Profile Local Linear Estimation of Generalized Semiparametric Regression Model for Longitudinal Data

Yanqing Sun
Department of Mathematics and Statistics
The University of North Carolina at Charlotte, Charlotte, NC 28223
yasun@uncc.edu
Liuquan Sun
Institute of Applied Mathematics
Academy of Mathematics and Systems Science, Beijing, China
slq@amt.ac.cn
Jie Zhou
Institute of Applied Mathematics
Academy of Mathematics and Systems Science, Beijing, China
zhoujie@amss.ac.cn

Abstract

This paper studies the generalized semiparametric regression model for longitudinal data where the covariate effects are constant for some and time-varying for others. Different link functions can be used to allow more flexible modelling of longitudinal data. The nonparametric components of the model are estimated using a local linear estimating equation and the parametric components are estimated through a profile estimating function. The method automatically adjusts for heterogeneity of sampling times, allowing the sampling strategy to depend on the past sampling history as well as possibly time-dependent covariates without specifically model such dependence. Large sample properties of the proposed estimators are investigated. Large sample pointwise and simultaneous confidence intervals for the regression coefficients are constructed. A formal hypothesis testing procedure is proposed to check whether the effect of a covariate is time-varying. A simulation study is conducted to examine the finite sample performances of the proposed estimation and hypothesis testing procedures. The method is illustrated with a data set from a HIV-1 RNA data set from an AIDS clinical trial.

Key Words: Asymptotics; Censored follow-up times; Cross-validation bandwidth selection; Kernel smoothing; Link function; Local linear estimating equation; Profile estimating equation; Sampling adjusted estimation; Testing time-varying effects; Weighted least squares.

1 Introduction

We consider semiparametric modeling of covariate effects on a longitudinal response process based on repeated measurements observed at a series of sampling times. Suppose that there is a random sample of n subjects. For the ith subject, let $Y_i(t)$ be the response process and let $Z_i(t)$ and $X_i(t)$ be the possibly time-dependent covariates of dimensions $p \times 1$ and $q \times 1$, respectively, over the time interval $[0, \tau]$. We consider the following generalized semiparametric regression model for $Y_i(t)$, $0 \le t \le \tau$,

$$\mu_i(t) = E\{Y_i(t)|X_i(t), Z_i(t)\} = g^{-1}\{\gamma^T(t)X_i(t) + \beta^T Z_i(t)\}, \qquad i = 1, \ldots, n, \tag{1}$$

where $g(\cdot)$ is a known link function, β is a p-dimensional vector of unknown parameters and $\gamma(t)$ is a q-dimensional vector of completely unspecified functions. The notation β^T represents transpose of a vector or matrix β. The first component of $X_i(t)$ is set to be 1, which gives a nonparametric baseline function. Under model (1), the effects of some covariates are constant while others are time-varying. Model (1) is more flexible than the parametric regression model where all the regression coefficients are time-independent and more desirable than the nonparametric model where every covariate effect is an unspecified function of time. Different link functions can be selected to provide a richer family of models for longitudinal data.

When the link function $g(\cdot)$ is the identity function, model (1) is known as the semiparametric additive model. The semiparametric additive model with longitudinal data has been studied extensively in recent years. These approaches include the nonparametric kernel smoothing by Hoover et al. (1998), the joint modelling of longitudinal responses and sampling times by Martinussen and Scheike (1999, 2000, 2001), Lin and Ying (2001), the backfitting method by Wu and Liang (2004) and the profile kernel smoothing approach by Sun and Wu (2005). Fan, Huang and Li (2007) proposed a profile local linear approach by imposing some correlation structure for the longitudinal data for improved efficiency. Fan and Li (2004) considered the profile local linear approach and the joint modelling for partially linear models. Hu, Wang and Carroll (2004) showed that for partially linear models, the backfitting is less efficient than the profile kernel method. When the link function is the natural logarithm function and $X_i(t) \equiv 1$, model (1) becomes the proportional means model. Data collected on the individual response processes at a finite set of sampling times are also called panel data. Zhang (2002) proposed a semiparametric pseudolikelihood method for the proportional means model under the assumption that the response is a nonhomogeneous Poisson process. For panel count data, the proportional means model has been studied by Sun and Wei (2000), Cheng and Wei (2000), and Hu, Sun and Wei (2003). Model (1) unifies the semiparametric additive model and the proportional means model under the same umbrella.

Although model (1) has been extensively studied for cross-sectional data, few have studied it with longitudinal data. Lin and Carroll (2001) studied model (1) when $X_i(t) \equiv 1$ by using profile-based generalized estimating equations (GEE) and a local linear approach. Lin, Song and Zhou (2007) proposed a local linear GEE method when all the regression coefficients are nonparametric functions of time. The GEE method with appropriately selected working covariance structure of the longitudinal data can lead to improved efficiency (Fan, Huang and Li (2007)). However, the selection of the working covariance can be difficult and the efficiency gain under an improperly selected working covariance structure is not clear. Further, there may be technique difficulties with the extension of the GEE method to more complicated sampling schemes. In both Lin and Carroll (2001) and Lin, Song and Zhou (2007), the sampling times are assumed to be independent of covariates and the situation of possible dropouts of the subjects in the follow-up is not considered. The extensions of their methods to more general sampling and censoring schemes would make these methods more useful in practice.

The marginal approach provides an important alternative to the longitudinal data analysis. It is more flexible in integrating complicated sampling and censoring schemes into the analysis. The powerful theories for empirical processes and counting processes facilitate such developments. Most of the existing marginal approaches for analyzing longitudinal data assume that the sampling times are independent of covariates or follow a proportional/additive mean rate model (Lin et al. (2000), Scheike (2002)) to account for possible dependence on the covariates; cf. Lin and Ying (2001), Martinussen and Scheike (1999, 2000, 2001). However, misspecifications of the sampling model may result in biased estimations and mislead the inferences for the response process. Sun and Wu (2005) proposed a profile kernel estimation procedure for the semiparametric additive model without having to specify a sampling model for the observation times. Similar approach was exploited by Sun (2010) for the proportional means model. This paper proposes a sampling adjusted profile local linear estimation method for the generalized semiparametric regression model (1). The paper has two main contributions. First, the proposed method automatically adjusts for heterogeneity of sampling times, allowing the sampling strategy to depend on the past sampling history as well as possibly time-dependent covariates without specifically model such dependence. Second, this paper presents an unified approach to the semiparametric model (1) with a general link function which has never been exploited for longitudinal data to the best of our knowledge. The local linear estimation technique has been shown to be design-adaptive and more efficient in correcting boundary bias than the kernel smoothing approach for the cross-sectional data; see Fan and Gijbels (1996). We show that these features preserve under the proposed approach for longitudinal data. The proposed method does not require time-varying covariates to be observed at all time, only the values at the sampling times are needed. Some hypothesis testing procedures are proposed to check whether the effect of a covariate is time-varying. This can lead to more efficient estimation when the effects of some covariates are not really time-varying.

The rest of the paper is organized as follows. In Section 2, a sampling adjusted profile-based local linear estimation method is proposed for model (1). Large sample properties are investigated in Section 3. Large sample pointwise and simultaneous confidence intervals for the regression coefficients are constructed. This section also presents some formal hypothesis testing procedures to check whether the effect of a covariate is time-varying. A cross-validation bandwidth selection approach is proposed to serve as a working tool for locating an appropriate bandwidth. A simulation study is conducted in Section 4 to examine the finite sample performances of the proposed statistical procedures. An application of the proposed methods to the analysis of a HIV-1 RNA data set from an AIDS clinical trial is given in Section 5, and some concluding remarks are made in Section 6. All proofs are given in the Appendix.

2 Profile local linear estimation approach

2.1 Prelimilaries

Suppose that the observations of the response process $Y_i(t)$ for the ith subject are taken at the sampling time points $0 \le t_{i1} < t_{i2} < \cdots < t_{in_i} \le \tau$, where n_i is the total number of observations on the ith subject and τ is the end of follow-up time. The sampling times are often irregular and depend on covariates. In addition, some subjects may drop out of the study early. Let $N_i(t) = \sum_{j=1}^{n_i} I(t_{ij} \le t)$ be the number of observations taken on the ith subject by time t, where $I(\cdot)$ is the indicator function. Let C_i be the end of follow-up time or censoring time whichever comes first. The responses for the ith subject can only be observed at the time points before C_i. Thus $N_i(t)$ can be written as $N_i^*(t \wedge C_i)$, where $N_i^*(t)$ is the counting process of sampling times. Let $X_i(t)$ and $Z_i(t)$ be the predictable covariate processes associated with the ith subject. We assume that $\{(Y_i(\cdot), X_i(\cdot), Z_i(\cdot), N_i(\cdot))\}$, $i = 1, \cdots, n$, are independent identically distributed random processes. In this section, we propose an estimation procedure for model (1) based on the observations $\{(Y_i(t_{ij}), X_i(t_{ij}),\ Z_i(t_{ij})); j = 1, \ldots, n_i, i = 1, \ldots, n.\}$. These are the values of $\{(Y_i(t), X_i(t), Z_i(t)),\ 0 \le t \le \tau\}$ observed at sampling times or the jump time points of $N_i(t) = N_i^*(t \wedge C_i)$, $i = 1, \ldots, n$.

Let $\mathcal{F}_t$ be the σ-field representing the history $N_i^*(\cdot)$, $X_i(\cdot)$ and $Z_i(\cdot)$ up to time t for $1 \le i \le n$. Let $\lambda_i(t)$ be the intensity process defined as follows

$$E\{dN_i^*(t)|\mathcal{F}_{t-}\} = \lambda_i(t)dt, \tag{2}$$

for $0 \le t \le \tau$. Thus $\lambda_i(t)$ is the sampling rate at time t conditional on the past $\mathcal{F}_{t-}$. Let $\alpha_i(t) = \alpha(t, X_i(t), Z_i(t))$ be the conditional mean rate of the sampling times such that $E\{dN_i^*(t)|X_i(t), Z_i(t)\} = \alpha(t, X_i(t), Z_i(t))\, dt$. Then $\alpha_i(t) = E\{\lambda_i(t)|X_i(t), Z_i(t)\}$ by the using the double expectation property.

Many existing methods such as Lin and Ying (2001), Martinussen and Scheike (1999, 2000, 2001) took the approach by modelling $\alpha_i(t)$. Lin and Ying (2001) assumed that the sampling process follows a proportional mean rate model (Lin, et al. (2000)). Martinussen and Scheike (1999, 2000) assumed that the intensity of the sampling process follows a multiplicative Aalen model (Aalen (1978)) $\lambda_i(t) = \eta_i(t)\alpha(t)$ where $\alpha(t)$ is an unknown deterministic function and $\eta_i(t)$ is a predictable process. Martinussen and Scheike (2001) considered the sampling adjusted approach by assuming that the intensity follows a nonparametric additive regression model $\lambda_i(t) = \eta_i(t)\alpha(t)^T X_i(t)$, where $\eta_i(t)$ is a predictable at risk indicator, $\alpha(t)$ is vector of unspecified time-dependent regression functions and $X_i(t)$ are predictable time varying covariates. For all these methods mentioned above, the misspecifications of the sampling model can lead to biased estimation of the mean longitudinal response since the expectations of the estimating equations may not be zero, which is also demonstrated in our simulation study in Section 4.

The proposed method in the following allows the sampling strategy to depend on the past $\mathcal{F}_{t-}$ as well as possibly time-dependent covariates without specifically model such dependence. The estimation procedure directly uses the sampling process $N_i(\cdot) = N_i^*(\cdot \wedge C_i)$

without modeling for $\lambda_i(t)$ or $\alpha_i(t)$.

2.2 Estimation procedures

We adopt a profile approach for the estimation of model (1). First, assuming β is known, the nonparametric component, $\gamma(t)$, of the model is estimated using the local linear estimating equations. The parametric component, β, is estimated through the weighted profile estimating equations. The details of the estimation procedure are described in the following.

At each t, let $\gamma(s) = \gamma(t) + \dot{\gamma}(t)(s-t) + O((s-t)^2)$ be the first order Taylor expansion of $\gamma(\cdot)$ for s in a neighborhood of t, where $\dot{\gamma}(t)$ is the derivative of $\gamma(t)$ with respect to t. Denote $\gamma_a(t) = (\gamma^T(t), \dot{\gamma}^T(t))^T$ and $\widetilde{X}_i(s, s-t) = X_i(s) \otimes (1,\, s-t)^T$ with $\otimes$ being the Kronecker product. Let $\tilde{\mu}_a(s, \gamma_a, \beta | X_i, Z_i) = \varphi\{\gamma_a^T(t)\widetilde{X}_i(s, s-t) + \beta^T Z_i(s)\}$, where $\varphi(x) = g^{-1}(x)$ is the inverse function of the link function $g(y)$. Let $W_i(t) = W(t, X_i(t), Z_i(t))$ be a nonnegative weight process that may depend on n. At each t and for fixed β, we consider the following estimating function for $\gamma_a(t)$:

$$U_a(\gamma_a, \beta) = \sum_{i=1}^{n} \int_0^{\tau} W_i(s)\{Y_i(s) - \tilde{\mu}_a(s, \gamma_a, \beta | X_i, Z_i)\} \widetilde{X}_i(s, s-t) K_h(s-t)\, d\, N_i(s), \qquad (3)$$

where $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a kernel function that weights smoothly down the contributions of remote data points and $h = h_n > 0$ is the bandwidth parameter that controls the size of a local neighborhood. The root of the equation $U_a(\gamma_a, \beta) = 0$ is denoted by $\tilde{\gamma}_a(t, \beta)$. Since the data used in (3) are localized in the neighborhood of t, a weight function for (3) will not have much effect on the local linear estimator.

Let $\dot{\varphi}(x)$ be the derivative of $\varphi(x) = g^{-1}(x)$ with respect to x. The estimating function $U_a(\gamma_a, \beta)$ can be obtained by setting $Q_i(s) = W_i(s)[\dot{\varphi}\{\gamma_a^T(t)\widetilde{X}_i(s, s-t) + \beta^T Z_i(s)\}]^{-1}$ in the derivative of the local weighted sum of the squares $\ell_a(\gamma_a, \beta) = \sum_{i=1}^n \int_0^\tau Q_i(s)\{Y_i(s) - \tilde{\mu}_a(s, \gamma_a, \beta\ | X_i, Z_i)\}^2 K_h(s-t)\, dN_i(s)$ with respect to γ_a. The expectation of $U_a(\gamma_a, \beta)$ is approximately zero for the true β and $\gamma(\cdot)$ as $h \to 0$ under the assumptions given in the Appendix. Let $\tilde{E}_{xx}(t) = n^{-1} \sum_{i=1}^n \int_0^\tau W_i(s) K_h(s-t)\ (\tilde{X}_i(s, s-t))^{\otimes 2}\, dN_i(s)$ and $\tilde{E}_{zx}(t) = n^{-1} \sum_{i=1}^n \int_0^\tau W_i(s) K_h(s-t) Z_i(s) (\tilde{X}_i(s, s-t))^T\, dN_i(s)$, where $v^{\otimes 2} = vv^T$ for a column vector v. $\tilde{E}_{yx}(t)$ is defined similarly to $\tilde{E}_{zx}(t)$ by replacing $Z_i(\cdot)$ with $Y_i(\cdot)$. Under the identity link function $g(x) = x$, a explicit solution for (3) can be derived as $\tilde{\gamma}_a(t,\beta) = \tilde{Y}_x^T(t) - \tilde{Z}_x^T(t)\beta$, where $\tilde{Y}_x(t) = \tilde{E}_{yx}(t)(\tilde{E}_{xx}(t))^{-1}$ and $\tilde{Z}_x(t) = \tilde{E}_{zx}(t)(\tilde{E}_{xx}(t))^{-1}$.

Let $\tilde{\gamma}(t,\beta)$ and $\tilde{\tilde{\gamma}}(t,\beta)$ be first and last q components of $\tilde{\gamma}_a(t,\beta)$, respectively. The profile estimating function for β is given by

$$U(\beta) = \sum_{i=1}^n \int_{t_1}^{t_2} W_i(s)[Y_i(s) - \varphi\{(\tilde{\gamma}(s,\beta))^T X_i(s) + \beta^T Z_i(s)\}]\{\frac{\partial \tilde{\gamma}(s,\beta)}{\partial \beta} X_i(s) + Z_i(s)\}\, dN_i(s), \tag{4}$$

where $[t_1, t_2] \subset (0,\tau)$. The subset $[t_1,t_2]$ is considered to avoid possible instability of $\tilde{\gamma}(t,\beta)$ near the boundary. In practice, this interval can be taken to be close to $[0,\tau]$. We estimate β by $\hat{\beta}$ that solves $U(\hat{\beta}) = 0$ and $\gamma(t)$ by $\hat{\gamma}(t) = \tilde{\gamma}(t,\hat{\beta})$.

The expression for the derivative $\frac{\partial \tilde{\gamma}(s,\beta)}{\partial \beta}$ in (4) is derived in the following. Since $U_a(\tilde{\gamma}_a(t,\beta),\beta) \equiv \mathbf{0}_{2q}$, $\tilde{\gamma}_a(t,\beta)$ satisfies

$$\left\{ \frac{\partial U_a(\gamma_a,\beta)}{\partial \gamma_a} \frac{\partial \tilde{\gamma}_a(t,\beta)}{\partial \beta} + \frac{\partial U_a(\gamma_a,\beta)}{\partial \beta} \right\} \Bigg|_{\gamma_a = \tilde{\gamma}_a(t,\beta)} = \mathbf{0}_{2q}.$$

It follows that

$$\frac{\partial \tilde{\gamma}_a(t,\beta)}{\partial \beta} = - \left\{ \frac{\partial U_a(\gamma_a,\beta)}{\partial \gamma_a} \right\}^{-1} \frac{\partial U_a(\gamma_a,\beta)}{\partial \beta} \Bigg|_{\gamma_a = \tilde{\gamma}_a(t,\beta)}, \tag{5}$$

where

$$-\frac{\partial U_a(\gamma_a,\beta)}{\partial \gamma_a} = \sum_{i=1}^n \int_0^\tau W_i(s)\dot{\varphi}\{\gamma_a^T \widetilde{X}_i(s,s-t) + \beta^T Z_i(s)\}\{\widetilde{X}_i(s,s-t)\}^{\otimes 2} K_h(s-t)\, d\, N_i(s), \tag{6}$$

$$-\frac{\partial U_a(\gamma_a,\beta)}{\partial \beta} = \sum_{i=1}^n \int_0^\tau W_i(s)\dot{\varphi}\{\gamma_a^T \widetilde{X}_i(s,s-t) + \beta^T Z_i(s)\}\widetilde{X}_i(s,s-t)(Z_i(s))^T K_h(s-t)\, d\, N_i(s). \tag{7}$$

The estimator $\hat{\beta}$ is a weighted least square estimator since the estimating function $U(\beta)$ can be obtained by setting $Q_i(t) = W_i(t)[\dot{\varphi}\{(\tilde{\gamma}(t,\beta))^T X_i(t) + \beta^T Z_i(t)\}]^{-1}$ in the derivative of the profile least squares function $\ell(\beta)$ with respect to β, where $\ell(\beta) = \sum_{i=1}^n \int_{t_1}^{t_2} Q_i(s)[Y_i(s) - \varphi\{(\tilde{\gamma}(s,\beta))^T X_i(s) + \beta^T Z_i(s)\}]^2\, d\, N_i(s)$.

2.3 Computational algorithm

The estimators $\hat{\beta}$ and $\hat{\gamma}(t)$ can be obtained through an iterated estimation procedure. Let $\hat{\beta}^{\{m-1\}}$ be the estimate of β at the $(m-1)$th step. The mth step estimator $\hat{\gamma}_a^{\{m\}}(t) =$

$\tilde{\gamma}_a(t, \hat{\beta}^{\{m-1\}})$ is the root of the estimating function (3) satisfying $U_a(\hat{\gamma}_a^{\{m\}}(t), \hat{\beta}^{\{m-1\}}) = 0$. The mth step estimator $\hat{\beta}^{\{m\}}$ is obtained by solving the estimating function for β:

$$
\begin{aligned}
U_m(\beta) &= \sum_{i=1}^{n} \int_{t_1}^{t_2} W_i(s)[Y_i(s) - \varphi\{(\tilde{\gamma}(s, \hat{\beta}^{\{m-1\}}))^T X_i(s) + \beta^T Z_i(s)\}] \\
&\qquad \times \{\frac{\partial \tilde{\gamma}(s, \hat{\beta}^{\{m-1\}})}{\partial \beta} X_i(s) + Z_i(s)\} \, dN_i(s), \qquad (8)
\end{aligned}
$$

where $\frac{\partial \tilde{\gamma}(t, \hat{\beta}^{\{m-1\}})}{\partial \beta}$ is calculated using the formula (5) at $\beta = \hat{\beta}^{\{m-1\}}$. The estimators $\hat{\gamma}_a^{\{m\}}(t)$ and $\hat{\beta}^{\{m\}}$ are updated at each iteration until convergence. The $\hat{\gamma}(t)$ is the first q components of $\hat{\gamma}_a(t) = \tilde{\gamma}_a(t, \hat{\beta})$.

3 Statistical inferences of semiparametric model

3.1 Asymptotic properties

This subsection investigates the asymptotic properties of the proposed estimators. These asymptotic results are used to construct confidence bands and formulate the test statistics for the regression coefficients in the subsequent subsections.

Let β_0 and $\gamma_0(t)$ be the true values of β and $\gamma(t)$ under model (1), respectively. Let $\mu_i(t) = \varphi\{\gamma_0^T(t) X_i(t) + \beta_0^T Z_i(t)\}$ and $\dot{\mu}_i(t) = \dot{\varphi}\{\gamma_0^T(t) X_i(t) + \beta_0^T Z_i(t)\}$. Let $w(t, x, z)$ be the deterministic limit of $W(t, x, z)$ in probability as $n \to \infty$. Define $e_{xx}(t) = E[w_i(t) \dot{\mu}_i(t) \{X_i(t)\}^{\otimes 2} \alpha_i(t)$ $\xi_i(t)]$ and $e_{xz}(t) = E[w_i(t) \dot{\mu}_i(t) X_i(t) \ \{Z_i(t)\}^T \alpha_i(t) \xi_i(t)]$, where $\xi_i(t) = I(C_i \geq t)$. Let $A = E[\int_{t_1}^{t_2} w_i(s) \dot{\mu}_i(s) \ \{Z_i(s) - (e_{xz}(s))^T (e_{xx}(s))^{-1} \ X_i(s)\}^{\otimes 2} \, dN_i(s)]$ and $\Sigma = E[\int_{t_1}^{t_2} w_i(s)$ $\{Y_i(s) - \mu_i(s)\}\{Z_i(s) - (e_{xz}(s))^T \ (e_{xx}(s))^{-1} X_i(s)\} \, dN_i(s)]^{\otimes 2}$, where $w_i(t) = w(t, X_i(t), Z_i(t))$.

Let $\hat{\mu}_i(s) = \varphi\{\hat{\gamma}^T(s) X_i(s) + \hat{\beta}^T Z_i(s)\}$ and $\hat{\dot{\mu}}_i(t) = \dot{\varphi}\{\hat{\gamma}^T(s) X_i(s) + \hat{\beta}^T Z_i(s)\}$. Let $\hat{E}_{xx}(t) = n^{-1} \sum_{i=1}^{n} \int_0^{\tau} W_i(s) K_h(s-t) \ \hat{\dot{\mu}}_i(s) (X_i(s))^{\otimes 2} \, dN_i(s)$ and $\hat{E}_{xz}(t) = n^{-1} \sum_{i=1}^{n} \int_0^{\tau} W_i(s) K_h(s - t) \hat{\dot{\mu}}_i(s) X_i(s) (Z_i(s))^T \, dN_i(s)$. The following theorem presents the consistency and asymptotic normality of $\hat{\beta}$.

Theorem 1. *Assume that Condition A holds. Then*

(a) $\hat{\beta} \xrightarrow{P} \beta_0$ *as* $n \to \infty$*;*

(b) $n^{1/2}(\hat{\beta}-\beta_0)\xrightarrow{\mathcal{D}}N(0, A^{-1}\Sigma A^{-1})$ *as* $nh^2\to\infty$ *and* $nh^5=O(1)$.

The matrix A *can be consistently estimated by*

$$\hat{A}=n^{-1}\sum_{i=1}^{n}\int_{t_1}^{t_2}W_i(s)\hat{\dot{\mu}}_i(s)\{Z_i(s)-(\hat{E}_{xz}(s))^T(\hat{E}_{xx}(s))^{-1}X_i(s)\}^{\otimes 2}\,dN_i(s),$$

and Σ *can be consistently estimated by*

$$\hat{\Sigma}=n^{-1}\sum_{i=1}^{n}\left(\int_{t_1}^{t_2}W_i(s)\{Y_i(s)-\hat{\mu}_i(s)\}\{Z_i(s)-(\hat{E}_{xz}(s))^T(\hat{E}_{xx}(s))^{-1}X_i(s)\}\,dN_i(s)\right)^{\otimes 2}.$$

Under Theorem 1, the proposed estimator $\hat{\beta}$ is consistent and asymptotically normal as long as the weight process $W(\cdot)$ converges in probability to a deterministic function $w(\cdot)$. The selection of $W(\cdot)$ plays a role in the variance of the estimator $\hat{\beta}$. Naturally, we would like to choose the optimal weight such that the asymptotic variance of $\hat{\beta}$ is minimized. This selection is usually difficult. It depends on the correlation structure of the longitudinal data among other things. Suppose that the repeated measurements of $Y_i(\cdot)$ within the same subject are independent and that $Y_i(\cdot)$ is independent of $N_i(\cdot)$ conditional on the covariates $X_i(t)$ and $Z_i(t)$. Let $\sigma_\epsilon^2(t|X_i,Z_i)=\text{Var}\{Y_i(t)|X_i(t),Z_i(t)\}$ be the conditional variance of $Y_i(t)$ given the covariates $X_i(t)$ and $Z_i(t)$ under model (1). Then the matrix $\Sigma=E[\int_{t_1}^{t_2}w_i^2(s)\sigma_\epsilon^2(s|X_i,Z_i)\ \{Z_i(s)-(e_{xz}(s))^T(e_{xx}(s))^{-1}\ X_i(s)\}^{\otimes 2}\alpha_i(s)\xi_i(s)\,ds]$. Let $\Sigma_0=E[\int_{t_1}^{t_2}\{\dot{\mu}_i(s)/\sigma_\epsilon(s|X_i,Z_i)\}^2\{Z_i(s)-(e_{xz}(s))^T(e_{xx}(s))^{-1}\ X_i(s)\}^{\otimes 2}\alpha_i(s)\xi_i(s)\,ds]$. We show in the Appendix that

$$A^{-1}\Sigma A^{-1}-\Sigma_0^{-1}\geq 0, \tag{9}$$

where $B\geq 0$ means that the matrix B is nonnegative definite. When $w_i(t)=\dot{\mu}_i(t)/\{\sigma_\epsilon(t|X_i,Z_i)\}^2$, $A=\Sigma=\Sigma_0$ and the equality in (9) holds. The situation often leads to asymptotically efficient estimators in many semiparametric models discussed by Bickel et al. (1993).

Next, we state an asymptotic result for the estimator $\hat{\gamma}(t)$. The result is useful for constructing confidence intervals for the mean response curve given the covariates. Denote $\dot{\gamma}_0(t)$, $\ddot{\gamma}_0(t)$ the first and second derivatives of $\gamma_0(t)$ with respect to t, respectively.

Theorem 2. *Under Condition A,* $\hat{\gamma}(t) \xrightarrow{P} \gamma_0(t)$,

$$\sqrt{nh}(\hat{\gamma}(t) - \gamma_0(t) - \frac{1}{2}\mu_2 h^2 \ddot{\gamma}_0^T(t)) \xrightarrow{\mathcal{D}} N\left(0, \Sigma_\gamma(t)\right),$$

as $nh^2 \to \infty$ *and* $nh^5 = O(1)$ *for* $t \in (0, \tau)$, *where* $\mu_2 = \int_{-1}^1 t^2 K(t)\, dt$, $\Sigma_\gamma(t) = (e_{xx}(t))^{-1}\Sigma_e(t)$ $(e_{xx}(t))^{-1}$, $\Sigma_e(t) = \lim_{n\to\infty} hE\{\int_0^\tau w_i(s)\{Y_i(s) - \mu_i(s)\}\ X_i(s)K_h(s-t)\, d\,N_i(s)\}^{\otimes 2}$. *The covariance matrix* $\Sigma_\gamma(t)$ *can be estimated consistently by* $\hat{\Sigma}_\gamma(t) = n^{-1}\sum_{i=1}^n \{\hat{g}_i(t)\}^{\otimes 2}$, *where*

$$\hat{g}_i(t) = h^{1/2}(\hat{E}_{xx}(t))^{-1}\int_0^\tau W_i(s)K_h(s-t)X_i(s)\{Y_i(s) - \hat{\mu}_i(s)\}\, d\,N_i(s) - h^{1/2}(\hat{E}_{xx}(t))^{-1}\hat{E}_{xz}(t)$$
$$\times \hat{A}^{-1}\int_{t_1}^{t_2} W_i(s)\{Z_i(s) - (\hat{E}_{xz}(s))^T(\hat{E}_{xx}(s))^{-1}X_i(s)\}\{Y_i(s) - \hat{\mu}_i(s)\}\, dN_i(s).$$

When the link function is the identity function, Sun and Wu (2005) showed that the asymptotic bias of using the profile kernel smoothing for $\gamma_0(t)$ is $\frac{1}{2}\mu_2 h^2\{\ddot{\gamma}_0^T(t) + 2(e_{xx}(t))^{-1}\dot{e}_{xx}(t)$ $\dot{\gamma}_0(t)\}$. This phenomenon parallels the situation described in Fan and Gijbels (1996, p.17) for the nonparametric regression with cross-sectional data that compares the Nadaraya-Watson estimator and the local linear estimator. The extra term in the bias of $\hat{\gamma}(t)$ using profile kernel smoothing depends on $(e_{xx}(t))^{-1}\dot{e}_{xx}(t)\dot{\gamma}_0(t)$. The bias of the profile kernel smoothing estimator can be large in the highly asymmetric design where $(e_{xx}(t))^{-1}\dot{e}_{xx}(t)\dot{\gamma}_0(t)$ is large. On the other hand, the bias of the profile local linear smoothing estimator only involves the second derivative $\ddot{\gamma}_0(t)$, thus is design-adaptive. Another advantage of the local linear smoothing over the kernel smoothing, as discussed in Fan and Gijbels (1996), is the automatic boundary adaption. The rate of convergence at boundary points using the local linear smoothing is same as for the interior points, which can be shown to hold for model (1) with longitudinal data as well.

Let $\Gamma_0(t) = \int_{t_1}^t \gamma_0(s)\, ds$ and $\hat{\Gamma}(t) = \int_{t_1}^t \hat{\gamma}(s)\, ds$. The following theorem presents a weak convergence result for $G_n(t) = n^{1/2}(\hat{\Gamma}(t) - \Gamma_0(t))$ over $t \in [t_1, t_2]$. This result provides theoretical justifications for testing the regression coefficient functions $\gamma(t)$ and for the construction of simultaneous confidence bands of $\Gamma(t) = \int_{t_1}^t \gamma(s)\, ds$ developed later.

Theorem 3. *Under Condition A,* $G_n(t) = n^{-1/2}\sum_{i=1}^n H_i(t) + o_p(1)$ *uniformly in* $t \in [t_1, t_2] \subset$

$(0,\tau)$ *as* $nh^2 \to \infty$ *and* $nh^5 \to 0$*, where*

$$\begin{aligned} H_i(t) &= \int_{t_1}^{t} (e_{xx}(s))^{-1} \int_0^{\tau} w_i(u) K_h(u-s) X_i(u) \{Y_i(u) - \mu_i(u)\}\, dN_i(u)\, ds \\ &\quad - \int_{t_1}^{t} (e_{xx}(s))^{-1} e_{xz}(s)\; ds\, A^{-1} \\ &\quad \times \int_{t_1}^{t_2} w_i(s) \{Z_i(s) - (e_{xz}(s))^T (e_{xx}(s))^{-1} X_i(s)\} \{Y_i(s) - \mu_i(s)\}\, dN_i(s). \end{aligned} \tag{10}$$

The processes $G_n(t)$ *converges weakly to a zero-mean Gaussian process* $G(t)$ *on* $[t_1, t_2]$*. The asymptotic covariance matrix of* $G_n(t)$ *can be estimated consistently by* $\hat{\Sigma}_G(t) = n^{-1} \sum_{i=1}^n \{\hat{H}_i(t)\}^{\otimes 2}$*, where*

$$\begin{aligned} \hat{H}_i(t) &= \int_{t_1}^{t} (\hat{E}_{xx}(s))^{-1} \int_0^{\tau} W_i(u) K_h(u-s) X_i(u) \{Y_i(u) - \hat{\mu}_i(u)\}\, dN_i(u)\, ds \\ &\quad - \int_{t_1}^{t} (\hat{E}_{xx}(s))^{-1} \hat{E}_{xz}(s)\; ds\, \hat{A}^{-1} \\ &\quad \times \int_{t_1}^{t_2} W_i(s) \{Z_i(s) - (\hat{E}_{xz}(s))^T (\hat{E}_{xx}(s))^{-1} X_i(s)\} \{Y_i(s) - \hat{\mu}_i(s)\}\, dN_i(s). \end{aligned} \tag{11}$$

3.2 Confidence intervals and simultaneous confidence bands

Let $\gamma^{(k)}(t)$ be the kth component of $\gamma(t)$. Similar notations are used throughout with the superscript (k) denoting the kth component of the corresponding vector. Assuming $nh^5 \to 0$, based on Theorem 2, the under-smoothing avoids estimating the second derivative $\ddot{\gamma}(t)$ and controls the size of the bias term. The large sample pointwise confidence intervals for $\gamma^{(k)}(t)$, $0 < t < \tau$, is obtained by

$$\hat{\gamma}^{(k)}(t) \pm (nh)^{-1/2} z_{\alpha/2} \left[n^{-1} \sum_{i=1}^n \{\hat{g}_i^{(k)}(t)\}^2 \right]^{1/2}. \tag{12}$$

By Theorem 3, the pointwise confidence intervals for $\Gamma^{(k)}(t)$, $0 < t < \tau$, is given by

$$\hat{\Gamma}^{(k)}(t) \pm n^{-1/2} z_{\alpha/2} \left[n^{-1} \sum_{i=1}^n \{\hat{H}_i^{(k)}(t)\}^2 \right]^{1/2}. \tag{13}$$

Furthermore, based on Theorem 3, simultaneous confidence bands and hypothesis tests related to the regression coefficient functions $\gamma(t)$ can be constructed. A key component is the estimation of confidence coefficients and the critical values. The Gaussian multiplier resampling method of Lin, Wei and Ying (1993) has been widely employed for this purpose and is described in the following.

Let $G_n^*(t) = n^{-1/2}\sum_{i=1}^n \hat{H}_i(t)\xi_i$, where $\xi_1, \xi_2, \ldots, \xi_n$ are independent identically distributed (iid) standard normal random variables independent from the observed data set. By Lemma 1 of Sun and Wu (2005), the processes $G_n(t)$ and $G_n^*(t)$ given the observed data sequence converge weakly to the same zero-mean Gaussian process on $[t_1, t_2]$. To approximate the distribution of $G_n(t)$, we simulate a large number of realizations from $G_n^*(t)$ by repeatedly generating $(\xi_1, \ldots, \xi_n)$ while fixing $\{Y_i(t), X_i(t), Z_i(t), N_i(t)), t \geq 0\}$ at their observed values. Let c_α be the $(1-\alpha)$-quantile of $\sup_{t_1 \leq t \leq t_2} |G_n^{*(k)}(t)/[\sum_{i=1}^n \{\hat{H}_i^{(k)}(t)\}^2/n]^{1/2}|$, which can be approximated by repeatedly generating independent normal samples $(\xi_1, \ldots, \xi_n)$. An asymptotic $1-\alpha$ simultaneous confidence bands for $\Gamma^{(k)}(t)$ on $[t_1, t_2]$ is given by

$$\hat{\Gamma}^{(k)}(t) \pm n^{-1/2} c_\alpha \left[n^{-1} \sum_{i=1}^n \{\hat{H}_i^{(k)}(t)\}^2 \right]^{1/2}. \quad (14)$$

3.3 Hypothesis testing of regression coefficients

The generalized semiparametric regression model (1) postulates that the covariates effects are constant for some and are time-varying for others. A formal hypothesis testing procedure can be established to check whether the effect of a covariate is time-varying under model (1). This can lead to more efficient estimation when the effects of some covariates are not really time-varying. We consider testing the null hypothesis H_0 that $\gamma^{(k)}(t)$ is constant for $0 \leq t \leq \tau$.

Under H_0, $\Gamma^{(k)}(t) - \frac{t-t_1}{t_2-t_1}\Gamma^{(k)}(t_2) = 0$ for $t \in [t_1, t_2]$. By Theorem 3 and the continuous mapping theorem,

$$n^{1/2}\left\{\hat{\Gamma}^{(k)}(t) - \frac{t-t_1}{t_2-t_1}\hat{\Gamma}^{(k)}(t_2)\right\} = n^{1/2}\left\{\hat{\Gamma}^{(k)}(t) - \Gamma^{(k)}(t)\right\} - \frac{t-t_1}{t_2-t_1} n^{1/2}\left\{\hat{\Gamma}^{(k)}(t_2) - \Gamma^{(k)}(t_2)\right\}$$

converges weakly to $G^{(k)}(t) - \frac{t-t_1}{t_2-t_1}G^{(k)}(t_2)$, where $G^{(k)}(t)$ is the kth component of the limiting Gaussian process $G(t)$ of $n^{1/2}\{\hat{\Gamma}(t) - \Gamma(t)\}$. The rationale leads to the following constructions of the test statistics:

$$S = \sup_{t_1 \le t \le t_2} n^{1/2} \left| \hat{\Gamma}^{(k)}(t) - \frac{t - t_1}{t_2 - t_1} \hat{\Gamma}^{(k)}(t_2) \right|$$

and

$$L = \int_{t_1}^{t_2} n \left\{ \hat{\Gamma}^{(k)}(t) - \frac{t - t_1}{t_2 - t_1} \hat{\Gamma}^{(k)}(t_2) \right\}^2 dt.$$

By the continuous mapping theorem, under H_0, the test statistic S converges in distribution to $\sup_{t_1 \le t \le t_2} \left| G^{(k)}(t) - \frac{t-t_1}{t_2-t_1} G^{(k)}(t_2) \right|$, and the test statistic S converges in distribution to $\int_{t_1}^{t_2} \left\{ G^{(k)}(t) - \frac{t-t_1}{t_2-t_1} G^{(k)}(t_2) \right\}^2 dt$. The two test statistics are commonly used in statistics literature with S referred as the supremum type and L as the integrated square type, cf., Martinussen and Scheike (2006).

Let

$$S^* = \sup_{t_1 \le t \le t_2} n^{1/2} \left| G_n^{*(k)}(t) - \frac{t - t_1}{t_2 - t_1} G_n^{*(k)}(t_2) \right|$$

and

$$L^* = \int_{t_1}^{t_2} n \left\{ G_n^{*(k)}(t) - \frac{t - t_1}{t_2 - t_1} G_n^{*(k)}(t_2) \right\}^2 dt.$$

The critical values of S and L can be approximated by simulating a number of copies of S^* and L^* obtained by repeatedly generating independent normal samples $(\xi_1, \ldots, \xi_n)$ while holding the observed data fixed. For example, the critical values of test statistics S and L at the significance level α can be estimated by the upper α quantile of, say 1000, copies of S^* and L^*, respectively. The p-values of the tests based on S and L are the percentages of S^* and L^* exceeding S and L, respectively. The null hypothesis is rejected if the p-values are less than α.

3.4 Cross-validation bandwidth selection

Let $\sigma^{(k)}(t)$ be the (k,k)th element of $\Sigma_\gamma(t)$. It follows from Theorem 2 that the mean integrated square error for estimating the kth component $\gamma^{(k)}(t)$ over $[t_1, t_2]$ is

$$\int_{t_1}^{t_2} \left[\frac{1}{4} \mu_2^2 \{ \ddot{\gamma}_0^{(k)}(t) \}^2 h^4 + \frac{1}{nh} \sigma^{(k)}(t) \right] dt.$$

The asymptotic optimal bandwidth is given by

$$h_{opt,k} = \left[\frac{\int_{t_1}^{t_2} \sigma^{(k)}(t)\, dt}{\int_{t_1}^{t_2} \mu_2^2 \{\ddot{\gamma}_0^{(k)}(t)\}^2} \right]^{1/5} n^{-1/5}.$$

The optimal theoretical bandwidth is difficult to achieve since it involves estimating the second derivative $\ddot{\gamma}_0^{(k)}(t)$. In practice, the appropriate bandwidth selection can be based on a cross-validation method. This approach is widely used in nonparametric function estimation literature, see Rice and Silverman (1991) for leave-one-subject-out cross-validation approach and Tian, Zucker and Wei (2005) for K-fold cross-validation approach.

An analog of the K-fold cross-validation approach in the current setting is to divide the data into K equal-sized groups. Let D_k denote the kth subgroup of data, then the kth prediction error is given by

$$PE_k(h) = \sum_{i \in D_k} \int_{t_1}^{t_2} \left[Y_i(t) - \varphi\{(\hat{\gamma}_{(-k)}(t))^T X_i(t) + \hat{\beta}_{(-k)}^T Z_i(t)\} \right]^2 dN_i(t), \qquad (15)$$

for $k = 1, \ldots, K$, where $\hat{\gamma}_{(-k)}(t)$ and $\hat{\beta}_{(-k)}$ are the estimators of $\gamma_0(t)$ and β_0 based on the data without the subgroup D_k. The data-driven bandwidth selection based on the K-fold cross-validation is to choose the bandwidth h that minimizes the total prediction error $PE(h) = \sum_{k=1}^{K} PE_k(h)$. As we show in Section 5 in the analysis of a HIV-1 RNA data set from an AIDS clinical trial, the K-fold cross-validation bandwidth selection provides a working tool for locating an appropriate bandwidth.

4 A simulation study

In this section, we examine finite sample properties of the estimation and hypothesis testing procedures proposed for model (1). The performances of the estimators for β and $\gamma(t)$ at a fixed time t are measured through the bias, the sample mean of the estimated standard errors (ESE), the sample standard error of the estimators (SEE) and the 95% empirical coverage probability (CP). To evaluate the overall performance of the estimator $\hat{\gamma}^{(k)}(t)$ on the interval $[h, \tau - h]$, we consider the square root of integrated mean square error $RMSE_k = \left\{ \frac{1}{N(\tau-2h)} \sum_{j=1}^{N} \int_h^{\tau-h} (\hat{\gamma}_j^{(k)}(t) - \gamma_0^{(k)}(t))^2\, dt \right\}^{1/2}$, where N is the repetition number, $\hat{\gamma}_j^{(k)}(t)$ is the jth estimate of $\gamma^{(k)}(t)$ for $j = 1, \ldots, N$. We use the unit weight function and the Epanechnikov kernel $K(u) = 0.75(1 - u^2)I(|u| \leq 1)$ throughout the simulation. We take $t_1 = 0$ and $t_2 = \tau$ in the estimating functions (4) and (8).

The performance of the estimators are examined under the following selected setting of model (1), in which we take the link function $g(x) = \ln(x)$:

$$Y_i(t) = \exp\{0.5\sqrt{t} + 0.5\sin(2t)X_i + \beta Z_i\} + \varepsilon_i(t), \qquad i = 1, \ldots, n, \tag{16}$$

for $0 \leq t \leq \tau$ with $\tau = 3.5$, where X_i is a Bernoulli random variable with the success probability of 0.5, Z_i is uniformly distributed on $(0, 1)$, $\varepsilon_i(t)$ is $N(\phi_i, 0.5^2)$ conditional on ϕ_i and ϕ_i is $N(0, 1)$. Here $\gamma(t) = (\gamma_1(t), \gamma_2(t))^T$ with $\gamma_1(t) = 0.5t^{1/2}$ and $\gamma_2(t) = 0.5\sin(2t)$.

We consider three models for the sampling times. The first model is a Poison process with the proportional mean rate

$$\alpha(t|X_i, Z_i) = 0.6\exp(0.7Z_i), \quad i = 1, \ldots, n. \tag{17}$$

The second model is a Poison process with the additive mean rate

$$\alpha(t|X_i, Z_i) = 0.4 + 0.9Z_i, \quad i = 1, \ldots, n. \tag{18}$$

To examine the performance of the proposed method when the sampling strategy depends on the past history, we consider a nonhomogeneous poisson process for the sampling times with the intensity function

$$\lambda(t|Z_i, Z_i^*) = 0.12t^{1/2}\exp\{2Z_i + Z_i^*(t)\}, \tag{19}$$

where Z_i is uniform on $(0, 1)$ and $Z_i^*(t) = 1$ if there was an event within the interval $[t-1, t)$ and 0 otherwise. For all the three sampling models, the censoring times C_i are generated from $U(1.5, 8)$. There are approximately 3 observations per subject in the interval $[0, \tau]$ and about 30% subjects are censored before $\tau = 3.5$.

Table 1 summarizes the bias, SEE, ESE and CP for β and RMSE for $\gamma(t)$ under the models (16) and (17). The integrals are evaluated on the grid points $s_i = 0.05i, i = 1, 2, \ldots, 69$. The summaries of performance of $\hat{\gamma}(t)$ at time points $0.5j$, $j = 1, \ldots, 6$, are given in Table 2.

The summaries of performances of $\hat{\beta}$ and $\hat{\gamma}(t)$ under the models (16) and (18) are presented in Tables 3 and 4. The summaries under the models (16) and (19) are presented in (5) and (6). Each entry of the tables is calculated based on 1000 repetitions.

It can be seen from Tables 1-6 that the proposed estimation procedures performed well for three sampling situations considered here. It appears that the estimates are unbiased and there is a good agreement between the estimated and empirical standard errors. The empirical coverage probabilities are reasonable and the results become better when the sample size increases from 100 to 300. Plots of $\hat{\gamma_1}(t)$ and $\hat{\gamma_2}(t)$ for model (16) are depicted in Figure 1 when $\beta = 0.5$ for $n = 100$ and $h = 0.3$. Figure 1 (a) and (b) in the first row are under the proportional sampling model (17), Figure 1 (c) and (d) in the second row are under the additive sampling model (18) and Figure 1 (e) and (f) in the third row are under the sampling model (19). The estimators $\hat{\gamma_1}(t)$ and $\hat{\gamma_2}(t)$ are essentially unbiased. These figures also show that the proposed estimation procedures perform well for the nonparametric components under these three different sampling models.

The following models are considered to evaluate the performance of the test statistics S and L:

$$Y_i(t) = \exp\{0.5\sqrt{t} + \{0.5 - \theta \sin(2t)\}X_i + 0.5Z_i\} + \varepsilon_i(t), \qquad i = 1, \ldots, n, \tag{20}$$

for $0 \leq t \leq \tau$, where the distributions of X_i, Z_i and $\varepsilon_i(t)$ are same as those given in model (16). Different values of θ are to be selected to examine the power of the tests.

Our null hypothesis is that the effect of X_i does not change with time. The observed sizes of the test statistics are calculated under $\theta = 0$. The powers of the tests are evaluated at $\theta = 0.1, 0.15$ and 0.2. Table 7 lists the empirical sizes and powers of the test statistics S and L at the significance level 0.05 under the sampling models (17), (18) and (19). Each entry is based on 1000 repetitions. Each p-value is estimated by generating 1000 independent Gaussian random samples. The empirical sizes of both the tests are reasonably close to the 0.05 nominal level. The empirical power increases when sample size increases. There is also an increased power when θ increases, which represents an increased time-varying effect under model (20). Again, the performances of the tests are robust to the models of sampling times.

Finally, we conduct a small simulation study under the identity link function to compare with the joint modelling method of Lin and Ying (2001) in which the sampling times are modelled through the proportional mean rate model. We consider the following model for the longitudinal response

$$Y_i(t) = \alpha(t) + Z_i + \varepsilon_i(t), \tag{21}$$

where Z_i and $\varepsilon_i(t)$ are same as those for model (16). Table 8 list the summaries of the estimation for $\beta = 1$ using the method of Lin and Ying (2001) (L&Y) and the proposed method with $h = 0.3$, 0.4 and 0.5 when the sampling times are generated from model (17), (18) and (19). Each entry is based on 1000 repetitions. The estimation of Lin and Ying (2001) has larger biases when the sampling model is mis-specified under (18) and (19), especially when the sampling strategy depends on the past history and the intercept $\alpha(t)$ varies more. In all the cases, Lin and Ying (2001) estimation yields large variances compared to the proposed method.

5 An application

We apply the methods developed in the previous sections to a HIV-1 RNA data set from an AIDS clinical trial for comparing a single protease inhibitor (PI) versus a double-PI antiretroviral regimens in treating HIV-infected patients. In this study, all subjects initiated the antiretroviral treatment at time 0 (baseline) and HIV-1 RNA levels in plasma (viral load) was measured repeatedly over time. The scheduled visits for the measurements were at weeks 0, 2, 4, 8, 16 and 24. However, the actual time of visits for individual subjects may vary around the scheduled visiting times. Some patients had prior antiviral treatment with non-nucleoside analogue reverse transcriptase inhibitors (NNRTIs) (indicated by a covariate $Z = 1$) and others did not have prior NNRTI treatment ($Z = 0$). The prior treatment experience is considered to be a factor that affects the antiviral response to the antiretroviral regimens in the current study. Let $X = 1$ to indicate the patients who received a double-PI treatment and $X = 0$ for patients who received a single-PI treatment in this study. Our interest is to compare the HIV viral load responses of the double-PI treatment with those of the single-PI treatment.

A total of 481 patients were enrolled in the study, with 2626 total visits. Owing to technical limitations, 175 responses were censored below the detection limit, or 6.67% and three responses were censored above the detection limit or 0.11%. The handling of the censored viral load data needs more complicated statistical methods and is out of the scope of this paper. We restrict our analysis to those responses within the detectable range for the purpose of illustrating the proposed methodologies. The average number of visits was 5.01 for treatment group $X = 1$ and 5.25 for treatment group $X = 0$. The scheduled durations between visits get longer at later times of the study. we consider the transformed time scale $t = \log_{10}(\text{day of actual visit } + 40) - \log_{10}(32)$. This kind of transformation is often used to make independent variables equally spaced for convenient bandwidth selections in non-parametric regression analysis. The transformation is also used to accommodate the fact that the data for actual visit times may have negative values. A value of -7 for the first actual visit indicates that the patient visited the clinic 7 days before the first scheduled visit. The response variable $Y(t)$ is the change of HIV-1 RNA level using a $\log_{10}$ scale at time t from the baseline. Here $\log_{10}$ scale of viral load is commonly used by AIDS researchers and is also good for stabilizing the variance of measurement errors.

The data is fitted to the following the model

$$Y_i(t) = \exp\{\gamma_1(t) + \gamma_2(t)X_i + \beta Z_i\} + \varepsilon_i(t), \tag{22}$$

for $0 \le t \le \tau$ with $\tau = 0.88$, the maximum of transformed observation times. We set $t_1 = 0.2$ and $t_2 = \tau - 0.2$ in (4) for the estimation of β. The bandwidth selected using $\mathcal{K}$-fold cross-validation method presented in Section 3.4 using $K = 13$ yields $h = 0.05$; see Figure 2 (a) for the plot of the total prediction error. With $h = 0.05$, the value of $\hat{\beta}$ is 0.1643 and the standard error is 0.0230. The estimators $\hat{\gamma}_1(t)$ and $\hat{\gamma}_2(t)$ and the 95% pointwise confidence intervals are plotted in the first row in Figure 3. The p-values for testing for time-dependence of $\gamma_2(t)$ are 0.003 and 0.005 for test statistics S and L, respectively, based on 1000 Gaussian samples.

To show how the estimates are affected by choices of bandwidth, we plot $\hat{\beta}$ against h in

Figure 2 (b). At $h = 0.11$, the value of $\hat{\beta}$ is 0.1599 and the standard error is 0.0233. The p-values for testing for time-dependence of $\gamma_2(t)$ are 0.051 and 0.035 for test statistics S and L, respectively. For $h = 0.2$, the value of $\hat{\beta}$ is 0.1569 and the standard error is 0.0233. The p-values for testing for time-dependence of $\gamma_2(t)$ are 0.0550 for both test statistics S and L. The estimators $\hat{\gamma}_1(t)$ and $\hat{\gamma}_2(t)$ and the 95% pointwise confidence intervals at $h = 0.11$ and 0.2 are plotted in the second and third rows of Figure 3, respectively.

Our hypothesis tests indicate that the treatment effect changes with time. The double-PI antiretroviral regimens works better than the single PI regimens in reducing viral load in treating HIV-infected patients and this effect becomes stronger over time during the course of the study as shown in Figure 3. The patients who had prior antiviral treatment with NNRTIs tend to have higher level of viral load than those who did not have the prior treatment.

Our experience shows that the "optimal" bandwidth that minimizes the total prediction error tends to be a little small to yield smoothed curves for the nonparametric regression coefficient functions. The values of the estimators of the parametric components are not greatly affected by the choices of the bandwidth and tend to stabilize for larger bandwidths. Nevertheless, the K-fold cross-validation bandwidth selection provides a working tool for locating an appropriate bandwidth.

6 Appendix

We assume the following conditions throughout the paper:

Condition A. The covariate processes $X_i(\cdot)$ and $Z_i(\cdot)$ are left continuous; The censoring time C_i is noninformative in the sense that $E\{dN_i^*(t)\,|X_i(t), Z_i(t), C_i \geq t\} = E\{dN_i^*(t)|X_i(t), Z_i(t)\}$ and $E\{Y_i(t)|X_i(t), Z_i(t), C_i \geq t\} = E\{Y_i(t)|X_i(t), Z_i(t)\}$; $dN_i^*(t)$ is independent of $Y_i(t)$ conditional on $X_i(t)$, $Z_i(t)$ and $C_i \geq t$; the processes $Y_i(t)$, $X_i(t)$, $Z_i(t)$ and $\alpha_i(t)$, $0 \leq t \leq \tau$, are bounded and their total variations are bounded by a constant; $E|N_i(t_2) - N_i(t_1)|^2 \leq L(t_2 - t_1)$ for $0 \leq t_1 \leq t_2 \leq \tau$, where $L > 0$ is a constant; the link function $g(y)$ is monotone and its inverse function $g^{-1}(x)$ is twice differentiable; $\gamma_0(t)$, $e_{xx}(t)$ and $e_{xz}(t)$ are twice differentiable; $(e_{xx}(t))^{-1}$ is bounded over $0 \leq t \leq \tau$; the matrices A and Σ are positive definite;

the weight process $W(t,x,z) \xrightarrow{P} w(t,x,z)$ uniformly in the range of (t,x,z); $w(t,x,z)$ is differentiable with uniformly bounded partial derivatives; the kernel function $K(\cdot)$ is symmetric with compact support on $[-1,1]$ and bounded variation; bandwidth $h \to 0$; $E|N_i(t+h) - N_i(t-h)|^{2+v} = O(h)$, for some $v > 0$; the limit $\lim_{n\to\infty} hE\{\int_0^\tau w_i(s)\{Y_i(s) - \mu_i(s)\} X_i(s)K_h(s-t)\, d\, N_i(s)\}^{\otimes 2} = \Sigma_e(t)$ exists and is finite.

Let $u_a(\gamma,\beta) = E([\varphi\{\gamma_0^T(t)X_i(t) + \beta_0^T Z_i(t)\} - \varphi\{\gamma^T(t)X_i(t) + \beta^T Z_i(t)\}]X_i(t)\xi_i(t)\alpha_i(t))$. Define $\gamma_\beta(t)$ as the unique root such that $u_a(\gamma_\beta,\beta) = 0$ for $\beta \in \mathcal{N}_\beta$ where $\mathcal{N}_\beta$ is a neighborhood of β_0. Let $e_{\beta,xx}(t) = E\big[w_i(t)\dot\varphi\{\gamma_\beta^T(t)X_i(t) + \beta^T Z_i(t)\}\{X_i(t)\}^{\otimes 2}\alpha_i(t)\xi_i(t)\big]$ and $e_{\beta,xz}(t) = E\big[w_i(t)\dot\varphi\{\gamma_\beta^T(t)X_i(t) + \beta^T Z_i(t)\}X_i(t)(Z_i(t))^T\alpha_i(t)\xi_i(t)\big]$. When $\beta = \beta_0$, we have $\gamma_\beta(t) = \gamma_0(t)$. In this case, $e_{\beta,xx}(t) = e_{xx}(t)$ and $e_{\beta,xz}(t) = e_{xz}(t)$. Let $\gamma_{a\beta}(t) = (\gamma_\beta^T(t), \mathbf{0}_q^T)^T$ where $\mathbf{0}_q$ is a $q \times 1$ vector of zeros.

Let $H = \mathrm{diag}\{I_q, hI_q\}$. The following lemmas are used in the proofs of the main theorems. The proofs of the lemmas make repeated applications of the Glivenko-Cantelli Theorem (Theorem 19.4 of van der Vaart, 1998). A sufficient condition for applying the Glivenko-Cantelli Theorem can be checked by estimating the order of the bracketing number, similar to the proof of Lemma 2 of Sun, Gilbert and McKeague (2009). This sufficient condition holds under the conditions provided in Condition A. The details are omitted to save space.

Lemma 1. *Assume that Condition A holds. Then as* $n \to \infty$*,* $H\tilde\gamma_a(t,\beta) \xrightarrow{P} \gamma_{a\beta}(t)$*,*

$$H\frac{\partial\tilde\gamma_a(t,\beta)}{\partial\beta} \xrightarrow{P} \begin{pmatrix} -(e_{\beta,xx}(t))^{-1}e_{\beta,xz}(t) \\ \mathbf{0}_q \end{pmatrix},$$

and $H\partial^2\tilde\gamma(t,\beta)/\partial\beta^2$ *converges in probability to a deterministic function of* (t,β) *of bounded variation, uniformly in* $t \in [t_1,t_2] \subset (0,\tau)$ *and* $\beta \in \mathcal{N}_\beta$ *at the rate* $n^{-1/2+\nu}$ *for* $\nu > 0$*.*

Proof of Lemma 1.

To simplify the presentations, we use the notations $\gamma_{a\beta}$ and γ_β for $\gamma_{a\beta}(t)$ and $\gamma_\beta(t)$, respectively. Let $\theta = H(\gamma_a - \gamma_{a\beta})$ and $\tilde\theta = H(\tilde\gamma_a(t,\beta) - \gamma_{a\beta})$. By (3), $\tilde\theta$ is the root of the following estimating function for fixed β:

$$U_a(\gamma_{a\beta}+H^{-1}\theta,\beta) = \sum_{i=1}^n \int_0^\tau W_i(s)\{Y_i(s) - \tilde\mu_a(s,\gamma_{a\beta}+H^{-1}\theta,\beta|X_i,Z_i)\}\widetilde{X}_i(s,s-t)K_h(s-t)\, d\, N_i(s), \tag{23}$$

where $\tilde{\mu}_a(s, \gamma_{a\beta} + H^{-1}\theta, \beta | X_i, Z_i) = \varphi\{\theta^T \widetilde{U}_i(s, s-t) + \gamma_{a\beta}^T(t)\widetilde{X}_i(s, s-t) + \beta^T Z_i(s)\}$ and $\widetilde{U}_i(s, s-t) = H^{-1}\widetilde{X}_i(s, s-t)$.

By the Glivenko-Cantelli theorem,

$$
\begin{aligned}
&n^{-1}\{U_a(\gamma_{a\beta} + H^{-1}\theta, \beta) - U_a(\gamma_{a\beta}, \beta)\} \\
&= -n^{-1}\sum_{i=1}^{n}\int_0^{\tau} W_i(s)\{\tilde{\mu}_a(s, \gamma_{a\beta} + H^{-1}\theta, \beta | X_i, Z_i) - \tilde{\mu}_a(s, \gamma_{a\beta}, \beta | X_i, Z_i)\}\widetilde{X}_i(s, s-t)K_h(s-t)\, dN_i(s) \\
&\xrightarrow{P} -E\left(\int_{-1}^{1} w_i(t)\dot{\mu}_{i\beta}(t)\theta^T\{X_i^T(t), uX_i^T(t)\}^T\{X_i^T(t), 0\}^T K(u)\alpha_i(t)\xi_i(t)\, du\right),
\end{aligned}
$$

uniformly in $t \in [t_1, t_2]$, $\beta \in \mathcal{N}_\beta$ and $\theta \in \mathcal{N}_0$, a neighborhood of $\mathbf{0}_{2q} \in R^{2q}$, where $\dot{\mu}_{i\beta}(t) = \dot{\varphi}\{\gamma_\beta^T(t)X_i(t) + \beta^T Z_i(t)\}$. The limit has a unique root at $\theta = \mathbf{0}_{2q}$.

By the Glivenko-Cantelli theorem and (3), $n^{-1}U_a(\gamma_{a\beta}, \beta) \xrightarrow{P} \{u_a^T(\gamma_\beta, \beta), \mathbf{0}_q^T\}^T = \mathbf{0}_{2q}$. It follows by Lemma 1 of Sun, Gilbert and McKeague (2009) that $\tilde{\theta} \xrightarrow{P} \mathbf{0}_{2q}$ uniformly in t and β. Thus

$$
H\tilde{\gamma}_a(t, \beta) - \gamma_{a\beta}(t) \xrightarrow{P} \mathbf{0}_{2q} \quad \text{uniformly in } t \in [t_1, t_2] \text{ and } \beta \in \mathcal{N}_\beta. \tag{24}
$$

Since $U_a(\tilde{\gamma}_a(t, \beta), \beta) \equiv \mathbf{0}_{2q}$, $\tilde{\gamma}_a(t, \beta)$ satisfies

$$
\left\{\frac{\partial U_a(\gamma_a, \beta)}{\partial \gamma_a}\frac{\partial \tilde{\gamma}_a(t, \beta)}{\partial \beta} + \frac{\partial U_a(\gamma_a, \beta)}{\partial \beta}\right\}\bigg|_{\gamma_a = \tilde{\gamma}_a(t, \beta)} = \mathbf{0}_{2q}. \tag{25}
$$

Note that

$$
\begin{aligned}
&-n^{-1}H^{-2}\frac{\partial U_a(\gamma_a, \beta)}{\partial \gamma_a} \\
&= n^{-1}\sum_{i=1}^{n}\int_0^{\tau} W_i(s)\dot{\varphi}\{\gamma_a^T\widetilde{X}_i(s, s-t) + \beta^T Z_i(s)\}H^{-2}\{\widetilde{X}_i(s, s-t)\}^{\otimes 2}K_h(s-t)\, dN_i(s) \qquad (26) \\
&= n^{-1}\sum_{i=1}^{n}\int_0^{\tau} W_i(s)\dot{\varphi}\{(H\gamma_a)^T H^{-1}\widetilde{X}_i(s, s-t) + \beta^T Z_i(s)\}H^{-2}\{\widetilde{X}_i(s, s-t)\}^{\otimes 2}K_h(s-t)\, dN_i(s).
\end{aligned}
$$

By the Glivenko-Cantelli theorem, the process

$$
n^{-1}\sum_{i=1}^{n}\int_0^{\tau} W_i(s)\dot{\varphi}\{\eta^T H^{-1}\widetilde{X}_i(s, s-t) + \beta^T Z_i(s)\}H^{-2}\{\widetilde{X}_i(s, s-t)\}^{\otimes 2}K_h(s-t)\, dN_i(s)
$$

converges in probability to

$$
E\left[\int_0^{\tau} w_i(t)\dot{\varphi}[\eta^T\{X_i^T(t), uX_i^T(t)\}^T + \beta^T Z_i(t)]\begin{pmatrix} 1 & u \\ u & u^2 \end{pmatrix} \otimes \{X_i(t)\}^{\otimes 2}\xi_i(t)\alpha_i(t)K(u)\, du\right],
$$

uniformly in $t \in [t_1, t_2]$, $\beta \in \mathcal{N}_\beta$ and η in a neighborhood of $\gamma_{a\beta}(t)$ at the rate $n^{-1/2+\nu}$ for $\nu > 0$.

It follows from (24) that

$$-n^{-1}H^{-2}\frac{\partial U_a(\gamma_a, \beta)}{\partial \gamma_a}\bigg|_{\gamma_a=\tilde{\gamma}_a(t,\beta)} \xrightarrow{P} E\left[w_i(t)\dot{\varphi}\{\gamma_\beta^T(t)X_i(t)+\beta^T Z_i(t)\}\begin{pmatrix}1 & 0\\ 0 & \mu_2\end{pmatrix}\otimes \{X_i(t)\}^{\otimes 2}\xi_i(t)\alpha_i(t)\right],$$

uniformly in $t \in [t_1, t_2]$ and $\beta \in \mathcal{N}_\beta$ at the rate $n^{-1/2+\nu}$ for $\nu > 0$.

Similarly,

$$\begin{aligned}
&-n^{-1}H^{-1}\frac{\partial U_a(\gamma_a, \beta)}{\partial \beta}\bigg|_{\gamma_a=\tilde{\gamma}_a(t,\beta)} \\
&= n^{-1}\sum_{i=1}^{n}\int_0^\tau W_i(s)\dot{\varphi}\{\gamma_a^T \widetilde{X}_i(s, s-t)+\beta^T Z_i(s)\}H^{-1}\widetilde{X}_i(s, s-t)(Z_i(s))^T K_h(s-t)dN_i(s)\bigg|_{\gamma_a=\tilde{\gamma}_a(t,\beta)} \\
&\xrightarrow{P} \begin{pmatrix} E\left[w_i(t)\dot{\varphi}\{\gamma_\beta^T(t)X_i(t)+\beta^T Z_i(t)\}X_i(t)(Z_i(t))^T\xi_i(t)\alpha_i(t)\right] \\ \mathbf{0}_q \end{pmatrix},
\end{aligned} \tag{27}$$

uniformly in $t \in [t_1, t_2]$ and $\beta \in \mathcal{N}_\beta$ at the rate $n^{-1/2+\nu}$ for $\nu > 0$. It follows from (25) that

$$H\frac{\partial \tilde{\gamma}_a(t, \beta)}{\partial \beta} \xrightarrow{P} \begin{pmatrix} -(e_{\beta,xx}(t))^{-1}e_{\beta,xz}(t) \\ \mathbf{0}_q \end{pmatrix}, \tag{28}$$

at the rate $n^{-1/2+\nu}$ for $\nu > 0$, uniformly in $t \in [t_1, t_2]$ and $\beta \in \mathcal{N}_\beta$.

By a similar argument, $H\partial^2\tilde{\gamma}(t, \beta)/\partial\beta^2$ converges in probability to a deterministic function of (t, β) of bounded variation, uniformly in $t \in [t_1, t_2]$ and $\beta \in \mathcal{N}_\beta$. □

Lemma 2. *Under Condition A, as $nh \to \infty$ and $nh^5 = O(1)$,*

$$(nh)^{1/2}\{\tilde{\gamma}(t, \beta_0) - \gamma_0(t) - \frac{1}{2}\mu_2 h^2\ddot{\gamma}_0^T(t)\} = (e_{xx}(t))^{-1}(nh)^{1/2}n^{-1}U_\gamma(\gamma_0, \beta_0) + o_p(1), \tag{29}$$

uniformly in $t \in [t_1, t_2] \subset (0, \tau)$, where $\mu_2 = \int_{-1}^1 t^2K(t)\,dt$ and

$$U_\gamma(\gamma_0, \beta_0) = \sum_{i=1}^{n}\int_0^\tau W_i(s)\{Y_i(s) - \mu_i(s)\}X_i(s)K_h(s-t)\,dN_i(s).$$

Further, $(nh)^{1/2}n^{-1}U_\gamma(\gamma_0, \beta_0) = O_p(1)$ uniformly in $t \in [t_1, t_2] \subset (0, \tau)$.

Proof of Lemma 2.

Let $\gamma_{0a}(t) = (\gamma_0^T(t), \dot{\gamma}_0^T(t))^T$, $\rho_n = (nh)^{1/2}$ and $\theta = \rho_n H(\gamma_a - \gamma_{0a}(t))$. By the first order Taylor expansion, we have

$$
\begin{aligned}
& n^{-1}\{U_a(\gamma_{0a} + \rho_n^{-1} H^{-1}\theta, \beta_0) - U_a(\gamma_{0a}, \beta_0)\} \\
&= -n^{-1} \sum_{i=1}^{n} \int_0^\tau W_i(s)\{\tilde{\mu}_a(s, \gamma_{0a} + \rho_n^{-1} H^{-1}\theta, \beta_0 | X_i, Z_i) - \tilde{\mu}_a(s, \gamma_{0a}, \beta_0 | X_i, Z_i)\} \\
&\qquad\qquad \times \widetilde{X}_i(s, s-t) K_h(s-t)\, dN_i(s) \\
&= -n^{-1} \sum_{i=1}^{n} \int_0^\tau W_i(s)\{\rho_n^{-1}\theta^T \widetilde{U}_i(s, s-t)\}\dot{\varphi}\{\gamma_{0a}^T \widetilde{X}_i(s, s-t) + \beta_0^T Z_i(s)\} \\
&\qquad\qquad \times \widetilde{X}_i(s, s-t) K_h(s-t)\, dN_i(s) + o_p(\rho_n^{-1}\theta) \\
&= -n^{-1} \sum_{i=1}^{n} \int_0^\tau W_i(s)(\widetilde{X}_i(s, s-t))^{\otimes 2} \rho_n^{-1} H^{-1}\theta \dot{\varphi}\{\gamma_{0a}^T \widetilde{X}_i(s, s-t) + \beta_0^T Z_i(s)\} K_h(s-t)\, dN_i(s) \\
&\quad + o_p(\rho_n^{-1}\theta),
\end{aligned}
$$

which holds uniformly in $t \in [t_1, t_2]$. Since $\tilde{\theta} = \rho_n H(\tilde{\gamma}_a(t, \beta_0) - \gamma_{0a}(t))$ is the root of $U_a(\gamma_{0a} + \rho_n^{-1} H^{-1}\theta, \beta_0)$, it follows that $\tilde{\theta}$ equals

$$
\begin{aligned}
&\Big(n^{-1} \textstyle\sum_{i=1}^{n} \int_0^\tau W_i(s)(\widetilde{X}_i(s, s-t))^{\otimes 2} H^{-1} \dot{\varphi}\{\gamma_{0a}^T \widetilde{X}_i(s, s-t) + \beta_0^T Z_i(s)\} K_h(s-t) dN_i(s) + o_p(\rho_n^{-1})\Big)^{-1} \\
&\qquad \times \rho_n n^{-1} U_a(\gamma_{0a}, \beta_0).
\end{aligned}
$$

The first q components of $\tilde{\theta}$ yields

$$
\rho_n(\tilde{\gamma}(t, \beta_0) - \gamma_0(t)) = (e_{xx}(t))^{-1} \rho_n n^{-1} U_1(\gamma_{0a}, \beta_0) + o_p(\rho_n^{-1}), \tag{30}
$$

uniformly in $t \in [t_1, t_2]$, where

$$
U_1(\gamma_{0a}, \beta_0) = \sum_{i=1}^{n} \int_0^\tau W_i(s)\{Y_i(s) - \tilde{\mu}_a(s, \gamma_{0a}, \beta_0 | X_i, Z_i)\} X_i(s) K_h(s-t)\, d\, N_i(s).
$$

By the local linear approximation for $\gamma_0(s)$ around t,

$$
\begin{aligned}
& \mu_i(s) - \tilde{\mu}_a(s, \gamma_{0a}, \beta_0 | X_i, Z_i) \\
&= \varphi\{\gamma_0^T(s) X_i(s) + \beta_0^T Z_i(s)\} - \varphi[\{\gamma_0^T(t) + \dot{\gamma}_0^T(t)(s-t)\} X_i(s) + \beta_0^T Z_i(s)] \\
&= \dot{\mu}_i(s) \left\{ \frac{1}{2} \ddot{\gamma}_0^T(t) X_i(s)(s-t)^2 + O((s-t)^3) \right\} (1 + o_p(1)),
\end{aligned}
$$

as $s \to t$, where $\dot{\mu}_i(s) = \dot{\varphi}\{\gamma_0^T(s)X_i(s) + \beta_0^T Z_i(s)\}$. It follows that

$$
\begin{aligned}
&\rho_n n^{-1} \sum_{i=1}^n \int_0^\tau W_i(s)\{\mu_i(s) - \tilde{\mu}_a(s, \gamma_{0a}, \beta_0 | X_i, Z_i)\} X_i(s) K_h(s-t)\, d\,N_i(s) \\
&= \frac{1}{2}\mu_2 \rho_n h^2 E\{w_i(t)\dot{\mu}_i(t) X_i(t) X_i^T(t)\xi_i(t)\alpha_i(t)\}\ddot{\gamma}_0(t) + o_p(\rho_n h^2) \\
&= \frac{1}{2}\mu_2 \rho_n h^2 e_{xx}(t)\ddot{\gamma}_0(t) + o_p(\rho_n h^2),
\end{aligned}
$$

uniformly in $t \in [t_1, t_2]$. Hence

$$
\begin{aligned}
&\rho_n n^{-1} U_1(\gamma_{0a}, \beta_0) \\
&= \rho_n n^{-1} \sum_{i=1}^n \int_0^\tau W_i(s)[Y_i(s) - \mu_i(s) + \{\mu_i(s) - \tilde{\mu}_a(s, \gamma_{0a}, \beta_0 | X_i, Z_i)\}] X_i(s) K_h(s-t)\, d\,N_i(s) \\
&= \rho_n n^{-1} U_\gamma(\gamma_0, \beta_0) + \frac{1}{2}\mu_2 \rho_n h^2 e_{xx}(t)\ddot{\gamma}_0(t) + o_p(\rho_n h^2), \qquad (31)
\end{aligned}
$$

uniformly in $t \in [t_1, t_2]$. By (30) and (31),

$$
\rho_n\{\tilde{\gamma}(t, \beta_0) - \gamma_0(t) - \frac{1}{2}\mu_2 h^2 \ddot{\gamma}_0^T(t)\} = (e_{xx}(t))^{-1}\, \rho_n n^{-1} U_\gamma(\gamma_0, \beta_0) + o_p(\rho_n^{-1}) + o_p(\rho_n h^2), \quad (32)
$$

uniformly in $t \in [t_1, t_2]$.

Following the same lines as the proof in Appendix A of Tian, Zucker and Wei (2005), we get $(nh)^{1/2} n^{-1} U_\gamma(\gamma_0, \beta_0) = O_p(1)$ uniformly in $t \in [t_1, t_2] \subset (0, \tau)$. □

Proof of Theorem 1.

By Lemma 1 and application of the Glivenko-Cantelli theorem to the estimating function defined in (4), we have

$$
\begin{aligned}
&n^{-1} U(\beta) \\
&\xrightarrow{P} E\Big\{ \int_{t_1}^{t_2} w_i(s) \Big[Y_i(s) - \varphi\{(\gamma_\beta(s))^T X_i(s) + \beta^T Z_i(s)\} \Big] \\
&\qquad \times \Big[-(e_{\beta,xz}(s))^T (e_{\beta,xx}(s))^{-1} X_i(s) + Z_i(s) \Big]\, dN_i(s) \Big\} \\
&= E\Big\{ \int_{t_1}^{t_2} w_i(s) \Big[\varphi\{(\gamma_0(s))^T X_i(s) + \beta_0^T Z_i(s)\} - \varphi\{(\gamma_\beta(s))^T X_i(s) + \beta^T Z_i(s)\} \Big] \\
&\qquad \times \Big[-(e_{\beta,xz}(s))^T (e_{\beta,xx}(s))^{-1} X_i(s) + Z_i(s) \Big] \xi_i(s)\alpha_i(s)\, ds \Big\} \\
&\equiv u(\beta),
\end{aligned}
$$

uniformly for $\beta \in \mathcal{N}_\beta$. Since $u(\beta_0) = 0$ and A is positive definite, β_0 is the unique root of $u(\beta)$. By Theorem 5.9 of van der Vaart (1998), $\hat{\beta} \xrightarrow{P} \beta_0$.

By Lemma 1 and the Glivenko-Cantelli theorem,

$$n^{-1} \sum_{i=1}^{n} \int_{t_1}^{t_2} W_i(s)[Y_i(s) - \varphi\{(\tilde{\gamma}(s, \beta_0))^T X_i(s) + \beta_0^T Z_i(s)\}] \frac{\partial^2 \tilde{\gamma}(s, \beta_0)}{\partial \beta^2} X_i(s)\, dN_i(s) \xrightarrow{P} 0.$$

It follows that

$$\begin{aligned}
&-n^{-1} \frac{\partial U(\beta)}{\partial \beta}\bigg|_{\beta=\beta_0} \\
&= n^{-1} \sum_{i=1}^{n} \int_{t_1}^{t_2} W_i(s) \dot{\varphi}\{(\tilde{\gamma}(s, \beta_0))^T X_i(s) + \beta_0^T Z_i(s)\} \\
&\qquad \times \left\{ \left(\frac{\partial \tilde{\gamma}(s, \beta_0)}{\partial \beta} \right)^T X_i(s) + Z_i(s) \right\}^{\otimes 2} dN_i(s) + o_p(1) \\
&\xrightarrow{P} E\{\int_{t_1}^{t_2} w_i(s) \dot{\varphi}\{(\gamma_0(s))^T X_i(s) + \beta_0^T Z_i(s)\}\{-(e_{xz}(s))^T (e_{xx}(s))^{-1} X_i(s) + Z_i(s)\}^{\otimes 2}\, dN_i(s)\} \\
&= A, \qquad (33)
\end{aligned}$$

uniformly in a neighborhood of β.

Now we show that $n^{-1/2} U(\beta_0)$ converges in distribution to a normal distribution. By Taylor expansion,

$$\begin{aligned}
&\varphi\{(\tilde{\gamma}(s, \beta_0))^T X_i(s) + \beta_0^T Z_i(s)\} - \varphi\{(\gamma_0(s))^T X_i(s) + \beta_0^T Z_i(s)\} \\
&= \dot{\mu}_i(s)\{(\tilde{\gamma}(s, \beta_0))^T - (\gamma_0(s))^T\} X_i(s) + O_p(\|\tilde{\gamma}(s, \beta_0) - \gamma_0(s)\|^2).
\end{aligned}$$

By Lemmas 1 and 2,

$$\begin{aligned}
&n^{-1/2} \sum_{i=1}^{n} \int_{t_1}^{t_2} W_i(s)[\varphi\{(\tilde{\gamma}(s, \beta_0))^T X_i(s) + \beta_0^T Z_i(s)\} - \varphi\{(\gamma_0(s))^T X_i(s) + \beta_0^T Z_i(s)\}] \\
&\qquad \times \left\{ (X_i(s))^T \frac{\partial \tilde{\gamma}(s, \beta_0)}{\partial \beta} + (Z_i(s))^T \right\} dN_i(s) \\
&= n^{-1/2} \sum_{i=1}^{n} \int_{t_1}^{t_2} W_i(s)[(\tilde{\gamma}(s, \beta_0))^T - (\gamma_0(s))^T] \\
&\qquad \times \dot{\mu}_i(s) X_i(s) \left\{ (X_i(s))^T \frac{\partial \tilde{\gamma}(s, \beta_0)}{\partial \beta} + (Z_i(s))^T \right\} dN_i(s) + O_p((nh^2)^{-1/2}) \\
&= o_p(1), \quad \text{as } nh^2 \to \infty.
\end{aligned}$$

Hence

$$
\begin{aligned}
&n^{-1/2}U(\beta_0) \qquad (34)\\
&= n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} W_i(s)[Y_i(s)-\mu_i(s)]\left\{\left(\frac{\partial\tilde{\gamma}(s,\beta_0)}{\partial\beta}\right)^T X_i(s)+Z_i(s)\right\} dN_i(s)+o_p(1)\\
&= n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2} w_i(s)[Y_i(s)-\mu_i(s)]\{Z_i(s)-(e_{xz}(s))^T(e_{xx}(s))^{-1}X_i(s)\}\, dN_i(s)+o_p(1),
\end{aligned}
$$

which converges in distribution to $N(0,\Sigma)$, where

$$
\Sigma = E\left(\int_{t_1}^{t_2} w_i(s)[Y_i(s)-\mu_i(s)]\{Z_i(s)-(e_{xz}(s))^T(e_{xx}(s))^{-1}X_i(s)\}\, dN_i(s)\right)^{\otimes 2}. \qquad (35)
$$

Since $n^{1/2}(\hat{\beta}-\beta_0) = -\left(n^{-1}\frac{\partial U(\beta_0)}{\partial\beta}\right)^{-1} n^{-1/2}U(\beta_0)+o_p(1)$, it follows from (33) and (34) that $n^{1/2}(\hat{\beta}-\beta_0) \xrightarrow{\mathcal{D}} N(0, A^{-1}\Sigma A^{-1})$ as $n\to\infty$. □

Proof of Theorem 2.

Since $\hat{\gamma}(t)=\tilde{\gamma}(t,\hat{\beta})$, we have $\hat{\gamma}(t)\xrightarrow{P}\gamma_0(t)$ uniform in $t\in[0,\tau]$ by Theorem 1 and Lemma 1. It also follows that $\partial\tilde{\gamma}(t,\beta^*)/\partial\beta\xrightarrow{P}-(e_{xx}(t))^{-1}e_{xz}(t)$ for β^* on the line segment between $\hat{\beta}$ and β_0. By Lemma 2 and (34),

$$
\begin{aligned}
&(nh)^{1/2}\{\hat{\gamma}(t)-\gamma_0(t)-\frac{1}{2}\mu_2 h^2\ddot{\gamma}_0^T(t)\}\\
&= (nh)^{1/2}\{\tilde{\gamma}(t,\beta_0)-\gamma_0(t)-\frac{1}{2}\mu_2 h^2\ddot{\gamma}_0^T(t)\}-(nh)^{1/2}(e_{xx}(t))^{-1}e_{xz}(t)\,(\hat{\beta}-\beta_0)+o_p(1)\\
&= n^{-1/2}\sum_{i=1}^{n} g_i(t)+o_p(1),
\end{aligned}
$$

where

$$
\begin{aligned}
g_i(t) \;=\; & h^{1/2}(e_{xx}(t))^{-1}\int_0^{\tau} w_i(s)K_h(s-t)X_i(s)\{Y_i(s)-\mu_i(s)\}dN_i(s) - h^{1/2}(e_{xx}(t))^{-1}e_{xz}(t)\\
&\times A^{-1}\int_{t_1}^{t_2} w_i(s)\{Z_i(s)-(e_{xz}(s))^T(e_{xx}(s))^{-1}X_i(s)\}\{Y_i(s)-\mu_i(s)\}\, dN_i(s).
\end{aligned}
$$

Following the arguments of Lemma 2 of Sun (2010),

$$
(nh)^{1/2}(\hat{\gamma}(t)-\gamma_0(t)-\frac{1}{2}\mu_2 h^2\ddot{\gamma}_0^T(t))\xrightarrow{\mathcal{D}} N\left(0,\Sigma_{\gamma}(t)\right), \qquad (36)
$$

as $nh^2 \to \infty$ and $nh^5 = O(1)$. The consistency of the variance estimator for $\Sigma_\gamma(t)$ follows from the proof of Theorem 2 of Sun (2010). $\square$

Proof of Theorem 3.

By (29), (33) and (34), we have

$$\begin{aligned}
G_n(t) &= n^{1/2}\int_{t_1}^t (\tilde\gamma(s;\beta_0)-\gamma_0(s))\,ds + n^{1/2}\int_{t_1}^t (\tilde\gamma(s;\hat\beta)-\tilde\gamma(s;\beta_0))\,ds \\
&= n^{1/2}\int_{t_1}^t (\tilde\gamma(s;\beta_0)-\gamma_0(s))\,ds - \int_{t_1}^t (e_{xx}(s))^{-1}e_{xz}(s)\,ds\, n^{1/2}(\hat\beta-\beta_0) + o_p(1) \\
&= n^{-1/2}\sum_{i=1}^n \Bigg\{ \int_{t_1}^t (e_{xx}(s))^{-1}\int_0^\tau K_h(u-s)w_i(s)X_i(u)\{Y_i(u)-\mu_i(u)\}\,dN_i(u)\,ds \\
&\quad - \int_0^t (e_{xx}(s))^{-1}e_{xz}(s)\,ds\,A^{-1} \\
&\quad \times \int_0^\tau w_i(s)\{Z_i(s)-(e_{xz}(s))^T(e_{xx}(s))^{-1}X_i(s)\}\{Y_i(s)-\mu_i(s)\}\,dN_i(s)\Bigg\} + o_p(1),
\end{aligned}$$

which converges weakly to a zero-mean Gaussian process by Lemma 1 of Sun and Wu (2005). $\square$

Proof of (9).

Note that $A = E[\int_{t_1}^{t_2} w_i(s)\dot\mu_i(s)\ \{Z_i(s)-(e_{xz}(s))^T(e_{xx}(s))^{-1}\ X_i(s)\}^{\otimes 2}\alpha_i(s)\,ds]$. Let

$$\begin{aligned}
D(s) &= A^{-1}w_i(s)\{Z_i(s)-(e_{xz}(s))^T(e_{xx}(s))^{-1}X_i(s)\}^T\sigma_\epsilon(s|X_i,Z_i)\alpha_i^{1/2}(s) \\
&\quad -\Sigma_0^{-1}\{Z_i(s)-(e_{xz}(s))^T(e_{xx}(s))^{-1}X_i(s)\}^T\{\dot\mu_i(s)/\sigma_\epsilon(s|X_i,Z_i)\}\alpha_i^{1/2}(s).
\end{aligned}$$

Then the matrix

$$\begin{aligned}
E\Bigg(\int_{t_1}^{t_2} D(s)D(s)^T\,ds\Bigg) &= A^{-1}\Sigma A^{-1} - A^{-1}A\Sigma_0^{-1} - \Sigma_0^{-1}AA^{-1} + \Sigma_0^{-1}\Sigma_0\Sigma_0^{-1} \\
&= A^{-1}\Sigma A^{-1} - \Sigma_0^{-1}
\end{aligned}$$

is nonnegative definite. $\square$

Acknowledgements

The first author's research was partially supported by NSF grants DMS-0604576 and DMS-0905777, NIH grant 2 R37 AI054165-09 and a fund provided by UNC Charlotte. The second author's research was partly supported by the National Natural Science Foundation of China Grants (No. 10731010, 10971015 and 10721101), the National Basic Research Program of China (973 Program) (No. 2007CB814902) and Key Laboratory of RCSDS, CAS (No.2008DP173182).

References

Aalen, O.O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics*, **6**, 701–726.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and adaptive estimation for semiparametric models*. Springer, New York.

Cheng, S. C. and Wei, L. J. (2000). Inferences for a semiparametric model with panel data. *Biometrika*, **87**, 89–97.

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

Fan, J., Huang, T. and Li, R. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function. *Journal of American Statistical Association*, **102**, 632–641.

Fan, J. and Li, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of American Statistical Association*, **99**, 710–723.

Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L. P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, **85**, 809–822.

Hu, X. J., Sun, J. and Wei, L. J. (2003). Regression parameter estimation from panel counts. *Scandinavian Journal of Statistics*, **30**, 25–43.

Hu, Z., Wang, N. and Carroll, R. J. (2004). Profile-kernel versus backfitting in the partially linear models for longitudinal/clustered data. *Biometrika*, **91**, 251–262.

Lin, D. Y., Wei, L. J. Yang, I. and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society, Series B*, **62**, Part 4, 711–730.

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

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

Lin, H., Song, P. X.-K. and Zhou, Q. M. (2007). Varying-coefficient marginal models and applications in longitudinal data analysis. *Sankhyă: The Indian Journal of Statistics*, **69**, 581–614.

Lin, X. and Carroll, R. J. (2001). Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association*, **96**, 1045–1056.

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

Martinussen, T. and Scheike, T. H. (2000). A nonparametric dynamic additive regression model for longitudinal data. *The Annals of Statistics*, **28**, 1000–1025.

Martinussen, T. and Scheike, T. H. (2001). Sampling adjusted analysis of dynamic additive regression models for longitudinal data. *Scandinavian Journal of Statistics*, **28**, 303–323.

Martinussen, T. and Scheike, T. H. (2006). *D*ynamic Regression Models for Survival Data. Springer, New York.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society,* *Series B*, **53**, 233–243.

Scheike, T. H. (2002). The additive nonparametric and semiparametric Aalen model as the rate function for a counting process. *Lifetime Data Analysis*, **8**, 247–262.

Sun, J. and Wei, L. J. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *Journal of the Royal Statistical Society, Series B*, **62**, 293–302.

Sun, Y. (2010). Estimation of semiparametric regression model with longitudinal data. *Lifetime Data Analysis*, **16**, 271–298.

Sun, Y., Gilbert, P. B. and McKeague, I. W. (2009). Proportional hazards models with continuous marks. *The Annals of Statistics*, **37**, 394–426.

Sun, Y. and Wu, H. (2005). Semiparametric time-varying coefficients regression model for longitudinal data. *Scandinavian Journal of Statistics*, **32**, 21–47.

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

Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

Wu, H. and Liang, H. (2004). Backfitting random varying-coefficient models with time-dependent smoothing covariates. *Scandinavian Journal of Statistics*, **31**, 3–19.

Zhang, Y. (2002). A semiparametric pseudolikelihood estimation method for panel count data. *Biometrika*, **89**, 39–48.

Table 1: Summary of Bias, SEE, ESE and CP for β and RMSE for $\gamma(t)$ under models (16) and (17).

β	n	h	Bias	SEE	ESE	CP	$RMSE_1$	$RMSE_2$
0	100	0.3	0.0002	0.1272	0.1177	93.1	0.2020	0.3124
		0.4	0.0068	0.1220	0.1173	93.8	0.1413	0.1795
		0.5	0.0016	0.1188	0.1173	94.8	0.1254	0.1611
	200	0.3	0.0019	0.0868	0.0830	94.4	0.1002	0.1321
		0.4	0.0015	0.0852	0.0832	94.2	0.0949	0.1227
		0.5	0.0009	0.0875	0.0836	94.3	0.0858	0.1114
	300	0.3	0.0034	0.0672	0.0683	95.1	0.0855	0.1122
		0.4	-0.0029	0.0709	0.0682	94.2	0.0770	0.0995
		0.5	0.0025	0.0692	0.0682	95.1	0.0716	0.0919
0.5	100	0.3	0.0013	0.0944	0.0930	93.5	0.1212	0.1520
		0.4	-0.0007	0.0972	0.0926	93.6	0.1099	0.1333
		0.5	-0.0081	0.0991	0.0924	92.7	0.1009	0.1206
	200	0.3	0.0012	0.0671	0.0657	94.1	0.0753	0.0960
		0.4	0.0020	0.0681	0.0660	94.7	0.0735	0.0907
		0.5	0.0031	0.0680	0.0664	94.4	0.0704	0.0878
	300	0.3	0.0028	0.0562	0.0540	94.4	0.0678	0.0849
		0.4	0.0011	0.0566	0.0539	94.4	0.0608	0.0758
		0.5	-0.0013	0.0561	0.0542	93.2	0.0570	0.0750
1.5	100	0.3	0.0015	0.0662	0.0618	92.8	0.0750	0.0848
		0.4	-0.0017	0.0658	0.0618	93.2	0.0674	0.0733
		0.5	0.0008	0.0668	0.0618	91.9	0.0659	0.0719
	200	0.3	0.0003	0.0442	0.0437	94.8	0.0459	0.0530
		0.4	0.0012	0.0446	0.0439	94.6	0.0461	0.0525
		0.5	-0.0008	0.0432	0.0439	95.2	0.0435	0.0573
	300	0.3	0.0017	0.0351	0.0355	95.5	0.0404	0.0462
		0.4	0.0006	0.0373	0.0356	94.2	0.0383	0.0456
		0.5	-0.0002	0.0374	0.0358	93.9	0.0366	0.0502

Table 2: Summary of Bias, SEE, ESE and CP for $\gamma(t)$ for $\beta = 0.5$ under models (16) and (17).

	$\gamma_1(t) = 0.5\sqrt{t}$				$\gamma_2(t) = 0.5\sin(2t)$			
t	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP
				$n = 100,\ h = 0.5$				
0.5	-0.0161	0.1186	0.1117	0.932	-0.0316	0.1219	0.1157	0.924
1.0	-0.0058	0.1046	0.0989	0.935	-0.0385	0.1016	0.0978	0.910
1.5	-0.0057	0.1022	0.0927	0.919	-0.0009	0.1078	0.0994	0.923
2.0	0.0003	0.0985	0.0889	0.909	0.0342	0.1244	0.1142	0.914
2.5	-0.0031	0.0921	0.0865	0.924	0.0481	0.1291	0.1159	0.881
3.0	-0.0037	0.0936	0.0847	0.913	0.0178	0.1040	0.0936	0.906
				$n = 200,\ h = 0.4$				
0.5	-0.0164	0.0876	0.0862	0.940	-0.0181	0.0940	0.0909	0.929
1.0	-0.0110	0.0792	0.0766	0.946	-0.0207	0.0771	0.0772	0.927
1.5	-0.0079	0.0708	0.0707	0.941	-0.0012	0.0836	0.0780	0.930
2.0	-0.0030	0.0698	0.0674	0.944	0.0221	0.0931	0.0916	0.945
2.5	-0.0040	0.0661	0.0662	0.958	0.0284	0.1004	0.0938	0.911
3.0	-0.0038	0.0661	0.0648	0.944	0.0073	0.0792	0.0754	0.932
				$n = 300,\ h = 0.3$				
0.5	-0.0112	0.0781	0.0782	0.945	-0.0129	0.0854	0.0845	0.944
1.0	-0.0083	0.0726	0.0693	0.937	-0.0115	0.0746	0.0719	0.929
1.5	-0.0063	0.0655	0.0633	0.938	-0.0016	0.0770	0.0733	0.938
2.0	-0.0036	0.0621	0.0612	0.943	0.0089	0.0901	0.0886	0.936
2.5	-0.0050	0.0626	0.0595	0.938	0.0142	0.0968	0.0904	0.929
3.0	-0.0019	0.0600	0.0581	0.935	0.0028	0.0743	0.0714	0.933

Table 3: Summary of Bias, SEE, ESE and CP for β and RMSE for $\gamma(t)$ under models (16) and (18).

β	n	h	Bias	SEE	ESE	CP	$RMSE_1$	$RMSE_2$
0	100	0.3	-0.0027	0.1288	0.1219	93.0	0.1639	0.2179
		0.4	0.0018	0.1289	0.1222	93.2	0.1464	0.1785
		0.5	0.0084	0.1259	0.1217	93.4	0.1299	0.1604
	200	0.3	0.0037	0.0877	0.0861	95.1	0.1003	0.1320
		0.4	0.0053	0.0893	0.0866	94.0	0.0973	0.1217
		0.5	0.0009	0.0881	0.0865	94.4	0.0881	0.1103
	300	0.3	0.0003	0.0738	0.0706	93.3	0.0889	0.1140
		0.4	0.0045	0.0708	0.0707	95.1	0.0790	0.0981
		0.5	-0.0002	0.0738	0.0706	93.7	0.0726	0.0919
0.5	100	0.3	0.0048	0.1074	0.0966	92.1	0.1249	0.1552
		0.4	0.0064	0.1011	0.0964	93.5	0.1113	0.1302
		0.5	-0.0017	0.1051	0.0964	92.5	0.1033	0.1197
	200	0.3	-0.0010	0.0676	0.0681	94.7	0.0754	0.0959
		0.4	0.0032	0.0709	0.0685	93.8	0.0758	0.0896
		0.5	0.0035	0.0693	0.0687	94.8	0.0695	0.0855
	300	0.3	-0.0043	0.0580	0.0557	93.5	0.0690	0.0843
		0.4	-0.0020	0.0555	0.0560	95.9	0.0600	0.0743
		0.5	0.0005	0.0553	0.0561	94.7	0.0569	0.0730
1.5	100	0.3	0.0029	0.0674	0.0632	93.3	0.0746	0.0808
		0.4	0.0031	0.0672	0.0630	94.5	0.0673	0.0704
		0.5	0.0013	0.0653	0.0633	93.9	0.0640	0.0689
	200	0.3	0.0005	0.0447	0.0445	94.8	0.0491	0.0547
		0.4	0.0025	0.0458	0.0447	94.6	0.0473	0.0508
		0.5	0.0010	0.0458	0.0447	93.8	0.0442	0.0529
	300	0.3	0.0008	0.0371	0.0364	94.0	0.0384	0.0427
		0.4	0.0007	0.0382	0.0364	95.3	0.0387	0.0439
		0.5	0.0004	0.0376	0.0365	93.8	0.0361	0.0490

Table 4: Summary of Bias, SEE, ESE and CP for $\gamma(t)$ for $\beta = 0.5$ under models (16) and (18).

	$\gamma_1(t) = 0.5\sqrt{t}$				$\gamma_2(t) = 0.5\sin(2t)$			
t	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP
				$n = 100$, $h = 0.5$				
0.5	-0.0135	0.1233	0.1125	0.936	-0.0376	0.1227	0.1136	0.924
1.0	-0.0090	0.1065	0.1017	0.940	-0.0404	0.1016	0.0969	0.894
1.5	-0.0067	0.1021	0.0959	0.925	-0.0018	0.1042	0.0981	0.931
2.0	-0.0050	0.0996	0.0918	0.930	0.0369	0.1214	0.1125	0.901
2.5	-0.0072	0.0984	0.0897	0.924	0.0483	0.1288	0.1147	0.885
3.0	-0.0051	0.0933	0.0876	0.926	0.0127	0.0989	0.0919	0.924
				$n = 200$, $h = 0.4$				
0.5	-0.0148	0.0893	0.0869	0.949	-0.0213	0.0929	0.0892	0.930
1.0	-0.0057	0.0804	0.0776	0.932	-0.0283	0.0776	0.0758	0.921
1.5	-0.0050	0.0745	0.0724	0.933	-0.0042	0.0806	0.0774	0.942
2.0	-0.0026	0.0711	0.0689	0.936	0.0187	0.0963	0.0904	0.925
2.5	-0.0035	0.0699	0.0675	0.937	0.0230	0.0941	0.0925	0.928
3.0	-0.0046	0.0699	0.0668	0.931	0.0070	0.0808	0.0748	0.923
				$n = 300$, $h = 0.3$				
0.5	-0.0049	0.0808	0.0781	0.935	-0.0145	0.0883	0.0829	0.927
1.0	-0.0019	0.0727	0.0698	0.937	-0.0136	0.0717	0.0709	0.943
1.5	0.0011	0.0682	0.0642	0.935	-0.0069	0.0768	0.0724	0.935
2.0	0.0038	0.0620	0.0618	0.949	0.0049	0.0898	0.0870	0.931
2.5	0.0021	0.0652	0.0600	0.928	0.0106	0.0972	0.0896	0.918
3.0	-0.0014	0.0629	0.0594	0.924	0.0092	0.0742	0.0709	0.938

Table 5: Summary of Bias, SEE, ESE and CP for β and RMSE for $\gamma(t)$ under models (16) and (19).

β	n	h	Bias	SEE	ESE	CP	$RMSE_1$	$RMSE_2$
0	100	0.3	-0.0009	0.1403	0.1364	0.924	0.3850	0.4263
		0.4	0.0055	0.1323	0.1245	0.933	0.2241	0.2371
		0.5	0.0004	0.1379	0.1255	0.932	0.1438	0.1548
	200	0.3	0.0028	0.0891	0.0887	0.941	0.1117	0.1335
		0.4	-0.0004	0.0926	0.0894	0.942	0.1035	0.1196
		0.5	-0.0018	0.0923	0.0895	0.933	0.0955	0.1074
	300	0.3	-0.0003	0.0757	0.0730	0.944	0.0906	0.1056
		0.4	0.0033	0.0728	0.0731	0.951	0.0827	0.0959
		0.5	-0.0025	0.0709	0.0727	0.944	0.0763	0.0884
0.5	100	0.3	-0.0009	0.1112	0.0987	0.920	0.2054	0.2218
		0.4	0.0036	0.1066	0.0990	0.925	0.1195	0.1244
		0.5	0.0019	0.1020	0.0997	0.933	0.1072	0.1122
	200	0.3	0.0004	0.0776	0.0703	0.923	0.0850	0.0914
		0.4	-0.0015	0.0730	0.0703	0.944	0.0803	0.0856
		0.5	-0.0025	0.0711	0.0705	0.948	0.0730	0.0799
	300	0.3	0.0013	0.0595	0.0575	0.945	0.0679	0.0751
		0.4	0.0024	0.0603	0.0581	0.935	0.0668	0.0703
		0.5	-0.0007	0.0589	0.0577	0.943	0.0605	0.0702
1.5	100	0.3	-0.0006	0.0671	0.0638	0.938	0.0737	0.0822
		0.4	-0.0010	0.0684	0.0641	0.930	0.0699	0.0627
		0.5	-0.0035	0.0650	0.0638	0.934	0.0634	0.0605
	200	0.3	0.0020	0.0457	0.0449	0.941	0.0477	0.0450
		0.4	-0.0007	0.0469	0.0452	0.939	0.0477	0.0456
		0.5	0.0010	0.0478	0.0453	0.933	0.0461	0.0490
	300	0.3	-0.0001	0.0368	0.0367	0.954	0.0403	0.0396
		0.4	-0.0006	0.0386	0.0369	0.934	0.0392	0.0396
		0.5	-0.0005	0.0369	0.0368	0.951	0.0363	0.0446

Table 6: Summary of Bias, SEE, ESE and CP for $\gamma(t)$ for $\beta = 0.5$ under models (16) and (19).

	$\gamma_1(t) = 0.5\sqrt{t}$				$\gamma_2(t) = 0.5\sin(2t)$			
t	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP
			$n = 100,\ h = 0.5$					
0.5	-0.0235	0.1462	0.1237	0.905	-0.0227	0.1487	0.1217	0.873
1.0	-0.0106	0.1152	0.1076	0.938	-0.0374	0.1033	0.0958	0.912
1.5	-0.0062	0.1039	0.1004	0.942	-0.0049	0.1021	0.0927	0.921
2.0	-0.0052	0.0969	0.0955	0.935	0.0352	0.1096	0.0998	0.895
2.5	-0.0033	0.0971	0.0917	0.919	0.0424	0.1083	0.0946	0.868
3.0	-0.0050	0.0932	0.0895	0.926	0.0114	0.0829	0.0723	0.907
			$n = 200,\ h = 0.4$					
0.5	-0.0097	0.1029	0.0947	0.927	-0.0193	0.1075	0.0961	0.921
1.0	-0.0078	0.0867	0.0808	0.920	-0.0225	0.0780	0.0749	0.927
1.5	-0.0033	0.0774	0.0754	0.936	-0.0028	0.0793	0.0739	0.933
2.0	-0.0023	0.0749	0.0706	0.924	0.0195	0.0864	0.0812	0.922
2.5	-0.0013	0.0715	0.0673	0.933	0.0318	0.0819	0.0765	0.895
3.0	-0.0001	0.0666	0.0649	0.941	0.0074	0.0606	0.0575	0.936
			$n = 300,\ h = 0.3$					
0.5	-0.0111	0.0942	0.0865	0.923	-0.0073	0.0972	0.0909	0.930
1.0	-0.0047	0.0736	0.0721	0.940	-0.0166	0.0739	0.0699	0.927
1.5	-0.0044	0.0702	0.0665	0.934	-0.0020	0.0711	0.0694	0.950
2.0	-0.0029	0.0646	0.0617	0.941	0.0114	0.0802	0.0772	0.937
2.5	-0.0041	0.0614	0.0587	0.948	0.0180	0.0790	0.0735	0.919
3.0	-0.0025	0.0577	0.0561	0.942	0.0033	0.0572	0.0547	0.934

Table 7: Empirical sizes and powers of the proposed tests at nominal level $\alpha = 0.05$ for model (16) under models (17), (18) and (19) for sampling times.

		size		power					
		$\theta = 0$		$\theta = 0.1$		$\theta = 0.15$		$\theta = 0.2$	
n	h	S	L	S	L	S	L	S	L
Under sampling times model (17)									
200	0.3	0.053	0.054	0.503	0.526	0.851	0.863	0.968	0.970
	0.4	0.055	0.057	0.542	0.566	0.851	0.859	0.980	0.981
	0.5	0.047	0.045	0.542	0.536	0.853	0.864	0.976	0.976
300	0.3	0.054	0.054	0.687	0.684	0.954	0.960	0.997	0.997
	0.4	0.060	0.054	0.699	0.702	0.950	0.952	0.999	0.998
	0.5	0.059	0.054	0.696	0.689	0.955	0.962	0.996	0.996
Under sampling times model (18)									
200	0.3	0.057	0.056	0.535	0.554	0.986	0.987	0.979	0.984
	0.4	0.053	0.049	0.529	0.542	0.861	0.862	0.975	0.977
	0.5	0.066	0.058	0.565	0.558	0.872	0.867	0.974	0.973
300	0.3	0.064	0.055	0.695	0.703	0.952	0.959	0.997	0.997
	0.4	0.066	0.069	0.717	0.718	0.956	0.956	1.000	1.000
	0.5	0.052	0.045	0.710	0.717	0.958	0.965	0.999	0.999
Under sampling times model (19)									
200	0.3	0.060	0.065	0.572	0.599	0.887	0.895	0.990	0.986
	0.4	0.065	0.066	0.584	0.578	0.886	0.885	0.986	0.989
	0.5	0.058	0.059	0.604	0.598	0.880	0.886	0.988	0.989
300	0.3	0.058	0.069	0.738	0.755	0.960	0.965	0.998	0.998
	0.4	0.065	0.066	0.760	0.768	0.967	0.968	0.999	0.998
	0.5	0.044	0.049	0.738	0.741	0.965	0.969	1.000	0.999

Table 8: Comparisons of the estimation for $\beta = 1$ using the proposed method and the method of Lin and Ying (2001) under model (21) with $n = 200$.

	$\alpha(t) = 1 + t$				$\alpha(t) = 1 + t^3$			
h	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP
			Under sampling times model (17)					
0.3	0.0008	0.1665	0.1648	0.952	0.0065	0.1692	0.1675	0.944
0.4	0.0013	0.1660	0.1649	0.947	0.0072	0.1693	0.1714	0.951
0.5	0.0013	0.1659	0.1649	0.944	0.0076	0.1693	0.1786	0.962
L&Y	0.0024	0.1772	0.1786	0.947	-0.0390	0.9055	0.8756	0.943
			Under sampling times model (18)					
0.3	0.0052	0.1724	0.1728	0.954	-0.0019	0.1801	0.1756	0.942
0.4	0.0053	0.1720	0.1728	0.952	-0.0022	0.1793	0.1794	0.951
0.5	0.0051	0.1717	0.1728	0.953	-0.0021	0.1788	0.1867	0.958
L&Y	0.0033	0.1802	0.1879	0.952	0.0182	0.9088	0.9081	0.949
			Under sampling times model (19)					
0.3	0.0062	0.1779	0.1765	0.948	0.0068	0.1801	0.1796	0.938
0.4	0.0061	0.1775	0.1764	0.950	0.0068	0.1797	0.1863	0.949
0.5	0.0059	0.1770	0.1764	0.951	0.0063	0.1797	0.1985	0.964
L&Y	0.0905	0.1885	0.2026	0.931	0.7316	0.9190	0.9414	0.871

(a) (b)

(c) (d)

(e) (f)

Figure 1: Plots of $\hat{\gamma}(t)$ for model (16) when $\gamma_1(t) = 0.5t^{1/2}$, $\gamma_2(t) = 0.5\sin(2t)$ and $\beta = 0.5$ for $n = 100$ and $h = 0.3$. The figures (a) and (b) in the first row are under the proportional sampling model (17), the figures (c) and (d) in the second row are under the additive sampling model (18) and the figures (e) and (f) in the third row are under the sampling model (19). The solid lines are the estimates and the doted lines are the true curves.

(a)

(b)

Figure 2: The curve of the total prediction error PE(h) is plotted against h in (a) and the change of $\hat{\beta}$ with h is shown in (b) for the HIV-1 RNA data.

(a) $\hat{\gamma}_1(t)$ for $h = 0.05$ (b) $\hat{\gamma}_2(t)$ for $h = 0.05$

(c) $\hat{\gamma}_1(t)$ for $h = 0.11$ (d) $\hat{\gamma}_2(t)$ for $h = 0.11$

(e) $\hat{\gamma}_1(t)$ for $h = 0.2$ (f) $\hat{\gamma}_2(t)$ for $h = 0.2$

Figure 3: The plots of the estimators $\hat{\gamma}_1(t)$ and $\hat{\gamma}_2(t)$ for three different bandwidths $h = 0.05, 0.11$ and 0.2 for the HIV-1 RNA data.