

# Analysis of asynchronous longitudinal data with partially linear models

Li Chen and Hongyuan Cao\*

*Department of Statistics, University of Missouri-Columbia,  
Columbia, MO 65211, U.S.A.*

*e-mail: [lichen@mail.missouri.edu](mailto:lichen@mail.missouri.edu); [caohong@missouri.edu](mailto:caohong@missouri.edu)*

**Abstract:** We study partially linear models for asynchronous longitudinal data to incorporate nonlinear time trend effects. Local and global estimating equations are developed for estimating the parametric and nonparametric effects. We show that with a proper choice of the kernel bandwidth parameter, one can obtain consistent and asymptotically normal parameter estimates for the linear effects. Asymptotic properties of the estimated nonlinear effects are established. Extensive simulation studies provide numerical support for the theoretical findings. Data from an HIV study are used to illustrate our methodology.

**MSC 2010 subject classifications:** Primary 62E20; secondary 62G05.

**Keywords and phrases:** Asynchronous longitudinal data, estimating estimations, local polynomials, partially linear models.

Received September 2016.

## Contents

1	Introduction . . . . .	1550
2	Model estimation and inference . . . . .	1551
2.1	Estimating equations . . . . .	1551
2.2	Asymptotic properties . . . . .	1553
2.3	Practical implementation . . . . .	1555
3	Numerical studies . . . . .	1557
3.1	Simulation . . . . .	1557
3.2	An application to HIV dataset . . . . .	1560
4	Concluding remarks . . . . .	1562
A	Appendix: Proofs of Theorems . . . . .	1563
A.1	Proof of Theorem 1 . . . . .	1563
A.2	Proof of Theorem 2 . . . . .	1566
A.3	Proof of Theorem 3 . . . . .	1567
	Acknowledgement . . . . .	1567
	References . . . . .	1567

---

\*The research of Hongyuan Cao was supported in part by a grant from University of Missouri Research Board.

## 1. Introduction

Asynchronous longitudinal data refers to the data structure that measurement times for the longitudinal response and longitudinal covariates are mismatched within individuals. For example, educational studies associate subjective evaluations of students' performance and objective test results. However, subjective information is usually collected through interviews or phone calls at different times from the exams. In clinical epidemiology, one may study the links between biomarkers, sampled repeatedly at lab visits, with self-reported measures of function and quality of life, captured via outpatient phone interviews. With the adoption of electronic medical records, an individual's information may be pooled from different sources to make treatment decisions, creating an asynchronous longitudinal data structure.

While regression analysis for synchronous longitudinal data (Liang and Zeger, 1986; Diggle et al., 2002) has been widely studied, there has been limited work on the analysis of regression models using asynchronous longitudinal data. Xiong and Dubin (2010) employed an ad hoc binning step to synchronize covariates and response measurements to use existing methods for synchronous data. Sentürk et al. (2012) explicitly addressed the asynchronous setting for generalized varying coefficient models with one covariate but did not provide the theoretical properties of the estimators. Cao et al. (2015) proposed a nonparametric kernel weighting approach for the generalized linear models to explicitly deal with the asynchronous structure and rigorously established the consistency and asymptotic normality of the resulting estimates. This was extended to a more general set up in Cao et al. (2016).

Despite these progresses, one limitation of marginal generalized linear models associating longitudinal response with corresponding longitudinal covariates is that non-linear time trends cannot be taken into account. Partially linear models provide a convenient vehicle to retain the parsimonious description of the relationship between the response variable and its covariates and to use a flexible nonparametric function to capture the nonlinear time trend effects. In this paper, we study the partially linear models for asynchronous longitudinal data:

$$E\{Y(t)|X(t)\} = \alpha(t) + X(t)^T\beta, \quad (1.1)$$

where  $t$  is a univariate time index,  $Y(t)$  is the response scalar,  $X(t)$  is the  $p \times 1$  covariate vector at time  $t$ ,  $\alpha(t)$  is an unspecified baseline function of  $t$  and  $\beta$  is vector of unknown regression coefficients. Model (1.1) does not require data analysts to parameterize the baseline function which may be difficult in practice. It keeps the flexibility of the nonparametric models for the baseline function, while maintaining the exploratory power of parametric models. Therefore, model (1.1) and its variations have been extensively studied in the literature for synchronous longitudinal data (Zeger and Diggle, 1994; Moyeed and Diggle, 1994; Lin and Carroll, 2001; Wang et al., 2005). Lin and Ying (2001) proposed estimation procedures under the framework of counting processes. Fan and Li (2004) developed difference-based method and profile least squares approach for the estimation of

$\beta$ . Due to the asynchronous data structure, their approaches cannot be directly used for our purpose.

In this paper, we propose a system of global and local estimating equations for model (1.1). Furthermore, under appropriate regularity conditions, we show that within a range of choices of the smoothing parameter (the kernel bandwidth), the estimator of  $\beta$  has the same rate of convergence as in the generalized linear model (Cao et al., 2015), slower than the parametric rate  $n^{-1/2}$  for synchronous data (Lin and Ying, 2001; Fan and Li, 2004). The asymptotic normality is also obtained for the estimates of the linear parameter and nonlinear function. Inferences on  $\beta$  are based on a resampling approach and we use bootstrap for the inference of the nonparametric function  $\alpha(\cdot)$ . What distinguishes our work from Cao et al. (2015) is that a non-parametric time trend effect is explicitly modeled with associated theoretical properties rigorously established. Unlike Lin and Ying (2001); Fan and Li (2004) for synchronous longitudinal data, a global estimating equation is proposed to down-weight those observations which are distant in time and enables the use of all covariate observations for each observed response with asynchronous longitudinal data. Our approach is related to some existing quasi/partial-likelihood based algorithms for partially linear models (Cai et al., 2007, 2008; Lu and Zhang, 2010). However, the asynchronous longitudinal data structure poses new challenges for both model estimation and inferences.

The remainder of the paper is organized as follows. In Section 2, we describe the global and local estimating equations for the parameter  $\beta$  and function  $\alpha(\cdot)$  in model (1.1), and provide the corresponding theoretical findings. In Section 3, we conduct simulation studies to examine the finite-sample properties of the new procedure and apply our method to dataset from an HIV study. Concluding remarks are given in Section 4. All technical proofs are relegated in the Appendix.

## 2. Model estimation and inference

In this section, we propose a system of estimating equations to simultaneously estimate  $\beta$  and  $\alpha(t)$ . The main idea is to use a global estimating equation for  $\beta$  by kernel weighting and the local polynomial technique (Fan and Gijbels, 1996) to construct a set of locally weighted estimating equations for  $\alpha(t)$ .

### 2.1. Estimating equations

Suppose there are  $n$  subjects in the study. We formulate the observation process using a bivariate counting process for the observation times of the covariate and response variables, similar to Martinussen and Scheike (1999); Lin and Ying (2001); Cao et al. (2015, 2016). Specifically, for subject  $i = 1, \dots, n$ , the observation times of the longitudinal covariate process  $X_i(t)$  and response process

$Y_i(s)$  are generated from a bivariate counting process.

$$N_i(t, s) = \sum_{j=1}^{L_i} \sum_{k=1}^{M_i} I(T_{ij} \leq t, S_{ik} \leq s)$$

counts the number of observation times up to  $t$  on the response and up to  $s$  on the covariates, where  $\{T_{ij}, j = 1, \dots, L_i\}$  are the observation times of the response and  $\{S_{ik}, k = 1, \dots, M_i\}$  are the observation times of the covariates.

If the nonparametric function  $\alpha(t)$  were known, partially linear models reduce to the linear models, and thus we can adopt methods in Cao et al. (2015) to estimate  $\beta$ . Motivated by this, we first propose an estimating equation to solve  $\beta$ , with  $\alpha(s)$  being fixed at its current value. The estimating equation is

$$U(\beta) = n^{-1} \sum_{i=1}^n \int_0^1 \int_0^1 K_h(t-s) X_i(s) \left[ Y_i(t) - \alpha(s) - X_i(s)^T \beta \right] dN_i(t, s) = 0, \quad (2.1)$$

where  $K_h(t) = K(t/h)/h$ ,  $K(t)$  is a symmetric kernel function, usually taken to be the Epanechnikov kernel  $K(t) = 0.75(1-t^2)_+$  and  $h$  is the bandwidth. In (2.1), longitudinal measurements of the nonparametric function  $\alpha(\cdot)$  are the same as the longitudinal covariates  $X(\cdot)$  without loss of generality. The proposed method equally works if longitudinal measurements of the nonparametric function  $\alpha(\cdot)$  are the same as the longitudinal response  $Y(\cdot)$ .

Next, in order to estimate  $\alpha(s)$ , we approximate it locally by a linear function

$$\alpha(s) \approx \gamma_0(t_0) + \gamma_1(t_0)(s - t_0) \quad (2.2)$$

for  $s$  in a neighborhood of  $t_0$ , where  $\gamma_0(t_0) = \alpha(t_0)$  and  $\gamma_1(t_0) = \dot{\alpha}(t_0)$ . The superscript dot denotes the first-order derivative. Given  $\beta$ , we propose to solve the following kernel-weighted local estimating equations for  $\alpha(\cdot)$ :

$$n^{-1} \sum_{i=1}^n \iint K_{h_1, h_2}(s - t_0, t - t_0) \{ Y_i(t) - \gamma_0(t_0) - \gamma_1(t_0)(s - t_0) - \beta^T X_i(s) \} dN_i(t, s) = 0 \quad (2.3)$$

$$n^{-1} \sum_{i=1}^n \iiint (s - t_0) K_{h_1, h_2}(s - t_0, t - t_0) \{ Y_i(t) - \gamma_0(t_0) - \gamma_1(t_0)(s - t_0) - \beta^T X_i(s) \} dN_i(t, s) = 0, \quad (2.4)$$

where  $K_{h_1, h_2}(t, s) = K(t/h_1, s/h_2)/(h_1 h_2)$  and  $K(t, s)$  is a bivariate kernel function, say, the product of univariate Epanechnikov kernels  $K(t, s) = 0.5625(1-t^2)_+(1-s^2)_+$ . Altogether, we need to solve three estimating equations (2.1), (2.3) and (2.4) interactively: solve (2.1) for  $\beta$  with  $\alpha(\cdot)$  being fixed, and solve (2.3)-(2.4) for  $\alpha(\cdot)$  with  $\beta$  fixed.

We now present an iterative algorithm to implement our estimation procedure. The algorithm is given as follows:

Step 0 (initialization step). Choose an initial estimate  $\hat{\alpha}(\cdot) = \hat{\alpha}^{(0)}(\cdot)$ , for example,  $\hat{\alpha}^{(0)}(t) = 0$ . Solve (2.1) to obtain  $\hat{\beta}^{(0)}$ . Set  $\hat{\beta} = \hat{\beta}^{(0)}$ .

Step 1. Fix  $\hat{\beta}$  at the current value. For  $t_0 = S_{ik}$ , solve (2.3) and (2.4) to obtain  $\hat{\gamma}_0(S_{ik})$  and  $\hat{\gamma}_1(S_{ik})$ ,  $i = 1, \dots, n$ ;  $k = 1, \dots, M_i$ .

Step 2. Update  $\hat{\beta}$  by solving equation (2.1) with fixed  $\alpha(S_{ik}) = \hat{\gamma}_0(S_{ik})$ ,  $i = 1, \dots, n$ ;  $k = 1, \dots, M_i$ .

Step 3. Repeat Steps 1 and 2 until convergence.

Step 4. Fix  $\hat{\beta}$  at its estimated value from Step 3. The final estimate of  $\alpha(t)$  is  $\hat{\gamma}_0(t) = \hat{\gamma}_0(t; h_1, h_2, \hat{\beta})$ , where  $\{\hat{\gamma}_0(t; h_1, h_2, \hat{\beta}), \hat{\gamma}_1(t; h_1, h_2, \hat{\beta})\}$  are obtained by solving (2.3) and (2.4).

The basic idea behind the foregoing algorithm is intuitive: estimate  $\alpha(s)$  locally via (2.3) and (2.4), and then use all of the data and (2.1) to estimate  $\beta$ , with  $\hat{\alpha}(s)$  replacing  $\alpha(s)$ . The proposed algorithm above is similar in spirit to Cai et al. (2007, 2008); Lu and Zhang (2010) for partially linear hazards regression models with multivariate survival data and transformation models with univariate survival data. All these algorithms, in their own contexts, alternatively optimize the global and local quasi-likelihood functions or estimating equations until convergence.

In the synchronous case where  $s = t$ , our algorithm is different from that in Fan and Li (2004), where a difference-based method or a profile least squares approach was used to get a closed form estimate of  $\beta$  with parametric rate of convergence. The baseline function  $\alpha(\cdot)$  was estimated by smoothing the partial residuals using a local linear fit. As the rate of convergence for  $\hat{\beta}$  is faster than that for the nonparametric estimator, the errors in estimating  $\beta$  were negligible and the value of  $\beta$  can be considered as known in the nonparametric estimation of  $\alpha(\cdot)$ . The asynchronous structure greatly complicates the estimation and inference procedure and we cannot obtain a closed form expression for  $\beta$  even in the linear model case. Therefore, inferential procedures in Fan and Li (2004) are not applicable.

The estimation procedure involves choosing smoothing parameters on two quite different levels. In step 1 of the algorithm the aim is estimation of the nonparametric part  $\alpha(t)$ . However, in step 2, the goal is to estimate the parametric part  $\beta$ . In step 1, we use bandwidth that falls into the theoretically allowable range specified in Theorem 2.2 below. In step 2, due to the asynchronous structure of the observed longitudinal data, commonly used approaches such as cross-validation do not work here. We propose to minimize the mean squared errors by calculating the bias and variance of  $\hat{\beta}$  separately. Simulation studies show that our procedure is quite robust with different choices of bandwidths.

## 2.2. Asymptotic properties

In this section, we establish the asymptotic properties of the estimators  $\hat{\beta}$  and  $\hat{\gamma}_0(t)$ . For subject  $i$ , we allow the observations of  $X_i(\cdot)$  and the observations of  $Y_i(\cdot)$  to be arbitrarily correlated. We specify our assumptions on the covariance structure as follows. For  $s, t \in [0, 1]$ , let  $\text{var}\{Y(t) \mid X(t)\} = \sigma\{t, X(t)\}^2$  and

$\text{cov}\{Y(s), Y(t) \mid X(s), X(t)\} = r\{s, t, X(s), X(t)\}$ . Observe that the conditional variances and correlations of  $Y$  are completely unspecified and may depend on  $X$ . We need the following conditions.

- (A1)  $N_i(t, s)$  is independent of  $(Y_i, X_i)$  and moreover,  $E\{dN_i(t, s)\} = \lambda(t, s)dt ds$ , where  $\lambda(t, s)$  is a twice-continuously differentiable function for any  $0 \leq t, s \leq 1$ . In addition, Lebesgue measure for  $\mathcal{G} = \{\lambda(s, s) > 0, s \in [0, 1]\}$  is strictly positive. For  $t_1 \neq s_1, t_2 \neq s_2$ ,  $P\{dN(t_1, t_2) = 1 \mid N(s_1, s_2) - N(s_1-, s_2-) = 1\} = f(t_1, t_2, s_1, s_2)dt_1 dt_2$  where  $f(t_1, t_2, s_1, s_2)$  is continuous for  $t_1 \neq s_1, t_2 \neq s_2$  and  $f(t_1 \pm, t_2 \pm, s_1 \pm, s_2 \pm)$  exists.
- (A2)  $\mathcal{F} = \{s \in [0, 1] \mid X(s) \neq 0\}$  has positive Lebesgue measure with probability 1 and for  $s \in \mathcal{F}$ ,  $X(s)$  is not a constant function with probability 1.
- (A3) For any  $(s, t) \in [0, 1]^{\otimes 2}$ ,  $E\{X(s)X(s)^T\}$  is continuously twice-differentiable. In addition,  $\alpha_0(s)$  and  $E\{X(s)\}$  are twice continuously differentiable for  $s \in [0, 1]$ , where  $\alpha_0(s)$  is the true nonparametric baseline function. Moreover,  $E\{\|X(t)\|^4\} < \infty$  for  $t \in [0, 1]$ . Furthermore,

$$\int E\left[\|X(s)X(s)^T\| \sigma\{s, X(s)\}^2\right] \lambda(s, s) ds < \infty \quad \text{and}$$

$$\int \|\text{var}\{X(s)\}\| \lambda(s, s) ds < \infty.$$

- (A4)  $K(\cdot)$  is a symmetric density function satisfying  $\int zK(z)dz = 0$ ,  $\int z^2K(z)dz < \infty$  and  $\int K(z)^2dz < \infty$ . Moreover,  $K(\cdot, \cdot)$  is a bivariate density function satisfying  $\int z_1K(z_1, z_2)dz_1 dz_2 = \int z_2K(z_1, z_2)dz_1 dz_2 = 0$ ,  $\iint z_1^2K(z_1, z_2)dz_1 dz_2 < \infty$  and  $\iint z_2^2K(z_1, z_2)dz_1 dz_2 < \infty$ .
- (A5)  $nh \rightarrow \infty$  and  $nh^5 \rightarrow 0$ .

Condition (A1) requires that the observation process to be independent of both the response and covariates. We require that  $\lambda(t, s)$  to be positive when  $s = t$  for non-zero measure of time  $s$ . The requirement on the intensity function  $\lambda(t, s)$  is quite mild. Condition (A2) ensures identifiability of  $\beta$  and condition (A3) posits smoothness assumptions on the expectation of some functionals of  $X(s)$  and  $\alpha_0(s)$ . It provides additional regularity conditions on the observation intensity  $\lambda(\cdot, \cdot)$ . These assumptions can be relaxed similar to Cao et al. (2016) that allows the intensity function to depend on covariate process  $X(\cdot)$ . Condition (A4) and (A5) specify valid kernels and bandwidths.

The following theorem, which is established in the appendix, states the asymptotic properties of  $\hat{\beta}$ .

**Theorem 2.1.** *Under conditions (A1)-(A5), the asymptotic distribution of  $\hat{\beta}$  satisfies*

$$\sqrt{nh}(\hat{\beta} - \beta_0) \rightarrow N(0, A^{-1}\Sigma(A^T)^{-1}), \quad (2.5)$$

where

$$A = \int \text{Var}\{X(s)\} \lambda(s, s) ds$$

and

$$\Sigma = \int K(z)^2 dz \int E \left[ X(s) X(s)^T \sigma^2 \{s, X(s)\} \right] \lambda(s, s) ds.$$

The asymptotic result has the same rate of convergence as that in Cao et al. (2015) when the identity link function is used but larger variance. This extra variability is due to estimation of the nonparametric function  $\alpha_0(\cdot)$ . In the special case that  $E\{X(t)\} = 0, \forall t \in [0, 1]$  our results coincide with those in Cao et al. (2015). The asymptotic bias is of order  $h^2$  from the proof in the Appendix and vanishes under condition (A5). Compared with partially linear models with synchronous longitudinal data (Fan and Li, 2004), we have a slower rate of convergence.

We need additional conditions to establish the limiting distribution of the nonparametric function  $\alpha(\cdot)$ .

$$(A5^*) \quad nh_1 h_2 \rightarrow \infty \text{ and } (nh_1 h_2)^{1/2} (h_1^2 + h_2^2) \rightarrow 0.$$

**Theorem 2.2.** *Under conditions (A1)-(A4) and (A5\*), we have*

$$\sqrt{nh_1 h_2} \{ \hat{\gamma}_0(s; \hat{\beta}) - \alpha_0(s) \} \xrightarrow{d} N(0, \Sigma^*(s)), \quad \text{as } n \rightarrow \infty, \quad (2.6)$$

where

$$\Sigma^*(s) = \frac{1}{\lambda(s, s)} \iint K(z_1, z_2)^2 dz_1 dz_2 E \left[ \sigma^2 \{s, X(s)\} \right].$$

The bias is of order  $h_1^2 + h_2^2$  and vanishes through condition (A5\*). The rate of convergence is slower than the regular nonparametric  $n^{2/5}$  rate. This is the price to pay for the asynchronous data structure. The variance has a similar form to that in Fan and Li (2004) except we have a bivariate counting process and intensity function.

### 2.3. Practical implementation

As shown in Theorem 1, the asymptotic variance of  $\hat{\beta}$  has a standard sandwich form  $A^{-1} \Sigma A^{-1}$ . However, it is challenging to estimate parameters  $A$  and  $\Sigma$  without imposing additional assumptions on the covariate process and intensity function of the observation times. Estimating equation based method used in Cao et al. (2015) is not applicable as  $A$  cannot be estimated with available data. Therefore, it is desirable to have a feasible computation approach to approximate the asymptotic variance of  $\hat{\beta}$ . In this section, we propose to use a resampling scheme similar to that in Lu and Zhang (2010) to approximate the asymptotic distribution of  $\hat{\beta}$ .

The resampling algorithm proceeds as follows. First, we generate  $n$  iid exponential random variables  $\{\xi_i, i = 1, \dots, n\}$  with mean 1 and variance 1. Fixing data at their observed values, we solve the following  $\xi_i$ -weighted estimating equations and denote the solutions as  $\beta^*$  and  $\alpha^*(s)$ :

$$n^{-1} \sum_{i=1}^n \xi_i \int_0^1 \int_0^1 K_h(t-s) X_i(s) \left[ Y_i(t) - \alpha(s) - X_i(s)^T \beta \right] dN_i(t, s) = 0,$$

$$\begin{aligned}
& n^{-1} \sum_{i=1}^n \xi_i \iint K_{h_1, h_2}(s-t_0, t-t_0) \{Y_i(t) - \gamma_0(t_0) - \gamma_1(t_0)(s-t_0) \\
& \quad - \beta^T X_i(s)\} dN_i(t, s) = 0, \\
& \sum_{i=1}^n \xi_i \iint (s-t_0) K_{h_1, h_2}(s-t_0, t-t_0) \{Y_i(t) - \gamma_0(t_0) - \gamma_1(t_0)(s-t_0) \\
& \quad - \beta^T X_i(s)\} dN_i(t, s) = 0.
\end{aligned}$$

The estimates  $\beta^*$  and  $\alpha^*(s)$  can be obtained using the same iterative algorithm proposed in Section 2.1. In the following theorem, we establish the validity of the proposed resampling method.

**Theorem 2.3.** *Under the regularity conditions (A1)-(A5), the conditional distribution of  $(nh)^{1/2}(\beta^* - \hat{\beta})$  given the observed data converges almost surely to the limiting distribution in (2.5).*

Based on Theorem 3, by repeatedly generating  $\{\xi_1, \dots, \xi_n\}$  many times, we may obtain a large number of realizations of  $\beta^*$ . The variance estimate of  $\hat{\beta}$  can be obtained by referring to the empirical distribution of  $\beta^*$ . Point-wise inference on  $\hat{\gamma}_0(t; \hat{\beta})$  can be similarly obtained by the resampling method and we omit the details here. The proposed Monte Carlo scheme is similar to the bootstrap in terms of sample space, convergence, etc. Specifically, the sample space for  $\beta^*$  is conditional on the data whereas that of  $\hat{\beta}$  is unconditional; the distribution of  $(nh)^{1/2}(\beta^* - \beta)$  converges to the same limit as that of  $(nh)^{1/2}(\hat{\beta} - \beta)$  for almost all realizations of the data. Bootstrap would be a possible alternative to the proposed approach. However, it is not clear how to justify that the bootstrap is valid for the present problem.

Our method depends on the selection of bandwidth. To estimate  $\beta_0$ , theoretically speaking, condition (A5) says that the bandwidth cannot be too small; otherwise, the variance will be quite large. On the other hand, to eliminate the asymptotic bias, one requires a small bandwidth. Theorem 2.1 indicates that the allowable range of bandwidth is  $(n^{-1}, n^{-1/5})$ . Our numerical studies show that the proposed method has robust performance in the range  $(n^{-0.7}, n^{-0.6})$ . In the estimation of  $\alpha_0(s)$ , if we let  $h_1 = h_2 = h_\alpha$ , based on condition (A5\*), a valid bandwidth is in the range  $(n^{-1/2}, n^{-1/6})$ . Our numerical studies show that bandwidths in the range  $(n^{-0.4}, n^{-0.3})$  have decent performance. In general, asynchronous estimators converge more slowly and are less efficient than those based on synchronous data.

An automatic bandwidth selection procedure for  $\hat{\beta}$  can be carried out by calculating the bias and variability of  $\hat{\beta}$  separately. To be specific, as the bias of  $\hat{\beta}$  is of order  $h^2$ , we first regress  $\hat{\beta}(h)$  on  $h^2$  in a reasonable range of  $h$  to obtain the slope estimate  $\hat{C}$ . To obtain the variance, we split the data randomly into two parts and obtain regression coefficient estimates  $\hat{\beta}_1(h)$  and  $\hat{\beta}_2(h)$  based on each half sample. The variance of  $\hat{\beta}(h)$  is estimated by  $\hat{V}(h) = (\hat{\beta}_1(h) - \hat{\beta}_2(h))^2/4$ . We thus calculate the mean squared error as  $\hat{C}^2 h^4 + \hat{V}(h)$  and select the bandwidth  $h$  minimizing this mean squared error.



### 3. Numerical studies

#### 3.1. Simulation

In this section, we examine the finite sample performance of the proposed estimators. We first study the performance of the parametric regression coefficient  $\hat{\beta}$ . We generated 1,000 datasets, each consisting of  $n = 200, 900$  or  $1,600$  subjects. The numbers of observation times for the response  $Y(t)$  and covariate  $X(t)$  were Poisson distributed with intensity rate 6. With these two numbers of measurements, the observation times for the response and covariate were generated from uniform distribution  $\mathcal{U}(0, 1)$  independently. The covariate process was Gaussian, with values at fixed time points being multivariate normal with mean 0, variance 1 and correlation  $e^{-|t_{ij}-t_{ik}|}$ , where  $t_{ij}$  was the  $j$ th measurement time and  $t_{ik}$  was the  $k$ th measurement time for the covariates, both on subject  $i$ . At the data generating stage, in order to generate the response, we included the response observation times with the covariate observation times when generating the covariates needed for simulating response at the response observation times. The responses were generated from

$$Y(t) = \alpha(t) + X(t)\beta + \epsilon(t), \quad (3.1)$$

where  $\beta$  is the regression coefficient,  $\epsilon(t)$  is Gaussian, with mean 0, variance 1, and  $\text{cov}\{\epsilon(s), \epsilon(t)\} = 2^{-|t-s|}$ , and  $\alpha(t)$  is the nonparametric baseline function. Once the response was generated, we removed the covariate measurements at the response times from the observed covariate values. In this simulation, we used the Epanechnikov kernel  $K(t) = 0.75(1-t^2)_+$ ,  $\beta = -2$ ,  $\alpha(t) = \sin(2\pi t)$ ,  $\sqrt{t}$  and  $0.4t + 0.5$  and the bandwidth for estimating  $\alpha(\cdot)$  was set to be  $h_1 = h_2 = h_\alpha = n^{-0.4}$ . Our goal was to assess the performance of  $\hat{\beta}$ . The results were very similar for other choices of kernel functions,  $\beta$ s, functional forms of  $\alpha(\cdot)$  and  $h_\alpha$ . For comparison, we implemented last value carried forward approach to synchronize covariates and response measurements and used method in Fan and Li (2004). Specifically, for a response observed at time  $t_{ij}$ , the covariate at time  $t_{ij}$  was taken to be the covariate observed at time  $s = \max(x \leq t_{ij}, x \in \{s_{i1}, \dots, s_{iM_i}\})$ . This corresponds to the most recent observation time relative to the response. For a response, if no covariate was observed prior to the response's observation time, the observed response was omitted from the analysis.

Table 1 summarized the simulation results. We can see that as sample size increases, biases and standard deviations decrease, across different functional forms of  $\alpha(\cdot)$ . The standard deviation estimate is close to the empirical one and coverage probabilities are close to the nominal one. This is evident even for moderately large sample size. In contrast, last value carried forward approach exhibits large bias, which does not attenuate with increased sample size. On the other hand, with last value carried forward approach, as sample size increases, variance estimate gets smaller, and consequently, the coverage probability deteriorates.

We next examine the performance of  $\hat{\alpha}(t)$ . Local linear regression is used to estimate the baseline function. Simulation set up is exactly the same as before.

TABLE 1  
1,000 simulation results for inference of  $\beta$

BD	$\alpha(t) = \sin(2\pi t)$				$\alpha(t) = \sqrt{t}$				$\alpha(t) = 0.4t + 0.5$			
	Bias	SD	SE	CP	Bias	SD	SE	CP	Bias	SD	SE	CP
<i>n</i> = 200												
$n^{-0.6}$	0.030	0.081	0.078	92	0.027	0.083	0.079	92	0.031	0.085	0.080	92
$n^{-0.7}$	0.021	0.091	0.086	93	0.014	0.094	0.087	92	0.016	0.088	0.086	94
auto	0.017	0.089	0.083	92	0.023	0.086	0.084	93	0.019	0.090	0.084	93
LVCF	0.239	0.082	0.079	15	0.247	0.081	0.080	14	0.242	0.087	0.081	17
<i>n</i> = 900												
$n^{-0.6}$	0.014	0.048	0.046	93	0.013	0.046	0.046	93	0.013	0.046	0.046	94
$n^{-0.7}$	0.008	0.060	0.056	94	0.007	0.058	0.056	93	0.005	0.060	0.056	93
auto	0.005	0.055	0.052	93	0.006	0.055	0.052	93	0.007	0.055	0.052	95
LVCF	0.234	0.037	0.038	0	0.241	0.040	0.038	0	0.235	0.037	0.038	0
<i>n</i> = 1,600												
$n^{-0.6}$	0.010	0.037	0.038	95	0.010	0.039	0.038	94	0.008	0.038	0.038	95
$n^{-0.7}$	0.001	0.049	0.048	94	0.005	0.050	0.048	93	0.005	0.052	0.049	93
auto	0.004	0.046	0.043	94	0.004	0.045	0.044	95	0.004	0.048	0.044	94
LVCF	0.238	0.029	0.029	0	0.232	0.029	0.029	0	0.237	0.029	0.029	0

Note: “BD” represents different bandwidths in the estimation of  $\beta$ , “Bias” is the empirical bias, “SD” is the sample standard deviation, “SE” is the average of the standard error estimates, “CP”/100 represents the coverage probability of the 95% confidence interval for  $\hat{\beta}$ , “auto” represents the results by automatic bandwidth selection procedure, and “LVCF” represents last value carried forward approach.

We assess the performance of  $\hat{\alpha}(\cdot)$  by the square root of average squared errors (RASE),

$$\text{RASE}^2 = n_{grid}^{-1} \sum_{k=1}^{n_{grid}} \{\hat{\alpha}(t_k) - \alpha(t_k)\}^2,$$

where  $\{t_k, k = 1, \dots, n_{grid}\}$  are the grid points at which the baseline function  $\alpha(\cdot)$  is estimated. In our simulation  $n_{grid} = 100$  with  $t_k, k = 1, \dots, n_{grid}$  equally spaced between  $(0, 1)$ . The bandwidth for estimating  $\beta$  is set to be  $h = n^{-0.6}$ . The RASEs for different functional forms of  $\alpha(\cdot)$  and bandwidths based on 100 replications are listed in Table 2. The results are very typical. RASE gets smaller with smaller bandwidths. The performance improves with increased sample sizes. Figure 1 depicts the typical estimated curves of  $\alpha(t)$  along with its point-wise confidence interval.

The standard error of  $\alpha(\cdot)$  is calculated through bootstrap (Efron and Tibshirani, 1993). Specifically, we draw 100 independent bootstrap samples from the generated data and compute  $\alpha(t)$  based on the bootstrap, denoted as  $\hat{\alpha}^{*(b)}(t)$ ,  $b = 1, \dots, 100$ . The standard error estimate of  $\alpha(t)$  is obtained through

$$\hat{\text{se}}\{\hat{\alpha}(t)\} = \left[ \frac{1}{99} \sum_{b=1}^{100} \{\hat{\alpha}^{*(b)}(t) - \hat{\alpha}^{*(\cdot)}(t)\}^2 \right]^{1/2},$$

TABLE 2  
100 simulation results for inference of  $\hat{\alpha}(\cdot)$  based on RASE

n	BD	$\alpha(t) = \sin(2\pi t)$		$\alpha(t) = \sqrt{t}$		$\alpha(t) = 0.4t + 0.5$	
		MEAN	SD	MEAN	SD	MEAN	SD
200	$n^{-0.5}$	0.206	0.044	0.192	0.041	0.191	0.041
	$n^{-0.4}$	0.168	0.046	0.154	0.040	0.150	0.039
900	$n^{-0.5}$	0.152	0.020	0.151	0.024	0.153	0.028
	$n^{-0.4}$	0.103	0.023	0.092	0.018	0.102	0.023
1600	$n^{-0.5}$	0.141	0.015	0.142	0.021	0.145	0.020
	$n^{-0.4}$	0.089	0.018	0.087	0.013	0.084	0.015

Note: “BD” represents different bandwidths in the estimation of  $\alpha(\cdot)$ , “MEAN” represents empirical average of RASE, and “SD” represents empirical standard deviation of RASE.

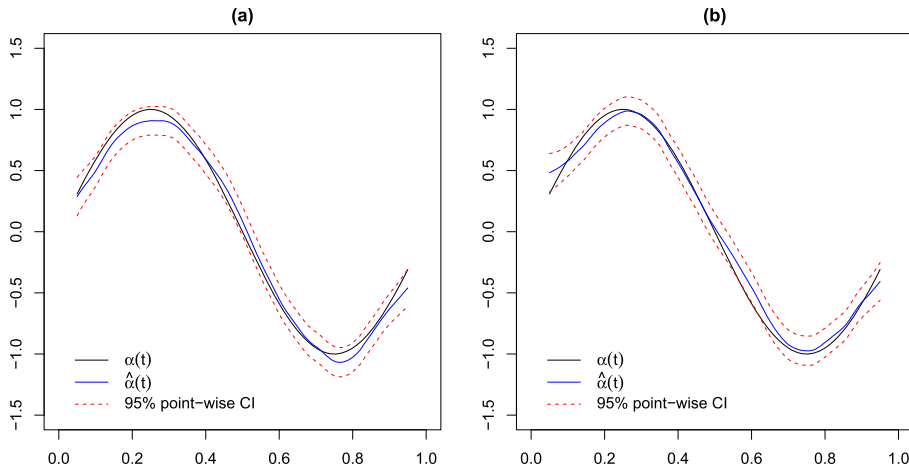


FIG 1. Typical estimated baseline curve and point wise confidence interval for  $\alpha(t) = \sin(2\pi t)$ . (a) corresponds to  $n = 900$  and (b) corresponds to  $n = 1600$ .

where

$$\hat{\alpha}^{*(\cdot)}(t) = \frac{1}{100} \sum_{b=1}^{100} \hat{\alpha}^{*(b)}(t).$$

We look at bias, variance and pointwise coverage probabilities at  $t = 0.25, 0.5$  and  $0.75$  with different functional forms of  $\alpha(\cdot)$  and bandwidth  $h_\alpha$ . Bandwidth for the estimation of  $\beta$  is set to be  $n^{-0.6}$ .

The results are summarized in Table 3. We can see that the biases are small, standard error is close to the sample standard deviation and coverage probabilities are close to the nominal 95%. This pattern is consistent across different time points, functional form of  $\alpha_0(t)$  and sample sizes. As sample size increases, both bias and variability attenuate. Therefore, we can conduct valid point-wise inference for the function  $\alpha(t)$ .

TABLE 3  
1,000 simulation results for inference of  $\alpha_0(t)$

BD	Bias	$t = 0.25$			$t = 0.5$			$t = 0.75$				
		SD	SE	CP	Bias	SD	SE	CP	Bias	SD	SE	CP
$\alpha_0(t) = \sin(2\pi t)$												
$n = 200$												
$n^{-0.4}$	-0.053	0.135	0.131	92	0.004	0.141	0.134	93	0.052	0.134	0.131	92
$n^{-0.5}$	-0.020	0.175	0.177	95	-0.005	0.185	0.182	95	0.007	0.178	0.178	95
$n = 900$												
$n^{-0.4}$	-0.018	0.086	0.086	96	-0.002	0.086	0.087	96	0.012	0.092	0.086	96
$n^{-0.5}$	-0.006	0.142	0.138	94	-0.003	0.144	0.140	95	0.013	0.139	0.139	95
$n = 1600$												
$n^{-0.4}$	-0.014	0.076	0.075	95	-0.001	0.076	0.075	95	0.011	0.075	0.075	95
$n^{-0.5}$	-0.005	0.128	0.129	95	-0.003	0.129	0.128	95	-0.003	0.132	0.128	95
$\alpha_0(t) = \sqrt{t}$												
$n = 200$												
$n^{-0.4}$	-0.013	0.142	0.134	94	-0.007	0.133	0.133	95	-0.005	0.133	0.133	95
$n^{-0.5}$	0.004	0.179	0.168	93	0.010	0.170	0.169	95	0.010	0.172	0.168	94
$n = 900$												
$n^{-0.4}$	0.001	0.086	0.087	95	-0.001	0.090	0.087	94	0.001	0.091	0.087	94
$n^{-0.5}$	-0.003	0.139	0.136	95	-0.005	0.134	0.136	95	0.005	0.138	0.138	95
$n = 1600$												
$n^{-0.4}$	0.001	0.075	0.074	94	0.004	0.076	0.075	94	-0.002	0.075	0.075	95
$n^{-0.5}$	0.003	0.128	0.127	94	-0.002	0.129	0.126	95	-0.002	0.134	0.130	95
$\alpha_0(t) = 0.4t + 0.5$												
$n = 200$												
$n^{-0.4}$	0.001	0.134	0.131	95	0.003	0.135	0.132	94	0.003	0.134	0.132	94
$n^{-0.5}$	0.004	0.173	0.167	94	0.003	0.171	0.171	95	-0.002	0.176	0.170	94
$n = 900$												
$n^{-0.4}$	0.001	0.091	0.087	94	0.001	0.086	0.087	95	0.004	0.088	0.086	95
$n^{-0.5}$	0.005	0.138	0.138	95	-0.006	0.136	0.138	96	-0.004	0.137	0.138	95
$n = 1600$												
$n^{-0.4}$	-0.002	0.075	0.075	95	-0.001	0.077	0.075	94	-0.001	0.074	0.075	95
$n^{-0.5}$	-0.002	0.134	0.130	95	-0.005	0.132	0.130	95	-0.008	0.129	0.128	95

Note: "BD" represents different bandwidths in the estimation of  $\alpha(\cdot)$ , "Bias" is the empirical bias, "SD" is the sample standard deviation, "SE" is the average of the standard error estimates and "CP"/100 represents the coverage probability of the 95% confidence interval for  $\hat{\beta}$ .

### 3.2. An application to HIV dataset

In this section, we apply our method to an HIV dataset. A total of 191 HIV patients were followed from July 1997 to September 2002 in a University hospital. Details of the study design, methods and medical implications are given in Wohl et al. (2005). During this study, all patients were scheduled to have their measurements taken during the semi-annual visits, with HIV viral load and CD4 counts obtained separately at different labs. Because many patients did not visit the clinic at the scheduled time or even missed those scheduled visits, and the HIV infection occurred randomly during the study, there are

TABLE 4  
Summary statistics for  $\hat{\beta}$  based on model (1.1)

$h(n^{-\gamma})$	$104(n^{-0.5})$	$62(n^{-0.6})$	62(auto)	LVCF
$\hat{\beta}$	-1.050	-0.906	-0.906	-0.828
SE( $\hat{\beta}$ )	0.287	0.433	0.433	0.199
z-value	-3.659	-2.092	-2.092	-4.160

unequal numbers of repeated measurement on HIV viral load and CD4 counts and the measurement times for these two variables are different.

Under missing at random assumption (Little and Rubin, 2002), we eliminate patients who do not have any records on either HIV or CD4 counts and use 181 subjects in our analysis. We take log transformed CD4 counts as covariate and log transformed HIV viral load as response. The bandwidth for nonparametric estimate of  $\alpha(t)$  is set at  $h_\alpha = 2(Q_3 - Q_1)n^{-0.35} = 226$ , where  $Q_3$  is the 0.75 quantile and  $Q_1$  is the 0.25 quantile of the pooled sample of measurement times for the covariate and response and  $n$  is the number of patients. This adjusts the range of observation times to be compatible with the simulations studies. To estimate  $\beta$ , we use bandwidths  $h = 2(Q_3 - Q_1)n^{-\gamma}$ , where  $\gamma = 0.5$  or  $0.6$ . Results based on last value carried forward implemented using methods proposed in Fan and Li (2004) are presented as a comparison. From Table 4, we can see the statistically significant negative relationship between CD4 counts and HIV viral load through the proposed method. This negative association is consistent with medical literature (Phillips et al., 2001). LVCF also produced a statistically significant negative association between CD4 counts and HIV viral load. Figure 2 depicts the estimated baseline function  $\alpha(t)$  using data adaptive bandwidth  $h = 62$  in the estimation of  $\beta$ . It also plots the estimated baseline function

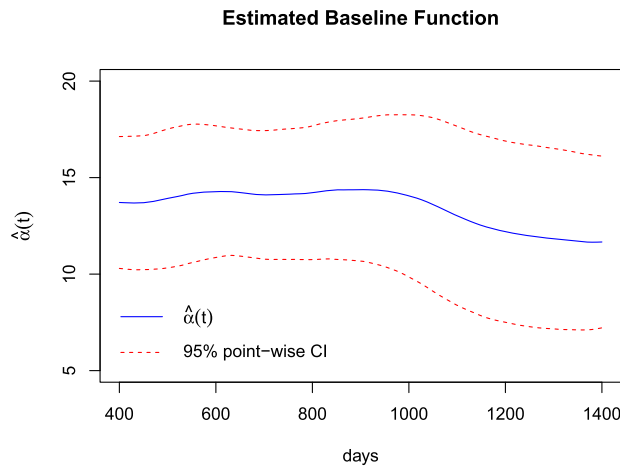


FIG 2. Estimated Baseline Function. The solid line represents the estimated baseline function; the dashed lines, the estimated baseline function plus/minus 1.96 standard errors.

plus/minus 1.96 standard errors, which can serve as a point-wise confidence interval ignoring the bias of the nonparametric fit. From Figure 2, we can see that there is no significant time trend effect within the first 3 years. However, there is a significant decreasing trend starting from around 3 years until the end of the study. Nonparametric estimation of the baseline function removes such time trend to avoid biased estimates of the regression coefficient  $\beta$  for the association between CD4 counts and HIV viral load.

#### 4. Concluding remarks

In this paper, we propose a semiparametric partially linear model for asynchronous longitudinal data analysis. This allows the exploration of curvature and nonlinear time trend effect of longitudinal response. Local polynomial is used to approximate the nonparametric function. Our analysis is based on a set of global and local estimating equations with an iterative algorithm that has good convergency properties. The rate of convergence for the parametric part remains the same as in the parametric model with asynchronous longitudinal data but the variance is larger due to the estimation of nonparametric baseline function. For asynchronous longitudinal data, longitudinal covariate and response are mismatched even for the same subject. Essentially for each longitudinal response, kernel smoothing approach only uses covariate in its neighborhood controlled by bandwidth  $h$ . Consequently, we have  $(nh)^{1/2}$  rate of convergence. It is not clear whether  $n^{1/2}$  rate of convergence could be achieved under stronger assumptions.

Numerical studies support our theoretical predictions and show the newly proposed method yield valid inference in contrast to the ad-hoc last value carried forward approach. On the other hand, we can simply use a weighted last value carried forward method to obtain valid inference. The main idea is that the further the last observed longitudinal covariate is from the current longitudinal response, the less it should contribute to the estimating equation. This can be handled formally by weighting the last observation as a decreasing function of the time between the most recently observed longitudinal covariate and current longitudinal response.

Our method is based on the working independence assumption similar to GEE (Liang and Zeger, 1986). Both GEE with synchronous data and our proposed analytical strategy for asynchronous longitudinal data are valid when the data are missing at random (Little and Rubin, 2002). In GEE, with time-dependent covariates, Pepe and Anderson (1994) showed that parameter estimates are generally biased unless (i) the mean for the response at time  $t$  given all past, present, and future covariate values is equal to that given the covariate values observed at  $t$  or unless (ii) independence estimating equations are used. The condition (i) is a strong assumption. When data are missing at random, (i) cannot be verified with the observed data and (ii) is a conservative approach which ensures valid inference using complete data observations regardless whether (i) holds. When (ii) is adopted, it is challenging to improve efficiency, since the correlation structure in the data cannot be exploited in the working covariance matrix.

Similar issues arise in our asynchronous longitudinal data set-up, with further work warranted to understand the extent to which valid estimation may be achieved with non-diagonal working covariance matrices and whether efficiency gains might be achievable.

In some applications, both time-independent and time-dependent covariates exist. Cao et al. (2015) requires that the observation times of time-dependent covariates are the same and treat time-independent covariates the same as time-dependent covariates. It is of great interest to explore whether improved rate of convergence could be obtained for the time-independent covariate in such mixed covariate models. To test whether  $\alpha(t)$  is really time-varying, a simultaneous confidence band will be needed. Cao et al. (2017) studied the construction of smooth simultaneous confidence band for varying coefficient models with sparse longitudinal data. Simultaneous inference for the nonparametric function  $\alpha(t)$  for asynchronous longitudinal data remains open and warrant additional development.

## Appendix A: Appendix: Proofs of Theorems

### A.1. Proof of Theorem 1

We first show the local consistency of  $\hat{\beta}$ . For any  $x$ , define

$$\begin{aligned} & nU_2(\gamma_0, \gamma_1, \beta)(x) \\ &= \left( \begin{array}{c} \sum_{i=1}^n \iint K_{h_1, h_2}(s-x, t-x) \{Y_i(t) - \gamma_0(x) - \gamma_1(x)(s-x) - \beta^T X_i(s)\} dN_i(t, s) \\ \sum_{i=1}^n \iint \frac{s-x}{h_1} K_{h_1, h_2}(s-x, t-x) \{Y_i(t) - \gamma_0(x) - \gamma_1(x)(s-x) - \beta^T X_i(s)\} dN_i(t, s) \end{array} \right) \quad (\text{A.1}) \end{aligned}$$

By Taylor expansion, we have

$$\begin{aligned} U_2(\hat{\gamma}_0, \hat{\gamma}_1, \beta)(x) &= U_2(\alpha, \dot{\alpha}, \beta)(x) - E_3(x) \begin{pmatrix} \hat{\gamma}_0(x) - \alpha(x) \\ h_1 \{\hat{\gamma}_1(x) - \dot{\alpha}(x)\} \end{pmatrix} \quad (\text{A.2}) \\ &+ o_p \left\{ \sqrt{(\hat{\gamma}_0(x) - \alpha(x))^2 + (\hat{\gamma}_1(x) - \dot{\alpha}(x))^2} \right\}. \end{aligned}$$

where

$$E_3(x) = \frac{1}{n} \sum_{i=1}^n \iint K_{h_1, h_2}(s-x, t-x) \begin{pmatrix} 1 \\ \frac{s-x}{h_1} \end{pmatrix} (1 \quad \frac{s-x}{h_1}) dN_i(t, s).$$

Combining all the equations above, we have

$$\begin{aligned} & E_3(x) \begin{pmatrix} \hat{\gamma}_0(x) - \alpha(x) \\ h_1 \{\hat{\gamma}_1(x) - \dot{\alpha}(x)\} \end{pmatrix} + o_p \left\{ \sqrt{(\hat{\gamma}_0(x) - \alpha(x))^2 + (\hat{\gamma}_1(x) - \dot{\alpha}(x))^2} \right\} \\ &= U_2(\alpha, \dot{\alpha}, \beta)(x). \end{aligned}$$

By standard nonparametric techniques and the law of large numbers, we can show that  $E_3(x)$  converges to a deterministic function

$$e_3(x) = \lambda(x, x) \begin{pmatrix} 1 & \iint z_1 K(z_1, z_2) dz_1 dz_2 \\ \iint z_1 K(z_1, z_2) dz_1 dz_2 & \iint z_1^2 K(z_1, z_2) dz_1 dz_2 \end{pmatrix}$$

$$= \lambda(x, x) \begin{pmatrix} 1 & 0 \\ 0 & k_2 \end{pmatrix},$$

where  $k_2 = \iint z_1^2 K(z_1, z_2) dz_1 dz_2$  and we take into account (A4). Therefore

$$\begin{aligned} \begin{pmatrix} \hat{\gamma}_0(x) - \alpha(x) \\ h_1\{\hat{\gamma}_1(x) - \dot{\alpha}(x)\} \end{pmatrix} &= e_3(x)^{-1} U_2(\alpha, \dot{\alpha}, \beta)(x) \\ &= \frac{1}{\lambda(x, x)} \frac{1}{k_2} \begin{pmatrix} k_2 & 0 \\ 0 & 1 \end{pmatrix} U_2(\alpha, \dot{\alpha}, \beta) \end{aligned}$$

Denote

$$U_2(\alpha, \dot{\alpha}, \beta)(x) = \{U_{21}(\alpha, \dot{\alpha}, \beta)(x), U_{22}(\alpha, \dot{\alpha}, \beta)(x)\}^T.$$

We have

$$\begin{aligned} \hat{\gamma}_0(x) - \alpha(x) &= \frac{1}{\lambda(x, x)} \frac{1}{k_2} \{k_2 U_{21}(\alpha, \dot{\alpha}, \beta)(x)\} \\ &= \frac{1}{\lambda(x, x)} U_{21}(\alpha, \dot{\alpha}, \beta)(x). \end{aligned} \quad (\text{A.3})$$

Note that the estimating equation for  $\beta$  is

$$U(\beta) = \frac{1}{n} \sum_{i=1}^n \int_0^1 \int_0^1 K_h(t-s) X_i(s) [Y_i(t) - \hat{\gamma}_0(s) - X_i(s)^T \beta] dN_i(t, s) = 0. \quad (\text{A.4})$$

Plugging (A.3) into (A.4), we have

$$\begin{aligned} U(\beta) &= \frac{1}{n} \sum_{i=1}^n \iint X_i(s) \{Y_i(t) - \alpha(s) - \beta^T X_i(s)\} K_h(t-s) dN_i(t, s) \\ &\quad - \frac{1}{n} \sum_{i=1}^n \iint X_i(s) \frac{1}{\lambda(s, s)} U_{21}(\alpha, \dot{\alpha}, \beta)(s) K_h(t-s) dN_i(t, s). \end{aligned} \quad (\text{A.5})$$

Note that

$$\alpha(s) = E\{Y(s) - \beta^T X(s)\} = \alpha_0(s) + E\{X(s)^T\}(\beta_0 - \beta) \quad (\text{A.6})$$

and

$$U_{21}(\alpha, \dot{\alpha}, \beta)(s) = n^{-1} \sum_{i=1}^n \iint K_h(t-s) \{Y_i(t) - \alpha(s) - X_i(s)^T \beta\} dN_i(t, s). \quad (\text{A.7})$$

For (A.7), we consider the class of functions

$$\left\{ \iint K_h(t-s) \{Y(t) - \alpha(s) - X(s)^T \beta\} dN(t, s) : |\beta - \beta_0| < \epsilon_1, |\alpha(s) - \alpha_0(s)| < \epsilon_2 \right\}$$

for given constants  $\epsilon_1$  and  $\epsilon_2$ . It can be shown that this class is a P-Donsker class (van der Vaart and Wellner, 1996). As a result, we obtain that for  $|\beta - \beta_0| < \epsilon_1, |\alpha(s) - \alpha_0(s)| < \epsilon_2$ ,

$$U_{21}(\alpha, \dot{\alpha}, \beta)(s) = U_{21}(\alpha_0, \dot{\alpha}_0, \beta_0)(s) + o_p(1) = o_p(1). \quad (\text{A.8})$$



Consequently,

$$U(\beta) = \frac{1}{n} \sum_{i=1}^n \iint X_i(s) \left[ Y_i(t) - \alpha_0(s) - E\{X_i(s)^T\}(\beta_0 - \beta) - X_i(s)^T \beta \right] K_h(t-s) dN_i(t, s) + o_p(1). \quad (\text{A.9})$$

Denote

$$\tilde{U}(\beta) = \frac{1}{n} \sum_{i=1}^n \iint X_i(s) \left[ Y_i(t) - \alpha_0(s) - E\{X_i(s)^T\}(\beta_0 - \beta) - X_i(s)^T \beta \right] K_h(t-s) dN_i(t, s) \quad (\text{A.10})$$

We have

$$\tilde{U}(\beta) \rightarrow u(\beta) \quad \text{as } n \rightarrow \infty, \quad (\text{A.11})$$

where

$$u(\beta) = \int \text{var}\{X(s)\} \lambda(s, s) ds (\beta_0 - \beta).$$

Under assumption (A1) and (A2),  $\beta_0$  is the unique solution to  $u(\beta)$ . As  $\hat{\beta}$  solves the estimating equation  $U(\beta) = 0$ , it follows from the convexity lemma (Ander- sen and Gill, 1982) that  $\hat{\beta} \rightarrow \beta_0$  in probability.

We next show the asymptotic normality of  $\hat{\beta}$ . The key idea is to establish the following relationship

$$\begin{aligned} & \sup_{|\beta - \beta_0| < M(nh)^{-1/2}} \left| (nh)^{1/2} \tilde{U}(\beta) - (nh)^{1/2} [\tilde{U}(\beta_0) - E\{\tilde{U}(\beta_0)\}] \right. \\ & \quad \left. - (nh)^{1/2} A(\beta - \beta_0) \right| \\ & = Dn^{1/2} h^{5/2} + o_p(n^{1/2} h^{5/2}) + o_p(1), \end{aligned} \quad (\text{A.12})$$

where  $A$  is given in Theorem 1 and  $D$  is a constant.

To obtain (A.12), first, using  $\mathcal{P}_n$  and  $\mathcal{P}$  to denote the empirical measure and the true probability measure respectively, we obtain

$$\begin{aligned} & (nh)^{1/2} \tilde{U}(\beta) \\ & = (nh)^{1/2} (\mathcal{P}_n - \mathcal{P}) \iint K_h(t-s) X(s) \left[ Y(t) - \alpha_0(s) - E\{X(s)^T\}(\beta_0 - \beta) \right. \\ & \quad \left. - X(s)^T \beta \right] dN(t, s) \\ & \quad + (nh)^{1/2} E \iint K_h(t-s) X(s) \left[ Y(t) - \alpha_0(s) - E\{X(s)^T\}(\beta_0 - \beta) \right. \\ & \quad \left. - X(s)^T \beta \right] dN(t, s) \\ & = I + II. \end{aligned} \quad (\text{A.13})$$

For the second term on the right hand side of (A.13), we have

$$II = -(nh)^{1/2} A(\beta - \beta_0) + Cn^{1/2} h^{5/2} + o_p(n^{1/2} h^{5/2}). \quad (\text{A.14})$$

From (A2),  $A$  is a positive-definite matrix, thus non-singular. For the first term on the right hand side of (A.13), we consider the class of functions

$$\left\{ h^{1/2} \iint X(s) \left[ Y(t) - \alpha_0(s) - E\{X(s)^T\}(\beta_0 - \beta) - X(s)^T \beta \right] K_h(t-s) dN(t,s) : |\beta - \beta_0| < \epsilon \right\}$$

for a given constant  $\epsilon$ . It can be shown that this class is a P-Donsker class (van der Vaart and Wellner, 1996). As a result, we obtain that the first term on the right hand side of (A.13) for  $|\beta - \beta_0| < M(nh)^{-1/2}$  is equal to

$$\begin{aligned} & (nh)^{1/2} (\mathcal{P}_n - \mathcal{P}) \iint K_h(t-s) X(s) \{Y(t) - \alpha_0(s) - X(s)^T \beta_0\} dN(t,s) + o_p(1) \\ &= (nh)^{1/2} \left[ \tilde{U}(\beta_0) - E\{\tilde{U}(\beta_0)\} \right] + o_p(1). \end{aligned} \quad (\text{A.15})$$

Combing (A.15) and (A.14) and by condition (A4), we obtain (A.12). Consequently,

$$(nh)^{1/2} A(\hat{\beta} - \beta_0) + Cn^{1/2}h^{5/2} + o_p(n^{1/2}h^{5/2}) = (nh)^{1/2} [\tilde{U}(\beta_0) - E\{\tilde{U}(\beta_0)\}]. \quad (\text{A.16})$$

On the other hand, following similar arguments as that in Cao et al. (2015), we can calculate

$$\begin{aligned} \Sigma &= \text{var} \left( h^{1/2} \iint X(s) \{Y(t) - \alpha_0(s) - X(s)^T \beta_0\} K_h(t-s) dN(t,s) \right) \\ &= \int K(z)^2 dz \int E[X(s)X(s)^T \sigma\{s, X(s)\}^2] \lambda(s, s) ds. \end{aligned}$$

To prove the asymptotic normality, we verify the Lyapunov condition. Define

$$\psi_i = (nh)^{1/2} n^{-1} \iint K_h(t-s) X_i(s) \{Y_i(t) - \alpha_0(s) - X(s)^T \beta_0\} dN_i(t,s).$$

Similar to the calculation of  $\Sigma$ ,

$$\sum_{i=1}^n E(|\psi_i - E\psi_i|^3) = nO\{(nh)^{3/2}n^{-3}h^{-2}\} = O\{(nh)^{-1/2}\}.$$

Therefore,

$$(nh)^{1/2} \left[ \tilde{U}(\beta_0) - E\{\tilde{U}(\beta_0)\} \right] \xrightarrow{d} N(0, \Sigma). \quad (\text{A.17})$$

Combining (A.9), (A.16) and (A.17), we finish the proof of Theorem 1.

## A.2. Proof of Theorem 2

From (A.3) and (A.6), we have

$$\hat{\gamma}_0(x; \hat{\beta}) - \alpha_0(x) = \frac{1}{\lambda(x, x)} U_{21}(\alpha, \hat{\alpha}, \hat{\beta})(x) - E\{X(x)^T\}(\hat{\beta} - \beta_0).$$

We now establish the limiting distribution of  $U_{21}(\alpha_0, \dot{\alpha}_0, \hat{\beta})(x)$ . Similar to (A.8), we have

$$U_{21}(\alpha, \dot{\alpha}, \hat{\beta}) = U_{21}(\alpha_0, \dot{\alpha}_0, \beta_0) + o_p(1).$$

and we have

$$\begin{aligned} & U_{21}(\alpha_0, \dot{\alpha}_0, \beta_0)(x) \\ &= n^{-1} \sum_{i=1}^n \iint K_{h_1, h_2}(s-x, t-x) \{Y_i(t) - \alpha_0(x) - \beta_0^T X_i(s)\} dN_i(t, s) \end{aligned}$$

From laws of large numbers, it has expectation 0 and the variance is

$$(nh_1h_2)^{-1} \iint K(z_1, z_2)^2 dz_1 dz_2 E[\sigma^2\{x, X(x)\}] \lambda(x, x).$$

Since  $\hat{\beta}$  has faster rate of convergence, therefore, we have

$$\sqrt{nh_1h_2} \{\hat{\gamma}_0(x; \hat{\beta}) - \alpha_0(x)\} \xrightarrow{d} N(0, \Sigma^*), \quad (\text{A.18})$$

where

$$\Sigma^* = \frac{1}{\lambda(x, x)} \iint K(z_1, z_2)^2 dz_1 dz_2 E[\sigma^2\{x, X(x)\}].$$

### A.3. Proof of Theorem 3

The proof of Theorem 3 can be similarly derived as that for Theorem 1 and hence is omitted here.

### Acknowledgement

Hongyuan Cao is partially supported by a grant from University of Missouri Research Board. We are grateful to Jason P. Fine and Donglin Zeng for stimulating discussions.

### References

- Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100–1120. [MR0673646](#)
- Cai, J., Fan, J., Jiang, J. and Zhou, H. (2007). Partially linear hazard regression for multivariate survival data. *J. Amer. Statist. Assoc.* **102**, 538–551. [MR2370851](#)
- Cai, J., Fan, J., Jiang, J. and Zhou, H. (2008). Partially linear hazard regression with varying-coefficients for multivariate survival data. *J. R. Stat. Soc. B* **70**, 141–158. [MR2412635](#)
- Cao, H., Zeng, D. and Fine, J. P. (2015). Regression analysis of sparse asynchronous longitudinal data. *J. R. Stat. Soc. B* **77**, 755–776. [MR3382596](#)

- Cao, H., Li, J. and Fine, J. P. (2016). On last observation carried forward and asynchronous longitudinal regression analysis. *Electronic Journal of Statistics* **10**, 1155–1180. [MR3499524](#)
- Cao, H., Liu, W. and Zhou, Z. (2017). Simultaneous nonparametric regression analysis of sparse longitudinal data. *Bernoulli*, invited revision submitted.
- Diggle, P., Heagerty, P., Liang, K. Y. and Zeger, S. L. (2002). *Analysis of Longitudinal Data (2nd ed.)*, Clarendon, TX: Clarendon Press. [MR2049007](#)
- Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, Chapman & Hall, New York. [MR1270903](#)
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications*. London: Chapman & Hall. [MR1383587](#)
- Fan, J. and Li, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *J. Amer. Statist. Assoc.* **99**, 710–723. [MR2090905](#)
- Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear model. *Biometrika* **73**, 13–22. [MR0836430](#)
- Lin, D. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *J. Amer. Statist. Assoc.* **96**, 103–113. [MR1952726](#)
- Lin, X. and Carroll, R. J. (2001). Semiparametric regression for clustered data using generalized estimating equations. *J. Amer. Statist. Assoc.* **96**, 1045–1056. [MR1947252](#)
- Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data (2nd ed.)*, New York: Wiley. [MR1925014](#)
- Lu, W. and Zhang, H. H. (2010). On estimation of partially linear transformation models. *J. Amer. Statist. Assoc.* **105**, 683–691. [MR2724852](#)
- Martinussen, T. and Scheike, T. (1999). A semiparametric additive regression model for longitudinal data. *Biometrika* **86**, 691–702. [MR1723787](#)
- Moyeed, R. A. and Diggle, P. J. (1994). Rates of convergence in semiparametric modeling of longitudinal data. *Austral. J. Statist.* **36**, 75–93. [MR1309507](#)
- Pepe, M. S. and Anderson, G. L. (1994) A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics - Simulation and Computation* **23**, 939–951.
- Phillips, A. N. et al. (2001). HIV viral load response to antiretroviral therapy according to the baseline CD4 cell count and viral load. *The Journal of American Medical Association* **286**, 2560–2567.
- Sentürk, D., Dalrymple, L. S., Mohammed, S. M., Kaysen, G. A. and Nguyen, D. V. (2012). Modeling time-varying effects with generalized and unsynchronized longitudinal data. *Statist. Med.* **32**, 2971–2987. [MR3073830](#)
- van der Vaart A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*. Springer, New York. [MR1385671](#)
- Wang, N., Carroll, R. J. and Lin, X. (2005). Efficient semiparametric marginal estimation for longitudinal/clustered data. *J. Amer. Statist. Assoc.* **100**, 147–157. [MR2156825](#)
- Wohl, D., Zeng, D., Stewart, P., Glomb, N., Alcorn, T., Jones, S., Handy, J., Fiscus, S., Weinberg, A., Gowda, D. and van der Horst, C. (2005). Cy-

tomegalovirus viremia, mortality and cmv end-organ disease among patients with AIDS receiving potent antiretroviral therapies. *Journal of AIDS* **38**, 538-544.

Xiong, X. and Dubin, J. A. (2010). A binning method for analyzing mixed longitudinal data measured at distinct time points. *Statist. Med.* **29**, 1919-1931. [MR2758463](#)

Zeger, S. L. and Diggle, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics* **50**, 689-699.