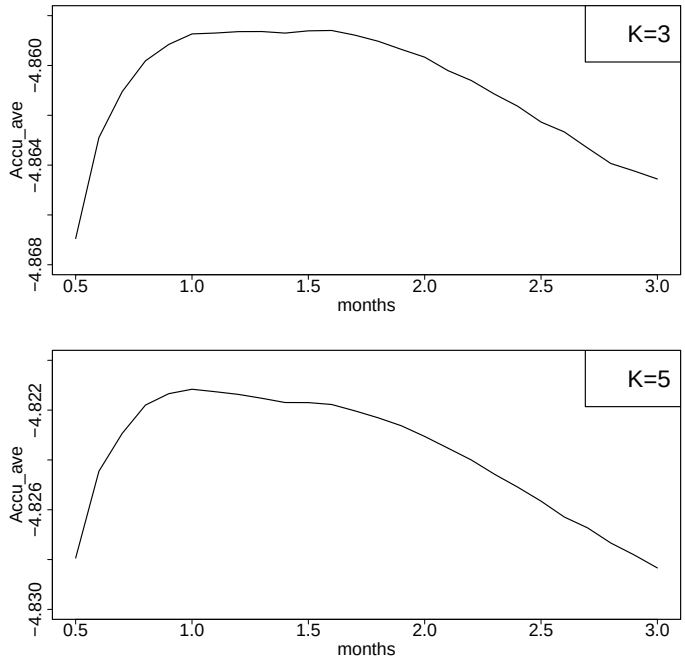

Figure 2.13: Average accuracy v.s. bandwidths when K=3 and K=5 under model 2.21.

Table 2.10: Estimates of parameters, their standard errors and p values under model 2.21, using bandwith $h = 1.42$.

	β	θ_0	θ_1
EST	-0.195	-0.639	0.024
ESE	0.030	0.078	0.007
p value	$\ll 0.001$	$\ll 0.001$	$\ll 0.001$

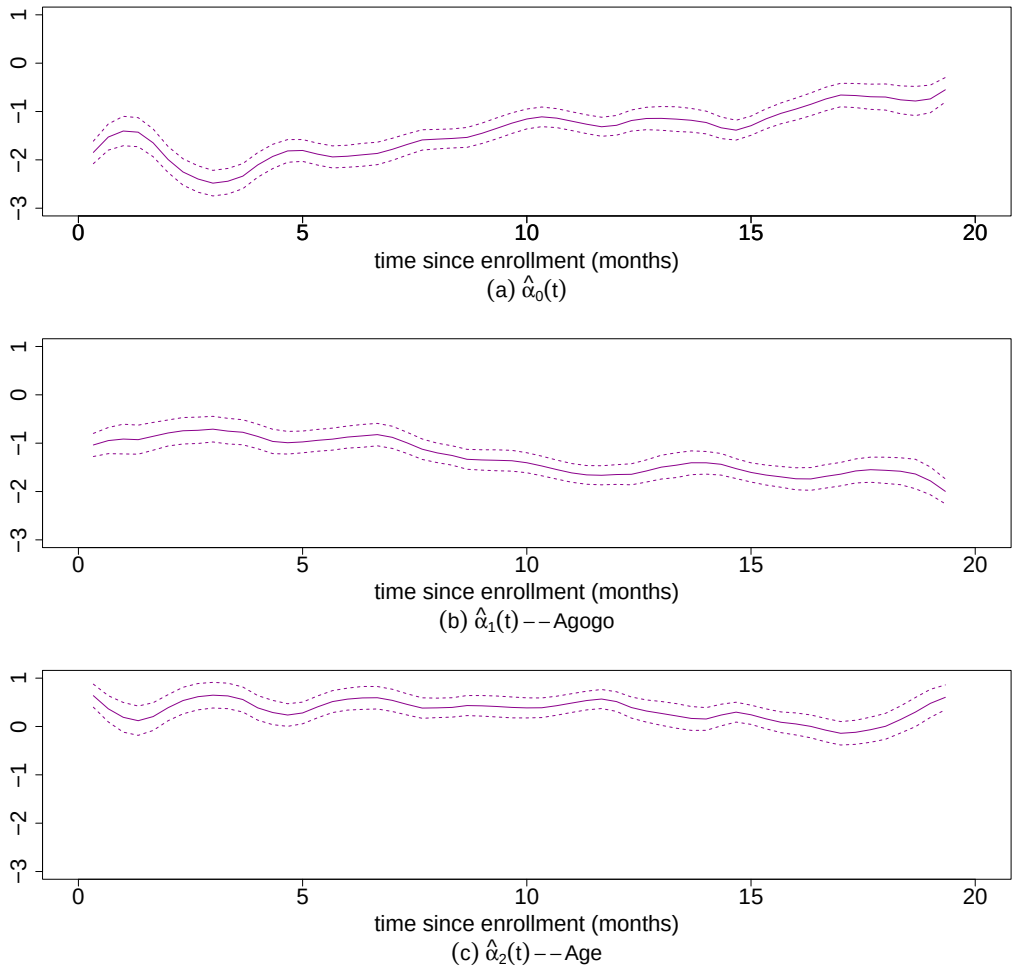

Figure 2.14: Estimation results of time-varying effects of covariates under model 2.21. The solid lines represent the point estimates, while the dashed lines signify the 95% pointwise confidence intervals.

From Figure 2.14, we draw similar conclusions as those in Section 2.6.1. Firstly, as time progresses, the baseline risk of infections increases. When comparing participants from the two study sites, we observe that the risk of infections is lower in Agogo. Additionally, older children are at a higher risk of infections.

The vaccine efficacy at time t can be defined as

$$\begin{aligned} \text{VE}(t) &= 1 - \frac{\lambda_i(t|\text{Vacc}_i = 1)}{\lambda_i(t|\text{Vacc}_i = 0)} \\ &= 1 - \exp\left\{\gamma(t - T^V_{iV_i(t^-)})I(V_i(t^-) > 0)\right\}. \end{aligned}$$

Figure 2.15 shows the estimated vaccine efficacy, the shaded area represents 95% confidence interval. The standard deviation is calculated by delta method, which equals

$$\exp\{\gamma(t - T^V_{iV_i(t^-)})I(V_i(t^-) > 0)\}\text{ESE}(\gamma(t - T^V_{iV_i(t^-)})),$$

where $\text{ESE}(\gamma(t - T^V_{iV_i(t^-)}))$ is the estimated standard error of $\gamma(t - T^V_{iV_i(t^-)})$. From Figure 2.15, we can tell the vaccine efficacy decreases with the time since the most recent vaccination, approximately from 40% to 20%.

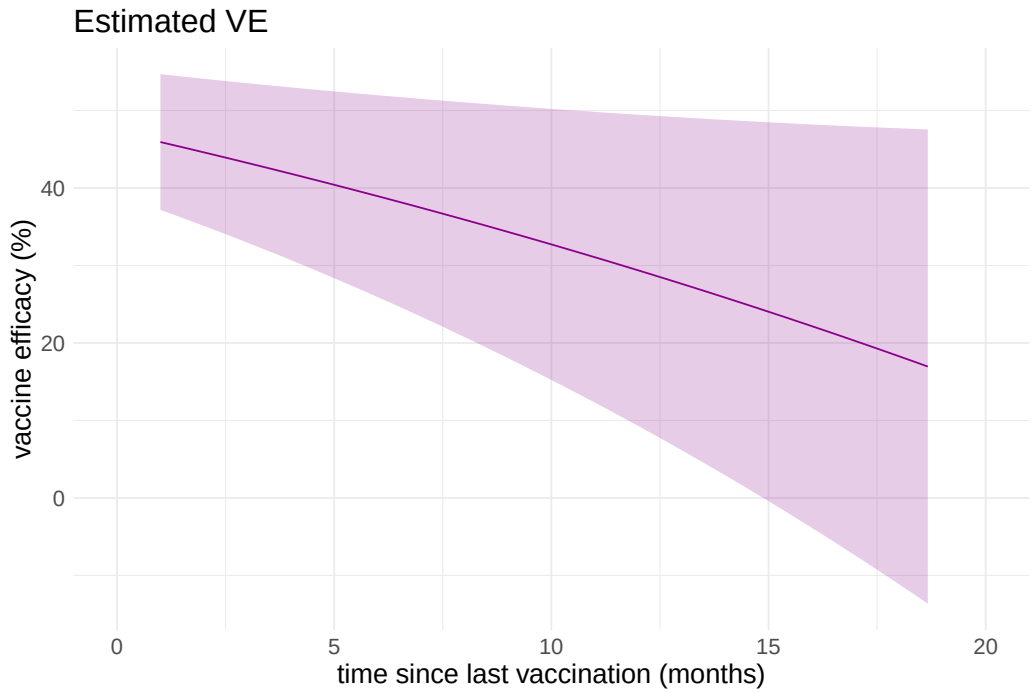

Figure 2.15: Estimated vaccine efficacy over time since last vaccination under model 2.21. The solid line represents the point estimate, while the shaded area signifies the 95% pointwise confidence interval.

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 significance level, both supremum and chi-square tests indicate acceptance of the linear hypothesis of $\gamma(u, \theta)$. Figure 2.16 are plots of test process and Gaussian multiplier processes. The red lines depict the test process $R(u, \hat{\eta})$ while the gray lines represent 500 realizations of Gaussian multiplier process $R^*(u)$. Observing the plots, we notice that the test processes for all parameters fall within the gray regions. This suggests no departure from the null hypothesis.

Table 2.11: Test statistics and p values under model 2.21. T_1 is supremum type test. T_{2g} and T_{2c} are chi-square type tests. The critical value of T_{2g} value is based on Gaussian multiplier distribution and the critical value of T_{2c} is based on chi-square distribution. The results are based on 500 Gaussian multiplier samples.

	T_1	T_{2g}	T_{2c}
Test Statistic	2.128	29.965	29.965
p value	0.158	0.188	0.186

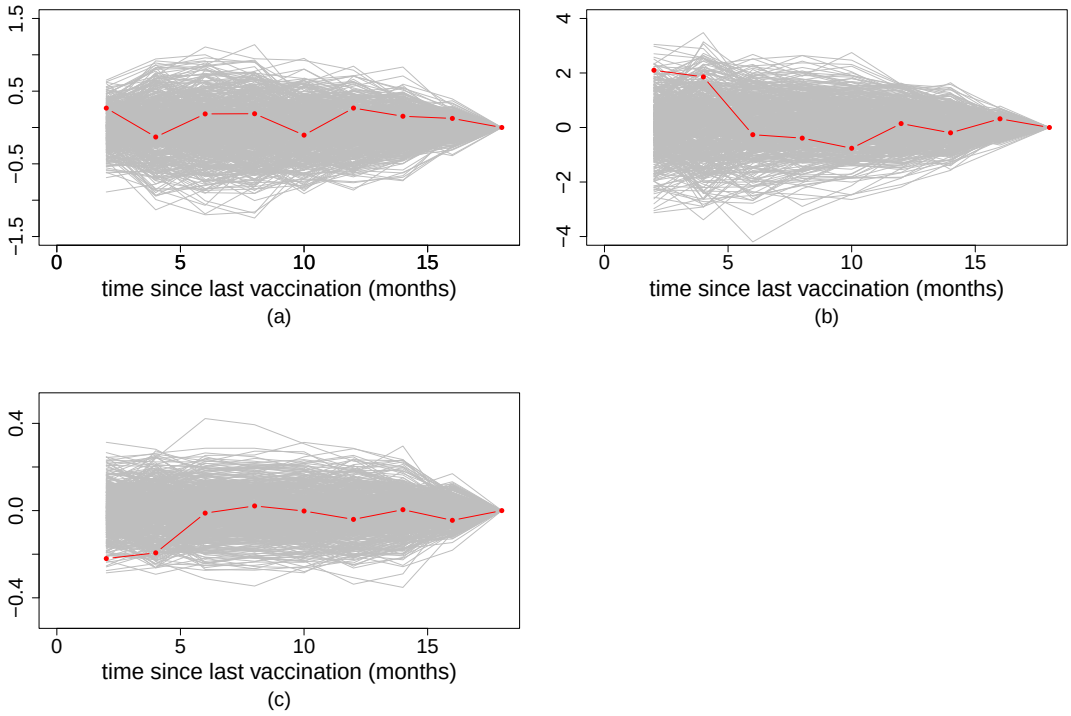

Figure 2.16: Test process $R(u, \hat{\eta})$ and 500 realizations of Gaussian multiplier process $R^*(u)$ under model 2.21. The plots in (a) to (c) correspond to the parameters β, θ_0 and θ_1, respectively.

2.7 Summary

In this chapter, we proposed a generalized class of semiparametric intensity models for recurrent event data. These models offer significant flexibility through various choices of link functions and parametric functions of covariate-varying effects.

Maximum log-likelihood estimation procedures were investigated through local linear approximation and profile likelihood method. Simulation results showed that the estimation procedure performs well under different scenarios. Asymptotic properties of the estimators were derived, including the consistency and asymptotic normality. Hypothesis test procedures were developed to test the parametric forms of the covariate-varying effects ($\gamma(u, \theta)$). We derived an Gaussian multiplier procedure to approximate the critical values of the test statistics. The simulations showed that this method works well with the observed sizes of the tests close to their nominal level 0.05.

We applied these methods to the 20 months follow-up MAL-094 vaccine trial data. Based on the results from the two models discussed in Section 2.6, we observed that site, age and hemoglobin have significant influence on the risk of infections. Specifically, participants in Agogo exhibited a significantly lower infection risk compared to those in Siaya. Additionally, older children were found to have a relatively higher risk of infections. Notably, we discovered a negative correlation between hemoglobin levels and infection risk, participants with higher hemoglobin levels tended to have a lower risk of infections.

From Section 2.6.1, we can tell that following prior infections, the risk of subsequent infection increased for both control and treatment groups. However, for participants in the treatment group, the increment in risk was lower than that observed in the control group after approximately 3 months since the most recent infection.

The vaccine efficacy against the first infections was about 40-50% at the first 5 months and decreased to about 30% after 5 months. The analysis also showed that the vaccine efficacy against reinfections varied between 10-40%. The pattern of fluctuation in vaccine efficacy seemed to be associated with the timing of vaccine boosters (see Figure 2.10). Hypothesis testing results showed that the parametric forms $\gamma_0(u, \theta)$ and $\gamma_1(u, \tilde{\theta})$ used in the analysis were adequate.

From Section 2.6.2, it was observed that the vaccine efficacy against infections decreased with time since the most recent vaccination, declining from approximately 40% to 20%. Hypothesis test was also conducted and showed that the parametric form $\gamma(u, \theta)$ used in the analysis was adequate as well.

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

3.1 Introduction

While the frailty model 1.1 proposed by Nileson *et al.* (1992) does not accommodate time-varying effects, it provides a foundational framework that can be extended to more specialized models tailored to address specific research questions. In this chapter, we build upon this idea by incorporating the frailty into a nonparametric dynamic intensity model for analyzing recurrent event data. This model assesses both time-varying and covariate-varying effects using unspecified functions. Frailty is incorporated multiplicatively into the intensity function to account for unobserved heterogeneity within the population.

In Section 3.2, we introduce the model and develop a maximum likelihood estimation procedure that utilizes local linear approximation through a double kernel approach. We employ an EM algorithm to maximize the likelihood. Additionally, in Section 3.2.4, we discuss an adaptive algorithm designed to address potential identifiability issues under specific settings. In Section 3.3, a weighted bootstrap method is presented for estimating the variance of the estimators. In Section 3.4, we validate the estimation procedure through finite sample simulations across various scenarios. In Section 3.5, the proposed frailty intensity model is applied to the MAL-094 malaria vaccine trial data.

3.2 Model and Estimation

3.2.1 Model Descriptions

Suppose there is a random sample of n subjects. τ is duration of study time. Denote n_i as the total number of events for subject i. The event time T_{ij}, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n_i$, represents the j th event time for subject i. We have $T_{i1} < T_{i2} < ... < T_{in_i} \leq \tau$. $X_i(t)$, $Z_i(t)$ and $U_i(t)$ 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, the intensity of counting process $N_i^*(t)$ is defined as $\lambda_i^f(t) = \lim_{\Delta t \downarrow 0} \Pr(\Delta N_i^*(t) = 1|\mathcal{F}_{it^-}^*)/\Delta t$. By definition, we have $E(dN_i^*(t)|\mathcal{F}_{it^-}^*) = \lambda_i^f(t)dt$.

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. Let $\mathcal{F}_{it^-} = \mathcal{F}_{it^-}^* \vee C_i$ be the filtration generated by the observed event history, covariate processes and censoring for subject i. The censoring is assumed to be non-informative in the sense that $E(dN_i(t)|\mathcal{F}_{it^-}) = E(dN_i(t)|\mathcal{F}_{it^-}^*) = Y_i(t)\lambda_i^f(t)dt$. The observed data consists of $\mathcal{D} = \Big\{N_i(t), Y_i(t), X_i(t), Z_i(t), U_i(t), t \in [0, \tau_i]\Big\}, (i = 1, ..., n)$.

The nonparametric dynamic intensity model incorporating frailty is proposed as:

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

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

The frailty term ξ_i, also known as a random effect, is independent with both the subjects' covariates and stochastic processes. The ξ_i are independent and identically distributed, following Gamma(θ, θ). The probability density function for ξ_i is given by

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

for $\xi > 0$, $\theta > 0$, where $\Gamma(\theta)$ is the Gamma function.

Constraining the two parameters of the Gamma distribution to be identical ensures scaling such that the resulting distribution has a mean of 1 and a variance of $\frac{1}{\theta}$, which helps avoid identifiability issues. The identifiability problem arises from the fact that multiplying ξ_i by a constant and dividing the entire exponential term by the same constant keeps ξ_i following a Gamma distribution, albeit with a new scale parameter(Nileson *et al.* (1992)).

The frailty term ξ_i seeks to model unobserved heterogeneity among the subjects and induce the dependence structure among recurrent events within 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.

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, allowing the intensity function to vary based on this history. As an illustration, consider setting $U_i(t) = t - T_{iN_i(t^-)}$ and $Z_i(t) = I(N_i(t^-) > 0)$. The covariate $U_i(t)$ represents the time since most recent event. The indicator $I(N_i(t^-) > 0)$ is essential because the gap time $U_i(t)$ is only meaningful once the subject has experienced at least one event. In another example, by defining $U_i(t) = t - V_i(t)$ and $Z_i(t) = I(t > V_i(t))$, where $V_i(t)$ represents the time of vaccination or intervention,

model 3.1 captures changes in intensity following the vaccination or intervention.

3.2.2 Nonparametric Maximum Likelihood Estimation

For ease of representation, we denote

$$\lambda_i(t) = \exp\{\alpha^{\mathsf{T}}(t)X_i(t) + \gamma^{\mathsf{T}}(U_i(t))Z_i(t)\}, \tag{3.2}$$

then the intensity function 3.1 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

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

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

Integrate over ξ_i, the observed data likelihood is given by

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

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

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

We use local linear approximation method to estimate the nonparametric estimators $\alpha(\cdot)$ and $\gamma(\cdot)$(Heng (2019)). 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 and second 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 frailty included intensity function localized at (t_0,u_0), we have

$$\lambda_i^{f*}(t,\vartheta^*|t_0,u_0) = \xi_i\exp\{\vartheta^{*T}(t_0,u_0)\tilde{X_i}^*(t|t_0,u_0)\}, \tag{3.3}$$

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

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

the approximated intensity function 3.3 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 obtain maximum likelihood estimators. In this process, ξ_i is treated as a latent variable. During the E-step, we calculate the conditional expectation of ξ_i, and in the M-step, we maximize the conditional expectation of the complete log-likelihood. This iterative method enables us to estimate model parameters $\alpha(\cdot)$, $\gamma(\cdot)$, and θ, the parameter governing the frailty term.

E-Step. Following Nileson *et al.* (1992), we can show that, conditional on observed data, ξ_i is conditional Gamma distributed and follows Gamma$(\theta + N_i(\tau), \theta + \int_0^\tau Y_i(t)\lambda_i(t)dt)$, where $\lambda_i(t)$ takes the form 3.2.

Denote $E(\xi_i|\mathcal{D})$ and $E(\log\xi_i|\mathcal{D})$ as the conditional expectations of ξ_i and $\log\xi_i$, respectively, given the observed data. We can express them in closed forms as follows:

$$E(\xi_i|\mathcal{D}) = \frac{\theta + N_i(\tau)}{\theta + \int_0^\tau Y_i(t)\lambda_i(t,\alpha(t),\gamma(U_i(t)))dt} \tag{3.4}$$

and

$$E(\log \xi_i|\mathcal{D}) = \frac{\Gamma'(\theta + N_i(\tau))}{\Gamma(\theta + N_i(\tau))} - \log\Big\{\theta + \int_0^\tau Y_i(t)\lambda_i(t, \alpha(t), \gamma(U_i(t)))dt\Big\}. \tag{3.5}$$

The estimated conditional expectations $\hat{E}(\xi_i|\mathcal{D})$ and $\hat{E}(\log \xi_i|\mathcal{D})$ can be obtained by replacing $\alpha(t)$, $\gamma(U_i(t))$ and θ by the estimates at the previous iteration.

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

$$\begin{aligned} E[l(\vartheta^*, \theta|t_0, u_0)] = \sum_{i=1}^n \int_0^\tau K_h(t - t_0)K_b(U_i(t) - u_0)\Big[\Big\{\hat{E}(\log \xi_i|\mathcal{D}) \\ + \log\{Y_i(t)\lambda_i^*(t, \vartheta^*|t_0, u_0)\}\Big\}dN_i(t) - Y_i(t)\hat{E}(\xi_i|\mathcal{D})\lambda_i^*(t, \vartheta^*|t_0, u_0)dt\Big] \\ + \sum_{i=1}^n \hat{E}\{\log f(\xi_i)|\mathcal{D}\}, \end{aligned} \tag{3.6}$$

where $K_h(.) = K_1(./h)/h$, $K_b(.) = K_2(./b)/b$, $K_1(.)$, $K_2(.)$ are kernel functions, h and b are bandwidth parameters.

Maximizing 3.6 can be conducted by maximizing the following two parts,

$$\begin{aligned} E[l_1^{c*}(\vartheta^*|t_0, u_0)|\mathcal{D}] = \sum_{i=1}^n \int_0^\tau K_h(t - t_0)K_b(U_i(t) - u_0)\Big[\Big\{\hat{E}(\log \xi_i|\mathcal{D}) \\ + \log\{Y_i(t)\lambda_i^*(t, \vartheta^*|t_0, u_0)\}\Big\}dN_i(t) - Y_i(t)\hat{E}(\xi_i|\mathcal{D})\lambda_i^*(t, \vartheta^*|t_0, u_0)dt\Big] \end{aligned} \tag{3.7}$$

and

$$\begin{aligned} E[l_2(\theta)|\mathcal{D}] &= \sum_i^n \hat{E}(\log f(\xi_i)|\mathcal{D}) \\ &= n\theta \log \theta - n \log \Gamma(\theta) + (\theta - 1)\sum_{i=1}^n \hat{E}(\log \xi_i|\mathcal{D}) - \theta \sum_{i=1}^n \hat{E}(\xi_i|\mathcal{D}). \end{aligned} \tag{3.8}$$

The estimated conditional expectations $\hat{E}(\xi_i|\mathcal{D})$ and $\hat{E}(\log \xi_i|\mathcal{D})$ can be calculated in

E-step based on the observed data and estimates at previous iteration.

Take derivative of $E[l_1^{c*}(\vartheta^*|t_0,u_0)|\mathcal{D}]$ with respect to ϑ^*, we get the local score function of $\vartheta^*(t_0,u_0)$,

$$U(\vartheta^*|t_0,u_0)=\sum_{i=1}^{n}\int_0^{\tau}K_h(t-t_0)K_b(U_i(t)-u_0)\tilde{X}_i^*(t|t_0,u_0)\\ \times\Big\{dN_i(t)-Y_i(t)\hat{E}(\xi_i|\mathcal{D})\lambda_i^*(t,\vartheta^*|t_0,u_0)dt\Big\}. \tag{3.9}$$

Set $U(\vartheta^*|t_0,u_0)=\mathbf{0}$, 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 $\hat{\gamma}(u_0)$:

$$\hat{\alpha}(t_0)=n^{-1}\sum_{1=1}^{n}\hat{\alpha}(t_0,U_i(t_0)), \tag{3.10}$$

and

$$\hat{\gamma}(u_0)=n_{u_0}^{-1}\sum_{t_{u_0}\in\mathcal{V}_{u_0}}\hat{\gamma}(t_{u_0},u_0), \tag{3.11}$$

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 $\mathcal{V}_{u_0}$.

Take derivative of $E[l_2(\theta)|\mathcal{D}]$ with respect to θ and set it to zero, getting

$$\log(\theta)-\frac{\Gamma'(\theta)}{\Gamma(\theta)}+\frac{1}{n}\sum_{i=1}^{n}\hat{E}(\log\xi_i|\mathcal{D})-\frac{1}{n}\sum_{i=1}^{n}\hat{E}(\xi_i|\mathcal{D})+1=0, \tag{3.12}$$

θ can be updated by solving equation 3.12.

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

3.2.3 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 3.4 and 3.5. 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

$$\begin{aligned}\hat{E}^{\{k\}}(\log \xi_i) =& \frac{\Gamma'(\hat{\theta}^{\{k-1\}} + N_i(\tau))}{\Gamma(\hat{\theta}^{\{k-1\}} + N_i(\tau))} - \log\Big\{\hat{\theta}^{\{k-1\}} \\ &+ \int_0^\tau Y_i(t)\lambda_i\{t, \hat{\alpha}^{\{k-1\}}(t), \hat{\gamma}^{\{k-1\}}(U_i(t))\}dt\Big\}.\end{aligned}$$

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 3.9 with $\hat{E}(\xi_i|\mathcal{D})$ replaced by $\hat{E}^{\{k\}}(\xi_i|\mathcal{D})$. Then we take corresponding components and aggregate through equation 3.10 and 3.11. Update $\hat{\theta}^{\{k\}}$ by solving equation 3.12 with $\hat{E}(\xi_i|\mathcal{D})$ and $\hat{E}(\log \xi_i|\mathcal{D})$ replaced by $\hat{E}^{\{k\}}(\xi_i|\mathcal{D})$ and $\hat{E}^{\{k\}}(\log \xi_i|\mathcal{D})$.

5. Repeat Step 3 and Step 4 iteratively until converge, the resulting 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.

3.2.4 An Adaptive Estimation Algorithm

In previous algorithm, we have an implicit assumption $X_i(t) \neq Z_i(t)$. However, 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 $X_i(t) = 1$, $U_i(t) = t - t_{iN_i(t^-)}$ 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 $(t_0, u_0) \in \Delta = \{0 \leq u \leq t \leq t^*\}$, where $t^* \leq h + b$, $\alpha(t_0)$ 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 estimate $\gamma(u)$ separately.

The overall estimation procedure follows a similar outline as sketched in Section 3.2.3, 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 3.9 using double kernel method, we need the following procedure to estimate $\alpha(t)$ and $\gamma(u)$ separately.

1. For $(t_0, u_0) \in \Delta = \{0 \leq u \leq t \leq t^*\}$, where $t^* \leq h + b$, we can estimate $\vartheta^*(t_0, u_0)$ by solving $U(\vartheta^*|t_0, u_0) = 0$ as we described in equation 3.9. 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 $\tilde{X}_i^{**}(t|t_0, u_0) = (X_i^\mathsf{T}(t)(t - t_0), Z_i^\mathsf{T}(t), Z_i^\mathsf{T}(t)(U_i(t) - u_0))^\mathsf{T}$. Denote $\lambda_i^{**}(t) = \exp\{\alpha^\mathsf{T}(t_0)X_i(t) + \vartheta^{**\mathsf{T}}(t_0, u_0)\tilde{X}_i^{**}(t|t_0, u_0)\}$.

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

$$
\begin{aligned}
E[l_1^{c*}(\alpha, \vartheta^{**}|t_0, u_0)|D] = \sum_{i=1}^{n} \int_0^\tau K_h(t - t_0)K_b(U_i(t) - u_0)\Big[\Big\{\hat{E}(\log \xi_i|\mathcal{D}) \\
+ \log\{Y_i(t)\lambda_i^{**}(t, \alpha, \vartheta^{**}|t_0, u_0)\}\Big\}dN_i(t) \\
- Y_i(t)\hat{E}(\xi_i|\mathcal{D})\lambda_i^{**}(t, \alpha, \vartheta^{**}|t_0, u_0)dt\Big]. \qquad (3.13)
\end{aligned}
$$

   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 $\hat{\vartheta}{**}(t_0, u_0) = (\hat{\dot{\alpha}}^\mathsf{T}(t_0), \hat{\gamma}^\mathsf{T}(u_0), \hat{\dot{\gamma}}^\mathsf{T}(u_0))^\mathsf{T}$ 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 3.11.

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

3.3 Variance Estimator

In this session, we employed the weighted bootstrap procedure (Ma and R.Kosorok (2005)) to get the variance estimators for $\hat{\alpha}(t)$, $\hat{\gamma}(u)$ and $\hat{\theta}$. The basic idea is to assign independent and identically distributed (i.i.d.) positive random weights to each observation. We then obtain weighted estimators from this bootstrap sample.

By repeating this process multiple times, we consider the variance estimators to be the sample variance of all the weighted estimators.

Let $\{\omega_1, \omega_2, ...\omega_n\}$ be n independent realizations of random variable Ω, which follows exponential distribution with mean 1. The weights $\{\omega_1, \omega_2, ...\omega_n\}$ are independent with the observed data $\mathcal{D}$.

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

$$\log l_n^{\circ\omega}(\alpha(\cdot), \gamma(\cdot), \theta) = \sum_{i=1}^{n} \omega_i \Big[\int_0^\tau \log(Y_i(t)\lambda_i(t)) dN_i(t) + \theta \log \theta - \log \Gamma(\theta) + \log \Gamma\{\theta + N_i(\tau)\} - \{\theta + N_i(\tau)\} \log \Big\{ \theta + \int_0^\tau Y_i(t)\lambda_i(t)dt \Big\} \Big]. \tag{3.14}$$

To maximize 3.14, we continue to use the EM algorithm. In the E-step, we update the conditional expectations $\hat{E}(\xi_i|\mathcal{D})$ and $\hat{E}(\log \xi_i|\mathcal{D})$ according to 3.4 and 3.5. These updates involve replacing $\alpha(t)$, $\gamma(U_i(t))$, and θ with the weighted estimates from the previous iteration. In M-step, we maximize the weighted version of conditional localized complete log-likelihood

$$E[l^\omega(\vartheta^*, \theta|t_0, u_0)|\mathcal{D}] = E[l_1^\omega(\vartheta^*|t_0, u_0)|\mathcal{D}] + E[l_2^\omega(\theta)|\mathcal{D}],$$

where

$$E[l_1^\omega(\vartheta^*|t_0, u_0)|\mathcal{D}] = \sum_{i=1}^{n} \omega_i \int_0^\tau K_h(t - t_0) K_b(U_i(t) - u_0) \Big[\Big\{ \hat{E}(\log \xi_i|\mathcal{D}) + \log\{Y_i(t)\lambda_i^*(t, \vartheta^*|t_0, u_0)\} \Big\} dN_i(t) - Y_i(t)\hat{E}(\xi_i|\mathcal{D})\lambda_i^*(t, \vartheta^*|t_0, u_0)dt \Big],$$

and

$$E[l_2^{\omega}(\theta)|\mathcal{D}] = \sum_{i=1}^{n} \omega_i \theta \log \theta - \sum_{i=1}^{n} \omega_i \log \Gamma(\theta) + (\theta - 1) \sum_{i=1}^{n} \omega_i \hat{E}(\log \xi_i|\mathcal{D}) - \theta \sum_{i=1}^{n} \omega_i \hat{E}(\xi_i|\mathcal{D}).$$

The estimation procedures follow the same approach as outlined in Section 3.2.2, but in a weighted version. Iterate between E-step and M-step until converge to obtain the weighted estimators. Suppose we undertake 100 repetitions of this weighted bootstrap, resulting in 100 weighted estimators. The sample variance of these weighted estimators serves as the estimated variance of $\hat{\alpha}(t)$, $\hat{\gamma}(u)$, and $\hat{\theta}$.

3.4 Simulation Studies

In this section, we perform simulations to demonstrate the effectiveness of the proposed methods. Section 3.4.1 focuses on the double kernel estimation method, and Section 3.4.2 delves into the adaptive method. We use Epanechnikov kernel function $K(x) = 3/4(1 - x^2)I\{|x| \leq 1\}$. All the variance estimators are obtained through 100 repetitions of weighted bootstrap that we illustrated in Section 3.3.

In the simulation studies in this section, we employ the following abbreviations. 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.

3.4.1 Simulation Studies Using Double Kernel Algorithm

Generate data from following models:

$$\lambda_i^f(t) = \xi_i \exp\{\alpha_0(t) + \alpha_1(t)X_i + \gamma(U_i(t))I(N_i(t^-) > 0)W_i\}, \tag{3.15}$$

for $t \in [0, \tau]$.

- $\tau = 4$, For each subject i, we generate censoring time $C_i \sim U(3, 8)$, the study time for subject i is the minimum of C_i and τ.
- $X_i \sim Ber(0.5)$, $W_i \sim U(0, 1)$, $U_i(t) = t - T_{iN_i(t^-)}$.
- $\alpha_0(t) = 1 - \log(1 + 0.2\log(1 + t))$, $\alpha_1(t) = -0.5 + 0.1t$ and $\gamma(u) = -\frac{0.3}{1+u}$.
- $\xi_i \sim \text{Gamma}(\theta, \theta)$.

Under these settings, the average number of events per subject is approximately 7. A smaller θ in the Gamma distribution implies greater variability in the ξ_i. In this subsection, we simulate three scenarios: $\theta = 1$, $\theta = 2$, and $\theta = 5$. Table 3.1 summarizes the estimation results for θ under these scenarios. Figures 3.1, 3.2, and 3.3 illustrate the estimation results for $\alpha_0(t)$, $\alpha_1(t)$, and $\gamma(u)$ when $\theta = 1$, $\theta = 2$, and $\theta = 5$, respectively.

Table 3.1: Bias, ESE, SEE and CP for estimator of θ under model 3.15 when $\theta = 1, 2,$ and 5 for sample size $n = 800$, 1000 and 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. The results are based on 500 simulations. The estimated standard error in each simulation is obtained via 100 weighted bootstrap samples.

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

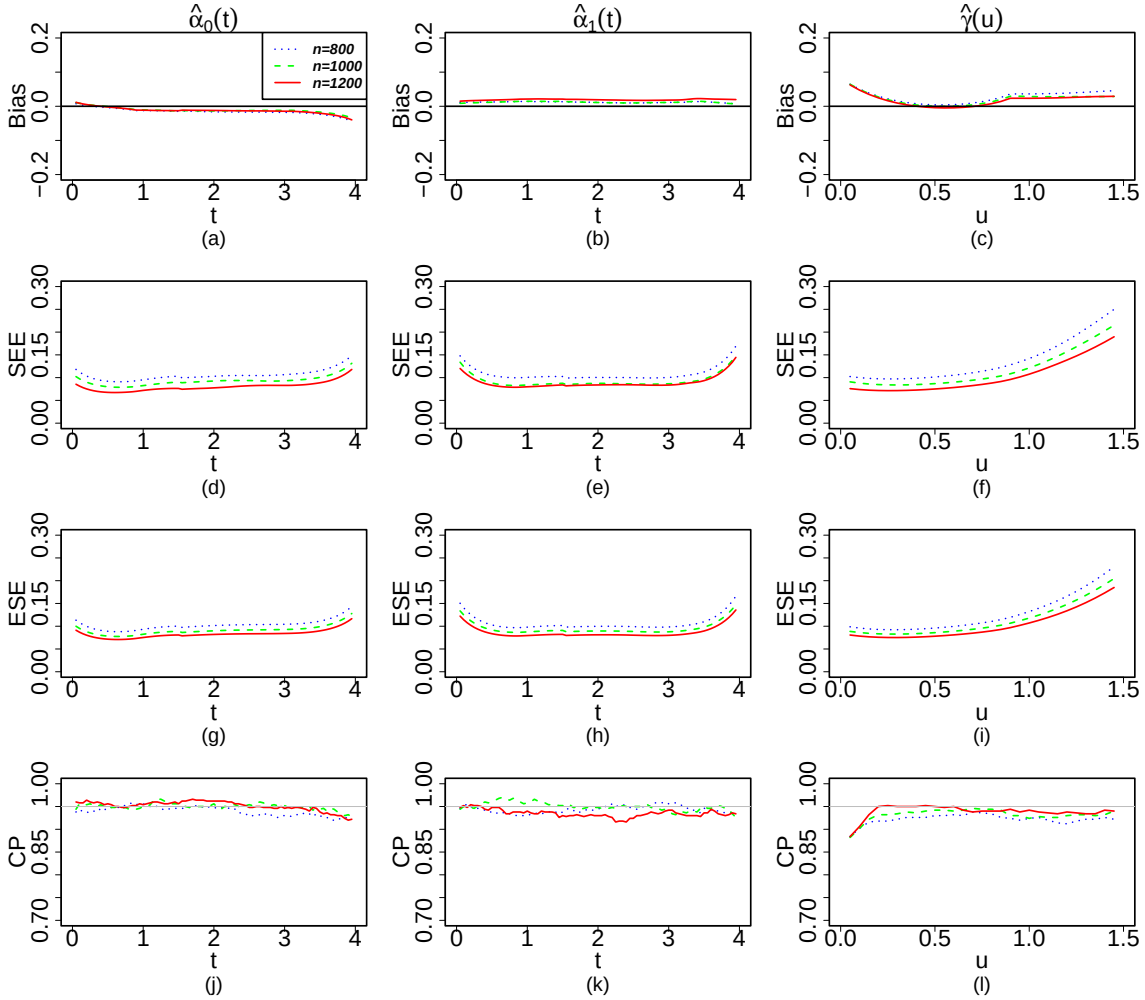

Figure 3.1: Bias, SEE, ESE and CP for estimators of $\alpha_0(t)$, $\alpha_1(t)$ and $\gamma(u)$ under model 3.15 when $\theta = 1$, using bandwidths $h = b = 0.9$. The dotted, dashed and solid lines represent sample size $n = 800$, 1000 and 1200, respectively. The results are based on 500 simulations. The estimated standard errors in each simulation are obtained via 100 weighted bootstrap samples.

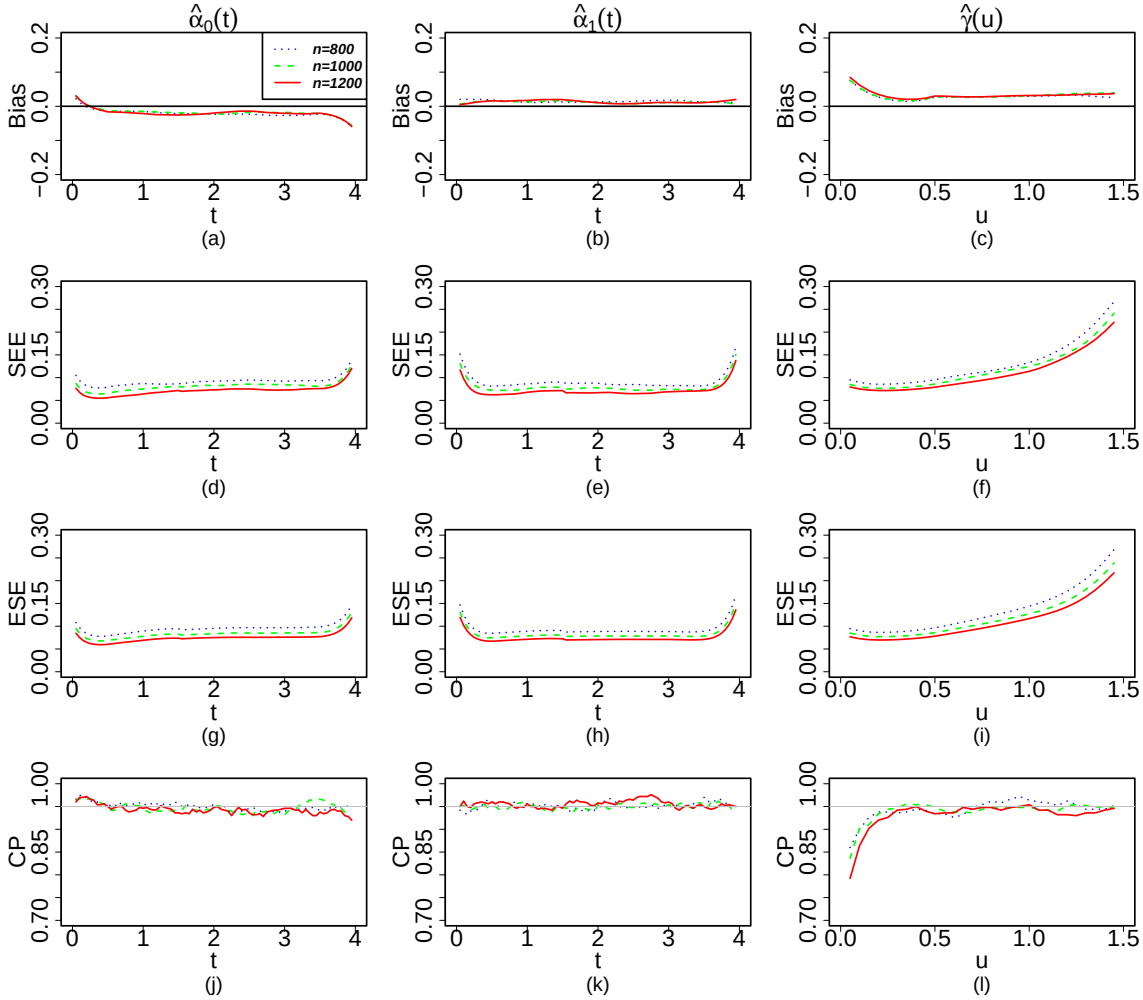

Figure 3.2: Bias, SEE, ESE and CP for estimators of $\alpha_0(t)$, $\alpha_1(t)$ and $\gamma(u)$ under model 3.15 when $\theta = 2$, using bandwidths $h = b = 0.5$. The dotted, dashed and solid lines represent sample size $n = 800$, 1000 and 1200, respectively. The results are based on 500 simulations. The estimated standard errors in each simulation are obtained via 100 weighted bootstrap samples.

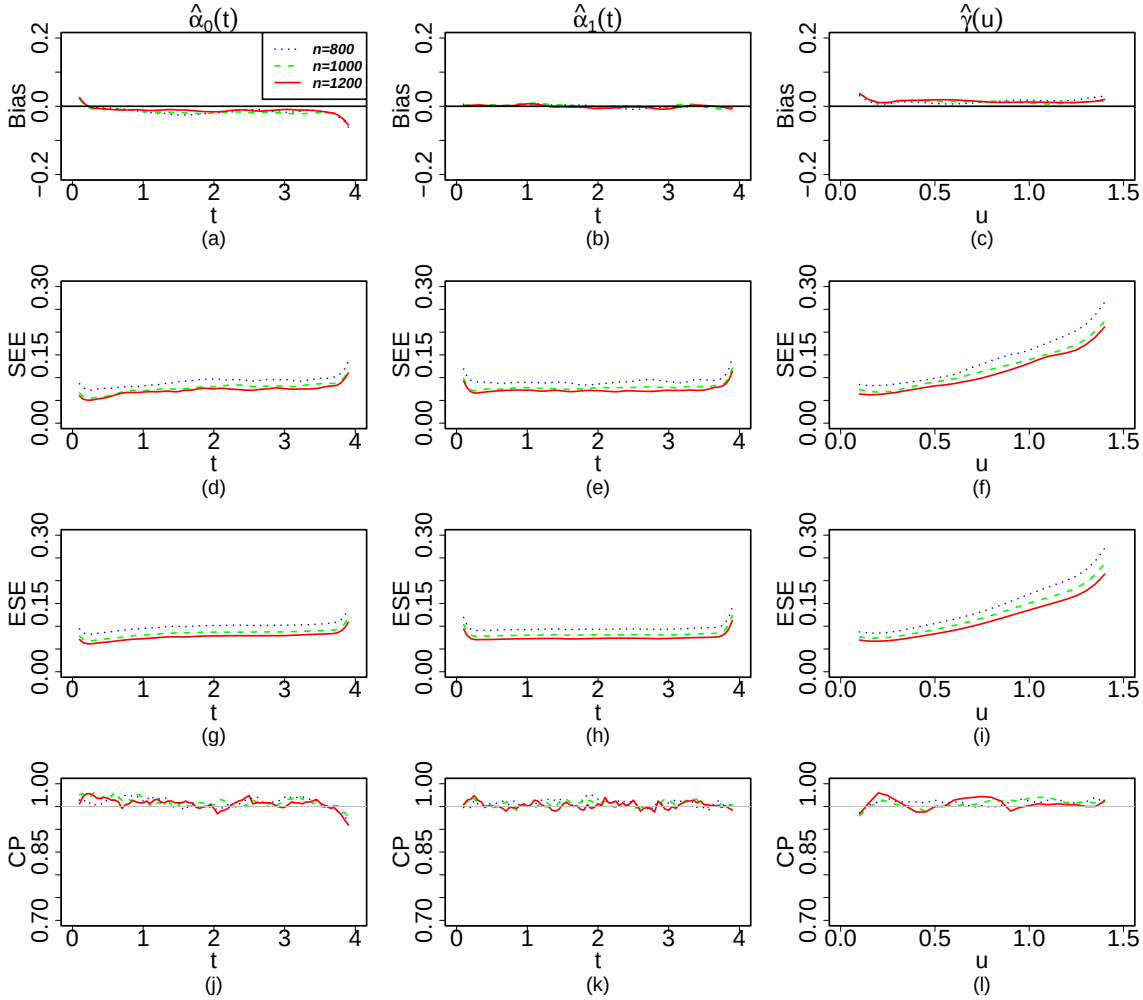

Figure 3.3: Bias, SEE, ESE and CP for estimators of $\alpha_0(t)$, $\alpha_1(t)$ and $\gamma(u)$ under model 3.15 when $\theta = 5$, using bandwidths $h = b = 0.3$. The dotted, dashed and solid lines represent sample size $n = 800$, 1000 and 1200, respectively. The results are based on 500 simulations. The estimated standard errors in each simulation are obtained via 100 weighted bootstrap samples.

We now want to compare the estimation results using different bandwidths for fixed parameter θ and sample size n. Table 3.2 presents the estimation results for θ when $\theta = 2$ and $n = 800$. Figure 3.4 displays the estimation results for $\alpha_0(t)$, $\alpha_1(t)$, and $\gamma(u)$ when $\theta = 2$ and $n = 800$. We observe that all three pairs of bandwidths perform well. The pair with larger bandwidths ($h = b = 0.5$) exhibits smaller biases, sample, and estimated standard errors for all the estimators of θ, $\alpha_0(t)$, $\alpha_1(t)$, and $\gamma(u)$. In practical applications, it is not necessary for h and b to be the same.

Table 3.2: Bias, ESE, SEE and CP for estimator of θ under model 3.15 when $\theta = 2$ and $n = 800$, using bandwidths $h = b = 0.3$, $h = b = 0.4$ and $h = b = 0.5$. The results are based on 500 simulations. The estimated standard error in each simulation is obtained via 100 weighted bootstrap samples.

$[h, b]$	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

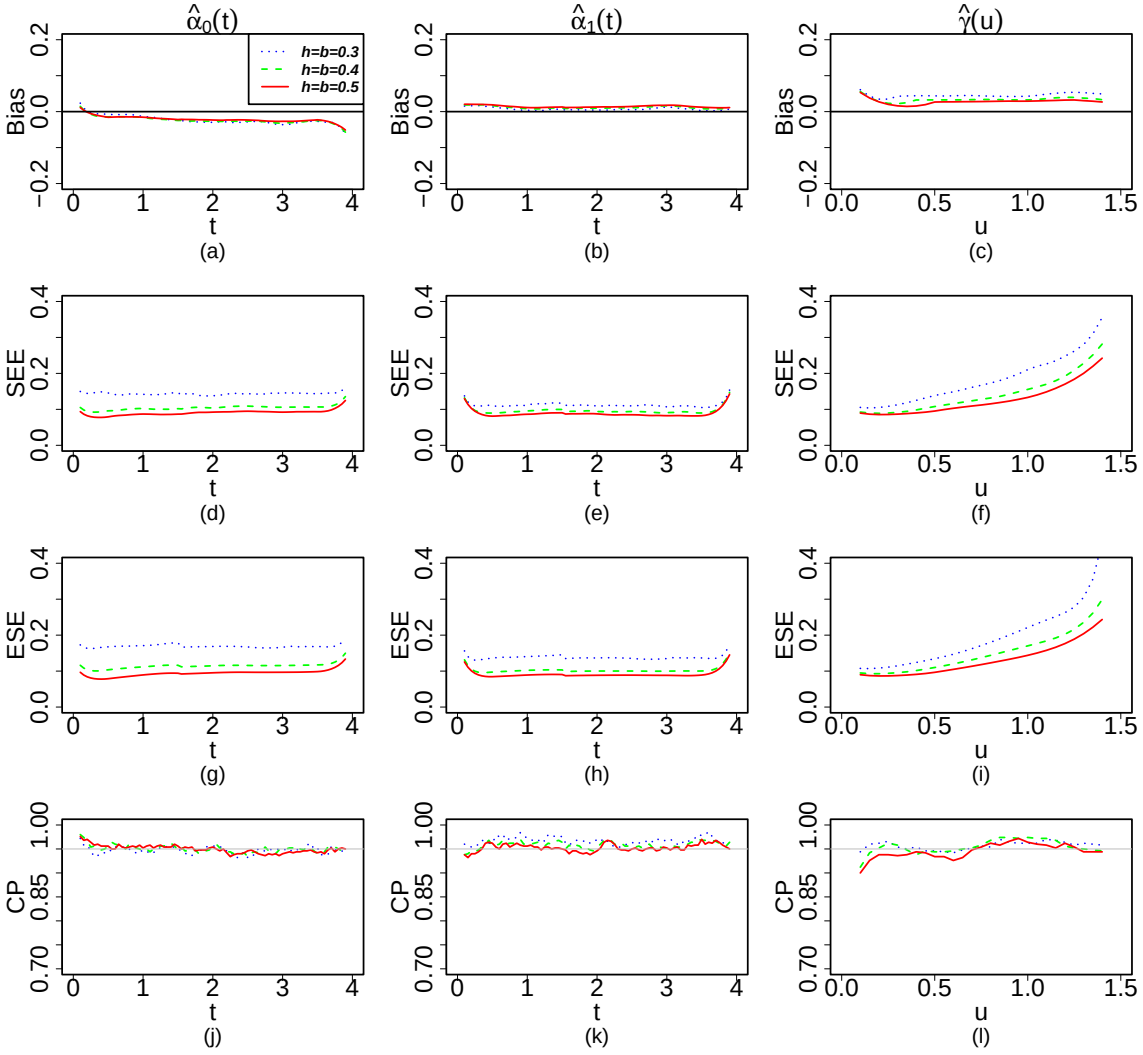

Figure 3.4: Bias, ESE, SEE and CP for estimators of $\alpha_0(t)$, $\alpha_1(t)$ and $\gamma(u)$ under model 3.15 when $\theta = 2$ and $n = 800$. The dotted, dashed and solid lines represent bandwidths $h = b = 0.3$, $h = b = 0.4$ and $h = b = 0.5$, respectively. The results are based on 500 simulations. The estimated standard errors in each simulation are obtained via 100 weighted bootstrap samples.

3.4.2 Simulations Studies Using Adaptive Algorithm

In this section, we conduct simulations using adaptive algorithm described in Section 3.2.4.

We generate data from the following intensity model

$$\lambda_i^f(t) = \xi_i \exp\{\alpha_0(t) + \gamma(U_i(t))I(N_i(t^- > 0)\}, \tag{3.16}$$

for $t \in [0, \tau]$.

- $\tau = 4$. For each subject i, we generate censoring time $C_i \sim U(3, 8)$, the study time is the minimum of C_i and τ.
- $U_i(t) = t - T_{iN_i(t^-)}$.
- $\alpha_0(t) = 2 - \log(1 + 0.2\log(1 + t))$ and $\gamma(u) = -\frac{1-u}{\exp((1-u)^2)}$.
- $\xi_i \sim \text{Gamma}(\theta, \theta)$ with $\theta = 2$.

The average number of events under these settings is approximately 17. As discussed in Section 3.2.4, when all events have been experienced, identifying $\alpha(t)$ and $\gamma(u)$ in specific local regions can be challenging. It is necessary to employ the adaptive algorithm for parameter estimations. Variance estimators are obtained from 100 repetitions of weighted bootstrap, as detailed in Section 3.3. 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 for estimator of θ under model 3.16 when $\theta = 2$ for sample size $n = 800, 1000$, and 1200, using bandwidths $h = b = 0.3$. The results are based on 500 simulations. The estimated standard error in each simulation is obtained via 100 weighted bootstrap samples.

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

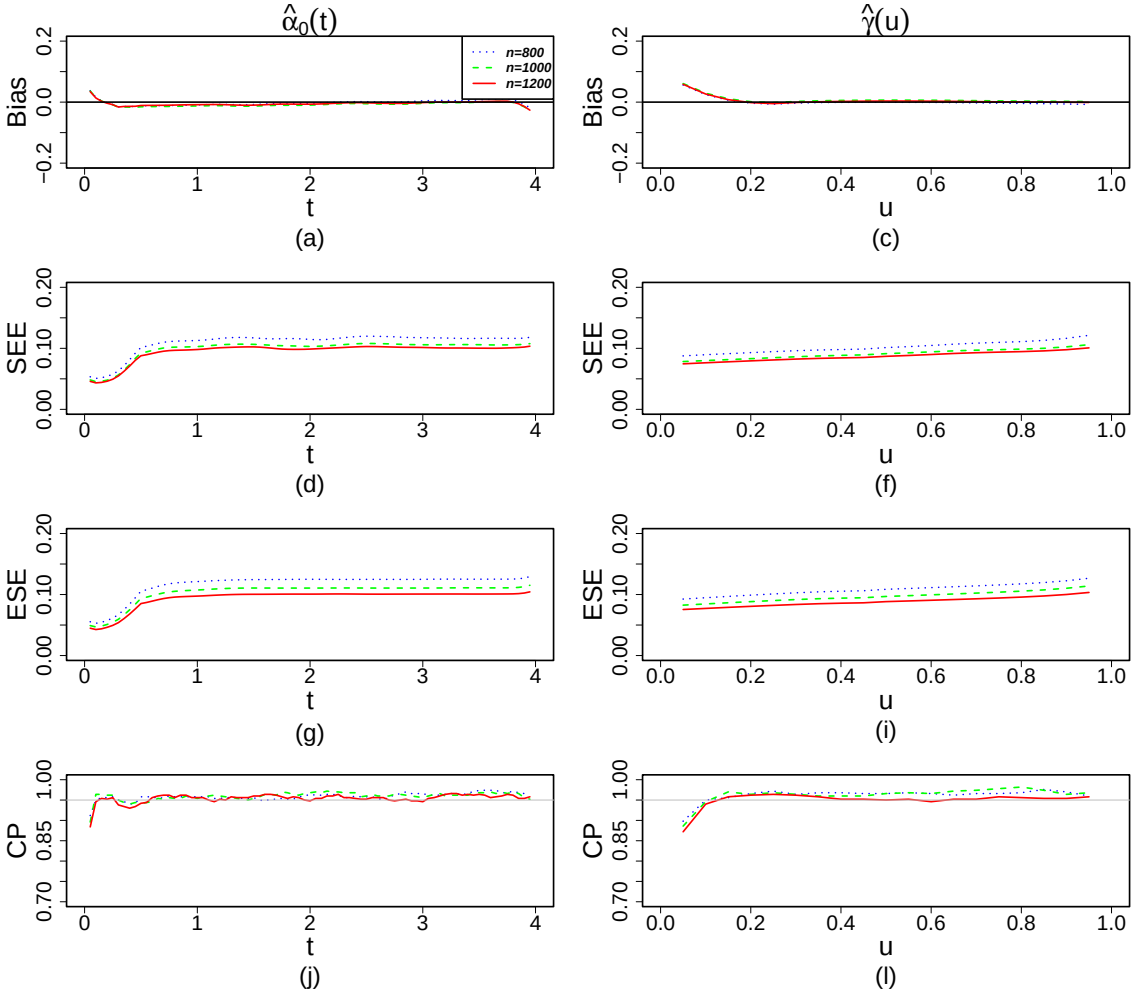

Figure 3.5: Bias, ESE, SEE and CP for the estimators of $\alpha_0(t)$ and $\gamma(u)$ under model 3.16 when $\theta = 2$, using bandwidths $h = b = 0.3$. The dotted, dashed and solid lines represent sample size $n = 800$, 1000 and 1200, respectively. The results are based on 500 simulations. The estimated standard errors in each simulation are obtained via 100 weighted bootstrap samples.

3.5 Data Application

In this section, we apply the proposed nonparametric frailty model to analyze the MAL-094 malaria vaccine trial data. Censoring is defined as follows: for each participant, visits occurring after three consecutive missed scheduled visits with no intervening unscheduled visits are considered censored. Over the 32 months follow-up time, 4633 malaria infections are observed among 1461 participants before censoring. Among these participants, 1065 of them have experienced at least one infection, with

the participant having the highest number of infections having had 34 infections.

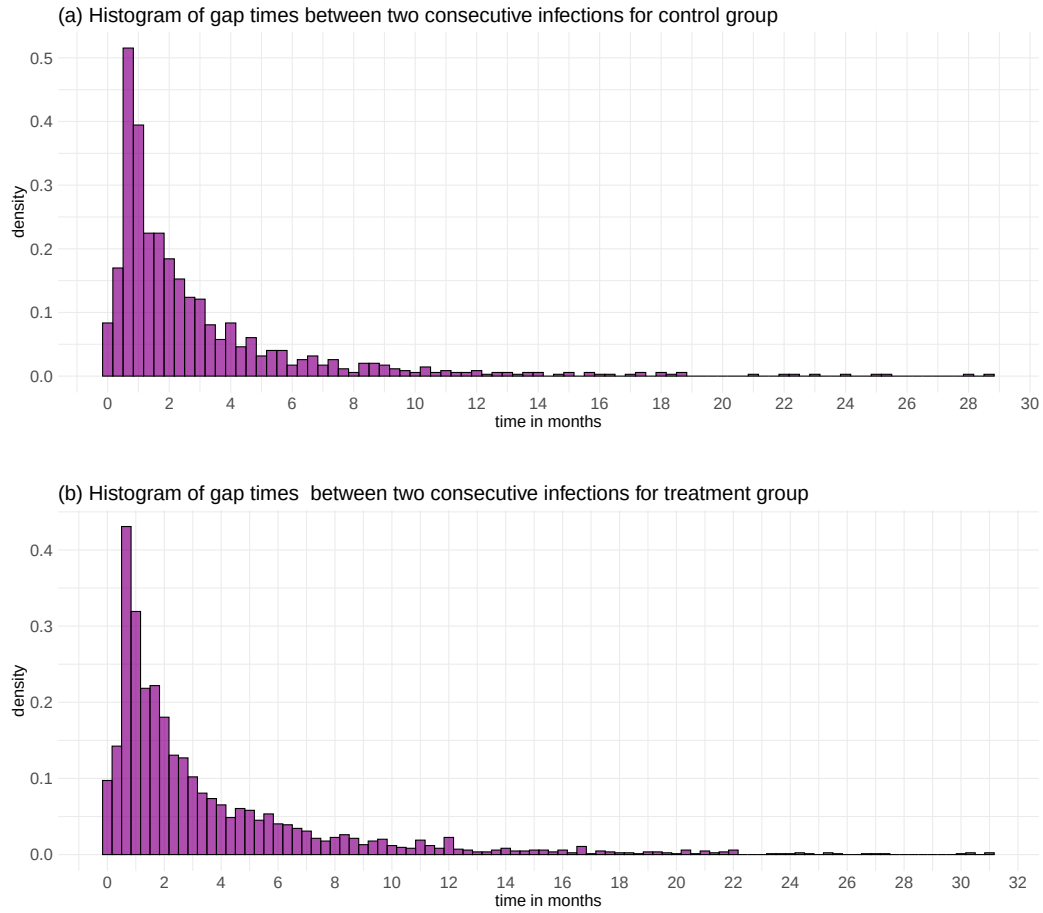

Figure 3.6: Histograms of gap times between consecutive infections for control and treatment groups in the 32 months follow-up data.

We define the observed infection times and counting process of the malaria infections. Define T_{ij} as the jth infection time we observed for subject i. Denote n_i as the total event number experienced by subject i before the end of study or censoring, whichever comes first. We have $T_{i1} < T_{i2} < ... < T_{in_i}$. The counting process $N_i(t) = \sum_{j=1}^{n_i} I(T_{ij} \leq t)$ registers the observed number of infections 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 risk of malaria infections is modeled by the intensity function $N_i(t)$. We combine the four vaccine arms into treatment group, treating the placebo arm as the

control group. The vaccine effect is then evaluated between the control group and the treatment group. We incorporate covariates such as site, age and hemoglobin level into the model.

To assess the effect of the RTS,S/AS01$_{\mathrm{E}}$ vaccine across two time scales, we derive a covariate for each participant from their event history, specifying the time since their most recent infection.

For each participant, we introduce a frailty term to represents the unobserved heterogeneity over the study period. A frailty term larger than 1 indicates that the participant is more likely to get infected, while a frailty term smaller than 1 indicates that the participant is less likely to get infected. These frailty terms capture individual differences in susceptibility to malaria infections that are not explained by observed covariates.

The intensity function of malaria infections is modeled as follows:

$$\begin{aligned}\lambda_i(t) =& \xi_i \exp\Big\{\alpha_0(t) + \alpha_1(t)\mathrm{Vacc}_i + \alpha_2(t)\mathrm{Agogo}_i + \alpha_3(t)\mathrm{Age}_i + \alpha_4(t)\mathrm{Hemo}_i \\ &+ \gamma_0(t - T_{iN_i(t^-)})I(N_i(t^-) > 0) + \gamma_1(t - T_{iN_i(t^-)})I(N_i(t^-) > 0)\mathrm{Vacc}_i\Big\},\end{aligned} \tag{3.17}$$

for $0 \leq t \leq 32$(months).

In model 3.17, $\alpha_0(t)$, $\alpha_1(t)$, $\alpha_2(t)$, $\alpha_3(t)$, $\alpha_4(t)$, $\gamma_0(u)$ and $\gamma_1(u)$ are unspecified functions. The variable ξ_i follows Gamma(θ, θ) 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). Agogo$_i$ is the study site indicator (1= Agogo, 0 = Siaya) and Age$_i$ is the age in years at enrollment. Hemo$_i$ is the standardized hemoglobin level. The model is fitted on $u \in [0, 8.33]$, where is the 90th percentile of the gap times between consecutive infections.

Figure 3.7 presents the estimation results of the nonparametric parameters. We

observe a slight increasing trend in the baseline risk over time. The risk in the treatment group is lower than that in the control group. Participants living in Agogo exhibit a lower risk compared to those in Siaya. Additionally, older children tend to have a higher infection risk. The ones with higher hemoglobin level tends to experience lower risk.

Furthermore, following prior infections, the risk of subsequent infections increases as indicated by $\hat{\gamma}_0(u) > 0$. However, the increment in risk for participants in the treatment group is lower than that for participants in the control group, demonstrated by $\hat{\gamma}_1(u) < 0$. In conclusion, participants in the treatment group show a lower infection risk both for initial infections and subsequent reinfections compared to the control group.

The estimate of θ is 3.128, with an estimated standard error obtained from 100 weighted bootstrap samples of 0.556. Test whether the variance of the frailty term equals 0, i.e. $H_0 : \frac{1}{\theta} = 0$. The p value is less than 10^{-3}, suggests there are meaningful associations among participants.

For each participant, we have the conditional expectation $E(\xi_i)$ at convergence. This allow us to incorporate heterogeneity when predicting the infection intensity for a particular individual. With all the covariates considered, $E(\xi_i) > 1$ indicates that the individual is more susceptible to malaria infections, while $E(\xi_i) < 1$ indicates that the individual is relatively less likely to get infected.

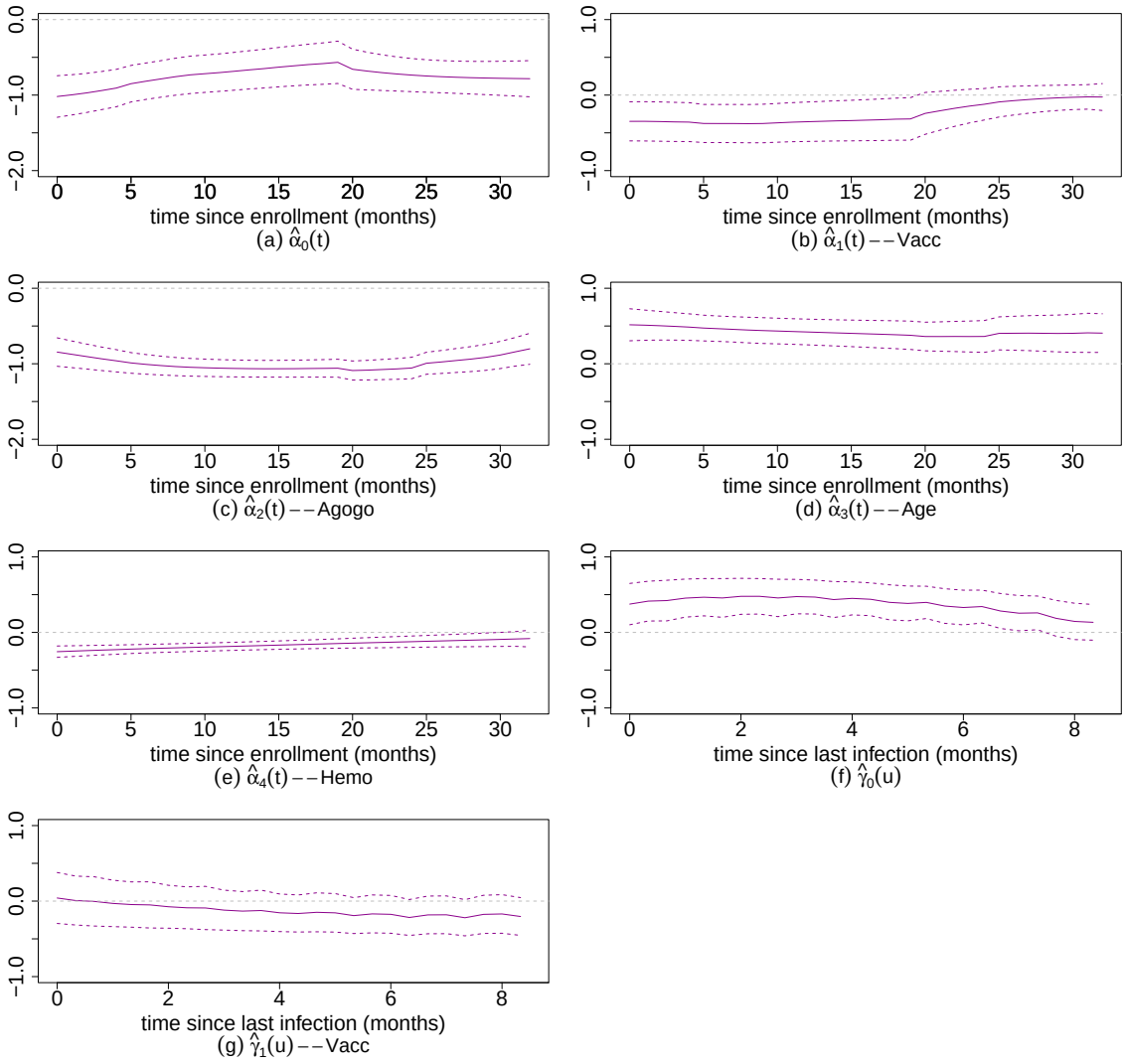

Figure 3.7: Estimation results of time-varying effects of covariates under model 3.17 using $h = 32$ months and $b = 8.33$ months. The solid lines represent the point estimates, while the dashed lines signify the 95% pointwise confidence intervals. The estimated standard errors are obtained via 100 weighted bootstrap samples.

3.6 Summary

In this chapter, we proposed a nonparametric intensity model with frailty for recurrent event data. The frailty follows a Gamma distribution with an unknown parameter. It acts multiplicatively on the intensity and measures the unobserved heterogeneity among individuals. We developed a maximum likelihood estimation procedure to estimate the effects of covariates in two time-scales using local linear

approximation method with double kernels. In E step, we estimated the posterior expectation of the Gamma frailty. In M step, we maximized the conditional complete log-likelihood using double kernels and aggregations. Variance estimators were obtained using a weighted bootstrap procedure. Simulation studies demonstrated the satisfactory performance of the proposed methods.

We applied the methods to the MAL-094 malaria vaccine trial data. Our analysis revealed that site, age and hemoglobin level have influence on the risk of malaria infections. Specifically, participants in Agogo exhibited lower infection risk compared to those in Siaya. Older children were observed to have a relatively higher risk of infections. Furthermore, the ones with higher hemoglobin levels are less likely to become infected.

Figure 3.7 (b) illustrated the risk of the first infections among participants in the treatment group was lower than that of control group. Additionally, Figure 3.7 (b) and (g) showed that the risk of reinfections was further decreased in the treatment group.

CHAPTER 4: CONCLUSION

In conclusion, we investigated two dynamic intensity models for recurrent event data. These proposed models provide frameworks for studying how event intensity evolves over time and how the occurrence of a prior event influences the likelihood of a future event. Extensive simulations demonstrated their validities.

In Chapter 2, we developed a generalized class of semiparametric intensity models. Through the choice of link function, the proposed models encompass a wide range of models such as the multiplicative intensity model and the additive intensity model. Maximum likelihood estimation procedure was investigated through local linear approximation and profile likelihood method. Asymptotic properties of the estimators have been derived. We developed hypothesis testing procedures to test the parametric forms of the covariate-varying effects. Additionally, we derived a Gaussian multiplier method to approximate the critical values of the test statistics.

In Chapter 3, we proposed a nonparametric frailty intensity model, which can measure the time-varying and covariate-varying effects after taking the unobserved heterogeneity into consideration. Maximum likelihood Estimation procedure was provided using local linear approximation method with double kernels. Maximization was achieved through an EM algorithm. Variance estimators were obtained using a weighted bootstrap procedure.

Both of the models were applied to the MAL-094 malaria vaccine trial data. It was found that site, hemoglobin and age have influence on the risk of malaria infections. Following prior infections, the risk of subsequent infections increased, but the increment in risk for participants in the treatment group was smaller than those in the control group. In conclusion, we observe the vaccine effects against both first

infections and reinfections.

The applications have demonstrated that the proposed models have effectively allowed us to tackle the inquiries within the MAL-094 malaria vaccine trial data. Additionally, these models could have potential utility in other medical studies involving recurrent events, especially in situations requiring the modeling of treatment effects across two time scales.

In the future, I plan to develop R packages for both models. Furthermore, I intend to integrate the frailty framework into the semiparametric models we discussed in Chapter 2 and establish hypothesis testing procedures to examine the Gamma frailty.

REFERENCES

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.

Heng, F. (2019). *Dynamic Modeling of Incomplete Event Histogy Data*. Ph.D. thesis, University of North Carolina at Chatlotte, Charlotte.

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

Lawless, J. F., Nadeau, C., and Cook, 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.

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.

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.

Silverman, B. (1986). Density estimation for statistics and data analysis, publisher chapman and hall. *London, UK*.

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 1 and Theorem 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 covariate processes $X_i(t)$, $Z_i(t)$, $U_i(t)$ and the intensity function $\lambda_i(t)$, $0 \leq t \leq \tau$, 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. When bandwidth $h \to 0$; $nh^2 \to \infty$ 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.

(7) 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 2.8,

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

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

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

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

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

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

$$\hat{\eta}\xrightarrow{\mathcal{P}}\eta_0.$$

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

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

Consider

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

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

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

So we have

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

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

$$\Sigma_\eta=E\Big[\int_{t_1}^{t_2}\frac{\dot{\lambda}_i(t,\eta_0)}{\lambda_i(t,\eta_0)}\Big\{\Big(\frac{\partial\zeta(U_i(t),\eta_0)}{\partial\eta}\Big)^{\mathsf{T}}P_i(t)-e_{12}(t)^{\mathsf{T}}(e_{11}(t))^{-1}X_i(t)\Big\}dM_i(t)\Big]^{\otimes2}.$$

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

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

The covariance matrix Σ_η can be estimated by

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

A_η can be estimated by

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

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 $\frac{\partial\tilde{\alpha}(t,\eta_m)}{\partial\eta}\xrightarrow{\mathcal{P}}-e_{11}^{-1}(t)e_{12}(t)$. So combining equation A.2 with equation A.5 and A.6, we can continue writing the first

two terms of equation A.5 as

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

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,

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

By Taylor expansion, we have

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

where ${\alpha_0}^*(t)=(\alpha_0(t),\dot{\alpha}_0(t))^{\mathsf{T}}$. 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

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

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

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

where $\Sigma_\alpha = e_{11}(t)\Sigma_e(t)(e_{11}(t))^{-1}$, with

$$\Sigma_e(t) = \lim_{n\to\infty} hE\Big\{\int_0^\tau K_h^2(s-t)\frac{\dot{\lambda}^2(s)}{\lambda(s)}[X_i(s)]^{\otimes 2}ds\Big\}.$$

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

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