Biostatistics Advance Access published October 1, 2013

Biostatistics (2013), *pp.* 1–14
doi:10.1093/biostatistics/kxt040

Sieve estimation in semiparametric modeling of longitudinal data with informative observation times

XINGQIU ZHAO*
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong
xingqiu.zhao@polyu.edu.hk

SHIRONG DENG, LI LIU
School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, China

LEI LIU
Department of Preventive Medicine and Robert H. Lurie Cancer Center, Northwestern University, Chicago, IL 60611, USA

Summary

Analyzing irregularly spaced longitudinal data often involves modeling possibly correlated response and observation processes. In this article, we propose a new class of semiparametric mean models that allows for the interaction between the observation history and covariates, leaving patterns of the observation process to be arbitrary. For inference on the regression parameters and the baseline mean function, a spline-based least squares estimation approach is proposed. The consistency, rate of convergence, and asymptotic normality of the proposed estimators are established. Our new approach is different from the usual approaches relying on the model specification of the observation scheme, and it can be easily used for predicting the longitudinal response. Simulation studies demonstrate that the proposed inference procedure performs well and is more robust. The analyses of bladder tumor data and medical cost data are presented to illustrate the proposed method.

Keywords: Asymptotic normality; Estimating equation; Informative observation process; Longitudinal medical costs; Polynomial spline.

1. Introduction

Longitudinal data occur frequently in a wide variety of settings, including epidemiological studies, clinical trials, and economic applications. While response variables are observed repeatedly at irregular time points for different subjects under study, the observations are independent between different subjects and may be correlated within each subject. Examples of such data include cancer recurrence and longitudinal medical costs, which will be described below.

*To whom correspondence should be addressed.

For the analysis of longitudinal data, parametric regression models have been studied by Laird and Ware (1982) and Liang and Zeger (1986) among others, and an excellent review has been provided by Diggle *and others* (2002). In addition, a number of semiparametric models with nice features have been considered for modeling longitudinal data. Zeger and Diggle (1994) proposed a semiparametric mixed model for longitudinal data and suggested a backfitting procedure for inference. Lin and Ying (2001) developed a novel and simple semiparametric and non-parametric method for the regression analysis of irregularly spaced longitudinal data by formulating the observation times within the framework of counting processes. A basic assumption behind these methods is that observation times are independent of the response variable, completely or given covariates. However, such an assumption can be violated in many applications, such as an example given by a set of longitudinal data arising from a bladder cancer follow-up study conducted by the Veterans Administration Cooperative Urological Research Group (Byar, 1980). All patients had superficial bladder tumors when they entered the study and these tumors were removed transurethrally, and then patients were randomly allocated to one of the three treatments, placebo, thiotepa, and pyridoxine. Many patients had multiple recurrences of new tumors during the study. One problem with the data set is that some patients in the study had significantly more clinical visits than others (Sun *and others*, 2005). This indicates that the number of clinical visits may contain some information about the tumor occurrence rate. Another example can be found in the longitudinal (monthly) medical costs of chronic heart failure (CHF) patients from the clinical data repository (CDR) at the University of Virginia Health System (Liu *and others*, 2008). One phenomenon from some preliminary analysis is that the patients visiting hospital more often tended to pay more for each visit, that is, the level of medical costs is associated with the frequency of observation times. Thus, an important question is how to take into account or make use of this information for inference about the tumor recurrence rate. To investigate this problem, two methods have been developed. One is the conditional modeling approach proposed by Sun *and others* (2005); another is the frailty-based approach proposed by Sun *and others* (2007), Liang *and others* (2009), Zhao *and others* (2012), among others. A common and key assumption in these two approaches is that the observation process follows a Poisson or mixed Poisson with the proportional intensity function. However, the fit of the Poisson model may be inadequate when the observation process displays under-dispersion or over-dispersion. In addition, the relation between the observation and response processes may vary with some covariates. For example, in the bladder cancer study, patients who received the thiotepa treatment may have less superficial bladder tumors, and thus may visit the doctor less often than those in the placebo group, which means that the correlation between the observation times and the tumor recurrence process may be different for different treatment groups. In the medical cost data, non-white patients were more likely to visit hospital, and paid more for their visits, which indicates that the patients' medical costs and visiting times are related with the race (Liu *and others*, 2008).

In this article, motivated by the characteristic of the two longitudinal data sets mentioned above, we propose a new class of semiparametric regression models that allows for the interaction between the observation history and some covariates, while leaving the patterns of the observation times to be arbitrary. For the non-parametric estimation of the baseline mean function, a B-spline approximation will be used following Huang (1999) and Lu *and others* (2007, 2009).

The remainder of this paper is organized as follows. We begin in Section 2 by introducing notation and describing models for longitudinal data. In Section 3, a spline-based least squares method is proposed for estimation of regression parameters and the baseline mean function. Section 3 also presents the asymptotic properties of the proposed estimators, including the consistency, rate of convergence, and asymptotic normality. In order to assess the finite-sample performance of the proposed inference procedure, we present some results obtained from simulation studies in Section 4. In Section 5, the proposed approaches are illustrated through the analysis of two data sets from a bladder tumor study and longitudinal medical costs. Some concluding remarks are made in Section 6.

2. Statistical model

Consider a longitudinal study that consists of a random sample of n subjects. For subject i, let $Y_i(t)$ denote the response variable and $\mathbf{X}_i$ denote a p-dimensional vector of covariates, $i = 1, \ldots, n$. Suppose that $Y_i(t)$ is observed at distinct time points $T_{K_i,1} < T_{K_i,2} < \cdots < T_{K_i,K_i}$, where K_i is the total number of observations on subject i. In the following, we regard these observation times arising from an underlying counting process $N_i^*(t)$ characterized by $N_i(t) = \sum_{j=1}^{K_i} I(T_{K_i,j} \leqslant t) = N_i^*(\min(t, C_i))$, where $I(\cdot)$ is the indicator function, and C_i is the follow-up or censoring time with $K_i = N_i^*(C_i)$ for subject i, $i = 1, \ldots, n$. Then, the process $Y_i(t)$ is observed only at the time points where $N_i(t)$ jumps.

Define $\mathcal{F}_{it} = \{N_i(s), 0 \leqslant s < t\}$. For the analysis, we assume that $Y_i(t)$ follows the marginal model

$$E\{Y_i(t)|\mathbf{X}_i, W_i, \mathcal{F}_{it}\} = \mu_0(t) + \beta'\mathbf{X}_i + \alpha' H(\mathcal{F}_{it}, W_i), \tag{2.1}$$

given $\mathbf{X}_i$, $\mathcal{F}_{it}$, and the covariate W_i, which is allowed to be a component of the vector $\mathbf{X}_i$, where $\mu_0(t)$ is an unspecified smooth function of t, β is a p-dimensional vector of unknown regression parameters, α is a q-dimensional vector of regression coefficients, and $H(\cdot)$ is a vector of known functions of the counting process $N_i(t)$ up to $t-$ and the covariate W_i, representing the interaction between the observation history and some covariates. In particular, in longitudinal follow-up clinical studies with different treatments, W_i's can be defined as the treatment indicators, and thus α represents the effect of interaction between the frequency of observation times and treatment group on the longitudinal response variable. In fact, our modeling approach is different from usual approaches. Here, the possible effect of the observation process is directly incorporated into the conditional model about the longitudinal process, no additional model assumption is needed for the observation process, and the fitted conditional model can be useful for prediction in longitudinal medical cost studies.

The model (2.1) specifies that the process $Y_i(t)$ depends on the observation process $N_i(t)$ through function H, which can be chosen according to situations. Following the discussion in Sun *and others* (2005), a natural and simple choice for H may be $H(\mathcal{F}_{it}, W_i) = N_i(t-)W_i$, which means that $Y_i(t)$ and $\mathcal{F}_{it}$ are related through the total number of observations made before time t and their relation may vary with covariate W_i. An alternative is that $Y_i(t)$ depends on $\mathcal{F}_{it}$ only through a recent number of observations, say, in u time units, and this corresponds to $H(\mathcal{F}_{it}, W_i) = (N_i(t-) - N_i(t-u))W_i$. One could define H as a vector given by the forgoing two choices if both the total and recent numbers of observations may contain information about $N_i(t)$.

In addition, we assume that

$$E\{Y_i(t)|\mathbf{X}_i, W_i, N_i(s), 0 \leqslant s \leqslant t\} = E\{Y_i(t)|\mathbf{X}_i, W_i, \mathcal{F}_{it}\},$$

which means that conditional on the covariates and the follow-up time, the mean of the response variable at time point t is only related to the observation history before t. The observation for each individual consists of $\mathbf{O} = (K, \tilde{T}_K, \tilde{Y}_K, \tilde{N}_K, \mathbf{X}, W, C)$, with $\tilde{T}_K = (T_{K,1}, \ldots, T_{K,K})$, $\tilde{Y}_K = (Y(T_{K,1}), \ldots, Y(T_{K,K}))$, $\tilde{N}_K = (N(T_{K,1}), \ldots, N(T_{K,K}))$. Throughout this paper, we will assume that we observe n i.i.d. copies, $\mathbf{O}_1, \ldots, \mathbf{O}_n$ of $\mathbf{O}$. The main purpose here is to estimate the regression coefficients α, β and the smooth baseline mean function $\mu_0(t)$.

3. ESTIMATION PROCEDURE

For inference about model (2.1), we define

$$L_n(\beta, \alpha, \mu) = \sum_{i=1}^{n}\sum_{j=1}^{K_i}[Y_i(T_{K_i,j}) - \mu(T_{K_i,j}) - \beta'\mathbf{X}_i - \alpha' H(\mathcal{F}_{iT_{K_i,j}}, W_i)]^2$$
$$= \sum_{i=1}^{n}\int_0^{\tau}\{Y_i(t) - \mu(t) - \beta'\mathbf{X}_i - \alpha' H(\mathcal{F}_{it}, W_i)\}^2 \,\mathrm{d}N_i(t). \quad (3.1)$$

To make an inference about $\mu_0(t)$, we propose to use B-splines to approximate it. For a finite closed interval $[0, \tau]$, let $\mathcal{I} = \{t_i\}_1^{m_n+2l}$, with

$$0 = t_1 = \cdots = t_l < t_{l+1} < \cdots < t_{m_n+l} < t_{m_n+l+1} = \cdots = t_{m_n+2l} = \tau$$

be a sequence of knots that partition $[0, \tau]$ into $m_n + 1$ subintervals and $m_n = O(n^{\upsilon})$ for $0 < \upsilon < 1/2$. Let $\{B_{il}, 1 \leqslant i \leqslant q_n\}$ denote the B-spline basis functions with $q_n = m_n + l$. Let $\Psi_{l,\mathcal{I}}$ (with order l and knots $\mathcal{I}$) be the class linearly spanned by the B-spline functions, that is,

$$\Psi_{l,\mathcal{I}} = \left\{\sum_{i=1}^{q_n} \gamma_i B_{il} : \gamma_i \in \mathbb{R},\ i = 1, \ldots, q_n\right\}.$$

Assume that $\mu_0(t)$ has a bounded rth derivative. According to Schumaker (1981), there exists a smooth spline $\mu_n(t) \in \Psi_{l,\mathcal{I}}$ such that

$$\| \mu_n - \mu_0 \|_{\infty} = \sup_{t\in[0,\tau]} |\mu_n(t) - \mu_0(t)| = O(n^{-\upsilon r}).$$

Define $\mu_n(t) = \gamma' B_l(t)$, where $\gamma = (\gamma_1, \ldots, \gamma_{q_n})'$ and $B_l(t) = (B_{1l}(t), \ldots, B_{q_n l}(t))'$. We see that $L_n(\beta, \alpha, \mu)$ in (3.1) is approximate to

$$L_n(\beta, \alpha, \gamma) = \sum_{i=1}^{n}\int_0^{\tau}\{Y_i(t) - \gamma' B_l(t) - \beta'\mathbf{X}_i - \alpha' H(\mathcal{F}_{it}, W_i)\}^2 \,\mathrm{d}N_i(t).$$

The resulting estimating function for β, α, and γ has the form

$$U(\beta, \alpha, \gamma) = \sum_{i=1}^{n}\int_0^{\tau}\begin{pmatrix}\mathbf{X}_i \\ H(\mathcal{F}_{it}, W_i) \\ B_l(t)\end{pmatrix} \times \{Y_i(t) - \gamma' B_l(t) - \beta'\mathbf{X}_i - \alpha' H(\mathcal{F}_{it}, W_i)\} \,\mathrm{d}N_i(t),$$

The solution to $U(\beta, \alpha, \gamma) = 0$ has a closed form

$$\begin{pmatrix}\hat\beta_n \\ \hat\alpha_n \\ \hat\gamma_n\end{pmatrix} = \left[\sum_{i=1}^{n}\int_0^{\tau}\begin{pmatrix}\mathbf{X}_i \\ H(\mathcal{F}_{it}, W_i) \\ B_l(t)\end{pmatrix}^{\otimes 2} \mathrm{d}N_i(t)\right]^{-1}\left[\sum_{i=1}^{n}\int_0^{\tau}\begin{pmatrix}\mathbf{X}_i \\ H(\mathcal{F}_{it}, W_i) \\ B_l(t)\end{pmatrix} Y_i(t)\,\mathrm{d}N_i(t)\right].$$

Then the resulting estimator for $\mu_0(t)$ is $\hat\mu_n(t) = \hat\gamma_n' B_l(t)$.

Let $\theta_0 = (\beta_0, \alpha_0, \mu_0)$ be the true value of $\theta = (\beta, \alpha, \mu)$, and $\hat\theta_n = (\hat\beta_n, \hat\alpha_n, \hat\mu_n)$ be the estimator of θ_0. Then one can show that under conditions C1–C5 stated in the Appendix, $\hat\theta_n$ is consistent with the $n^{(1-\upsilon)/2}$ rate of convergence. When $\upsilon = 1/(1 + 2r)$, then $n^{(1-\upsilon)/2} = n^{r/(1+2r)}$, and it follows from Stone (1980, 1982) that the rate of convergence of the estimator $\hat{\mu}_n$ is the optimal rate in non-parametric regression. Furthermore, one can obtain that $\sqrt{n}((\hat{\beta}_n - \beta_0)', (\hat{\alpha}_n - \alpha_0)')'$ converges in distribution to $N(\mathbf{0}, \Sigma)$, where $\Sigma = A^{-1}B(A^{-1})'$ with

$$A = E\left[\sum_{j=1}^{K}\left\{\begin{pmatrix}\mathbf{X}\\ H(\mathcal{F}_{T_{K,j}}, W)\end{pmatrix} - E\left(\begin{pmatrix}\mathbf{X}\\ H(\mathcal{F}_{T_{K,j}}, W)\end{pmatrix}\middle| K, T_{K,j}\right)\right\}^{\otimes 2}\right]$$

and

$$\begin{aligned}B = E\Bigg[&\sum_{j=1}^{K}\sum_{j'=1}^{K}\{Y(T_{K,j}) - \mu_0(T_{K,j}) - \beta_0'\mathbf{X} - \alpha_0' H(\mathcal{F}_{T_{K,j}}, W)\}\\ &\times \{Y(T_{K,j'}) - \mu_0(T_{K,j'}) - \beta_0'\mathbf{X} - \alpha_0' H(\mathcal{F}_{T_{K,j'}}, W)\}\\ &\times \left\{\begin{pmatrix}\mathbf{X}\\ H(\mathcal{F}_{T_{K,j}}, W)\end{pmatrix} - E\left(\begin{pmatrix}\mathbf{X}\\ H(\mathcal{F}_{T_{K,j}}, W)\end{pmatrix}\middle| K, T_{K,j}\right)\right\}\\ &\times \left\{\begin{pmatrix}\mathbf{X}\\ H(\mathcal{F}_{T_{K,j'}}, W)\end{pmatrix} - E\left(\begin{pmatrix}\mathbf{X}\\ H(\mathcal{F}_{T_{K,j'}}, W)\end{pmatrix}\middle| K, T_{K,j'}\right)\right\}'\Bigg].\end{aligned}$$

Also, one can derive the asymptotic normality of linear functionals of $\hat{\mu}_n$. The proofs are outlined in the Appendix of supplementary material available at *Biostatistics* online.

4. Simulation studies

In this section, simulation studies were conducted to assess the finite sample properties of the proposed estimators. We generated the response variable from the following random-effects model:

$$Y_i(t) = \mu_0(t) + \beta_1 X_{1i} + \beta_2 X_{2i} + \alpha H(\mathcal{F}_{it}, W_i) + \epsilon_i(t), \tag{4.1}$$

where X_{1i} and X_{2i} are generated from Bernoulli distribution with success probability 0.5 and the uniform distribution over $(-1, 1)$, respectively, $\epsilon_i(t)$'s are independent standard normal variables, and $H(\mathcal{F}_{it}, W_i) = N_i(t-)W_i$ with $W_i = X_{1i}$. The follow-up time C_i is generated from the uniform distribution over interval $(\tau/2, \tau)$ with $\tau = 6$.

For the objective of the study, we considered two cases of the observation process as follows:

Case 1. The number of observation times K_i was assumed to follow the Poisson distribution with mean $C_i \exp(-0.25X_{1i} + 0.5X_{2i})$ and the observation times $(T_{K_i,1}, \ldots, T_{K_i,K_i})$ were taken to be the order statistics of a random sample of size K_i from the uniform distribution over $(0, C_i)$. That is, the observation process satisfies the assumed model by Sun *and others* (2005).

Case 2. The K_i was assumed to follow the uniform distribution over $\{1, 2, 3\}$ if $X_{1i} = 0$ and the uniform distribution over $\{4, 5, 6\}$ otherwise, and the observation times $(T_{K_i,1}, \ldots, T_{K_i,K_i})$ were generated in the same way as in Case 1.

The true parameter values used in our simulation studies are $\beta_0 = (\beta_{10}, \beta_{20})' = (-1, 1)'$, $\alpha_0 = -1.5, -1, 0, 1$ or 1.5, and $\mu_0(t) = \log(1+t)$. To estimate $\mu_0(t)$, the cubic B-splines are used in computing the spline estimators. To choose the number of interior knots, we partitioned the range of v, $(0, 0.5)$ into 20 equal subintervals and chose v to be the partition points. For each value of v, we took $m_n = n^v$ denoting the number of interior knots. For determining locations of knots, there are two commonly used data-driven methods. One is the equally spaced knots, which are given by

$T_{\min} + k(T_{\max} - T_{\min})/(m_n + 1)$, $k = 0, 1, \ldots, m_n + 1$, with $T_{\min}$ and $T_{\max}$ being the respective minimum and maximum values of distinct observation times. Another is the partitions according to quantiles of the observation times, i.e. the $k/(m_n + 1)$ quantiles ($k = 0, 1, \ldots, m_n + 1$) of the distinct observation times are chosen to be the knots. The value of υ that minimizes the Bayesian information criteria (BIC) was selected. Here $\mathrm{BIC} = 2\log(L) + \log(n)(q_n + 3)$ where $L = L_n(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n)$ as defined in (3.1) and q_n is the number of the B-spline basis functions. We carried out simulations for the different situations of the number and placement of knots with $W = X_1$, $\alpha = 1$, $\mu_0(t) = \log(1 + t)$, and $n = 50$, and found that the estimation results are very similar and the method is insensitive to the selection of number and placement of knots, where the value of υ was selected by the BIC as $\frac{1}{8}$. Thus, in the following, we took the number of interior knots as $n^{1/8}$ and equally spaced knots.

Tables 1 and 2 present the simulation results on estimation of β_0 and α_0 with the sample size $n = 50$ or 100 in the two cases. In the tables, we compared the proposed method with a competing method developed by Sun *and others* (2005), to demonstrate the robustness of the proposed method. All the tables include the estimated bias (BIAS) given by the average of the estimates minus the true value, the bootstrap standard errors of the estimates (BSE), the sample standard deviation of the estimates (SSE), and the bootstrap 95% coverage probabilities (CP) obtained from 1000 independent runs. Here, we used 200 replications in bootstrap to estimate the standard errors. Figure 1 shows the estimate of $\mu_0(t) = \log(1 + t)$ for case 2 of the observation process with $\alpha = 1$. In the figure, the solid line represents the true curve of $\mu_0(t)$, and the point line represents the estimated curve of $\mu_0(t)$.

Based on our simulation results, we have the following findings: (i) In case 1 of the observation process, it can be seen from Table 1 that the proposed method performs slightly better than the method of Sun *and others* (2005). The possible reason is that the estimating function for β and α proposed by Sun *and others* (2005) include the estimators of observation process model parameters and the approximated longitudinal outcomes as the measurement at the time point nearest to t. Our method is designed for directly estimating the longitudinal model parameters and is easy to implement. So as a result, both the estimated bias and standard error of our estimators may be slightly smaller than those obtained by the method of Sun *and others* (2005). (ii) In case 2 of the observation process, both the proposed estimator and the SPSZ's estimator are approximately unbiased when $\alpha = 0$. When $\alpha \neq 0$, the proposed estimator is approximately unbiased while the SPSZ's estimator yields biased estimates and the bias could be larger as α diverges from 0. In other words, the proposed estimation procedure seems to be more robust. The possible reason is that our estimation method is model-free for the observation process, while their estimation procedure relies on the model assumption about the observation process. (iii) The estimated curve of $\mu_0(t)$ is close to its true curve with the moderate sample size, indicating that the B-spline estimator for $\mu_0(t)$ works well. (iv) The sample standard errors and the bootstrap standard errors of the proposed estimators are close to each other. Also, the bootstrap 95% coverage rates are close to the nominal level, that is, the proposed spline semiparametric bootstrap procedure provides reasonable estimates and the normal approximation seems to be appropriate.

We also conducted some sensitivity analysis to evaluate the performance of the proposed estimators for β and $\mu_0(t)$ when the interaction term is misspecified. Specifically, we generated $Y_i(t)$ from model (4.1) with $\mu_0(t) = \log(1 + t)$ and the interaction term $H(\mathcal{F}_{it}, W_i) = (N_i(t-) - N_i(t - 4))X_{1i}$, where X_{1i}, X_{2i}, C_i, and the observation times were generated from the same set-up as given above. We considered the interaction term H as misspecified by three possible forms: (i) $H^{(1)}(\mathcal{F}_{it}, W_i) = N_i(t-)X_{1i}$; (ii) $H^{(2)}(\mathcal{F}_{it}, W_i) = (N_i(t-) - N_i(t - 3))X_{1i}$; (iii) $H^{(3)}(\mathcal{F}_{it}, W_i) = (N_i(t-) - N_i(t - 5))X_{1i}$. We applied the proposed estimation procedure to the true and misspecified models by using the generated data from the true model. The simulation results are summarized in Table 3. It can be seen from the table that the estimators for β are still approximately unbiased for the misspecified situations considered here. We drew the figure for the estimates of $\mu_0(t)$, omitted here for the sake of space, and found that there are some discrepancies between the estimated and true baseline mean functions at later times. The possible reason

Table 1. *Simulation results for the Poisson observation process in case* 1

			$\hat{\beta}_1$				$\hat{\beta}_2$				$\hat{\alpha}$			
n	α	Method	BIAS	SSE	BSE	CP	BIAS	SSE	BSE	CP	BIAS	SSE	BSE	CP
50	−1.5	Proposed	−0.0092	0.3316	0.3190	0.9410	−0.0193	0.2884	0.2771	0.9350	−0.0014	0.0956	0.0870	0.9280
		SPSZ	−0.0242	0.3330			−0.0205	0.2894			0.0172	0.0969		
	−1	Proposed	0.0028	0.3430	0.3197	0.9320	0.0011	0.2939	0.2800	0.9350	−0.0019	0.1014	0.0884	0.9120
		SPSZ	−0.0073	0.3454			0.0006	0.2946			0.0106	0.1017		
	0	Proposed	−0.0136	0.3256	0.3163	0.9460	0.0131	0.2969	0.2779	0.9390	−0.0021	0.0981	0.0859	0.9140
		SPSZ	−0.0114	0.3252			0.0139	0.2982			−0.0039	0.0966		
	1	Proposed	0.0097	0.3261	0.3219	0.9420	0.0107	0.3051	0.2837	0.9390	0.0011	0.0976	0.0892	0.9290
		SPSZ	0.0225	0.3302			0.0141	0.3060			−0.0138	0.1011		
	1.5	Proposed	−0.0020	0.3258	0.3167	0.9430	0.0010	0.2922	0.2795	0.9420	0.0001	0.0963	0.0873	0.9260
		SPSZ	0.0206	0.3295			0.0064	0.2939			−0.0221	0.0982		
100	−1.5	Proposed	0.0027	0.2358	0.2267	0.9430	−0.0036	0.2032	0.2004	0.9440	0.0010	0.0690	0.0617	0.9220
		SPSZ	−0.0087	0.2384			−0.0058	0.2038			0.0124	0.0708		
	−1	Proposed	−0.0008	0.2257	0.2288	0.9580	0.0174	0.2076	0.1987	0.9370	0.0025	0.0689	0.0625	0.9240
		SPSZ	−0.0100	0.2266			0.0161	0.2076			0.0106	0.0683		
	0	Proposed	−0.0074	0.2381	0.2284	0.9420	0.0088	0.2064	0.2016	0.9430	0.0021	0.0692	0.0619	0.9280
		SPSZ	−0.0056	0.2405			0.0103	0.2063			0.0007	0.0690		
	1	Proposed	0.0022	0.2337	0.2268	0.9490	−0.0139	0.2086	0.1985	0.9380	−0.0010	0.0669	0.0622	0.9300
		SPSZ	0.0143	0.2378			−0.0094	0.2090			−0.0112	0.0682		
	1.5	Proposed	0.0124	0.2278	0.2261	0.9520	0.0055	0.2072	0.1997	0.9340	−0.0011	0.0680	0.0623	0.9270
		SPSZ	0.0292	0.2331			0.0101	0.2090			−0.0156	0.0710		

SPSZ, method in Sun *and others* (2005).

Table 2. *Simulation results for the non-Poisson observation process in case* 2

			$\hat{\beta}_1$				$\hat{\beta}_2$				$\hat{\alpha}$			
n	α	Method	BIAS	SSE	BSE	CP	BIAS	SSE	BSE	CP	BIAS	SSE	BSE	CP
50	−1.5	Proposed	0.0057	0.3582	0.3327	0.9400	−0.0009	0.2785	0.2706	0.9490	0.0008	0.0898	0.0857	0.9330
		SPSZ	−0.2463	0.3681			−0.0010	0.2778			0.1341	0.0935		
	−1	Proposed	0.0109	0.3514	0.3348	0.9270	−0.0033	0.2819	0.2694	0.9370	−0.0005	0.0902	0.0856	0.9390
		SPSZ	−0.1466	0.3507			−0.0033	0.2821			0.0836	0.0900		
	0	Proposed	0.0039	0.3346	0.3356	0.9490	0.0027	0.2937	0.2689	0.9320	0.0012	0.0867	0.0858	0.9480
		SPSZ	0.0334	0.3341			0.0014	0.2937			−0.0130	0.0827		
	1	Proposed	−0.0130	0.3540	0.3369	0.9320	−0.0096	0.2842	0.2716	0.9490	−0.0005	0.0897	0.0853	0.9340
		SPSZ	0.2036	0.3508			−0.0110	0.2849			−0.1129	0.0874		
	1.5	Proposed	0.0145	0.3446	0.3350	0.9380	−0.0036	0.2916	0.2707	0.9320	−0.0033	0.0939	0.0863	0.9350
		SPSZ	0.3239	0.3471			−0.0005	0.2916			−0.1658	0.0946		
100	−1.5	Proposed	0.0055	0.2542	0.2364	0.9400	0.0073	0.1920	0.1905	0.9540	−0.0011	0.0623	0.0611	0.9390
		SPSZ	−0.2511	0.2569			0.0078	0.1937			0.1292	0.0650		
	−1	Proposed	0.0004	0.2435	0.2371	0.9430	0.0058	0.1889	0.1887	0.9500	0.0012	0.0623	0.0610	0.9500
		SPSZ	−0.1596	0.2441			0.0052	0.1898			0.0829	0.0604		
	0	Proposed	−0.0047	0.2435	0.2357	0.9390	−0.0017	0.2010	0.1895	0.9350	0.0017	0.0656	0.0608	0.9240
		SPSZ	0.0235	0.2422			−0.0008	0.2020			−0.0124	0.0621		
	1	Proposed	−0.0108	0.2469	0.2359	0.9460	0.0063	0.1989	0.1899	0.9330	0.0011	0.0630	0.0606	0.9390
		SPSZ	0.2126	0.2469			0.0058	0.1985			−0.1108	0.0632		
	1.5	Proposed	−0.0020	0.2404	0.2377	0.9460	0.0019	0.1968	0.1907	0.9430	−0.0013	0.0641	0.0610	0.9390
		SPSZ	0.3155	0.2411			0.0004	0.1969			−0.1608	0.0641		

SPSZ, method in Sun *and others* (2005).

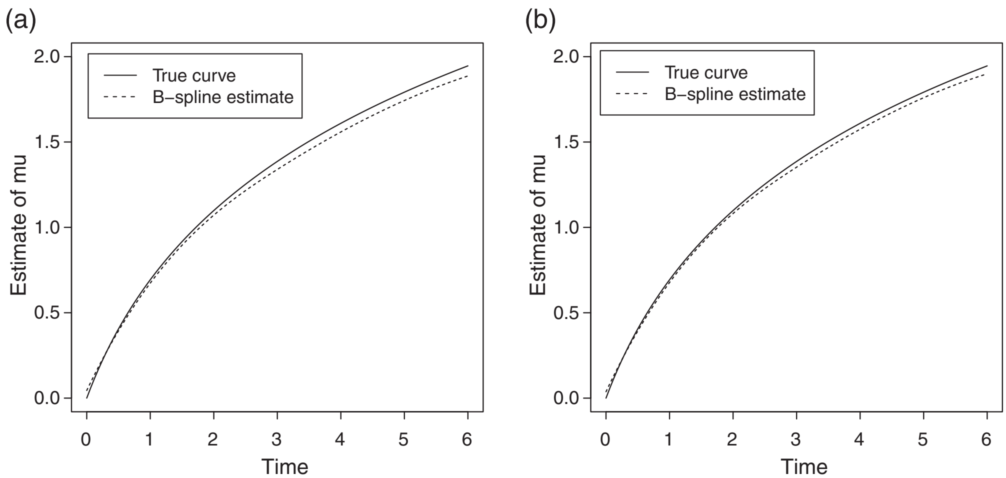

Fig. 1. Estimate of $\mu_0(t) = \log(1+t)$ with $\alpha = 1$ and $H(\mathcal{F}_{it}, W_i) = N_i(t-)X_{1i}$ under non-Poisson observation processes. (a) Sample size $n = 50$ and (b) Sample size $n = 100$.

is that less information can be available at times close to the end of study period. Thus, such misspecifications have a little influence on estimation of μ_0. We also considered other misspecified situations and obtained similar conclusions.

5. Applications

5.1 *Analysis of bladder cancer data*

This subsection presents an analysis of the bladder cancer data by applying our proposed methods. There were 116 subjects with superficial bladder tumors randomized into one of three treatment groups: placebo, thiotepa, and pyridoxine. In the following, we restrict our attention to the placebo and thiotepa groups with respective sizes of 47 and 38. For each patient, the observed information includes times when he or she made clinical visits and the numbers of recurrent tumors between clinical visits. Two baseline covariates were observed and they are the number of initial tumors and the size of the largest initial tumor.

To analyze the data, for patient i, define x_{1i} to be equal to 1 if the ith patient was given the thiotepa treatment and 0 otherwise, x_{2i} the number of initial tumors, and x_{3i} the size of the largest initial tumor, with $i = 1, \ldots, 85$. We define the response $Y_i(t)$ to be the natural logarithm of the cumulated new tumor numbers of patient i up to time t plus 1 to avoid 0. Let $N_i(\cdot)$ represent the accumulated observation numbers of patient i over the study period. Assume that $\{Y_i(t)\}$ can be described by model (2.1) with $H(\mathcal{F}_{it}, W_i) = N_i(t-)X_{1i}$, meaning that the relation between recurrence rate of bladder tumors and observation times may vary with different treatments, i.e.

$$E\{Y_i(t)|X_{1i}, X_{2i}, X_{3i}, \mathcal{F}_{it}\} = \mu_0(t) + \beta_1' X_{1i} + \beta_2' X_{2i} + \beta_3' X_{3i} + \alpha' N_i(t-)X_{1i}.$$

Here, we took the last visit time of patient i as C_i in the analysis. For estimation of $\mu_0(t)$, we use the cubic B-spline approximation by taking the number of interior knots m_n as n^v with $v = 0.1$.

The application of the estimation procedure proposed in the previous sections gave $\hat{\beta}_1 = -0.3445$, $\hat{\beta}_2 = 0.1730$, $\hat{\beta}_3 = -0.0325$, and $\hat{\alpha} = -0.0288$ with the bootstrap standard errors being 0.1369, 0.0450, 0.0470,

Table 3. *Sensitivity analysis for the misspecification of the interaction term H*

			$\hat{\alpha}$				$\hat{\beta}_1$				$\hat{\beta}_2$			
	n	α	BIAS	BIAS$^{(1)}$	BIAS$^{(2)}$	BIAS$^{(3)}$	BIAS	BIAS$^{(1)}$	BIAS$^{(2)}$	BIAS$^{(3)}$	BIAS	BIAS$^{(1)}$	BIAS$^{(2)}$	BIAS$^{(3)}$
Poisson	50	−1	0.0019	−0.0208	0.1416	−0.0183	−0.0091	−0.0321	−0.0746	−0.0297	0.0025	−0.0130	−0.0221	−0.0124
		0	0.0032	0.0032	0.0032	0.0032	−0.0057	−0.0057	−0.0057	−0.0057	−0.0072	−0.0072	−0.0072	−0.0072
		1	−0.0009	0.0273	−0.1351	0.0248	−0.0020	0.0207	0.0633	0.0183	0.0100	−0.0014	0.0077	−0.0020
	100	−1	0.0005	−0.0303	0.1342	−0.0269	0.0045	−0.0244	−0.0650	−0.0215	−0.0068	0.0017	−0.0089	0.0021
		0	−0.0030	−0.0030	−0.0030	−0.0030	0.0022	0.0022	0.0022	0.0022	0.0065	0.0065	0.0065	0.0065
		1	0.0037	0.0243	−0.1401	0.0210	0.0102	0.0288	0.0693	0.0258	−0.0114	0.0113	0.0219	0.0109
Non-Poisson	50	−1	−0.0017	0.0097	0.0874	0.0091	−0.0125	−0.0832	−0.0130	−0.0770	0.0024	0.0022	0.0020	0.0018
		0	−0.0030	−0.0030	−0.0030	−0.0030	−0.0030	−0.0030	−0.0030	−0.0030	0.0018	0.0018	0.0018	0.0018
		1	0.0025	−0.0157	−0.0934	−0.0152	0.0057	0.0772	0.0070	0.0710	−0.0085	0.0014	0.0016	0.0018
	100	−1	0.0024	0.0128	0.0906	0.0117	−0.0057	−0.0771	−0.0022	−0.0689	0.0052	0.0062	0.0074	0.0064
		0	−0.0001	−0.0001	−0.0001	−0.0001	0.0082	0.0082	0.0082	0.0082	0.0070	0.0070	0.0070	0.0070
		1	−0.0001	−0.0116	−0.0872	−0.0103	−0.0081	0.0719	−0.0048	0.0634	−0.0099	0.0005	−0.0006	0.0005

BIAS, the estimated bias of the parameter estimate with the true interaction term H; BIAS$^{(1)}$, the estimated bias of the parameter estimate with misspecified interaction term $H^{(1)}$; BIAS$^{(2)}$, the estimated bias of the parameter estimate with misspecified interaction term $H^{(2)}$; BIAS$^{(3)}$, the estimated bias of the parameter estimate with misspecified interaction term $H^{(3)}$.

and 0.0109, which correspond to p-values of 0.0118, 0.0001, 0.4888, and 0.0079, respectively, based on the asymptotic results of the estimators. Here $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ represent the estimated regression coefficients corresponding to the treatment indicator, the number of initial tumors, and the size of the largest initial tumor, respectively, while $\hat{\alpha}$ represents the estimated effect of the interaction between the observation process and the treatment indicator on the tumor recurrence rate. These results indicate that the response process and the interaction between the observation process and the treatment indicator are significantly negatively correlated. Just as explained in Sun *and others* (2005), there are two reasons for this finding. One is that the more often the patient visited the clinic, had tumors removed, and received treatment, the lower was the tumor recurrence rate; another one is that more visits means less time for tumor growth. Furthermore, the thiotepa treatment significantly reduces the occurrence rate of bladder tumors, and the number of initial tumors has a significant positive effect on the tumor recurrence rate. However, the occurrence rates of bladder tumors do not seem to be significantly related to the size of the largest initial tumor. These conclusions are consistent with those presented in Sun *and others* (2005, 2007) and Liang *and others* (2009). Compared with the models in Sun *and others* (2005, 2007) and Liang *and others* (2009), our fitted model may provide more information about the correlation between the tumor recurrence rate and observation times over treatment groups and also could be useful to estimate the future recurrence rate based on the observation history.

5.2 *Analysis of medical cost data*

We also apply the proposed method to analyzing medical costs for CHF patients in the United States. The data source is the CDR database from the University of Virginia Health System, available online at http://cdr.virginia.edu/cdr. This study includes a total of 1475 patients who were at least 60 years old and first diagnosed and treated in 2004 with heart failure (ICD9 diagnosis code beginning with 428). For each patient, the information recorded in the CDR included clinical visiting times and the medical cost for each hospital visit and some covariates, e.g. age, gender (male $= 1$, female $= 0$), race (white $= 1$, non-white $= 0$). The follow-up time was taken as patient's last hospital admission (up to July 31, 2006), or date of death extracted from Death Certificate Data at the Virginia Department of Vital Statistics.

Looking at the data, one can see that patients visiting hospital more often tended to pay more for each visit, which implies that observation times are informative of medical costs, as pointed out by Liu *and others* (2008). To analyze the data, following Liu *and others* (2008), for patient i, we define x_{1i} to be the square of the normalized age (centered at its mean of 72 years and divided by its standard error 7.7441), x_{2i} gender, and x_{3i} race, with $i = 1, \ldots, 1475$. We define the response $Y_i(t)$ to be the natural logarithm of the medical cost of patient i at time t. Let $N_i(\cdot)$ represent the accumulated numbers of visiting times for patient i over the study period. Define $\tilde{N}_i(t) = (N_i(t-) - 10)/8$ and assume that $\{Y_i(t)\}$ can be described by model (2.1) with $H(\mathcal{F}_{it}, W_i) = \tilde{N}_i(t)X_{2i} + \tilde{N}_i(t)X_{3i}$, meaning that the relation between medical cost and the clinical times may vary with different gender and race, i.e.

$$E\{Y_i(t)|X_{1i}, X_{2i}, X_{3i}, \mathcal{F}_{it}\} = \mu_0(t) + \beta_1' X_{1i} + \beta_2' X_{2i} + \beta_3' X_{3i} + \alpha_1' \tilde{N}_i(t) X_{2i} + \alpha_2' \tilde{N}_i(t) X_{3i}.$$

For estimation of $\mu_0(t)$, we use the cubic B-spline approximation by taking the number of interior knots m_n as n^v with $v = 0.1$.

By applying the estimation procedure proposed in the previous sections, we obtained that $\hat{\beta}_1 = -0.1335$, $\hat{\beta}_2 = 0.1655$, $\hat{\beta}_3 = -0.2598$, and $\hat{\alpha}_1 = 0.2281$, $\hat{\alpha}_2 = 0.1789$ with the bootstrap standard errors being 0.0347, 0.0788, 0.0867, 0.0599, and 0.0540, which correspond to p-values of 0.0001, 0.0357, 0.0027, 0.0001, and 0.0009, respectively, based on the asymptotic results of the estimators. Here $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ represent the estimated regression coefficients corresponding to the square of age, the gender indicator, and the race indicator, respectively, while $\hat{\alpha}_1$ and $\hat{\alpha}_2$ represent the estimated effects of the interaction between the observation process and the gender indicator on the medical cost, and the interaction between the observation process and the race indicator on the medical cost, respectively. These results indicate that: (i) each hospital medical cost is significantly decreasing with age^2. This fact may be due to two reasons. One is that the higher frequency of older patients to visit hospital diluted the medical cost for each visit; another is that the less aggressive treatment for older CHF patients resulted in lower medical cost. (ii) The gender effect is significant on the medical cost as males spent 18% more than females for each visit. (iii) Significant lower costs were spent by white patients (−0.2598 in log scale, or 22% lower in dollar value). (iv) The interaction effects between the visiting times and the gender or race indicator are significantly positively correlated with the medical cost, which can be explained by the reason that patients visiting hospital more often may have more serious conditions and thus may pay more for each visit no matter gender or race. These results are consistent with those obtained by Liu *and others* (2008) using the joint random-effects modeling approach. In fact, our fitted model can provide more insights into how the medical cost and observation times are related and also be easily used to predict the future medical cost based on the observation history.

6. Concluding remarks

For the statistical analysis of longitudinal data, we have proposed a new semiparametric model for the situations where the correlation between the response process and the observation process may vary with the covariates, including the conditional model of Sun *and others* (2005) as a special case. The new model allows for the interaction between the observation history and some components of the covariates, which is different from the joint models of Sun *and others* (2007) and Liang *and others* (2009) through latent variables to characterize the correlation between the response process and the observation times. For inference about model parameters, a spline-based least squares estimation procedure has been proposed, and the asymptotic properties of the resulting estimators of both the regression parameters and the non-parametric baseline mean function have been established. Another key difference between the approach developed here and those presented in Sun *and others* (2005, 2007) and Liang *and others* (2009) is that our method is designed for directly estimating the longitudinal model parameters but leaving the patterns of the observation times to be arbitrary and is easy to implement, whereas their estimation procedures rely on the model specification for observation processes. As demonstrated in our simulation studies, the proposed approaches are more flexible and robust. In particular, the proposed estimation procedure performs better than the method of Sun *and others* (2005) for the situations of Poisson and non-Poisson observation processes considered here. An important application of the proposed modeling approach is that the estimation results can be easily used to make a prediction about the longitudinal response process such as forecasting longitudinal medical costs. Further studies are needed to extend the proposed method to the analysis of irregularly observed longitudinal data involving both an informative observation scheme and a dependent terminal event and perform deep analysis an the medical cost data. Moreover, we can extend our model to a class of conditional time-varying coefficient models as follows:

$$E\{Y_i(t)|\mathbf{X}_i, W_i, \mathcal{F}_{it}\} = \mu_0(t) + \beta(t)'\mathbf{X}_i + \alpha(t)'H(\mathcal{F}_{it}, W_i).$$

For inference about the above model, B-spline function approximations can be used to estimate the time-varying coefficients and the smooth baseline mean function simultaneously, and then the asymptotic properties of spline-based estimators could be established by using similar arguments.

Supplementary material

Supplementary material is available online at http://biostatistics.oxfordjournals.org.

Acknowledgments

We are very grateful to the editor, Professor Geert Molenberghs, the associate editor, and the referee for their insightful comments and suggestions that greatly improved the paper. We are also very grateful to Professor Zhiliang Ying for offering his many valuable comments and suggestions during preparation of the manuscript. We thank the University of Virginia for providing us with the medical cost data. *Conflict of Interest:* Professor Lei Liu is a consultant to Celladon Corporation.

Funding

X.Z.'s research is partly supported by the Research Grant Council of Hong Kong (PolyU 5040/11P) and the Hong Kong Polytechnic University. L.L.'s research is partly supported by the Natural Science Foundation of China (11101315, 11171262). Lei Liu's research is partly supported by AHRQ grant R01 HS020263.

References

Byar, D. P. (1980). The veterans administration study of chemoprophylaxis for recurrent stage i bladder tumors: comparisons of placebo, pyridoxine and topical thiotepa. In: Pavane-Macaluso, M., Smith, P. H., Edsmyr, F. (editors), *Bladder Tumors and Other Topics in Urological Oncology*. New York: Plenum, pp. 363–370.

Diggle, P. J., Heagerty, P., Liang, K. Y. and Zeger, S. L. (2002). *Analysis of Longitudinal Data*, 2nd edition. Oxford, UK: Oxford University Press.

Huang, J. (1999). Efficient estimation of the partly linear additive Cox model. *Journal of the American Statistical Association* **27**, 1536–1563.

Laird, N M. and Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics* **38**, 963–974.

Liang, Y., Lu, W. and Ying, Z. L. (2009). Joint modeling and analysis of longitudinal data with informative observation times. *Biometrics* **65**, 377–384.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

Lin, D. Y. and Ying, Z. L. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association* **96**, 103–113.

Liu, L., Huang, X. and O'Quigley, J. (2008). Analysis of longitudinal data in the presence of informative observation times and a dependent terminal event with application to medical cost data. *Biometrics* **64**, 950–958.

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

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

Schumaker, L. (1981). *Spline Functions: Basic Theory*. New York: Wiley.

Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics* **8**, 1048–1360.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics* **10**, 1040–1053.

Sun, J., Park, D., Sun, L. and Zhao, X. (2005). Semiparametric regression analysis of longitudinal data with informative observation times. *Journal of the American Statistical Association* **100**, 882–889.

SUN, J., SUN, L. AND LIU, D. (2007). Regression analysis of longitudinal data in the presence of informative observation and censoring times. *Journal of the American Statistical Association* **102**, 1397–1406.

ZEGER, S. L. AND DIGGLE, P. J. (1994). Semiparametric models for longitudinal data with application to cd4 cell numbers in HIV seroconverters. *Biometrics* **50**, 689–699.

ZHAO, X., TONG, X. AND SUN, L. (2012). Joint analysis of longitudinal data with dependent observation times. *Statistica Sinica* **22**, 317–336.

[*Received May 12, 2013; revised July 30, 2013; accepted for publication August 30, 2013*]

Supplementary Materials for Sieve Estimation in Semiparametric Modeling of Longitudinal Data with Informative Observation Times

XINGQIU ZHAO

Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong

xingqiu.zhao@polyu.edu.hk

SHIRONG DENG, LI LIU

School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, China

LEI LIU

Department of Preventive Medicine and Robert H. Lurie Cancer Center, Northwestern University, Chicago, IL 60611, USA

Appendix: Proofs of Asymptotic Properties of $\hat{\theta}_n$

To establish the asymptotic properties of the estimators, we need the following regularity conditions.

C1. The maximum spacing of the knots satisfies $\triangle = \max_{l+1<i<m_n+l+1} \mid t_i - t_{i-1} \mid = O(n^{-v})$.

C2. The parameter spaces of $(\beta', \alpha')'$, $\mathcal{R}$ is bounded and convex on $\mathbb{R}^{p+q}$, and the true parameter $(\beta_0, \alpha_0, \mu_0) \in \mathcal{R}^\circ \times \mathcal{F}_r$, where $\mathcal{R}^\circ$ is the interior of $\mathcal{R}$, and

$$\mathcal{F}_r = \{\mu : [0,\infty) \longrightarrow \mathbb{R} \Big| |\mu^{(l)}(s) - \mu^{(l)}(t)| \leq M|s-t|^\zeta, s, t \in [0,\tau]\},$$

where $r = l + \zeta > 0.5$, M is a positive constant and $f^{(l)}$ denotes the lth derivative of function f.

C3. $P(\|\mathbf{X}\| \leq M_1) = 1$ for a positive constant M_1, that is, the covariate vector is uniformly bounded.

C4. There exists a positive integer M_2 such that $P(K \leq M_2) = 1$, that is, the number of the observation is finite.

C5. If with probability 1, $\mathbf{h}_1'\mathbf{X} + \mathbf{h}_2' H(\mathcal{F}_t, W) + h_3(t) = 0$ for some deterministic function h_3, and $\mathbf{h}_1 \in \mathbb{R}^p$ and $\mathbf{h}_2 \in \mathbb{R}^q$, then $\mathbf{h}_1 = 0$, $\mathbf{h}_2 = 0, h_3(t) = 0$.

Next, we introduce more notation. Let

$$\mathcal{F} = \left\{ f : [0, \infty) \longrightarrow \mathbb{R} : \ ||f||_2 \equiv \left[E\left\{ \int_0^\tau |f(t)|^2 dN(t) \right\} \right]^{1/2} < \infty \right\}.$$

Let $Z = \{Z(t, W) \equiv H(\mathcal{F}_t, W), 0 \leq t \leq \tau\}$ represent a q-dimensional bounded random process index by t. Here, without loss of generality, we assume that W is one-dimensional. Define

$$\mathcal{G} \equiv \{z(t, w) : [0, \tau] \times [-M_1, M_1] \longrightarrow \mathcal{M}\},$$

where $\mathcal{M}$ is a bounded set on $\mathbb{R}^q$, and for function $f(\ \mathbf{x}, z, t) : [-M_1, M_1]^p \times \mathcal{G} \times [0, \tau] \longrightarrow \mathbb{R}$, define

$$\|f\|_2 \equiv \left[E\left\{ \sum_{j=1}^{K} |f(\mathbf{X}, Z(T_{K,j}, W), T_{K,j})|^2 \right\} \right]^{1/2}.$$

Set $M_n(g) = n^{-1} L_n(\beta, \alpha, \mu) = \mathbb{P}_n m_g(\mathbf{O})$, where $g(\mathbf{x}, z, t) = \beta'\mathbf{x} + \alpha' z(t, w) + \mu(t)$ and

$$m_g(\mathbf{O}) = \sum_{j=1}^{K} [Y(T_{K,j}) - g(\mathbf{X}, Z(T_{K,j}, W), T_{K,j})]^2,$$

and $M(g) = P m_g(\mathbf{O})$, where Pf and $\mathbb{P}_n f$ represent $\int f dP$ and $n^{-1} \sum_{i=1}^n f(\mathbf{O}_i)$, respectively.

Since $\mathcal{F}$ is a Hilbert space, and $\mathcal{F}_r \subset \mathcal{F}$, by the Hilbert Projection Theorem (Stakgold, 1998, page 288), for $x_j \in \mathcal{F}$, there is a unique $a_j^* \in \mathcal{F}_r$, s.t. $(x_j - a_j^*) \perp \mathcal{F}_r$, for $j = 1, \cdots, p$. Let $z_l(t, w)$ be the lth component of $H(\mathcal{F}_t, W), l = 1, \cdots, q$. Then for $z_l(t, w) \in \mathcal{F}$, there is a unique $b_l^*(t) \in \mathcal{F}_r$, s.t. $(z_l - b_l^*) \perp \mathcal{F}_r$, for $l = 1, \cdots, q$. Let $\mathbf{a}^* = (a_1^*, \cdots, a_p^*)'$ and $\mathbf{b}^* = (b_1^*, \cdots, b_q^*)'$. Furthermore, we need the following condition.

C6. $E\left[\int_0^\tau \begin{pmatrix} \mathbf{X} - \mathbf{a}^* \\ H(\mathcal{F}_t, W) - \mathbf{b}^*(t) \end{pmatrix}^{\bigotimes 2} dN(t) \right]$ is nonsingular.

Here C1 is the same condition as required by Lu et al. (2007). C2 is a common assumption in the nonparametric smoothing estimation problem. C3 and C4 are mild conditions. C5 is needed to establish the identifiability of the model. C6 is a technical condition.

Proof of the consistency of $\hat{\theta}_n$.

Let $\mu_n(t)$ be the B-spline function approximation of $\mu_0(t)$ with $||\mu_n - \mu_0||_\infty = O(n^{-vr})$, $g_n(\mathbf{x}, z, t) = \beta_0'\mathbf{x} + \alpha_0' z(t, w) + \mu_n(t)$, $\hat{g}_n(\mathbf{x}, z, t) = \hat{\beta}_n'\mathbf{x} + \hat{\alpha}_n' z(t, w) + \hat{\mu}_n(t)$, and $g_0(\mathbf{x}, z, t) = \beta_0'\mathbf{x} + \alpha_0' z(t, w) + \mu_0(t)$. Without loss of generality, we assume that $\mu_n > \mu_0$. Thus $g_n > g_0$, and $\|g_n - g_0\|_\infty = O(n^{-vr})$. Choose a $\phi_n \in \Psi_{l,\mathcal{I}}$ and b_1 and b_2, such that $h_n \equiv {b_1}'\mathbf{x} + {b_2}'z + \phi_n$, and $\|h_n\|_2^2 = O(n^{-vr} + n^{-\frac{1-v}{2}})$. Then for any $\lambda > 0$, $\|g_n - g_0 + \lambda h_n\|_2^2 = O(n^{-vr} + n^{-\frac{1-v}{2}})$. Let

$$
\begin{aligned}
&H_n(\lambda) \equiv M_n(g_n + \lambda h_n) \\
=&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{K_i}[Y_i(T_{K_i,j}) - (g_n + \lambda h_n)(\mathbf{X}_i, Z_i(T_{K_i,j}, W_i), T_{K_i,j})]^2,
\end{aligned}
$$

then $H_n'(\lambda)$ is a nondecreasing function. Therefore, to prove the convergence of $\hat{g}_n$ to g_0, it is sufficient to show that $\forall \lambda_0 > 0, H_n'(\lambda_0) > 0$ and $H_n'(-\lambda_0) < 0$ except on an event with probability converging to zero. This can be proved by using the arguments similar to those in the proof of Lemma 5.1 of Huang (1999). Thus $\hat{g}_n$ must be between $g_n - \lambda_0 h_n$ and $g_n + \lambda_0 h_n$, and so $\|\hat{g}_n - g_n\|_2^2 \leq \lambda_0^2 \|h_n\|_2^2 = O(n^{-vr} + n^{-\frac{1-v}{2}})$. Then we have $\|\hat{g}_n - g_0\|_2 \leq \|\hat{g}_n - g_n\|_2 + \|g_n - g_0\|_2 = O((n^{-vr} + n^{-\frac{1-v}{2}})^{1/2})$, and

$$
\begin{aligned}
&\|\hat{g}_n - g_0\|_2 \\
&= \|(\hat{\beta}_n - \beta_0)'(\mathbf{x} - \mathbf{a}^*) + (\hat{\alpha}_n - \alpha_0)'(z - \mathbf{b}^*)\|_2 + \|(\hat{\beta}_n - \beta_0)'\mathbf{a}^* + (\hat{\alpha}_n - \alpha_0)'\mathbf{b}^* + (\hat{\mu}_n - \mu_0)\|_2.
\end{aligned}
$$

By C6, we can get $\|\hat{\beta}_n - \beta_0\| \longrightarrow 0$ and $\|\hat{\alpha}_n - \alpha_0\| \longrightarrow 0$ almost surely from the first term of the right hand side of the above equality and thus it follows that $\|\hat{\mu}_n - \mu_0\|_2 \longrightarrow 0$ almost surely. This completes the proof of the consitency.

Proof of the rate of convergence of $\hat{\theta}_n$.

For any $\eta > 0$, let

$$\mathcal{F}_\eta \equiv \{g = \beta'\mathbf{x} + \alpha' z + \mu : \|\beta - \beta_0\| \leq \eta, \|\alpha - \alpha_0\| \leq \eta, \mu \in \Psi_{l,\mathcal{I}}, \|\mu - \mu_0\|_2 \leq \eta\}.$$

Similar to Lemma A.2 in Huang (1999), for any $\varepsilon \leq \eta$,

$$\log N_{[\,]}(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2) \leq c_1 q_n \log(\eta/\varepsilon)$$

for a constant c_1. Thus, for $\varepsilon > 0$, there exists a set of brackets $\{[g_i^l, g_i^r], i = 1, \cdots, (\frac{\eta}{\varepsilon})^{c_1 q_n}\}$ such that, for each $g \in \mathcal{F}_\eta$, there is a $[g_s^l, g_s^r]$, s.t. $g_s^l(\mathbf{x}, z, t) \leq g(\mathbf{x}, z, t) \leq g_s^r(\mathbf{x}, z, t)$, for all $\mathbf{x}$, $t \in [0, \tau]$ and $z \in \mathcal{G}$, and $\|g_s^r - g_s^l\|_2^2 \leq \varepsilon^2$.

By the consistency of $\hat{\theta}_n$, $\hat{g}_n \in \mathcal{F}_\eta$, for any $\eta > 0$ and sufficiently large n.

Next, consider the class $\mathcal{M}_\eta \equiv \{m_g(\mathbf{O}) - m_{g_0}(\mathbf{O}) : g \in \mathcal{F}_\eta\}$.

For $i = 1, \cdots, (\frac{\eta}{\varepsilon})^{c_1 q_n}$, define

$$\begin{aligned} m_i^l(\mathbf{O}) = \sum_{j=1}^{K} \Big\{ &[\min\{|g_i^l(\mathbf{X}, Z(T_{K,j}, W)|, T_{K,j}), |g_i^r(\mathbf{X}, Z(T_{K,j}, W), T_{K,j})|\}]^2 \\ &- 2Y(T_{K,j})\{g_i^r(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) I(Y(T_{K,j}) \geq 0) \\ &+ g_i^l(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) I(Y(T_{k,j}) < 0)\} \\ &+ 2Y(T_{K,j}) g_0(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) - g_0^2(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) \Big\} \end{aligned}$$

and

$$\begin{aligned} m_i^r(\mathbf{O}) = \sum_{j=1}^{K} \Big\{ &[\max\{|g_i^l(\mathbf{X}, Z(T_{K,j}, W), T_{K,j})|, |g_i^r(\mathbf{X}, Z(T_{K,j}, W), T_{K,j})|\}]^2 \\ &- 2Y(T_{K,j})\{g_i^l(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) I(Y(T_{K,j}) \geq 0) \\ &+ g_i^r(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) I(Y(T_{K,j}) < 0)\} \\ &+ 2Y(T_{K,j}) g_0(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) - g_0^2(\mathbf{X}, Z(T_{K,j}, W), T_{K,j}) \Big\}. \end{aligned}$$

It is easy to show that $||m_i^r(\mathbf{O}) - m_i^l(\mathbf{O})||_{P,B}^2 \le c_2\varepsilon^2$ with a positive constant c_2, where $\|\cdot\|_{P,B}$ is the "Bernstein norm" defined as $\|f\|_{P,B} = \{2P(e^{|f|} - 1 - |f|)\}^{1/2}$ by van der Vaart and Wellner (1996). Thus $\{[m_i^l(\mathbf{O}), m_i^r(\mathbf{O})], i = 1, \cdots, (\frac{\eta}{\varepsilon})^{c_1 q_n}\}$ is the set of brackets for $\mathcal{M}_\eta$, which implies that

$$\log N_{[]}(\varepsilon, \mathcal{M}_\eta, ||\cdot||_{P,B}) \le c_1 q_n \log(\eta/\varepsilon).$$

Similarly, we can verify that $||m_g(\mathbf{O}) - m_{g_0}(\mathbf{O})||_{P,B}^2 \le c_3\eta^2$ for any $g \in \mathcal{F}_\eta$ by C4. Therefore, by Lemma 3.4.3 of van der Vaart and Wellner (1996), we obtain

$$E\|n^{1/2}(\mathbb{P}_n - P)\|_{\mathcal{M}_\eta} \le c_4 J_{[]}(\eta, \mathcal{M}_\eta, ||\cdot||_{P,B}) \left\{1 + \frac{J_{[]}(\eta, \mathcal{M}_\eta, ||\cdot||_{P,B})}{\eta^2 n^{1/2}}\right\}, \qquad (A1)$$

where $\|n^{1/2}(\mathbb{P}_n - P)\|_{\mathcal{M}_\eta} = \sup_{f\in\mathcal{M}_\eta} |n^{1/2}(\mathbb{P}_n - P)f|$, and

$$J_{[]}(\eta, \mathcal{M}_\eta, ||\cdot||_{P,B}) = \int_0^\eta \{1 + \log N_{[]}(\varepsilon, \mathcal{M}_\eta, ||\cdot||_{P,B})\}^{1/2} d\varepsilon \le c_5 q_n^{1/2}\eta.$$

The right hand side of (A1) yields $\varphi_n(\eta) = c_5(q_n^{1/2}\eta + q_n/n^{1/2})$. It is easy to see that $\varphi_n(\eta)/\eta$ is decreasing in η, and

$$r_n^2\varphi_n(\frac{1}{r_n}) = r_n q_n^{1/2} + r_n^2 q_n/n^{1/2} \le 2n^{1/2}, \text{ for } r_n = n^{\frac{1-v}{2}} \text{ and } 0 < v < 1/2.$$

Note that $Pm_g(\mathbf{O}) - Pm_{g_0}(\mathbf{O}) = \|g - g_0\|_2^2$. Thus, by Theorem 3.2.5 of van der Vaart and Wellner (1996), $n^{\frac{1-v}{2}}\|\hat{g}_n - g_0\|_2 = O_p(1)$. Therefore by the similar arguments as those in the proof of consistency of $\hat{\beta}_n$, $\hat{\alpha}_n$, and $\hat{\mu}_n$, we can get the rate of convergence of $\hat{\beta}_n$, $\hat{\alpha}_n$, and $\hat{\mu}_n$ as $n^{(1-v)/2}$.

Proof of the asymptotic normality of $\hat{\theta}_n$.

Let $\mathcal{H} \equiv \{h = (\mathbf{h}_1,\ \mathbf{h}_2, h_3) : (\mathbf{h}_1', \mathbf{h}_2')' \in \mathcal{R}, h_3 \in \mathcal{F}_r, \|\mathbf{h}_1\| \le 1, \|\mathbf{h}_2\| \le 1, \|h_3\|_\infty \le 1\}$. We define a sequence of maps S_n mapping a neighborhood of $(\beta_0, \alpha_0, \mu_0)$, denoted by $\mathcal{U}$, in the parameter space for (β, α, μ) into $l^\infty(\mathcal{H})$ as

$$\begin{aligned} & S_n(\beta,\alpha,\mu)[\mathbf{h}_1,\mathbf{h}_2,h_3] \\ \equiv\ & n^{-1}\frac{d}{d\varepsilon}L_n(\beta+\varepsilon\mathbf{h}_1,\alpha+\varepsilon\mathbf{h}_2,\mu+\varepsilon h_3)\Big|_{\varepsilon=0} \\ =\ & -\frac{2}{n}\sum_{i=1}^n\int_0^\tau[Y_i(t)-\beta'\mathbf{X}_i-\alpha' H(\mathcal{F}_{it},W_i)-\mu(t)][\mathbf{h}_1'\ \mathbf{X}_i+\mathbf{h}_2' H(\mathcal{F}_{it},W_i)+h_3(t)]dN_i(t) \\ \equiv\ & \mathbb{P}_n\psi(\theta;\mathbf{O})[\mathbf{h}_1,\ \mathbf{h}_2,h_3], \end{aligned}$$

where

$$\psi(\theta;\mathbf{O})[\mathbf{h}_1,\ \mathbf{h}_2,h_3]=-2\int_0^\tau[Y(t)-\beta'\mathbf{X}-\alpha' H(\mathcal{F}_t,W)-\mu(t)][\mathbf{h}_1'\ \mathbf{X}+\mathbf{h}_2' H(\mathcal{F}_t,W)+h_3(t)]dN(t)$$

with $\theta=(\beta,\alpha,\mu)$.

Correspondingly, we define the limit map $S:\mathcal{U}\longrightarrow l^\infty(\mathcal{H})$ as

$$S(\theta)[\mathbf{h}_1,\ \mathbf{h}_2,h_3]=P\psi(\theta;\mathbf{O})[\mathbf{h}_1,\ \mathbf{h}_2,h_3],$$

where $l^\infty(\mathcal{H})$ is the space of bounded functionals on $\mathcal{H}$ under the supermum norm $\|f\|_\infty = \sup_{h\in\mathcal{H}}|f(h)|$. Also, we define the derivative of $S(\beta,\alpha,\mu)[\mathbf{h}_1,\ \mathbf{h}_2,h_3]$ at (β_0,α_0,μ_0), denoted by $\dot{S}(\beta_0,\alpha_0,\mu_0)[\mathbf{h}_1,\ \mathbf{h}_2,h_3]$, as a map from the space $\{(\beta-\beta_0,\alpha-\alpha_0,\mu-\mu_0):(\beta,\alpha,\mu)\in\mathcal{U}\}$ to $l^\infty(\mathcal{H})$ and

$$\begin{aligned} & \dot{S}(\beta_0,\alpha_0,\mu_0)(\beta-\beta_0,\alpha-\alpha_0,\mu-\mu_0)[\mathbf{h}_1,\mathbf{h}_2,h_3] \\ =\ & \frac{d}{d\epsilon}S(\beta_0+\epsilon(\beta-\beta_0),\alpha_0+\epsilon(\alpha-\alpha_0),\mu_0+\epsilon(\mu-\mu_0))[\mathbf{h}_1,\mathbf{h}_2,h_3]\Big|_{\epsilon=0} \\ \equiv\ & \sigma_1(\mathbf{h}_1,\mathbf{h}_2,h_3)'(\beta-\beta_0)+\sigma_2(\mathbf{h}_1,\mathbf{h}_2,h_3)'(\alpha-\alpha_0)+\int_0^\tau(\mu-\mu_0)(t)d\sigma_3(\mathbf{h}_1,\mathbf{h}_2,h_3)(t), \end{aligned}$$

where

$$\sigma_1(\mathbf{h}_1,\mathbf{h}_2,h_3)=2P\int_0^\tau[\mathbf{h}_1'\ \mathbf{X}+\mathbf{h}_2' H(\mathcal{F}_t,W)+h_3(t)]\mathbf{X}dN(t),$$

$$\sigma_2(\mathbf{h}_1,\mathbf{h}_2,h_3)=2P\int_0^\tau[\mathbf{h}_1'\ \mathbf{X}+\mathbf{h}_2' H(\mathcal{F}_t,W)+h_3(t)]H(\mathcal{F}_t,W)dN(t),$$

and

$$\sigma_3(\mathbf{h}_1, \mathbf{h}_2, h_3)(t) = 2P \int_0^t [\mathbf{h}_1' \ \mathbf{X} + \mathbf{h}_2' H(\mathcal{F}_s, W) + h_3(s)] dN(s).$$

To derive the asymptotic normality of the estimators $(\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n)$, following the proof of Theorem 3.3.1 of van der Vaart and Wellner (1996, page 310) and the proof of Theorem 2 in Zeng et al. (2005), we can show that

$$\begin{aligned}
&\sqrt{n}\dot{S}(\beta_0, \alpha_0, \mu_0)(\hat{\beta}_n - \beta_0, \hat{\alpha}_n - \alpha_0, \hat{\mu}_n - \mu_0)[\mathbf{h}_1, \mathbf{h}_2, h_3] \\
&= \sigma_1(\mathbf{h}_1, \mathbf{h}_2, h_3)' \sqrt{n}(\hat{\beta}_n - \beta_0) + \sigma_2(\mathbf{h}_1, \mathbf{h}_2, h_3)' \sqrt{n}(\hat{\alpha}_n - \alpha_0) \\
&\quad + \int_0^\tau \sqrt{n}(\hat{\mu}_n - \mu_0)(t) d\sigma_3(\mathbf{h}_1, \mathbf{h}_2, h_3)(t) \\
&= -\sqrt{n}(S_n - S)(\beta_0, \alpha_0, \mu_0)[\mathbf{h}_1, \mathbf{h}_2, h_3] + o_p(1),
\end{aligned} \tag{A2}$$

for any $(\mathbf{h}_1, \mathbf{h}_2, h_3) \in \mathcal{H}$.

Next, we will derive the asymptotic normality of $\hat{\theta}_n$ from (A2).

(i) Note that

$$\begin{aligned}
&\sigma_3(\mathbf{h}_1, \mathbf{h}_2, h_3)(t) \\
&= 2E\left[\int_0^\tau I(s \le t)\left\{\mathbf{h}_1' \ \mathbf{X} + \mathbf{h}_2' H(\mathcal{F}_s, W) + h_3(s)\right\} dN(s)\right] \\
&= 2E\left[\sum_{j=1}^K I(T_{K,j} \le t)\left\{\mathbf{h}_1' \ \mathbf{X} + \mathbf{h}_2' H(\mathcal{F}_{T_{K,j}}, W) + h_3(T_{K,j})\right\}\right] \\
&= 2E\left[\sum_{j=1}^K E\left\{I(T_{K,j} \le t)\left(\mathbf{h}_1' \ \mathbf{X} + \mathbf{h}_2' H(\mathcal{F}_{T_{K,j}}, W) + h_3(T_{K,j})\right) | K\right\}\right] \\
&= 2E\left[\sum_{j=1}^K E\left\{I(T_{K,j} \le t)\left(E\left(\mathbf{h}_1' \ \mathbf{X} + \mathbf{h}_2' H(\mathcal{F}_{T_{K,j}}, W) | K, T_{K,j}\right) + h_3(T_{K,j})\right) \Big| K\right\}\right].
\end{aligned}$$

Thus, we can take

$$\begin{aligned}
h_3(T_{K,j}) &= -E\left\{\mathbf{h}_1' \ \mathbf{X} + \mathbf{h}_2' H(\mathcal{F}_{T_{K,j}}, W) \Big| K, T_{K,j}\right\} \\
&= -\begin{pmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{pmatrix}' E\left[\begin{pmatrix} \mathbf{X} \\ H(\mathcal{F}_{T_{K,j}}, W) \end{pmatrix} \Big| K, T_{K,j}\right]
\end{aligned}$$

such that $\sigma_3(\mathbf{h}_1, \mathbf{h}_2, h_3)(t) \equiv 0$. Furthermore, for this h_3, we have

$$\begin{pmatrix} \sigma_1(\mathbf{h}_1, \mathbf{h}_2, h_3) \\ \sigma_2(\mathbf{h}_1, \mathbf{h}_2, h_3) \end{pmatrix} = 2A \begin{pmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{pmatrix},$$

and then it follows from (A2) that

$$2 \begin{pmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{pmatrix}' A\sqrt{n} \begin{pmatrix} \hat{\beta}_n - \beta_0 \\ \hat{\alpha}_n - \alpha_0 \end{pmatrix} \longrightarrow N(0, \sigma^2),$$

where

$$\sigma^2 = P[\psi(\beta_0, \alpha_0, \mu_0; \mathbf{O})[\mathbf{h}_1, \mathbf{h}_2, h_3]]^2 = 4 \begin{pmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{pmatrix}' B \begin{pmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{pmatrix},$$

and A and B are given in Section 3. Therefore, $A\sqrt{n} \begin{pmatrix} \hat{\beta}_n - \beta_0 \\ \hat{\alpha}_n - \alpha_0 \end{pmatrix} \longrightarrow N(0, B)$, which implies $\sqrt{n} \begin{pmatrix} \hat{\beta}_n - \beta_0 \\ \hat{\alpha}_n - \alpha_0 \end{pmatrix} \longrightarrow N(\mathbf{0}, A^{-1}B(A^{-1})')$.

(ii) Note that

$$\begin{pmatrix} \sigma_1(\mathbf{h}_1, \mathbf{h}_2, h_3) \\ \sigma_2(\mathbf{h}_1, \mathbf{h}_2, h_3) \end{pmatrix} = 2\Gamma \begin{pmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{pmatrix} + 2E \left\{ \sum_{j=1}^{K} \begin{pmatrix} \mathbf{X} \\ H(\mathcal{F}_{T_{K,j}}, W) \end{pmatrix} h_3(T_{K,j}) \right\},$$

where

$$\Gamma = E \left[\sum_{j=1}^{K} \begin{pmatrix} \mathbf{X} \\ H(\mathcal{F}_{T_{K,j}}, W) \end{pmatrix}^{\otimes 2} \right].$$

Then, we can take

$$\begin{pmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{pmatrix} = -\Gamma^{-1} E \left\{ \sum_{j=1}^{K} \begin{pmatrix} \mathbf{X} \\ H(\mathcal{F}_{T_{K,j}}, W) \end{pmatrix} h_3(T_{K,j}) \right\}$$

such that $\sigma_1(\mathbf{h}_1, \mathbf{h}_2, h_3) = 0$ and $\sigma_2(\mathbf{h}_1, \mathbf{h}_2, h_3) = 0$. Thus it follows from (A2) that $\sqrt{n} \int_0^\tau (\hat{\mu}_n(t) -$

$\mu_0(t))dG_g(t)$ converges in distribution to $N(0, \sigma_g^2)$, where for $g \in \mathcal{F}_r$,

$$
\begin{aligned}
&G_g(t) \\
&= P\left[\int_0^t \left\{ g(s) - \begin{pmatrix} \mathbf{X} \\ H(\mathcal{F}_s, W) \end{pmatrix}' \Gamma^{-1} P\left(\int_0^\tau \begin{pmatrix} \mathbf{X} \\ H(\mathcal{F}_u, W) \end{pmatrix} g(u)dN(u)\right) \right\} dN(s)\right],
\end{aligned}
$$

and

$$
\begin{aligned}
\sigma_g^2 &= E\left[\sum_{j=1}^K \{Y(T_{K,j}) - \mu_0(T_{K,j}) - \beta_0'\mathbf{X} - \alpha_0' H(\mathcal{F}_{T_{K,j}}, W)\}\right. \\
&\quad \left.\times \left\{ g(T_{K,j}) - \begin{pmatrix} \mathbf{X} \\ H(\mathcal{F}_{T_{K,j}}, W) \end{pmatrix}' \Gamma^{-1} P\left(\int_0^\tau \begin{pmatrix} \mathbf{X} \\ H(\mathcal{F}_u, W) \end{pmatrix} g(u)dN(u)\right) \right\}\right]^2.
\end{aligned}
$$

References

Huang, J. (1999). Efficient estimation of the partly linear additive Cox model. *Journal of the American Statistical Association* **27,** 1536–1563.

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

Stakgold, I. (1998). *Green's Functions and Boundary Value Problems.* John Wiley & Sons, Inc.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* New York: Springer-Verlag.

Zeng, D., Lin, D. and Yin, G. S. (2005). Maximum likelihood estimation for the proportional odds model with random effects. *Journal of the American Statistical Association* **100**, 470–482.