Title

Unified Inference for Sparse and Dense Longitudinal Data in Time-varying Coefficient Models

Journal

Scandinavian Journal of Statistics, 44(1)

ISSN

0303-6898

Authors

Chen, Yixin

Yao, Weixin

Publication Date

2017-03-01

DOI

10.1111/sjos.12253

Peer reviewed

Unified Inference for Sparse and Dense Longitudinal Data in Time-Varying Coefficient Models

Yixin Chen and Weixin Yao*

Abstract

Time-varying coefficient models are widely used in longitudinal data analysis. These models allow the effects of predictors on response to vary over time. In this article, we consider a mixed-effects time-varying coefficient model to account for the within subject correlation for longitudinal data. We show that when kernel smoothing is used to estimate the smooth functions in time-varying coefficient models for sparse or dense longitudinal data, the asymptotic results of these two situations are essentially different. Therefore, a subjective choice between the sparse and dense cases might lead to erroneous conclusions for statistical inference. In order to solve this problem, we establish a unified self-normalized central limit theorem, based on which a unified inference is proposed without deciding whether the data are sparse or dense. The effectiveness of the proposed unified inference is demonstrated through a simulation study and an analysis of Baltimore MACS data.

*Weixin Yao is corresponding author, Associate Professor, Department of Statistics, University of California, Riverside, California 92521. Email: weixin.yao@ucr.edu.

Key words: Longitudinal data; Kernel smoothing; Time-varying coefficient models; Self-normalization.

1 Introduction

Longitudinal data sets arise in biostatistics and life-time testing problems when the responses of the individuals are recorded repeatedly over a period of time. Examples can be found in clinical trials, follow-up studies for monitoring disease progression, and observational cohort studies. In many longitudinal studies, repeated measurements of the response variable are collected at irregular and possibly subject-specific time points. Therefore, the measurements within each subject are possibly correlated with each other and data are often highly unbalanced, but different subjects can be assumed to be independent. Typically, the scientific interest is either in the pattern of change over time of the outcome measures or more simply in the dependence of the outcome on the covariates.

A useful nonparametric model to quantify the influence of covariates other than time is the time-varying coefficient model, in which coefficients are allowed to change smoothly over time. Let $\{(y_{ij}, \mathbf{x}_i(t_{ij}), t_{ij}); i = 1, 2, ..., n; j = 1, 2, ..., n_i\}$ be a longitudinal sample from n randomly selected subjects, where t_{ij} is the time when the jth measurement of the ith subject is made and assumed to have bounded support, n_i is the number of repeated measurements of the ith subject, y_{ij} is the response, and $\mathbf{x}_i(t_{ij}) = \mathbf{x}_{ij} = (x_i^0, x_i^1(t_{ij}), ..., x_i^k(t_{ij}))^T$ are the $(k+1)$-dimensional covariates for the ith subject at time t_{ij}, where $x_i^0 = 1$ is an optional intercept. The total number of observations in this sample is $N = \sum_{i=1}^n n_i$. The time-varying coefficient model can be written as

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta}(t_{ij}) + \epsilon_i(t_{ij}), \tag{1.1}$$

where $\boldsymbol{\beta}(t) = (\beta_0(t), \beta_1(t), ..., \beta_k(t))^T$ for all $t \geqslant 0$ are smooth functions of t, $\epsilon_i(t)$ is a

realization of a zero-mean stochastic process $\epsilon(t)$, and $\mathbf{x}_{ij}$ and ϵ_i are independent. It allows the time-varying intercept to exist when $x^0(t) \equiv 1$.

To better account for the local correlation structure of the longitudinal data, similar to the nonparametric mixed-effects model used by Wu and Zhang (2002) and Kim and Zhao (2013), we add a subject-specific random trajectory $v_i(\cdot)$ to model (1.1) and consider the following *mixed-effects time-varying coefficient model*

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta}(t_{ij}) + v_i(t_{ij}) + \sigma(t_{ij})\epsilon_{ij}, \tag{1.2}$$

where $v_i(t)$ is a realization of a mean 0 process with a covariance function $\gamma(t, t') = \text{cov}\left\{v_i(t), v_i(t')\right\} = \text{E}[v_i(t)v_i(t')]$, ϵ_{ij} are errors with $\text{E}(\epsilon_{ij}) = 0$ and $\text{E}(\epsilon_{ij}^2) = 1$, and $v_i(t)$ and ϵ_{ij} are assumed to be independent. Model (1.2) is basically the same as the model of Hoover, et al. (1998) and is a special case of the model investigated by Liang, et al. (2003) and Tian and Wu (2014) which also include subject-specific coefficients. Our primary goal in this article is to estimate the varying coefficients $\boldsymbol{\beta}(t)$ and construct confidence intervals for them.

Longitudinal data can be identified as sparse or dense according to the number of measurements within each subject. Statistical analyses for sparse or dense longitudinal data have been a subject of intense investigation in the recent ten years. Please see, for example, Yao, et al. (2005) and Ma, et al. (2012) for the studies of the sparse longitudinal data when n_i is assumed to be bounded or follows a given distribution with $\text{E}(n_i) < \infty$; and see, for example, Fan and Zhang (2000) and Zhang and Chen (2007) for the studies of the dense longitudinal data when $n_i \to \infty$. Kim and Zhao (2013) specified the sparse and dense cases clearly. Here we adopt their assumptions for the number of repeated measurements of each subject under these two scenarios:

- Sparse longitudinal data: $n_1, n_2, ..., n_n$ are independent and identically distributed positive-integer-valued random variables with $\text{E}(n_i) < \infty$;

- Dense longitudinal data: $\min_{1\leq i\leq n}(n_i) \geqslant M_n$ for some $M_n \to \infty$ as $n \to \infty$.

Other assumptions regarding the number of repeated measurements within each subject were also used to study asymptotic behaviors of local polynomial estimators in varying coefficient models. These assumptions are more or less similar to sparse or dense longitudinal data definitions described above. For example, in Hoover, et al. (1998) and Wu and Chiang (2000) the asymptotic inference was established under the assumption of $\max_{1\leq i\leq n}(n_i N^{-1}) \to 0$ as $n \to \infty$. This assumption covers the sparse longitudinal data condition defined above but does not meet the definition of dense longitudinal data. In practice, it is well known that the boundary between sparse and dense cases is not always clear. A subjective choice between sparse and dense cases may pose challenges for statistical inference. Furthermore, asymptotic properties of estimators could be different under sparse and dense assumptions. For example, as pointed out in Wu and Chiang (2000), estimators proposed in Hoover, et al. (1998) may not be consistent under the dense data setting. Li and Hsing (2010) established a uniform convergence rate for weighted local linear estimation of mean and variance functions for functional/longitudinal data. Nevertheless, Kim and Zhao (2013) showed that convergence rates and limiting variances under sparse and dense assumptions are different. This motivated them to develop unified nonparametric approaches to perform longitudinal data analysis without deciding whether the data are dense or sparse. However, Kim and Zhao (2013) only considered estimating the mean response curve without the presence of covariates.

In this article, we use the mixed-effects time-varying coefficient model (1.2) to take covariates other than time into account. The time-varying coefficient part, $\beta(t)$, in this model describes the effect of interest. The model considered by Kim and Zhao (2013) is a special case of ours if $\mathbf{x}_{ij} = 1$. We first show that when using kernel smoothing to estimate smooth functions for sparse or dense longitudinal data, asymptotic results of these two situations are essentially different. Therefore, a subjective choice between sparse and dense cases could

lead to wrong conclusions for statistical inference. In order to solve this problem, motivated by Kim and Zhao (2013), we establish a unified self-normalized central limit theorem, based on which a unified inference is proposed that can adapt to both sparse and dense cases. The resulting unified confidence interval is simple to compute and use in practice. The effectiveness of the proposed unified inference is demonstrated through a simulation study and an analysis of Baltimore MACS data.

This article is organized as follows. In Section 2, we first introduce a sample-size weighted local constant estimator of the smooth functions $\boldsymbol{\beta}(t)$ and provide asymptotic properties for both sparse and dense longitudinal data. Under the mixed-effects time-varying coefficient model setting, we then propose a unified convergence theory based on a self-normalization technique. In Section 3, we provide numerical results from a simulation study and use the Baltimore MACS data to demonstrate the performance of the proposed unified approach. Section 4 contains some discussion. Regularity conditions and proofs are assembled in the Appendix.

2 A Unified Approach for Longitudinal Data

2.1 Estimation Method

Hoover, et al. (1998) proposed a local constant fit for the time-varying coefficient model. However, they did not consider the effect of repeated measurements for each subject. Similar to Li and Hsing (2010), we consider a sample-size weighted local constant estimation method for the model (1.2). Let $f(\cdot)$ be the density function of t_{ij} and let t be an interior point of the support of $f(\cdot)$. The weighted local constant estimator we consider is

$$\hat{\boldsymbol{\beta}}(t) = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} \left[y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}(t) \right]^2 K\left(\frac{t_{ij} - t}{h_n} \right) = \mathbf{H}_n^{-1} \mathbf{g}_n, \tag{2.1}$$

where $K(\cdot)$ is a kernel function which is symmetric about 0 and satisfies $\int_{\mathbb{R}} K(u)\mathrm{d}u = 1$ and $h_n > 0$ is a bandwidth, depending on n, with

$$\mathbf{H}_n = \sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{x}_{ij}\mathbf{x}_{ij}^T K(\frac{t_{ij}-t}{h_n}), \quad \mathbf{g}_n = \sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{x}_{ij} y_{ij} K(\frac{t_{ij}-t}{h_n}). \tag{2.2}$$

Similar to the estimator considered by Kim and Zhao (2013), the above estimator does not take within-subject correlations into account for the simplicity of explanation. However the statistical inference we establish in this article takes within-subject correlations into account and is based on the model assumption (1.2). Based on Lin and Carroll (2000), the working independence kernel regression estimate $\hat{\boldsymbol{\beta}}(t)$ of (2.1) is still consistent and can achieve optimal convergence rate. However, the working independence estimate might lose some efficiency compared to many proposed methods that incorporate within-subject correlations into nonparametric regression estimator. See, for example, Fan, et al. (2007), Fan and Wu (2008), Pourahmadi (2007), Pan and Mackenzie (2003), Ye and Pan (2006), Zhang and Leng (2012), Yao and Li (2013), and Zhang et al. (2015).

2.2 Asymptotic Properties for Sparse and Dense Longitudinal Data

Based on sparse and dense cases specified in Kim and Zhao (2013), we will show that convergence rates and limiting variances of $\hat{\boldsymbol{\beta}}(t)$ are different for sparse and dense longitudinal data. To gain intuition about this, we decompose the difference between the estimated value $\hat{\boldsymbol{\beta}}(t)$ and the true value $\boldsymbol{\beta}(t)$ in the following way:

$$\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t) - \mathbf{H}_n^{-1} \sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{x}_{ij} \left[\mathbf{x}_{ij}^T \boldsymbol{\beta}(t_{ij}) - \mathbf{x}_{ij}^T \boldsymbol{\beta}(t)\right] K(\frac{t_{ij}-t}{h_n}) = \mathbf{H}_n^{-1} \sum_{i=1}^{n} \boldsymbol{\xi}_i, \tag{2.3}$$

where the asymptotic distribution of $\hat{\boldsymbol{\beta}}(t)$ is determined by the right hand side, with

$$\boldsymbol{\xi}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\boldsymbol{\xi}_{ij}, \quad \boldsymbol{\xi}_{ij} = \mathbf{x}_{ij}\left[v_i(t_{ij}) + \sigma(t_{ij})\epsilon_{ij}\right]K(\frac{t_{ij}-t}{h_n}). \tag{2.4}$$

Based on the previous definition $\gamma(t,t^{'}) = \mathrm{cov}\left\{v_i(t), v_i(t^{'})\right\} = \mathrm{E}\left[v_i(t)v_i(t^{'})\right]$, and $\mathrm{E}(\boldsymbol{\xi}_{ij}\boldsymbol{\xi}_{ij^{'}}^T) = \mathrm{E}\left\{\mathrm{E}\left(\boldsymbol{\xi}_{ij}\boldsymbol{\xi}_{ij^{'}}^T \mid t_{ij}, t_{ij^{'}}\right)\right\}$, we have, for $j \neq j^{'}$,

$$\mathrm{E}(\boldsymbol{\xi}_{ij}\boldsymbol{\xi}_{ij^{'}}^T) = \mathrm{E}\left\{\mathbf{G}(t_{ij}, t_{ij^{'}})\gamma(t_{ij}, t_{ij^{'}})K(\frac{t_{ij}-t}{h_n})K(\frac{t_{ij^{'}}-t}{h_n})\right\} \approx h_n^2\mathbf{G}(t,t)f^2(t)\gamma(t,t), \tag{2.5}$$

where $\mathbf{G}(t_{ij}, t_{ij^{'}}) = \mathrm{E}(\mathbf{x}_{ij}\mathbf{x}_{ij^{'}}^T \mid t_{ij}, t_{ij^{'}})$ and $\mathbf{G}(t,t) = \lim_{t^{'}\to t}\mathbf{G}(t,t^{'})$. Throughout this article, $a_n \approx b_n$ means that $a_n/b_n \to 1$. For the same subject and same time point,

$$\mathrm{E}(\boldsymbol{\xi}_{ij}\boldsymbol{\xi}_{ij}^T) = \mathrm{E}\left\{\boldsymbol{\Gamma}(t_{ij})\left[\gamma(t_{ij}, t_{ij}) + \sigma^2(t_{ij})\right]K^2(\frac{t_{ij}-t}{h_n})\right\} \approx \boldsymbol{\Gamma}(t)h_n f(t)\psi_k\left[\gamma(t,t) + \sigma^2(t)\right], \tag{2.6}$$

where $\boldsymbol{\Gamma}(t_{ij}) = \mathrm{E}(\mathbf{x}_{ij}\mathbf{x}_{ij}^T|t_{ij})$ and $\psi_K = \int_{\mathbb{R}} K^2(u)\mathrm{d}u$. Since

$$\mathrm{var}(\boldsymbol{\xi}_i|n_i) = n_i^{-2}\left\{\sum_{j=1}^{n_i}\mathrm{E}(\boldsymbol{\xi}_{ij}\boldsymbol{\xi}_{ij}^T) + \sum_{1\leqslant j\neq j^{'}\leqslant n_i}\mathrm{E}(\boldsymbol{\xi}_{ij}\boldsymbol{\xi}_{ij^{'}}^T)\right\},$$

then by (2.5) and (2.6), we have the following result,

$$\mathrm{var}(\boldsymbol{\xi}_i|n_i) \approx \frac{1}{n_i}\boldsymbol{\Gamma}(t)h_n f(t)\psi_K\left[\gamma(t,t) + \sigma^2(t)\right] + (1-\frac{1}{n_i})\mathbf{G}(t,t)h_n^2 f^2(t)\gamma(t,t). \tag{2.7}$$

Under the sparse assumption that n_i's are independently and identically distributed with $\mathrm{E}(n_i) < \infty$, we have, $\mathrm{var}(\boldsymbol{\xi}_i|n_i) \approx \boldsymbol{\Gamma}(t)h_n f(t)\psi_K\left[\gamma(t,t) + \sigma^2(t)\right]/n_i$ as $h_n \to 0$; under the dense assumption that $\min_{1\leq i\leq n}(n_i) \geqslant M_n$ for some $M_n \to \infty$ as $n \to \infty$, we have, $\mathrm{var}(\boldsymbol{\xi}_i|n_i) \approx \mathbf{G}(t,t)h_n^2 f^2(t)\gamma(t,t)$ with $M_n h_n \to \infty$. Therefore, limiting variances for sparse and dense cases are substantially different. We state asymptotic properties for these two

scenarios in the following theorem.

Theorem 2.1. *Let*

$$\boldsymbol{\rho}(t) = \left[\frac{\boldsymbol{\beta}^{'}(t) f^{'}(t)}{f(t)} + \frac{\boldsymbol{\beta}^{''}(t)}{2} + \boldsymbol{\Gamma}^{-1}(t) \boldsymbol{\Gamma}^{'}(t) \boldsymbol{\beta}^{'}(t) \right] \int_{\mathbb{R}} u^2 K(u) \mathrm{d}u.$$

Based on the regularity conditions in the Appendix, we have the following asymptotic results.

- *Sparse data: Assume $nh_n \to \infty$ and $sup_n nh_n^5 < \infty$. Then*

$$\sqrt{nh_n} \left[\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t) - h_n^2 \boldsymbol{\rho}(t) \right] \to N\left(\boldsymbol{0}_{k+1}, \boldsymbol{\Sigma}_{sparse}(t) \right), \tag{2.8}$$

  where $\boldsymbol{\Sigma}_{sparse}(t) = \boldsymbol{\Gamma}^{-1}(t) \psi_K \left[\gamma(t,t) + \sigma^2(t) \right] \tau / f(t)$, $\boldsymbol{0}_{k+1}$ is a $(k+1) \times 1$ vector with each entry being 0, and $\tau = E(1/n_1)$.

- *Dense data: Assume $\gamma(t,t) \neq 0$, $n_i \geq M_n$, $M_n h_n \to \infty$, $nh_n \to \infty$ and $sup_n nh_n^4 < \infty$. Then*

$$\sqrt{n} \left[\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t) - h_n^2 \boldsymbol{\rho}(t) \right] \to N\left(\boldsymbol{0}_{k+1}, \boldsymbol{\Sigma}_{dense}(t) \right), \tag{2.9}$$

  where $\boldsymbol{\Sigma}_{dense}(t) = \boldsymbol{\Gamma}^{-1}(t) \boldsymbol{G}(t,t) \gamma(t,t) \boldsymbol{\Gamma}^{-1}(t)$.

Based on Theorem 2.1, $\hat{\boldsymbol{\beta}}(t)$ has the traditional nonparametric convergence rate if the data are sparse but has root n convergence rate if the data are dense. In addition, note that if $\mathbf{x} = 1$, then Theorem 2.1 simplifies to asymptotic results provided by Kim and Zhao (2013). Based on the asymptotic normality in Theorem 2.1, confidence intervals for $\boldsymbol{\beta}(t)$ are different under sparse and dense assumptions. Let $z_{1-\alpha/2}$ be the $1 - \alpha/2$ standard normal quantile. Then an asymptotic $1 - \alpha$ confidence interval for the smooth function $\beta_l(t)$, $l = 0, \ldots, k$ is

$$\hat{\beta}_l(t) - h_n^2 \hat{\rho}_l(t) \pm z_{1-\alpha/2} (nh_n)^{-1/2} \left\{ \left[\hat{\boldsymbol{\Gamma}}^{-1}(t) \psi_K \left[\hat{\gamma}(t,t) + \hat{\sigma}^2(t) \right] \hat{\tau} / \hat{f}(t) \right]^{1/2} \right\}_{l,l} \tag{2.10}$$

for sparse data, or

$$\hat{\beta}_l(t) - h_n^2 \hat{\rho}_l(t) \pm z_{1-\alpha/2} n^{-1/2} \left\{ \left[\hat{\mathbf{\Gamma}}^{-1}(t) \hat{\mathbf{G}}(t,t) \hat{\gamma}(t,t) \hat{\mathbf{\Gamma}}^{-1}(t) \right]^{1/2} \right\}_{l,l} \tag{2.11}$$

for dense data, where $\boldsymbol{\beta}(t) = (\beta_0(t), \beta_1(t), \ldots, \beta_k(t))^T$, $\hat{\beta}_l(t)$ is the $(l+1)$th element of $\hat{\boldsymbol{\beta}}(t)$, $\hat{\rho}_l(t)$ is the $(l+1)$th element of $\hat{\boldsymbol{\rho}}(t)$ and the subscript (l,l) refers to the $(l+1)$th diagonal element of a matrix. In the above formulas, $\hat{\tau} = n^{-1} \sum_{i=1}^{n} n_i^{-1}$, $\hat{\gamma}(t,t)$, $\hat{\sigma}^2(t)$, $\hat{f}(t)$, $\hat{\rho}_l(t)$, $\hat{\mathbf{\Gamma}}^{-1}(t)$, and $\hat{\mathbf{G}}(t,t)$ are consistent estimates of τ, $\gamma(t,t)$, $\sigma^2(t)$, $f(t)$, $\rho_l(t)$, $\mathbf{\Gamma}^{-1}(t)$, and $\mathbf{G}(t,t)$. In practice, $f(t)$ can be estimated by kernel density estimate, $\hat{f}(t) = N^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n_i} K_{h_n}(t_{ij} - t)$, where $K_{h_n}(t) = h_n^{-1} K(t/h_n)$. The nonparametric mean functions $\mathbf{\Gamma}(t)$, $\gamma(t,t')$, and $\mathbf{G}(t,t')$ can be estimated by kernel smoothing methods. For example, $\hat{\Gamma}_{lm}(t) = N^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n_i} x_{ijl} x_{ijm} K_{h_n}(t_{ij} - t)$, where $\Gamma_{lm}(t)$ is the (l,m)th element of $\mathbf{\Gamma}(t)$ and x_{ijl} is the lth element of $\mathbf{x}_{ij}$. Then $\sigma(t)$ and $\boldsymbol{\rho}(t)$ can be easily estimated by noting that $\text{var}\{y(t_{ij})\} = \sigma^2(t_{ij}) + \gamma(t_{ij}, t_{ij})$.

2.3 Proposed Unified Approach

From Section 2.2, asymptotic results for sparse and dense longitudinal data are essentially different and thus a subjective choice between these two situations poses challenges for statistical inference, which motivates us to find a unified approach. In this section, we propose a unified self-normalized central limit theorem which can adapt to both sparse and dense cases for the mixed-effects time-varying coefficient model (1.2). Let $\mathbf{U}_n(t) = \mathbf{H}_n^{-1} \mathbf{W}_n \mathbf{H}_n^{-1}$, where $\mathbf{H}_n$ is defined in (2.2), and

$$\mathbf{W}_n = \sum_{i=1}^{n} \left\{ \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{x}_{ij} \left[y_{ij} - \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}(t_{ij}) \right] K(\frac{t_{ij} - t}{h_n}) \right\} \left\{ \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{x}_{ij}^T \left[y_{ij} - \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}(t_{ij}) \right] K(\frac{t_{ij} - t}{h_n}) \right\}.$$

We have the following unified central limit theorem.

Theorem 2.2. *Assume* $nh_n/\log n \to \infty$ *and* $\sup_n nh_n^5 < \infty$ *for sparse data, or* $n_i \geq M_n$, $M_n h_n \to \infty$, $nh_n^2/\log n \to \infty$ *and* $\sup_n nh_n^4 < \infty$ *for dense data. Under the regularity conditions in the Appendix,*

$$\boldsymbol{U}_n(t)^{-1/2}\left[\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t) - h_n^2\boldsymbol{\rho}(t)\right] \to N(\boldsymbol{0}_{k+1}, \boldsymbol{I}_{k+1})$$

in both sparse and the dense settings, where $\boldsymbol{I}_{k+1}$ *is the* $(k+1)\times(k+1)$ *identity matrix.*

Note that the central limit theorem proposed in Kim and Zhao (2013) is a special case of Theorem 2.2 if $\mathbf{x} = 1$ is assumed in model (1.2). Based on Theorem 2.2, a unified asymptotic pointwise $1-\alpha$ confidence interval for $\beta_l(t)$, $l = 0, \ldots, k$ can be written as follows:

$$\hat{\beta}_l(t) - h_n^2\hat{\rho}_l(t) \pm z_{1-\alpha/2}\left[\mathbf{U}_n(t)^{1/2}\right]_{l,l}. \tag{2.12}$$

The confidence intervals (2.10) and (2.11) in Section 2.2 require estimates of the within-subject covariance function $\gamma(t,t)$, the overall noise variance function $\sigma^2(t)$, and the conditional expectation $\mathbf{G}(t,t)$, which need extra smoothing procedures; but (2.12) does not need such estimates and can be used for both sparse and dense cases through the self-normalizer $\mathbf{U}_n(t)^{1/2}$.

Due to the bias term $h_n^2\rho_l(t)$, it is possible that the estimate $\hat{\beta}_l(t)$ is outside the confidence interval. Since it is difficult to estimate the bias $h_n^2\boldsymbol{\rho}(t)$ in practice due to unknown derivatives f', $\boldsymbol{\beta}'$, $\boldsymbol{\beta}''$ and $\boldsymbol{\Gamma}'$, we use the same kernel function as in Kim and Zhao (2013), $K(u) = 2G(u) - G(u/\sqrt{2})/\sqrt{2}$, where $G(u)$ is the standard normal density. Then $\int_{\mathbb{R}} u^2K(u)\mathrm{d}u = 0$ and therefore $\boldsymbol{\rho}(t) = \mathbf{0}_{k+1}$. This obviously does not solve the bias problem. For instance, if f, $\boldsymbol{\beta}$ and $\boldsymbol{\Gamma}$ are four times differentiable, then we have the higher order bias term $O(h_n^4)$. As Kim and Zhao (2013) stated, the bias problem is an inherently difficult problem and no good solutions exist so far. Our simulation results in Section 3.1 demonstrate that the new proposed self-normalized confidence interval works well.

For kernel regression, the selection of bandwidth is generally more important than the selection of kernel functions. As stated in Wu and Chiang (2000), under-smoothing or over-smoothing is mainly caused by inappropriate bandwidth choices in practice, but is rarely influenced by kernel shapes. The asymptotic optimal bandwidth depends on n and n_i and it should be able to balance the asymptotic bias term $h_n^2\boldsymbol{\rho}(t)$ and the asymptotic variance term $\mathbf{U}_n(t)$. However, as proved in Theorem 2.1, asymptotic properties of the variance term $\mathbf{U}_n(t)$ depend on whether the data are dense or sparse and how n_i increases with n. Therefore, it is not easy to derive a unified asymptotic optimal bandwidth. To select the bandwidth for $\hat{\boldsymbol{\beta}}$ in practice, we use the idea of "leave-one-subject-out" cross-validation procedure suggested by Rice and Silverman (1991). Let $\hat{\boldsymbol{\beta}}_{-i}(t)$ be a kernel estimator of $\boldsymbol{\beta}(t)$ computed using the data with all repeated measurements of the ith subject left out, and define

$$\mathrm{CV}(h_n) = \sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} \left\{ y_{ij} - \mathbf{x}_{ij}^T \hat{\boldsymbol{\beta}}_{-i}(t_{ij}) \right\}^2 \tag{2.13}$$

to be the subject-based cross-validation. The optimal bandwidth is then defined to be the unique minimizer of $\mathrm{CV}(h_n)$. Based on Remark 2.3 of Wu and Chiang (2000), the above CV bandwidth approximately minimizes the following average squared error

$$\mathrm{ASE}(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} \left[\mathbf{x}_{ij}^T \left\{ \boldsymbol{\beta}(t_{ij}) - \hat{\boldsymbol{\beta}}(t_{ij}) \right\} \right]^2.$$

3 Simulation and Real Data Application

3.1 Simulation Study

We follow Kim and Zhao (2013) to construct the subject-specific random trajectory $v_i(\cdot)$. Consider the model

$$y_{ij} = \sum_{l=0}^{2} \beta_l(t_{ij}) x_{ijl}(t_{ij}) + \sum_{m=1}^{3} \alpha_{im} \Phi_m(t_{ij}) + \sigma \epsilon_{ij}, \quad i = 1, ..., n;\ j = 1, ..., n_i,$$

where $\alpha_{im} \sim N(0, \omega_m)$ and $\epsilon_{ij} \sim N(0, 1)$. Let $\beta_0(t) = 5(t - 0.6)^2$, $\beta_1(t) = cos(3\pi t)$, $\beta_2(t) = sin(2\pi t)$, $\Phi_1(t) = 1$, $\Phi_2(t) = \sqrt{2}sin(2\pi t)$, $\Phi_3(t) = \sqrt{2}cos(2\pi t)$, $(\omega_1, \omega_2, \omega_3) = (0.6, 0.3, 0.1)$, and $\sigma = 1$. Then the variance function $\gamma(t, t) = 0.6 + 0.6sin^2(2\pi t) + 0.2cos^2(2\pi t)$. The time points t_{ij} are uniformly distributed on $[0, 1]$. To generate covariates, let $b_{i1} \sim N(0, 0.3)$, $b_{i2} \sim N(0, 0.3)$, $\eta_{ij} \sim N(0, 1)$, $\delta_{ij} \sim N(0, 1)$ and $\varphi(t) = \sqrt{2}(t + 1)$, then set $x_{ij0} = 1$, $x_{ij1} = b_{i1}\varphi(t_{ij}) + \eta_{ij}$ and $x_{ij2} = b_{i2}\varphi(t_{ij}) + \delta_{ij}$ for $i = 1, ..., n$ and $j = 1, ..., n_i$. We consider two sample sizes, $n = 200$ or 400. Under this setting, we have the following conditional expectations: $\mathbf{\Gamma}(t_{ij}) = \mathrm{E}(\mathbf{x}_{ij}\mathbf{x}_{ij}^T \mid t_{ij}) = \mathrm{diag}\{1, 0.6(t_{ij} + 1)^2 + 1, 0.6(t_{ij} + 1)^2 + 1\}$ and

$$\mathbf{G}(t_{ij}, t_{ij}) = \lim_{t_{ij'} \to t_{ij}} \mathrm{E}(\mathbf{x}_{ij}\mathbf{x}_{ij'}^T \mid t_{ij}, t_{ij'}) = \mathrm{diag}\{1, 0.6(t_{ij} + 1)^2, 0.6(t_{ij} + 1)^2\}.$$

For the vector $(n_1, n_2, ..., n_n)$ of the number of repeated measurements on each subject, we consider four cases

$$N_1: \quad n_i \sim U[\{5, 6, ..., 15\}]; \qquad N_2: \quad n_i \sim U[\{15, 16, ..., 35\}]; \tag{3.1}$$

$$N_3: \quad n_i \sim U[\{80, 81, ..., 120\}]; \qquad N_4: \quad n_i \sim U[\{150, 151, ..., 250\}]. \tag{3.2}$$

Here $U[D]$ represents the discrete uniform distribution on a finite set D. Five confidence intervals are compared in our simulation study:

1. the self-normalization based confidence interval in (2.12) (SN);

2. the asymptotic normality based confidence interval (2.10) for sparse data (NS);

3. the asymptotic normality based confidence intervals (2.11) for dense data (ND);

4. the bootstrap confidence interval with 200 bootstrap replications from sampling subjects with replacement (BS);

5. the infeasible confidence interval (NSD)

$$\hat{\beta}_l(t) - h_n^2 \hat{\rho}_l(t) \pm z_{1-\alpha/2} n^{-1/2} \mathbf{S}_{l,l}, \tag{3.3}$$

where $\mathbf{S} = \left\{ \mathbf{\Gamma}^{-1}(t) \mathbf{G}(t,t) \mathbf{\Gamma}^{-1}(t) (1-\hat{\tau}) \gamma(t,t) + \mathbf{\Gamma}^{-1}(t) \hat{\tau} \psi_K \left[\gamma(t,t) + \sigma^2(t) \right] / \left[h_n f(t) \right] \right\}^{1/2}$.

The confidence interval NSD is used as a benchmark to compare the performance of the other confidence intervals, since NSD uses the true theoretical limiting variance function (2.7). Note, however, that NSD is practically infeasible, since it depends on many unknown functions. Similar to Kim and Zhao (2013), we use the true functions $\gamma(t,t)$, $\sigma^2(t)$, $f(t)$, $\mathbf{\Gamma}(t)$, and $\mathbf{G}(t,t)$ for NS, ND, and NSD, which gives an advantage to these three methods and removes the impact of different estimation methods. Note that the proposed self-normalization based confidence interval SN only requires a point estimate of $\boldsymbol{\beta}(t)$ and thus is very easy to implement. We would like to demonstrate that our new method SN works comparably or better than NS and ND even when true functions are used for NS and ND.

To measure the performance of different confidence intervals, we use the following two criteria: empirical coverage probabilities and lengths of confidence intervals. Let $t_1 < \cdots < t_{20}$ be 20 grid points evenly spaced on $[0.1, 0.9]$. For each grid point t_j $(j = 1, \ldots, 20)$ and a given confidence level, we construct confidence intervals for smooth functions $\beta_0(t_j)$, $\beta_1(t_j)$, and $\beta_2(t_j)$, and compute the empirical coverage probabilities based on 500 replications. Using 500 replications is restricted by the computing time based on a personal computer

with Intel(R) Core(TM) i5 CPU, 4GB installed memory, and 32-bit operating system. For each of the five confidence intervals, the empirical coverage probabilities and lengths are averaged at 20 grid points. The bandwidth used for each replicate is the average of 20 optimal bandwidths in (2.13) based on 20 replications (Kim and Zhao, 2013).

The results for $\beta_1(t)$ and $\beta_2(t)$ are showed in Table 1 to 4. The bandwidths for N_1, N_2, N_3 and N_4 when $n = 200$ are 0.0548, 0.0471, 0.0359 and 0.0334, respectively. The bandwidths for N_1, N_2, N_3 and N_4 when $n = 400$ are 0.0498, 0.0428, 0.0327 and 0.0273, respectively. It can easily be seen that the performance of the confidence intervals NS and ND for $\beta_1(t)$ and $\beta_2(t)$ strongly depends on the spareness or denseness of the data for both sample sizes. When the number of repeated measurements on each subject is increased from the sparse setting N_1 to the dense setting N_4, the performance of the confidence interval NS assuming the sparse data becomes worse, while the confidence interval ND assuming the dense data becomes better. These two confidence intervals only perform well under their corresponding sparse or dense setting, which further confirms the theoretical results in Theorem 2.1.

Note that the confidence interval ND assuming dense data gives same widths for each simulation setting at a certain nominal level. This is because asymptotic variances at 20 grid points assuming dense data are the same for each simulation setting. In addition, since we use the same way to generate two covariates x_{ij1} and x_{ij2}, the diagonal elements in $\mathbf{\Gamma}(t)$ and $\mathbf{G}(t,t)$ corresponding to $\beta_1(t)$ and $\beta_2(t)$ in (2.10), (2.11), and (3.3) are the same at a given grid point. Hence the widths of the confidence intervals of $\beta_1(t)$ and $\beta_2(t)$ are the same for NSD, NS, and ND .

Compared to NS and ND, the proposed self-normalization based confidence interval SN provides more stable and better performance. First, it has similar widths and coverage probabilities as the bootstrap confidence interval (BS) and both of them perform closely to the infeasible confidence interval NSD; second, its computing time is much shorter than for the bootstrap confidence interval; finally, asymptotic properties of the self-normalization

method have been established in this article, whereas theoretical properties of the bootstrap procedure for longitudinal data have not been developed as far as we know. We also did simulation studies on some larger sample sizes, for example, $n = 3000$, and the proposed self-normalized method still works very well and performs better than sparse and dense intervals and has similar performance to the bootstrap method under all cases we tried.

3.2 Application to Baltimore MACS Data

In this section, we apply the self-normalization based confidence interval to the HIV part of the Baltimore MACS data which came from the Baltimore MACS Public Data Set Release PO4 (1984 - 1991) provided by Dr. Alfred Saah. CD4 cells can be destroyed by human immune-deficiency virus(HIV) and thus the percentage of the CD4 cells in the blood of a human body will change after HIV infection. Because of this, CD4 cell count and the percentage in the blood are the most popular used markers to monitor the progression of the disease.

The HIV status of 283 homosexual men who were infected with HIV during the follow-up period between 1984 and 1991 was included in this data set. All individuals were scheduled to have measurements made twice a year. Since many patients missed some of their scheduled visits and HIV infections happened randomly during the study, numbers of repeated measurements for each patient are not equal and their measurement times are different. Further details about the design, methods, and medical implications of the study can be found in Kaslow, et al. (1987).

The response variable is the CD4 percentage over time after HIV infection. Three covariates are: patient's age, smoking status with 1 as smoker and 0 as nonsmoker, and the CD4 cell percentage before the infection. The aim of our statistical analysis is to evaluate the effects of smoking, pre-HIV infection CD4 percentage, and age at HIV infection on the mean CD4 percentage after the infection. Define t_{ij} to be the time (in years) of the jth mea-

surement of the ith individual after HIV infection. In this data set, patients have minimum 1 and maximum 14 measurements. Let Y_{ij} be the ith individual's CD4 percentage at time t_{ij} and X_{1i} be the smoking status for the ith individual. We center age and pre-infection CD4 percentage using the sample average. Then we construct the time-varying coefficient model as follows:

$$Y_{ij} = \beta_0(t_{ij}) + \beta_1(t_{ij})X_{1i} + \beta_2(t_{ij})X_{2i} + \beta_3(t_{ij})X_{3i} + \epsilon_{ij},$$

where $\beta_0(t)$ represents the baseline CD4 percentage and can be interpreted as the mean CD4 percentage at time t for a nonsmoker with average pre-infection CD4 percentage and average age at HIV infection. Therefore, $\beta_1(t)$, $\beta_2(t)$, and $\beta_3(t)$ represent time-varying effects for smoking, age at HIV infection, and pre-infection CD4 percentage, respectively, on the post-infection CD4 percentage at time t.

We use the kernel smoothing method stated in (2.1) to estimate smoothing functions $\beta_0(t)$, $\beta_1(t)$, $\beta_2(t)$, and $\beta_3(t)$. The bandwidth was chosen by using the leave-one-subject-out cross-validation method, and its value is 0.7074. This real data set is most likely to be the sparse case. However, based on Table 1 to 4, even for the case of N_1, the proposed self-normalized method (SN) provides better confidence interval than the sparse confidence interval (NS). In addition, as we discussed before, the sparse confidence interval (2.10) requires estimates of many unknown quantities and some of them are not easy to estimate, while the self-normalized confidence interval (2.12) does not require any additional estimates besides the estimates of regression coefficients. Therefore, self-normalization based 95% confidence intervals were constructed for $\beta_0(t), \ldots, \beta_3(t)$ at 100 equally spaced time points between 0.1 and 5.9 years. We also constructed bootstrap 95% confidence intervals at the same 100 time points, based on 1000 bootstrap replications. Figure 1 depicts fitted coefficient functions (solid curves) with 95% self-normalization based confidence intervals (dashed curves) and bootstrap confidence intervals (dotted curves). It can easily be seen that self-normalization

based confidence intervals are very close to bootstrap confidence intervals. Indeed, they almost overlap with each other. However, the computing time for the self-normalization based confidence interval is much shorter than the bootstrap confidence interval. The former one only takes approximately 5 seconds, whereas the latter one needs almost 50 minutes based on a personal computer with Intel(R) Core(TM) i5 CPU, 4GB installed memory, and 32-bit operating system.

Based on the constructed confidence intervals, the mean baseline CD4 percentage of the population decreases with time, but at a rate that appears to be slowing down at four years after the infection. Since confidence intervals for smoking and age of HIV infection cover 0 most of the time, these two covariates do not significantly affect the post-infection CD4 percentage. The pre-infection CD4 percentage appears to be positively associated with higher post-infection CD4 percentage, which is expected. The above findings basically agree with Wu and Chiang (2000), Fan and Zhang (2000), and Huang, et al. (2002), and Qu and Li (2006).

4 Discussion

In this article, we proposed a unified inference for the time-varying coefficient model (1.2) for the longitudinal data based on the new established unified self-normalized central limit theorem. The new inference tool allows us to do inference for the longitudinal data without subjectively deciding whether the data are sparse or dense. The effectiveness of the proposed unified inference is demonstrated through a simulation study and an analysis of Baltimore MACS data. However, we want to point out that our method only unifies the inference of the sparse and dense situations discussed in our article. It requires more research to provide a unified inference that is applicable to all cases.

The weighted local constant estimators that we considered in this article only use one

smoothing parameter, which may not be able to provide adequate smoothing for all coefficient curves at the same time. Wu and Chiang (2000) proposed the componentwise local least squares criteria to estimate time-varying coefficients using different amounts of smoothing. The reason that we use one smoothing parameter is for the simplicity of computation and our proposed unified inference can be extended to the case of different smoothing parameters as well.

For time-varying coefficient models, commonly asked questions are whether coefficient functions $\boldsymbol{\beta}(\cdot)$ vary over time and whether certain covariates are significant. Therefore, we may wish to test whether a certain component of $\boldsymbol{\beta}(\cdot)$ is identically zero or constant. The generalized likelihood ratio statistics for the nonparametric testing problems proposed in Fan, et al. (2001) might be considered, but the theoretical and practical aspects for longitudinal data would require substantial development.

5 Acknowledgments

We would like to thank an Associate Editor and two referees for their helpful comments and suggestions which have greatly improved the presentation of this paper. We are also grateful to Dr. Alfred Saah for providing the Baltimore MACS Public Data Set Release PO4 (1984-1991). Yao's research is supported by NSF grant DMS-1461677.

Appendix

The following conditions are imposed to facilitate the proof and are adopted from Wu and Chiang (2000), Huang, et al. (2002), and Kim and Zhao (2013).

Regularity Conditions:

1. The observation time points follow a random design in the sense that t_{ij}, for $j = 1, ..., n_i$ and $i = 1, ..., n$, are chosen independently from an unknown distribution with a density

$f(\cdot)$ on a finite interval. The density function $f(\cdot)$ is continuously differentiable in a neighborhood of t and is uniformly bounded away from 0 and infinity.

2. In a neighborhood of t, $\boldsymbol{\beta}(\cdot)$ is twice continuously differentiable, $\sigma^2(\cdot)$ is continuously differentiable. In a neighborhood of (t,t), $\gamma(t,t^{'}) = \text{cov}\{v_i(t), v_i(t^{'})\}$ is continuously differentiable and $\gamma(t,t) = \lim_{t^{'}\to t}\text{cov}\{v_i(t), v_i(t^{'})\}$. Furthermore, $\sigma^2(t) < \infty$ and $\gamma(t,t) < \infty$.

3. $\{v_i(\cdot)\}_i$, $\{t_{ij}\}_{ij}$, $\{\epsilon_{ij}\}_{ij}$ are independent and identically distributed and mutually independent.

4. $\{\mathbf{x}_{ij}\}_{ij}$, $\{v_i(\cdot)\}_i$, $\{\epsilon_{ij}\}_{ij}$ are mutually independent. $\{\mathbf{x}_{ij}\}_i$ are independent and identically distributed. For the same i, $\mathbf{x}_{i1}, ..., \mathbf{x}_{in_i}$ have identical distribution and can be correlated. $\text{E}\left[\|\mathbf{x}_{ij}\| \cdot \left\|\mathbf{x}_{ij^{'}}\right\| \cdot \left\|\mathbf{x}_{ij^{''}}\right\| |t_{ij}, t_{ij^{'}}, t_{ij^{''}}\right] < \infty$ for $1 \leqslant j \neq j^{'} \neq j^{''} \leqslant n_i$.

5. $\boldsymbol{\Gamma}(t)$ is invertible and differentiable.

6. $\text{E}\{|v_i(\cdot) + \sigma(\cdot)\epsilon_{ij}|^4\}$ is continuous in a neighborhood of t and $\text{E}\{|v_i(\cdot) + \sigma(\cdot)\epsilon_{ij}|^4\} < \infty$.

7. $K(\cdot)$ is bounded, symmetric, and has bounded support and bounded derivative.

Since $\sigma^2(t)$ and $\gamma(t,t)$ are unknown in most applications and the unified approach that we propose does not need the specific structures of $\sigma^2(t)$ and $\gamma(t,t)$, therefore, we do not require further specific structures for $\sigma^2(t)$ and $\gamma(t,t)$, except for their continuity in the above condition 2. These conditions are not weakest possible conditions. For instance, in condition 7, actually we only need the first moment of $K(\cdot)$ to be 0 so that the bias term containing the first order of h will be 0. $K(\cdot)$ is allowed to be negative. The symmetry assumption is traditionally used for kernel function and will automatically satisfy the condition of zero first moment. In addition, the requirement for bounded support for the kernel $K(\cdot)$ could be relaxed as well. All asymptotic results still hold if we put a restriction on the tail of $K(\cdot)$.

For example, $\limsup_{t\to\infty}|K(t)t^5| < \infty$ (Fan and Gijbels, 1992). In the following, without confusing, we will omit the subscript n of h_n for the simplicity of notation.

Proof of Theorem 2.1. Based on (2.3), asymptotic results for sparse or dense longitudinal data depend on the limiting distribution of $\boldsymbol{\xi}_i$ which is defined in (2.4). In order to obtain the limiting distribution of $\boldsymbol{\xi}_i$, we define the following notation.

$$\mathbf{H}_n = \sum_{i=1}^{n} \mathbf{V}_i, \qquad \mathbf{V}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \mathbf{V}_{ij}, \qquad \mathbf{V}_{ij} = \mathbf{x}_{ij}\mathbf{x}_{ij}^T K(\frac{t_{ij}-t}{h}),$$

$$\mathbf{b}_n = \sum_{i=1}^{n} \boldsymbol{\zeta}_i, \quad \boldsymbol{\zeta}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \boldsymbol{\zeta}_{ij}, \quad \boldsymbol{\zeta}_{ij} = \mathbf{x}_{ij}\left[\mathbf{x}_{ij}^T\boldsymbol{\beta}(t_{ij}) - \mathbf{x}_{ij}^T\boldsymbol{\beta}(t)\right] K(\frac{t_{ij}-t}{h}),$$

$$\boldsymbol{\Gamma}(t_{ij}) = \mathrm{E}(\mathbf{x}_{ij}\mathbf{x}_{ij}^T|t_{ij}), \quad \Gamma_1(t_{ij}) = \mathrm{E}(x_{ijl}^2 x_{ijr}^2|t_{ij}), \quad \boldsymbol{\Gamma}_2(t_{ij}) = \mathrm{E}(X_{ijm}^2\mathbf{x}_{ij}\mathbf{x}_{ij}^T|t_{ij}),$$

where $l, r, m = 0, ..., (k+1)$. Throughout this article, we consider the element-wise variance of a matrix. Based on regularity conditions 1, 2, 3, 4, 5, 7, Taylor's expansion and symmetry of the kernel function $K(\cdot)$, we have the following results,

$$\begin{aligned}
\mathrm{E}(\mathbf{V}_{ij}) =& \mathrm{E}\left\{\mathrm{E}(\mathbf{V}_{ij}|t_{ij})\right\} = \mathrm{E}\left\{\mathrm{E}(\mathbf{x}_{ij}\mathbf{x}_{ij}^T|t_{ij})K(\frac{t_{ij}-t}{h})\right\} \\
=& h\int \left[\boldsymbol{\Gamma}(t) + \boldsymbol{\Gamma}^{'}(t)ht_0 + o(h)\right] K(t_0)\left[f(t) + f^{'}(t)ht_0 + o(h)\right] \mathrm{d}t_0 \\
=& \boldsymbol{\Gamma}(t)hf(t)\left[1 + O(h^2)\right],
\end{aligned}$$

and

$$\begin{aligned}\text{var}(\mathbf{V}_{ij}(l,r)) &= \text{var}\left(x_{ijl}x_{ijr}K(\frac{t_{ij}-t}{h})\right)\\ &= \text{E}\left\{\left[x_{ijl}x_{ijr}K(\frac{t_{ij}-t}{h})\right]^2\right\} - \left\{\text{E}\left[x_{ijl}x_{ijr}K(\frac{t_{ij}-t}{h})\right]\right\}^2\\ &= \text{E}\left[\Gamma_1(t_{ij})K^2(\frac{t_{ij}-t}{h})\right] - O(h^2) = h\Gamma_1(t)f(t)\psi_K + o(h) - O(h^2) = O(h),\end{aligned}$$

where (l,r) refers to the element of $\mathbf{V}_{ij}$ in the lth row and rth column. Therefore, $\text{var}(\mathbf{V}_{ij}) = O(h)$. Similarly, we have the following results for $\boldsymbol{\zeta}_{ij}$,

$$\begin{aligned}\text{E}(\boldsymbol{\zeta}_{ij}) &= \text{E}\left\{\text{E}\left\{\mathbf{x}_{ij}\left[\mathbf{x}_{ij}^T\boldsymbol{\beta}(t_{ij}) - \mathbf{x}_{ij}^T\boldsymbol{\beta}(t)\right]K(\frac{t_{ij}-t}{h})|t_{ij}\right\}\right\}\\ &= \text{E}\left\{\boldsymbol{\Gamma}(t_{ij})\left[\boldsymbol{\beta}(t_{ij}) - \boldsymbol{\beta}(t)\right]K(\frac{t_{ij}-t}{h})\right\}\\ &= h^3 f(t)\boldsymbol{\Gamma}(t)\left[\frac{\boldsymbol{\beta}^{'}(t)f^{'}(t)}{f(t)} + \frac{\boldsymbol{\beta}^{''}(t)}{2} + \boldsymbol{\Gamma}^{-1}(t)\boldsymbol{\Gamma}^{'}(t)\boldsymbol{\beta}^{'}(t)\right]\int t_0^2K(t_0)\text{d}t_0 + o(h^3)\\ &= \boldsymbol{\Gamma}(t)h^3f(t)\boldsymbol{\rho}(t) + o(h^3),\end{aligned}$$

and

$$\begin{aligned}\text{var}(\zeta_{ijm}) &= \text{var}\left\{x_{ijm}\left[\mathbf{x}_{ij}^T\boldsymbol{\beta}(t_{ij}) - \mathbf{x}_{ij}^T\boldsymbol{\beta}(t)\right]K(\frac{t_{ij}-t}{h})\right\}\\ &= \text{E}\left\{\text{E}\left\{\left[x_{ijm}\mathbf{x}_{ij}^T\left[\boldsymbol{\beta}(t_{ij}) - \boldsymbol{\beta}(t)\right]K(\frac{t_{ij}-t}{h})\right]^2|t_{ij}\right\}\right\} - \left[O(h^3)\right]^2\\ &= \int\left[\boldsymbol{\beta}(t_{ij}) - \boldsymbol{\beta}(t)\right]^T\boldsymbol{\Gamma}_2(t_{ij})\left[\boldsymbol{\beta}(t_{ij}) - \boldsymbol{\beta}(t)\right]K^2(\frac{t_{ij}-t}{h})f(t_{ij})\text{d}t_{ij} - O(h^6)\\ &= O(h^3),\end{aligned}$$

where $\boldsymbol{\rho}(t) = \left[\frac{\boldsymbol{\beta}^{'}(t)f^{'}(t)}{f(t)} + \frac{\boldsymbol{\beta}^{''}(t)}{2} + \boldsymbol{\Gamma}^{-1}(t)\boldsymbol{\Gamma}^{'}(t)\boldsymbol{\beta}^{'}(t)\right]\int_{\mathbb{R}}u^2K(u)\text{d}u$, ζ_{ijm} and x_{ijm} are the m^{th} elements of $\boldsymbol{\zeta}_{ij}$ and $\mathbf{x}_{ij}$, respectively. Therefore, $\text{var}(\boldsymbol{\zeta}_{ij}) = O(h^3)$. In both sparse and

the dense cases, $\mathrm{E}(\mathbf{V}_i|n_i)$ is not random, so we have $\mathrm{var}(\mathbf{V}_i) = \mathrm{E}\left\{\mathrm{var}(\mathbf{V}_i|n_i)\right\} \leqslant \mathrm{var}(\mathbf{V}_{ij})$. Therefore, $\mathrm{var}(\mathbf{H}_n) = O(nh)$. Then

$$\mathbf{H}_n = \mathrm{E}(\mathbf{H}_n) + O_p\left(\sqrt{\mathrm{var}(\mathbf{H}_n)}\right) = \left[1 + O_p\left\{h^2 + \frac{1}{\sqrt{nh}}\right\}\right] n\mathbf{\Gamma}(t)hf(t).$$

Similarly, $\mathbf{b}_n = n\mathbf{\Gamma}(t)h^3 f(t)\boldsymbol{\rho}(t) + o_p(nh^3) + O_p(\sqrt{nh^3})$. Hence,

$$\mathbf{H}_n^{-1}\mathbf{b}_n = \frac{\mathbf{\Gamma}^{-1}(t)\left[n\mathbf{\Gamma}(t)h^3 f(t)\boldsymbol{\rho}(t) + o_p(nh^3) + O_p(\sqrt{nh^3})\right]}{\left[1 + O_p(h^2 + \sqrt{\frac{1}{nh}})\right] nhf(t)} = h^2\boldsymbol{\rho}(t) + \boldsymbol{\delta}_n,$$

where $\boldsymbol{\delta}_n = o_p(h^2) + O_p(\sqrt{\frac{h}{n}})$.

For dense longitudinal data, under the given conditions we have $\boldsymbol{\delta}_n = o_p(1/\sqrt{n})$ and $[nh^2 f^2(t)]^{-1}\,\mathrm{var}(\sum_{i=1}^n \boldsymbol{\xi}_i) \approx \mathbf{G}(t,t)\gamma(t,t)$. For any unit vector $\mathbf{d} \in \mathbb{R}^{k+1}$, let $\mathbf{d}^T\sum_{i=1}^n \boldsymbol{\xi}_i = \sum_{i=1}^n \mathbf{d}^T\boldsymbol{\xi}_i = \sum_{i=1}^n \boldsymbol{\theta}_i$, where $\boldsymbol{\theta}_i = \mathbf{d}^T\boldsymbol{\xi}_i$. Then we have $\mathrm{E}(\boldsymbol{\theta}_i^2) = O(h^2)$ and $\mathrm{E}(\boldsymbol{\theta}_i^3) = O(h^3)$ based on the regularity condition 4. By the Lyapunov central limit theorem, $\frac{\sum_{i=1}^n \boldsymbol{\xi}_i}{h\sqrt{n}f(t)} \to N(\mathbf{0}_{k+1}, \mathbf{G}(t,t)\gamma(t,t))$. Similarly, for sparse longitudinal data, since $\boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_n$ are independent and identically distributed, the result follows from $\boldsymbol{\delta}_n = o_p(1/\sqrt{nh})$ and $\mathrm{var}(\sum_{i=1}^n \boldsymbol{\xi}_i) \approx nh\tau\psi_K f(t)[\gamma(t,t) + \sigma^2(t)]\mathbf{\Gamma}(t)$. $\square$

Proof of Theorem 2.2. Based on Theorem 2.1, if we can show $n\mathbf{U}_n(t) \to \mathbf{\Sigma}_{\text{dense}}(t)$ and $nh\mathbf{U}_n(t) \to \mathbf{\Sigma}_{\text{sparse}}(t)$, the Theorem 2.2 can be proved.

Denote $K_{ij} = K(\frac{t_{ij}-t}{h})$. Let

$$\begin{aligned}
\mathbf{W}_n &= \sum_{i=1}^n \left\{\frac{1}{n_i}\sum_{j=1}^{n_i} \mathbf{x}_{ij}\left[y_{ij} - \mathbf{x}_{ij}^T\hat{\boldsymbol{\beta}}(t_{ij})\right]K_{ij}\right\}\left\{\frac{1}{n_i}\sum_{j=1}^{n_i} \mathbf{x}_{ij}^T\left[y_{ij} - \mathbf{x}_{ij}^T\hat{\boldsymbol{\beta}}(t_{ij})\right]K_{ij}\right\} \\
&= \sum_{i=1}^n \left(\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T + \boldsymbol{\xi}_i\boldsymbol{\alpha}_i^T + \boldsymbol{\alpha}_i\boldsymbol{\xi}_i^T + \boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^T\right),
\end{aligned}$$

where $\boldsymbol{\xi}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\mathbf{x}_{ij}\left[v_i(t_{ij})+\sigma(t_{ij})\epsilon_{ij}\right]K_{ij}$, and $\boldsymbol{\alpha}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\mathbf{x}_{ij}\left[\mathbf{x}_{ij}^T\boldsymbol{\beta}(t_{ij})-\mathbf{x}_{ij}^T\hat{\boldsymbol{\beta}}(t_{ij})\right]K_{ij}$. Similarly as Kim and Zhao (2013), by Theorem 3.1 in Li and Hsing (2010), $\left|\hat{\boldsymbol{\beta}}(z)-\boldsymbol{\beta}(z)\right| = O_p(l_n)\mathbf{1}_{k+1}$ uniformly for z in the neighborhood of t, where $l_n = h^2+\sqrt{\frac{\log n}{n}}$ for dense data, $l_n = h^2+\sqrt{\frac{\log n}{nh}}$ for sparse data and $\mathbf{1}_{k+1}$ is a $(k+1)\times 1$ vector with all elements equal to 1. Then $\boldsymbol{\alpha}_i = O_p(|\boldsymbol{\alpha}_i|) = O_p(l_n)\frac{1}{n_i}\sum_{j=1}^{n_i}\left|\mathbf{x}_{ij}\mathbf{x}_{ij}^T\mathbf{1}_{k+1}K_{ij}\right|$. Since $\boldsymbol{\xi}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\boldsymbol{\xi}_{ij}$ which is defined in (2.4), we can get

$$\begin{aligned}\sum_{i=1}^{n}\left|\boldsymbol{\xi}_i\boldsymbol{\alpha}_i^T+\boldsymbol{\alpha}_i\boldsymbol{\xi}_i^T+\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^T\right| &= O_p(l_n)\sum_{i=1}^{n}\frac{1}{n_i^2}\sum_{j=1}^{n_i}\left|\boldsymbol{\xi}_{ij}\right|\sum_{j=1}^{n_i}\left|\mathbf{x}_{ij}^T(\mathbf{x}_{ij}^T\mathbf{1}_{k+1})K_{ij}\right| \\ &\quad +O_p(l_n)\sum_{i=1}^{n}\frac{1}{n_i^2}\sum_{j=1}^{n_i}\left|\mathbf{x}_{ij}\mathbf{x}_{ij}^T\mathbf{1}_{k+1}K_{ij}\right|\sum_{j=1}^{n_i}\left|\boldsymbol{\xi}_{ij}^T\right| \\ &\quad +O_p(l_n^2)\sum_{i=1}^{n}\frac{1}{n_i^2}\sum_{j=1}^{n_i}\left|\mathbf{x}_{ij}\mathbf{x}_{ij}^T\mathbf{1}_{k+1}K_{ij}\right|\sum_{j=1}^{n_i}\left|\mathbf{x}_{ij}^T(\mathbf{x}_{ij}^T\mathbf{1}_{k+1})K_{ij}\right|.\end{aligned}$$

Based on the proof of Theorem 2.1,

$$\boldsymbol{\xi}_{ij} = \mathrm{E}(\boldsymbol{\xi}_{ij})+O_p(\sqrt{\mathrm{var}(\boldsymbol{\xi}_{ij})}) = O_p(\sqrt{\mathrm{E}(\boldsymbol{\xi}_{ij}\boldsymbol{\xi}_{ij}^T)}) = O_p(\sqrt{h}),$$

$$\mathbf{x}_{ij}\mathbf{x}_{ij}^TK_{ij} = \mathbf{V}_{ij} = \mathrm{E}(\mathbf{V}_{ij})+O_p(\sqrt{\mathrm{var}(\mathbf{V}_{ij})}) = O_p(h)+O_p(\sqrt{h}) = O_p(\sqrt{h}).$$

Since $\mathbf{x}_{ij}\mathbf{x}_{ij}^T\mathbf{1}_{k+1}K_{ij} = \mathbf{x}_{ij}\mathbf{x}_{ij}^TK_{ij}\mathbf{1}_{k+1}$, $\mathbf{x}_{ij}^T(\mathbf{x}_{ij}^T\mathbf{1}_{k+1})K_{ij} = \mathbf{1}_{k+1}^T\mathbf{x}_{ij}\mathbf{x}_{ij}^TK_{ij}$ and $l_n^2 = o(l_n)$, then $\sum_{i=1}^{n}\left|\boldsymbol{\xi}_i\boldsymbol{\alpha}_i^T+\boldsymbol{\alpha}_i\boldsymbol{\xi}_i^T+\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^T\right| = O_p(nhl_n)$. Recall that $\boldsymbol{\xi}_1,...,\boldsymbol{\xi}_n$ are independent, then

$$\mathbf{W}_n = \mathrm{E}\left(\sum_{i=1}^{n}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T\right)+O_p\left(\sqrt{\mathrm{var}(\sum_{i=1}^{n}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T)}\right)+O_p(nhl_n) = \sum_{i=1}^{n}\mathrm{E}(\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T)+O_p(x_n),$$

where $x_n = \sqrt{\sum_{i=1}^{n}\mathrm{var}(\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T)}+nhl_n$. By Theorem 2.1, we have $n\mathbf{H}_n^{-1}\sum_{i=1}^{n}\mathrm{E}(\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T)\mathbf{H}_n^{-1}\to\boldsymbol{\Sigma}_{\mathrm{dense}}(t)$ for dense data or $nh\mathbf{H}_n^{-1}\sum_{i=1}^{n}\mathrm{E}(\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T)\mathbf{H}_n^{-1}\to\boldsymbol{\Sigma}_{\mathrm{sparse}}(t)$ for sparse data. Therefore, it remains to show that $x_n = o(nh^2)$ for dense data and $x_n = o(nh)$ for sparse data.

For the dense data, we have $\sum_{i=1}^{n}\mathrm{var}(\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T) = O(nh^4)$ based on the regularity condition

6 and thus $x_n = O(\sqrt{n}h^2 + nh^3 + h\sqrt{n\log n}) = o(nh^2)$. For the sparse data, we have $\sum_{i=1}^n \mathrm{var}(\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T) = O(nh)$ and therefore $x_n = O(\sqrt{nh} + nh^3 + \sqrt{nh\log n}) = o(nh)$.

References

Fan J. and Gijbels, I. (1992). Variable Bandwidth and Local Linear Regression Smoothers. *The Annals of Statistics*, 20, 2008-2036.

Fan, J., Huang, T., and Li, R. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function. *Journal of American Statistical Association*, 102, 632-640.

Fan, J. and Wu, Y. (2008). Semiparametric estimation of covariance matrices for longitudinal data. *Journal of American Statistical Association*, 103, 1520-1533.

Fan, J., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *The Annals of Statistics*, 29, 153-193.

Fan, J. and Zhang, J. T. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *Journal of the Royal Statistical Society*, Ser. B, 62, 303-322.

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

Huang, J. Z., Wu, C. O. and Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika*, 89, 111-128.

Kaslow, R. A., Ostrow, D. G., Detels, R., Phair, J. P., Polk, B. F. and Rinaldo, C. R. (1987). The multicenter AIDS cohort study - rationale, organization, and selected characteristics of the participants. *American Journal of Epidemiology*, 126, 310-318.

Kim, S. and Zhao, Z. (2013). Unified inference for sparse and dense longitudinal models. *Biometrika*, 100, 203-212.

Liang, H., Wu, H. and Carroll R.J. (2003). The relationship between virologic responses in AIDS clinical research using mixed-effects varying-coefficient models with measurement error. *Biostatistics*, 4, 297-312.

Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics*, 38, 3321-3351.

Lin, X. and Carroll, R. J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of American Statistical Association*, 95, 520-534.

Ma, S., Yang, L. and Carroll, R. J. (2012). A simultaneous confidence band for sparse longitudinal regression. *Statistica Sinica*, 22, 95-122.

Pan, J. and Mackenzie, G. (2003). Model selection for joint mean-covariance structures in longitudinal studies. *Biometrika*, 90, 239-244.

Pourahmadi, M. (2007). Cholesky decompositions and estimation of a covariance matrix: Orthogonality of variance-correlation parameters. *Biometrika*, 94, 1006-1013.

Qu, A. and Li, R. (2006). Quadratic inference functions for varying-coefficient models with longitudinal data. *Biometrics*, 62, 379-391.

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

Tian, X. and Wu, C. O. (2014). Estimation of rank-tracking probabilities using nonparametric mixed-effects models for longitudinal data. *Statistics and Its Interface*, 7, 87-99.

Wu, C. O. and Chiang, C.-T. (2000). Kernel smoothing on varying coefficient models with longitudinal dependent variable. *Statistica Sinica*, 10, 433-456.

Wu, H. and Zhang, J. T. (2002). Local polynomial mixed-effects models for longitudinal data. *Journal of the American Statistical Association*, 97, 883-897.

Yao, F., Müller, H. G. and Wang, J. L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100, 577-590.

Yao, W. and Li, R. (2013). New local estimation procedure for a non-parametric regression function for longitudinal data. *Journal of the Royal Statistical Society*, Series B, 75, 123-138.

Ye, H. and Pan, J. (2006). Modelling covariance structures in generalized estimating equations for longitudinal data. *Biometrika*, 93, 927-941.

Zhang, J. T. and Chen, J. (2007). Statistical inferences for functional data. *The Annals of Statistics*, 35, 1052-1079.

Zhang, W. and Leng, C. (2012). A moving average cholesky factor model in covariance modeling for longitudinal data. *Biometrika*, 99, 141-150.

Zhang, W., Leng, C., and Tang C. (2015). A joint modelling approach for longitudinal studies. *Journal of the Royal Statistical Society*, Ser. B, 77, 219-238.

Table 1: Average empirical coverage percentages and lengths, in parentheses, for $\beta_1(t)$ of five confidence intervals when n=200.

$1-\alpha$	N	SN	NS	ND	NSD	BS
90%	N_1	85.7(0.218)	82.1(0.198)	56.4(0.115)	87.4(0.226)	88.8(0.238)
	N_2	87.1(0.169)	76.6(0.132)	70.4(0.115)	88.5(0.174)	88.7(0.177)
	N_3	88.6(0.133)	61.2(0.074)	82.6(0.115)	89.8(0.136)	89.0(0.135)
	N_4	89.2(0.126)	54.5(0.057)	86.3(0.115)	90.1(0.128)	89.2(0.126)
95%	N_1	91.4(0.261)	88.4(0.236)	64.1(0.137)	92.6(0.270)	93.7(0.283)
	N_2	92.7(0.201)	84.1(0.157)	78.1(0.137)	93.4(0.207)	93.7(0.210)
	N_3	93.2(0.159)	68.8(0.088)	88.9(0.137)	94.1(0.163)	93.5(0.161)
	N_4	93.7(0.150)	61.4(0.068)	91.9(0.137)	94.6(0.153)	93.7(0.151)

SN, the self-normalized confidence interval in (2.12); NS and ND, the asymptotic normality based confidence intervals (2.10) and (2.11) assuming sparse and dense data, respectively; NSD, the infeasible confidence interval in (3.3); BS, the bootstrap confidence interval; N_i, i=1,...,4, the number of measurements on individual subjects defined in (3.1) and (3.2).

Table 2: Average empirical coverage percentages and lengths, in parentheses, for $\beta_2(t)$ of five confidence intervals when n=200.

$1-\alpha$	N	SN	NS	ND	NSD	BS
90%	N_1	86.6(0.219)	83.0(0.198)	57.1(0.115)	88.2(0.226)	88.9(0.232)
	N_2	86.9(0.169)	77.1(0.132)	70.6(0.115)	88.2(0.174)	88.1(0.174)
	N_3	88.5(0.134)	61.6(0.074)	82.8(0.115)	89.6(0.136)	88.7(0.135)
	N_4	88.9(0.126)	54.0(0.057)	85.8(0.115)	90.1(0.128)	89.0(0.127)
95%	N_1	92.0(0.260)	89.3(0.236)	65.2(0.137)	93.5(0.270)	93.8(0.276)
	N_2	93.0(0.201)	84.4(0.157)	78.6(0.137)	94.0(0.207)	93.7(0.208)
	N_3	93.5(0.160)	69.8(0.088)	89.3(0.137)	94.1(0.163)	93.8(0.161)
	N_4	93.7(0.150)	60.0(0.068)	91.3(0.137)	94.2(0.153)	93.6(0.150)

SN, the self-normalized confidence interval in (2.12); NS and ND, the asymptotic normality based confidence intervals (2.10) and (2.11) assuming sparse and dense data, respectively; NSD, the infeasible confidence interval in (3.3); BS, the bootstrap confidence interval; N_i, i=1,...,4, the number of measurements on individual subjects defined in (3.1) and (3.2).

Table 3: Average empirical coverage percentages and lengths, in parentheses, for $\beta_1(t)$ of five confidence intervals when n=400.

$1-\alpha$	N	SN	NS	ND	NSD	BS
90%	N_1	86.5(0.163)	82.9(0.147)	54.9(0.082)	87.3(0.166)	88.8(0.174)
	N_2	87.9(0.125)	78.1(0.098)	69.2(0.082)	88.7(0.126)	89.4(0.130)
	N_3	88.7(0.097)	63.6(0.055)	82.2(0.082)	89.2(0.098)	89.1(0.098)
	N_4	88.4(0.091)	53.7(0.042)	84.8(0.082)	89.2(0.092)	88.5(0.091)
95%	N_1	92.3(0.194)	89.5(0.175)	63.1(0.097)	93.3(0.198)	94.1(0.207)
	N_2	93.6(0.149)	85.6(0.117)	78.0(0.097)	94.1(0.150)	94.5(0.154)
	N_3	94.2(0.116)	72.0(0.065)	88.9(0.097)	94.6(0.117)	94.5(0.117)
	N_4	94.0(0.108)	62.2(0.051)	91.0(0.097)	94.4(0.109)	94.0(0.109)

SN, the self-normalized confidence interval in (2.12); NS and ND, the asymptotic normality based confidence intervals (2.10) and (2.11) assuming sparse and dense data, respectively; NSD, the infeasible confidence interval in (3.3); BS, the bootstrap confidence interval; N_i, i=1,...,4, the number of measurements on individual subjects defined in (3.1) and (3.2).

Table 4: Average empirical coverage percentages and lengths, in parentheses, for $\beta_2(t)$ of five confidence intervals when n=400.

$1-\alpha$	N	SN	NS	ND	NSD	BS
90%	N_1	86.4(0.162)	82.8(0.147)	55.0(0.082)	87.6(0.166)	88.1(0.170)
	N_2	88.2(0.125)	77.8(0.098)	68.8(0.082)	88.9(0.126)	88.9(0.128)
	N_3	88.8(0.097)	62.8(0.055)	81.7(0.082)	89.0(0.098)	88.8(0.097)
	N_4	90.2(0.091)	56.0(0.042)	86.4(0.082)	90.8(0.092)	90.1(0.091)
95%	N_1	93.1(0.194)	90.4(0.175)	63.9(0.097)	93.9(0.198)	94.4(0.203)
	N_2	93.8(0.148)	85.7(0.117)	77.5(0.097)	94.2(0.150)	94.4(0.152)
	N_3	94.0(0.115)	71.1(0.065)	88.8(0.097)	94.5(0.117)	94.1(0.116)
	N_4	95.0(0.108)	64.0(0.051)	92.5(0.097)	95.5(0.109)	95.0(0.109)

SN, the self-normalized confidence interval in (2.12); NS and ND, the asymptotic normality based confidence intervals (2.10) and (2.11) assuming sparse and dense data, respectively; NSD, the infeasible confidence interval in (3.3); BS, the bootstrap confidence interval; N_i, i=1,...,4, the number of measurements on individual subjects defined in (3.1) and (3.2).

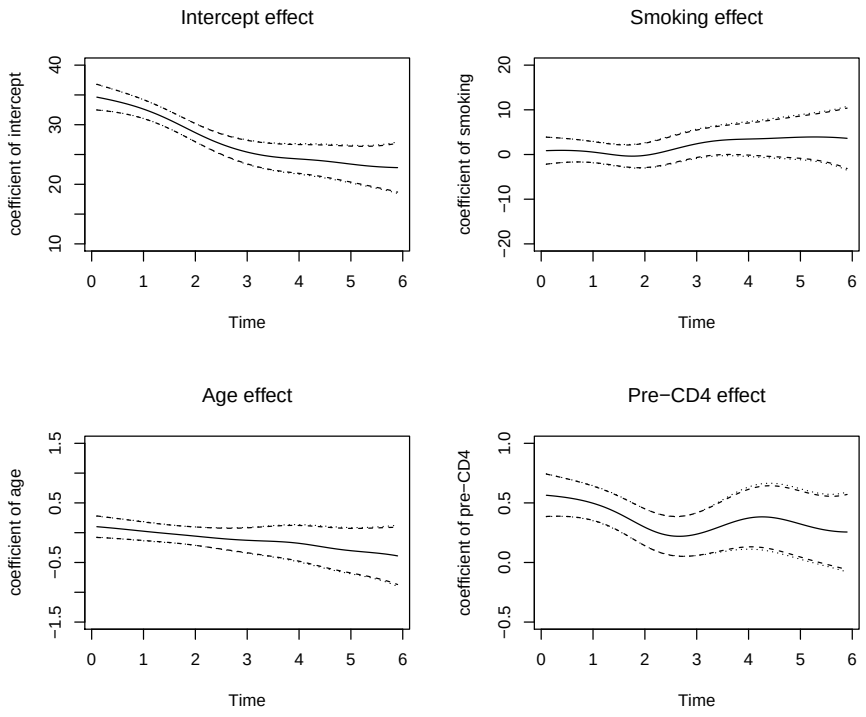

Figure 1: Application to AIDS data. Estimated coefficient curves for the baseline CD4 percentage and the effects of smoking, age and pre-infection CD4 percentage on the percentage of CD4 cells. The value of the selected bandwidth is 0.7074. Solid curves, estimated effects; dashed curves, 95% self-normalization based confidence intervals; dotted curves, 95% bootstrap pointwise confidence intervals.