Backfitting and Local Likelihood Methods for Nonparametric Mixed-Effects Models with Longitudinal Data

Jeong-Gun Park

Frontier Science & Technology Research Foundation, Boston, USA

and Hulin Wu

University of Rochester, Rochester, USA

January, 2004

Corresponding Author:

Dr. Hulin Wu, Professor

Department of Biostatistics and Computational Biology

University of Rochester School of Medicine and Dentistry

601 Elmwood Avenue, Box 630, Rochester, New York 14642, USA

Phone: (585) 275-6767

Fax: (585) 273-1031

Email: hwu@bst.rochester.edu

Backfitting and Local Likelihood Methods for Nonparametric Mixed-Effects Models with Longitudinal Data

Jeong-Gun Park and Hulin Wu

Abstract

For longitudinal data analyses, it is important to estimate both population (mean) response and subject-specific individual responses. We consider a nonparametric mixed-effects model that characterizes both population effect and random effects as nonparametric functions, i.e., $y_i(t_{ij}) = \eta(t_{ij}) + v_i(t_{ij}) + \varepsilon_i(t_{ij}),\ i = 1, \ldots, m,\ j = 1, \ldots, n_i$. Although similar models have been studied by ourselves and others, in this paper we propose a novel approach to use the local likelihood concept and a backfitting algorithm to combine the local polynomial regression techniques and the linear mixed-effects (LME) model methods for efficiently estimating both population mean and individual curve functions. The asymptotic properties of the proposed estimators are established for two scenarios: (i) the number of subjects (m) and the number of measurements per subject (n_i) tend to infinity; and (ii) m tends to infinity while n_i is finite. The simulation studies are carried out to compare the performance of our proposed estimator with that of the two existing estimators, local polynomial LME (LLME) estimator proposed by Wu and Zhang (2002) and local polynomial GEE (LGEE) estimator proposed by Lin and Carroll (2000), and we show that our approach performs better than the existing methods in the sense of mean squared errors (MSE). We illustrate our estimation methods with an application to an AIDS clinical study.

Key Words: Longitudinal Data; Linear Mixed-Effects Model; Nonparametric Mixed-Effects Model; Local Likelihood; Cross Validation; Local Polynomial Regression.

1 Introduction

Longitudinal data frequently arise in many scientific studies, and longitudinal data modeling/analysis has played an important role in scientific investigations. See Diggle, Liang and Zeger (1994), Davidian and Giltinan (1995), and Vonesh and Chinchilli (1996) for many examples. Longitudinal data are collected repeatedly over a time period on the target subjects, so that the measurements within each subject may be correlated with each other but between-subjects are usually assumed to be uncorrelated. As an example, the number or percentage of CD4 cells which play a vital role in the immune function of human body is monitored in HIV infected patients over time throughout the disease incubation period or during clinical studies.

In this paper, we consider a general nonparametric mixed-effects model for longitudinal data (Shi, Weiss and Taylor, 1996; Rice and Wu, 2001; Wu and Zhang, 2002). Let m denote the number of subjects, with the i^{th} subject having n_i observations over time. Assume that $y_i(t_{ij})$ is the response for the i^{th} subject at time point t_{ij} and follows a nonparametric mixed-effects (NPME) model

$$y_i(t_{ij}) = \eta(t_{ij}) + v_i(t_{ij}) + \varepsilon_i(t_{ij}),\ i = 1, \ldots, m,\ j = 1, \ldots, n_i, \tag{1.1}$$

where $\eta(t_{ij})$ and $v_i(t_{ij})$ represent population mean curve over time and individual curve variations from $\eta(t_{ij})$ over time, respectively, and $\varepsilon_i(t_{ij})$ are the measurement errors. The $v_i(t_{ij})$'s are considered to be realizations of a zero mean process with a covariance function $\gamma(j, j') = E[v_i(t_{ij})v_i(t_{ij'})]$ which assumes that the observations within the same subject might be correlated, $\varepsilon_i(t_{ij})$ are assumed to be an uncorrelated zero mean process with a variance function $\sigma^2(t)$, and $v_i(t_{ij})$ and $\varepsilon_i(t_{ij})$ are assumed to be mutually independent.

For general nonparametric regression models with longitudinal data, Hoover et al (1998) proposed smoothing spline and kernel smoothing methods. Lin and Carroll (2000) introduced a kernel GEE estimator that is shown to be most efficient asymptotically when the working covariance is assumed to be diagonal (the working-independence assumption). Huang, Wu, and Zhou (2002) studied the regression spline method. Fan and Zhang (2000) suggested a simple two-step approach. However, all these methods only concerned the population mean estimation of unknown nonparametric functions under the longitudinal study framework. The typical correlation structure of longitudinal data was not efficiently considered. The between-subject and within-subject variations were not carefully modelled. The estimators of unknown nonparametric functions for individual subjects are usually not available from these existing methods. Shi et al (1996) and Rice and Wu (2001) introduced mixed-effects regression spline models for longitudinal data analysis. These models are essentially standard parametric mixed-effects models which are easy to understand and implement. Wang (2003) recently proposed an iterative kernel GEE estimator in which the within-subject correlation is directly incorporated. She showed that this new estimator uniformly outperforms the working independence estimator. Wu and Zhang (2002) extended the powerful mixed-effects modelling idea to local polynomial smoothing of longitudinal data. They also provided some numerical evidences for better performance of their local polynomial linear mixed-effects (LLME) model estimator. Staniswalis and Lee (1998) applied a similar model to study a dose-response problem. In their paper, a different approach was used.

The mixed-effects modelling idea allows us to estimate the population mean function $\eta(t)$ and individual curve variations $v_i(t)$ from the mean function, or individual curve functions $\varphi_i(t) \equiv \eta(t) + v_i(t)$. It has been known that both population mean and individual curve functions are important, especially in clinical or biomedical studies. The population mean function represents the overall progress of the underlying population process and individual curve functions represent the responses from individual subjects. For instance, the mean response of CD4 cells to a treatment is critical in AIDS clinical studies and the individual response of CD4 cells from each patient can provide better ideas for clinicians to individualize the treatment.

In this article, we propose a new method to estimate $\eta(t)$ and $\varphi_i(t)$ based on a local likelihood technique combined with an iterative approach of backfitting described in later sections. The idea of local likelihood estimation was originally introduced by Tibshirani and Hastie (1987), and we extend their approach to mixed-effects models for longitudinal data. The proposed backfitting algorithm repeatedly corrects the estimates of the variance-covariance parameters using E-M algorithm and updates the population mean and individual curve estimates based on the new variance-covariance estimates. One advantage of using our proposed approach is to preserve the properties of the local maximum likelihood approach and improve the variance-covariance estimates using the proposed backfitting method. From simulation studies, we can clearly see that our local likelihood estimators greatly improve the estimates of population mean functions of the working-independent kernel GEE estimator (Lin and Carroll, 2000), and outperform the individual curve estimates proposed by Wu and Zhang (2002).

Recently Wang (2003) also proposed an iterative kernel GEE estimator to improve the working-independent kernel GEE estimator by incorporating the within-subject correlation directly in the estimating equation. It has been shown (Lin et al. 2003) that Wang's estimator has a locality property at cluster or subject level, but it is non-local at observation level. These properties are critical for this estimator to achieve asymptotic consistency and asymptotic efficiency. Comparing to Wang's estimator, in this paper we try to improve the finite sample efficiency of the working-independent estimator (Lin and Carroll, 2000) by incorporating the within-subject correlation in a local likelihood framework. Since the correlation (or covariance matrix) in our estimator is simultaneously estimated with the mean functions based on the local data, it is efficient and has an 'adaptive' feature. In addition, our method also provides the estimates of responses for individual subjects using the empirical Bayesian approach (the best linear unbiased predictor or BLUP), while Wang's method and others only focus on the population estimate. The good performance of our estimators is attributed to the mixed-effects modelling idea.

The remainder of this article is organized as follows. In section 2, we indroduce a general form of the local likelihood in mixed-effects models for longitudinal data. The non-parametric functions are approximated using the local polynomial approach. We propose new estimation methods to estimate the population mean function $\eta(t)$ and individual curve functions $\varphi_i(t)$. Section 3 investigates bandwidth selection methods and proposes a back-fitting estimation procedure. We study asymptotic theories of the proposed estimators in Section 4. In Section 5, we perform simulation experiments to evaluate the proposed method by comparing to the existing methods. The new method is applied to a real example of a longitudinal data set obtained from an AIDS clinical study for illustration in Section 6. We conclude with some final remarks in Section 7.

2 Methods of Estimation: Local likelihood Approach

Tibshirani and Hastie (1987) first proposed the local likelihood method. Staniswalis (1989) and Fan, Farmen, and Gijbels (1998) further studied the properties of local kernel-weighted likelihood estimators. In this section, we generalize the local likelihood ideas into longitudinal data in which within-subject correlations commonly exist.

Suppose that $\mathbf{y}_i = (y_i(t_{i1}), \cdots, y_i(t_{in_i}))^T$ is a $(n_i \times 1)$ vector of observations obtained from the i^{th} subject at time points $t_{i1}, \cdots, t_{in_i}$ and has a distribution $f_{i,\ (i=1,\cdots,m)}$. Then the contributions from the i^{th} subject to the overall log-likelihood is $l_i(\boldsymbol{\theta}_i; \mathbf{y}_i) = log f_i(\mathbf{y}_i; \boldsymbol{\theta}_i)$, where $\boldsymbol{\theta}_i$ are unknown parameters to be estimated. The log-likelihood of the observations from all the m subjects is then given by

$$\sum_{i=1}^{m} l_i(\boldsymbol{\theta}_i; \mathbf{y}_i). \tag{2.1}$$

Let $K_{ih} = diag\{K_{i1}, \cdots, K_{in_i}\}$, where $K_{ij} = K([t_{ij} - t_0]/h)/h$ are kernel-weight functions in the neighborhood of t_0 and $h > 0$ is a smoothing parameter, called bandwidth. Then the kernel-weighted log-likelihood is defined by

$$l^*(\boldsymbol{\theta}; \mathbf{y}) = \sum_{i=1}^{m} l_i(\boldsymbol{\theta}_i; \mathbf{y}_i, K_{ih}), \tag{2.2}$$

which is a function of K_{ih}.

As an example, if $n_i = 1$ and $l_i(\boldsymbol{\theta}_i; \mathbf{y}_i, K_{ih}) = l_i(\boldsymbol{\theta}_i; \mathbf{y}_i)K_{ih}$, then the kernel-weighted log-likelihood can be written as

$$l^*(\boldsymbol{\theta}; \mathbf{y}) \;\; = \;\; \sum_{i=1}^{m} \Big\{ log f_i(y_i, \boldsymbol{\theta}_i) \; \frac{1}{h} K\Big(\frac{t_i - t_0}{h}\Big) \Big\},$$

which is the standard local likelihood function for independent data as discussed by Staniswalis (1989) and Fan, et. al. (1998). For the case that there is no within-subject correlations, Equation (2.2) can be written as $\sum_{i=1}^{m} \sum_{j=1}^{n_i} l_i(\boldsymbol{\theta}_i; y_{ij}) K_{ij}$. This coincides with the cases that Hoover et al. (1998) and Lin and Carroll (2000) considered.

However, in general, the form of the local log-likelihood (2.2) is problem-specific. The application of the kernel weight in different ways may result in different estimators. The following section illustrates the application of the definition (2.2) under the setting of nonparametric mixed-effects models for different scenarios.

2.1 Local Polynomial Approximation

In this section, we provide a general framework to construct a local approximation of the NPME model (1.1) by a linear mixed-effects model as first proposed by Wu and Zhang

(2002). Assume that $\eta(t)$ and $v_i(t)$ have $(p+1)^{th}$ continuous derivatives, for any fixed t. By applying Taylor expansions, the population mean function and individual curve functions at t_{ij} can be approximated by the local polynomial of order p in a neighborhood of t and can be written as

$$\eta(t_{ij}) \approx \sum_{k=0}^{p} \frac{\eta^{(k)}(t)}{k!}(t_{ij}-t)^k \quad \text{and} \quad v_i(t_{ij}) \approx \sum_{k=0}^{p} \frac{v^{(k)}(t)}{k!}(t_{ij}-t)^k. \tag{2.3}$$

By letting

$$\begin{aligned} X_{ij} &= [1, t_{ij}-t, \ldots, (t_{ij}-t)^p]^T, \\ \boldsymbol{\beta} &= [\eta(t), \eta^{'}(t), \ldots, \frac{\eta^{(p)}(t)}{p!}]^T, \\ \mathbf{b}_i &= [v_i(t), v_i^{'}(t), \ldots, \frac{v^{(p)}(t)}{p!}]^T. \end{aligned} \tag{2.4}$$

Model (1.1) can be approximated by a linear mixed-effects (LME) model within a neighborhood of t:

$$y_i(t_{ij}) = X_{ij}^T(\boldsymbol{\beta} + \mathbf{b}_i) + \varepsilon_i(t_{ij}), \;\; j = 1, \ldots, n_i, \;\; i = 1, \ldots, m, \tag{2.5}$$

where the vector $\boldsymbol{\varepsilon}_i = (\varepsilon_i(t_{ij}), \ldots, \varepsilon_i(t_{in_i}))^T$ has mean zero with a covariance matrix $R_i = E[\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T] = diag[\sigma^2(t_{i1}), \ldots, \sigma^2(t_{in_i})]$ and the vector $\mathbf{b}_i$ of random effects has mean zero with a covariance matrix $D \equiv D(t) = E[\mathbf{b}_i \mathbf{b}_i^T]$. For inferences on the parameters $\boldsymbol{\beta}$ and $\mathbf{b}_i$ in model (2.5), we apply the previously introduced ideas of local likelihood approach to the LME model.

2.2 Local Marginal Likelihood Estimation

In this section, we introduce a local marginal likelihood method for estimating the population mean curve $\eta(t)$. For the model (2.5), let $X_i = [X_{i1}, \ldots, X_{in_i}]^T$ and assume that $\mathbf{b}_i$ and $\boldsymbol{\varepsilon}_i$ are independent and normally distributed. Then, the marginal distribution of $\mathbf{y}_i$ is normal with a mean of $X_i\boldsymbol{\beta}$ and variance of $V_i = X_i D X_i^T + R_i$. It thus yields the log-likelihood function for $\boldsymbol{\beta}$ is

$$l(\boldsymbol{\beta}; \mathbf{y}) = -\frac{1}{2}\sum_{i=1}^{m}\Big\{[\mathbf{y}_i - X_i\boldsymbol{\beta}]^T V_i^{-1}[\mathbf{y}_i - X_i\boldsymbol{\beta}] + C_{i1}\Big\}, \tag{2.6}$$

where $C_{i1} = log(|V_i|) + n_i log(2\pi)$. The local marginal log-likelihood function for estimating $\boldsymbol{\beta}$ can be written as

$$l^*(\boldsymbol{\beta};\mathbf{y}) = -\frac{1}{2}\sum_{i=1}^{m}\Big\{[\mathbf{y}_i - X_i\boldsymbol{\beta}]^T K_{ih}^{1/2} V_i^{-1} K_{ih}^{1/2}[\mathbf{y}_i - X_i\boldsymbol{\beta}] + (\mathbf{1}_{n_i}^T K_{ih}\mathbf{1}_{n_i})C_{i1}\Big\}, \qquad (2.7)$$

where $K_{ih} = diag[K_h(t_{i1} - t), \ldots, K_h(t_{in_i} - t))]$ is a $(n_i \times n_i)$ matrix of kernel functions which weight the residuals $(\mathbf{y}_i - X_i\boldsymbol{\beta})$ symmetrically.

For a given variance V_i, the differentiation of (2.7) with respect to $\boldsymbol{\beta}$ yields the estimating equation for $\boldsymbol{\beta}$,

$$\mathbf{X}^T\Omega_h\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^T\Omega_h\mathbf{y}, \qquad (2.8)$$

where $\mathbf{X} = [X_1, \ldots, X_n]^T$, $\mathbf{y} = [\mathbf{y}_1, \ldots, \mathbf{y}_n]^T$, and $\Omega_h = diag[\Omega_{1h}, \ldots, \Omega_{nh}]$ with $\Omega_{ih} = K_{ih}^{1/2} V_i^{-1} K_{ih}^{1/2}$. Thus, a closed-form estimator for $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}}_M = \Big[\sum_{i=1}^{m}\{X_i^T K_{ih}^{\frac{1}{2}} V_i^{-1} K_{ih}^{\frac{1}{2}} X_i\}\Big]^{-1}\Big[\sum_{i=1}^{m}\{X_i^T K_{ih}^{\frac{1}{2}} V_i^{-1} K_{ih}^{\frac{1}{2}}\mathbf{y}_i\}\Big]. \qquad (2.9)$$

When V_i is known, the estimator (2.9) can be obtained by fitting the following model using the Splus function LM or SAS procedure PROC GLM,

$$V_i^{-\frac{1}{2}} K_{ih}^{\frac{1}{2}}\mathbf{y}_i = V_i^{-\frac{1}{2}} K_{ih}^{\frac{1}{2}} X_i\boldsymbol{\beta} + \mathbf{e}_i, \qquad (2.10)$$

where $\mathbf{e}_i$ has a mean 0 and variance $\sigma_e^2 I_{n_i}$. The model (2.10) is a standard linear model with a response variable $V_i^{-\frac{1}{2}} K_{ih}^{\frac{1}{2}}\mathbf{y}_i$ and a covariate $V_i^{-\frac{1}{2}} K_{ih}^{\frac{1}{2}} X_i$.

The local marginal likelihood estimator of $\eta(t)$ can be found as

$$\hat{\eta}_M(t) = \boldsymbol{\xi}_{p+1}^T\hat{\boldsymbol{\beta}}_M, \qquad (2.11)$$

where $\boldsymbol{\xi}_{p+1}$ is a $(p+1)$ dimensional vector with the first element being 1, and 0 elsewhere.

The covariance matrix V_i has been assumed to be known in order to obtain the closed-form estimator (2.9). In practice, we, however, commonly encounter real examples where the covariances are unknown and need to be estimated. The estimation of covariances as well as random effects curves will be introduced in later sections. When V_i is a known diagonal matrix, the estimator $\hat{\boldsymbol{\beta}}$ is reduced to the local polynomial GEE (LGEE) estimator proposed by Lin and Carroll (2000).

2.3 Local Joint Likelihood Estimation

In this section, an alternative estimation approach is proposed to estimate the parameters in the local LME model with longitudinal data. For the LME model (2.5), we derive the joint log-likelihood function of $\{(\mathbf{y}_i, \mathbf{b}_i), i = 1, \cdots, m\}$ based on the assumptions of Gaussian distributions for $\mathbf{b}_i$ and $\boldsymbol{\varepsilon}_i$. Under the Gaussian assumptions, we have $\mathbf{y}_i|\mathbf{b}_i \sim N(X_i\boldsymbol{\beta} + X_i\mathbf{b}_i, R_i)$ and $\mathbf{b}_i \sim N(0, D)$. Thus, the joint log-likelihood function of $(\mathbf{y}_i, \mathbf{b}_i)$ is

$$l(\boldsymbol{\beta}, \mathbf{b}; \mathbf{y}) = -\frac{1}{2}\sum_{i=1}^{m}\left\{\left[\mathbf{y}_i - X_i(\boldsymbol{\beta} + \mathbf{b}_i)\right]^T R_i^{-1}\left[\mathbf{y}_i - X_i(\boldsymbol{\beta} + \mathbf{b}_i)\right] + \mathbf{b}_i^T D^{-1}\mathbf{b}_i + C_{i3}\right\}, \quad (2.12)$$

where $\mathbf{b} = (\mathbf{b}_1, \cdots, \mathbf{b}_m)$ and $C_{i3} = log(|R_i|) + log(|D|) + 2n_i log(2\pi)$ is free from $\boldsymbol{\beta}$ and $\mathbf{b}_i$. The local kernel-weighted log-likelihood at the neighborhood of time t can be considered in two different ways:

$$\begin{aligned} l_1^*(\boldsymbol{\beta}, \mathbf{b}; \mathbf{y}) &= -\frac{1}{2}\sum_{i=1}^{m}\left\{\left[\mathbf{y}_i - X_i(\boldsymbol{\beta} + \mathbf{b}_i)\right]^T K_{ih}^{1/2} R_i^{-1} K_{ih}^{1/2}\left[\mathbf{y}_i - X_i(\boldsymbol{\beta} + \mathbf{b}_i)\right]\right. \\ &\qquad \left. + \mathbf{b}_i^T D^{-1}\mathbf{b}_i + C_{i3}\right\}, \end{aligned} \quad (2.13)$$

and

$$\begin{aligned} l_2^*(\boldsymbol{\beta}, \mathbf{b}; \mathbf{y}) &= -\frac{1}{2}\sum_{i=1}^{m}\left\{\left[\mathbf{y}_i - X_i(\boldsymbol{\beta} + \mathbf{b}_i)\right]^T K_{ih}^{1/2} R_i^{-1} K_{ih}^{1/2}\left[\mathbf{y}_i - X_i(\boldsymbol{\beta} + \mathbf{b}_i)\right]\right. \\ &\qquad \left. + \mathbf{1}_{n_i}^T K_{ih}\mathbf{1}_{n_i}\left[\mathbf{b}_i^T D^{-1}\mathbf{b}_i + C_{i3}\right]\right\}, \end{aligned} \quad (2.14)$$

where I_{n_i} is the identity matrix with dimension n_i. In (2.13), the kernel weights are symmetrically applied only to the first term of the likelihood function, i.e., the residual term $\mathbf{y}_i - X_i(\boldsymbol{\beta} + \mathbf{b}_i)$ which only involves the observed data, while, in (2.14), the kernel weights are applied to the entire likelihood function of (2.12) in which the random-effect term $\mathbf{b}_i^T D^{-1}\mathbf{b}_i$ is also multiplied by the kernel weight. These two different kernel weighting methods result in two different estimators.

For given D, R_i, and K_{ih}, the estimators obtained by maximizing (2.13) and (2.14) are the solutions of the following mixed-model normal equations reference with $\omega_{ih} = 1$ and $\omega_{ih} = \mathbf{1}_{n_i}^T K_{ih}\mathbf{1}_{n_i}$, respectively.

$$\begin{cases} \sum_{i=1}^{m} X_i^T \Psi_{ih} X_i(\boldsymbol{\beta} + \mathbf{b}_i) = \sum_{i=1}^{m} X_i^T \Psi_{ih}\mathbf{y}_i, \\ X_i^T \Psi_{ih} X_i\boldsymbol{\beta} + (X_i^T \Psi_{ih} X_i + \omega_{ih} D^{-1})\mathbf{b}_i = X_i^T \Psi_{ih}\mathbf{y}_i, \ i = 1, \cdots, m, \end{cases} \quad (2.15)$$

where $\Psi_{ih} = K_{ih}^{1/2} R_i^{-1} K_{ih}^{1/2}$. By solving the above normal equations (2.15), the estimators for $\boldsymbol{\beta}$ and $\mathbf{b_i}$, under the assumptions of known D and R_i, can be written into the following closed forms:

$$\hat{\boldsymbol{\beta}}_J = \Big[\sum_{i=1}^{m} X_i^T \Sigma_{ih}^{-1} X_i\Big]^{-1} \Big[\sum_{i=1}^{m} X_i^T \Sigma_{ih}^{-1} \mathbf{y}_i\Big], \tag{2.16}$$

and

$$\hat{\mathbf{b}}_i = \Big[X_i^T \Psi_{ih} X_i + \omega_{ih} D^{-1}\Big]^{-1} \Big[X_i^T \Psi_{ih}(\mathbf{y}_i - X_i\boldsymbol{\beta})\Big], \tag{2.17}$$

where $\Sigma_{ih} = \omega_{ih}^{-1} K_{ih}^{\frac{1}{2}} X_i D X_i^T K_{ih}^{\frac{1}{2}} + R_i$. Thus, the estimators of $\eta(t)$ and $\upsilon_i(t)$ can be found as

$$\hat{\eta}_J(t) = \boldsymbol{\xi}_{p+1}^T \hat{\boldsymbol{\beta}}_J \quad \text{and} \quad \hat{\upsilon}_i(t) = \boldsymbol{\xi}_{p+1}^T \hat{\mathbf{b}}_i, \tag{2.18}$$

where $\boldsymbol{\xi}_{p+1}$ is a $(p+1)$ dimensional vector with the first element being 1, and 0 elsewhere. The objective local likelihood function (2.13) results in the exact LLME estimators proposed by Wu and Zhang (2002) which are the closed forms (2.16) and (2.17) with the weight $\omega_{ih} = 1$.

One may notice the difference between the local marginal likelihood estimator (2.9) and the estimator (2.16) for the population parameter $\boldsymbol{\beta}$ due to different weight functions. In the estimates of random-effects parameters (2.17), the population parameter $\boldsymbol{\beta}$ can be replaced by any consistent estimators, such as (2.9) and (2.16). In fact, $\mathbf{b}_i$ is an empirical Bayesian estimator or a best linear unbiased predictor (BLUP). See Davidian and Giltinan (1995) and Vonesh and Chinchilli (1996) for details. The estimates of random effects allow us to capture the individual response curves, $\varphi_i(t) = \eta(t) + \upsilon_i(t)$, which is a great advantage of the NPME modelling approach. One can also easily see that, from (2.15) with $\omega_{ih} = 1$ and $\omega_{ih} = \mathbf{1}_{n_i}^T K_{ih} \mathbf{1}_{n_i}$, the application of kernel weights in different ways may result in different local likelihood estimators. These estimators may have different properties and efficiencies.

3 Bandwidth Selection and Backfitting Algorithm

When local polynomial techniques are used to fit nonparametric models, an important issue is to determine the smoothing parameter h (bandwidth) and the order of polynomial approximations. Fan and Gijbels (1996) suggested that the bandwidth selection is far more important than the determination of the polynomial order, since the local linear fit usually can give a satisfactory result in most applications. In practical implementation, the smoothing parameter can be chosen by the subjective examination of scatter plots and the fitted curves. However, it might be more reasonable to use a data-driven method to select the bandwidth. In this section, we focus on the method of cross validation to choose the bandwidth h for a given kernel function $K_h(\cdot)$. Note that the kernel function $K_h(\cdot)$ is usually a symmetric function. The choice of the kernel functions is not as critical as the selection of the bandwidth. Some kernel functions and their properties are reviewed by Fan and Gijbels (1996). For simplicity, we use normal density kernel in this paper. The bandwidth h controls the goodness of fit and the roughness of the estimated function. The explicit forms of estimators allow us to use a backfitting algorithm to estimate $\eta(t)$, $v_i(t)$, and the unknown variance-covariance parameters iteratively. Our proposed backfitting algorithm estimates these unknown parameters with suitable bandwidths.

3.1 Cross-Validation (CV)

For estimating the mean function nonparametrically, a leave-one-subject-out cross validation (SJCV) which involves deletion of data from each subject is suggested for choosing an "optimal" bandwidth by Rice and Silverman (1991). As they pointed out, the main advantage of the method SJCV does not rely on the specific structure of within-subject correlations. However, a leave-one-time-point-out cross validation (PTCV) may be more appropriate for estimating the individual curve functions as proposed by Wu and Zhang (2002) because, for a given subject, the within-subject measurement errors are assumed to be uncorrelated. The presence of within-subject correlation is not caused by the measurement errors $\varepsilon_i(t)$, but by the random-effects in our model setting. This justifies the use of PTCV for estimating the individual curves, since it is a standard nonparametric estimation problem with *iid* measurement errors. For more details, see Wu and Zhang (2002).

We now define the SJCV and PTCV criteria for bandwidth selections. Let $\hat{\eta}^{(-i)}(t)$ be the estimator of $\eta(t)$ based on the data with the measurements of the i^{th} subject entirely removed. The SJCV criterion is defined as

$$CV_{SJ}(h) = \frac{1}{m}\sum_{i=1}^{m}\left\{\frac{1}{n_i}\sum_{j=1}^{n_i}\left[y_i(t_{ij}) - \hat{\eta}^{(-i)}(t_{ij})\right]^2\right\}. \tag{3.1}$$

The optimal bandwidth h_{SJ}^* for estimating the mean curve function $\eta(t)$ is obtained by minimizing $CV_{SJ}(h)$.

For estimation of the random curves, the following criterion $CV_{PT}(h)$ for obtaining the optimal bandwidth h_{PT}^*, is defined as follows. Define $\{t_{k,\ k=1,\ldots,T}\}$, be all the distinct design time points in the whole data set, and $y_l(t_k)$ be the measurements obtained from the l^{th} subject at time point t_k for $l = 1, \ldots, M_k$.

$$CV_{PT}(h) = \frac{1}{T}\sum_{k=1}^{T}\left\{\frac{1}{M_k}\sum_{l=1}^{M_k}\left[y_l(t_k) - \hat{\varphi}_l^{(-t_k)}(t_k)\right]^2\right\}, \tag{3.2}$$

where $\varphi_l(t_k)$ is such that $y_l(t_k) = \varphi_l(t_k) + \varepsilon_l(t_k)$ and $\hat{\varphi}_l^{(-t_k)}(t_k)$ is the estimator of $\varphi_l(t_k)$ based on the data after excluding the entire measurements obtained at the time point t_k. The optimal bandwidth h_{PT}^* for random-effects curves can be obtained by minimizing $CV_{PT}(h)$.

The estimators of the population mean function $\eta(t)$ and individual curve function $\varphi_i(t)$ are obtained with the optimal bandwidths based on the criteria (3.1) and (3.2). As pointed out by Wu and Zhang (2002), the methods based on the SJCV criterion to estimate both curves $\eta(t)$ and $\varphi_i(t)$ or based on the PTCV criterion to estimate both curve functions perform poorly. Wu and Zhang (2002) showed that the use of h_{SJ}^* for estimating population mean function $\eta(t)$ and h_{PT}^* for estimating random-effects curve functions $v_i(t)$ is more appropriate. We use similar ideas in our backfitting algorithm proposed in the following.

3.2 Backfitting Algorithm

In this section, a backfitting algorithm is proposed to obtain population mean function estimate $\hat{\eta}(t)$ and individual curve estimates $\hat{v}_i(t)$ as well as the estimates of variance-covariance components. For convenience, we introduce our algorithm using the local marginal likelihood estimator $\hat{\eta}_M$ for population mean estimator and $\hat{v}_i(t)$ with $\omega_{ih} = 1$ for random-effects curve estimators. Other estimators such as $\hat{\eta}_J$ and $\hat{v}_i(t)$ with $\omega_{ih} = \mathbf{1}_{n_i}^T K_{ih} \mathbf{1}_{n_i}$ can also be used in the algorithm. For simplicity, we assume that $R_i = \sigma^2 I_{n_i}$. The backfitting algorithm for estimating $\eta(t)$ and $v_i(t)$ is proposed as follows.

Step I: Find initial estimates $\hat{D}_0$ and $\hat{R}_{i0} = \hat{\sigma}_0^2 I_{n_i}$ for D and $R_i = \sigma^2 I_{n_i}$, respectively, and let $D = \hat{D}_0$ and $\sigma^2 = \hat{\sigma}_0^2$.

Step II: Based on the SJCV bandwidth selection method, obtain the estimate $\hat{\boldsymbol{\beta}}_M$ using the formula (2.9).

Step III: Based on the PTCV bandwidth selection method, obtain the estimate $\hat{\mathbf{b}}_i$ using the formula (2.17) with replacing $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}_M$ obtained from Step II.

Step IV: With the estimate $\hat{\mathbf{b}}_i$ obtained from Step III, find the estimates $\hat{\sigma}^2$ and $\hat{D}$ introduced in (3.7) and (3.8) in the following based on the PTCV bandwidth selection method.

Step V: Replace D and σ^2 by $\hat{D}$ and $\hat{\sigma}^2$, respectively, and repeat Step II through Step IV until convergence.

Step Vi: Obtain $\hat{\eta}_M(t) = \boldsymbol{\xi}_{p+1}^T \hat{\boldsymbol{\beta}}_M$, $\hat{v}_i(t) = \boldsymbol{\xi}_{p+1}^T \hat{\mathbf{b}}_i$, and $\hat{\varphi}_i(t) = \hat{\eta}_M(t) + \hat{v}_i(t)$.

The backfitting algorithm allows us to estimate the population mean function and random-effects curves separately using different bandwidths at each iteration step. This is more efficient than the estimation procedure proposed by Wu and Zhang (2002) in which the population mean function and random-effects curves are simultaneously estimated using the same bandwidth.

It is very important to get good estimates of variance-covariance parameters such as D and σ^2, since the estimates of both population mean and random-effects curve functions involve these variance-covariance components. In simple linear mixed-effects models, Laird, Lange and Stram (1987) proposed using the EM algorithm coupled with the restricted maximum likelihood (REML) approach to estimate variance-covariance parameters. We follow similar ideas to derive our local estimates of variance-covariance parameters. Suppose that $\mathbf{b}_i$ and $\boldsymbol{\varepsilon}_i$ were known, the estimates of variance-covariance parameters would be, by the maximum likelihood approach with normality assumptions,

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{m} \boldsymbol{\varepsilon}_i^T \boldsymbol{\varepsilon}_i, \tag{3.3}$$

$$\hat{D} = \frac{1}{m}\sum_{i=1}^{m} \mathbf{b}_i \mathbf{b}_i^T, \tag{3.4}$$

where $N = \sum_{i=1}^{m} n_i$. In real applications, $\mathbf{b}_i$ and $\boldsymbol{\varepsilon}_i$ are not known, and the EM algorithm can be applied (Davidian and Giltinan, 1995). In the REML estimation, the 'E-step' replaces $\sum_{i=1}^{m} \boldsymbol{\varepsilon}_i^T \boldsymbol{\varepsilon}_i$ and $\sum_{i=1}^{m} \mathbf{b}_i \mathbf{b}_i^T$ by their expectations conditional on the observed values $\mathbf{y}$, under the current estimates of the variance-covariance components, i.e.,

$$E\left[\sum_{i=1}^{m} \boldsymbol{\varepsilon}_i^T \boldsymbol{\varepsilon}_i | \mathbf{y}_i\right] = \sum_{i=1}^{m} \left\{\tilde{\boldsymbol{\varepsilon}}_i^T \tilde{\boldsymbol{\varepsilon}}_i + tr(Cov[\boldsymbol{\varepsilon}_i|\mathbf{y}_i])\right\}, \tag{3.5}$$

$$E\left[\sum_{i=1}^{m} \mathbf{b}_i \mathbf{b}_i^T | \mathbf{y}_i\right] = \sum_{i=1}^{m} \left\{\tilde{\mathbf{b}}_i \tilde{\mathbf{b}}_i^T + Cov[\mathbf{b}_i|\mathbf{y}_i]\right\}, \tag{3.6}$$

where $\tilde{\boldsymbol{\varepsilon}}_i = E[\boldsymbol{\varepsilon}_i|\mathbf{y}_i]$ and $\tilde{\mathbf{b}}_i = E[\mathbf{b}_i|\mathbf{y}_i]$. The 'M-step' consists of solving the equations (3.3) and (3.4) with the unknown quantities being replaced by their current expectations given in (3.5) and (3.6). Therefore, the REML estimators of the variance-covariance parameters in the kernel-weighted LME model $K_h^{1/2}(t_{ij} - t)y_i(t_{ij}) = K_h^{1/2}(t_{ij} - t)X_{ij}^T(\boldsymbol{\beta} + \mathbf{b}_i) + \varepsilon_i(t_{ij})$ are formed as follows.

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{m} \left\{ \left[\mathbf{y}_i - X_i(\hat{\boldsymbol{\beta}} + \hat{\mathbf{b}}_i)\right]^T K_{ih} \left[\mathbf{y}_i - X_i(\hat{\boldsymbol{\beta}} + \hat{\mathbf{b}}_i)\right] + \sigma_0^2 tr\{I_{n_i} - \sigma_0^2 \Lambda_i\} \right\}, \tag{3.7}$$

and

$$\hat{D} = \frac{1}{m}\sum_{i=1}^{m} \left\{ \hat{\mathbf{b}}_i \hat{\mathbf{b}}_i^T + D_0\left(I - X_i^T K_{ih}^{\frac{1}{2}} \Lambda_i K_{ih}^{\frac{1}{2}} X_i D_0\right) \right\}, \tag{3.8}$$

where $\Lambda_i = V_i^{*-1}\{I_{n_i} - K_{ih}^{\frac{1}{2}} X_i P^{-1} X_i^T K_{ih}^{\frac{1}{2}} V_i^{*-1}\}$, $P = \sum_{i=1}^{n}\{X_i^T K_{ih}^{\frac{1}{2}} V_i^{*-1} K_{ih}^{\frac{1}{2}} X_i\}$, $V_i^* = \sigma_0^2 I_{n_i} + K_{ih}^{\frac{1}{2}} X_i D_0 X_i^T K_{ih}^{\frac{1}{2}}$, and $N = \sum_{i=1}^{m} n_i$. Note that, for the backfitting algorithm, the initial values of D_0 and σ_0^2 in Step I can be taken from SAS Procedure PROC MIXED or Splus function LME with fitting the linear mixed-effects model $\mathbf{y}_i = X_i\boldsymbol{\beta} + X_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i$ or the local linear mixed-effects model $K_{ih}\mathbf{y}_i = K_{ih}X_i\boldsymbol{\beta} + K_{ih}X_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i$.

4 Large Sample Theory

In this section, we study the asymptotic properties and inferences of the local marginal likelihood estimator $\hat{\eta}_M(t)$ of the population mean function at a fixed time point $t \in \mathcal{T}$, where $\mathcal{T}$ is the collection of all the design time points. Similar techniques can be used to study the asymptotic properties for $\hat{\eta}_J(t)$ which will be reported elsewhere. For simplicity, here we only consider the local constant ($p = 0$) and local linear ($p = 1$) approximations. The assumptions necessary for studying the asymptotic properties of $\hat{\eta}_M(t)$ are listed as follows.

a. The design time points $j = 1, 2, \cdots, n_i$, t_{ij}, $i = 1, 2, \ldots, m$ are iid with density function $f(\cdot)$.

b. $f(t) \neq 0$, for any time point t in the interior of the support of $f(t)$, and the continuous second derivative $f''(t)$ exists.

c. The population curve $\eta(t)$ has twice-continuous derivatives at t, i.e., $\eta''(t)$ exists and is continuous.

d. $\gamma(s,t) = cov[v_i(s), v_i(t)]$ has twice-continuous derivatives at s and t.

e. The variance function $\sigma^2(\cdot)$ is continuous at t.

f. The kernel function $K(\cdot)$ is a bounded and symmetric density function with bounded support $[-1, 1]$.

g. $\tilde{n} = m/(\sum_{i=1}^m \frac{1}{n_i})$, $N = \sum_{i=1}^m n_i$, $m \to \infty$, $h \to 0$, and

(a) when $p = 0$, $\begin{cases} n_i h^3 \to \infty \text{ and } n_i h^5 \to 0, \text{ for } n_i \to \infty \\ N h^3 \to \infty \text{ and } N h^5 \to 0, \text{ for } n_i < \infty \end{cases}$

(b) when $p = 1$, $\begin{cases} n_i h \to \infty \text{ and } n_i h^3 \to 0, \text{ for } n_i \to \infty \\ N h \to \infty \text{ and } N h^3 \to 0, \text{ for } n_i < \infty \end{cases}$.

h. When $p = 1$, for technical and notational convenience, the matrix D is assumed to be a diagonal matrix with non-zero diagonal elements $\delta_1^2(t)$ and $\delta_2^2(t)$, i.e, $D = diag\{\delta_1^2(t), \delta_2^2(t)\}$.

The asymptotic bias and variance of $\hat{\eta}_M(t)$ are investigated based on the two scenarios, $Case\ I:\ n_i \to \infty$ and $Case\ II:\ n_i < \infty$. For both cases, we have the assumptions $m \to \infty$ and $h \to 0$. For notation simplicity, we define $B_{(r,s)}(K) = \int K^r(u)u^s du$.

Theorem 4.1 ($Case\ I:\ n_i \to \infty$) *Under the assumptions* $(\mathbf{a} - \mathbf{g})$, *the following results are obtained.*

(i) The asymptotic bias and variance of the local likelihood estimator $\hat{\eta}_M(t)$ *with* $p = 0$ *are respectively*

$$\begin{aligned} bias[\hat{\eta}_M(t)] &= \frac{h^2}{2}\left\{ \left(2\eta^{'}(t)f^{'}(t)/f(t) + \eta^{''}(t)\right) B_{(1,2)}(K)\left[1 + O_p\left((\tilde{n}h^3)^{-\frac{1}{2}}\right)\right]\right\}, \\ Var[\hat{\eta}_M(t)] &= \frac{\gamma(t,t)}{m}\left[1 + O_p\left((\tilde{n}h)^{-\frac{1}{2}}\right)\right]. \end{aligned}$$

(ii) The asymptotic mean squared error (MSE) of the local likelihood estimator $\hat{\eta}_M(t)$ *with* $p = 0$ *and the asymptotic optimal bandwidth at time* t *are respectively*

$$\begin{aligned} MSE[\hat{\eta}_M(t)] &= \frac{\gamma(t,t)}{m} + \frac{h^4}{4}\left(2\eta^{'}(t)f^{'}(t)/f(t) + \eta^{''}(t)\right)^2 B^2_{(1,2)}(K) \\ &\quad + O_p\left((m^2\tilde{n}h)^{-\frac{1}{2}}\right) + O_p\left(\sqrt{h^5/\tilde{n}}\right), \\ h_{opt} &= \left\{\frac{1}{m^2\tilde{n}}\left(\frac{C_0^2(t)}{4\left(2\eta^{'}(t)f^{'}(t)/f(t) + \eta^{''}(t)\right)^4 B^4_{(1,2)}(K)}\right)\right\}^{\frac{1}{9}}\left[1 + o_p(1)\right], \end{aligned}$$

where $C_0(t)$ *is a bounded function such that* $O_p\left((m^2\tilde{n}h)^{-\frac{1}{2}}\right) = \frac{1}{\sqrt{m^2\tilde{n}h}}C_0(t)$ *in the function of MSE.*

Theorem 4.2 ($Case\ I:\ n_i \to \infty$) *Under the assumptions* $(\mathbf{a} - \mathbf{h})$, *the following results are obtained.*

(i) The asymptotic bias and variance of the local likelihood estimator $\hat{\eta}_M(t)$ *with* $p = 1$ *are respectively*

$$\begin{aligned} bias[\hat{\eta}_M(t)] &= \frac{h^2}{2}\eta^{''}(t)B_{(1,2)}(K)\left[1 + O_p\left((\tilde{n}h)^{-\frac{1}{2}}\right)\right], \\ Var[\hat{\eta}_M(t)] &= \frac{\gamma(t,t)}{m}\left[1 + O_p\left((\tilde{n}h)^{-\frac{1}{2}}\right)\right]. \end{aligned}$$

(ii) The asymptotic mean squared error (MSE) of the local likelihood estimator $\hat{\eta}_M(t)$ with $p = 1$ and the asymptotic optimal bandwidth at time t are respectively

$$\begin{aligned} MSE[\hat{\eta}_M(t)] &= \frac{\gamma(t,t)}{m} + \frac{h^4}{4}\eta''^2(t)B^2_{(1,2)}(K) + O_p\Big((m^2\tilde{n}h)^{-\frac{1}{2}}\Big) + O_p\Big(\sqrt{h^7/\tilde{n}}\Big), \\ h_{opt} &= \left\{\frac{1}{m^2\tilde{n}}\left(\frac{C_1(t)}{\eta''^4(t)B^4_{(1,2)}(K)}\right)\right\}^{\frac{1}{9}}\Big[1 + o_p(1)\Big], \end{aligned}$$

where $C_1(t)$ is a bounded function such that $O_p\Big((n^2\tilde{n}h)^{-\frac{1}{2}}\Big) = \frac{1}{\sqrt{m^2\tilde{n}h}}C_1(t)$ in the function of MSE.

Note that, if the variance of $v_i(t)$ equals to zero, i.e., $\gamma(t,t) = 0$, then the random effects in model (1.1) are zero and within-subject correlations are also zero. In this case, the variance V_i becomes a diagonal matrix and our estimator $\hat{\eta}_M(t)$ reduces to the working-independent estimator proposed by Lin and Carroll (2000). If $\gamma(t,t)$ is not equal to zero, then our estimator is $\sqrt{m}$-consistent and asymptotically unbiased. Here we note that $(m^2\tilde{n}h)^{-\frac{1}{2}}$ is a lower order than $\sqrt{\frac{h^5}{\tilde{n}}}$ when $p = 0$ and is a higher order than $\sqrt{\frac{h^7}{\tilde{n}}}$ when $p = 1$. When $m \to \infty$ and $n_i \to \infty$, the asymptotic optimal bandwidth for obtaining the local constant and local linear estimators of the population mean curve $\eta(t)$ is in the order $(m^2\tilde{n})^{-\frac{1}{9}}$. We also note that, under the same assumptions $\mathbf{a} - \mathbf{h}$, the asymptotic biases and variances of $\hat{\eta}_M(t)$ in Theorem 4.1 and Theorem 4.2 are the same as those of the local joint likelihood estimator $\hat{\eta}_J(t)$ which is the LLME estimator proposed by Wu and Zhang (2002).

For many longitudinal studies, the assumption $n_i \to \infty$ may not be realistic since the number of measurements per subject is usually small compared to the number of subjects. Thus, in the following, we also give the asymptotic results under the assumption of finite n_i.

Theorem 4.3 $(Case\ II: \ n_i < \infty)$ *Let $\tau^2(t) = \gamma(t,t) + \sigma^2(t)$. Under the assumptions* $(\mathbf{a} - \mathbf{g})$*, the following results are obtained.*

(i) The asymptotic bias and variance of the local likelihood estimator $\hat{\eta}_M(t)$ with $p = 0$ are respectively

$$bias[\hat{\eta}_M(t)] = \frac{h^2}{2}\Big(2\eta'(t)f'(t)/f(t) + \eta''(t)\Big)B_{(1,2)}(K)\Big[1 + O_p\Big((Nh^3)^{-\frac{1}{2}}\Big)\Big],$$

$$Var[\hat{\eta}(t)] = \frac{\tau^2(t)}{Nh}\Big(B_{(2,0)}(K)/f(t)\Big)\Big[1+O_p(h)\Big].$$

(ii) The asymptotic mean squared error (MSE) of the local likelihood estimator $\hat{\eta}_M(t)$ *with* $p = 0$ *and the asymptotic optimal bandwidth at time* t *are respectively*

$$MSE[\hat{\eta}_M(t)] = \frac{\tau^2(t)}{Nh}\Big(B_{(2,0)}(K)/f(t)\Big) + \frac{h^4}{4}\Big(2\eta'(t)f'(t)/f(t) + \eta''(t)\Big)^2 B^2_{(1,2)}(K) + O_p(N^{-1}),$$

$$h_{opt} = \left\{\frac{1}{N}\left(\frac{\tau^2(t)B_{(2,0)}(K)}{f(t)\Big(2\eta'(t)f'(t)/f(t)+\eta''(t)\Big)^2 B^2_{(1,2)}(K)}\right)\right\}^{\frac{1}{5}}\Big[1+o_p(1)\Big].$$

Theorem 4.4 *(Case II :* $n_i < \infty$*) Let* $\tau^2(t) = \gamma(t,t) + \sigma^2(t)$*. Under the assumptions* $(\mathbf{a} - \mathbf{h})$*, the following results are obtained.*

(i) The asymptotic bias and variance of the local likelihood estimator $\hat{\eta}_M(t)$ *with* $p = 1$ *are respectively*

$$bias[\hat{\eta}_M(t)] = \frac{h^2}{2}\eta''(t)B_{(1,2)}(K)\Big[1+O_p\Big((Nh)^{-\frac{1}{2}}\Big)\Big],$$

$$Var[\hat{\eta}_M(t)] = \frac{\tau^2(t)}{Nh}\Big(B_{(2,0)}(K)/f(t)\Big)\Big[1+O_p\Big((Nh)^{-\frac{1}{2}}\Big)\Big].$$

(ii) The asymptotic mean squared error (MSE) of the local likelihood estimator $\hat{\eta}_M(t)$ *with* $p = 1$ *and the asymptotic optimal bandwidth at time* t *are respectively*

$$MSE[\hat{\eta}_M(t)] = \frac{\tau^2(t)}{Nh}\Big(B_{(2,0)}(K)/f(t)\Big) + \frac{h^4}{4}\eta''^2(t)B^2_{(1,2)}(K) + O_p\Big((Nh)^{-\frac{3}{2}}\Big),$$

$$h_{opt} = \left\{\frac{1}{N}\left(\frac{\tau^2(t)B_{(2,0)}(K)}{f(t)\eta''^2(t)B^2_{(1,2)}(K)}\right)\right\}^{\frac{1}{5}}\Big[1+o_p(1)\Big].$$

The above two theorems shows that, with n_i finite, the local likelihood estimator $\hat{\eta}_M(t)$ is $\sqrt{Nh}$ consistent, even when the variance of $v_i(t)$ is equal to zero, i.e., $\gamma(t,t) = 0$ or the random-effects curve $v_i(t)$ itself is 0. This indicates that the within-subject correlation does not affect the convergence of $\hat{\eta}_M(t)$ when n_i is finite, unlike the convergence property of $\hat{\eta}_M(t)$ with $n_i \to \infty$ that the convergence of the estimator depends on the variance of $v_i(t)$.

When $m \to \infty$ with $n_i < \infty$, the asymptotic optimal bandwidth for obtaining the local constant and local linear estimators of the population mean curve $\eta(t)$ is in the order $N^{-\frac{1}{5}}$.

We also summarize the results of asymptotic distribution of $\hat{\eta}_{_M}(t)$ in the following theorem.

Theorem 4.5 *Under the assumptions* $(\mathbf{a}-\mathbf{h})$ *and further assume that* $y_i(t_{ij}) - \eta(t_{ij}) = \upsilon_i(t_{ij}) + \varepsilon_i(t_{ij})$ *are uniformly bounded for all* i *and* j. *For both* $p = 0$ *and* $p = 1$, *the following results are obtained.*

(a) (Case I : $n_i \to \infty$*) For any given time* t *in the interior point of the support of* f *and for both local constant and local linear estimators,* $\hat{\eta}_{_M}(t)$ *follow asymptotically Gaussian distribution with mean* $\eta(t)$ *and variance* $\gamma(t,t)/m$*; that is,*

$$\sqrt{m}\Big(\hat{\eta}_{_M}(t) - \eta(t)\Big) \;\to\; N\Big(0, \gamma(t,t)\Big) \text{ in distribution.} \tag{4.1}$$

(b) (Case II : $n_i < \infty$*) For any given time* t *in the interior point of the support of* f *and for both local constant and local linear estimators,* $\hat{\eta}_{_M}(t)$ *follows asymptotically Gaussian distribution with mean* $\eta(t) + bias[\hat{\eta}(t)]$ *and variance* $\frac{\tau^2(t)}{Nhf(t)}\Big(B_{(2,0)}(K)/f(t)\Big)$, *where* $bias[\hat{\eta}(t)]$ *is presented in Theorem 4.3; that is,*

$$\sqrt{Nh}\Big(\hat{\eta}_{_M}(t) - \eta(t) - bias[\hat{\eta}(t)]\Big) \;\to\; N\bigg(0, \tau^2(t)\Big(B_{(2,0)}(K)/f(t)\Big)\bigg) \text{ in distribution.} \tag{4.2}$$

The asymptotic distributions for both local constant and local linear likelihood estimators of $\eta(t)$ shown in Theorem 4.5 are the same as the asymptotic distributions of LLME estimator of $\eta(t)$ given by Wu and Zhang (2002). For the local likelihood estimator $\hat{\eta}_{_M}(t)$ with $p > 1$, the derivation of the asymptotic distribution requires tedious computations with higher dimensional matrices which are intractable. The asymptotic theories for random effects curve estimators are difficult to study, but are worthy future research topics.

All the sketch proofs for the above theorems are given in Appendix.

5 Simulation Studies

The finite-sample performance of the proposed estimators is evaluated by comparing to the existing estimators using simulation experiments according to the following model,

$$y_i(t) = (\alpha_0 + \alpha_{i0}) + (\alpha_1 + \alpha_{i1})cos(2\pi t) + (\alpha_2 + \alpha_{i2})sin(2\pi t) + \varepsilon_i(t), \ i = 1, \cdots, m, \quad (5.1)$$

where the vector $\boldsymbol{\alpha}_i = (\alpha_{i0}, \alpha_{i1}, \alpha_{i2})^T$ is normally distributed with mean 0 and variance-covariance matrix $diag(\sigma_0^2, \sigma_1^2, \sigma_2^2)$, and $\varepsilon_i(t)$ are normally distributed with mean 0 and variance $\sigma_\varepsilon^2(1+t)$. The design time points are taken between 0 and 1 as $t_j = j/(n+1)$, $j = 1, \cdots, n$, and $\boldsymbol{\alpha}_i$ and $\varepsilon_i(t)$ are independent. In order to obtain imbalanced data, some data points were randomly removed with missing rate λ, and hence the average number of measurements per subject is $\bar{n}_i = n(1-\lambda)$.

For the model (5.1) with $t \neq s$, the within-subject correlation can be calculated as

$$\rho(t,s) \equiv corr\Big[y_i(t), y_i(s)\Big] = \frac{\sigma_0^2 + \sigma_1^2 cos(2\pi t)cos(2\pi s) + \sigma_2^2 sin(2\pi t)sin(2\pi s)}{\sqrt{AB}}, \quad (5.2)$$

where $A = \sigma_0^2 + \sigma_1^2 cos^2(2\pi t) + \sigma_2^2 sin^2(2\pi t) + \sigma_\varepsilon^2(1+t)$ and $B = \sigma_0^2 + \sigma_1^2 cos^2(2\pi s) + \sigma_2^2 sin^2(2\pi s) + \sigma_\varepsilon^2(1+s)$. For simplicity, we assumed that $\sigma_1^2 = \sigma_2^2 \equiv \sigma_*^2$ for our simulations. Under this assumption, the range of the correlation $\rho(t,s)$ can be determined as

$$\frac{\sigma_0^2 - \sigma_*^2}{\sigma_0^2 + \sigma_*^2 + 2\sigma_\varepsilon^2} \leq \rho(t,s) \leq \frac{\sigma_0^2 + \sigma_*^2}{\sigma_0^2 + \sigma_*^2 + \sigma_\varepsilon^2}. \quad (5.3)$$

In our simulation studies, we assumed that $\sigma_*^2 = \sigma_\varepsilon^2 \equiv 1$. Two different values for σ_0 were taken as 1.3 and 3 to control within-subject correlations as (0.15,0.73) and (0.67, 0.91), respectively.

For this particular model (5.1), the population mean and random-effects curve functions are $\eta(t) = \alpha_0 + \alpha_1 cos(2\pi t) + \alpha_2 sin(2\pi t)$ and $\upsilon_i(t) = \alpha_{i0} + \alpha_{i1} cos(2\pi t) + \alpha_{i2} sin(2\pi t)$, respectively. In order to evaluate the performance of different estimators for the population mean function $\eta(t)$ and the individual curve functions $\varphi_i(t) = \eta(t) + \upsilon_i(t)$, we compute the average squared error (ASE) as

$$ASE(\hat{\eta}) = \frac{1}{n} \sum_{i=1}^{m} \left[\frac{1}{n_i} \sum_{j=1}^{n_i} \Big\{ \eta(t_{ij}) - \hat{\eta}(t_{ij}) \Big\}^2 \right], \quad (5.4)$$

and

$$ASE(\hat{\varphi}) = \frac{1}{n}\sum_{i=1}^{m}\left[\frac{1}{n_i}\sum_{j=1}^{n_i}\left\{\varphi_i(t_{ij}) - \hat{\varphi}_i(t_{ij})\right\}^2\right]. \tag{5.5}$$

The following statistics ASEPR and ASEPS are used to evaluate the relative performance of two different estimators:

$$ASEPR(M1/M2) = Average\left\{\frac{ASE_{M1} - ASE_{M2}}{ASE_{M1}}\right\}, \tag{5.6}$$

$$ASEPS(M1/M2) = \frac{\#\{ASE_{M1} > ASE_{M2}\}}{Total\ \#\{Simulation\ Samples\}}, \tag{5.7}$$

where ASE_{M1} and ASE_{M2} are ASEs from the M1 method and the M2 method, respectively. The statistics $ASEPR(M1/M2)$ represents the proportion of ASE reduction by M2 method compared to M1 method and $ASEPS(M1/M2)$ represents the proportion of simulation runs that the ASEs from M2 method are smaller than that from M1 method. If $ASEPR(M1/M2) > 0$ and/or $ASEPS(M1/M2) > 0.5$, it indicates that the M2 method performs better than M1 method in terms of ASE.

In this simulation, the performance of our proposed estimators is evaluated by comparing with the estimators from the two existing methods as following.

Method 1. Local LME (LLME) method (Wu and Zhang, 2002): Obtain the estimators $\hat{\boldsymbol{\beta}}_{WZ}$ and $\hat{\mathbf{b}}_{i,WZ}$ based on SJCV criterion and PTCV criterion, respectively, by fitting a LME model locally.

Method 2. Local GEE (LGEE) method (Lin and Carroll, 2000): Obtain $\hat{\boldsymbol{\beta}}_{LC}$ using LGEE approach proposed by Lin and Carroll (2000) based on SJCV criterion. Note that the LGEE method does not provide individual curve estimators.

Method 3. Local Marginal Likelihood method with Backfitting (LMBF): Obtain $\hat{\boldsymbol{\beta}}_M$ in (2.9) and $\hat{\mathbf{b}}_{i,M}$ using the backfitting algorithm proposed in Section 3.2 based on SJCV criterion and PTCV criterion, respectively.

In this simulation study, we investigate the performance of the three methods listed above using local constant kernel estimators for simplicity. From the LLME and LMBF

Table 1: *The average (standard error) of ASEs for estimating the mean curve function $\eta(t)$ in 500 simulations.*

$(m,\bar{n}_i)$	ρ	LLME	LGEE	LMBF
(30,5)	(0.15,0.73)	0.144 (0.005)	0.147 (0.005)	0.144 (0.005)
	(0.67,0.91)	0.382 (0.018)	0.435 (0.019)	0.380 (0.019)
(30,20)	(0.15,0.73)	0.111 (0.004)	0.114 (0.004)	0.113 (0.004)
	(0.67,0.91)	0.335 (0.017)	0.355 (0.017)	0.338 (0.017)
(70,5)	(0.15,0.73)	0.068 (0.002)	0.071 (0.002)	0.065 (0.002)
	(0.67,0.91)	0.172 (0.009)	0.188 (0.009)	0.171 (0.009)
(70,20)	(0.15,0.73)	0.051 (0.002)	0.053 (0.002)	0.052 (0.002)
	(0.67,0.91)	0.151 (0.008)	0.162 (0.008)	0.154 (0.008)

Table 2: *The ASEPR (standard error) and ASEPS (standard error) for $\hat{\eta}(t)$ by LMBF method with respect to LGEE method and LLME method in 500 simulations.*

		ASEPR		ASEPS	
$(m,\bar{n}_i)$	ρ	LGEE/LMBF	LLME/LMBF	LGEE/LMBF	LLME/LMBF
(30, 5)	(0.15,0.73)	0.017 (0.009)	-0.029 (0.012)	0.610 (0.029)	0.522 (0.022)
	(0.67,0.91)	0.177 (0.019)	-0.042 (0.022)	0.814 (0.017)	0.542 (0.022)
(30, 20)	(0.15,0.73)	0.003 (0.003)	-0.050 (0.009)	0.564 (0.022)	0.452 (0.022)
	(0.67,0.91)	0.086 (0.010)	-0.045 (0.011)	0.798 (0.018)	0.456 (0.022)
(70, 5)	(0.15,0.73)	0.069 (0.012)	0.002 (0.014)	0.742 (0.020)	0.586 (0.022)
	(0.67,0.91)	0.059 (0.019)	-0.136 (0.027)	0.708 (0.020)	0.478 (0.022)
(70, 20)	(0.15,0.73)	0.000 (0.005)	-0.052 (0.009)	0.588 (0.022)	0.408 (0.022)
	(0.67,0.91)	0.089 (0.007)	-0.085 (0.012)	0.850 (0.016)	0.394 (0.022)

methods, we obtained both population mean function and individual curve function estimates. However, the LGEE method only provides the population mean function estimate. To evaluate the performance of these methods, 500 replications were run with each of different sample size scenarios, $(m, \bar{n}_i) = (30, 5)$, $(30, 20)$, $(70, 5)$, and $(70, 20)$. The missing rate λ was chosen to be 0.15.

We report the ASEs of the three estimation methods for the population mean function $\eta(t)$ and their relative performance results in Tables 1 and 2. These results show that LMBF and LLME methods are quite similar in terms of ASEs (Wilcoxon Signed Rank test indicates that the difference in ASEs between these two methods is not significant). However, the proposed LMBF method performs better than the working independence LGEE method in terms of ASEs, especially when the within-subject correlation is strong (Wilcoxon Signed Rank test indicates that there is a significant difference in ASEs between the two methods when $0.67 < \rho < 0.91$). Table 2 shows that, in all simulation scenarios, the LMBF method produced smaller ASEs in more than 50% of the 500 simulation runs compared to the LGEE method, and the ASE improvement is up to 18%. Comparing the LMBF method to the LLME method, for some scenarios, the LMBF method is slightly better, and for some other scenarios, the LLME method is slightly better.

We present the simulation results for random and individual curve function estimates as well as the comparisons between the LMBF and LLME methods in Tables 3 and 4 (the individual curve estimates are not available from the LGEE method). Interestingly, for the individual curve function estimation, our proposed LMBF method performs better than the LLME method in all simulation scenarios. The improvement in the ASEPR ranges $6 - 65\%$ and in the ASEPS, ranges $73 - 92\%$.

In summary, our limited simulation studies show that, for the estimation of population mean function, the proposed LMBF method has a similar performance to the LLME method, but a better performance compared to the LGEE method. For the estimation of individual curve functions, our LMBF method greatly improves the LLME method. We believe that the better performance of the LMBF method may be due to the two factors. One is that we derived our estimators based on more powerful likelihood approach, and

Table 3: *The average (standard error) of ASEs for estimating the random-effects function $\upsilon_i(t)$ and the individual curve $\varphi_i(t)$ in 500 simulations.*

$(m,\bar{n}_i)$	ρ	LLME ($\upsilon_i(t)$)	LMBF ($\upsilon_i(t)$)	LLME ($\varphi_i(t)$)	LMBF ($\varphi_i(t)$)
(30,5)	(0.15,0.73)	0.596 (0.007)	0.551 (0.007)	0.739 (0.005)	0.695 (0.005)
	(0.67,0.91)	0.399 (0.019)	0.354 (0.020)	0.781 (0.005)	0.734 (0.005)
(30,20)	(0.15,0.73)	0.182 (0.005)	0.169 (0.005)	0.293 (0.002)	0.282 (0.002)
	(0.67,0.91)	-0.043 (0.017)	-0.052 (0.017)	0.292 (0.002)	0.286 (0.002)
(70,5)	(0.15,0.73)	0.635 (0.004)	0.616 (0.003)	0.703 (0.003)	0.681 (0.003)
	(0.67,0.91)	0.579 (0.010)	0.559 (0.009)	0.751 (0.004)	0.730 (0.003)
(70,20)	(0.15,0.73)	0.232 (0.002)	0.226 (0.002)	0.283 (0.001)	0.278 (0.001)
	(0.67,0.91)	0.137 (0.008)	0.129 (0.008)	0.288 (0.001)	0.283 (0.001)

Table 4: *The ASEPR (standard error) and ASEPS (standard error) for $\hat{\upsilon}_i(t)$ and $\hat{\varphi}_i(t)$ by LMBF method with respect to LLME method in 500 simulations.*

		LLME/LMBF			
$(m,\bar{n}_i)$	ρ	ASEPR ($\upsilon_i(t)$)	ASEPS ($\upsilon_i(t)$)	ASEPR ($\varphi_i(t)$)	ASEPS ($\varphi_i(t)$)
(30, 5)	(0.15,0.73)	0.069 (0.008)	0.386 (0.017)	0.646 (0.075)	0.866 (0.015)
	(0.67,0.91)	0.046 (0.120)	0.686 (0.021)	0.338 (0.026)	0.918 (0.012)
(30, 20)	(0.15,0.73)	-0.731 (0.810)	0.790 (0.018)	0.201 (0.016)	0.840 (0.016)
	(0.67,0.91)	0.136 (0.166)	0.630 (0.022)	0.056 (0.006)	0.806 (0.018)
(70, 5)	(0.15,0.73)	0.029 (0.002)	0.774 (0.019)	0.609 (0.047)	0.840 (0.016)
	(0.67,0.91)	-0.004 (0.027)	0.750 (0.019)	0.360 (0.030)	0.850 (0.016)
(70, 20)	(0.15,0.73)	0.025 (0.002)	0.812 (0.017)	0.173 (0.013)	0.836 (0.017)
	(0.67,0.91)	-0.490 (0.435)	0.684 (0.021)	0.087 (0.011)	0.730 (0.020)

another is that we used a backfitting approach with EM algorithm to get better estimates for variance-covariance parameters and better bandwidths for population mean and random curve functions. Also we would like to mention that the local marginal estimator (η_M) and local joint estimator (η_J) are compared via numerical simulations (data not shown). The two estimators perform similarly.

6 Application to an AIDS Clinical Study

We apply the proposed local likelihood approach equipped with a backfitting algorithm to an AIDS clinical study developed by the AIDS Clinical Trials Group (ACTG). The study enrolled 517 HIV-1 infected patients in three antiviral treatments. We illustrate the proposed method by applying the methodologies to model CD4 cell response in one of the three treatment arms. In this treatment arm, 166 patients were treated with an highly active antiretroviral therapy (HAART) for 120 weeks during which CD4 cell counts were monitored at weeks 4, 8, and every 8 weeks thereafter. However, each individual patient might not exactly follow the designed schedule, and missing clinical visits for CD4 cell measurements frequently occurred which makes the data set to be a typical longitudinal data set. Since CD4 cell count is an important marker for assessing immunologic response of an antiviral regimen, our objective for the analysis is to characterize the population and individual responses in terms of CD4 cell counts recoveries during the antiviral treatment using the proposed method in this paper. The CD4 cell count data from 166 individual patients during 120 weeks of treatment are plotted in Figure 1.

The number of CD4 cell count measurements per patient varies from 2 to 19 observations. The CD4 cell count ranges from 1 to 1354. For computational stability, we stabilize the large variations of CD4 cell counts and computational algorithms by making transformations $LogCD4 = log\,(\{CD4\ cell\ count\}/100)$ for our model fitting. By fitting the nonparametric mixed-effects model (1.1), we used the LLME and LMBF methods to obtain the estimates of $\eta(t)$ and $v_i(t)$. For estimation of the population mean curve $\eta(t)$, we found that the optimal bandwidths h^*_{SJ} for both methods were 2 weeks, and for estimation of individual variation curves $v_i(t)$, we found that the optimal bandwidths h^*_{PT} for LLME and LMBF

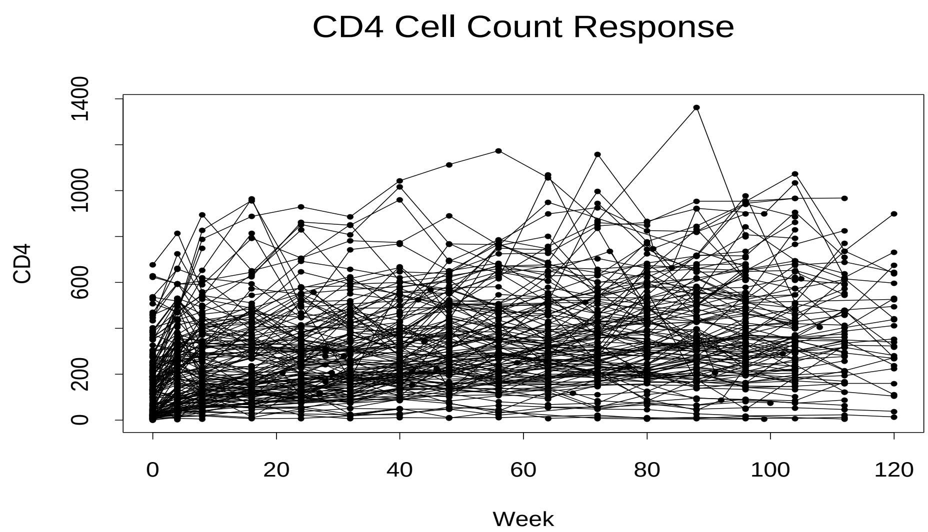

Figure 1: CD4+ Cell Counts from 166 individual patients for 120 weeks of antiviral treatment.

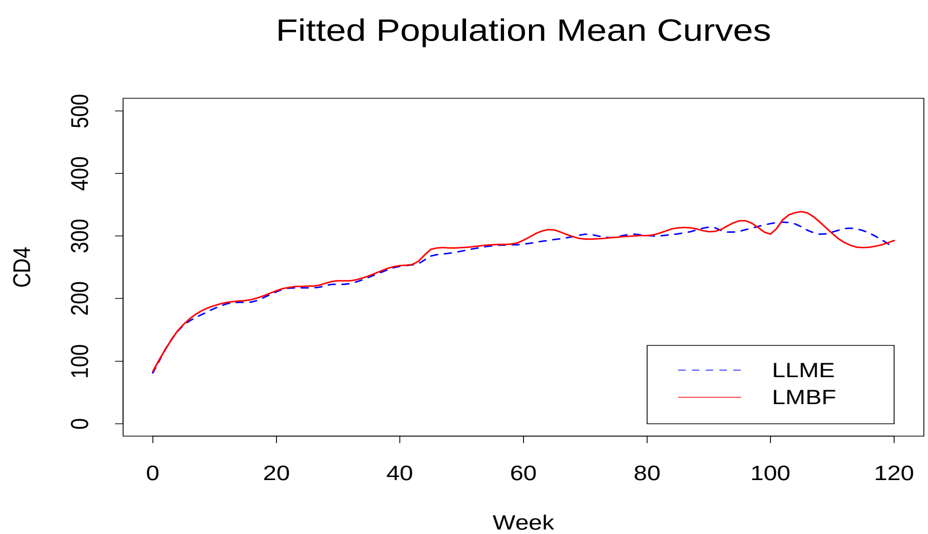

Figure 2: Population Mean Curve Estimates of CD4+ Cell Counts for 166 individual patients

methods were 14.5 and 26.5 weeks, respectively.

In order to assess the differences among the estimates from the three methods, the

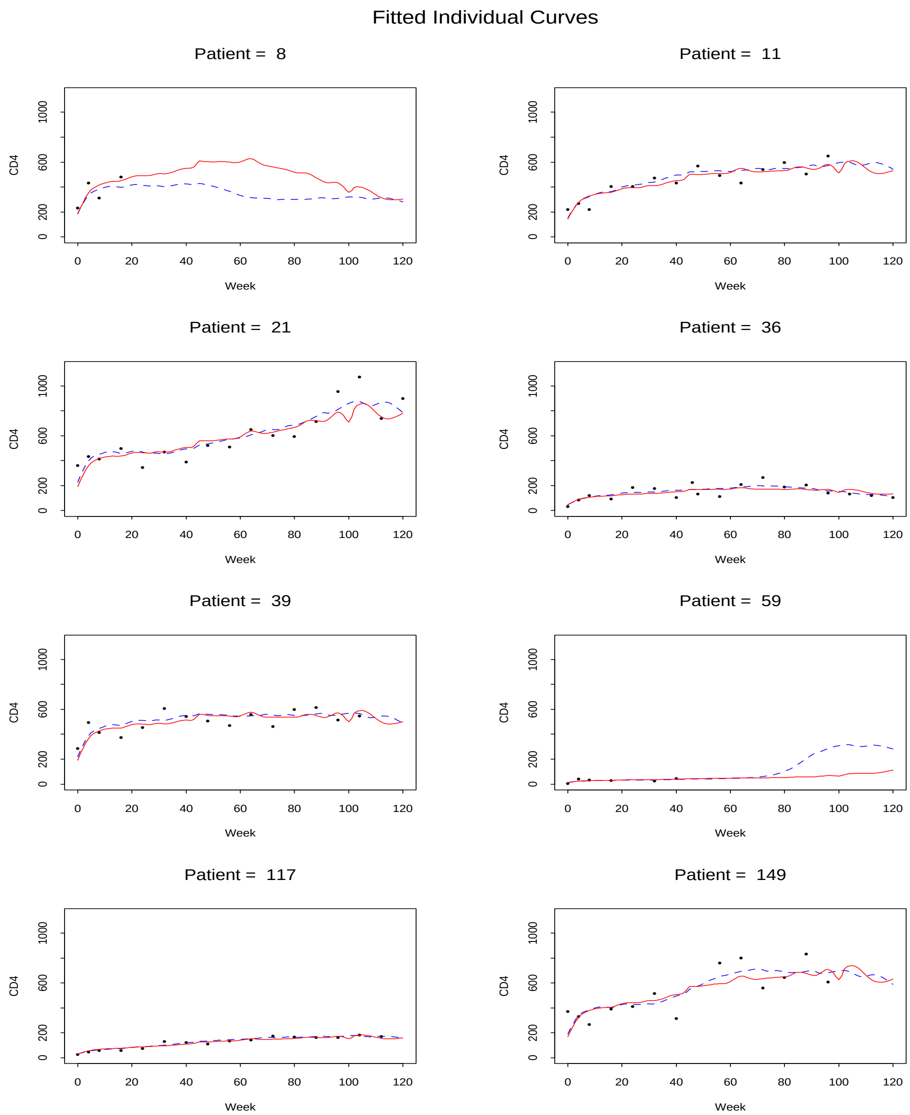

Figure 3: Individual Curve Estimates of CD4 Cell Counts for selected 8 individual patients. —— LMBF; - - - - LLME.

estimated population mean curves $\hat{\eta}(t)$ and estimated individual curve functions $\hat{\varphi}_i(t) = \hat{\eta}(t) + \hat{v}_i(t)$ from selected subjects are plotted in Figure 2 and Figure 3, respectively. From Figure 2, one can see that the estimated population mean curve from the proposed method LMBF is different from the LLME method. Figure 3 presents individual curve estimates from selected subjects which show that the two methods LLME and LMBF give a little difference, but produce very similar trends, except for Patients 8 and 59 who dropped the study early without CD4 cell measurements after 16 and 40 weeks, respectively. In fact, the fitted curves for Patient 8 and 59 after 16 and 40 weeks, respectively, provide the predicted results from the two different methods. The difference in the predicted curves from the two methods may be caused by their difference in the estimated mean curve functions which provide important information for individual subject predictions.

7 Concluding Remarks

We have proposed local likelihood methods and a backfitting algorithm for the NPME models with longitudinal data. We have shown that the combination of the local likelihood method and the backfitting algorithm improves the estimators of nonparametric mean and individual curves for longitudinal data compared to the recently proposed LLME and LGEE methods (Wu and Zhang 2002, Lin and Carroll 2000). This is due mainly to the fact that the local likelihood approach preserves the properties of the likelihood and the proposed backfitting algorithm enables us to apply the optimal bandwidths for different parameters and to efficiently estimate variance-covariance parameters. Furthermore, our local kernel likelihood approach is flexible to apply kernel functions to the likelihood functions in different ways which may result in different estimators. A better estimator can be selected by comparing the asymptotic properties and finite-sample performance of those estimators.

It is also interesting to notice that in a recent paper, Wang (2003) proposed a clever kernel GEE estimator that incorporates the within-subject correlation by an iterative smoothing procedure. This estimator is shown to be more efficient than the working independent kernel GEE estimator (Lin and Carroll 2000). Lin et al. (2003) compared the Wang's estimator to the smoothing spline estimator in which the within-subject correlation is also considered. They concluded that, if the within-subject correlation is incorporated, the kernel GEE estimator or Wang's estimator is asymptotically equivalent to the smoothing spline estimator and the two estimators also have similar behavior in finite samples. This suggests that the local kernel estimator can achieve the same efficiency as the 'global' or 'non-local' estimators if the within-subject correlation is appropriately considered in longitudinal data analysis.

Acknowledgements

We thank Drs. Jinting Zhang and Hua Liang for their helpful discussions on the results in this paper. We are also grateful to Professors Xinhong Lin and Naisyin Wang for sharing their unpublished manuscripts with us. We appreciate the helpful comments and suggestions from the referee and the editor. This work was partially supported by the National Institutes of Health (NIH) research grants RO1 AI45356, RO1 AI052765, RO1 AI055290 and U01 AI38855.

Appendix

Proof of Theorem 4.1 and 4.2 From the objective function (2.7) in Section 2, we denote

$$Q(\boldsymbol{\beta}) \equiv l'(\boldsymbol{\beta}) = -\sum_{i=1}^{m} X_i^T K_{ih}^{1/2} V_i^{-1} K_{ih}^{1/2} (y_i - X_i\boldsymbol{\beta}). \tag{7.1}$$

Since $Q(\hat{\boldsymbol{\beta}}) = 0$ and by Taylor expansion, $0 \approx Q(\boldsymbol{\beta}) + Q'(\boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$, and we have $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = -(Q'(\boldsymbol{\beta}))^{-1}Q(\boldsymbol{\beta})$, where

$$\begin{aligned} Q(\boldsymbol{\beta}) &= -\sum_{i=1}^{m} X_i^T K_{ih}^{1/2} V_i^{-1} K_{ih}^{1/2} (\mathbf{y}_i - X_i\boldsymbol{\beta}), \\ Q'(\boldsymbol{\beta}) &= \sum_{i=1}^{m} X_i^T K_{ih}^{1/2} V_i^{-1} K_{ih}^{1/2} X_i. \end{aligned}$$

To compute the bias and variance of the estimator $\hat{\boldsymbol{\beta}}$, we let $\mathcal{T}$ be the collection of all the design time points and obtain $E[Q(\boldsymbol{\beta})|\mathcal{T}]$ and $Var[Q(\boldsymbol{\beta})|\mathcal{T}]$. We let $\eta_i = \left(\eta_{i1}, \cdots, \eta_{in_i}\right)^T$, $G_{x1i} = X_i^T K_{ih}^{1/2} R_i^{-1} K_{ih}^{1/2} X_i$, $G_{x2i} = X_i^T K_{ih}^{1/2} R_i^{-1} X_i$, $G_{\eta 1i} = X_i^T K_{ih}^{1/2} R_i^{-1} K_{ih}^{1/2} \eta_i$, and $G_{\eta 2i} = X_i^T R_i^{-1} K_{ih}^{1/2} \eta_i$, where $\eta_{ij} = \eta(t_{ij}) - \sum_{r=0}^{p} (t_{ij} - t)^r \eta^{(r)}(t)/r!$. Then,

$$\begin{aligned} E[Q(\boldsymbol{\beta})|\mathcal{T}] &= -\sum_{i=1}^{m} X_i^T K_{ih}^{1/2} V_i^{-1} K_{ih}^{1/2} \eta_i \\ &= -\sum_{i=1}^{m} \left\{ G_{\eta 1i} - G_{x2i}(D^{-1} + X_i^T R_i^{-1} X_i)^{-1} G_{\eta 2i} \right\}, \end{aligned}$$

$$
\begin{aligned}
Var[Q(\boldsymbol{\beta})|\mathcal{T}] &= \sum_{i=1}^{m}\Big\{X_i^T K_{ih}^{1/2} V_i^{-1} K_{ih}^{1/2}\Big(\Gamma_i + R_i\Big) K_{ih}^{1/2} V_i^{-1} K_{ih}^{1/2} X_i\Big\} \\
&= \sum_{i=1}^{m}\Big[X_i^T K_{ih}^{1/2} R_i^{-1} K_{ih}^{1/2}\Big(\Gamma_i + R_i\Big) K_{ih}^{1/2} R_i^{-1} K_{ih}^{1/2} X_i \\
&\quad - 2G_{x2i}\Lambda_i^{-1}(t) X_i^T R_i^{-1} K_{ih}^{1/2}\Big(\Gamma_i + R_i\Big) K_{ih}^{1/2} R_i^{-1} K_{ih}^{1/2} X_i \\
&\quad + G_{x2i}\Lambda_i^{-1}(t) X_i^T R_i^{-1} K_{ih}^{1/2}\Big(\Gamma_i + R_i\Big) K_{ih}^{1/2} R_i^{-1} X_i \Lambda_i^{-1}(t) G_{x2i}\Big],
\end{aligned}
$$

where $\Lambda_i(t) = D^{-1} + X_i^T R_i^{-1} X_i$.

For $p = 0$,

$$
Q'(\eta(t)) = \sum_{i=1}^{m}\Big\{\sum_{j=1}^{n_i} K_{ij}/\sigma^2(t_{ij}) - \Lambda_i^{-1}(t)\Big(\sum_{j=1}^{n_i} K_{ij}^{1/2}/\sigma^2(t_{ij})\Big)^2\Big\}, \tag{7.2}
$$

$$
E[Q(\eta(t))|\mathcal{T}] = -\sum_{i=1}^{m}\Big\{\sum_{j=1}^{n_i} K_{ij}\eta_{ij}/\sigma^2(t_{ij}) - \Lambda_i^{-1}(t)\Big(\sum_{j=1}^{n_i} K_{ij}^{1/2}/\sigma^2(t_{ij})\Big)\Big(\sum_{j=1}^{n_i} K_{ij}^{1/2}\eta_{ij}/\sigma^2(t_{ij})\Big)\Big\}, \tag{7.3}
$$

$$
\begin{aligned}
Var[Q(\boldsymbol{\beta})|\mathcal{T}] &= \sum_{i=1}^{m}\Bigg[\sum_{j=1}^{n_i}\left(\frac{K_{ij}^2}{\sigma^4(t_{ij})}\right)\tau^2(t_{ij}) + \sum_{j=1}^{n_i}\sum_{j'\neq j}^{n_i}\left(\frac{K_{ij}}{\sigma^2(t_{ij})}\right)\gamma(t_{ij}, t_{ij'})\left(\frac{K_{ij'}}{\sigma^2(t_{ij'})}\right) \\
&\quad - 2\Lambda_i^{-1}(t)\left(\sum_{j=1}^{n_i}\frac{K_{ij}^{1/2}}{\sigma^2(t_{ij})}\right)\Bigg\{\sum_{j=1}^{n_i}\left(\frac{K_{ij}^{3/2}}{\sigma^4(t_{ij})}\right)\tau^2(t_{ij}) \\
&\quad + \sum_{j=1}^{n_i}\sum_{j'\neq j}^{n_i}\left(\frac{K_{ij}^{1/2}}{\sigma^2(t_{ij})}\right)\gamma(t_{ij}, t_{ij'})\left(\frac{K_{ij'}}{\sigma^2(t_{ij'})}\right)\Bigg\} \\
&\quad + \Lambda^{-2}(t)\left(\sum_{j=1}^{n_i}\frac{K_{ij}^{1/2}}{\sigma^2(t_{ij})}\right)^2\Bigg\{\sum_{j=1}^{n_i}\left(\frac{K_{ij}}{\sigma^4(t_{ij})}\right)\tau^2(t_{ij}) \\
&\quad + \sum_{j=1}^{n_i}\sum_{j'\neq j}^{n_i}\left(\frac{K_{ij}^{1/2}}{\sigma^2(t_{ij})}\right)\gamma(t_{ij}, t_{ij'})\left(\frac{K_{ij'}^{1/2}}{\sigma^2(t_{ij'})}\right)\Bigg\}\Bigg]
\end{aligned}
$$

where $\Lambda_i(t) = \delta^{-2}(t) + \sum_{j=1}^{n_i} 1/\sigma^2(t_{ij})$ and $\tau^2(t_{ij}) = \gamma(t_{ij}, t_{ij}) + \sigma^2(t_{ij})$.

Note that for any random variable R, we have $R = E[R] + O_p\Big(\sqrt{Var(R)}\Big)$, and applying this approximation leads to the followings. For $r, s = 0, \frac{1}{2}, 1, \frac{3}{2}, 2$,

$$
\sum_{j=1}^{n_i} K_{ij}^r/\sigma^2(t_{ij}) = \frac{n_i h^{1-r} f(t)}{\sigma^2(t)} B_{(r,0)}(K)\Big[1 + O_p\Big((n_i h)^{-\frac{1}{2}}\Big)\Big],
$$

$$\sum_{j=1}^{n_i} K_{ij}^r \eta_{ij}/\sigma^2(t_{ij}) = \frac{n_i h^{3-r}}{\sigma^2(t)} \left(\eta^{'}(t) f^{'}(t) + \frac{1}{2}\eta^{''}(t) f(t)\right) B_{(r,2)}(K) \left[1 + O_p((n_i h^3)^{-\frac{1}{2}})\right],$$

$$\begin{aligned}\sum_{j=1}^{n_i} \sum_{j' \neq j}^{n_i} \left(\frac{K_{ij}^r}{\sigma^2(t_{ij})}\right) \gamma(t_{ij}, t_{ij'}) \left(\frac{K_{ij'}^s}{\sigma^2(t_{ij'})}\right) &= \frac{n_i(n_i-1)h^{2-r-s} f^2(t)}{\sigma^4(t)} \\ &\quad \cdot \gamma(t,t) B_{(r,0)}(K) B_{(s,0)}(K) \left[1 + O_p(h^2)\right],\end{aligned}$$

$$\begin{aligned}\Lambda_i(t) &= \delta^{-2}(t) + \sum_{j=1}^{n_i} 1/\sigma^2(t_{ij}) \\ &= n_i \int \frac{1}{\sigma^2(s)} f(s) ds \left[1 + o_p(1)\right] \\ &= n_i \kappa(\sigma^2) \left[1 + o_p(1)\right],\end{aligned}$$

where $B_{(r,s)}(K) = \int K^r(u) u^s du$ and $\kappa(\sigma^2) = \int \frac{1}{\sigma^2(u)} f(u) du$.

Thus, $Q^{'}(\eta(t))$, $E[Q(\eta(t))|\mathcal{T}]$ and $Var[Q(\eta(t))|\mathcal{T}]$ are found as

$$\begin{aligned}Q^{'}(\eta(t)) &= \frac{N f(t)}{\sigma^2(t)} \Big[1 - \frac{h f(t)}{\sigma^2(t)\kappa(\sigma^2)} B^2_{(\frac{1}{2},0)}(K) + O_p\Big((\tilde{n}h)^{-\frac{1}{2}}\Big)\Big], \\ E[Q(\eta(t))|\mathcal{T}] &= -\frac{N h^2}{\sigma^2(t)} \left(\eta^{'}(t) f^{'}(t) + \frac{1}{2}\eta^{''}(t) f(t)\right) B_{(1,2)}(K) \Big[1 + O_p\Big((\tilde{n}h^3)^{-\frac{1}{2}}\Big)\Big], \\ Var[Q(\eta(t))|\mathcal{T}] &= \frac{(\sum_{i=1}^m n_i^2) f^2(t)}{\sigma^4(t)} \gamma(t,t) \Big\{\Big(1 - \frac{h f(t)}{\sigma^2(t)\kappa(\sigma^2)} B^2_{(\frac{1}{2},0)}(K)\Big)^2 + O_p(h^2)\Big\}.\end{aligned}$$

Therefore, the rest of the proof progresses with simple algebra. Theorem 1 is proved.

When $p = 1$, the matrices G_{x1i}, G_{x2i} and $\Lambda_i(t)$ are 2×2, and the vectors $G_{\eta 1i}$ and $G_{\eta 2i}$ are 2×1. These can be shown as

$$\begin{aligned}G_{x1i} &= \begin{pmatrix} \sum_{j=1}^{n_i} K_{ij}/\sigma^2(t_{ij}) & \sum_{j=1}^{n_i} (t_{ij}-t) K_{ij}/\sigma^2(t_{ij}) \\ \sum_{j=1}^{n_i} (t_{ij}-t) K_{ij}/\sigma^2(t_{ij}) & \sum_{j=1}^{n_i} (t_{ij}-t)^2 K_{ij}/\sigma^2(t_{ij}) \end{pmatrix}, \\ G_{x2i} &= \begin{pmatrix} \sum_{j=1}^{n_i} K_{ij}^{1/2}/\sigma^2(t_{ij}) & \sum_{j=1}^{n_i} (t_{ij}-t) K_{ij}^{1/2}/\sigma^2(t_{ij}) \\ \sum_{j=1}^{n_i} (t_{ij}-t) K_{ij}^{1/2}/\sigma^2(t_{ij}) & \sum_{j=1}^{n_i} (t_{ij}-t)^2 K_{ij}^{1/2}/\sigma^2(t_{ij}) \end{pmatrix}, \\ G_{\eta 1i} &= \begin{pmatrix} \sum_{j=1}^{n_i} K_{ij}\eta_{ij}/\sigma^2(t_{ij}) \\ \sum_{j=1}^{n_i} (t_{ij}-t) K_{ij}\eta_{ij}/\sigma^2(t_{ij}) \end{pmatrix}, \\ G_{\eta 2i} &= \begin{pmatrix} \sum_{j=1}^{n_i} K_{ij}^{1/2}\eta_{ij}/\sigma^2(t_{ij}) \\ \sum_{j=1}^{n_i} (t_{ij}-t) K_{ij}^{1/2}\eta_{ij}/\sigma^2(t_{ij}) \end{pmatrix},\end{aligned}$$

$$\Lambda_i(t) = \begin{pmatrix} \delta_1^{-2}(t) + \sum_{j=1}^{n_i} 1/\sigma^2(t_{ij}) & \sum_{j=1}^{n_i} 1(t_{ij}-t)/\sigma^2(t_{ij}) \\ \sum_{j=1}^{n_i}(t_{ij}-t)/\sigma^2(t_{ij}) & \delta_2^{-2}(t) + \sum_{j=1}^{n_i}(t_{ij}-t)^2/\sigma^2(t_{ij}) \end{pmatrix}.$$

Using the approximation techniques as in the proof of Theorem 1 above and some matrix computations leads to the completion of proof for Theorem 4.

Proof of Theorem 4.3 and 4.4 When n_i is finite and $p = 0$, we have the following results for $r, s = 0, \frac{1}{2}, 1, \frac{3}{2}, 2$, with applying the approximation $R = E[R] + O_p\Big(\sqrt{Var(R)}\Big)$ for a random variable R.

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i} K_{ij}^r/\sigma^2(t_{ij}) = \frac{Nh^{1-r}f(t)}{\sigma^2(t)} B_{(r,0)}(K)\Big[1 + O_p\Big((Nh)^{-\frac{1}{2}}\Big)\Big],$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i} K_{ij}^r \eta_{ij}/\sigma^2(t_{ij}) = \frac{Nh^{3-r}}{\sigma^2(t)}\Big(\eta^{'}(t)f^{'}(t) + \frac{1}{2}\eta^{''}(t)f(t)\Big) B_{(r,2)}(K)\Big[1 + O_p((Nh^3)^{-\frac{1}{2}})\Big],$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sum_{j'\neq j}\left(\frac{K_{ij}^r}{\sigma^2(t_{ij})}\right)\gamma(t_{ij}, t_{ij'})\left(\frac{K_{ij'}^s}{\sigma^2(t_{ij'})}\right) = \frac{n_i(n_i-1)h^{2-r-s}f^2(t)}{\sigma^4(t)}\gamma(t,t)B_{(r,0)}(K)B_{(s,0)}(K) \cdot\left[1 + O_p\Big(\big(\sum_{i=1}^{m} n_i(n_i-1)h^2\big)^{-\frac{1}{2}}\Big)\right],$$

where $\eta^{'}(t)$ and $\eta^{''}(t)$ stand for the first and second derivatives of $\eta(t)$, respectively. And the $\Lambda_i(t)$ is finite with $n_i < \infty$ and can be written as

$$\Lambda_i(t) = \delta^{-2}(t) + \sum_{j=1}^{n_i}\frac{1}{\sigma^2(t_{ij})} = \delta^{-2}(t)\Lambda_{\delta i}(t),$$

where $\Lambda_{\delta i}(t) = 1 + \delta^2(t)\sum_{j=1}^{n_i}\frac{1}{\sigma^2(t_{ij})}$. Now, $\Lambda_{\delta i}^{-1}(t)$ can be expressed as $\Lambda_{\delta i}^{-1}(t) = 1 - \phi_1(t)$, where $0 < \phi_i(t) < 1$. Thus we have

$$Q^{'}(\eta(t)) = \sum_{i=1}^{m}\Big\{G_{x1i} - \delta^2(t)\Big[G_{x2i}^2 - \phi_i(t)G_{x2i}^2\Big]\Big\},$$

$$E[Q(\boldsymbol{\beta})|\mathcal{T}] = -\sum_{i=1}^{m}\Big\{G_{\eta 1i} - \delta^2(t)\Big[G_{x2i}^2 - \phi_i(t)G_{\eta 2i}^2\Big]\Big\},$$

$$Var[Q(\boldsymbol{\beta})|\mathcal{T}] = \sum_{i=1}^{m}\Big[X_i^T K_{ih}^{1/2} R_i^{-1} K_{ih}^{1/2}\Big(\Gamma_i + R_i\Big)K_{ih}^{1/2} R_i^{-1} K_{ih}^{1/2} X_i - 2\delta^2(t)G_{x2i}X_i^T R_i^{-1} K_{ih}^{1/2}\Big(\Gamma_i + R_i\Big)K_{ih}^{1/2} R_i^{-1} K_{ih}^{1/2} X_i$$

$$
\begin{aligned}
&+ \delta^2(t) G_{x2i} X_i^T R_i^{-1} K_{ih}^{1/2}\Big(\Gamma_i + R_i\Big) K_{ih}^{1/2} R_i^{-1} X_i G_{x2i} \\
&+ \phi_i(t)\Big\{ 2\delta^2(t) G_{x2i} X_i^T R_i^{-1} K_{ih}^{1/2}\Big(\Gamma_i + R_i\Big) K_{ih}^{1/2} R_i^{-1} K_{ih}^{1/2} X_i \\
&- \delta^2(t) G_{x2i} X_i^T R_i^{-1} K_{ih}^{1/2}\Big(\Gamma_i + R_i\Big) K_{ih}^{1/2} R_i^{-1} X_i G_{x2i} \Big\}\Big],
\end{aligned}
$$

Thus, simple algebra shows the following results.

$$
\begin{aligned}
Q^{'}(\eta(t)) &= \frac{Nf(t)}{\sigma^2(t)}\Big[\Big(1 - \frac{\delta^2(t)}{\sigma^2(t)}\Big) + \frac{(\sum_{i=1}^m n_i(n_i-1))hf(t)\delta^2(t)}{N\sigma^2(t)} B^2_{(\frac{1}{2},0)}(K) + O_p\Big((Nh)^{-\frac{1}{2}}\Big)\Big], \\
E[Q(\eta(t))|\mathcal{T}] &= -\frac{Nh^2}{\sigma^2(t)}\Big(\eta^{'}(t)f^{'}(t) + \frac{1}{2}\eta^{''}(t)f(t)\Big)B_{(1,2)}(K)\Big[\Big(1 - \frac{\delta^2(t)}{\sigma^2(t)}\Big) \\
&\quad + \frac{(\sum_{i=1}^m n_i(n_i-1))hf(t)\delta^2(t)}{N\sigma^2(t)} \frac{B_{(\frac{1}{2},0)}(K)B_{(\frac{1}{2},2)}(K)}{B_{(1,2)}(K)} + O_p\Big((Nh^3)^{-\frac{1}{2}}\Big)\Big], \\
Var[Q(\boldsymbol{\beta})|\mathcal{T}] &= \frac{Nf(t)}{\sigma^4(t)h}\tau^2(t)B_{(2,0)}(K)\Big[\Big(1 - \frac{\delta^2(t)}{\sigma^2(t)}\Big)^2 + O_p\Big(h\Big)\Big].
\end{aligned}
$$

Therefore, by using the relations $bias(\hat{\eta}(t)) = -(Q^{'}(\eta(t)))^{-1}E[Q(\eta(t))|\mathcal{T}]$ and $Var[\hat{\eta}(t)] = (Q^{'}(\eta(t)))^{-2}Var[Q(\eta(t))|\mathcal{T}]$, the proof of Theorem 3 is completed.

For $p = 1$, the procedure of proof for Theorem 4 is similar to the proof for Theorem 3, with noticing that $I + DX_i^T R_i^{-1} X_i$ is symmetric and positive definite and the inverse can be written as $(I + DX_i^T R_i^{-1} X_i)^{-1} = I - A_i(t)$, where $A_i(t)$ is a matrix with the determinant between 0 and 1. This completes the proof.

Proof of Theorem 4.5 To show part (a), Lindeberg central limit theorem is applied. Let ξ_i is a $(m \times 1)$ vector of random variables with mean 0 and covariance I.

$$
\begin{aligned}
\hat{\eta}(t) - E[\hat{\eta}(t)] &= \Big(Q^{'}(\eta(t))\Big)^{-1}\Big\{\sum_{i=1}^m X_i^T K_{ih}^{1/2} V_i^{-1} K_{ih}^{1/2}(\mathbf{y}_i - \eta_i)\Big\} \\
&= \frac{1}{m}\sum_{i=1}^m U_i,
\end{aligned}
$$

where $U_i = m\Big(Q^{'}(\eta(t))\Big)^{-1}\Big(X_i^T K_{ih}^{1/2} V_i^{-1} K_{ih}^{1/2}(\mathbf{y}_i - \eta_i)\Big)$.

Let $C_i^2 = \sum_{i=1}^m Var[U_i]$. Then we have

$$
C_m^2 = \sum_{i=1}^m m^2\Big(Q^{'}(\eta(t))\Big)^{-1}\Big[X_i^T K_{ih}^{1/2} V_i^{-1} K_{ih}^{1/2}(\Gamma_i + R_i) K_{ih}^{1/2} V_i^{-1} K_{ih}^{1/2} X_i\Big]\Big(Q^{'}(\eta(t))\Big)^{-1}
$$

$$
\begin{aligned}
&= m^2 \frac{\sum_{i=1}^m n_i^2}{N^2} \gamma(t,t) \Big[1 + o_p(1) \Big] \\
&\longrightarrow \infty, \text{ as } m \to \infty,
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{Var[U_i]}{C_m^2} &= \frac{n_i^2}{m^2 (\sum_{i=1}^m n_i^2)} \gamma(t,t) \Big[1 + o_p(1) \Big] \\
&\longrightarrow 0, \text{ as } m \to \infty.
\end{aligned}
$$

And, since $y_i(t_{ij}) - \eta(t_{ij}) = \upsilon_i(t_{ij}) + \varepsilon_i(t_{ij}) < \infty$ for all i and j, Lindeberg condition satisfies. Because $E[\hat{\eta}(t)] - \eta(t) = o_p(1)$, it completes the proof. For proof of part (b) will follow the similar lines.

References

Davidian, M., and Giltinan, D.M. (1995), *Nonlinear Models for Repeated Measurement Data*, London: Chapman and Hall.

Diggle, P.J., Liang, K.Y., and Zeger, S.L. (1994), *Analysis of Longitudinal data*, Oxford: Oxford University Press.

Fan, J., Farmen, M., and Gijbels, I. (1998), "Local Maximum Likelihood Estimation and Inference," *Journal of the Royal Statistical Society*, Ser. B, 60, 591-608.

Fan, J., and Zhang, J.T. (2000), "Two-Step Estimation of Functional Linear Models with Applications to Longitudinal Data," *Journal of Royal Statistical Society* Ser. B, 62, 303-322.

Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and its Applications*, London: Chapman and Hall.

Hoover, D.R., Rice, J.A., Wu, C.O., and Yang, L.P. (1998), "Nonparametric Smoothing Estimates of Time-Varying Coefficients Models With Longitudinal Data," *Biometrika*, 85, 809-822.

Huang, J.Z., Wu, C.O., and Zhou, L. (2002), "Varing-Coefficient Models and Basis Function Approximations for the Analysis of Repeated Measurements." *Biometrika*, 89, 111-128.

Laird, N., Lange, N., and Stram, D. (1987), "Maximum Likelihood Computations With Repeated Measures: Application of the EM Algorithm", *Journal of the American Statistical Association*, 82, 97-105.

Lin, X., and Carroll, R.J. (2000), "Nonparametric Function Estimation for Clustered Data When the Predictor is Mesured Without/With Error," *Journal of the American Statistical Association*, 95, 520-534.

Lin, X., Wang, N., Welsh, A.H., and Carroll, R.J. (2003), "Equivalent Kernels of Smoothing Splines in Nonparametric Regression for Cluster/Longitudinal Data," Preprint, downloaded from *http://www.bepress.com/umichbiostat/*.

Rice, J.A., and Silverman, B.W. (1991), "Estimating the Mean and Covariance Structure Non-parametrically When the Data are Curves," *Journal of the Royal Statistical Society*, Ser. B, 53, 233-243.

Rice, J.A., and Wu, C.O. (2001), "Nonparameteric Mixed Effects Models for Unequally Sampled Noisy Curves," *Biometrics*, 57, 253-259.

Shi, M., Weiss, R.E., and Taylor, J.M.G. (1996), "An Analysis of Paediatric CD4 counts for Acquired Immune Deficiency Syndrome using Flexible Random Curves," *Applied Statistics*, 45, 151-163.

Staniswalis, J.G. (1989), "The Kernel Estimate of a Regression Function in Likelihood-Based Models," *Journal of the American Statistical Association*, 84, 276-283.

Staniswalis, J.G. and Lee, J.J. (1998), "Nonparametric Regression Analysis of Longitudinal Data," *Journal of the American Statistical Association*, 93, 1403-1418.

Tibshirani, R. and Hastie, T. (1987), "Local Likelihood Estimation," *Journal of the American Statistical Association*, 82, 559-567.

Vonesh, E.F., and Chinchilli, V.M. (1996), *Linear and Nonlinear Models for the Analysis of Repeated Measurements*, New York: Marcel Dekker, Inc.

Wang, N. (2003), "Marginal Nonparametric Kernel Regression Accounting for within-subject correlation," *Biometrika*, 90, 43-52.

Wild, C.J., and Yee, T.W. (1996), "Additive Extensions to Generalized Estimating Equation Methods," *Journal of the Royal Statistical Society*, Ser. B, 58, 711-725.

Wu, H., and Zhang, J.T. (2002), "Local Polynomial Mixed-Effects Models for Longitudinal Data," *Journal of the American Statistical Association*, 97, 883-897.