SCIENCE CHINA
Mathematics

• ARTICLES •

December 2016 Vol. 59 No. 12: 2393–2410
doi: 10.1007/s11425-016-0112-6

Time-varying latent model for longitudinal data with informative observation and terminal event times

In memory of Professor Xiru Chen (1934–2005)

PEI YanBo[1], DU Ting[2] & SUN LiuQuan[2,*]

[1]*School of Statistics, Capital University of Economics and Business, Beijing* 100070*, China;*
[2]*Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing* 100190*, China*
Email: peiyanbo@cueb.edu.cn, duting@amss.ac.cn, slq@amt.ac.cn

Received February 19, 2016; accepted October 10, 2016; published online October 31, 2016

Abstract Longitudinal data often occur in follow-up studies, and in many situations, there may exist informative observation times and a dependent terminal event such as death that stops the follow-up. We propose a semiparametric mixed effect model with time-varying latent effects in the analysis of longitudinal data with informative observation times and a dependent terminal event. Estimating equation approaches are developed for parameter estimation, and asymptotic properties of the resulting estimators are established. The finite sample behavior of the proposed estimators is evaluated through simulation studies, and an application to a bladder cancer study is provided.

Keywords estimating equations, informative observation times, joint modeling, longitudinal data, terminal event, time-varying effect

MSC(2010) 62N02, 62G20

Citation: Pei Y B, Du T, Sun L Q. Time-varying latent model for longitudinal data with informative observation and terminal event times. Sci China Math, 2016, 59: 2393–2410, doi: 10.1007/s11425-016-0112-6

1 Introduction

Longitudinal data frequently occur in clinical trials, epidemiological studies and observational investigations. Various methods have been developed for analyzing these data [2, 4, 6, 10, 26, 28]. For example, Diggle et al. [3] gave a summary of commonly used methods including random effect model and estimating equation approaches. Lin and Ying [13] and Welsh et al. [25] provided some semiparametric and nonparametric regression analyses of longitudinal data. However, all these methods focus only on the situation where the longitudinal response variable and the observation times are independent given covariates.

In recent years, a number of authors have considered the situation where the longitudinal response variable is correlated with the observation times, i.e., the observation times are informative about the longitudinal responses [5, 12, 14, 19–22, 30, 31]. For example, Sun et al. [22] suggested a joint model for the longitudinal process and the observation times through a shared latent variable. Liang et al. [12] and Zhou et al. [31] considered joint modeling of longitudinal data via latent variables. All these methods are designed for the analysis of longitudinal data with informative observation times in the absence of a dependent terminal event.

*Corresponding author

In practice, there may exist a terminal event such as death that stops the follow-up, and the terminal event is usually strongly correlated with both the longitudinal responses and the observation times. For example, patients in a severe disease stage often die in a shorter period, and the longitudinal medical costs may be related to both hospital visits and death [15]. Liu et al. [16] and Liu et al. [17] indicated that ignoring the dependent terminal event would lead to biased estimates for the analysis of longitudinal medical costs. Thus, there is clearly a need to develop suitable models for longitudinal data, which account for informative observation times and a dependent terminal event. Note that parametric models are sensitive to the assumed distribution, and misspecifying the parametric models often leads to erroneous inference. Nonparametric models are subject to curse of dimensionality, and it may be difficult to obtain reliable estimates for moderate or small data sets. Thus, semiparametric models are good compromises and retain nice features of both the parametric and nonparametric models. However, there exists some limited research on the semiparametric regression analysis of these data. For example, Liu et al. [15] studied a joint random effects model, where the distributions of the random effects are specified. He et al. [8] proposed some shared frailty models, where one random effect is assumed to be normally distributed. Sun et al. [23] suggested a joint modeling approach via two latent variables, where the dependence structure between two latent variables are left unspecified. The aforementioned methods assume that the effects of latent variables on the longitudinal responses are time-invariant.

In many applications, the effects of latent variables on the longitudinal responses may vary over time. One motivating example is the bladder cancer study [1,12]. Cai et al. [1] showed that the latent variable effect is truly time-dependent. In reality, it is important to know the temporal effects of latent variables, and the semiparametric model with time-varying latent effects provides a nice graphical summary of time dynamics of latent variables. For example, in the bladder cancer data, Cai et al. [1] showed that the time-varying latent effect is negative for small t, but its magnitude diminishes and eventually it becomes positive as t increases. Recently, Cai et al. [1] considered a semiparametric model with time-varying latent effects in the analysis of longitudinal data with informative observation times, but it was not designed to handle the presence of a terminal event. How to characterize the time-dependent behavior of latent variables shared by the longitudinal responses, the informative observation times and the terminal event is of main interest here. In this paper, we propose a semiparametric mixed effect model with time-varying latent effects in the analysis of longitudinal data with informative observation times and a dependent terminal event. The new model offers great flexibility in formulating the effects of latent variables on the longitudinal response variable while adjusting its association with the observation times and the terminal event. In addition, based on the estimates of the time-varying latent effects, the proposed model can summarize and explain the time-varying latent effects more clearly, and further allow for inference about the effects of latent variables.

The rest of the article is organized as follows. In Section 2, we introduce joint modeling of the longitudinal response, the observation times and the terminal event. Section 3 presents estimation procedures for regression parameters of interest, and the asymptotic properties of the proposed estimators are established. Some simulation results to evaluate the proposed methods are reported in Section 4. An application to a bladder tumor study is provided in Section 5. Some concluding remarks are given in Section 6. Proof of Theorem 3.1 is given in Appendix.

2 Model specifications

Consider a longitudinal study with n independent subjects. For subject $i, i = 1, \ldots, n$, let $Y_i(t)$ denote the longitudinal response variable of interest at time t. Also let $X_i(t)$ and $Z_i(t)$ be the $p \times 1$ and $q \times 1$ vector of possibly time-dependent covariates, respectively. In addition, let D_i be the terminal event time and C_i be the censoring time. Define $T_i = C_i \wedge D_i$, $\delta_i = I(D_i \leqslant C_i)$ and $\Delta_i(t) = I(T_i \geqslant t)$, where $a \wedge b = \min(a, b)$, and $I(\cdot)$ is the indicator function. Let $N_i(t)$ be the counting process denoting the number of the observation times before or at time t. The longitudinal variable $Y_i(t)$ is observed only at the time points where $N_i(t)$ jumps for $t \leqslant T_i$.

We consider the following semiparametric mixed-effect model for the longitudinal response variable:

$$Y_i(t) = \mu_0(t) + \beta_0' X_i(t) + u_i \alpha_0(t)' Z_i(t) + \varepsilon_i(t), \tag{2.1}$$

where $\mu_0(t)$ is an unspecified functions of t, β_0 is a $p \times 1$ vector of unknown regression parameters, u_i is a subject-specific random effect, $\alpha_0(t)$ is a $q \times 1$ vector of time-varying regression coefficients, and $\varepsilon_i(t)$ is a zero mean measurement error process. Here $u_i\alpha_0(t)$ denotes the time-varying latent effects for $Z_i(t)$.

Let W_i be a vector of r-dimensional time-independent covariates, and ν_i be a gamma random variable with mean 1 and variance σ_0^2. For the observation process, it will be assumed that conditional on W_i and ν_i, $N_i(t)$ is a nonstationary Poisson process with the intensity function

$$E\{dN_i(t) \mid W_i, \nu_i\} = \nu_i \exp\{\gamma_0' W_i\} d\Lambda_0(t), \tag{2.2}$$

where γ_0 is a vector of unknown regression parameters, and $\Lambda_0(t)$ is an unspecified baseline cumulative intensity function.

Following Huang and Wang [9], we specify the proportional hazards frailty model for the terminal event time D_i as

$$d\Lambda^D(t \mid W_i, \nu_i) = \nu_i \exp\{\eta_0' W_i\} d\Lambda_0^D(t), \tag{2.3}$$

where η_0 is a vector of unknown regression parameters, and $\Lambda_0^D(t)$ is an unspecified baseline cumulative hazard function.

Remark 1. Model (2.1) is different from Liang et al. [12] in that $Z_i(t)$ can be included in $X_i(t)$ or not, and the effects of latent variables are time-varying. For example, in the bladder cancer data studied in Section 5, $X_i(t) = (X_{i1}, X_{i2})'$ and $Z_i(t) = X_{i1}$, where X_{i1} is the treatment variable, and X_{i2} is the logarithm of the number of initial tumors. In addition, when $Z_i(t) \equiv 1$, Model (2.1) reduces to the time-varying latent effect model studied by Cai et al. [1] in the absence of a terminal event. In practice, in order to decide which covariates to have the time-varying latent effects, we can use the focused information criterion to choose $Z_i(t)$ as discussed in [31]. Here we only consider a frailty model with time-independent covariates in (2.2) and (2.3). However, the proposed estimation procedure can be extended in a straightforward manner to deal with time-dependent covariates in (2.2) and (2.3) at the cost of having more complicated formulas.

The following assumptions are required for making inference:

(i) Given the frailty ν_i and the covariates $\{X_i(\cdot), Z_i(\cdot), W_i\}$, $Y_i(\cdot)$, $N_i(\cdot)$ and D_i are mutually independent.

(ii) The censoring time C_i is noninformative in the sense that given the covariates $\{X_i(\cdot), Z_i(\cdot), W_i\}$, C_i is independent of $Y_i(\cdot)$, $N_i(\cdot)$ and D_i.

(iii) The two random effects are assumed to satisfy $E(u_i \mid \nu_i) = \nu_i - 1$.

Remark 2. There are two purposes for Condition (iii): one is to characterize the association among the longitudinal response, the observation times and the terminal event via the two latent variables u_i and ν_i; and the other is for identifiability of (2.1). Note that $\alpha_0(t)$ is unspecified. Thus, in order to ensure identifiability of model parameters, it must be assumed that $E(u_i)$ or $E(u_i \mid \nu_i)$ is fixed. As in [12], we assume Condition (iii) for simplicity and computational convenience. In fact, the proposed method can be extended to the case that $E(u_i \mid v_i) = g(v_i)$, where $g(v_i)$ is a q-dimensional vector with each component being a specified function.

3 Estimation procedures

Note that given W_i and ν_i, the observation process is a nonhomogeneous Poisson process. Let m_i denote the total number of observations for subject i before T_i. It follows that given $\{\nu_i, W_i, T_i\}$, m_i has a Poisson distribution with mean $\nu_i \Lambda_0(T_i) \exp\{\gamma_0' W_i\}$. In what follows, when there is no ambiguity, we will suppress $\{X_i(t), Z_i(t), W_i\}$ in the conditional expectation. Similar to [12], we have

$$E\{\Delta_i(t) dN_i(t) \mid \nu_i, m_i, T_i\} = \Delta_i(t) m_i \frac{d\Lambda_0(t)}{\Lambda_0(T_i)}.$$

Let $\mathcal{A}_{10}(t) = \int_0^t \mu_0(s)d\Lambda_0(s)$, $\mathcal{A}_{20}(t) = \int_0^t \alpha_0(s)d\Lambda_0(s)$, and $B_i(t) = Z_i(t)E\{(\nu_i - 1) \mid m_i, T_i, \delta_i\}$. Then under Model (2.1) and Assumptions (i)–(iii), we have

$$\begin{aligned}
&E[\{Y_i(t) - \beta_0' X_i(t)\}\Delta_i(t)dN_i(t) \mid m_i, T_i, \delta_i] \\
&\quad = E(E[\{Y_i(t) - \beta_0' X_i(t)\}\Delta_i(t)dN_i(t) \mid \nu_i, m_i, T_i, \delta_i] \mid m_i, T_i, \delta_i) \\
&\quad = E(E[\{Y_i(t) - \beta_0' X_i(t)\} \mid \nu_i]E[\Delta_i(t)dN_i(t) \mid \nu_i, m_i, T_i] \mid m_i, T_i, \delta_i) \\
&\quad = E\left[\{\mu_0(t) + (\nu_i - 1)\alpha_0(t)' Z_i(t)\}\Delta_i(t) m_i \frac{d\Lambda_0(t)}{\Lambda_0(T_i)} \,\middle|\, m_i, T_i, \delta_i\right] \\
&\quad = \Delta_i(t) m_i \frac{d\mathcal{A}_{10}(s)}{\Lambda_0(T_i)} + \Delta_i(t) m_i B_i(t)' \frac{d\mathcal{A}_{20}(t)}{\Lambda_0(T_i)}.
\end{aligned}$$

Define

$$M_i(t; \beta, \mathcal{A}_1, \mathcal{A}_2, \Lambda) = \int_0^t \left[\{Y_i(s) - \beta' X_i(s)\}\Delta_i(s)dN_i(s) - \Delta_i(s) m_i \frac{d\mathcal{A}_1(s)}{\Lambda(T_i)} - \Delta_i(s) m_i B_i(s)' \frac{d\mathcal{A}_2(s)}{\Lambda(T_i)}\right].$$

Then it follows that $E\{M_i(t; \beta_0, \mathcal{A}_{10}, \mathcal{A}_{20}, \Lambda_0) \mid T_i, m_i, \delta_i\} = 0$, which implies that $M_i(t; \beta_0, \mathcal{A}_{10}, \mathcal{A}_{20}, \Lambda_0)$ are zero-mean stochastic processes. Thus, for given Λ and B_i, by applying the generalized estimating equation approach [11], we can use the following estimating equations to estimate $\mathcal{A}_{10}(t)$, $\mathcal{A}_{20}(t)$ and β_0:

$$\sum_{i=1}^n \begin{pmatrix} 1 \\ B_i(t) \end{pmatrix} dM_i(t; \beta, \mathcal{A}_1, \mathcal{A}_2, \Lambda) = 0, \quad 0 \leqslant t \leqslant \tau, \tag{3.1}$$

$$\sum_{i=1}^n \int_0^\tau X_i(t) dM_i(t; \beta, \mathcal{A}_1, \mathcal{A}_2, \Lambda) = 0, \tag{3.2}$$

where τ is a prespecified constant such that $P(T_i \geqslant \tau) > 0$. Here we do not estimate $\mu_0(t)$ and $\alpha_0(t)$ directly, we just consider the estimates of $\mathcal{A}_{10}(t)$ and $\mathcal{A}_{20}(t)$ without the kernel smoothing technique. Define

$$R_1(t; \Lambda, B) = \begin{pmatrix} \sum_{i=1}^n \Delta_i(t) m_i \Lambda(T_i)^{-1} & \sum_{i=1}^n \Delta_i(t) m_i \Lambda(T_i)^{-1} B_i(t)' \\ \sum_{i=1}^n \Delta_i(t) m_i B_i(t) \Lambda(T_i)^{-1} & \sum_{i=1}^n \Delta_i(t) m_i \Lambda(T_i)^{-1} B_i(t) B_i(t)' \end{pmatrix}$$

and

$$R_2(t; \Lambda, B) = \left(\sum_{i=1}^n \Delta_i(t) m_i \Lambda(T_i)^{-1} X_i(t), \sum_{i=1}^n \Delta_i(t) m_i \Lambda(T_i)^{-1} X_i(t) B_i(t)'\right),$$

where $B = (B_1, \ldots, B_n)'$. Let $\hat{\mathcal{A}}_1(t; \beta, \Lambda, B)$ and $\hat{\mathcal{A}}_2(t; \beta, \Lambda, B)$ denote the solutions to the estimating equation (3.1), which have the following closed forms:

$$\begin{pmatrix} \hat{\mathcal{A}}_1(t; \beta, \Lambda, B) \\ \hat{\mathcal{A}}_2(t; \beta, \Lambda, B) \end{pmatrix} = \int_0^t R_1(s; \Lambda, B)^{-1} \begin{pmatrix} \sum_{i=1}^n \{Y_i(s) - \beta' X_i(s)\}\Delta_i(s) dN_i(s) \\ \sum_{i=1}^n \{Y_i(s) - \beta' X_i(s)\} B_i(s) \Delta_i(s) dN_i(s) \end{pmatrix}.$$

Plugging $\hat{\mathcal{A}}_1(t; \beta, \Lambda, B)$ and $\hat{\mathcal{A}}_2(t; \beta, \Lambda, B)$ into the estimating equation (3.2), we obtain

$$U^*(\beta; \Lambda, B) = \sum_{i=1}^n \int_0^\tau \left[X_i(t) - R_2(t; \Lambda, B) R_1(t; \Lambda, B)^{-1} \begin{pmatrix} 1 \\ B_i(t) \end{pmatrix}\right] \{Y_i(t) - \beta' X_i(t)\}\Delta_i(t) dN_i(t).$$

Note that for any function $g(\cdot)$,

$$\frac{1}{n} \sum_{i=1}^n \int_0^\tau \left[X_i(t) - R_2(t; \Lambda_0, B) R_1(t; \Lambda_0, B)^{-1} \begin{pmatrix} 1 \\ B_i(t) \end{pmatrix}\right] g(t) \Delta_i(t) dN_i(t)$$

converges to zero in probability. Then the estimating function $U^*(\beta; \Lambda, B)$ can be extended to the following estimating function:

$$U_g^*(\beta; \Lambda, B) = \sum_{i=1}^n \int_0^\tau \left[X_i(t) - R_2(t; \Lambda, B) R_1(t; \Lambda, B)^{-1} \begin{pmatrix} 1 \\ B_i(t) \end{pmatrix}\right] \{Y_i(t) - \beta' X_i(t) - g(t)\}\Delta_i(t) dN_i(t).$$

Following [13], in order to reduce the variance of $U_g^*(\beta;\Lambda,B)$, $g(t)$ can be chosen as $\bar{Y}^*(t;\Lambda)-\beta'\bar{X}(t;\Lambda)$, where

$$\bar{X}(t;\Lambda)=\frac{\sum_{i=1}^n \Delta_i(t)m_i\Lambda(T_i)^{-1}X_i(t)}{\sum_{i=1}^n \Delta_i(t)m_i\Lambda(T_i)^{-1}},$$
$$\bar{Y}^*(t;\Lambda)=\frac{\sum_{i=1}^n \Delta_i(t)m_i\Lambda(T_i)^{-1}Y_i^*(t)}{\sum_{i=1}^n \Delta_i(t)m_i\Lambda(T_i)^{-1}}$$

and $Y_i^*(t)$ is the measurement of $Y_i(\cdot)$ at the time point nearest to t. Thus, for given Λ and B_i, we specify the following estimating function for β_0:

$$U(\beta;\Lambda,B)=\sum_{i=1}^n\int_0^\tau\left[X_i(t)-R_2(t;\Lambda,B)R_1(t;\Lambda,B)^{-1}\begin{pmatrix}1\\B_i(t)\end{pmatrix}\right]$$
$$\times\{Y_i(t)-\bar{Y}^*(t;\Lambda)-\beta'\{X_i(t)-\bar{X}(t;\Lambda)\}\}\Delta_i(t)dN_i(t). \tag{3.3}$$

In practice, however, Λ_0 and B_i are unknown, and we cannot directly use the estimating function (3.3). For this, consider (2.2) and (2.3), and it can be checked that

$$\tilde{M}_i^D(t)=N_i^D(t)-\int_0^t\Delta_i(s)H_{1i}(s;\sigma_0^2,\eta_0,\Lambda_0^D)\exp(\eta_0'W_i)d\Lambda_0^D(s) \tag{3.4}$$

and

$$\tilde{M}_i(t)=\int_0^t\Delta_i(s)dN_i(s)-\int_0^t\Delta_i(s)H_{2i}(s;\sigma_0^2,\eta_0,\Lambda_0^D)\exp(\gamma_0'W_i)d\Lambda_0(t) \tag{3.5}$$

are zero-mean stochastic processes, where $H_{1i}(t;\eta,\sigma^2,\Lambda^D)=E\{\nu_i\mid D_i\geqslant t\}=\{1+\sigma^2\exp(\eta'W_i)\Lambda^D(t)\}^{-1}$ and $H_{2i}(t;\eta,\sigma^2,\Lambda^D)=E\{\nu_i\mid T_i,\delta_i\}=(1+\sigma^2\delta_i)\{1+\sigma^2\exp(\eta'W_i)\Lambda^D(t)\}^{-1}$. Thus, for given H_{1i} and H_{2i}, using the generalized estimating equation approach [11], we can estimate η_0, γ_0, $\Lambda_0^D(t)$ and $\Lambda_0(t)$. However, the weight functions H_{1i} and H_{2i} also include unknown parameter σ^2, which must be estimated. In order to estimate σ^2, we can use the observed likelihood, which is given by Ye et al. [27],

$$L(\sigma^2;\eta_0,\gamma_0,\Lambda_0^D,\Lambda_0)\propto\prod_{i=1}^n\frac{\Gamma(m_i+\delta_i+1/\sigma^2)}{\Gamma(1/\sigma^2)(\sigma^2)^{1/\sigma^2}(d_i+1/\sigma^2)^{(m_i+\delta_i+1/\sigma^2)}}, \tag{3.6}$$

where $d_i=\int_0^\infty\Delta_i(s)[\exp(\eta_0'W)d\Lambda_0^D(s)+\exp(\gamma_0'W_i)d\Lambda_0(s)]$. Differentiating the logarithm of $L(\sigma^2;\eta_0,\gamma_0,\Lambda_0^D,\Lambda_0)$ with respect to σ^2 gives the estimating equation for σ^2.

Let $\theta=(\eta',\gamma',\sigma^2,\Lambda_0^D,\Lambda_0)'$. For $k=0$ and 1, define

$$S_1^{(k)}(t;\eta,\sigma^2,\Lambda^D)=\frac{1}{n}\sum_{i=1}^n\Delta_i(t)H_{1i}(t;\eta,\sigma^2,\Lambda^D)W_i^{\otimes k}\exp(\eta'W_i),$$
$$S_2^{(k)}(t;\eta,\gamma,\sigma^2,\Lambda^D)=\frac{1}{n}\sum_{i=1}^n\Delta_i(t)H_{2i}(t;\eta,\sigma^2,\Lambda^D)W_i^{\otimes k}\exp(\gamma'W_i),$$

where $a^{\otimes 0}=1$ and $a^{\otimes 1}=a$ for any vector a. In view of (3.4)–(3.6), we propose to estimate θ using the solutions to the equations $\tilde{U}(\theta)=(\tilde{U}_1',\tilde{U}_2',\tilde{U}_3,\tilde{U}_4,\tilde{U}_5)'=0$, where

$$\tilde{U}_1=\sum_{i=1}^n\int_0^\tau\left\{W_i-\frac{S_1^{(1)}(t;\eta,\sigma^2,\Lambda^D)}{S_1^{(0)}(t;\eta,\sigma^2,\Lambda^D)}\right\}dN_i^D(t),$$
$$\tilde{U}_2=\sum_{i=1}^n\int_0^\tau\left\{W_i-\frac{S_2^{(1)}(t;\eta,\gamma,\sigma^2,\Lambda^D)}{S_2^{(0)}(t;\eta,\gamma,\sigma^2,\Lambda^D)}\right\}\Delta_i(t)dN_i(t),$$
$$\tilde{U}_3=\frac{\partial\log L(\sigma^2;\eta,\gamma,\Lambda^D,\Lambda)}{\partial\sigma^2},$$

$$\tilde{U}_4 = \sum_{i=1}^{n} [dN_i^D(t) - \Delta_i(t) H_{1i}(t; \sigma^2, \eta, \Lambda^D) \exp(\eta' W_i) d\Lambda^D(t)],$$
$$\tilde{U}_5 = \sum_{i=1}^{n} [\Delta_i(t) dN_i(t) - \Delta_i(t) H_{2i}(t; \sigma^2, \eta, \Lambda^D) \exp(\gamma' W_i) d\Lambda(t)].$$

Let $\hat{\eta}, \hat{\gamma}, \hat{\sigma}^2, \hat{\Lambda}_0^D(t)$ and $\hat{\Lambda}_0(t)$ denote the solutions to $\tilde{U}(\theta) = 0$, where the estimates $\hat{\Lambda}_0^D(t)$ and $\hat{\Lambda}_0(t)$ will be a piecewise constant function with jumps only at the observation times (across all subjects) and the observed terminal event times, respectively.

To estimate $B_i(t)$, we need to calculate the conditional expectation $\nu_i - 1$ given (m_i, T_i, δ_i). Note that given (m_i, T_i, δ_i), the conditional density of ν_i is $p(\nu_i \mid m_i, T_i, \delta_i) \propto p(\nu_i; \sigma^2) \nu_i^{m_i + \delta_i} \exp\{-\nu_i [\Lambda_0(T_i) \exp(\gamma_0' W_i) + \exp(\eta_0' W_i) \Lambda_0^D(T_i)]\}$, where $p(\nu_i; \sigma^2)$ is the probability density of ν_i. Then $B_i(t)$ can be written as

$$B_i(t) = \left\{ \frac{\sigma_0^2 (m_i + \delta_i) + 1}{\sigma_0^2 [\exp(\gamma_0' W_i) \Lambda_0(T_i) + \exp(\eta_0' W_i) \Lambda_0^D(T_i)] + 1} - 1 \right\} Z_i(t).$$

Thus, $B_i(t)$ can be estimated by

$$\hat{B}_i(t) = \left\{ \frac{\hat{\sigma}^2 (m_i + \delta_i) + 1}{\hat{\sigma}^2 [\exp(\hat{\gamma}' W_i) \hat{\Lambda}_0(T_i) + \exp(\hat{\eta}' W_i) \hat{\Lambda}_0^D(T_i)] + 1} - 1 \right\} Z_i(t).$$

By replacing Λ and B_i with $\hat{\Lambda}_0$ and $\hat{B}_i$ in the estimating function (3.3), and we propose to estimate β_0 using the solution to the equation $U(\beta; \hat{\Lambda}_0, \hat{B}) = 0$, where $\hat{B} = (\hat{B}_1, \ldots, \hat{B}_n)'$. Denote this estimator by $\hat{\beta}$, which can be expressed as

$$\hat{\beta} = \left[\sum_{i=1}^{n} \int_0^{\tau} \left\{ X_i(t) - R_2(t; \hat{\Lambda}_0, \hat{B}) R_1(t; \hat{\Lambda}_0, \hat{B})^{-1} \begin{pmatrix} 1 \\ \hat{B}_i(t) \end{pmatrix} \{X_i(t) - \bar{X}(t)\} \right\}' \Delta_i(t) dN_i(t) \right]^{-1}$$
$$\times \sum_{i=1}^{n} \int_0^{\tau} \left\{ X_i(t) - R_2(t; \hat{\Lambda}_0, \hat{B}) R_1(t; \hat{\Lambda}_0, \hat{B})^{-1} \begin{pmatrix} 1 \\ \hat{B}_i(t) \end{pmatrix} \right\} \{Y_i(t) - \bar{Y}^*(t)\} \Delta_i(t) dN_i(t),$$

where $\bar{X}(t) = \bar{X}(t; \hat{\Lambda}_0)$, $\bar{Y}^*(t) = \bar{Y}^*(t; \hat{\Lambda}_0)$. The corresponding estimators of $\mathcal{A}_{10}(t)$ and $\mathcal{A}_{20}(t)$ are given by $\hat{\mathcal{A}}_1(t) \equiv \hat{\mathcal{A}}_1(t; \hat{\beta}, \hat{\Lambda}_0, \hat{B})$ and $\hat{\mathcal{A}}_2(t) \equiv \hat{\mathcal{A}}_2(t; \hat{\beta}, \hat{\Lambda}_0, \hat{B})$.

As discussed in [27], $\hat{\eta}, \hat{\gamma}, \hat{\sigma}^2, \hat{\Lambda}_0^D(t)$ and $\hat{\Lambda}_0(t)$ are consistent. Then using the uniform strong law of large numbers, one can show that $\hat{\beta}$ is consistent, and $\hat{\mathcal{A}}_1(t)$ and $\hat{\mathcal{A}}_2(t)$ are uniformly consistent for $t \in [0, \tau]$. The asymptotic distributions of $\hat{\beta}$, $\hat{\mathcal{A}}_1(t)$ and $\hat{\mathcal{A}}_2(t)$ are given in the following theorem with the proof given in Appendix.

Theorem 3.1. *Under the regularity conditions* (C1)–(C4) *stated in Appendix,* $n^{1/2}(\hat{\beta} - \beta_0)$ *is asymptotically normal with mean zero and covariance matrix* $D^{-1} \Sigma (D')^{-1}$, *where* D *and* Σ *are defined in the Appendix. Furthermore,* $n^{1/2}\{\hat{\mathcal{A}}_1(t) - \mathcal{A}_{10}(t)\}$ *and* $n^{1/2}\{\hat{\mathcal{A}}_2(t) - \mathcal{A}_{20}(t)\}$ *jointly converge weakly to a zero-mean Gaussian vector process for* $t \in [0, \tau]$.

The asymptotic covariance matrix can be consistently estimated by $\hat{D}^{-1} \hat{\Sigma} (\hat{D}')^{-1}$, where $\hat{D}$ and $\hat{\Sigma}$ are obtained by the usual plug-in method. Note that $\hat{\Sigma}$ is complicated and has no explicit form. Here, we propose to use the bootstrap method to estimate the covariance matrix of $\hat{\beta}$. The accuracy of the bootstrap method depends on the sample size and the number of bootstrap samples. A large number of bootstrap samples yield a high accuracy. In the following simulation studies with the sample size $n = 200$, we used 500 bootstrap samples and found the variance estimation to be fairly accurate. Of course, if the sample size and the number of bootstrap samples are too large, the computation will be time-consuming. In addition, since estimation of each parameter depends on a subset of the other parameters in $\tilde{U}(\theta)$, the solutions can be obtained through an iterative procedure. Here, we propose the following iterative algorithm to solve $\tilde{U}(\theta) = 0$, which is robust and effective in the simulation studies in Section 4.

Step 0. Choose initial estimates $\sigma^{2(0)}$, $\eta^{(0)}$ and $\Lambda_0^{D(0)}(t)$.

Step 1. Let $H_{1i}^{(0)}(t) = \frac{1}{1+\Lambda_0^{D(0)}(t)\exp(W_i'\eta^{(0)})\sigma^{2(0)}}$, and $H_{2i}^{(0)}(t) = \frac{1+\sigma^{2(0)}\delta_i}{1+\Lambda_0^{D(0)}(t)\exp(W_i'\eta^{(0)})\sigma^{2(0)}}$. Put $H_{1i}^{(0)}(t)$ and $H_{2i}^{(0)}(t)$ into $\tilde{U}_1 = 0$, $\tilde{U}_2 = 0$, $\tilde{U}_4 = 0$ and $\tilde{U}_5 = 0$, and solve the resulting equations for updated estimates $\eta^{(1)}$, $\gamma^{(1)}$, $\Lambda_0^{D(1)}(t)$ and $\Lambda_0^{(1)}(t)$.

Step 2. For given $\eta^{(1)}$, $\gamma^{(1)}$, $\Lambda_0^{D(1)}(t)$ and $\Lambda_0^{(1)}(t)$, obtain $\sigma^{2(1)}$ by solving $\tilde{U}_3 = 0$.

Step 3. Return to Step 1 with updated estimates until convergence.

Note that many choices can be used for the initial estimates $\sigma^{2(0)}$, $\eta^{(0)}$ and $\Lambda_0^{D(0)}(t)$. Typically, we can take $\sigma^{2(0)} = 1$, $\eta^{(0)} = 0$, and set $\Lambda_0^{D(0)}(t)$ to be the Nelson-Aalen type estimate of the cumulative baseline hazard function. For the convergence, also several criteria can be applied, and in the simulation studies below, we used the absolute differences $\leqslant 10^{-3}$ between the iterative estimates of the parameters.

4 Simulation study

Simulation studies were conducted to examine the finite sample properties of the proposed estimators with the focus on estimating β_0, $\mathcal{A}_{10}(t)$ and $\mathcal{A}_{20}(t)$, respectively. In the study, let $X_i = (X_{i1}, X_{i2})'$, where X_{i1} follows a Bernoulli distribution with success probability 0.5, and X_{i2} follows a uniform distribution on $(0,1)$. Set $Z_i = X_{i2}$ in (2.1), and $W_i = X_i$ in (2.2) and (2.3). The terminal event time D_i was generated from (2.3) with $\Lambda_0^D(t) = t^2/8$, $\eta_0 = (0.5, 0.5)'$, and $\sigma_0^2 = 0.5$ or 1. The censoring time C_i was taken as $\min(C_i^*, \tau)$, where C_i^* follows a uniform distribution on $(1,6)$ and $\tau = 5$, yielding about 36% censoring for the terminal event. The observation times were generated from (2.2) with $\Lambda_0(t) = 2.1t$ and $\gamma_0 = (0.2, 0.5)'$, which correspond to 6 observations per subject on average. Given ν_i, let $u_i = \nu_i - 1 + e_i$, where e_i is the standard normal random variable. The longitudinal response variable $Y_i(t)$ was generated from (2.1) with $\mu_0(t) = t^{1/2}$, and $\beta_0 = (\beta_1, \beta_2)' = (1,1)'$, where ε_i is the standard normal random variable. For $\alpha_0(t)$, we considered four functions: $\alpha_0(t) = 0, 1, t/4$ or $1/(1+t)$. The results presented below are based on 500 replications with sample size $n = 200$. The asymptotic variance was estimated using the bootstrap method with 500 bootstrap samples, which were found to be adequate.

Table 1 presents the simulation results for estimation of $\beta_0 = (\beta_1, \beta_2)'$. In the table, Bias is the sample mean of the estimate minus the true value, SE is the sampling standard error of the estimate, SEE is

Table 1 Simulation results for estimation of β_1 and β_2

			β_1				β_2			
σ^2	$\alpha_0(t)$	Method	Bias	SE	SEE	CP	Bias	SE	SEE	CP
0.5	0	Ours	−0.0039	0.0649	0.0671	0.952	0.0047	0.1149	0.1235	0.958
		CLZ	−0.0055	0.0698	0.1334	0.982	0.0054	0.1357	0.2623	0.986
	1	Ours	0.0171	0.2097	0.2002	0.944	0.0247	0.3503	0.3380	0.932
		CLZ	0.0099	0.2798	0.3272	0.980	0.8618	0.4825	0.5698	0.656
	$1/(1+t)$	Our	0.0004	0.1403	0.1283	0.932	0.0110	0.2200	0.2181	0.960
		CLZ	−0.0013	0.1836	0.2592	0.966	0.5097	0.3166	0.4522	0.770
	$t/4$	Ours	0.0047	0.0800	0.0829	0.960	0.0200	0.1604	0.1500	0.916
		CLZ	−0.0035	0.1057	0.1725	0.976	0.2222	0.2353	0.3032	0.900
1.0	0	Ours	0.0009	0.0627	0.0657	0.964	−0.0058	0.1149	0.1189	0.940
		CLZ	0.0003	0.0694	0.1798	0.996	−0.0069	0.1337	0.3125	0.982
	1	Ours	0.0047	0.1713	0.1571	0.928	0.0293	0.2852	0.2780	0.940
		CLZ	−0.0121	0.2029	0.3000	0.976	0.4580	0.3617	0.4987	0.850
	$1/(1+t)$	Ours	−0.0004	0.1097	0.1040	0.942	−0.0123	0.1823	0.1829	0.936
		CLZ	−0.0062	0.1395	0.2082	0.972	0.2318	0.2424	0.3733	0.932
	$t/4$	Ours	0.0062	0.0796	0.0761	0.938	0.0293	0.1448	0.1393	0.922
		CLZ	0.0010	0.0940	0.1893	0.996	0.1343	0.1692	0.3587	0.954

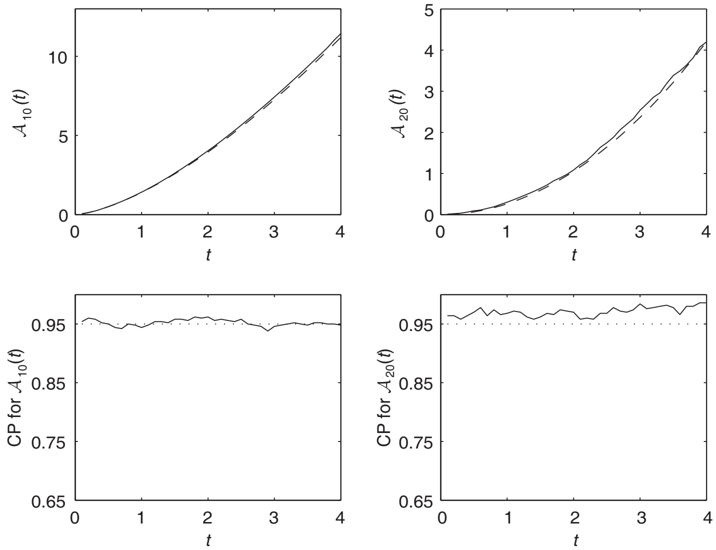

Figure 1 Pointwise biases (top row) and pointwise coverage probabilities (bottom row) for $\mathcal{A}_{10}(t)$ and $\mathcal{A}_{20}(t)$ when $\alpha_0(t) = t/4$. The dashed lines are the true functions and the solid lines are the estimates of $\mathcal{A}_{10}(t)$ and $\mathcal{A}_{20}(t)$, respectively

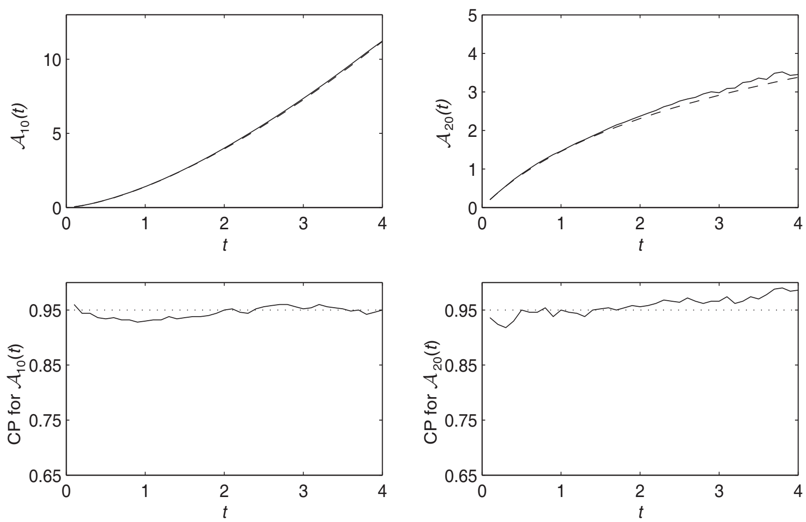

Figure 2 Pointwise biases (top row) and pointwise coverage probabilities (bottom row) for $\mathcal{A}_{10}(t)$ and $\mathcal{A}_{20}(t)$ when $\alpha_0(t) = 1/(1+t)$. The dashed lines are the true functions and the solid lines are the estimates of $\mathcal{A}_{10}(t)$ and $\mathcal{A}_{20}(t)$, respectively

the sample mean of the standard error estimate, and CP is the 95% empirical coverage probability for β_1 and β_2 based on the normal approximation. It can be seen from the table that our proposed method performed well for the situations considered here. Specifically, the proposed estimators were virtually unbiased, and the standard error estimators were very accurate based on the bootstrap method. The 95% empirical coverage probabilities were reasonable.

For comparison, we also considered the method of Cai et al. [1] (denoted by CLZ), who studied (2.1) and (2.2) without the terminal event and assumed that $Z_i(t) \equiv 1$. Under the same setup as above, the simulation results for β_1 and β_2 are also reported in Table 1. The results indicate that the CLZ' method may lead to biases, especially for estimation of β_2. In addition, the CLZ' method tended to cause an inflated SEE, and yielded improper coverage probabilities.

In the same simulation studies as reported in Table 1 with $\sigma_0^2 = 1$, we investigated the performance of the estimates of $\mathcal{A}_{10}(t)$ and $\mathcal{A}_{20}(t)$. We computed pointwise biases and pointwise coverage probabilities of the 95% confidence intervals on $(0, 4]$ with 40 grid points $0.1k$, $k = 1, \ldots, 40$. The pointwise biases and pointwise coverage probabilities are depicted in Figures 1 and 2 with

$$\alpha_0(t) = \frac{t}{4} \quad \text{and} \quad \alpha_0(t) = \frac{1}{(1+t)},$$

respectively. These results show that the proposed estimators performed quite well and essentially provided unbiased estimates of $\mathcal{A}_{10}(t)$ and $\mathcal{A}_{20}(t)$. The asymptotic standard errors presented a reasonable description of the variability for the estimates of $\mathcal{A}_{10}(t)$ and $\mathcal{A}_{20}(t)$, and the proposed estimation procedures are reliable. We also considered other setups and the results were similar to those given above.

5 An application

For the illustration purpose, we applied the proposed methods to the longitudinal bladder cancer data (see [21,22]). In this study, the patients were randomly assigned to placebo and thiotepa treatment groups. During the study, many patients had multiple recurrences of the bladder tumors and all recurrent tumors between visits were recorded and removed at clinical visits. There are 85 bladder cancer patients, 47 in the placebo group and 38 in the thiotepa treatment group. For each patient, the observed information includes the clinical visit times in month and the number of bladder tumors that occurred between clinical visits. The frequency of visits ranges from 1 to 38, and the average visiting numbers are 13.5 for the treatment group and 8.7 for the control group. About 25.9% of patients died during follow-up, and the total follow-up is 53 months. In addition, two baseline covariates were measured: the number of initial tumors before entering the study and the size of the largest initial tumor. These data have been analyzed by Sun et al. [21], Sun et al. [22], Liang et al. [12] and Zhou et al. [31], among others. In particular, without modeling the terminal event, Liang et al. [12] used a mixed random effect model to analyze the data. Cai et al. [1] showed that the latent variable effect is truly time-dependent, and proposed a time-varying latent effect model to analyze the data. Here we analyze the data using the proposed time-varying latent effect model, and the main goal is to investigate the time-varying latent effects shared by the longitudinal responses, the observation times and the terminal event. Since the size of the largest initial tumor had been shown to have no effect on the tumor recurrence rate (see [22, 30]), we focus on the effects of thiotepa treatment and number of initial tumors on the tumor recurrence process with informative observation and terminal event times.

Following Liang et al. [12], for subject i ($i = 1, \ldots, 85$), let $Y_i(t)$ stand for the natural logarithm of the number of observed tumors at time t plus 1 to avoid 0. For covariates, let $X_{i1} = 1$ if the patient was from the thiotepa group, and 0 if the patient was from the placebo group, and X_{i2} be the logarithm of the number of initial tumors. We first estimated the regression parameters in (2.2) for the visiting process, and the results are summarized in Table 2. These results suggest that the thiotepa treatment has a significant effect on the visiting process, but the number of initial tumors seems to have no significant effect on the visiting process.

Based on the above analysis, we chose $Z_i = X_{i1}$ in (2.1) because the thiotepa treatment is significantly related to the visiting process, which was also verified by Zhou et al. [31] using the focused information criterion. The asymptotic variance was estimated using bootstrap method with 500 bootstrap samples. The estimates of $\beta_0 = (\beta_1, \beta_2)'$ and their estimated standard errors are presented in Table 2. We can observe that both the thiotepa treatment and the number of initial tumors have significant effects on the tumor recurrence process. In particular, the thiotepa treatment significantly reduce the tumor recurrence rate, the number of initial tumor has the detrimental effect, i.e., the patients with the higher number of initial tumors tend to have a higher tumor occurrence rate. In addition, for the estimates of γ_1, γ_2 and β_1, the estimated standard errors from our proposed method are smaller than those from the methods of Liang et al. [12] and Cai et al. [1]; while for the estimate of β_2, the three methods provide comparable estimated standard errors. Thus, our proposed method is more efficient than the methods of Liang et al. [12] and Cai et al. [1].

Table 2 Joint analysis of the bladder cancer data

	Est	SE	p-value
γ_1	0.455	0.129	4.396×10^{-4}
γ_2	−0.139	0.151	0.358
β_1	−0.160	0.002	6.103×10^{-5}
β_2	0.194	0.067	3.600×10^{-3}

Note. Est is the estimate of the parameter, and SE is the standard error estimate.

6 Concluding remarks

In this paper, we proposed a semiparametric mixed effect model with time-varying latent effects in the analysis of longitudinal data with informative observation times and a dependent terminal event. The proposed model provided a flexible way of modeling the effects of latent variables on the longitudinal response variable while adjusting its association with the observation times and the terminal event. An estimation procedure was proposed to obtain consistent and asymptotically normal estimators. The simulation results demonstrated that the proposed method performed well, and an illustrative example was provided.

We used the proportional hazards frailty model for the terminal event. Other competing models, such as the additive hazards frailty model, the proportional odds frailty model, and the linear transformation model with frailties (see [29]), could be used as well. In addition, (2.3) can be generalized to

$$d\Lambda^D(t \mid W_i, \nu_i) = \nu_i^b \exp\{\eta_0' W_i\} d\Lambda_0^D(t),$$

where b is an unknown parameter. It seems not to be straightforward to generalize the proposed approach to this situation, and further research is needed to address this issue.

For (2.1), if one is interested in $\mu_0(t)$ and $\alpha_0(t)$ directly, the estimates of $\mathcal{A}_{10}(t)$ and $\mathcal{A}_{20}(t)$ may be used as base for estimation. Specifically, let $\mu_0^*(t) = d\mathcal{A}_{10}(t)/dt$, $\alpha_0^*(t) = d\mathcal{A}_{20}(t)/dt$ and $\lambda_0(t) = d\Lambda_0(t)/dt$. Using the kernel smoothing technique, we can obtain the estimates $\hat{\mu}_0^*(t)$, $\hat{\alpha}_0^*(t)$ and $\hat{\lambda}_0(t)$ of $\mu_0^*(t)$, $\alpha_0^*(t)$ and $\lambda_0(t)$ based on $\hat{\mathcal{A}}_1(t)$, $\hat{\mathcal{A}}_2(t)$ and $\hat{\Lambda}_0(t)$, respectively. Then $\mu_0(t)$ and $\alpha_0(t)$ can be estimated by $\hat{\mu}_0^*(t)/\hat{\lambda}_0(t)$ and $\hat{\alpha}_0^*(t)/\hat{\lambda}_0(t)$, respectively. It would be worthwhile to study the properties of these estimators in future research.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 11231010, 11171330 and 11201315), Key Laboratory of Random Complex Structures and Data Science, Chinese Academy of Sciences (Grant No. 2008DP173182) and Beijing Center for Mathematics and Information Interdisciplinary Sciences. The authors thank the associate editor-in-chief, the associate editor and the two reviewers for their constructive and insightful comments and suggestions that greatly improved the paper.

References

1 Cai N, Lu W, Zhang H. Time-varying latent effect model for longitudinal data with informative observation times. Biometrics, 2012, 68: 1093–1102
2 Davis C S. Statistical Methods for the Analysis of Repeated Measurements. New York: Springer, 2002
3 Diggle P J, Liang K Y, Zeger S L. The Analysis of Longitudinal Data. Oxford: Oxford University Press, 1994
4 Fan J, Li R. New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. J Amer Statist Assoc, 2004, 99: 710–723
5 Fang S, Zhang H, Sun L. Joint analysis of longitudinal data with additive mixed effect model for informative observation times. J Statist Plann Inference, 2016, 169: 43–55
6 Fitzmaurice G M, Laird N M, Ware J H. Applied Longitudinal Analysis. New York: John Wiley and Sons, 2004
7 Gill R D, Johansen S. A survey of product-integration with a view toward application in survival analysis. Ann Statist, 1990, 18: 1501–1555

8 He X, Tong X, Sun J. Semiparametric analysis of panel count data with correlated observation and follow-up times. Lifetime Data Anal, 2009, 15: 177–196
9 Huang C Y, Wang M C. Joint modeling and estimation for recurrent event processes and failure time data. J Amer Statist Assoc, 2004, 99: 1153–1165
10 Laird N M, Ware J H. Random-effects models for longitudinal data. Biometrics, 1982, 38: 963–974
11 Liang K Y, Zeger S L. Longitudinal data analysis using generalized linear models. Biometrika, 1986, 73: 13–22
12 Liang Y, Lu W, Ying Z. Joint modeling and analysis of longitudinal data with informative observation times. Biometrics, 2009, 65: 377–384
13 Lin D Y, Ying Z. Semiparametric and nonparametric regression analysis of longitudinal data. J Amer Statist Assoc, 2001, 96, 103–126
14 Lin H, Scharfstein D O, Rosenheck R A. Analysis of longitudinal data with irregular outcome-dependent follow-up. J Roy Statist Soc Ser B, 2004, 66: 791–813
15 Liu L, Huang X, O'Quigley J. Analysis of longitudinal data in the presence of informative observational times and a dependent terminal event, with application to medical cost data. Biometrics, 2008, 64: 950–958
16 Liu L, Wolfe R A, Huang X. Shared frailty models for recurrent events and a terminal event. Biometrics, 2004, 60: 747–756
17 Liu L, Wolfe R A, Kalbfleisch J D. A shared random effects model for censored medical costs and mortality. Stat Med, 2007, 26: 139–155
18 Pollard D. Empirical Processes: Theory and Applications. Hayward: Institute of Mathematical Statistics, 1990
19 Ryu D, Sinha D, Mallick B, et al. Longitudinal studies with outcome-dependent follow-up: Models and Bayesian regression. J Amer Statist Assoc, 2007, 102: 952–961
20 Song X, Mu X, Sun L. Regression analysis of longitudinal data with time-dependent covariates and informative observation times. Scand J Statist, 2012, 39: 248–258
21 Sun J, Park D-H, Sun L, et al. Semiparametric regression analysis of longitudinal data with informative observation times. J Amer Statist Assoc, 2005, 100: 882–889
22 Sun J, Sun L, Liu D. Regression analysis of longitudinal data in the presence of informative observation and censoring times. J Amer Statist Assoc, 2007, 102: 1397–1406
23 Sun L, Song X, Zhou J, et al. Joint analysis of longitudinal data with informative observation times and a dependent terminal event. J Amer Statist Assoc, 2012, 107: 688–700
24 van der Vaart A W, Wellner J A. Weak Convergence and Empirical Processes. New York: Springer, 1996
25 Welsh A H, Lin X, Carroll R J. Marginal longitudinal nonparametric regression: Locality and efficiency of spline and kernel methods. J Amer Statist Assoc, 2002, 97: 482–493
26 Wu M C, Carroll R J. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. Biometrics, 1988, 44: 175–188
27 Ye Y, Kalbfleisch J D, Schaubel D E. Semiparametric analysis of correlated recurrent and terminal events. Biometrics, 2007, 63: 78–87
28 Zeger S L, Diggle P J. Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. Biometrics, 1994, 50: 689–699
29 Zeng D, Lin D. Semiparametric transformation models with random effects for joint analysis of recurrent and terminal events. Biometrics, 2009, 65: 746–752
30 Zhao X, Tong X, Sun L. Joint analysis of longitudinal data with dependent observation times. Statist Sinica, 2012, 22: 317–336
31 Zhou J, Zhao X, Sun L. A new inference approach for joint models of longitudinal data with informative observation and censoring times. Statist Sinica, 2013, 23: 571–593

Appendix A: Proof of Theorem 3.1

In order to study the asymptotic properties of the proposed estimators, we need the following regularity conditions:

(C1) $\{Y_i(\cdot), N_i(\cdot), T_i, \delta_i, X_i(\cdot), Z_i(\cdot), W_i\}, i = 1, \ldots, n$, are independent and identically distributed.

(C2) $E\{N_i(\tau)\} < \infty$, and $P(T_i \geqslant \tau) > 0$.

(C3) $X_i(t)$ and $Z_i(t)$ are of bounded variation on $[0, \tau]$, and W_i is bounded.

(C4) $\mu_0(t)$ and $\alpha_0(t)$ are right continuous with left-hand limits, and have bounded total variation on $[0, \tau]$.

(C5) D is nonsingular, where

$$D = E\left[\int_0^\tau \left\{X_i(t) - R_2(t;\Lambda_0,B)\{R_1(t;\Lambda_0,B)\}^{-1}\begin{pmatrix}1\\ B_i(t)\end{pmatrix}\right\}\{X_i(t)-\bar{x}(t)\}\right],$$

and $\bar{x}(t)$ is the limit of $\bar{X}(t;\Lambda_0)$.

Conditions (C1)–(C3) are standard for regression methods in analyzing longitudinal data [23], in which Condition (C2) implies that there is a positive probability for the longitudinal responses to be observed in $[0,\tau]$. Condition (C4) is a technical assumption for the existence of $\mathcal{A}_{10}(t)$ and $\mathcal{A}_{20}(t)$. Note that D is the limit of $-\partial U(\beta_0;\hat{\Lambda}_0,\hat{B})/\partial\beta$. Condition (C6) is needed for the existence and uniqueness of the estimator $\hat{\beta}$.

A.1 Asymptotic results for $\hat{\eta}$, $\hat{\gamma}$, $\hat{\sigma}^2$, $\hat{\Lambda}_0^D(t)$ and $\hat{\Lambda}_0(t)$

Define

$$\hat{\Lambda}_0^D(t;\eta,\sigma^2,\Lambda^D) = \frac{1}{n}\sum_{i=1}^n\int_0^t \frac{dN_i^D(u)}{S_1^{(0)}(u;\eta,\sigma^2,\Lambda^D)}$$

and

$$\hat{\Lambda}_0(t;\eta,\gamma,\sigma^2,\Lambda^D) = \frac{1}{n}\sum_{i=1}^n\int_0^t \frac{\Delta_i(u)dN_i(u)}{S_2^{(0)}(u;\eta,\gamma,\sigma^2,\Lambda^D)}.$$

Let $\tilde{\Lambda}_0^D(t) = \hat{\Lambda}_0^D(t;\eta_0,\sigma_0^2,\Lambda_0^D)$, $\tilde{\Lambda}_0(t) = \hat{\Lambda}_0(t;\eta_0,\gamma_0,\sigma_0^2,\Lambda_0^D)$, $\tilde{S}_1^{(k)}(t;\Lambda^D) = S_1^{(k)}(t;\eta_0,\sigma_0^2,\Lambda^D)$, $\tilde{S}_2^{(k)}(t;\Lambda^D) = S_2^{(k)}(t;\eta_0,\gamma_0,\sigma_0^2,\Lambda^D)$ $(k=0,1)$, $\tilde{H}_{1i}(t;\Lambda^D) = H_{1i}(t;\eta_0,\sigma_0^2,\Lambda^D)$, $\tilde{H}_{2i}(t;\Lambda^D) = H_{2i}(t;\eta_0,\sigma_0^2,\Lambda^D)$, $\bar{W}_1(t;\Lambda^D) = \tilde{S}_1^{(1)}(t;\Lambda^D)/\tilde{S}_1^{(0)}(t;\Lambda^D)$, $\bar{W}_2(t;\Lambda^D) = \tilde{S}_2^{(1)}(t;\Lambda^D)/\tilde{S}_2^{(0)}(t;\Lambda^D)$, $\bar{N}^D(t) = n^{-1}\sum_{i=1}^n N_i^D(t)$ and $\bar{N}(t) = n^{-1}\sum_{i=1}^n\int_0^t\Delta_i(u)N_i(u)$. Also let $s_1^{(0)}(t)$, $s_2^{(0)}(t)$, $\bar{w}_1(t)$ and $\bar{w}_2(t)$ be the limits of $\tilde{S}_1^{(0)}(t;\Lambda_0^D)$, $\tilde{S}_2^{(0)}(t;\Lambda_0^D)$, $\bar{W}_1(t;\Lambda_0^D)$ and $\bar{W}_2(t;\Lambda_0^D)$, respectively. Set

$$\tilde{G}_1(t) = \frac{1}{n}\sum_{i=1}^n\int_0^t \sigma_0^2\Delta_i(u)\tilde{H}_{1i}(u;\Lambda_0^D)\tilde{H}_{1i}(u;\tilde{\Lambda}_0^D)\exp(2\eta_0'W_i)\frac{d\Lambda_0^D(u)}{\tilde{S}_1^{(0)}(u;\tilde{\Lambda}_0^D)}.$$

It can be shown that

$$\tilde{\Lambda}_0^D(t) - \Lambda_0^D(t) = \int_0^t\{\tilde{\Lambda}_0^D(u) - \Lambda_0^D(u)\}dG_1(u) + \frac{1}{n}\sum_{i=1}^n\int_0^t\frac{d\tilde{M}_i^D(u)}{\tilde{S}_1^{(0)}(u;\tilde{\Lambda}_0^D)},$$

which is a linear Volterra integral equation, and the solution is

$$\tilde{\Lambda}_0^D(t) - \Lambda_0^D(t) = \frac{1}{n}\sum_{i=1}^n\int_0^t\frac{\tilde{Q}(u-)}{\tilde{Q}(t)}\frac{d\tilde{M}_i^D(u)}{\tilde{S}_1^{(0)}(u;\tilde{\Lambda}_0^D)},$$

where $\tilde{Q}(t) = \prod_{s\leqslant t}\{1-d\tilde{G}_1(s)\}$ is the product-integral of $\tilde{G}_1(s)$ over $[0,t]$ (see [7]). Thus, using the asymptotic properties of the product-integral [7], the uniform strong law of large numbers [18], and [13, Lemma A.1], we obtain that uniformly in $t\in[0,\tau]$,

$$\tilde{\Lambda}_0^D(t) - \Lambda_0^D(t) = \frac{1}{n}\sum_{i=1}^n\psi_{1i}(t) + o_p(n^{-1/2}), \tag{A.1}$$

where

$$\psi_{1i}(t) = \int_0^t\frac{Q(u-)}{Q(t)}\frac{d\tilde{M}_i^D(u)}{s_1^{(0)}(u)},$$

and $Q(t)$ is the limit of $\tilde{Q}(t)$. Let

$$\tilde{G}_2(t) = \frac{1}{n}\sum_{i=1}^n\int_0^t\sigma_0^2\Delta_i(u)\tilde{H}_{1i}(u;\Lambda_0^D)H_{2i}(u;\tilde{\Lambda}_0^D)\exp\{(\eta_0+\gamma_0)'W_i\}\frac{d\Lambda_0(u)}{\tilde{S}_2^{(0)}(u;\tilde{\Lambda}_0)},$$

and $G_2(t)$ be the limit of $\tilde{G}_2(t)$. It then follows from (A.1) that

$$\tilde{\Lambda}_0(t)-\Lambda_0(t)=\int_0^t\{\tilde{\Lambda}_0^D(u)-\Lambda_0^D(u)\}d\tilde{G}_2(u)+\frac{1}{n}\sum_{i=1}^n\int_0^t\frac{d\tilde{M}_i(u)}{\tilde{S}_2^{(0)}(u;\tilde{\Lambda}_0^D)}$$
$$=\frac{1}{n}\sum_{i=1}^n\psi_{2i}(t)+o_p(n^{-1/2}), \tag{A.2}$$

where

$$\psi_{2i}(t)=\int_0^t\psi_{1i}(u)dG_2(u)+\int_0^t\frac{d\tilde{M}_i(u)}{s_2^{(0)}(u)}.$$

Define

$$\tilde{U}_1(\eta;\sigma^2)=\sum_{i=1}^n\int_0^\tau\bigg\{W_i-\frac{S_1^{(1)}(t;\eta,\sigma^2,\hat{\Lambda}_0^D)}{S_1^{(0)}(t;\eta,\sigma^2,\hat{\Lambda}_0^D)}\bigg\}dN_i^D(t)$$

and

$$\tilde{U}_2(\gamma;\eta,\sigma^2)=\sum_{i=1}^n\int_0^\tau\bigg\{W_i-\frac{S_2^{(2)}(t;\eta,\gamma,\sigma^2,\hat{\Lambda}_0^D)}{S_2^{(0)}(t;\eta,\gamma,\sigma^2,\hat{\Lambda}_0^D)}\bigg\}\Delta_i(t)dN_i(t).$$

Note that

$$\tilde{U}_1(\eta_0;\sigma_0^2)=\sum_{i=1}^n\int_0^\tau\{W_i-\bar{W}_1(t;\Lambda_0^D)\}d\tilde{M}_i^D(t)+\sum_{i=1}^n\int_0^\tau\{\bar{W}_1(t;\Lambda_0^D)-\bar{W}_1(t;\tilde{\Lambda}_0^D)\}dN_i^D(t).$$

Similar to (A.2), we have

$$\tilde{U}_1(\eta_0;\sigma_0^2)=\sum_{i=1}^n\zeta_{1i}+o_p(n^{1/2}), \tag{A.3}$$

where

$$\zeta_{1i}=\int_0^\tau\{W_i-\bar{w}_1(t)\}d\tilde{M}_i^D(t)+\int_0^\tau\psi_{1i}(t)dG_3(t),$$

and $G_3(t)$ is the limit of $\tilde{G}_3(t)$ with

$$\tilde{G}_3(t)=\frac{1}{n}\sum_{i=1}^n\int_0^t\sigma_0^2\Delta_i(u)\tilde{H}_{1i}(u;\Lambda_0^D)\tilde{H}_{1i}(u;\tilde{\Lambda}_0^D)\exp(2\eta_0'W_i)\{W_i-\bar{W}_1(u;\Lambda_0^D)\}\frac{d\bar{N}^D(u)}{\tilde{S}_1^{(0)}(u;\tilde{\Lambda}_0^D)}.$$

In a similar manner, we obtain

$$\tilde{U}_2(\gamma_0;\eta_0,\sigma_0^2)=\sum_{i=1}^n\zeta_{2i}+o_p(n^{1/2}), \tag{A.4}$$

where

$$\zeta_{2i}=\int_0^\tau\{W_i-\bar{w}_2(t)\}d\tilde{M}_i(t)+\int_0^\tau\psi_{1i}(t)dG_4(t),$$

and $G_4(t)$ is the limit of $\tilde{G}_4(t)$ with

$$\tilde{G}_4(t)=\frac{1}{n}\sum_{i=1}^n\int_0^t\sigma_0^2\Delta_i(u)\tilde{H}_{1i}(u;\Lambda_0^D)\tilde{H}_{2i}(u;\tilde{\Lambda}_0^D)\exp\{(\eta_0+\gamma_0)'W_i\}\{W_i-\bar{W}_2(u;\Lambda_0^D)\}\frac{d\bar{N}(u)}{\tilde{S}_2^{(0)}(u;\tilde{\Lambda}_0^D)}.$$

Define

$$\widetilde{U}_3(\sigma^2;\eta,\gamma)=\sum_{i=1}^n\bigg[h_i(\sigma^2)+\frac{m_i+\delta_i+1/\sigma^2}{\sigma^4(\tilde{d}_i(\eta,\gamma)+1/\sigma^2)}+\frac{1}{\sigma^4}\log\{\tilde{d}_i(\eta,\gamma)+1/\sigma^2\}\bigg],$$

where

$$h_i(\sigma^2)=\frac{1}{\sigma^4}\bigg\{-\frac{\Gamma^{(1)}(m_i+\delta_i+1/\sigma^2)}{\Gamma(m_i+\delta_i+1/\sigma^2)}+\frac{\Gamma^{(1)}(1/\sigma^2)}{\Gamma(1/\sigma^2)}+\log(\sigma^2)-1\bigg\},$$

$$\tilde{d}_i(\eta,\gamma)=\int_0^\infty \Delta_i(u)[\exp(\eta' W)d\hat{\Lambda}_0^D(u;\eta,\sigma^2,\Lambda^D)+\exp(\gamma' W_i)d\hat{\Lambda}_0(u;\eta,\sigma^2,\Lambda^D)]$$

and $\Gamma^{(1)}(x)=\partial\Gamma(x)/\partial x$. From an argument similar to that in the proof of (A.3), we get

$$\tilde{U}_3(\sigma_0^2;\gamma_0,\eta_0)=\sum_{i=1}^n \zeta_{3i}+o_p(n^{1/2}), \tag{A.5}$$

where

$$\zeta_{3i}=\left[h_i(\sigma_0^2)+\frac{m_i+\delta_i+1/\sigma_0^2}{\sigma_0^4(d_i+1/\sigma_0^2)}+\frac{1}{\sigma_0^4}\log\{d_i+1/\sigma_0^2\}\right]+\frac{d_i+m_i+\delta_i+1/\sigma_0^2}{\sigma_0^4(d_i+2/\sigma_0^2)^2}$$
$$\times\int_0^\infty \Delta_i(u)\{\exp(\eta_0' W)d\psi_{1i}(u)+\exp(\gamma_0' W_i)d\psi_{2i}(u)\}.$$

Let $\alpha=(\eta',\gamma',\sigma^2)'$, $\alpha_0=(\eta_0',\gamma_0',\sigma_0^2)'$, $\hat{\alpha}=(\hat{\eta}',\hat{\gamma}',\hat{\sigma}^2)')$, and

$$\tilde{U}_\alpha(\alpha)=(\tilde{U}_1(\eta;\sigma^2)',\tilde{U}_1(\gamma;\eta,\sigma^2)',\tilde{U}_3(\sigma^2;\eta,\gamma)').$$

Note that $\tilde{U}_\alpha(\hat{\alpha})=0$. Then it follows from (A.3)–(A.5) and the Taylor expansion that

$$\hat{\alpha}-\alpha_0=\frac{1}{n}\sum_{i=1}^n \Omega^{-1}\zeta_i+o_p(n^{-1/2}),$$

where Ω is the limit of $-n^{-1}\partial\tilde{U}_\alpha(\alpha_0)/\partial\alpha$, and $\zeta_i=(\zeta_{1i}',\zeta_{2i}',\zeta_{3i})'$. For notational convenience, denote $(\xi_{1i}',\xi_{2i}',\xi_{3i})'=\Gamma^{-1}\zeta_i$. Thus,

$$\hat{\eta}-\eta_0=\frac{1}{n}\sum_{i=1}^n \xi_{1i}+o_p(n^{-1/2}), \tag{A.6}$$

$$\hat{\gamma}-\gamma_0=\frac{1}{n}\sum_{i=1}^n \xi_{2i}+o_p(n^{-1/2}) \tag{A.7}$$

and

$$\hat{\sigma}^2-\sigma_0^2=\frac{1}{n}\sum_{i=1}^n \xi_{3i}+o_p(n^{-1/2}). \tag{A.8}$$

Let

$$\tilde{G}_5(t)=\frac{1}{n}\sum_{i=1}^n\int_0^t \sigma_0^2\Delta_i(u)\tilde{H}_{1i}^2(u;\Lambda_0^D)\exp(2\eta_0' W_i)W_i\frac{d\Lambda_0^D(u)}{\tilde{S}_1^{(0)}(u;\Lambda_0^D)}-\int_0^t \bar{W}_1(u;\Lambda_0^D)d\Lambda_0^D(u),$$
$$\tilde{G}_6(t)=\frac{1}{n}\sum_{i=1}^n\int_0^t \sigma_0^2\Lambda_0^D(u)\Delta_i(u)\tilde{H}_{1i}(u;\Lambda_0^D)\tilde{H}_{2i}(u;\Lambda_0^D)\exp\{(\eta_0+\gamma_0)' W_i\}W_i\frac{d\Lambda_0(u)}{S_2^{(0)}(u;\Lambda_0^D)},$$
$$\tilde{G}_7(t)=\frac{1}{n}\sum_{i=1}^n\int_0^t \Delta_i(u)\tilde{H}_{1i}(u;\Lambda_0^D)\big[\tilde{H}_{2i}(u;\Lambda_0^D)\exp(\eta_0' W_i)\Lambda_0^D(u)-\delta_i\big]\exp(\gamma_0' W_i)\frac{d\Lambda_0(u)}{S_2^{(0)}(u;\Lambda_0^D)}$$

and $G_5(t)$, $G_6(t)$ and $G_7(t)$ be the limits of $\tilde{G}_5(t)$, $\tilde{G}_6(t)$ and $\tilde{G}_7(t)$, respectively. Note that $\hat{\Lambda}_0^D(t)\equiv\hat{\Lambda}_0^D(t;\hat{\eta},\hat{\sigma}^2,\hat{\Lambda}_0^D)$, and $\hat{\Lambda}_0(t)\equiv\hat{\Lambda}_0(t;\hat{\eta},\hat{\gamma},\sigma^2,\hat{\Lambda}_0^D)$. Following similar arguments to that in (A.3), together with (A.6)–(A.8), we have

$$\hat{\Lambda}_0^D(t)-\Lambda_0^D(t)=\frac{1}{n}\sum_{i=1}^n \psi_{3i}(t)+o_p(n^{-1/2}), \tag{A.9}$$

where

$$\psi_{3i}(t)=\psi_{1i}(t)+\xi_{1i}'\int_0^t\frac{Q(u-)}{Q(t)}dG_5(u)+\xi_{3i}\sigma_0^{-2}\int_0^t\frac{Q(u-)}{Q(t)}dG_1(u).$$

Likewise,

$$\hat{\Lambda}_0(t) - \Lambda_0(t) = \frac{1}{n}\sum_{i=1}^{n} \psi_{4i}(t) + o_p(n^{-1/2}), \tag{A.10}$$

where

$$\psi_{4i}(t) = \int_0^t \frac{d\tilde{M}_i(u)}{s_2^{(0)}(u)} + \int_0^t \psi_{3i}(u)dG_2(u) + \xi_{1i}' G_6(t) - \xi_{2i}' \int_0^t \bar{w}_2(u)d\Lambda_0(u) + \xi_{3i} G_7(t).$$

A.2 Asymptotic result for $U(\beta_0; \hat{\Lambda}_0, \hat{B})$

Define $K_{10}(t) = \Lambda_0(t)$, $K_{20}(t) = 0$, and

$$M_i^*(t; K_1, K_2, \Lambda, B_i) = \int_0^t \Delta_i(u)N_i(u) - \Delta_i(t)\frac{m_i}{\Lambda_0(T_i)}dK_1(t) - \Delta_i(t)\frac{m_i}{\Lambda_0(T_i)}B_i(t)'K_2(t).$$

Then

$$E\{M_i^*(t; K_{10}, K_{20}, \Lambda_0, B_i) \mid m_i, T_i, \delta_i\} = 0.$$

Thus, we can use the following estimating equations to estimate $K_{10}(t)$ and $K_{20}(t)$:

$$\sum_{i=1}^{n} \begin{pmatrix} 1 \\ \hat{B}_i(t) \end{pmatrix} dM_i^*(t; K_1, K_2, \hat{\Lambda}_0, \hat{B}_i) = 0.$$

The solutions to the above equations are given by

$$\begin{pmatrix} \hat{K}_1(t) \\ \hat{K}_2(t) \end{pmatrix} = \int_0^t R_1(u; \hat{\Lambda}_0, \hat{B})^{-1} \begin{pmatrix} 1 \\ \hat{B}_i(u) \end{pmatrix} \Delta_i(u)dN_i(u).$$

Let $M_i(t) = M_i(t; \beta_0, \mathcal{A}_{10}, \mathcal{A}_{20}, \Lambda_0)$, $M_i^*(t) = M_i^*(t; K_{10}, K_{20}, \Lambda_0, B_i)$, $\bar{Y}^*(t) = \bar{Y}^*(t; \hat{\Lambda}_0)$ and $\bar{X}(t) = \bar{X}(t; \hat{\Lambda}_0)$. It can be checked that

$$\begin{aligned} U(\beta_0; \hat{\Lambda}_0, \hat{B}) &= \sum_{i=1}^{n} \int_0^\tau X_i(t)dM_i(t) - \sum_{i=1}^{n} \int_0^\tau X_i(t)\{\bar{Y}^*(t) - \beta_0'\bar{X}(t)\}dM_i^*(t) \\ &\quad - \int_0^\tau \Delta_i(t)X_i(t)\left[\frac{m_i}{\hat{\Lambda}_0(T_i)}d\hat{\mathcal{A}}_1(t; \beta_0, \hat{\Lambda}_0, \hat{B}) - \frac{m_i}{\Lambda_0(T_i)}d\mathcal{A}_{10}(t)\right] \\ &\quad - \sum_{i=1}^{n} \int_0^\tau \Delta_i(t)X_i(t)\left[\frac{m_i\hat{B}_i(t)'}{\hat{\Lambda}_0(T_i)}d\hat{\mathcal{A}}_2(t; \beta_0, \hat{\Lambda}_0, \hat{B}) - \frac{m_iB_i(t)'}{\Lambda_0(T_i)}d\mathcal{A}_{20}(t)\right] \\ &\quad + \sum_{i=1}^{n} \int_0^\tau \Delta_i(t)\{\bar{Y}^*(t) - \beta_0'\bar{X}(t)\}X_i(t)\left[\frac{m_i}{\hat{\Lambda}_0(T_i)}d\hat{K}_1(t) - \frac{m_i}{\Lambda_0(T_i)}dK_{10}(t)\right] \\ &\quad + \sum_{i=1}^{n} \int_0^\tau \Delta_i(t)\{\bar{Y}^*(t) - \beta_0'\bar{X}(t)\}X_i(t)\left[\frac{m_i\hat{B}_i(t)'}{\hat{\Lambda}_0(T_i)}d\hat{K}_2(t) - \frac{m_iB_i(t)'}{\Lambda_0(T_i)}dK_{20}(t)\right]. \end{aligned} \tag{A.11}$$

Let $\Psi_{1i}(t)$, $\Psi_{2i}(t)$, $\Psi_{3i}(t)$, $\Psi_{4i}(t)$ and $\Psi_{5i}(t)$ be the derivatives and the Hadamard derivatives of $B_i(t)$ with respect to η, γ, σ^2, Λ^D and Λ, respectively. Note that

$$\frac{1}{n}\sum_{i=1}^{n} M_i(t; \beta_0, \hat{\mathcal{A}}_1(\cdot; \beta_0, \hat{\Lambda}_0, \hat{B}), \hat{\mathcal{A}}_2(\cdot; \beta_0, \hat{\Lambda}_0, \hat{B}), \hat{\Lambda}) = 0.$$

Then it follows from the functional delta method and (A.6)–(A.10) that

$$\frac{1}{n}\sum_{i=1}^{n} M_i(t) = \frac{1}{n}\sum_{i=1}^{n}\left[\int_0^t \frac{\Delta_i(s)m_i}{\hat{\Lambda}_0(T_i)}d\hat{\mathcal{A}}_1(s; \beta_0, \hat{\Lambda}_0, \hat{B}) - \int_0^t \frac{\Delta_i(s)m_i}{\Lambda_0(T_i)}d\mathcal{A}_{10}(s)\right]$$

$$+\frac{1}{n}\sum_{i=1}^{n}\left[\int_0^t \frac{\Delta_i(s)m_i}{\hat{\Lambda}_0(T_i)}\hat{B}_i(s)'d\mathcal{A}_2(s;\beta_0,\hat{\Lambda}_0,\hat{B}) - \int_0^t \frac{\Delta_i(s)m_iB_i(s)'}{\Lambda_0(T_i)}d\mathcal{A}_{20}(s)\right]$$
$$=\frac{1}{n}\sum_{i=1}^{n}L_{1i}(t) + \int_0^t \kappa_1(s)d\{\mathcal{A}_1(s;\beta_0,\hat{\Lambda}_0,\hat{B}) - \mathcal{A}_{10}(s)\}$$
$$+\int_0^t \kappa_2(s)'d\{\mathcal{A}_2(s;\beta_0,\hat{\Lambda}_0,\hat{B}) - \mathcal{A}_{20}(s)\} + o_p(n^{-\frac{1}{2}}), \tag{A.12}$$

where

$$L_{1i}(t) = J_1(t)'\xi_{1i} + J_2(t)'\xi_{2i} + J_3(t)\xi_{3i} + \int_0^\tau J_4(t,s)\,d\psi_{3i}(s) + \int_0^\tau J_5(t,s)d\psi_{4i}(s)$$

with

$$J_1(t) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\int_0^t \Delta_i(u)\frac{m_i}{\Lambda_0(T_i)}\Psi_{1i}(u)'d\mathcal{A}_{20}(u),$$
$$J_2(t) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\int_0^t \Delta_i(u)\frac{m_i}{\Lambda_0(T_i)}\Psi_{2i}(u)'d\mathcal{A}_{20}(u),$$
$$J_3(t) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\int_0^t \Delta_i(u)\frac{m_i}{\Lambda_0(T_i)}\Psi_{3i}(u)'d\mathcal{A}_{20}(u),$$
$$J_4(t,s) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\int_0^t \Delta_i(s)\Delta_i(u)\frac{m_i}{\Lambda_0(s)}\Psi_{4i}(u)'d\mathcal{A}_{20}(u),$$
$$J_5(t,s) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\int_0^t \Delta_i(s)\Delta_i(u)m_i\left[-\frac{d\mathcal{A}_{10}(u)}{\Lambda_0^2(s)} + \left\{\Psi_{5i}(u) - \frac{B_i(u)}{\Lambda_0(s)}\right\}'\frac{d\mathcal{A}_{20}(u)}{\Lambda_0(s)}\right],$$
$$\kappa_1(s) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Delta_i(s)\frac{m_i}{\Lambda_0(T_i)},$$
$$\kappa_2(s) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Delta_i(s)\frac{m_i}{\Lambda_0(T_i)}B_i(s).$$

Similarly,

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^t B_i(s)dM_i(s) = \frac{1}{n}\sum_{i=1}^{n}L_{2i}(t) + \int_0^t \kappa_2(s)d\{\mathcal{A}_1(s;\beta_0,\hat{\Lambda}_0,\hat{B}) - \mathcal{A}_{10}(s)\}$$
$$+\int_0^t \kappa_3(s)d\{\mathcal{A}_2(s;\beta_0,\hat{\Lambda}_0,\hat{B}) - \mathcal{A}_{20}(s)\} + o_p(n^{-\frac{1}{2}}), \tag{A.13}$$

where

$$L_{2i}(t) = J_1^*(t)\xi_{1i} + J_2^*(t)\xi_{2i} + J_3^*(t)\xi_{3i} + \int_0^\tau J_4^*(t,s)\,d\psi_{3i}(s) + \int_0^\tau J_5^*(t,s)d\psi_{4i}(s)$$

with

$$J_1^*(t) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\int_0^t \Delta_i(u)\frac{m_i}{\Lambda_0(T_i)}B_i(u)\{\Psi_{1i}(u)'d\mathcal{A}_{20}(u)\}',$$
$$J_2^*(t) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\int_0^t \Delta_i(u)\frac{m_i}{\Lambda_0(T_i)}B_i(u)\{\Psi_{2i}(u)'d\mathcal{A}_{20}(u)\}',$$
$$J_3^*(t) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\int_0^t \Delta_i(u)\frac{m_i}{\Lambda_0(T_i)}B_i(u)\Psi_{3i}(u)'d\mathcal{A}_{20}(u),$$
$$J_4^*(t,s) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\int_0^t \Delta_i(s)\Delta_i(u)\frac{m_i}{\Lambda_0(t)}B_i(u)\Psi_{4i}(u)'d\mathcal{A}_{20}(u),$$

$$J_5^*(t,s) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \int_0^t \Delta_i(s)\Delta_i(u) m_i B_i(u)\bigg[-\frac{d\mathcal{A}_{10}(u)}{\Lambda_0^2(s)} + \bigg\{\Psi_{5i}(u) - \frac{B_i(u)}{\Lambda_0(s)}\bigg\}'\frac{d\mathcal{A}_{20}(u)}{\Lambda_0(s)}\bigg],$$

$$\kappa_3(s) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \Delta_i(s)\frac{m_i}{\Lambda_0(T_i)}B_i(s)B_i(s)'.$$

It follows from (A.12) and (A.13) that

$$\mathcal{A}_1(t;\beta_0,\hat{\Lambda}_0,\hat{B}) - \mathcal{A}_{10}(t) = \frac{1}{n}\sum_{i=1}^n \Gamma_{1i}(t) + o_p(n^{-\frac{1}{2}}), \tag{A.14}$$

$$\mathcal{A}_2(t;\beta_0,\hat{\Lambda}_0,\hat{B}) - \mathcal{A}_{20}(t) = \frac{1}{n}\sum_{i=1}^n \Gamma_{2i}(t) + o_p(n^{-\frac{1}{2}}), \tag{A.15}$$

where

$$\begin{aligned}\Gamma_{1i}(t) = \int_0^t &\{\kappa_1(s) - \kappa_2(s)'\kappa_4(s)^{-1}\kappa_3(s)\}^{-1}[\{1-\kappa_2(s)\kappa_4(s)^{-1}B_i(s)\}dM_i(s)\\ &- dL_{1i}(s) + \kappa_2(s)'\kappa_4(s)^{-1}dL_{2i}(s)],\\ \Gamma_{2i}(t) = \int_0^t &\{\kappa_4(s) - \kappa_3(s)\kappa_1(s)^{-1}\kappa_2(s)'\}^{-1}[\{B_i(s) - \kappa_3(s)\kappa_1(s)^{-1}\}dM_i(s)\\ &+ \{\kappa_3(s)\kappa_1(s)^{-1}dL_{1i}(s)\} - dL_{2i}(s)].\end{aligned}$$

In a similar manner, we have

$$\hat{K}_1(t) - K_{10}(t) = \frac{1}{n}\sum_{i=1}^n \Gamma_{3i}(t) + o_p(n^{-\frac{1}{2}}), \tag{A.16}$$

$$\hat{K}_2(t) - K_{20}(t) = \frac{1}{n}\sum_{i=1}^n \Gamma_{4i}(t) + o_p(n^{-\frac{1}{2}}), \tag{A.17}$$

where $\Gamma_{3i}(t)$ and $\Gamma_{4i}(t)$ are obtained by replacing $M_i(t)$, $\mathcal{A}_{10}$ and $\mathcal{A}_{20}$ in $\Gamma_{1i}(t)$ and $\Gamma_{2i}(t)$ with $M_i^*(t)$, K_{10} and K_{20}, respectively. Let $\bar{y}^*(t)$ and $\bar{x}(t)$ be the limits of $\bar{Y}^*(t)$ and $\bar{X}(t)$, respectively. Thus, it follows from (A.11) and (A.14)–(A.17) that

$$U(\beta_0;\hat{\Lambda}_0,\hat{B}) = \sum_{i=1}^n \phi_i + o_p(n^{1/2}), \tag{A.18}$$

where

$$\begin{aligned}\phi_i = &\int_0^\tau X_i(t)\,dM_i(t) - \int_0^\tau \{\bar{y}^*(t) - \beta_0'\bar{x}(t)\}X_i(t)dM_i^*(t) + P_1\xi_{i1} + P_2\xi_{i2} + P_3\xi_{i3}\\ &+ \int_0^\tau P_4(t)d\psi_{i3}(t) + \int_0^\tau P_5(t)d\psi_{i4}(t) + \int_0^\tau \Phi_1(t)d\Gamma_{1i}(t) + \int_0^\tau \Phi_2(t)d\Gamma_{2i}(t)\\ &- \int_0^\tau \{\bar{y}^*(t) - \beta_0'\bar{x}(t)\}[\Phi_1(t)d\Gamma_{3i}(t) + \Phi_2(t)d\Gamma_{4i}(t)],\end{aligned}$$

with

$$P_1 = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\int_0^\tau \Delta_i(s)\frac{m_iX_i(s)}{\Lambda_0(T_i)}[\{\bar{y}^*(s) - \beta_0'\bar{x}(s)\}\Psi_{1i}(s)'dK_{20}(s) - \Psi_{1i}(s)'d\mathcal{A}_{20}(s)]',$$

$$P_2 = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\int_0^\tau \Delta_i(s)\frac{m_iX_i(s)}{\Lambda_0(T_i)}[\{\bar{y}^*(s) - \beta_0'\bar{x}(s)\}'\Psi_{2i}(s)'dK_{20}(s) - \Psi_{2i}(s)'d\mathcal{A}_{20}(s)]',$$

$$P_3 = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\int_0^\tau \Delta_i(s)\frac{m_iX_i(s)}{\Lambda_0(T_i)}[\{\bar{y}^*(s) - \beta_0'\bar{x}(s)\}\Psi_{3i}(s)'dK_{20}(s) - \Psi_{3i}(s)'d\mathcal{A}_{20}(s)],$$

$$P_4(t) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} \Delta_i(t)\Delta_i(s)\frac{m_iX_i(s)}{\Lambda_0(t)}[\{\bar{y}^*(s) - \beta_0'\bar{x}(s)\}\Psi_{4i}(s)'dK_{20}(s) - \Psi_{4i}(s)'d\mathcal{A}_{20}(s)],$$

$$P_5(t) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} \Delta_i(t)\Delta_i(s)\frac{m_iX_i(s)}{\Lambda_0(t)}[\{\bar{y}^*(s) - \beta_0'\bar{x}(s)\}\Psi_{5i}(s)'dK_{20}(s) - \Psi_{5i}(s)'d\mathcal{A}_{20}(s)]$$
$$+ \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} \Delta_i(t)\Delta_i(s)\frac{m_iX_i(s)}{\Lambda_0^2(t)}[d\mathcal{A}_1(s) + B_i(s)'d\mathcal{A}_2(s) - \{\bar{y}^*(s) - \beta_0'\bar{x}(s)\}$$
$$\times [dK_1(s) + B_i(s)'dK_2(s)]],$$

$$\Phi_1(t) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\Delta_i(t)\frac{m_i}{\Lambda_0(T_i)}X_i(t),$$

$$\Phi_2(t) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\Delta_i(t)\frac{m_i}{\Lambda_0(T_i)}X_i(t)B_i(t)'.$$

By the multivariate central limit theorem, $n^{-1/2}U(\beta_0;\hat{\Lambda}_0,\hat{B})$ is asymptotically normal with mean zero and covariance matrix $\Sigma = E\{\phi_i\phi_i'\}$. Note that $-\partial U(\beta;\hat{\Lambda}_0,\hat{B})/\partial\beta|_{\beta=\beta_0}$ converges in probability to D. Thus, the Taylor expansion of $U(\beta;\hat{\Lambda}_0,\hat{B})$ yields that $n^{1/2}(\hat{\beta}-\beta_0)$ is asymptotically zero-mean normal with covariance matrix $D^{-1}\Sigma(D')^{-1}$.

To show the weak convergence of $n^{1/2}\{\hat{\mathcal{A}}_1(t)-\mathcal{A}_{10}(t)\}$ and $n^{1/2}\{\hat{\mathcal{A}}_2(t)-\mathcal{A}_{20}(t)\}$, first note that $R_1(t;\hat{B})$ converges in probability to $R_1^*(t)$ uniformly in $t\in[0,\tau]$, where

$$R_1^*(t) = \begin{pmatrix} \kappa_1(t) & \kappa_2(t)' \\ \kappa_2(t) & \kappa_3(t) \end{pmatrix}.$$

It can be checked that the derivative of $(\hat{\mathcal{A}}_1(t),\hat{\mathcal{A}}_2(t)')'$ with respect to β evaluated at β_0 converge in probability to $\Upsilon(t)$ uniformly in $t\in[0,\tau]$, where

$$\Upsilon(t) = \int_0^t R_1^*(s)\begin{pmatrix}\Phi_1(s)' \\ \Phi_2(s)'\end{pmatrix}d\Lambda_0(s).$$

It follows from the Taylor expansion, (A.14), (A.15) and (A.18) that

$$n^{1/2}\begin{pmatrix} \hat{\mathcal{A}}_1(t)-\mathcal{A}_{10}(t) \\ \hat{\mathcal{A}}_2(t)-\mathcal{A}_{20}(t) \end{pmatrix} = n^{1/2}\begin{pmatrix} \hat{\mathcal{A}}_1(t;\beta_0,\hat{\Lambda},\hat{B})-\mathcal{A}_{10}(t) \\ \hat{\mathcal{A}}_2(t;\beta_0,\hat{\Lambda},\hat{B})-\mathcal{A}_{20}(t) \end{pmatrix} + \Upsilon(t)n^{1/2}(\hat{\beta}-\beta_0)$$
$$= n^{-1/2}\sum_{i=1}^{n}\Theta_i(t) + o_p(1),$$

where

$$\Theta_i(t) = \begin{pmatrix}\Gamma_{1i}(t) \\ \Gamma_{2i}(t)\end{pmatrix} + \Upsilon(t)D^{-1}\phi_i.$$

Because $\Theta_i(t)$ $(i=1,\dots,n)$ are independent zero-mean random variables for each t, the multivariate central limit theorem implies that $n^{1/2}\{\hat{\mathcal{A}}_1(t)-\mathcal{A}_{10}(t)\}$ and $n^{1/2}\{\hat{\mathcal{A}}_2(t)-\mathcal{A}_{20}(t)\}$ jointly converge in finite-dimensional distributions to a zero-mean Gaussian process. Since $\Theta_i(t)$ can be written as sums or products of monotone functions of t and are thus tight [24]. Thus, $n^{1/2}\{\hat{\mathcal{A}}_1(t)-\mathcal{A}_{10}(t)\}$ and $n^{1/2}\{\hat{\mathcal{A}}_2(t)-\mathcal{A}_{20}(t)\}$ are tight and jointly converge weakly to a zero-mean Gaussian vector process whose covariance function at (s,t) is given by $E\{\Theta_i(s)\Theta_i(t)'\}$.