Published in final edited form as:

Lifetime Data Anal. 2013 July ; 19(3): 317–349. doi:10.1007/s10985-013-9251-y.

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, USA

Liuquan Sun, and

Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Beijing, China

Jie Zhou

Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Beijing, China

Yanqing Sun: yasun@uncc.edu; Liuquan Sun: slq@amt.ac.cn; Jie Zhou: 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. A K -fold cross-validation bandwidth selection is proposed as a working tool for locating an appropriate bandwidth. A criteria for selecting the link function is proposed to provide better fit of the data. Large sample properties of the proposed estimators are investigated. Large sample pointwise and simultaneous confidence intervals for the regression coefficients are constructed. Formal hypothesis testing procedures are proposed to check for the covariate effects and whether the effects are time-varying. A simulation study is conducted to examine the finite sample performances of the proposed estimation and hypothesis testing procedures. The methods are illustrated with a data example.

Keywords

Asymptotics; Kernel smoothing; Link function; Sampling adjusted estimation; Testing time-varying effects; Weighted least squares

1 Introduction

We study 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 i th 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 $Y_i(t)$, $0 \leq t \leq \tau$,

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

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. 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. We refer to Hoover et al. (1998), Martinussen and Scheike (1999, 2000, 2001), Lin and Ying (2001), Wu and Liang (2004), Fan and Li (2004), Hu et al. (2004), Sun and Wu (2005) and Fan et al. (2007), among others. 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. For panel count data, the proportional means model has been studied by Sun and Wei (2000), Cheng and Wei (2000), Zhang (2002) and Hu et al. (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 et al. (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 et al. 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 et al. (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.

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. This presents an opportunity for model selection of the link function. A criteria for selecting the link function is proposed to provide a better fit of the 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 Sect. 2, a sampling adjusted profile-based local linear estimation method is proposed for model (1). Large sample properties are investigated in Sect. 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. The procedures for selecting bandwidth and the link function are proposed in Sect. 3.4. A simulation study is conducted in Sect. 4 to examine the finite sample performances of the proposed statistical procedures. The proposed methods are illustrated with the analysis of a HIV-1 RNA data set from an AIDS clinical trial in Sect. 5. Some concluding remarks are made in Sect. 6. All proofs are given in the Appendix.

2 Profile local linear estimation approach

2.1 Preliminaries

Suppose that the observations of the response process $Y_i(t)$ for the i th subject are taken at the sampling time points $0 < t_{i1} < t_{i2} < \dots < t_{in_i} \leq \tau$, where n_i is the total number of observations on the i th 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 $N_i(t) = \sum_{j=1}^{n_i} I(t_{ij} \leq t)$ be the number of observations taken on the i th 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 i th 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 i th subject. We assume that $\{(Y_i(\cdot), X_i(\cdot), Z_i(\cdot), N_i(\cdot))\}$, $i = 1, \dots, n$, are independent identically distributed random processes. In this section, we propose an estimation procedure for model (1) based on the observations $\{(Y_i(t_{ij}), X_i(t_{ij}), Z_i(t_{ij})); j = 1, \dots, n_i, i = 1, \dots, n\}$. These are the values of $\{(Y_i(t), X_i(t), Z_i(t)), 0 \leq t \leq \tau\}$ observed at sampling times or the jump time points of $N_i(t) = N_i^*(t \wedge C_i)$, $i = 1, \dots, n$.

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 $1 \leq i \leq n$. Let $\lambda_i(t)$ be the intensity process defined as follows

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

for $0 \leq t \leq \tau$. 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 Sect. 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 $\lambda_j(t)$ or $\alpha_j(t)$.

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 $\tilde{X}_i(s, s-t) = (X_i^T(s), (s-t)X_i^T(s))^T$. Let

$\tilde{\mu}_a(s, \gamma_a, \beta | X_i, Z_i) = \phi\{\gamma_a^T(t)\tilde{X}_i(s, s-t) + \beta^T Z_i(s)\}$, where $\phi(x) = g^{-1}(x)$ 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 $\gamma_a(t)$:

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

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 $\dot{\phi}(x)$ be the derivative of $\phi(x) = g^{-1}(x)$ with respect to x . The estimating function $U_a(\gamma_a, \beta)$ can be obtained by setting $Q_i(s) = W_i(s) [\dot{\phi}\{\gamma_a^T(t)\tilde{X}_i(s, s-t) + \beta^T Z_i(s)\}]^{-1}$ 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 $h \rightarrow 0$ 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

$$\tilde{\gamma}_a(t, \beta) = \tilde{Y}_x^T(t) - \tilde{Z}_x^T(t) \beta \text{ where } \tilde{Y}_x(t) = \tilde{E}_{yx}(t) (\tilde{E}_{xx}(t))^{-1} \text{ and } \tilde{Z}_x(t) = \tilde{E}_{zx}(t) (\tilde{E}_{xx}(t))^{-1}.$$

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

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

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 $\hat{\gamma}(t)$ by $\hat{\gamma}(t) = \tilde{\gamma}(t, \hat{\beta})$.

The expression for the derivative $\frac{\partial \tilde{\gamma}_a(t, \beta)}{\partial \beta}$ 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

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

It follows that

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

where

$$-\frac{\partial U_a(\gamma_a, \beta)}{\partial \gamma_a} = \sum_{i=1}^n \int_0^\tau W_i(s) \dot{\phi} \{ \gamma_a^T \tilde{X}_i(s, s-t) + \beta^T Z_i(s) \} \times \{ \tilde{X}_i(s, s-t) \}^{\otimes 2} K_h(s-t) dN_i(s), \quad (6)$$

$$-\frac{\partial U_a(\gamma_a, \beta)}{\partial \beta} = \sum_{i=1}^n \int_0^\tau W_i(s) \dot{\phi} \{ \gamma_a^T \tilde{X}_i(s, s-t) + \beta^T Z_i(s) \} \times \tilde{X}_i(s, s-t) \{ Z_i(s) \}^T K_h(s-t) dN_i(s), \quad (7)$$

The estimator $\hat{\beta}$ is a weighted least square estimator since the estimating function $U(\beta)$ can be obtained by setting $Q_i(t) = W_i(t) [\dot{\phi} \{ (\tilde{\gamma}(t, \beta))^T X_i(t) + \beta^T Z_i(t) \}]^{-1}$ in the derivative of the profile least squares function $\ell(\beta)$ with respect to β , where

$$\ell(\beta) = \sum_{i=1}^n \int_{t_1}^{t_2} Q_i(s) [Y_i(s) - \phi\{\tilde{\gamma}(s, \beta)^T X_i(s) + \beta^T Z_i(s)\}]^2 dN_i(s).$$

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 m th 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 m th step estimator $\hat{\beta}^{(m)}$ is obtained by solving the estimating function for β :

$$U_m(\beta) = \sum_{i=1}^n \int_{t_1}^{t_2} W_i(s) [Y_i(s) - \phi\{\tilde{\gamma}(s, \hat{\beta}^{(m-1)})^T X_i(s) + \beta^T Z_i(s)\}] \times \left\{ \frac{\partial \tilde{\gamma}(s, \hat{\beta}^{(m-1)})}{\partial \beta} X_i(s) + Z_i(s) \right\} dN_i(s), \quad (8)$$

where $\frac{\partial \tilde{\gamma}(s, \hat{\beta}^{(m-1)})}{\partial \beta}$ 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 $\hat{\gamma}_a(t) = \tilde{\gamma}_a(t, \hat{\beta})$. The estimation of β requires that both $\hat{\gamma}_a^{(m)}(t)$ and $\frac{\partial \tilde{\gamma}(s, \hat{\beta}^{(m-1)})}{\partial \beta}$ be evaluated at the combined sampling points of all subjects or the jump points of $\{N_i(\cdot), i = 1, \dots, n\}$. The estimate $\hat{\gamma}(t)$ at the last iteration can be obtained by solving $U_a(\hat{\gamma}_a^{(m)}(t), \hat{\beta}^{(m-1)}) = 0$ at the grid points fine enough such that their plots look reasonably smooth.

2.4 Estimation under the fixed designs

Model (2) assumes existence of intensity for the counting processes that record the sampling time points. This formulation excludes sampling at predetermined time points, i.e., the fixed design. However, the method developed in Sect. 2.2 can be extended to the fixed designs

with some modifications. Let $t_1, \dots, t_k$ be the fixed sampling time points at which the responses and covariates may be observed. For the fixed designs, estimation of model (1) does not involve the kernel neighborhood smoothing. In particular, for the fixed designs, the counting process is $N_i(t) = \sum_{j=1}^k I(t_j \leq t \wedge C_i)$, where C_i is the censoring time for subject i . The equation (3) should be replaced by

$$U_a(\gamma(t), \beta) = \sum_{i=1}^n \{Y_i(t) - \phi(\gamma^T(t)X_i(t) + \beta^T Z_i(t))\} X_i(t) I(C_i \geq t). \quad (9)$$

Let $\tilde{\gamma}(t, \beta)$ solve the equation $U_a(\gamma(t), \beta)$ at the fixed time points $t = t_1, \dots, t_k$ for each fixed β . The estimator $\hat{\beta}$ solves $U(\hat{\beta}) = 0$ where $U(\beta)$ is the profile estimating equation under the fixed design having the same expression as (4). The regression coefficient function $\gamma(t)$ is estimated by $\hat{\gamma}(t) = \tilde{\gamma}(t, \hat{\beta})$.

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

$\mu_i(t) = \phi\{\gamma_0^T(t)X_i(t) + \beta_0^T Z_i(t)\}$ and $\dot{\mu}_i(t) = \dot{\phi}\{\gamma_0^T(t)X_i(t) + \beta_0^T Z_i(t)\}$. Let $w(t, x, z)$ be the deterministic limit of $W(t, x, z)$ in probability as $n \rightarrow \infty$. Define $e_{xx}(t) = E[w(t, X_i(t))\dot{\mu}_i(t)\{X_i(t)\}^{\otimes 2}\alpha_i(t)\xi_i(t)]$ and $e_{xz}(t) = E[w(t, X_i(t))\dot{\mu}_i(t)X_i(t)\{Z_i(t)\}^T\alpha_i(t)\xi_i(t)]$, 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 $\Sigma = E[\int_{t_1}^{t_2} w_i(s)\{Y_i(s) - \mu_i(s)\}\{Z_i(s) - (e_{xz}(s))^T(e_{xx}(s))^{-1}X_i(s)\}^{\otimes 2} dN_i(s)]$, where $w_i(t) = w(t, X_i(t), Z_i(t))$.

Let $\hat{\mu}_i(s) = \phi\{\hat{\gamma}^T(s)X_i(s) + \hat{\beta}^T Z_i(s)\}$ and $\hat{\dot{\mu}}_i(s) = \dot{\phi}\{\hat{\gamma}^T(s)X_i(s) + \hat{\beta}^T Z_i(s)\}$. Let

$\widehat{E}_{xx}(t) = n^{-1} \sum_{i=1}^n \int_0^t W_i(s)K_h(s-t)\hat{\dot{\mu}}_i(s)(X_i(s))^{\otimes 2} dN_i(s)$ and

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

Theorem 1—Assume that Condition A holds. Then

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

The matrix A can be consistently estimated by

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

and Σ can be consistently estimated by

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

Under Theorem 1, the proposed estimator $\widehat{\beta}$ 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 $\widehat{\beta}$. Naturally, we would like to choose the optimal weight such that the asymptotic variance of $\widehat{\beta}$ 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 $\sigma_\varepsilon^2(t|X_i, Z_i) = \text{Var}\{Y_i(t)|X_i(t), Z_i(t)\}$ be the conditional variance of $Y_i(t)$ given the covariates $X_i(t)$ and $Z_i(t)$ under model (1). Then the matrix

$$\begin{aligned} \Sigma &= E \left[\int_{t_1}^{t_2} w_i^2(s) \sigma_\varepsilon^2(s|X_i, Z_i) \{Z_i(s) - (e_{xz}(s))^T (e_{xx}(s))^{-1} X_i(s)\}^{\otimes 2} \alpha_i(s) \xi_i(s) ds \right]. \text{ Let} \\ \Sigma_0 &= E \left[\int_{t_1}^{t_2} \{\dot{\mu}_i(s)/\sigma_\varepsilon(s|X_i, Z_i)\}^2 \{Z_i(s) - (e_{xz}(s))^T (e_{xx}(s))^{-1} X_i(s)\}^{\otimes 2} \alpha_i(s) \xi_i(s) ds \right]. \end{aligned}$$

We show in the Appendix that

$$A^{-1} \Sigma A^{-1} - \Sigma_0^{-1} \geq 0, \quad (10)$$

where $B \geq 0$ means that the matrix B is nonnegative definite. When $w_i(t) = \dot{\mu}_i(t)/\{\sigma_\varepsilon(t|X_i, Z_i)\}^2$, $A = \Sigma = \Sigma_0$ and the equality in (10) 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 $\widehat{\gamma}(t)$. 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, $\widehat{\gamma}(t) \xrightarrow{P} \gamma_0(t)$,

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

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

$$\begin{aligned} \widehat{g}_i(t) &= h^{1/2} (\widehat{E}_{xx}(t))^{-1} \int_0^\tau W_i(s) K_h(s-t) X_i(s) \{Y_i(s) - \widehat{\mu}_i(s)\} dN_i(s) - h^{1/2} (\widehat{E}_{xx}(t))^{-1} \widehat{E}_{xz}(t) \\ &\quad \times \widehat{A}^{-1} \int_{t_1}^{t_2} W_i(s) \{Z_i(s) - (\widehat{E}_{xz}(s))^T (\widehat{E}_{xx}(s))^{-1} X_i(s)\} \{Y_i(s) - \widehat{\mu}_i(s)\} dN_i(s). \end{aligned}$$

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 $\widehat{\Gamma}(t) = \int_{t_1}^t \widehat{\gamma}(s) ds$. The following theorem presents a weak convergence result for $G_n(t) = n^{1/2}(\widehat{\Gamma}(t) - \Gamma_0(t))$ 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 $nh^2 \rightarrow \infty$ and $nh^5 \rightarrow 0$, where

$$\begin{aligned} H_i(t) = & \int_{t_1}^t (e_{xx}(s))^{-1} \int_0^\tau w_i(u) K_h(u-s) X_i(u) \{Y_i(u) - \mu_i(u)\} dN_i(u) ds \\ & - \int_{t_1}^t (e_{xx}(s))^{-1} \dot{e}_{xx}(s) ds A^{-1} \\ & \times \int_{t_1}^{t_2} w_i(s) \{Z_i(s) - (e_{xz}(s))^T (e_{xx}(s))^{-1} X_i(s)\} \{Y_i(s) - \mu_i(s)\} dN_i(s). \end{aligned} \quad (11)$$

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

$\widehat{\Sigma}_G(t) = n^{-1} \sum_{i=1}^n \{\widehat{H}_i(t)\}^{\otimes 2}$, where

$$\begin{aligned} \widehat{H}_i(t) = & \int_{t_1}^t (\widehat{E}_{xx}(s))^{-1} \int_0^\tau W_i(u) K_h(u-s) X_i(u) \{Y_i(u) - \widehat{\mu}_i(u)\} dN_i(u) ds \\ & - \int_{t_1}^t (\widehat{E}_{xx}(s))^{-1} \widehat{E}_{xz}(s) ds \widehat{A}^{-1} \\ & \times \int_{t_1}^{t_2} W_i(s) \{Z_i(s) - (\widehat{E}_{xz}(s))^T (\widehat{E}_{xx}(s))^{-1} X_i(s)\} \{Y_i(s) - \widehat{\mu}_i(s)\} dN_i(s). \end{aligned} \quad (12)$$

Remark—For the estimation of model (1) under the fixed designs, the asymptotic results similar to those in Theorem 1 can be established. Without the kernel neighborhood smoothing, one needs to replace, $e_{xz}(t) = E[w_i(t) \dot{\mu}_i(t) X_i(t) \{Z_i(t)\}^T a_i(t) \xi_i(t)]$ by $e_{xz}(t) = E[\dot{\mu}_i(t) X_i(t) \{Z_i(t)\}^T \xi_i(t)]$, and $\widehat{E}_{xz}(t) = n^{-1} \sum_{i=1}^n \int_0^\tau W_i(s) K_h(s-t) \widehat{\mu}_i(s) X_i(s) \{Z_i(s)\}^T dN_i(s)$ by

$\widehat{E}_{xz}(t) = n^{-1} \sum_{i=1}^n \widehat{\mu}_i(t) X_i(t) (Z_i(t))^T \xi_i(t)$. Similar replacements hold for $e_{xx}(t)$, $\widehat{E}_{xx}(t)$, $e_{yx}(t)$ and $\widehat{E}_{yx}(t)$. The following asymptotic results can be established for $\gamma(t)$ at $t = t_1, \dots, t_k$: $\sqrt{n}(\widehat{\gamma}(t) - \gamma_0(t)) \xrightarrow{\mathcal{D}} N\left(0, \sum_{\gamma}(t)\right)$, where $\Sigma_{\gamma}(t) = (e_{xx}(t))^{-1} \Sigma_e(t) (e_{xx}(t))^{-1}$, $\Sigma_e(t) = E\{(Y_i(t) - \mu_i(t))X_i(t)\xi_i(t)\}^{\otimes 2}$.

3.2 Confidence intervals and simultaneous confidence bands

Let $\gamma^{(k)}(t)$ be the k th component of $\gamma(t)$. Similar notations are used throughout with the superscript (k) denoting the k th component of the corresponding vector. Assuming $nh^5 \rightarrow 0$, 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

$$\widehat{\gamma}^{(k)}(t) \pm (nh)^{-1/2} z_{\alpha/2} \left[n^{-1} \sum_{i=1}^n \{\widehat{g}_i^{(k)}(t)\}^2 \right]^{1/2}. \quad (13)$$

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

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

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 et al. (1993) has been widely employed for this purpose and is described in the following.

Let $G_n^*(t) = n^{-1/2} \sum_{i=1}^n \widehat{H}_i(t) \xi_i$, where $\xi_1, \xi_2, \dots, \xi_n$ 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 $(\xi_1, \dots, \xi_n)$ while fixing $\{Y_i(t), X_i(t), Z_i(t), N_i(t), t = 0\}$ at their observed

values. Let c_{α} be the $(1 - \alpha)$ -quantile of $\sup_{t_1 \leq t \leq t_2} |G_n^{*(k)}(t)| / \left[\sum_{i=1}^n \{\widehat{H}_i^{(k)}(t)\}^2 / n \right]^{1/2}$, which can be approximated by repeatedly generating independent normal samples $(\xi_1, \dots, \xi_n)$. An asymptotic $1 - \alpha$ simultaneous confidence bands for $\Gamma^{(k)}(t)$ on $[t_1, t_2]$ is given by

$$\widehat{\Gamma}^{(k)}(t) \pm n^{-1/2} c_{\alpha} \left[n^{-1} \sum_{i=1}^n \{\widehat{H}_i^{(k)}(t)\}^2 \right]^{1/2}. \quad (15)$$

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_{01} that $\gamma^{(k)}(t)$ is constant for $0 \leq t \leq \tau$.

Under H_{01} , $\Gamma^{(k)}(t) - \frac{t-t_1}{t_2-t_1} \Gamma^{(k)}(t_2) = 0$ for $t \in [t_1, t_2]$. By Theorem 3 and the continuous mapping theorem,

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

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

$$S_1 = \sup_{t_1 \leq t \leq t_2} n^{1/2} \left| \widehat{\Gamma}^{(k)}(t) - \frac{t-t_1}{t_2-t_1} \widehat{\Gamma}^{(k)}(t_2) \right|$$

and

$$L_1 = \int_{t_1}^{t_2} n \left\{ \widehat{\Gamma}^{(k)}(t) - \frac{t-t_1}{t_2-t_1} \widehat{\Gamma}^{(k)}(t_2) \right\}^2 dt.$$

By the continuous mapping theorem, under H_{01} , the test statistic S_1 converges in distribution to $\sup_{t_1 \leq t \leq t_2} \left| G^{(k)}(t) - \frac{t-t_1}{t_2-t_1} G^{(k)}(t_2) \right|$, and the test statistic S_1 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_1 referred as the supremum type and L_1 as the integrated square type, cf., Martinussen and Scheike (2006).

Let

$$S_1^* = \sup_{t_1 \leq t \leq t_2} n^{1/2} \left| G_n^{*(k)}(t) - \frac{t-t_1}{t_2-t_1} G_n^{*(k)}(t_2) \right|$$

and

$$L_1^* = \int_{t_1}^{t_2} n \left\{ G_n^{*(k)}(t) - \frac{t-t_1}{t_2-t_1} G_n^{*(k)}(t_2) \right\}^2 dt.$$

The critical values of S_1 and L_1 can be approximated by simulating a number of copies of S_1^* and L_1^* obtained by repeatedly generating independent normal samples $(\xi_1, \dots, \xi_n)$ while holding the observed data fixed. For example, the critical values of test statistics S_1 and L_1 at the significance level α can be estimated by the upper α quantile of, say 1,000, copies of S_1^* and L_1^* , respectively. The p -values of the tests based on S_1 and L_1 are the percentages of

S_1^* and L_1^* exceeding S_1 and L_1 , respectively. The null hypothesis is rejected if the p -values are less than α .

The tests of the null hypothesis H_{02} that $\gamma^{(k)}(t) = 0$ for $0 \leq t \leq \tau$ can also be constructed similarly. In particular, one may consider the test statistics: $S_2 = \sup_{t_1 \leq t \leq t_2} n^{1/2} |\hat{\Gamma}^{(k)}(t)|$ and $L_2 = \int_{t_1}^{t_2} n \{\hat{\Gamma}^{(k)}(t)\}^2 dt$. The reference distributions of S_2 and L_2 can be generated based on $S_2^* = \sup_{t_1 \leq t \leq t_2} n^{1/2} |G_n^{*(k)}(t)|$ and $L_2^* = \int_{t_1}^{t_2} n \{G_n^{*(k)}(t)\}^2 dt$, respectively.

3.4 Selections of bandwidth and link function

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 k th 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

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

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 et al. (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 k th subgroup of data, then the k th prediction error is given by

$$PE_k(h) = \sum_{i \in D_k} \int_{t_1}^{t_2} \left[Y_i(t) - \phi\{\hat{\gamma}_{(-k)}(t)\}^T X_i(t) + \hat{\beta}_{(-k)}^T Z_i(t) \right]^2 dN_i(t), \quad (16)$$

for $k = 1, \dots, K$, 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 Sect. 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.

Our estimation procedure for model (1) holds for a wide class of link functions. This presents an opportunity to select the most appropriate link function for a particular application. In some applications the choice may be based on prior knowledge, but more

often it will be a pragmatic choice based on what gives the “best fit”. One natural criterion for accessing the model fit is the regression deviation defined as

$$RD(g(\cdot), h_{cv}) = \sum_{i=1}^n \int_{t_1}^{t_2} [Y_i(t) - g^{-1}\{\widehat{\gamma}_g(t)^T X_i(t) + \widehat{\beta}_g^T Z_i(t)\}]^2 dN_i(t), \quad (17)$$

where h_{cv} is the bandwidth selected based on the K -fold cross-validation method for the given link function $g(\cdot)$ described above, and $\widehat{\gamma}_g(t)$ and $\widehat{\beta}_g$ are the estimators of $\gamma_0(t)$ and β_0 under model (1) with the bandwidth h_{cv} . In practice, the link function $g(\cdot)$ can be selected to minimize the regression deviation. This approach is illustrated through a data example in Sect. 5.

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 $\widehat{\gamma}^{(k)}(t)$ on the interval $[h, \tau - h]$, we consider the square root of integrated mean square error

$RMSE_k = \left\{ \frac{1}{N(\tau-2h)} \sum_{j=1}^N \int_h^{\tau-h} (\widehat{\gamma}_j^{(k)}(t) - \gamma_0^{(k)}(t))^2 dt \right\}^{1/2}$, where N is the repetition number, $\widehat{\gamma}_j^{(k)}(t)$ is the j th estimate of $\gamma^{(k)}(t)$ for $j = 1, \dots, N$. We use the unit weight function $W_j(t) = 1$ 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 $g(x) = \ln(x)$:

$$Y_i(t) = \exp\{0.5\sqrt{t} + 0.5\sin(2t)X_i + \beta Z_i\} + \varepsilon_i(t), \quad i=1, \dots, n, \quad (18)$$

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 $N(\varphi_i, 0.5^2)$ 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 Poisson process with the proportional mean rate

$$\alpha(t|X_i, Z_i) = 0.6 \exp(0.7Z_i), \quad i=1, \dots, n. \quad (19)$$

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

$$\alpha(t|X_i, Z_i) = 0.4 + 0.9Z_i, \quad i=1, \dots, n. \quad (20)$$

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

$$\lambda(t|Z_i, Z_i^*) = 0.12t^{1/2} \exp\{2Z_i + Z_i^*(t)\}, \quad (21)$$

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 longitudinal model (18) with $\beta = 0.5$ and with the sampling times models (19)–(21). The integrals are evaluated on the grid points $s_j = 0.05j$, $j = 1, 2, \dots, 69$. The summaries of performance of $\hat{\gamma}(t)$ at time points $0.5j$, $j = 1, \dots, 6$, are given in Table 2. Each entry of the tables is calculated based on 1,000 repetitions. Table 1 and Table 2 include the simulation results for $\beta = 0.5$ and $n = 200$ and 300. The expanded simulations for $\beta = 0.0$ and 1.5 and at $n = 100$ are also conducted but not reported here. The simulation studies demonstrate that the proposed estimation procedures perform 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 for both sample sizes 200 and 300. Plots of $\hat{\gamma}_1(t)$ and $\hat{\gamma}_2(t)$ for model (18) are depicted in Fig. 1 when $\beta = 0.5$ for $n = 100$ and $h = 0.3$. Figure 1a, b in the first row are under the proportional sampling model (19), Fig. 1c, d in the second row are under the additive sampling model (20) and Fig. 1e, f in the third row are under the sampling model (21). 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_1 and L_1 for testing H_{01} :

$$Y_i(t) = \exp\{0.5\sqrt{t} + \{0.5 - \theta \sin(2t)\}X_i + 0.5Z_i\} + \varepsilon_i(t), \quad i = 1, \dots, n, \quad (22)$$

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 (18). Different values of θ are to be selected to examine the power of the tests.

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 3 lists the empirical sizes and powers of the test statistics S_1 and L_1 at the significance level 0.05 under the sampling models (19)–(21). Each entry is based on 1,000 repetitions. Each p -value is estimated by generating 1,000 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 (22). 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

$$Y_i(t) = \alpha(t) + \beta Z_i + \varepsilon_i(t), \quad (23)$$

where Z_i and $\varepsilon_i(t)$ are same as those for model (18), $\beta = 1$, and $\alpha(t)$ is taken to be $1+t$ or $1+t^2$. Table 4 list the summaries of the estimation for β for two different choices of $\alpha(t)$ 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 (19)–(21). Each entry is based on 1,000 repetitions. The estimation of Lin and Ying (2001) has larger biases when the sampling

model is mis-specified under (20) and (21), 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

In this section, we apply the proposed methods to a real data example. We demonstrate how to select the link function that provides better fit of the data using the procedures given in Sect. 3.4. The estimation and inference are then carried out using the selected link function. We consider the analysis of a HIV-1 RNA data set from an AIDS clinical trial. In this study, all subjects initiated the antiretroviral treatment at time 0 (the baseline). Some subjects received a single protease inhibitor (PI) while others received a double-PI antiretroviral regimens in treating HIV-infected patients. HIV-1 RNA levels in plasma was measured repeatedly during the follow-up. The scheduled visit times were at weeks 0, 2, 4, 8, 16 and 24. But the actual visit times of individuals may vary around the scheduled visiting times. Some patients had prior antiviral treatment with non-nucleoside analogue reverse transcriptase inhibitors (NNRTI) and others did not have prior NNRTI treatment. The prior NNRTI treatment is considered to be a factor that affects the antiviral response to the antiretroviral regimens in the current study.

A total of 481 patients were enrolled in the study, with 2, 626 total visits. Owing to technical limitations, 175 measurements of HIV-1 RNA levels were censored below the detection limit, and three were censored above the detection limit. We restrict our analysis to those responses within the detectable range. This data set has been analyzed by Sun and Wu (2005). Here we use the same transformed time scale $t = \log_{10}(\text{day of actual visit} + 40) - \log_{10}(32)$ of the actual visits so that the transformed sampling time points are more evenly distributed suitable for bandwidth selection. The maximum of transformed sampling times is $\tau = 0.88$. The response variable $Y(t)$ is the change of HIV-1 RNA level using a $\log_{10}$ scale at time $t \in [0, \tau]$ from the baseline. We refer to Sun and Wu (2005) for the detailed discussions of the transformations. Let $X = 1$ denote the patients who received a double-PI treatment and $X = 0$ for patients who received a single-PI treatment. Let Z be the indicator of the prior antiviral treatment with NNRTI, with 1 for having had NNRTI and 0 for having not received NNRTI.

Analysis of Sun and Wu (2005) shows that the effect of treatment (double-PI versus single-PI) is time-varying after adjusting for the prior NNRTI antiviral treatment experience under the semiparametric additive regression model. Here we consider to fit the following generalized semiparametric model

$$\mu_i(t) = g^{-1}\{\gamma_1(t) + \gamma_2(t)X_i + \beta Z_i\}, \quad (24)$$

for $0 \leq t \leq \tau$, where $g(\cdot)$ is a known link function. In the following we illustrate the selection of $g(\cdot)$ based on the criteria (17) among the two commonly used link functions, the identity link function and the logarithm link function. We use the unit weight function $W_i(t) = 1$ and set $t_1 = 0$ and $t_2 = \tau$ in (4) for the estimation of β .

Based on the procedure given in Sect. 3.4, the κ -fold cross-validation method with $K = 50$ for the identity link function yields the bandwidth $h_{cv} = 0.06$. The plot of the total prediction error is given in Fig. 2a. The regression deviation is $RD = 3.3495 \times 10^3$ using (17). For the logarithm link function, the κ -fold cross-validation with $K = 50$ yields $h_{cv} = 0.07$. The plot of the total prediction error is similar to Fig. 2a. The corresponding regression deviation is $RD = 3.3801 \times 10^3$, which is larger than that under the identity link function. This suggests that model (24) with the identity link function provides better fit of the data.

Under the identity link function, the estimate $\hat{\beta}$ is 0.6243 with the standard error 0.0883 for $h = 0.06$. The estimates $\hat{\gamma}_1(t)$ and $\hat{\gamma}_2(t)$ and the 95 % pointwise confidence intervals are plotted in the first row in Fig. 3. The p -values for testing for time-dependence of $\gamma_2(t)$ are 0.009 and 0.022 using the test statistics S_1 and L_1 , respectively, based on 1,000 Gaussian samples. Our experience shows that the “optimal” bandwidth that minimizes the total prediction error tends to be a little small for yielding smoothed curves for the nonparametric regression function estimation. The estimation of the parametric components are not greatly affected by the choices of the bandwidth. The plot of $\hat{\beta}$ against h is given in Fig. 2b. For example, $\hat{\beta}$ is 0.6150 with the standard error 0.0888 for $h = 0.09$ and $\hat{\beta}$ is 0.6064 with the standard error 0.0894 for $h = 0.12$. The p -values for testing for time-dependence of $\gamma_2(t)$ are 0.014 and 0.034 for $h = 0.09$ using the test statistics S_1 and L_1 , respectively. The corresponding p -values are 0.039 and 0.022 for $h = 0.12$ based on the test statistics S_1 and L_1 , respectively. The estimates $\hat{\gamma}_1(t)$ and $\hat{\gamma}_2(t)$ and the 95 % pointwise confidence intervals with $h = 0.09$ and 0.12 are plotted in the second and third rows of Fig. 3, respectively. Our hypothesis tests indicate that the treatment effect changes with time. The p -values for testing $\gamma_2(t) = 0$ using the test statistics S_2 and L_2 are 0.036 and 0.044, respectively for $h = 0.06$. 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 Fig. 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.

6 Discussion

In this paper we study the generalized semiparametric regression model for longitudinal data. The semiparametric model (1) allows the covariate effects to be constant for some and time-varying for others. We proposed an estimation method that automatically adjusts for heterogeneity of sampling times. The nonparametric components of the model are estimated using the local linear estimating equations and the parametric components are estimated through the weighted profile estimating functions. Unlike the profile-based estimation methods of Lin and Carroll (2001), Lin et al. (2007) and Fan et al. (2007), the proposed method applies to the situations where the conditional distributions of the sampling times may depend on the past sampling history as well as covariates and where the subjects may dropout in the follow-up studies. Also unlike the joint modelling approaches of Martinussen and Scheike (1999), Martinussen and Scheike (2000, 2001), Lin and Ying (2001), Hu et al. (2003) and Fan and Li (2004), the proposed method utilizes the sampling process directly without having to model its intensity function $\lambda_j(t)$ or the mean rate function $\alpha_j(t)$. The model specifications of $\lambda_j(t)$ or $\alpha_j(t)$ can be very difficult and problematic especially when the sampling strategy may depend on the past sampling history. Our simulation studies show that the proposed method works well under a variety of sampling models including the proportional mean rate model, additive mean rate model and the sampling model that depends on the past sampling history. Simulations also show that, for some existing methods, misspecification of the sampling model can result in large estimation bias.

This paper presents an unified approach to the semiparametric model (1) with a general link function. Different link functions can be specified for more flexible modelling of longitudinal data. Both the categorical and continuous longitudinal responses can be modelled with appropriately chosen link functions. For example, the identity and logarithm link functions can be used for the continuous response variables while the logit link function can be used for the binary responses. Model (1) presents an opportunity for model selection of the link function. A criteria is proposed for selecting the link function that provides better fit. The procedures are illustrated with a data example.

The estimation method developed in Sects. 2.1–2.3 assumes existence of intensity for the counting processes that record the random sampling time points. The method is extended to the fixed designs in Sect. 2.4 where data are observed at the planned sampling time points. The asymptotic properties of the estimators for both cases are derived in Sect. 3.1. In the situation of a “mixed design” where some observations are made at fixed time points with positive probability while other observations are made at random sampling points, we believe that the estimating functions (3) and (4) are still valid yielding consistent estimators. However, the asymptotic properties need to be carefully worked out since the asymptotic rate of $\hat{\gamma}(t)$ is $(nh)^{-1/2}$ under the random design and is $n^{-1/2}$ under the fixed design. The derivation of the asymptotic results require clear classifications whether a observation time point is from a planned visiting time or from a random sampling time.

The proposed estimation method is a marginal approach that does not take into consideration of the correlations between the repeated measurements. Following the discussions in Sect. 3.1, the efficiency of the estimation can be improved by selecting $W_i(t) = W(t, X_i(t), Z_i(t))$, where $W(t, x, z)$ converges in probability to $w(t, x, z) = \dot{\mu}_i(t)/\{\sigma_\varepsilon^2(t|x, z)\}^2$. Since model (1) does not specify the conditional variance structure for $\sigma_\varepsilon^2(t|x, z)$, such selection can be difficult. When the dimension of covariates is small, a two-stage estimation procedure can be considered. In the first stage, estimate β and $\gamma(t)$ with the identity weight function to obtain $\hat{\beta}_I$ and $\hat{\gamma}_I(t)$. In the second stage, the updated estimation of β and $\gamma(t)$ is obtained by choosing the weight $W_i(t) = \hat{\mu}_i(t)/\{\hat{\sigma}_\varepsilon^2(t|x, z)\}^2$, where $\hat{\mu}_i(t) = \phi\{\hat{\gamma}_I^T(t)X_i(t) + \hat{\beta}_I^T Z_i(t)\}$ and $\{\sigma_\varepsilon^2(t|x, z)\}^2$ equals

$$\frac{\sum_{j=1}^n \int_0^\tau \{Y_j(s) - \phi(\hat{\gamma}_I^T(s)X_j(s) + \hat{\beta}_I^T Z_j(s))\}^2 K_h(s-t) \tilde{K}_b(X_j-x, Z_j-z) dN_j(s)}{\sum_{j=1}^n \int_0^\tau K_h(s-t) \tilde{K}_b(X_j-x, Z_j-z) dN_j(s)},$$

where $\tilde{K}_B(u) = |B|^{-1/2} \tilde{K}(B^{-1/2}u)$ and $\tilde{K}(u)$ is a multivariate kernel function and B is the bandwidth $(p+q) \times (p+q)$ positive definite matrix. The strategy may be difficult to implement when the dimension of covariates is large because of the curse of dimensionality. Further research is warranted.

Acknowledgments

The authors thank the reviewers for their constructive comments that have improved the presentation and content of the paper. The research of Yanqing Sun was partially supported by NSF grants DMS-0905777 and DMS-1208978, NIH NIAID grant 2 R37 AI054165-10 and a fund provided by UNC Charlotte. The research of Liuquan Sun 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 OO. Nonparametric inference for a family of counting processes. *Ann Stat.* 1978; 6:701–726.
- Bickel, PJ.; Klaassen, CAJ.; Ritov, Y.; Wellner, JA. Efficient and adaptive estimation for semiparametric models. Springer; New York: 1993.
- Cheng SC, Wei LJ. Inferences for a semiparametric model with panel data. *Biometrika.* 2000; 87:89–97.
- Fan, J.; Gijbels, I. Local polynomial modelling and its applications. Chapman and Hall; London: 1996.
- Fan J, Li R. New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *J Am Stat Assoc.* 2004; 99:710–723.

- Fan J, Huang T, Li R. Analysis of longitudinal data with semiparametric estimation of covariance function. *J Am Stat Assoc.* 2007; 102:632–641. [PubMed: 19707537]
- Hoover DR, Rice JA, Wu CO, Yang LP. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika.* 1998; 85:809–822.
- Hu XJ, Sun J, Wei LJ. Regression parameter estimation from panel counts. *Scand J Stat.* 2003; 30:25–43.
- Hu Z, Wang N, Carroll RJ. Profile-kernel versus backfitting in the partially linear models for longitudinal/clustered data. *Biometrika.* 2004; 91:251–262.
- Lin X, Carroll RJ. Semiparametric regression for clustered data using generalized estimating equations. *J Am Stat Assoc.* 2001; 96:1045–1056.
- Lin DY, Ying Z. Semiparametric and nonparametric regression analysis of longitudinal data (with discussion). *J Am Stat Assoc.* 2001; 96:103–113.
- Lin DY, Wei LJ, Ying Z. Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika.* 1993; 80:557–572.
- Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. *J R Stat Soc Ser B.* 2000; 62(Part 4):711–730.
- Lin H, Song PX-K, Zhou QM. Varying-coefficient marginal models and applications in longitudinal data analysis. *Sankhya.* 2007; 69:581–614.
- Martinussen T, Scheike TH. A semiparametric additive regression model for longitudinal data. *Biometrika.* 1999; 86:691–702.
- Martinussen T, Scheike TH. A nonparametric dynamic additive regression model for longitudinal data. *Ann Stat.* 2000; 28:1000–1025.
- Martinussen T, Scheike TH. Sampling adjusted analysis of dynamic additive regression models for longitudinal data. *Scand J Stat.* 2001; 28:303–323.
- Martinussen, T.; Scheike, TH. *Dynamic regression models for survival data.* Springer; New York: 2006.
- Rice JA, Silverman BW. Estimating the mean and covariance structure nonparametrically when the data are curves. *J R Stat Soc Ser B.* 1991; 53:233–243.
- Sun Y. Estimation of semiparametric regression model with longitudinal data. *Lifetime Data Anal.* 2010; 16:271–298. [PubMed: 19890712]
- Sun J, Wei LJ. Regression analysis of panel count data with covariate-dependent observation and censoring times. *J R Stat Soc Ser B.* 2000; 62:293–302.
- Sun Y, Wu H. Semiparametric time-varying coefficients regression model for longitudinal data. *Scand J Stat.* 2005; 32:21–47.
- Sun Y, Gilbert PB, McKeague IW. Proportional hazards models with continuous marks. *Ann Stat.* 2009; 37:394–426. [PubMed: 19838313]
- Tian L, Zucker D, Wei LJ. On the Cox model with time-varying regression coefficients. *J Am Stat Assoc.* 2005; 100:172–183.
- Van der Vaart, AW. *Asymptotic statistics.* Cambridge University Press; Cambridge: 1998.
- Wu H, Liang H. Backfitting random varying-coefficient models with time-dependent smoothing covariates. *Scand J Stat.* 2004; 31:3–19.
- Zhang Y. A semiparametric pseudolikelihood estimation method for panel count data. *Biometrika.* 2002; 89:39–48.

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 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 $h \rightarrow 0$; $E|N_i(t+h) - N_i(t-h)|^{2+\nu} = O(h)$, for some $\nu > 0$; the limit

$$\lim_{n \rightarrow \infty} h E \left\{ \int_0^\tau w_i(s) \{Y_i(s) - \mu_i(s)\} X_i(s) K_h(s-t) dN_i(s) \right\}^{\otimes 2} = \sum_e(t) \text{ exists and is finite.}$$

Let $u_a(\gamma, \beta) = E([\phi\{\gamma_0^T(t)X_i(t) + \beta_0^T Z_i(t)\} - \phi\{\gamma^T(t)X_i(t) + \beta^T Z_i(t)\}]X_i(t)\xi_i(t)\alpha_i(t))$. 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 $e_{\beta,xx}(t) = E[w_i(t) \phi\{\gamma_\beta^T(t)X_i(t) + \beta^T Z_i(t)\} \{X_i(t)\}^{\otimes 2} \alpha_i(t)\xi_i(t)]$ and $e_{\beta,xz}(t) = E[w_i(t) \phi\{\gamma_\beta^T(t)X_i(t) + \beta^T Z_i(t)\} X_i(t)(Z_i(t))^T \alpha_i(t)\xi_i(t)]$. 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 $H = \text{diag}\{I_q, h I_q\}$. 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 et al. (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 $n \rightarrow \infty$, $H\tilde{\gamma}_a(t, \beta) \xrightarrow{P} \gamma_{a\beta}(t)$,

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

and $H^2 \tilde{\gamma}(t, \beta) / \beta^2$ 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 :

$$U_a(\gamma_{a\beta} + H^{-1}\theta, \beta) = \sum_{i=1}^n \int_0^\tau W_i(s) \{Y_i(s) - \tilde{\mu}_a(s, \gamma_{a\beta} + H^{-1}\theta, \beta | X_i, Z_i)\} \times \tilde{X}_i(s, s-t) K_h(s-t) dN_i(s), \quad (25)$$

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

By the Glivenko-Cantelli theorem,

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

uniformly in $t \in [t_1, t_2]$, $\beta \in \mathcal{N}_\beta$ and $\theta \in \mathcal{N}_\theta$, a neighborhood of $\mathbf{0}_{2q} \in \mathbb{R}^{2q}$, where $\dot{\mu}_{i\beta}(t) = \dot{\phi} \{ \gamma_\beta^T(t) X_i(t) + \beta^T Z_i(t) \}$. 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 et al. (2009) that $\tilde{\theta} \xrightarrow{P} \mathbf{0}_{2q}$ uniformly in t and β . Thus

$$H\tilde{\gamma}_a(t, \beta) - \gamma_{a\beta}(t) \xrightarrow{P} \mathbf{0}_{2q} \quad \text{uniformly in } t \in [t_1, t_2] \text{ and } \beta \in \mathcal{N}_\beta. \quad (26)$$

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

$$\left\{ \frac{\partial U_a(\gamma_a, \beta)}{\partial \gamma_a} \frac{\partial \tilde{\gamma}_a(t, \beta)}{\partial \beta} + \frac{\partial U_a(\gamma_a, \beta)}{\partial \beta} \right\} \Big|_{\gamma_a = \tilde{\gamma}_a(t, \beta)} = \mathbf{0}_{2q}. \quad (27)$$

Note that

$$\begin{aligned}
& -n^{-1} H^{-2} \frac{\partial U_a(\gamma_a, \beta)}{\partial \gamma_a} = n^{-1} \sum_{i=1}^n \int_0^\tau W_i(s) \dot{\phi} \{ \gamma_a^T \tilde{X}_i(s, s-t) + \beta^T Z_i(s) \} H^{-2} \{ \tilde{X}_i(s, s-t) \}^{\otimes 2} K_h(s-t) dN_i(s) \\
&= n^{-1} \sum_{i=1}^n \int_0^\tau W_i(s) \dot{\phi} \{ (H\gamma_a)^T H^{-1} \tilde{X}_i(s, s-t) + \beta^T Z_i(s) \} H^{-2} \{ \tilde{X}_i(s, s-t) \}^{\otimes 2} K_h(s-t) dN_i(s).
\end{aligned} \quad (28)$$

By the Glivenko-Cantelli theorem, the process

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

converges in probability to

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

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 (26) that

$$\xrightarrow{P} E \left[w_i(t) \dot{\phi} \{ \gamma_{\beta}^T(t) X_i(t) + \beta^T Z_i(t) \} \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix} \otimes \{ X_i(t) \}^{\otimes 2} \xi_i(t) \alpha_i(t) \right],$$

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,

$$\begin{aligned} & -n^{-1} H^{-1} \frac{\partial U_a(\gamma_a, \beta)}{\partial \beta} \Big|_{\gamma_a = \tilde{\gamma}_a(t, \beta)} \\ &= n^{-1} \sum_{i=1}^n \int_0^{\tau} W_i(s) \dot{\phi} \{ \gamma_a^T \tilde{X}_i(s, s-t) + \beta^T Z_i(s) \} H^{-1} \tilde{X}_i(s, s-t) (Z_i(s))^T K_h(s-t) dN_i(s) \Big|_{\gamma_a = \tilde{\gamma}_a(t, \beta)} \quad (29) \\ & \xrightarrow{P} \begin{pmatrix} E \left[w_i(t) \dot{\phi} \{ \gamma_{\beta}^T(t) X_i(t) + \beta^T Z_i(t) \} X_i(t) (Z_i(t))^T \xi_i(t) \alpha_i(t) \right] \\ \mathbf{0}_q \end{pmatrix}, \end{aligned}$$

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 (27) that

$$H \frac{\partial \tilde{\gamma}_a(t, \beta)}{\partial \beta} \xrightarrow{P} \begin{pmatrix} -(e_{\beta, xx}(t))^{-1} e_{\beta, xz}(t) \\ \mathbf{0}_q \end{pmatrix}, \quad (30)$$

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, $H^{-2} \tilde{\gamma}(t, \beta) / \beta^2$ 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 $nh \rightarrow \infty$ and $nh^5 = O(1)$,

$$(nh)^{1/2} \{ \tilde{\gamma}(t, \beta_0) - \gamma_0(t) - \frac{1}{2} \mu_2 h^2 \dot{\gamma}_0^T(t) \} = (e_{xx}(t))^{-1} (nh)^{1/2} n^{-1} U_{\gamma}(\gamma_0, \beta_0) + o_p(1), \quad (31)$$

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

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

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

$$\left(n^{-1}\sum_{i=1}^n\int_0^\tau W_i(s)(\tilde{X}_i(s,s-t))^{\otimes 2}H^{-1}\dot{\phi}\{\gamma_{0a}^T\tilde{X}_i(s,s-t)+\beta_0^TZ_i(s)\}K_h(s-t)dN_i(s)+o_p(\rho_n^{-1}) \right)^{-1} \times \rho_n n^{-1}U_a(\gamma_{0a},\beta_0).$$

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

$$\rho_n(\tilde{\gamma}(t,\beta_0)-\gamma_0(t))=(e_{xx}(t))^{-1}\rho_n n^{-1}U_1(\gamma_{0a},\beta_0)+o_p(\rho_n^{-1}), \quad (32)$$

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

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

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

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

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

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

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

$$\begin{aligned}
\rho_n n^{-1}U_1(\gamma_{0a},\beta_0) &= \rho_n n^{-1}\sum_{i=1}^n\int_0^\tau W_i(s)[Y_i(s)-\mu_i(s)+\{\mu_i(s)-\tilde{\mu}_a(s,\gamma_{0a},\beta_0|X_i,Z_i)\}]X_i(s)K_h(s-t)dN_i(s) \\
&= \rho_n n^{-1}U_\gamma(\gamma_0,\beta_0)+\frac{1}{2}\mu_2\rho_n h^2 e_{xx}(t)\ddot{\gamma}_0(t)+o_p(\rho_n h^2),
\end{aligned} \quad (33)$$

uniformly in $t \in [t_1, t_2]$. By (32) and (33),

$$\rho_n\{\tilde{\gamma}(t, \beta_0) - \gamma_0(t) - \frac{1}{2}\mu_2 h^2 \ddot{\gamma}_0^T(t)\} = (e_{xx}(t))^{-1} \rho_n n^{-1} U_\gamma(\gamma_0, \beta_0) + o_p(\rho_n^{-1}) + o_p(\rho_n h^2), \quad (34)$$

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

Following the same lines as the proof in Appendix A of Tian et al. (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

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

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), $\widehat{\beta} \xrightarrow{P} \beta_0$.

By Lemma 1 and the Glivenko-Cantelli theorem,

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

It follows that

$$\begin{aligned} & -n^{-1} \frac{\partial U(\beta)}{\partial \beta} \Big|_{\beta=\beta_0} \\ & = n^{-1} \sum_{i=1}^n \int_{t_1}^{t_2} W_i(s) \dot{\phi}\{\tilde{\gamma}(s, \beta_0)\}^T X_i(s) + \beta_0^T Z_i(s) \times \left\{ \left(\frac{\partial \tilde{\gamma}(s, \beta_0)}{\partial \beta} \right)^T X_i(s) + Z_i(s) \right\}^{\otimes 2} dN_i(s) + o_p(1) \\ & \xrightarrow{P} E \left\{ \int_{t_1}^{t_2} w_i(s) \dot{\phi}\{\gamma_0(s)\}^T X_i(s) + \beta_0^T Z_i(s) \{-(e_{xz}(s))^T (e_{xx}(s))^{-1} X_i(s) + Z_i(s)\}^{\otimes 2} dN_i(s) \right\} \\ & = A, \end{aligned} \quad (35)$$

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,

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

By Lemmas 1 and 2,

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

Hence

$$\begin{aligned} n^{-1/2} U(\beta_0) &= n^{-1/2} \sum_{i=1}^n \int_{t_1}^{t_2} W_i(s) [Y_i(s) - \mu_i(s)] \left\{ \left(\frac{\partial \tilde{\gamma}(s, \beta_0)}{\partial \beta} \right)^T X_i(s) + Z_i(s) \right\} dN_i(s) + o_p(1) \\ &= n^{-1/2} \sum_{i=1}^n \int_{t_1}^{t_2} W_i(s) [Y_i(s) - \mu_i(s)] \{Z_i(s) - (e_{xz}(s))^T (e_{xx}(s))^{-1} X_i(s)\} dN_i(s) + o_p(1), \end{aligned} \quad (36)$$

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

$$\Sigma = E \left(\int_{t_1}^{t_2} W_i(s) [Y_i(s) - \mu_i(s)] \{Z_i(s) - (e_{xz}(s))^T (e_{xx}(s))^{-1} X_i(s)\} dN_i(s) \right)^{\otimes 2}. \quad (37)$$

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 (35) and (36) that $n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{\mathcal{D}} N(0, A^{-1} \Sigma A^{-1})$ as $n \rightarrow \infty$.

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 (36),

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

where

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

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

$$(nh)^{1/2}(\widehat{\gamma}(t) - \gamma_0(t) - \frac{1}{2}\mu_2 h^2 \ddot{\gamma}_0^T(t)) \xrightarrow{\mathcal{D}} N(0, \sum_{\gamma}(t)), \quad (38)$$

as $nh^2 \rightarrow \infty$ 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).

Proof of Theorem 3

By (31), (35) and (36), we have

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

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

Proof of (10)

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

$$\begin{aligned} D(s) &= A^{-1} w_i(s) \{Z_i(s) - (e_{xz}(s))^T (e_{xx}(s))^{-1} X_i(s)\}^T \sigma_{\varepsilon}(s|X_i, Z_i) \alpha_i^{1/2}(s) \\ &\quad - \sum_0^{-1} \{Z_i(s) - (e_{xz}(s))^T (e_{xx}(s))^{-1} X_i(s)\}^T \{\dot{\mu}_i(s) / \sigma_{\varepsilon}(s|X_i, Z_i)\} \alpha_i^{1/2}(s). \end{aligned}$$

Then the matrix

$$\begin{aligned} E \left(\int_{t_1}^{t_2} D(s) D(s)^T ds \right) &= A^{-1} \sum A^{-1} - A^{-1} A \sum_0^{-1} - \sum_0^{-1} A A^{-1} + \sum_0^{-1} \sum_0 \sum_0^{-1} \\ &= A^{-1} \sum A^{-1} - \sum_0^{-1} \end{aligned}$$

is nonnegative definite.

Fig. 1.

Plots of $\hat{\gamma}(t)$ for model (18) when $\gamma_1(t) = 0.5t^{1/2}$, $\gamma_2(t) = 0.5 \sin(2t)$ and $\beta = 0.5$ for $n = 100$ and $h = 0.3$. **a, b** in the *first row* are under the proportional sampling model (19), **c, d** in the *second row* are under the additive sampling model (20) and **e, f** in the *third row* are under the sampling model (21). The *solid lines* are the estimates and the *dashed lines* are the true curves

Fig. 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) under the identity link function for the HIV-1 RNA data

Fig. 3.

The plots of the estimates $\hat{\gamma}_1(t)$ and $\hat{\gamma}_2(t)$ and the 95 % pointwise confidence intervals for three different bandwidths $h = 0.06, 0.09$ and 0.12 under the identity link function for the HIV-1 RNA data

Table 1

Summary of bias, SEE, ESE and CP for $\gamma(\theta)$ under the longitudinal model (18) with $\beta = 0.5$, $\gamma_1(t)=0.5\sqrt{t}$ and $\gamma_2(\theta) = 0.5 \sin(2\theta)$ and with the sampling times models (19)–(21)

n	h	Bias	SEE	ESE	CP	$RM\ SE_1$	$RM\ SE_2$
Under sampling times model (19)							
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
Under sampling times model (20)							
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
Under sampling times model (21)							
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

Table 2

Summary of bias, SEE, ESE and CP for $\gamma(\theta)$ at $t = 0.5, 1.0, 1.5, 2.0, 2.5, 3.0$ for $n = 200$ and $h = 0.4$ under the longitudinal model (18) with $\beta = 0.5$, $\gamma_1(t) = 0.5\sqrt{t}$ and $\gamma_2(\theta) = 0.5\sin(2\theta)$ and with the sampling times models (19)–(21)

t	$\gamma_1(t) = 0.5\sqrt{t}$				$\gamma_2(\theta) = 0.5\sin(2\theta)$			
	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP
Under sampling times model (19)								
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
Under sampling times model (20)								
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
Under sampling times model (21)								
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

Table 3

Empirical sizes and powers of the tests based on S_1 and L_1 at nominal level $\alpha = 0.05$ under the longitudinal model (22) with the sampling times models (19)–(21)

n	h	Size		Power					
		$\theta = 0$		$\theta = 0.1$		$\theta = 0.15$		$\theta = 0.2$	
		S_1	L_1	S_1	L_1	S_1	L_1	S_1	L_1
Under sampling times model (19)									
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 (20)									
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 (21)									
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 4

Comparisons of the estimation for β using the proposed method and the method of Lin and Ying (2001) under model (23) with $\beta = 1$ and two different choices of $\alpha(t)$ for $n = 200$

h	$\alpha(t) = 1 + t$				$\alpha(t) = 1 + t^3$			
	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP
Under sampling times model (19)								
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 (20)								
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 (21)								
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