Electronic Journal of Statistics
Vol. 10 (2016) 1155–1180
ISSN: 1935-7524
DOI: 10.1214/16-EJS1141

On last observation carried forward and asynchronous longitudinal regression analysis

Hongyuan Cao*

Department of Statistics,
University of Missouri-Columbia,
Columbia, MO 65211, U.S.A
e-mail: caohong@missouri.edu

Jialiang Li†

Department of Statistics and Applied Probability,
National University of Singapore,
Singapore 117546
e-mail: stalj@nus.edu.sg

and

Jason P. Fine‡

Department of Biostatistics,
University of North Carolina at Chapel Hill,
Chapel Hill, NC 27514, U.S.A.
e-mail: jfine@email.unc.edu

Abstract: In many longitudinal studies, the covariates and response are often intermittently observed at irregular, mismatched and subject-specific times. Last observation carried forward (LOCF) is one of the most commonly used methods to deal with such data when covariates and response are observed asynchronously. However, this can lead to considerable bias. In this paper, we propose a weighted LOCF estimation using asynchronous longitudinal data for the generalized linear model. We further generalize this approach to utilize previously observed covariates in addition to the most recent observation. In comparison to earlier methods, the current methods are valid under weaker assumptions on the covariate process and allow informative observation times which may depend on response even conditional on covariates. Extensive simulation studies provide numerical support for the theoretical findings. Data from an HIV study is used to illustrate our methodology.

MSC 2010 subject classifications: Primary 60G20; secondary 60G05.
Keywords and phrases: Asynchronous longitudinal data, Kernel weighted estimation, last observation carried forward, nonparametric regression.

Received November 2015.

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

†The research of Jialiang Li was supported in part by grants ARF R-155-000-130-112 and NMRC/CBRG/0014/2012

‡The research of Jason P. Fine was supported in part by NIH grant P01 CA142538

Contents

1. Introduction

Longitudinal data arise in many scientific inquiries, such as epidemiological studies, clinical trials and educational studies, among others. In such studies, data are often collected at subject specific time points and the number of measurements varies across subjects. Last observation carried forward (LOCF) is one of the most commonly used approaches for analyzing incomplete longitudinal data. This method imputes the most recent observation as the current observation and then employs standard analyses treating the imputed covariate as the true covariate. Such analyses are problematic, see Lavori [4] and Molenberghs et al. [9]. First, it is assumed that the longitudinal measurement does not change from the time of the last measurement. Second, no distinction is made between those subjects who had a valid measurement and those subjects with imputed values, artificially increasing the amount of information in the data. These issues can induce substantial biases in parameter estimates and lead to inaccurate inferences, see Verbeke and Molenberghs [18]. To circumvent these problems, likelihood based approaches such as Verbeke and Molenberghs [18] and Cook et al. [2], inverse probability weighting such as Robins et al. [12] and Robins et al. [13] and multiple imputation such as Rubin [14] have been proposed as more principled and preferred methods for analysis. However, such methods impose stringent modeling assumptions and the inferences they produce are typically highly dependent on untestable and often implicit assumptions regarding the distribution of the unobserved measurements given the observed measurements.

In this paper, our focus is regression with so-called asynchronous longitudinal data as in Cao et al. [1], where the measurement times for a longitudinal response and a longitudinal covariate are mismatched. We propose an intuitively appealing weighting approach which retains the simplicity of LOCF imputation for the current value of the covariate, enabling the use of methods for synchronous data where response and covariate are measured at the same time points. As an example, in a dialysis study of end-stage renal disease patients, infection-related hospitalization status and serum C-reactive protein are obtained at distinct time points within each patient. In clinical epidemiology, measurements of vital signs and lab tests are often conducted at different times for the same individual. In electronic medical records, a subject's information may be pooled from different sources to make treatment decisions, creating an asynchronous longitudinal data structure. We consider the estimation of generalized linear models which relate the current value of a longitudinal outcome to the current value of a longitudinal covariate. We modify estimating equation techniques in Liang and Zeger [5] for synchronous data to obtain unbiased inferences based on LOCF imputation of the current value of the longitudinal covariate with mismatched measurement times.

While regression analysis using estimating equations for synchronous longitudinal data such as Liang and Zeger [5] has been widely studied, there has been limited work on the analysis of regression models using asynchronous longitudinal data. Xiong and Dubin [20] employed an ad hoc binning step to synchronize covariates and response measurements to use existing methods for synchronous data. Sentürk et al. [15] explicitly addressed the asynchronous setting for generalized varying coefficient model with one covariate but did not provide the theoretical properties of the estimators. Cao et al. [1] proposed a nonparametric kernel weighting approach for the generalized linear model to explicitly deal with the asynchronous structure and rigorously established the consistency and asymptotic normality of the resulting estimates. In this paper, we formalize simple LOCF in a rigorous manner using weighting techniques similar to those in Cao et al. [1]. We show that the weighted LOCF is also consistent and asymptotically normal but is valid under weaker assumptions on the covariate and observation time processes, as detailed in the sequel.

The main idea of the weighting is that the further the last observation is from the current observation, the less it should contribute to the estimating equation. This is handled formally by weighting the last observation as a decreasing function of the time between the observed and missing measurement occasions. We show that this may be generalized using the half kernel weighting to utilize all previously observed covariates in addition to the most recent observation. In contrast to full kernel method in Cao et al. [1] which uses both previous and future covariate measurements, the proposed estimators are valid under weaker assumptions on the covariate processes and allow observation times to depend on the response even conditionally on covariates. We also relax the conditions for the validity of the full kernel methods, permitting covariate processes with independent increments and allowing observation times to depend on covariates but not responses, similarly to the conditional independence approach to longitudinal data in Lin and Ying [7], Lin et al. [6] and Sun et al. [16]. Interestingly, under independent increments, the rate of convergence of the estimators differs from that without independent increments.

The paper is organized as follows. In section 2, we recap results from Cao et al. [1], discuss the proposed weighted LOCF and half kernel estimators and corresponding theoretical findings. Section 3 reports simulation studies that compare the proposed methods with Cao et al. [1]. The new methods demonstrate improved performance in situations where the assumptions from the previously proposed estimator are violated, particularly with informative observation times. Interestingly, there is little loss of information in weighted LOCF versus using all previously observed covariate values in the estimating equations. Application to an HIV dataset illustrates the practical utility of the methods. Concluding remarks are given in Section 4. Proofs of results from Section 2 are relegated in the Appendix.

2. Main results

2.1. Full kernel estimation

This subsections presents the main results from [1]. We consider the generalized linear model:

$$E\{Y(t)|X(t)\} = g\{X(t)^T\beta\}, \tag{2.1}$$

where g is a known, strictly increasing and continuously twice-differentiable function, t is a univariate time index, $X(t)$ is a vector of time-varying covariates plus intercept term, $Y(t)$ is a time-varying response and β is an unknown time-invariant regression parameter. For subject $i = 1, \ldots, n$, the observation times of the longitudinal covariate process $X_i(t)$ and response process $Y_i(t)$ may be generated from a bivariate counting process like [7], where

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

counts the number of observation times up to t on the response and up to s on the covariates, where $\{T_{ij}, j = 1, \ldots, L_i\}$ are the observation times of the response and $\{S_{ik}, k = 1, \ldots, M_i\}$ are the observation times of the covariates. In order to use existing methods for synchronous longitudinal data, where $L_i = M_i$ and $T_{ij} = S_{ij}, j = 1, \ldots, L_i$, for each observed response, one may carry forward the most recently observed covariate. This ad hoc approach incurs substantial bias as shown in [1]. Furthermore, [1] proposed an estimating equation for β in (2.1)

$$U_n^f(\beta) = n^{-1}\sum_{i=1}^{n}\int_0^1\int_0^1 K_h(t-s)X_i(s)\Big[Y_i(t) - g\{X_i(s)^T\beta\}\Big]dN_i(t, s), \tag{2.2}$$

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. The response $Y_i(t)$ may be a continuous, categorical, or count variable, while the covariate $X_i(t)$ may include time-independent covariates, such as an intercept term, in addition to time-varying covariates. The main requirement for the validity of (2.2) is that if the time-varying covariates in $X_i(t)$ are multivariate, then the different covariates are measured at the same time points. The kernel weighting accounts for the fact that the covariate and response are mismatched and permits contributions to $U_n^f(\beta)$ from all possible pairings of response and covariate observations. We solve $U_n^f(\beta)$ to obtain an estimate for β, denoted $\hat{\beta}_f$.

We next present the asymptotic properties of $\hat{\beta}^f$. We specify our assumptions on the covariance structure as follows. For $s, t \in [0, \tau]$, let $\text{var}\{Y(t)|X(t)\} = \sigma\{t, X(t)\}^2$ and $cov\{Y(s), Y(t)|X(s), X(t)\} = r\{s, t, X(s), X(t)\}$, where τ is the maximum follow-up time.

We need the following conditions.

(A1) $N_i(t, s)$ is independent of (Y_i, X_i) and moreover, $E\{dN_i(t, s)\} = \lambda(t, s)dtds$, where $\lambda(t, s)$ is a twice-continuous differentiable function for any $0 \le t, s \le \tau$. In addition, Borel measure for $\mathcal{G} = \{\lambda(t, t) > 0, t \in [0, \tau]\}$ is strictly positive. For $t_1 \ne s_1, t_2 \ne s_2$, $\text{P}\{dN(t_1, t_2) = 1|N(s_1, s_2) - N(s_1-, s_2-) = 1\} = f(t_1, t_2, s_1, s_2)dt_1dt_2$ where $f(t_1, t_2, s_1, s_2)$ is continuous for $t_1 \ne s_1, t_2 \ne s_2$ and $f(t_1\pm, t_2\pm, s_1\pm, s_2\pm)$ exists.

(A2) If there exits a vector γ such that $\gamma^T X(s) = 0$ for any $s \in \mathcal{G}$ with probability one, then $\gamma = 0$.

(A3) For any β in a neighborhood of β_0, the true value of β, $E[X(s)g\{X(t)^T\beta\}]$ is continuously twice-differentiable in $(t, s) \in [0, \tau]^{\otimes 2}$ and $|g'(X(t)^T\eta)| \le q(\|X(t)\|)$ for some $q(\cdot)$ satisfying that $E[\|X(t)\|^4 q(\|X(t)\|)^2]$ is uniformly bounded in t. Additionally, $E\{\|X(t)\|^4\} < \infty$. Furthermore, $E[X(s_1)X(s_2)^T r\{t_1, t_2, X(t_1), X(t_2)\}]$ and $E[X(s_1)X(s_2)^T g\{X(t_1)^T\beta_0\}g\{X(t_2)^T\beta_0\}]$ are continuously twice differentiable in $(s_1, s_2, t_1, t_2) \in [0, \tau]^{\otimes 4}$. Moreover,

$$\int E\Big[\|X(s)X(s)^T\|\sigma\{s, X(s)\}^2\Big]\lambda(s, s)ds < \infty, \quad \text{and}$$
$$\int E\Big[\|X(s)X(s)^T\|g^{'}\{X(s)^T\beta_0\}\Big]\lambda(s, s)ds < \infty.$$

(A4) $K(\cdot)$ is a symmetric density function satisfying $\int z^2K(z)dz < \infty$ and $\int K(z)^2dz < \infty$. Additionally, $nh \to \infty$.

(A5) $nh^5 \to 0$.

The following theorem states the asymptotic properties of $\hat{\beta}_f$.

Theorem 1. *Under conditions (A1)-(A4), the asymptotic distribution of $\hat{\beta}_f$ satisfies:*

$$(nh)^{1/2}\big\{A(\beta_0)(\hat{\beta}_f - \beta_0) + Ch^2\big\} \to N(0, \Sigma), \tag{2.3}$$

where $A(\beta_0) = \int_s E[X(s)g'\{X(s)^T\beta_0\}X(s)^T]\lambda(s, s)ds$, β_0 is the true regression coefficient and C is a constant, which can be found in [1]. The asymptotic variance

$$\Sigma = \int K(z)^2dz \int E\Big[X(s)X(s)^T\sigma\{s, X(s)\}^2\Big]\lambda(s, s)ds. \tag{2.4}$$

If the bandwidth is further restricted by condition (A5), then the asymptotic bias in (2.3) vanishes and $\hat{\beta}_f$ is consistent:

Corollary 1. *Under conditions (A1)-(A5), $\hat{\beta}_f$ is consistent and converges to a mean zero normal distribution given in Theorem 1.*

2.2. Weighted LOCF estimation

In this subsection, we propose a weighted LOCF for the asynchronous longitudinal data. For the LOCF approach using generalized estimating equations for synchronous data as in [3], for a response at time t_{ij}, the covariate at time t_{ij} is taken to be the covariate observed at time $s = \max(x < t_{ij}, x \in \{s_{i1}, \ldots, s_{im_i}\})$. This method assumes that either the subject's response or the subject's covariate is constant from the most recent observation time and does not account for the variability inherent in this imputation. These assumptions may not hold in practice and violations can confound covariates with time, which in turn can bias estimates of covariate effects and their standard errors. As a result, the magnitude and even the direction of bias from LOCF is extremely difficult, if not impossible, to determine a priori.

We propose to remedy this bias by adopting a simple weighting strategy, downweighing imputed values which are far in time from the current response. To be specific, for a sample of n independent subject, the weighted generalized estimating equation for β is

$$\begin{aligned} U_n(\beta) \quad &= \quad n^{-1}\sum_{i=1}^{n}\int_0^1\int_0^1 K_h(t-s) I\Big\{s<t, \int_s^t dN_i(t,u)=0\Big\} \\ &\qquad X_i(s)\Big[Y_i(t)-g\{X_i(s)^T\beta\}\Big]dN_i(t,s), \end{aligned} \tag{2.5}$$

where $K_h(x) = K(x/h)/h, K(x)$ is a symmetric kernel function defined on $[-1, 1]$ and h is the bandwidth. Covariates are aggregated into the estimating equation (2.5) and for a response, there are multiple covariates contributing to the estimating equation with different weights. If there is no covariate measured before a response, such response does not contribute to the estimating equation (2.5). As the measurement times for the covariates and response are random and asynchronous, incorporating the correlation structure into the estimating equation to improve efficiency is unclear.

For $s < t$, the measurement times are allowed to depend on covariates through

$$E\Big[dN(t,s)I\{N(t,t)-N(t,s+)=0\}|X(s),Y(t)\Big] = \lambda\{t,s;X(s)\}dtds. \tag{2.6}$$

This assumption permits dependence on $Y(t)$ at times $t < s$, that is, future covariate observation times may depend on previous values of the response. The estimator presented above is valid under such informative observation times, which differs from [1], which did not allow dependence of the bivariate observation process on $Y(t)$ at any time points as can be seen from (A1). Additional regularity conditions are stated in condition (C0) below.

Before we present our asymptotic results, we need some notations and assumptions. The observations of $X(\cdot)$ can be arbitrarily correlated. We specify our assumptions on the covariance structure as follows. For $s, t, \in [0,1]$, let $\text{var}\{Y(t)|X(t)\} = \sigma\{t, X(t)\}^2$ and $\text{cov}\{Y(s), Y(t)|X(s), X(t)\} = r\{s, t, X(s), X(t)\}$. Furthermore, denote

$$F_\beta(s,t) = E[X(s)g\{X(t)^T\beta\}\lambda\{t, s; X(s)\}].$$

Its first order partial right and left derivative are

$$\dot{F}_\beta(s, s+) = \lim_{u\to 0+} u^{-1}\{F_\beta(s, s+u) - F_\beta(s,s)\},$$

and

$$\dot{F}_\beta(s, s-) = \lim_{u\to 0-} u^{-1}\{F_\beta(s, s+u) - F_\beta(s,s)\}.$$

By the same token, its second order partial right and left derivative are

$$\ddot{F}_\beta(s, s+) = \lim_{u\to 0+} (0.5u^2)^{-1}\{F_\beta(s, s+u) - F_\beta(s,s) - \dot{F}_\beta(s, s+)u\}$$

and

$$\ddot{F}_\beta(s, s-) = \lim_{u\to 0-} (0.5u^2)^{-1}\{F_\beta(s, s+u) - F_\beta(s,s) - \dot{F}_\beta(s, s-)u\}.$$

Moreover, denote $K_\beta(s,t) = E[X(s)g\{X(s)^T\beta\}\lambda\{t, s; X(s)\}]$ and its first order partial right and left derivative are defined in exactly the same way and denoted by $\dot{K}_\beta(s, s+)$ and $\dot{K}_\beta(s, s-)$. Let $\ddot{K}_\beta(s, s+)$ and $\ddot{K}_\beta(s, s-)$ be its second order partial right and left derivative. In addition, denote $G_\beta(s_1, s_2, t_1, t_2) = E\Big[X(s_1)X(s_2)^T g\{X(t_1)^T\beta\}g\{X(t_2)^T\beta\}\lambda\{t_2, s_2; X(s_2)\}\Big]$, its first order partial right and left derivative as $\dot{G}_\beta(s_1, s_2, s_1+, s_2+)$ and $\dot{G}_\beta(s_1, s_2, s_1-, s_2-)$; denote $J_\beta(s_1, s_2, t_1, t_2) = E\Big[X(s_1)X(s_2)^T r\{t_1, t_2, X(t_1), X(t_2)\}\lambda\{t_2, s_2; X(s_2)\}\Big]$, its first order partial right and left derivative as $\dot{J}_\beta(s_1, s_2, s_1+, s_2+)$ and $\dot{J}_\beta(s_1, s_2, s_1-, s_2-)$.

We need the following conditions.

(C0) For $s < t$, (2.6) is satisfied by $N(t,s)$. In addition, for any β in a neighborhood of β_0, the true regression coefficient, $F_\beta(s,t)$, $\dot{F}_\beta(s, s+)$, $\ddot{F}_\beta(s, s+)$, $K_\beta(s,t)$, $\dot{K}_\beta(s, s+)$, and $\ddot{K}_\beta(s, s+)$ are continuous functions for $(s,t) \in [0,1]^{\otimes 2}$. Moreover, $G_\beta(s_1, s_2, t_1, t_2)$, $\dot{G}_\beta(s_1, s_2, s_1+, s_2+)$, $J_\beta(s_1, s_2, t_1, t_2)$ and $\dot{J}_\beta(s_1, s_2, s_1+, s_2+)$ are continuous functions for $(s_1, s_2, t_1, t_2) \in [0,1]^{\otimes 4}$. Furthermore, $\text{P}\{dN(t_1, t_2) = 1 | N(s_1, s_2) - N(s_1-, s_2-) = 1\} = f(t_1, t_2, s_1, s_2)dt_1dt_2$ for $t_1 \neq s_1, t_2 \neq s_2$, where $f(t_1, t_2, s_1, s_2)$ is continuous for $t_1 \neq s_1, t_2 \neq s_2$ and $f(t_1\pm, t_2\pm, s_1\pm, s_2\pm)$ exist.

(C1) $\mathcal{F} \equiv \{s \in [0,1] \mid X(s) \neq 0\}$ has positive Borel measure with probability 1 and for $s \in \mathcal{F}, X(s)$ is not a constant function with probability 1. In addition, with positive probability, Borel measure for $\mathcal{G} = \Big[\lambda\{t,t;X(t)\} > 0, t \in \mathcal{F}\Big]$ is strictly positive.

(C2) For any β in a neighborhood of β_0, $|g'\{X(t)^T\beta\}| \le q(\|X(t)\|)$ for some $q(\cdot)$ satisfying $E\{\|X(t)\|^4 q(\|X(t)\|)^2\}$ is uniformly bounded in t. Moreover, $E\Big[X(s)X(s)^T\sigma\{t,X(t)\}^2\lambda\{t,s;X(s)\}\Big]$ is continuous and has partial right and left derivative with respect to t. Additionally, $E\{\|X(t)\|^4\} < \infty$,

$$\int_0^1 E\Big[\|X(s)X(s)^T\|\sigma^2\{s,X(s)\}\lambda\{s,s;X(s)\}\Big]ds < \infty,$$

and

$$\int_0^1 E\Big[\|X(s)X(s)^T\|g'\{X(s)^T\beta_0\}\lambda\{s,s;X(s)\}\Big]ds < \infty.$$

(C3) $K(\cdot)$ is a symmetric density function defined on $[-1,1]$ satisfying $|\int_0^1 zK(z)dz| = |\int_{-1}^0 zK(z)dz| < \infty$ and $\int_{-1}^1 K(z)^2dz < \infty$. Additionally, $nh \to \infty$ and $nh^5 \to 0$.

For $s < t$, the condition (C0) requires conditionally independent observation times in which the expectation of the bivariate counting process at times t and s is conditionally independent of the responses at time t and s given the covariates at times t and s. No assumptions are needed for $t < s$, unlike that specified in (A1). In addition, this condition assumes the existence of right continuous derivatives on $F_\beta(s,t)$. Such assumptions are satisfied by $X(t)$ with independent increments, such as Poisson process and Brownian motion. These stochastic processes do not satisfy (A3). We show below that the estimator's asymptotic behavior, in particular, the rate of convergence, may depend critically on the smoothness of the covariate processes. The other assumptions (C1)–(C3) are similar to those in Theorem 1.

The following theorem, which is proved in the appendix, states the asymptotic properties of $\hat\beta$ from $U_n(\beta)$ in (2.5) under the weaker conditions specified above.

Theorem 2. *Under (C0)–(C3), we have*

$$(nh)^{1/2}\big\{A(\beta_0)(\hat\beta - \beta_0) + Ch\big\} \to N(0,\Sigma), \tag{2.7}$$

where

$$A(\beta_0) = \frac{1}{2}\int_0^1 E[X(s)g'\{X(s)^T\beta_0\}X(s)^T\lambda\{s,s;X(s)\}]ds,$$
$$C = 2\int_0^1 zK(z)dz\int_0^1\{\dot F_{\beta_0}(s,s+) - \dot K_{\beta_0}(s,s+)\}ds$$

and

$$\Sigma = \int_0^1 K(z)^2dz\int_0^1 E\Big[X(s)X(s)^T\sigma^2\{s,X(s)\}\lambda\{s,s;X(s)\}\Big]ds.$$

From Theorem 2, the bias is generally of higher order h and we achieve a rate of convergence $n^{1/3}$, slower than $n^{2/5}$ specified in Theorem 1, where the bias is of order h^2. This is the price to pay for only requiring right continuous differentiability of certain functionals as specified in (C0). The increased bias resembles the boundary bias phenomenon in classical nonparametric regresion due to the boundary asymmetric kernel. To reduce the bias, one might employ boundary adjustment approaches which have been well studied in the nonparametric literature.

Regarding the computation, once the kernel function K has been chosen and the bandwidth has been fixed, the estimating equation can be solved using a standard Newton-Raphson implementation for generalized linear models, with good convergence properties.

Along the lines of [1], the variance of the estimators may be obtained using a sandwich formula $\{\partial U_n(\beta)/\partial\beta \mid_{\beta=\hat{\beta}}\}^{-1}\hat{\Sigma}[\{\partial U_n(\beta)/\partial\beta \mid_{\beta=\hat{\beta}}\}^{-1}]^T$, where $\hat{\Sigma} = n^{-2}\sum_{i=1}^n \Big(\int_0^1\int_0^1 K_h(t-s)X_i(s)I\{s<t, \int_s^t dN_i(t,u)=0\}[Y_i(t)-g\{X_i(s)^T\hat{\beta}\}] dN_i(t,s)\Big)^{\otimes 2}$. The consistency proof of $\hat{\Sigma}$ is in the appendix. Automatic bandwidth selection may be achieved as in [1], with bias of order h, as described in the appendix.

2.3. Half kernel estimation

To improve efficiency, the weighted LOCF can be extended to include information on all previously observed covariate, not only the most recently observed covariate. This is achieved by applying kernel weighting to all covariates observed before the response. The half kernel estimating equation is

$$U_n^*(\beta) = n^{-1}\sum_{i=1}^n \int_0^1\int_0^1 K_h^*(t-s)X_i(s)\Big[Y_i(t)-g\{X_i(s)^T\beta\}\Big]dN_i(t,s), \quad (2.8)$$

where

$$K_h^*(x) = \begin{cases} 2K_h(x) & \text{if } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases}$$

This accounts for the fact that the covariates and response are mismatched and only covariates that are observed before the response are used. If the observation times for covariates and response are close to each other, the kernel weight is close to 1; on the other hand, if they are far apart, the contribution to the estimating equation (2.8) may be 0. We solve $U_n^*(\beta)=0$ to obtain an estimate for β, denoted by $\hat{\beta}_*$. We modify the assumption on the bivariate counting process for simple weighted LOCF using half kernel estimation.

For $s \leq t$, the bivariate counting process $N(t,s)$ satisfies

$$\mathrm{E}\{dN(t,s) \mid X(s), Y(t), s\leq t\} = \mathrm{E}\{dN(t,s)\mid X(s)\} = \lambda^*\{t,s;X(s)\}dtds. \quad (2.9)$$

Similar to condition (2.6), (2.9) allows an informative observation process for times $t < s$. Hence, the half kernel estimation procedure above enjoys a robustness similar to that of weighted LOCF, which is not shared by the full kernel approach as in Theorem 1.

The asymptotic distribution of $\hat{\beta}_*$ is stated in the following theorem and proved in the appendix.

Theorem 3. *Under (C0) with (2.6) replaced by (2.9), (C1)–(C3), we have*

$$(nh)^{1/2}\big\{A^*(\beta_0)(\hat{\beta}_* - \beta_0) + C^* h\big\} \to N(0, \Sigma^*), \tag{2.10}$$

where $A^*(\beta_0)$ is obtained by replacing $\lambda\{s, s; X(s)\}$ by $\lambda^*\{s, s; X(s)\}$ in $A(\beta_0)$,

$$C^* \;=\; 2\int_0^1 zK(z)dz \int_0^1 \{\dot{F}^*_{\beta_0}(s, s+) - \dot{K}^*_{\beta_0}(s, s+)\}ds,$$

where $\dot{F}^*_{\beta_0}$ and $\dot{K}^*_{\beta_0}$ are obtained by replacing $\lambda\{t, s; X(s)\}$ by $\lambda^*\{t, s; X(s)\}$ in $\dot{F}_{\beta_0}$ and $\dot{K}_{\beta_0}$, respectively, and

$$\Sigma^* = \int_0^1 K(z)^2 dz \int_0^1 E\Big[X(s)X(s)^T \sigma^2\{s, X(s)\}\lambda^*\{s, s; X(s)\}\Big]ds.$$

For the half kernel approach, the bias and variance are generally of the same order as in the weighted LOCF. Improved bias properties are also possible under the following condition which may be satisfied by processes with independent increments:

(C4) $\dot{F}^*_{\beta_0}(s, s+) = 0$ and $\dot{K}^*_{\beta_0}(s, s+) = 0$.

When $X(t)$ follows a homogeneous Poisson process or the Brownian motion and g is the identify function, $\dot{F}^*_{\beta}(s, s+) = 0$ holds for all β. If $\lambda^*\{t, s; X(s)\}$ in (2.9) is constant, $\dot{K}^*_{\beta}(s, s+) = 0$ for all β. Consequently, $C^* = 0$ and the estimation bias for the half kernel based estimator $\hat{\beta}_*$ is of order $O(h^2)$ as specified in the following corollary.

Corollary 2. *Under same conditions as in Theorem 3 and with the addition of (C4), we have*

$$(nh)^{1/2}\big\{A^*(\beta_0)(\hat{\beta}_* - \beta_0) + C_2 h^2\big\} \to N(0, \Sigma^*), \tag{2.11}$$

where

$$C_2 \;=\; \int_0^1 z^2 K(z)dz \int_0^1 \{\ddot{F}^*_{\beta_0}(s, s+) - \ddot{K}^*_{\beta_0}(s, s+)\}ds,$$

where $\ddot{F}^*_{\beta_0}$ and $\ddot{K}^*_{\beta_0}$ are obtained by replacing $\lambda\{t, s; X(s)\}$ by $\lambda^*\{t, s; X(s)\}$ in $\ddot{F}_{\beta_0}$ and $\ddot{K}_{\beta_0}$, respectively, and Σ^* is the same as in Theorem 2.

This improvement in the convergence rate is shared by the weighted LOCF estimator. One might expect that half kernel estimation has smaller variance than the weighted LOCF, owing to the use of all previously observed covariates. While it is not possible to show that the half kernel estimator generally has smaller theoretical variance than weighted LOCF, simulations reported in Section 3 evidence some improvements. Interestingly, these differences are fairly small and diminish with large sample sizes. Variance estimation and bandwidth selection for half kernel estimation follow that for the weighted LOCF case and are omitted.

2.4. Revisiting full kernel estimation

For further efficiency improvement, a full kernel approach may be employed, as in section 2.1. We here consider the properties of this estimator under weaker conditions than those specified in section 2.1 We relax (A1) by allowing the measurement times to depend on covariates through

$$E\{dN(t,s) \mid X(s), Y(t)\} = E\{dN(t,s) \mid X(s)\} = \lambda^f\{t, s; X(s)\}dtds. \quad (2.12)$$

We solve $U_n^f(\beta) = 0$ to obtain an estimate for β, denoted by $\hat{\beta}_f^*$.

We require a stronger assumption than (C0) for weighted LOCF and for half kernel estimation, as specified below.

(C0*) The intensity function of the counting process $N(t,s)$ is specified in (2.12) for all s, t. In addition, for any β in a neighborhood of β_0, the true regression coefficient, $F_\beta^f(s,t)$, $\dot{F}_\beta^f(s,s+)$, $\dot{F}_\beta^f(s,s-)$, $\ddot{F}_\beta^f(s,s+)$, $\ddot{F}_\beta^f(s,s-)$, $K_\beta^f(s,t)$, $\dot{K}_\beta^f(s,s+)$, $\dot{K}_\beta^f(s,s-)$, $\ddot{K}_\beta^f(s,s+)$ and $\ddot{K}_\beta^f(s,s-)$ are continuous functions for $(s,t) \in [0,1]^{\otimes 2}$, where $\lambda\{t,s;X(s)\}$ is replaced by $\lambda^f\{t,s; X(s)\}$ in their respective definitions. $G_\beta^f(s_1,s_2,t_1,t_2)$, $\dot{G}_\beta^f(s_1,s_2,s_1+, s_2+)$, $\dot{G}_\beta^f(s_1,s_2,s_1-,s_2-)$, $J_\beta^f(s_1,s_2,t_1,t_2)$, $\dot{J}_\beta^f(s_1,s_2,s_1+,s_2+)$ and $\dot{J}_\beta^f(s_1,s_2,s_1-,s_2-)$ are continuous functions for $(s_1,s_2,t_1,t_2) \in [0,1]^{\otimes 4}$, where $\lambda\{t,s;X(s)\}$ is replaced by $\lambda^f\{t,s;X(s)\}$, respectively. $\mathrm{P}\{dN(t_1,t_2) = 1 | N(s_1,s_2) - N(s_1-,s_2-) = 1\} = f(t_1,t_2,s_1,s_2)dt_1dt_2$ for $t_1 \neq s_1, t_2 \neq s_2$, where $f(t_1,t_2,s_1,s_2)$ is continuous for $t_1 \neq s_1, t_2 \neq s_2$ and $f(t_1\pm, t_2\pm, s_1\pm, s_2\pm)$ exist.

This assumption is stronger than those for weighted LOCF and half kernel estimation. This assumption does not permit dependence of the bivariate observation process on the response $Y(t)$ at any times s and t. It is weaker than (A3) in that it permits the observation process to depend on the covariate process. It also weakens (A3) by relaxing the smoothness conditions on $X(s)$, covering the important special case of independent increments. The following theorem, which is proved in the appendix, states the asymptotic properties of the full kernel estimator under these more general conditions.

Theorem 4. *Under (C0*), (C1)–(C3), we have*

$$(nh)^{1/2}\{A^f(\beta_0)(\hat{\beta}_f^* - \beta_0) + C^f h\} \to N(0, \Sigma^f), \tag{2.13}$$

where

$$A^f(\beta_0) = \int_0^1 E[X(s)g'\{X(s)^T\beta_0\}X(s)^T\lambda^f\{s,s;X(s)\}]ds,$$

$$C^f = \int_0^1 zK(z)dz\Big[\int_0^1 \{\dot{F}_{\beta_0}^f(s,s+) - \dot{F}_{\beta_0}^f(s,s-)\}ds - \int_0^1 \{\dot{K}_{\beta_0}^f(s,s+)ds - \dot{K}_{\beta_0}^f(s,s-)\}ds\Big],$$

and

$$\Sigma^f = \int_{-1}^1 K(z)^2 dz \int_0^1 E\Big[X(s)X(s)^T\sigma^2\{s,X(s)\}\lambda^f\{s,s;X(s)\}\Big]ds.$$

For full kernel approach, the bias is of the same order as that in weighted LOCF and half kernel approach. However, the variance is smaller if $\lambda\{s,s;X(s)\} = \lambda^*\{s,s;X(s)\} = \lambda^f\{s,s;X(s)\}$. This is accomplished through utilizing both lagged and forward observations.

A special case of Theorem 4 gives the result in Theorem 1.

Corollary 3. *Under the special case that* $\lambda^f\{t,s;X(s)\} = \lambda(t,s)$, *a twice continuously differentiable function in* $[0,1]^{\otimes 2}$ *and* $E[X(s)g\{X(t)^T\beta\}]$ *is twice continuously differentiable for any* β *in a neighborhood of* β_0, $\dot{F}_{\beta_0}^f(s,s+) = \dot{F}_{\beta_0}^f(s,s-)$ *and* $\dot{K}_{\beta_0}^f(s,s+) = \dot{K}_{\beta_0}^f(s,s-)$. *Theorem 3 is the same as Theorem 1.*

3. Simulation studies and a real example

We conducted extensive simulation studies to evaluate the properties of the proposed estimators in practical settings. We first study the performance of LOCF estimate, the proposed weighted LOCF estimate, half kernel estimate and full kernel estimate when assumptions in Cao et al. [1] hold. We generate $1,000$ dataset, each consisting of $n = 400$ or 1000 subjects. The numbers of observation times for the response $Y(t)$ and covariate $X(t)$ are generated from Poisson distribution with intensity rate 5. The observation times for the response and covariates are generated from uniform distribution $\mathcal{U}(0,1)$ independently. The covariate process is Gaussian, with values at observed time points being multivariate normal with mean 0, variance 1 and correlation $e^{-|t_{ij}-t_{ik}|}$, where t_{ij} and t_{ik} are jth and kth measurement time for the response, both on subject i. The response process was generated from

$$Y(t) = \beta_0 + X(t)\beta_1 + \epsilon(t),$$

where β_0 is the intercept, β_1 is regression coefficient and $\epsilon(t)$ is Gaussian, with mean 0, variance 1 and $\text{cov}\{\epsilon(s), \epsilon(t)\} = 2^{-|t-s|}$. Once the response is generated, we remove the covariate measurements at the response observation times to create the asynchronous data structure. In this simulation, we set $\beta_0 = 1.5$ and $\beta_1 = 1.5$ and assess the performance of $\hat{\beta}_1$. The results are very similar for other choices of βs.

For weighted LOCF, half kernel and full kernel estimation, the kernel function is the Epanechnikov kernel, which is $K(x) = 0.75(1-x^2)_+$, with the automatic bandwidth selection described in the appendix used in the estimation. Similar results were obtained using other kernel functions. A total of 1000 simulated dataset were analyzed.

Table 1A and 2A summarize the results of these simulations. Additional results with $n = 200$ can be found in the Appendix. For standard LOCF, the biases and coverage probabilities are -0.192 and 4.4% when $n = 400$ and -0.189 and 0 when $n = 1000$. Weighted LOCF, half kernel estimation, and full kernel estimation perform satisfactorily in terms of bias, variance, and coverage probability, particularly with larger sample sizes. In this setting, where the assumptions in Theorem 1 are satisfied, full kernel estimation exhibits smaller bias and variance than either weighted LOCF or half kernel estimation.

We next study the case that covariates follow Poisson process with intensity 3. The independent increments set-up violates the assumptions in Theorem 1 and our theory suggests an improved rate of convergence for half kernel estimation versus full kernel estimation. The data generation scheme is otherwise the same. For LOCF, the biases and coverage probabilities are -0.053 and 69.6% when $n = 400$ and -0.055 and 38.0% when $n = 1000$. From Table 1B and 2B, we observe that the full kernel approach has substantially larger bias than the weighted LOCF and half kernel approach, as predicted by Corollary 2. The empirical variances and variance estimates are in good agreement. The coverage probabilities for the weighted LOCF and half kernel approach are close to the nominal level. Those for the full kernel are much lower than the nominal level, owing to the large biases.

We then study informative observation times depending on responses. We first generate the observation time of one response t_0, which is $\mathcal{U}(0,1)$ distributed. The raw observation times for covariates follows Poisson$\{\exp(3)\}$ uniformly distributed in the 0.3 neighborhood of t_0. The rest of data generation is exactly the same as before. We use a thinning algorithm to determine whether to keep the covariate observation times, where the probability of keeping covariates observed before the response is 0.2 and after the response is $\min[1, 15\exp\{Y(t)/3\}]$. For LOCF, the biases and coverage probabilities are -0.123 and 14.2% when $n = 1000$ and -0.123 and 0 when $n = 5000$. The assumptions in Theorems 2 and 3 on the bivariate intensity process for weighted LOCF and half kernel estimation are satisfied whereas those for full kernel estimation in Theorem 4 are violated. This is evidenced in Table 1C and 2C, where the bias is larger and the coverage probability is poor for full kernel. Weighted LOCF and half kernel approach have small bias, good agreement between estimated and empirical standard errors and coverage probabilities close to the nominal level. As full kernel approach uses more data, it has smaller standard error compared with weighted LOCF and half kernel approach.

Table 1

Results of 1000 simulations under different scenarios: A: assumptions in Cao et al. [1] are satisfied; B: covariates follow Poisson process; C: informative observation time.

BD	Bias	RB	SD	SE	CP	Bias	RB	SD	SE	CP
Weighted LOCF										
A										
$n = 400$						$n = 1000$				
0.005	0.1	0.1	0.168	0.155	91	-0.5	-0.4	0.102	0.099	94
0.015	-0.8	-0.5	0.094	0.097	95	-0.7	-0.5	0.062	0.062	94
0.025	-1.4	-0.9	0.078	0.080	95	-1.2	-0.8	0.051	0.051	93
0.035	-1.9	-1.3	0.071	0.071	94	-1.7	-1.1	0.046	0.046	92
auto	-0.7	-0.5	0.163	0.151	92	-0.6	-0.4	0.103	0.099	93
B										
$n = 400$						$n = 1000$				
0.05	0.1	0.1	0.043	0.042	95	-0.1	-0.1	0.027	0.027	95
0.10	-0.1	-0.1	0.036	0.036	95	-0.2	-0.2	0.023	0.023	95
0.15	-0.3	-0.2	0.034	0.035	95	-0.4	-0.3	0.022	0.022	94
0.20	-0.6	-0.4	0.033	0.034	95	-0.7	-0.5	0.022	0.022	92
auto	-0.2	-0.1	0.043	0.042	94	-0.1	-0.1	0.028	0.027	93
C										
$n = 1000$						$n = 5000$				
0.015	-0.9	-0.6	0.104	0.100	94	-0.7	-0.5	0.045	0.046	95
0.020	-1.1	-0.7	0.092	0.088	94	-1.1	-0.7	0.040	0.040	94
0.025	-1.3	-0.8	0.083	0.080	94	-1.4	-0.9	0.036	0.036	94
0.030	-1.6	-1.1	0.077	0.075	94	-1.7	-1.1	0.033	0.034	92
auto	-0.9	-0.6	0.101	0.100	93	-0.7	-0.5	0.043	0.046	95
Half kernel										
A										
$n = 400$						$n = 1000$				
0.005	-0.6	-0.4	0.165	0.154	93	0.1	0.1	0.102	0.100	93
0.015	-0.9	-0.6	0.101	0.098	93	-0.5	-0.3	0.064	0.063	94
0.025	-1.3	-0.9	0.084	0.081	94	-1.2	-0.8	0.066	0.065	95
0.035	-1.9	-1.2	0.075	0.073	94	-1.8	-1.2	0.048	0.046	92
auto	3.1	2.0	0.135	0.156	92	1.1	0.8	0.104	0.100	92
B										
$n = 400$						$n = 1000$				
0.05	-1.1	-0.7	0.043	0.044	96	0.5	0.3	0.027	0.028	94
0.10	-0.8	-0.5	0.038	0.038	95	0.4	0.3	0.022	0.024	96
0.15	-1.0	-0.7	0.036	0.036	95	0.0	0.0	0.022	0.023	95
0.20	-1.3	-0.9	0.035	0.036	97	-0.5	-0.3	0.022	0.023	94
auto	-0.0	-0.0	0.037	0.038	96	0.4	0.3	0.027	0.026	94
C										
$n = 1000$						$n = 5000$				
0.015	-0.6	-0.4	0.104	0.103	94	-1.1	-0.7	0.047	0.047	94
0.020	-1.0	-0.7	0.092	0.091	93	-1.3	-0.9	0.042	0.041	94
0.025	-1.3	-0.9	0.084	0.083	93	-1.6	-1.0	0.038	0.038	92
0.030	-1.6	-1.1	0.078	0.078	93	-1.8	-1.2	0.036	0.035	92
auto	-2.2	-1.5	0.086	0.076	93	-0.6	-0.4	0.046	0.048	95

Note: "BD" represents different bandwidths, "Bias' (%)' is the empirical bias, "RB (%)" is the "Bias" divided by the true β_1, "SD" is the sample standard deviation, "SE" is the average of the standard error estimates and "CP (%)" represents the coverage probability of the 95% confidence interval for $\hat{\beta}_1$.

Per the request of the referee, we conducted additional simulations to compare with IPW as in [12, 13]. Strictly speaking, the assumptions for IPW are

TABLE 2

Results of 1000 simulations under different scenarios: A: assumptions in Cao et al. [1] are satisfied; B: covariates follow Poisson process; C: informative observation time.

BD	Bias	RB	SD	SE	CP	Bias	RB	SD	SE	CP
					Full kernel					
A										
$n = 400$						$n = 1000$				
0.005	-0.2	-0.1	0.120	0.110	94	-0.5	-0.4	0.075	0.074	94
0.015	-0.8	-0.5	0.078	0.075	93	-0.9	-0.6	0.049	0.048	94
0.025	-1.2	-0.8	0.066	0.065	95	-1.4	-0.9	0.042	0.041	93
0.035	-1.7	-1.2	0.061	0.060	94	-1.9	-1.3	0.038	0.038	92
auto	-0.7	-0.5	0.104	0.116	96	-0.9	-0.6	0.074	0.074	92
B										
$n = 400$						$n = 1000$				
0.05	-2.6	-1.7	0.035	0.036	90	-2.0	-1.3	0.022	0.023	86
0.10	-4.7	-3.1	0.034	0.033	67	-4.3	-2.9	0.022	0.021	42
0.15	-7.4	-4.9	0.033	0.032	34	-6.9	-4.6	0.022	0.020	9
0.20	-10.4	-6.9	0.032	0.032	11	-9.8	-6.5	0.022	0.020	1
auto	-3.8	-2.6	0.032	0.033	72	-2.6	-1.7	0.021	0.022	81
C										
$n = 1000$						$n = 5000$				
0.015	-4.2	-2.8	0.069	0.069	91	-4.2	-2.8	0.033	0.031	73
0.020	-7.4	-4.9	0.059	0.059	76	-7.5	-5.0	0.028	0.027	22
0.025	-9.4	-6.3	0.054	0.054	58	-9.5	-6.4	0.025	0.024	3
0.030	-10.8	-7.2	0.051	0.050	43	-11.0	-7.3	0.023	0.023	0
auto	-12.3	-8.2	0.049	0.050	28	-3.4	-2.7	0.033	0.033	79

Note: "BD" represents different bandwidths, "Bias' (%)' is the empirical bias, "RB (%)" is the "Bias" divided by the true β_1, "SD" is the sample standard deviation, "SE" is the average of the standard error estimates and "CP (%)" represents the coverage probability of the 95% confidence interval for $\hat{\beta}_1$.

TABLE 3

Results of 1000 simulations based on IPW. A: last observed covariate; B: nearest covariate.

n	Bias	RB	SD	SE	CP
A					
400	-0.1893	-9.1262	0.0512	0.0513	3.90
1000	-0.1906	-0.1271	0.0325	0.0325	0.00
B					
400	-0.1301	-0.0867	0.0462	0.0461	19.28
1000	-0.1312	-0.0874	0.0294	0.0292	0.40

Note: "Bias" is the empirical bias, "RB" is the "Bias" divided by the true β_1, "SD" is the sample standard deviation, "SE" is the average of the standard error estimates and "CP (%)" represents the coverage probability of the 95% confidence interval for $\hat{\beta}_1$.

not satisfied. For any subject, the covariate and response observation times are mismatched and the probability of observing complete data is 0. The data generation and implementation are detailed in the Appendix and the results are summarized in Table 3. We can see that IPW incurs substantial bias, which does not attenuate as sample size increases and therefore should not be used to analyze asynchronous longitudinal data.

Table 4
Summary statistics for $\hat{\beta}_1$ based on (2.2), (2.5) and (2.8).

h	$n^{-0.3}$	$n^{-0.5}$	$n^{-0.7}$	auto
full kernel	290	102	36	67
$\hat{\beta}_1$	-1.202	-1.148	-1.074	-1.112
SE($\hat{\beta}_1$)	0.153	0.177	0.228	0.195
z-value	-7.866	-6.481	-4.710	-5.692
weighted LOCF	290	102	36	87
$\hat{\beta}_1$	-1.195	-1.118	-1.030	-1.117
SE($\hat{\beta}_1$)	0.157	0.222	0.326	0.234
z-value	-7.593	-5.033	-3.169	-4.779
half kernel	290	102	36	36
$\hat{\beta}_1$	-1.257	-1.165	-1.030	-1.030
SE($\hat{\beta}_1$)	0.190	0.219	0.326	0.326
z-value	-6.604	-5.312	-3.168	-3.168

We now illustrate the proposed inferential procedures on a dataset from an HIV study [19], previously analyzed in [1]. A total of 190 patients were followed from July 1997 to September 2002 in a university hospital. There are unequal numbers of repeated measurements on viral load and CD4 count and there are different measurement times for these two variables. In our analysis, we take log transformed CD4 counts as covariate and log transformed HIV viral load as response. We use estimating equations (2.5) and (2.8) with bandwidths $h = 2(Q_3 - Q_1)n^{-\gamma}$, 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, n is the number of patients and $\gamma = 0.3, 0.5, 0.7$. The results are summarized in Table 3 with fixed bandwidths and data adaptive bandwidth. Results based on Theorem 1 are presented for comparison.

Earlier work has shown that LOCF produces a weak positive association between CD4 counts and HIV viral load, which is in an opposite direction to the known relationship between these variables. From Table 4, we see that weighted LOCF and half kernel produce similar point estimates and standard deviations, especially when bandwidths are small ($n^{-0.5}$ or $n^{-0.7}$). Full kernel approach in [1] has similar point estimates, but smaller standard deviation, due to the fact that it uses both forward and lagged covariates. Analysis based on newly proposed procedure and earlier method in [1] all showed statistically significant association between CD4 counts and HIV viral load, consistent with findings in the medical literature [11].

4. Concluding remarks

In this paper, we provide an intuitively appealing and rigorous formalization of LOCF for regression analysis with asynchronous longitudinal data. The resulting estimators are consistent and asymptotically normal, but with a rate of convergence which is slower than the usual parametric rate. The procedure performed well in simulations, evidencing substantial improvements in bias and coverage properties over the naïve LOCF. Its ease of implementation suggests that it has the potential to be practically useful in applications where LOCF is currently the method of choice, requiring only the addition of a weight to the generalized estimating equations.

Interestingly, the simulation studies demonstrated only a small loss of efficiency relative to half kernel estimation, which utilizes all previous covariate observations. Our intuition is that without stronger assumptions than those in the current paper, the most recently observed covariate contains the majority of information about the previous covariate values. As the sample size increases and the bandwidth shrinks, the half kernel estimation procedure only uses the most recent values of the covariates, similarly to weighted LOCF. Further work is needed for a rigorous comparison of the theoretical variances of these estimators.

Both weighted LOCF and half kernel estimation performed well with informative observation times, while full kernel estimation exhibited large biases and poor coverage. This lack of robustness should be weighted against the improved efficiency which may occur when necessary regularity conditions are satisfied. A related loss of efficiency may occur with covariates having independent increments, in which case both theoretical and simulation studies point to the superior performance of weighted LOCF and half kernel estimation. Additional numerical work would be valuable in further elucidating these issues.

Both GEE with synchronous data and our proposed approach for asynchronous data are valid when the data are missing completely at random as in [5, 8]. In GEE, with time-dependent covariates, [10] 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 the 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 estimation 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 data set-up, with further work needed to understand the extent to which valid estimation may be achieved with non-diagonal working covariance matrices and whether efficiency gains might be achievable. The efficiency issue is complicated by the fact that the asynchronous data estimators converge more slowly than the usual parametric rate obtained by GEE with synchronous data.

A key assumption for the proposed estimators is that the measurement times for previous covariates are independent of the current and future observed responses. This assumption excludes certain missing at random settings under which GEE with synchronous data might yield valid estimation based only on complete observations. On the other hand, GEE does not allow non-ignorable missingness. This contrasts with our approach, in which the missingness mechanism is specified by the bivariate intensity for the measurement times for the responses and covariates. Our assumptions allow non-ignorable missingness when the probability of observing a covariate, e.g., the measurement time intensity, depends on the value of the missing covariate. Our approach fails with non-ignorable missingness when the probability of observing a response depends on the value of the missing response.

Appendix A: Appendix

This appendix includes proofs of Theorems 2–4, practical implementation of the proposed method and its proof and additional simulations.

A.1. Proof of Theorem 2

Proof. The key idea is to establish the following relationship

$$\sup_{|\beta-\beta_0|<M(nh)^{-1/2}} \Big|(nh)^{1/2}U_n(\beta) - (nh)^{1/2}[U_n(\beta_0) - E\{U_n(\beta_0)\}] + (nh)^{1/2} A(\beta_0)(\beta-\beta_0)\Big| Cn^{1/2}h^{3/2} + o_p(n^{1/2}h^{3/2}) + o_p\{1+(nh)^{1/2}|\beta-\beta_0|\}, \quad \text{(A.1)}$$

where $A(\beta_0)$ is given in Theorem 2 and

$$C = 2\int_0^1 zK(z)dz \int_0^1 \{\partial F_{\beta_0}(s,t)/\partial t_{t=s+} - \partial K_{\beta_0}(s,t)/\partial t_{t=s+}\}ds.$$

To obtain (A.1), first, using $\mathcal{P}_n$ and $\mathcal{P}$ to denote the empirical measure and true probability measure respectively, we obtain

$$\begin{aligned}
&(nh)^{1/2}U_n(\beta) = (nh)^{1/2}(\mathcal{P}_n - \mathcal{P}) \\
&\int_0^1\int_0^1 K_h(t-s)I\{s<t, dN_i(t,u)=0\}X(s)\left[Y(t) - g\{X(s)^T\beta\}\right]dN(t,s) \\
&+ (nh)^{1/2}E \\
&\int_0^1\int_0^1 K_h(t-s)I\{s<t, dN_i(t,u)=0\}X(s)\left[Y(t) - g\{X(s)^T\beta\}\right]dN(t,s) \\
&= I + II. \qquad \text{(A.2)}
\end{aligned}$$

For the second term on the right-hand side of (A.2), we have

$$\begin{aligned}
II &= (nh)^{1/2}\int_0^1\int_0^1 K_h(t-s)E\Big(\left[X(s)g\{X(t)^T\beta_0\} - X(s)g\{X(s)^T\beta\}\right] \\
&\quad \lambda\{t,s;X(s)\}\Big)dtds \\
&= (nh)^{1/2}\int_0^1\int_0^1 K(z)E\Big(\left[X(s)g\{X(s+hz)^T\beta_0\} - X(s)g\{X(s)^T\beta\}\right] \\
&\quad \lambda\{s+hz,s;X(s)\}\Big)dzds.
\end{aligned}$$

Note $F_{\beta_0}(s,t) = E[X(s)g\{X(t)^T\beta_0\}\lambda\{t,s;X(s)\}$, $K_\beta(s,t) = E[X(s)g\{X(s)^T\beta\}\lambda\{t,s;X(s)\}]$. After Taylor expansion of $F_{\beta_0}(s,s+hz)$ and $K_\beta(s,s+hz)$, we obtain

$$\begin{aligned} II &= (nh)^{1/2}\Big[\int_{s=0}^1\int_{z=0}^1 K(z)dz\{F_{\beta_0}(s,s)-K_\beta(s,s)\}ds\Big] \\ &+ n^{1/2}h^{3/2}\Big[\int_{s=0}^1\int_0^1 zK(z)dz\{\partial F_{\beta_0}(s,s+)/\partial t - \partial K_\beta(s,s+)/\partial t\}ds \\ &+ O_p(n^{1/2}h^{5/2}), \end{aligned}$$

where we use the assumptions that $F_{\beta_0}(s,t)$ and $K_\beta(s,t)$ are continuous functions for $(s,t)\in[0,1]^{\otimes 2}$ and they have continuous left and right derivatives specified in (C0). Let $h\to 0$ and since $\int_0^1 K(z) = 0.5$, we extract the main terms

$$\begin{aligned} II &= -\frac{1}{2}(nh)^{1/2}\int_{s=0}^1 E\Big[X(s)g'\{X(s)^T\beta_0\}X(s)^T\lambda\{s,s;X(s)\}\Big]ds(\beta-\beta_0) \\ &\quad + Cn^{1/2}h^{3/2} + o_p(n^{1/2}h^{3/2}) + (nh)^{1/2}o(|\beta-\beta_0|) \qquad\qquad (A.3) \\ &\equiv -(nh)^{1/2}A(\beta_0)(\beta-\beta_0) + Cn^{1/2}h^{3/2} + o_p(n^{1/2}h^{3/2}) + (nh)^{1/2}o(|\beta-\beta_0|). \end{aligned}$$

Moreover, $A(\beta_0)$ is a positive-definite matrix from (C1), thus non-singular.

For the first term on the right-hand side of (A.2), we consider the class of functions

$$\{\int\int h^{1/2}K_h(t-s)I\{s<t,\int_s^t dN_i(t,u)=0\}X(s)\left[Y(t)-g\{X(s)^T\beta\}\right] \\ dN(t,s) : |\beta-\beta_0|<\epsilon\}$$

for a given constant ϵ. Note that the functions in this class are Lipschitz continuous in β and the Lipschitz constant is uniformly bounded by

$$M_1 = \sup_{|\beta-\beta_0|<\epsilon}\int\int h^{1/2}K_h(t-s)\|X(s)\|^2\|g'(X(s)^T\beta)\|dN(t,s).$$

Since by (C2),

$$M_1^2 \le \int\int hK_h(t-s)^2\|X(s)\|^4 q(\|X(s)\|)^2 dN(t,s)N(\tau,\tau),$$

we have

$$\begin{aligned} \{M_1^2|N(\cdot,\cdot)\} &\le N(\tau,\tau)\int\int hK_h(t-s)^2E\{\|X(s)\|^4q(\|X(s)\|)^2\}dN(t,s) \\ &\le M_2N(\tau,\tau)\int\int hK_h(t-s)^2dN(t,s) \end{aligned}$$

for some constant M_2. Conditional on $N(\tau,\tau)$, $E\{\int\int hK_h(t-s)^2dN(t,s)|N(\tau,\tau)\}$ can be easily verified to be finite. Thus, $E(M_1^2)$ is finite. Therefore, this class is a P-Donsker class by the Jain-Marcus theorem [17]. As the result, we obtain that the first term in the right-hand side of (A.2) for $|\beta-\beta_0| < M(nh)^{-1/2}$ is equal to

$$
\begin{aligned}
& (nh)^{1/2}(\mathcal{P}_n - \mathcal{P}) \iint K_h(t-s) \\
& I\{s < t, \int_s^t dN_i(t,u) = 0\} X(s) \left[Y(t) - g\{X(s)^T \beta_0\} \right] dN(t,s) + o_p(1) \\
= \; & (nh)^{1/2} \Big[U_n(\beta_0) - E\left\{ U_n(\beta_0) \right\} \Big] + o_p(1). \qquad (A.4)
\end{aligned}
$$

Combining (A.3) and (A.4) and by condition (C3), we obtain (A.1). Consequently,

$$
\begin{aligned}
& (nh)^{1/2} A(\beta_0)(\hat{\beta} - \beta_0) + C n^{1/2} h^{3/2} + o_p(n^{1/2} h^{3/2}) \\
+ \; & o_p\{1 + (nh)^{1/2} |\hat{\beta} - \beta_0|\} = (nh)^{1/2} \left[U_n(\beta_0) - E\left\{ U_n(\beta_0) \right\} \right]. \qquad (A.5)
\end{aligned}
$$

On the other hand, following the similar argument as before, we can calculate

$$
\begin{aligned}
\Sigma = & \, h\text{var} \iint K_h(t-s) I\{s < t, \int_s^t dN_i(t,u) = 0\} X(s) \left[Y(t) - g\{X(s)^T \beta_0\} \right] \\
& dN(t,s) == hE \left[\text{var} \left\{ D \middle| X(s), s \in [0,\tau]; N(t,s), (t,s) \in [0,1]^{\otimes 2} \right\} \right] \\
& + h\text{var} \left(E\{ D \middle| X(s), s \in [0,\tau]; N(t,s), (t,s) \in [0,1]^{\otimes 2} \} \right) \\
& = hE \Big(\iint \iint K_h(t_1 - s_1) K_h(t_2 - s_2) X(s_1) X(s_2)^T [r\{t_1, t_2, X(t_1), X(t_2)\} \\
& + g\{X(t_1)^T \beta_0\} g\{X(t_2)^T \beta_0\}] I\{s_1 < t_1, \int_{s_1}^{t_1} dN(t_1, u_1) = 0\} \\
& \{s_2 < t_2, \int_{s_2}^{t_2} dN(t_2, u_2) = 0\} dN(t_1, s_1) dN(t_2, s_2) \Big) \\
& - hE \Big[\iint K_h(t-s) I\{s < t, \int_s^t dN(t,u) = 0\} X(s) g\{X(t)^T \beta_0\} dN(t,s) \Big]^2 \\
& + hE \iint \iint K_h(t_1 - s_1) \Big[g\{X(t_1)^T \beta_0\} - g\{X(s_1)^T \beta_0\} \Big] X(s_1) X(s_2)^T \\
& I\{s_1 < t_1, \int_{s_1}^{t_1} dN(t_1, u_1) = 0\} \Big[g\{X(t_2)^T \beta_0\} - g\{X(s_2)^T \beta_0\} \Big] \\
& I\{s_2 < t_2, \int_{s_2}^{t_2} dN(t_2, u_2) = 0\} K_h(t_2 - s_2) dN(t_1, s_1) dN(t_2, s_2) \\
& - h\{ \iint K_h(t-s) I(s < t) E(X(s)[g\{X(t)^T \beta_0\} - g\{X(s)^T \beta_0\}] \\
& \lambda\{t, s; X(s)\} dt ds) \}^2 \\
& = I_1 - I_2 + I_3 - I_4.
\end{aligned}
$$

Using conditioning argument, we obtain

$$
I_1 \;=\; h \int_{t_1 \neq t_2} \int_{s_1 \neq s_2} I(s_1 < t_1, s_2 < t_2) K_h(t_1 - s_1) K_h(t_2 - s_2)
$$

$$
\begin{aligned}
& E\Big(X(s_1)X(s_2)^T[r\{t_1,t_2,X(t_1),X(t_2)+g\{X(t_1)^T\beta_0\}g\{X(t_2)^T\beta_0\}] \\
& \lambda\{t_2,s_2;X(s_2)\}\Big)f(t_1,t_2,s_1,s_2)dt_1ds_1dt_2ds_2 \\
+ \quad & h\int_{t_1}\int_{s_1\neq s_2} I(s_1<t_1,s_2<t_1)K_h(t_1-s_1)K_h(t_1-s_2) \\
& E\Big\{X(s_1)X(s_2)^T[\sigma^2\{t_1,X(t_1)\} \\
+ \quad & g\{X(t_1)^T\beta_0\}^2]\lambda\{t_1,s_2;X(s_2)\}\Big\}f(t_1,t_1,s_1,s_2)dt_1ds_1ds_2 \\
+ \quad & h\int_{t_1\neq t_2}\int_{s_1} I(s_1<t_1,s_1<t_2)K_h(t_1-s_1)K_h(t_2-s_1) \\
& E\Big(X(s_1)X(s_1)^T[r\{t_1,t_2,X(t_1),X(t_2)\} \\
+ \quad & g\{X(t_1)^T\beta_0\}g\{X(t_2)^T\beta_0\}]\lambda\{t_2,s_1;X(s_1)\}\Big)f(t_1,t_2,s_1,s_1)dt_1ds_1dt_2 \\
+ \quad & h\int_{t_1}\int_{s_1} I(s_1<t_1)K_h(t_1-s_1)^2E\Big(X(s_1)X(s_1)^T \\
& [\sigma^2\{t_1,X(t_1)\}+g\{X(t_1)^T\beta_0\}^2]\lambda\{t_1,s_1;X(s_1)\}\Big)dt_1ds_1.
\end{aligned}
$$

After change of variable and incorporating conditions (C2) and (C3), the first three terms in I_1 are all of order $O(h)$ and the last term equals to

$$
\int_0^1 K(z)^2dz\int E\Big(X(s)X(s)^T[\sigma^2\{s,X(s)\}+g\{X(s)^T\beta_0\}^2]\lambda\{s,s;X(s)\}\Big)ds.
$$

So we have

$$
\begin{aligned}
I_1 \quad = \quad & \int_0^1 K(z)^2dz\int E\Big(X(s)X(s)^T[\sigma^2\{s,X(s)\}+g\{X(s)^T\beta_0\}^2] \\
& \lambda\{s,s;X(s)\}\Big)ds+O(h).
\end{aligned}
$$

Similarly, it can be shown that

$$
I_2=\int_0^1 K(z)^2dz\int E\Big[X(s)X(s)^Tg\{X(s)^T\beta_0\}^2\lambda\{s,s;X(s)\}\Big]ds+O(h), \quad \text{(A.6)}
$$

and $I_3-I_4=O(h)$. Therefore, we have

$$
\Sigma=\int_0^1 K(z)^2dz\int E\Big[X(s)X(s)^T\sigma^2\{s,X(s)\}\lambda\{s,s;X(s)\}\Big]ds. \qquad \text{(A.7)}
$$

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

$$
\begin{aligned}
\psi_i=(nh)^{1/2}n^{-1}\iint I\{s<t,\int_s^t dN_i(t,u)=0\}K_h(t-s)X_i(s) \\
\big[Y_i(t)-g\{X_i(s)^T\beta_0\}\big]\,dN_i(t,s).
\end{aligned}
$$

Similar to the calculation of Σ,

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

Therefore,

$$(nh)^{1/2}\Big[U_n(\beta_0) - E\{U_n(\beta_0)\}\Big] \to_d N(0, \Sigma). \tag{A.8}$$

Combing with (A.5), we finish the proof of Theorem 2.

A.2. Proofs of Theorem 3 and Theorem 4

Proof. The proofs of Theorem 3 and Theorem 4 are very similar to the proof of Theorem 2 and thus omitted.

A.3. Consistency of variance estimation

For statistical inference, we estimate Σ by

$$\hat{\Sigma} = n^{-2}\sum_{i=1}^{n}\Big(\int_0^1\int_0^1 I\{s<t, \int_s^t dN_i(s,u)=0\}$$
$$K_h(t-s)X_i(s)[Y_i(t) - g\{X_i(s)^T\hat{\beta}\}]dN_i(t,s)\Big)^{\otimes 2}$$

and estimate the variance of $\hat{\beta}$ by the sandwich formula $\{\partial U_n(\beta)/\partial\beta\ |_{\beta=\hat{\beta}}\}^{-1}\hat{\Sigma}[\{\partial U_n(\beta)/\partial\beta\ |_{\beta=\hat{\beta}}\}^{-1}]^T$. Similar estimator can be obtained for $\hat{\Sigma}^*$ and $\hat{\Sigma}^f$. Denote $L_\beta(s,t) = \mathrm{E}\Big[X(s)\sigma^2\{t, X(t)\}X(s)^T\lambda\{t,s;X(s)\}\Big]$ and its first order partial right and left derivative with respect to $t \in [0,1]$ as $\dot{L}_\beta(s,s+)$ and $\dot{L}_\beta(s,s-)$. We need the following assumptions.

- (C5) For any β in a neighborhood of β_0, $L_\beta(s,t)$ and $\dot{L}(s,s+)$ are continuous functions.
- (C5*) For any β in a neighborhood of β_0, $L_\beta(s,t), \dot{L}_\beta(s,s+)$ and $\dot{L}_\beta(s,s-)$ are continuous functions.

Theorem 5. *Under (C5) and the assumptions in Theorem 1, $\hat{\Sigma}$ is consistent for Σ; under (C5) and assumptions in Theorem 2, $\hat{\Sigma}^*$ is consistent for Σ^*; and under (C5*) and assumptions in Theorem 3, $\hat{\Sigma}^f$ is consistent for Σ^f.*

Proof. To begin with, we have

$$\frac{\partial U_n(\beta)}{\partial\beta} = n^{-1}\sum_{i=1}^{n}\iint I\{s<t, \int_s^t dN_i(s,u)=0\}$$
$$K_n(t-s)X_i(s)\Big[-g^{'}\{X_i(s)^T\beta\}X_i(s)^T\Big]dN_i(t,s).$$

Using a similar argument to obtain equation (A.4), we show

$$\{\iint I\{s<t, \int_s^t dN(s,u)=0\}K_n(t-s)X(s)\Big[-g^{'}\{X(s)^T\beta\}X(s)^T\Big]$$

$$dN(t,s) : |\beta - \beta_0| < \epsilon\}$$

is a P-Glivenko-Cantelli class. Therefore,

$$\sup_{|\beta-\beta_0|<\epsilon} |\frac{\partial U_n(\beta)}{\partial \beta}|_{\beta=\hat{\beta}} - E\{\frac{\partial U_n(\beta)}{\partial \beta}\}|_{\beta=\hat{\beta}}| \to 0$$

in probability. Since $\hat{\beta}$ is consistent for β_0, by continuous mapping theorem, $\frac{\partial U_n(\beta)}{\partial \beta}|_{\beta=\hat{\beta}}$ converges in probability to $-A(\beta_0)$. Similarly, let

$$\hat{\Sigma}(\beta) = n^{-2} \sum_{i=1}^{n}$$

$$\Big(\iint I\{s < t, \int_s^t dN(s,u) = 0\} K_h(t-s) X_i(s) [Y_i(t) - g\{X_i(s)^T \beta\}] dN_i(t,s) \Big)^{\otimes 2}$$

then $\sup_{|\beta-\beta_0|<\epsilon} |\hat{\Sigma}(\beta) - E\{\hat{\Sigma}(\beta)\}| \to 0$ in probability. On the other hand,

$$E\{\hat{\Sigma}(\beta)\} = n^{-1}$$

$$E\Big(\iint I\{s<t, \int_s^t dN(s,u)=0\} K_h(t-s) X(s) [Y(t) - g\{X(s)^T \beta\}] dN_i(t,s) \Big)^{\otimes 2}.$$

After change of variables, and by (C0),

$$\begin{aligned}
& E\{\hat{\Sigma}(\beta)\} \\
= \ & (nh)^{-1} \iint \Big\{ K(z)^2 E\Big(X(s) \Big[\sigma^2\{s+hz, X(s+hz)\} + g\{X(s+hz)^T \beta\}^2 \\
- \ & 2g\{X(s+hz)^T \beta\} g\{X(s)^T \beta\} + g\{X(s)^T \beta\}^2 \Big] X(s)^T \lambda\{s+hz, s; X(s)\} \Big) \\
+ \ & o(h) \Big\} ds dz \\
= \ & (nh)^{-1} \Big\{ \iint K(z)^2 E[X(s) \sigma^2\{s, X(s)\} X(s)^T \lambda\{s,s;X(s)\}] dz ds + o(h) \Big\} \\
= \ & (nh)^{-1} \Sigma.
\end{aligned}$$

Therefore,

$$(nh)\hat{\Sigma} \xrightarrow{p} \Sigma \quad \text{as} \quad nh \to \infty.$$

The consistency of variance estimate follows.

Similarly, we can show that $\hat{\Sigma}^* \to \Sigma^*$ in probability and $\hat{\Sigma}^f \to \Sigma^f$ in probability.

A.4. Automatic bandwidth selection

Our method depends on the selection of bandwidth. We propose a data adaptive bandwidth selection procedure due to the fact that traditional cross-validation methods are not applicable as we have asynchronous measurement times for the covariates and response. Based on (2.7), (2.10) and (2.13), we first regress

$\hat{\beta}(h)$ on h 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 $\hat{V}(h) = \{\hat{\beta}_1(h) - \hat{\beta}_2(h)\}^2/4$. Using both $\hat{C}$ and $\hat{V}(h)$, we thus calculate the mean squared error as $\hat{C}^2h^2 + \hat{V}(h)$. Finally, we select the optimal bandwidth h minimizing this mean squared error.

A.5. Additional simulations

We report additional simulations with sample size $n = 200$ in Table 5.

The data generation for IPW is as follows. For each subject, the number of measurements is 5 for both covariate and response. Covariate observation times $s_{ik}, k = 1, \ldots, M_i$ and response observation times $t_{ij}, j = 1, \ldots, L_i$ are independently generated as $\mathcal{U}(0, 1)$. s_{ik} and t_{ij} are pooled together to have measurement times for both covariate and response. For each subject, covariate $X(t)$ is generated as a multivariate normal distribution with mean 0, variance 1 and covariance $e^{-|t_1-t_2|}$ between $X(t_1)$ and $X(t_2)$; error term $\epsilon(t)$ is generated as a multivariate normal distribution with mean 0, variance 1 and covariance $2^{-|t_1-t_2|}$ between $\epsilon(t_1)$ and $\epsilon(t_2)$. For subject i, $Y_i(t) = 1.5 + 1.5X_i(t) + \epsilon_i(t)$. We remove observations $Y_i(s_{ik}), k = 1, \ldots, M_i$ to create asynchronous longitudinal data structure. The estimating equation with IPW is

$$U_1(\beta) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{L_i} \frac{\Delta_{ij}\{Y_i(t_{ij}) - \beta_0 - X_i(t_{ij}^0)\beta_1\}}{p(\Delta_{ij} = 1)}, \tag{A.9}$$

where Δ_{ij} is the indicator function for the complete pair $(X_i(t_{ij}^0), Y_i(t_{ij}))$. For the last observation carried forward, $t_{ij}^0 = \max(x \le t_{ij}, x \in \{s_{i1}, \ldots, s_{iM_i}\})$ and for the nearest covariate, we have $t_{ij}^0 = \{x \in (s_{i1}, \ldots, s_{iM_i}), \text{such that} |x - t_{ij}| \text{is minimized}\}$. X_i singletons are missing responses. We then fit a logistic regression to the missing data variable defined in this way using covariate measurement times to estimate $p(\Delta_{ij} = 1)$. The results are summarized in Table 3. It can be seen that IPW incurs substantial bias, which does not attenuate as sample size increases and therefore should be not used to analyze asynchronous longitudinal data.

References

[1] 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

[2] Cook, R. J., Zeng, L. and Yi, G. Y. (2004). Marginal analysis of incomplete longitudinal binary data: a cautionary note on LOCF imputation. *Biometrics* **60**, 820–838. MR2101446

[3] Diggle, P., Heagerty, P., Liang, K. Y. and Zeger, S. L. (2002). *Analysis of Longitudinal Data (2nd ed.)*, Clarendon, TX: Clarendon Press. MR2049007

TABLE 5
Results of 1000 *simulations with* $n = 200$. *A: assumptions in Cao et al. (2015) hold; B: covariates follow Poisson process.*

BD	Bias	RB	SD	SE	CP
A					
weighted LOCF					
0.005	0.0117	0.0078	0.2423	0.2115	89.61
0.015	-0.0098	-0.0065	0.1424	0.1336	92.51
0.025	-0.0165	-0.0110	0.1168	0.1106	92.61
0.035	-0.0217	-0.0144	0.1055	0.0994	92.41
auto	-0.0187	-0.0124	0.2440	0.2053	88.71
half kernel					
0.005	-0.0047	-0.0031	0.2462	0.2085	87.21
0.015	-0.0097	-0.0065	0.1452	0.1349	92.91
0.025	-0.0142	-0.0095	0.1174	0.1128	93.11
0.035	-0.0188	-0.0125	0.1045	0.1013	93.51
auto	-0.0013	-0.0009	0.2087	0.1837	89.51
full kernel					
0.005	-0.0122	-0.0081	0.1754	0.1574	91.91
0.015	-0.0119	-0.0079	0.1079	0.1043	93.61
0.025	-0.0181	-0.0120	0.0940	0.0896	92.81
0.035	-0.0227	-0.0151	0.0874	0.0823	91.21
auto	-0.0040	-0.0027	0.1594	0.1458	91.01
B					
weighted LOCF					
0.05	-0.0054	-0.0036	0.0637	0.0589	92.61
0.1	-0.0054	-0.0036	0.0536	0.0507	92.21
0.15	-0.0073	-0.0048	0.0504	0.0483	93.21
0.2	-0.0097	-0.0065	0.0492	0.0475	93.51
auto	0.0014	0.0010	0.0619	0.0591	94.31
half kernel					
0.05	-0.0060	-0.0040	0.0653	0.0603	93.01
0.1	-0.0057	-0.0038	0.0560	0.0527	91.71
0.15	-0.0082	-0.0055	0.0530	0.0508	92.91
0.2	-0.0114	-0.0076	0.0519	0.0505	93.61
auto	-0.0019	-0.0012	0.0614	0.0597	93.81
full kernel					
0.05	-0.0207	-0.0138	0.0510	0.0496	92.11
0.1	-0.0452	-0.0301	0.0454	0.0454	81.92
0.15	-0.0717	-0.0478	0.0442	0.0444	63.44
0.2	-0.1012	-0.0674	0.0444	0.0444	38.56
auto	-0.0703	-0.0469	0.0466	0.0443	63.34

Note: "BD" represents different bandwidths, "Bias" is the empirical bias, "RB" is the "Bias" divided by the true β_1, "SD" is the sample standard deviation, "SE" is the average of the standard error estimates and "CP (%)" represents the coverage probability of the 95% confidence interval for $\hat{\beta}_1$.

[4] Lavori, P. W. (1992). Clinical trials in psychiatry: should protocol deviation censor patient data? *Neuropsychopharmacology* **6**, 39–48.

[5] Liang, K.-Y. and Zeger, S. L. (1986) Longitudinal data analysis using generalized linear model. *Biometrika* **73**, 13–22. MR0836430

[6] Lin, H., Scharfstein, D. O. and Rosenheck, R. A. (2004). Analysis of longitudinal data with irregular, outcome-dependent follow-up. *J. Roy. Statist. Soc. Ser. B* **66**, 791–813. MR2088782

[7] Lin, D. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *J. Amer. Statist. Assoc.* **96**, 103–113. MR1952726

[8] Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data (2nd ed.),* New York: Wiley. MR1925014

[9] Molenberghs, G., Thijs, H., Jansen, I., Beunckens, C., Kenward, M. G., Mallinckrodt, C. and Carroll, R. J. (2004). Analyzing incomplete longitudinal clinical trail data. *Biostatistics* **5**, 445–464.

[10] 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.

[11] 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.

[12] Robins, J. M., Rotnitzky, A., Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89**, 846–866. MR1294730

[13] Robins, J. M., Rotnitzky, A., Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J. Amer. Statist. Assoc.* **90**, 106–121.. MR1325118

[14] Rubin, D. (1996). Multiple imputation after 18+ years. *J. Amer. Statist. Assoc.* **91**, 473–489.

[15] 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

[16] Sun, J., Park, D-H., Sun, L. and Zhao, X. (2005). Semiparametric regression analysis of longitudinal data with informative observation times. *J. Amer. Statist. Assoc.* **100**, 882–889. MR2201016

[17] van der Vaart A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*. New York: Springer. MR1385671

[18] Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data.* New York: Springer. MR1880596

[19] 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). Cytomegalovirus viremia, mortality and cmv end-organ disease among patients with AIDS receiving potent antiretroviral therapies. *Journal of AIDS* **38**, 538–544.

[20] 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