Ann Inst Stat Math (2009) 61:27–46
DOI 10.1007/s10463-007-0146-0

Generalized partially linear mixed-effects models incorporating mismeasured covariates

Hua Liang

Received: 7 February 2006 / Revised: 18 January 2007 / Published online: 10 August 2007

Abstract In this article we consider a semiparametric generalized mixed-effects model, and propose combining local linear regression, and penalized quasilikelihood and local quasilikelihood techniques to estimate both population and individual parameters and nonparametric curves. The proposed estimators take into account the local correlation structure of the longitudinal data. We establish normality for the estimators of the parameter and asymptotic expansion for the estimators of the nonparametric part. For practical implementation, we propose an appropriate algorithm. We also consider the measurement error problem in covariates in our model, and suggest a strategy for adjusting the effects of measurement errors. We apply the proposed models and methods to study the relation between virologic and immunologic responses in AIDS clinical trials, in which virologic response is classified into binary variables. A dataset from an AIDS clinical study is analyzed.

Keywords AIDS clinical trial · Generalized linear mixed-effects models · Linear mixed-effects model · Local linear · Local quasilikelihood · Longitudinal data · Measurement error · Penalized quasilikelihood

1 Introduction

In recent years, the relation between viral load and CD4 cell counts has been well studied (Wu et al. 1999; Wu 2002; Liang et al. 2003). These studies have investigated the concordance and discordance between virologic and immunologic variables and may help clinicians more deeply understand AIDS pathogenesis and improve therapy

H. Liang (✉)
Department of Biostatistics and Computational Biology, University of Rochester,
Rochester, NY 14642, USA
e-mail: hliang@bst.rochester.edu

regimes. Analysis of these AIDS data poses statistical challenges. A common feature of the markers, such as viral load measurements and CD4+ cell counts, used to evaluate antiviral therapy is that their measurement produces longitudinal data-that is, a series of independent and dependent variables obtained by taking repeated measurements, over time, from any one subject. Consideration of the features of longitudinal data, such as within-subject and between-subject variations and the correlation structure, is of practical interest.

Antiretroviral therapy for HID-1 infected patients has greatly improved in recent years. Administration of drug cocktails consisting of three or more drugs may reduce and maintain virus load copies below the detection limit in many patients, although it is unlikely that combination therapies alone can eradicate HID in infected patients because long-lived infected cells and sites within the body exist where drugs may not be effective. Because of the success of highly active antiretroviral therapy for HID infection, viral load in AIDS patients is suppressed and reaches to magnitudes that are below the limit of quantification, and AIDS becomes a chronic disease. Clinicians and patients are therefore sometimes only interested in maintaining the viral load below the detection limit, and in how the immune system (measured by CD4 cell counts) relates to this suppression.

Our primary goal in this paper is to propose a generalized partially linear mixed-effects model (GPLMeM) to study the relation between binary viral load measurement and CD4 cell count. Our focus is not only on the population characteristic but also on individual diversities. This is particularly important in AIDS research because that there is generally a large between-subject variation, which indicates the importance of estimating each individual's parameters for the individualization of treatment management and care of patients with AIDS. We propose a new method to explore the population and individual characteristics by combining local linear regression (Fan and Gijbels 1996), penalized quasilikelihood (PQL, Breslow and Clayton 1993), and local quasilikelihood (Severini and Staniswalis 1994) techniques.

For longitudinal binary data analysis, generalized linear mixed-effects models (GLMeM) have been proposed to incorporate the between-subject and within-subject variations. See, for example, Breslow and Clayton (1993) and Schall (1991) for the PQL method, Zeger and Karim (1991) for the Gibbs sampler. Wang et al. (1998) considered generalized linear mixed models when one of the covariates is measured with error, and they proposed a simulation extrapolation (SIMEX, Stefanski and Cook 1995) estimation method.

To weaken model assumption for possible misspecification and to avoid the curse of dimensionality of fully nonparametric regression in the presence of several predictor variables, semiparametric models have been studied and used for longitudinal data analysis in literature (Severini and Staniswalis 1994). Lin and Carroll (2001) studied marginal semiparametric mixed-effects models and used the generalized estimating equation to estimate the parameter of interest. Wu and Zhang (2002) and Ke and Wang (2001) proposed semiparametric nonlinear mixed-effects models. Lin and Zhang (1999), Zhang (2004), and Zhang et al. (1998) used the smoothing spline to fit semiparametric mixed-effects models. These authors considered between-subject and within-subject variations in their estimation methods, but they mainly aimed at delineating population features, while individual characteristics are ignored.

The article is organized as follows. In Sect. 2 we formally introduce the model's framework, propose an estimation algorithm, and develop asymptotic results; extend to the case that X is measured with error; and discuss practical implementation of the methods. We illustrate the methods by conducting a small simulation experiment and extensively analyzing a data set from an AIDS study in Sect. 3 and provide a discussion in Sect. 4. The proof of the theoretical results is given in the Appendix.

2 Model and estimation

2.1 Model

Suppose that data are obtained from n independent subjects with outcome variables Y_{ij}, linear covariates X_{ij} $(p \times 1)$ and A_{ij} $(q \times 1)$, and nonlinear scalar Z_{ij}, where $i = 1, \ldots, n$ indicates the subjects; $j = 1, \ldots, n_i$ indicates the observation of the ith subject. Given the covariates X_{ij}, Z_{ij}, A_{ij}, and unobserved $q \times 1$ random effects vectors $\mathbf{b}_i$ and unobserved random effects function $c_i(\cdot)$, the observations Y_{ij} are assumed to be independent with means μ_{ij} and variance $\varphi\omega_{ij}V(\mu_{ij})$, where ω_{ij} is a known weight and $V(\cdot)$ is a known variance function. The GPLMeM of Y given X, Z is specified by

$$g(\mu_{ij}) = X_{ij}^{\mathrm{T}}\boldsymbol{\beta} + A_{ij}^{\mathrm{T}}\mathbf{b}_i + \theta(Z_{ij}) + c_i(Z_{ij}), \tag{1}$$

where $g(\cdot)$ is a known monotonic differentiable link function, β is a $p \times 1$ vector, and $\theta(\cdot)$ and $c_i(\cdot)$ are unknown smoothing functions. $c_i(\cdot)$ indicates the random-effects, which can be regarded as a realization of a stochastic process with mean zero. The random effects $\mathbf{b}_i$ are independent of $c_i(\cdot)$ and are independently distributed $N(0, \Sigma_b)$, where the covariance matrix Σ_b is determined by finite parameters. Given the covariates X_{ij}, A_{ij} and Z_{ij}, and $\mathbf{b}_i$ and $c_i(\cdot)$, the observations Y_{ij} and Y_{ik} are independent; i.e., $E(Y_{ij}|X_{ij}, Z_{ij}) = E\{Y_{ij}|X_{ij}, Z_{ij}, (X_{ik}, Z_{ik})_{k\neq j}\}$. In longitudinal data, Z_{ij} are often the treatment times, and X_{ij}, A_{ij}, and Y_{ij} are the observations of the ith subject at time Z_{ij}, for example, in the AIDS data set studied later.

2.2 Estimation

To estimate the parametric and nonparametric parts $\boldsymbol{\beta}$, $\mathbf{b}_i$, $\theta(\cdot)$, and $c_i(\cdot)$, we combine local linear regression, the PQL technique, and the quasilikelihood principle. We estimate these parameters by considering local PQL, then we update the estimate of $\boldsymbol{\beta}$ and $\mathbf{b}_i$, using all data points, by relying on the global PQL. In local PQL, we approximate $\theta(\cdot)$ and $c_i(\cdot)$ by linear functions:

$$\begin{aligned}\theta(z) &\approx \theta(z_0) + \theta'(z_0)(z - z_0) \equiv \alpha_0 + \alpha_1(z - z_0)\\ c_i(z) &\approx c_i(z_0) + c_i'(z_0)(z - z_0) \equiv \nu_{i0} + \nu_{i1}(z - z_0).\end{aligned}$$

Denote $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)^{\mathrm{T}}$, $\boldsymbol{\nu}_i = (\nu_{i0}, \nu_{i1})^{\mathrm{T}}$, $\Lambda_{ij} = \begin{pmatrix} 1 \\ Z_{ij} - z_0 \end{pmatrix}$. In a neighborhood of z_0, model (1) can be approximated by a GLMeM and described as

$$g(\mu_{ij}) = X_{ij}^{\mathrm{T}}\boldsymbol{\beta} + A_{ij}^{\mathrm{T}}\mathbf{b}_i + \boldsymbol{\alpha}^{\mathrm{T}}\Lambda_{ij} + \boldsymbol{\nu}_i^{\mathrm{T}}\Lambda_{ij},$$

where $\boldsymbol{\nu}_i \sim N(0, \Sigma_\nu)$ are random effects. In a conventional GLMeM setting, we can estimate $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\mathbf{b}_i$, and $\boldsymbol{\nu}_i$ by using the PQL principle; i.e., maximize

$$-\sum_{i=1}^{n}\left\{\sum_{j=1}^{n_i} Q(\widetilde{\mu}_{ij}, y_{ij}) + \boldsymbol{\nu}_i^{\mathrm{T}}\Sigma_\nu^{-1}\boldsymbol{\nu}_i + \mathbf{b}_i^{\mathrm{T}}\Sigma_b^{-1}\mathbf{b}_i\right\}$$

with respect to $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\mathbf{b}_i$, and $\boldsymbol{\nu}_i$, where $\widetilde{\mu}_{ij} = g^{-1}(X_{ij}^{\mathrm{T}}\boldsymbol{\beta} + A_{ij}^{\mathrm{T}}\mathbf{b}_i + \Lambda_{ij}^{\mathrm{T}}\boldsymbol{\alpha} + \Lambda_{ij}^{\mathrm{T}}\boldsymbol{\nu}_i)$ and $Q(s, y) = -2\int_y^s V^{-1}(t)(y - t)\mathrm{d}t$.

To consider the localization, we propose a locally PQL approach. The motivation of doing so is essentially the same as localization in conventional nonparametric regression.

Denote $\rho_k(t) = \{dg^{-1}(t)/dt\}^k/[\sigma^2 V\{g^{-1}(t)\}]$ and $q_k(t, y) = \partial^k Q\{g^{-1}(t), y\}/\partial t^k$ for $k = 1, 2, 3$. Let $\kappa_j = \int u^j K(u)\mathrm{d}u$, $\mu_j = \int u^j K^2(u)\mathrm{d}u$ for $j = 1, 2$, $N = \sum_{i=1}^n n_i$, and $M = \sum_{i=1}^n n_i^2$.

Step 1. For each fixed z_0, we consider the maximization

$$-\sum_{i=1}^{n}\left\{\sum_{j=1}^{n_i} K_h(Z_{ij} - z_0)Q(\widetilde{\mu}_{ij}, y_{ij}) + \boldsymbol{\nu}_i^{\mathrm{T}}\Sigma_\nu^{-1}\nu_i + \mathbf{b}_i^{\mathrm{T}}\Sigma_b^{-1}\mathbf{b}_i\right\} \tag{2}$$

with respect to $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\mathbf{b}_i$, and $\boldsymbol{\nu}_i$. $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a kernel function, and h is a bandwidth.

Take the differential of (2) on $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\mathbf{b}_i$, and $\boldsymbol{\nu}_i$. Step 1 is equivalent to solving the equations

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i} K_h(Z_{ij} - z_0)(Y_{ij} - \widetilde{\mu}_{ij})\rho_1(\widetilde{\mu}_{ij})\begin{pmatrix} 1 \\ Z_{ij} - z_0 \\ X_{ij} \end{pmatrix} = 0 \tag{3}$$

and, for each i,

$$\sum_{j=1}^{n_i} K_h(Z_{ij} - z_0)(Y_{ij} - \widetilde{\mu}_{ij})\rho_1(\widetilde{\mu}_{ij})\begin{pmatrix} A_{ij} \\ 1 \\ Z_{ij} - z_0 \end{pmatrix} = \begin{pmatrix} \Sigma_b^{-1} & 0 \\ 0 & \Sigma_\nu^{-1} \end{pmatrix}\begin{pmatrix} \mathbf{b}_i \\ \boldsymbol{\nu}_i \end{pmatrix}. \tag{4}$$

Denote $Y_{ij}^* = g(\widetilde{\mu}_{ij}) + (Y_{ij} - \widetilde{\mu}_{ij})g'(\widetilde{\mu}_{ij})$, $\widetilde{\mathbf{Y}}_i = (Y_{i1}^*, \ldots, Y_{in_i}^*)^{\mathrm{T}}$, $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{in_i})^{\mathrm{T}}$, and $\mathbf{A}_i$ similarly. Let $\mathbf{K}_{ih} = \mathrm{diag}\{K_h(Z_{i1} - z_0), \ldots, K_h(Z_{in_i} - z_0)\}$, $\mathbf{R}_i = \mathrm{diag}(R_{i1}, \ldots, R_{in_i})$ with $R_{ij} = \{g'(\widetilde{\mu}_{ij})\}^2 V(\widetilde{\mu}_{ij})$.

Equations (3) and (4) are approximately equivalent to iteratively fitting the following linear mixed-effects (Laird and Ware 1982) model:

$$\mathbf{K}_{ih}^{1/2}\widetilde{\mathbf{Y}}_i = \mathbf{K}_{ih}^{1/2}\{\mathbf{J}_i(\boldsymbol{\alpha}+\boldsymbol{\nu}_i)+\mathbf{X}_i\boldsymbol{\beta}+\boldsymbol{A}_i\mathbf{b}_i\}+\boldsymbol{\varepsilon}_i, \tag{5}$$

where $\boldsymbol{\varepsilon}_i \sim (0, \mathbf{R}_i)$ and $\mathbf{J}_i = (\Lambda_{i1}, \ldots, \Lambda_{in_i})^{\mathrm{T}}$.

Consequently, the estimates defined by step 1 are approximately given in closed-form expressions (Davidian and Giltinan 1995):

$$\begin{pmatrix}\widehat{\boldsymbol{\alpha}}\\ \widehat{\boldsymbol{\beta}}\end{pmatrix} = \left\{\sum_{i=1}^{n}(\mathbf{J}_i,\mathbf{X}_i)^{\mathrm{T}}\boldsymbol{\Omega}_i(\mathbf{J}_i,\mathbf{X}_i)\right\}^{-1}\left\{\sum_{i=1}^{n}(\mathbf{J}_i,\mathbf{X}_i)^{\mathrm{T}}\boldsymbol{\Omega}_i\widetilde{\mathbf{Y}}_i\right\}$$

$$\begin{pmatrix}\widehat{\mathbf{b}}_i\\ \widehat{\boldsymbol{\nu}}_i\end{pmatrix} = \begin{pmatrix}\Sigma_b & 0\\ 0 & \Sigma_\nu\end{pmatrix}(\mathbf{J}_i,\mathbf{A}_i)^{\mathrm{T}}\boldsymbol{\Omega}_i\left\{\widetilde{\mathbf{Y}}_i-(\mathbf{J}_i,\mathbf{A}_i)\begin{pmatrix}\widehat{\boldsymbol{\alpha}}\\ \widehat{\boldsymbol{\beta}}\end{pmatrix}\right\},$$

where $\boldsymbol{\Omega}_i = \mathbf{K}_{ih}^{1/2}\Sigma_i^{-1}\mathbf{K}_{ih}^{1/2}$ with $\Sigma_i = \mathbf{R}_i + \mathbf{K}_{ih}^{1/2}(\mathbf{J}_i\Sigma_\nu\mathbf{J}_i^{\mathrm{T}} + \mathbf{A}_i\Sigma_b\mathbf{A}_i^{\mathrm{T}})\mathbf{K}_{ih}^{1/2}$. $\theta(z_0)$ and $c_i(z_0)$ are estimated by $\widehat{\theta}(z_0) = \widehat{\alpha}_0$ and $\widehat{c}_i(z_0) = \widehat{\nu}_{i0}$.

Note that $\boldsymbol{\beta}$ and $\mathbf{b}_i$ are global parameters and their estimates given in step 1 are based only on local information. We update their estimate, using all data, by considering a global penalized quasilikelihood procedure:

Step 2. Update estimates of $\boldsymbol{\beta}$ and $\mathbf{b}_i$ by maximizing the object function

$$-\sum_{i=1}^{n}\left\{\sum_{j=1}^{n_i}Q(\widetilde{\mu}_{ij}^{*}, y_{ij}) + \mathbf{b}_i^{\mathrm{T}}\Sigma_b^{-1}\mathbf{b}_i\right\} \tag{6}$$

with respect to $\boldsymbol{\beta}$ and $\mathbf{b}_i$, where $\widetilde{\mu}_{ij}^{*} = g^{-1}\{X_{ij}^{\mathrm{T}}\boldsymbol{\beta} + A_{ij}^{\mathrm{T}}\mathbf{b}_i + \widehat{\alpha}_0(Z_{ij}) + \widehat{\nu}_{i0}(Z_{ij})\}$.

The expressions containing matrices Σ_b, Σ_ν, and $\mathbf{R}_i$, which are generally unknown and need to be estimated. To estimate these matrices, one may use the maximum and the restricted maximum likelihood methods to estimate the unknown components of Σ_b, Σ_ν, and $\mathbf{R}_i$ under the normality assumption. To implement these non-trivial methods, the EM-algorithm and Newton–Raphson methods have been proposed (Laird and Ware 1982; Davidian and Giltinan 1995). The standard statistical software packages such as SAS and Splus/R provide user-friendly functions to implement these methods(`lme` Splus/R function and the SAS procedure SAS MIXED). After obtaining the point estimates of the unknown components, we have $\widehat{\Sigma}_b$, $\widehat{\Sigma}_\nu$, and $\widehat{\mathbf{R}}_i$, and then $\widehat{\Sigma}_i$ and $\widehat{\Omega}_i$. The estimators of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\mathbf{b}_i$, and $\boldsymbol{\mu}_i$ thus can be obtained by substitution of $\widehat{\Sigma}_i$ and $\widehat{\Omega}_i$ in steps 1 and 2.

2.3 Asymptotic results

We claim the following condition, which is standard in the literature describing the generalized partially linear models.

Condition

(a) The function $q_2(t, y) < 0$ for $t \in R$ and y in the range of the response variable.
(b) The random-effects functions $c_i(z)$ are iid and have zero mean gaussian marginal distribution, and have the same distribution as a random-effects curve $c(z)$, a two times continuously differentiable function.
(c) The density function $f(z)$ of Z is positive and continuous at the point z_0.
(d) The functions $\theta(\cdot)$ and $\theta^{(2)}(\cdot)$ are continuous at the point z_0.
(e) With $R = X^{\mathrm{T}}\boldsymbol{\beta} + A^{\mathrm{T}}\mathbf{b} + \theta(Z) + c(Z)$, $E\{q_1^2(R, Y)|z\}$, $E\{q_1^2(R, Y)X|z\}$, and $E\{q_1^2(R, Y)XX^{\mathrm{T}}|z\}$ are twice differentiable in z.
(f) $E\{q_2^2(R, Y)\} < \infty$ and $E\{q_1^{2+\delta}(R, Y)\} < \infty$, for some $\delta > 2$.
(g) $Mh/N \to \lambda$, a finite nonnegative constant.
(h) Let $\gamma_y(z_1, z_2)$ be the covariance of $[Y - g^{-1}\{X^{\mathrm{T}}\boldsymbol{\beta} + A^{\mathrm{T}}\mathbf{b} + \theta(Z) + c(Z)\}]$ $\rho_1\{X^{\mathrm{T}}\boldsymbol{\beta} + A^{\mathrm{T}}\mathbf{b} + \theta(Z) + c(Z)\}$ for $Z = z_1$ and z_2. $\gamma_y(z_1, z_2)$ is continuous on z_1 and z_2.
(i) The functions $V''(\cdot)$ and $g'''(\cdot)$ are continuous.

Theorem 1 *Consider the estimator $\widehat{\theta}(z_0)$ given in step 1. Then, as $n \to \infty$, $h \to 0$ and $Nh \to \infty$, and under the specified condition, we have the asymptotic expansion*

$$\widehat{\theta}(z_0) - \theta(z_0) = b_\theta(z_0) + \frac{1}{Nf(z_0)E\{\rho_2(R)|Z = z_0\}} \sum_{i=1}^{n}\sum_{j=1}^{n_i} q_1(\eta_{ij}, y_{ij}) \times K_h(Z_{ij} - z_0) + o_P\left\{(Nh)^{-1/2} + h^2\right\}, \tag{7}$$

and hence

$$(Nh)^{1/2}\left\{\widehat{\theta}(z_0) - \theta(z_0) - b_\theta(z_0)\right\} \xrightarrow{D} N\{0, \sigma_\theta^2(z_0)\}, \tag{8}$$

where

$$b_\theta(z_0) = \frac{h^2}{2}\kappa_2\left[\theta''(z_0) + \frac{E\{c''(Z)\rho_2(R)|Z = z_0\}}{E\{\rho_2(R)|Z = z_0\}}\right]$$

and

$$\sigma_\theta^2(z_0) = \frac{\mu_0 E\{\rho_2(R)|Z = z_0\} + \lambda f(z_0)\gamma_y(z_0, z_0)}{f(z_0)[E\{\rho_2(R)|Z = z_0\}]^2}.$$

Theorem 2 *Let $\widehat{\boldsymbol{\beta}}$ be the estimate given in step 2. Under the condition, as $n \to \infty$, $Nh^4 \to 0$ and $Nh^2/log(1/h) \to \infty$,*

$$N^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N(\mathbf{0}, \mathbf{B}^{-1}\boldsymbol{\Sigma}_1\mathbf{B}^{-1}), \tag{9}$$

where $\mathbf{B} = E\left[\rho_2(R)XX^{\mathrm{T}}\right]$ and $\Sigma_1 = cov\left(q_1(R, Y)\left[X - \frac{E\{\rho_2(R)X|Z\}}{E\{\rho_2(R)|Z\}}\right]\right)$.

As in generalized partially linear models for cross-sectional data, Theorem 2 indicates that in order to estimate $\boldsymbol{\beta}$ at the root-n rate, one must undersmooth the nonparametric part $\theta(\cdot)$. This undersmoothing request is standard in the kernel literature, although ordinary bandwidth rates are permissible in partially linear models.

2.4 Implementation

To obtain the estimates by fitting (5), we give an initial value for $\widetilde{\mu}_{ij}$, say $\widetilde{\mu}_{ij}^{(0)}$, and then have the working response variable

$$\widetilde{Y}_{ij}^{(0)} = g(\widetilde{\mu}_{ij}^{(0)}) + (Y_{ij} - \widetilde{\mu}_{ij}^{(0)})g'(\widetilde{\mu}_{ij}^{(0)}).$$

For a given bandwidth h, we fit model (5) and obtain estimates $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\beta}}$, $\widehat{\mathbf{b}}_i$, and $\widehat{\boldsymbol{\nu}}_i$. Update $g(\widehat{\mu}_{ij}) = \widehat{\alpha}_0 + \widehat{\nu}_{i0} + X_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}} + A_{ij}^{\mathrm{T}}\widehat{\mathbf{b}}_i$ and $g'(\widehat{\mu}_{ij})$. Correspondingly, $\widetilde{Y}_{ij}$ can be updated as $g(\widehat{\mu}_{ij}) + (y_{ij} - \widehat{\mu}_{ij})g'(\widehat{\mu}_{ij})$, and fit (5) again. Repeat these procedures until convergence, and collect the final estimates $\widehat{\theta}(z_0) = \widehat{\alpha}_0$, $\widehat{c}_i(z_0) = \widehat{\nu}_{i0}$.

During the process of the implementation, one should keep in mind that the bandwidth needs to be selected in the estimation procedure involved in step 1. Theorem 2 indicates that undersmoothing the nonparametric part is necessary to guarantee to estimate $\boldsymbol{\beta}$ at the root-n rate. Because only the rates of convergence for the bandwidth h are necessary for the same limiting distribution for the estimators of $\boldsymbol{\beta}$, we select h by using the empirical bias bandwidth selection method (Ruppert 1997) in our data analysis. In our numerical analysis below, we select initial values by fitting the data by a generalized linear mixed-effects model. Because iteration is needed, we set the maximum number of iteration to be 30. Computation results for our numerical analysis are obtained within 18 iterations. This adoption leads to potentially burdensome computation, which may be attenuated with a powerful computation platform.

2.5 Measurement errors in covariates

In practice, the covariates may not be exactly observable, and like CD4 cell counts in our AIDS dataset analyzed later, are measured with substantial errors. The estimators and inference may be biased if one ignores these measurement errors. We need to adjust the resulting bias. In this special case, we denote X_{ij} by $X_i(Z_{ij})$, the true covariates of subject i, and W_{ij} by $W_i(Z_{ij})$, the observed values of X_{ij} at time Z_{ij}. Suppose we have the measurement error model,

$$W_i(z) = X_i(z) + u_i(z), \tag{10}$$

where $W_i(z)$ is the observed value and $X_i(z)$ is the underlying true value for the ith patient at treatment time z. The error $u_i(z)$ represents measurement error in the CD4+ cell counts. We assume that $u_i(z)$ is a mean zero process, and $\{X_i(z), u_i(z)\}$ are mutually independent. Fuller (1987) and Carroll et al. (1995) provided a good survey of measurement error models. Buonaccorsi et al. (2000) studied measurement errors in

linear mixed-effects models. Higgins et al. (1997) proposed a two-step approach to deal with measurement errors in nonlinear mixed-effects models. Wang et al. (1998) studied generalized linear mixed measurement error models. Liang and Ren (2005) investigated generalized partially linear models with errors-in-variables and proposed a measurement error calibration based on a SIMEX algorithm. The computation burden in a similar platform is tremendously expensive, although its implementation is straightforward. In this paper we adjust measurement error by using a mixed-effects regression spline model, as done by Higgins et al. (1997); a similar idea was used by Liang et al. (2003).

The data points repeatedly measured along time z are similar to replications if we can assume that the measured variable is a smooth function of z. Shi et al. (1996) and Rice and Wu (2001) have modeled the natural history of the CD4+ cell process in untreated HID-infected patients by using a mixed-effects regression spline model. Model the CD4+ cell counts as follows:

$$X_i(Z_{ij}) = \sum_{k=0}^{p} \xi_k \psi_k(Z_{ij}) + \sum_{k=0}^{q} \eta_{ik} \phi_k(Z_{ij}) = \boldsymbol{\Psi}_i(Z_{ij})\boldsymbol{\xi} + \boldsymbol{\Phi}_i(Z_{ij})\boldsymbol{\eta}_i, \quad (11)$$

where $\boldsymbol{\Psi}_i(z) = \{\psi_0(z), \psi_1(z), \ldots, \psi_p(z)\}$ and $\boldsymbol{\Phi}_i(z) = \{\phi_0(z), \phi_1(z), \ldots, \phi_q(z)\}$ are basis functions such as cubic B-spline basis, $\boldsymbol{\xi} = (\xi_0, \xi_1, \ldots, \xi_p)^{\mathrm{T}}$ is a fixed-effect parameter vector, and $\boldsymbol{\eta}_i = (\eta_{i0}, \eta_{i1}, \ldots, \eta_{iq})^{\mathrm{T}}$ is a random effect vector with mean zero and covariance matrix Σ_η (Σ_η may be unstructured or can be specified with a special structure). The selection of the number (p and q) and locations of knots for regression splines can be achieved by using cross-validation (Eubank 1999). Rice and Wu (2001) suggest setting $p = q$ and $\psi_k(z) = \phi_k(z)$. Model (11) is a standard linear mixed-effects model, which can be fitted by using the LME function of Splus (Pinheiro and Bates 2000).

Higgins et al. (1997) have proposed a two-step approach to deal with measurement errors in time-dependent covariates in nonlinear mixed-effects models. The first step is to estimate the covariate function by fitting an appropriate model for covariate processes, and then fit the nonlinear mixed-effects model by plugging in the estimates of the covariates in the second step. This is essentially similar to the regression calibration idea (Carroll et al. 1995). By applying a similar idea to our model for measurement error in CD4+ cell counts, we fit a LME model,

$$W_i(Z_{ij}) = \boldsymbol{\Psi}_i(Z_{ij})\boldsymbol{\xi} + \boldsymbol{\Phi}_i(Z_{ij})\boldsymbol{\eta}_i + \boldsymbol{u}_i, \quad (12)$$

and obtain an estimate $\widehat{W}_{ij}$ of W_{ij}, which is regarded as X_{ij}.

3 Numerical illustration

In this section, we first conduct a small simulation study for an illustration. We then use the proposed methods to analyze a data set from an AIDS study. In our numerical analysis, we calculate the naive estimates, i.e., those obtained by ignoring measurement

errors, the proposed estimates. We use the quartic kernel $k(u) = 15/16(1-u^2)^2 I_{(|u|\le 1)}$ for nonparametric regression.

3.1 A small simulation experiment

We performed a small simulation experiment and the data were generated from a logistic model, whose log(odd) can be expressed as

$$\begin{aligned} \text{logit}\{Y_{ij} = 1|X_{ij}, Z_{ij}\} &= X_{ij}(\beta + b_i) + \theta_i(Z_{ij}), \\ W_{ij} &= X_{ij} + U_{ij} \qquad i = 1, \ldots, n, \quad j = 1, \ldots, m, \end{aligned}$$

where X_{ij} is independent uniform $(0, 1)$ components, Z_{ij} is uniformly distributed $U(0, 1)$, and U_i is normally distributed $N(0, \sigma_u^2)$. The parameter β is equal to 1.85, and the nonparametric function is $\theta_i(z) = (1 + c_{1i})\cos(2\pi z) + (2 + c_{2i})\sin(2\pi z)$ with (c_{1i}, c_{2i}) being 2-dimensional normal distribution $N(0, \sigma_b^2 \mathbf{I}_2)$.

We consider a combination of $(n, m) = (20, 30), (40, 50), (60, 70), \sigma_u = 0.1, 0.25$, and $\sigma_b = 0.1$. In each of configurations, we run 500 replications. We calculate the proposed estimates and the naive estimates (i.e., we ignore the measurement errors). The estimated values of β are summarized in Table 1.

The results are in accord with the theory for measurement error models. The simulation results based on the proposed method indicate that the naive estimator of β is

Table 1 The estimates with standard errors (SE) of the parameter β for the simulated data. "Mean": the means of the estimated values; "SE": sample standard errors based on 500 replications

n	m	σ_u	Proposed		Naive	
			Mean	SE	Mean	SE
20	30	0.1	1.876	0.355	1.784	0.352
40			1.873	0.256	1.759	0.243
60			1.887	0.305	1.742	0.299
20	50		1.831	0.338	1.65	0.337
40			1.822	0.227	1.67	0.266
60			1.848	0.18	1.706	0.201
20	70		1.824	0.238	1.672	0.238
40			1.796	0.214	1.627	0.208
60			1.813	0.18	1.628	0.185
20	30	0.25	1.89	0.343	1.298	0.354
40			1.885	0.317	1.257	0.326
60			1.891	0.264	1.253	0.264
20	50		1.822	0.258	1.128	0.254
40			1.799	0.237	1.1	0.248
60			1.832	0.207	1.179	0.206
20	70		1.821	0.239	1.127	0.219
40			1.79	0.197	1.128	0.194
60			1.809	0.192	1.089	0.195

markedly biased, while the estimated values based on the proposed method is much closer to the true value than the naive estimate of β.

3.2 Data analysis

Let Y be binary viral load, X be CD4 cell counts and Z be the treatment time. An ordinary logistic mixed-effects model says that the logit of $Y = 1$ satisfies

$$\text{logit}\{E(Y_{ij}|X_{ij}, Z_{ij}, b_{1i}, b_{2i})\} = \alpha_0 + \alpha_1 X_{ij} + \alpha_2 Z_{ij} + b_{1i} X_{ij} + b_{2i} Z_{ij}, \quad (13)$$

where b_{1i} and b_{2i} reflect the individual variations. The merit of this model includes its computational convenience and easy interpretation of the model parameter. However, this model may not be able to capture some curvature, as we shall see later because drug resistance, and noncompliance probably contribute nonlinearly along the treatment time. An alternative is a partially logistic mixed-effects model, described as

$$\text{logit}\{E(Y_{ij}|X_{ij}, Z_{ij}, b_{1i}, c_i)\} = X_{ij}\beta + X_{ij} b_{1i} + \theta(Z_{ij}) + c_i(Z_{ij}), \quad (14)$$

where $\theta(z)$ and $c_i(z)$ are unknown smoothing functions, which describe the population characteristic and individual diversities.

In this section, we present the results of analysis of an AIDS clinical study conducted by the Pediatric AIDS Clinical Trials Group (PACTG 345, Scott et al. 2001). In this study, 33 patients were enrolled as cohort II. Specimens were obtained on days 0, 1, 3, 7, 14, 28, 56, ..., 1115, and 559 observations were obtained with 256 HID-1 RNA measurements below the detection limit of 400 copies/mL: 45% viral load observations were therefore suppressed below the detection limit. Figure 1 presents the individual observations of plasma HID RNA concentration (viral load) after initial antiretroviral treatments. A main objective of the treatment is to suppress the viral load below the limit of detection.

The purpose of this analysis focuses on whether the viral load is suppressed and how the of CD4 cell counts change during the treatment. We apply the model and estimation method described in Sect. 2 to explore this data set and address the concerns mentioned in Sect. 1 by modeling the dynamic relation between the binary response (with or without viral suppression) and CD4 cell counts during the treatment period of about 3 years.

Let $Y_{ij} = 1$ if the viral load of the ith subject at time Z_{ij} is below the detection limit, and 0 otherwise. In this analysis, we consider four scenarios: (i) model (13) and ignoring consideration of measurement error in CD4+ cell counts; (ii) model (13) with consideration of measurement error in CD4+ cell counts; (iii) model (14), ignoring measurement error in CD4+ cell counts; and (iv) model (14) with consideration of measurement error in CD4+ cell counts. When considering measurement errors in CD4+ cell counts and using the method proposed in Sect. 2.5 to produce the "true" CD4+ cell counts, the smoothing parameters (the number of knots) are selected by the model selection criterion BIC and the location of knots is selected at the quantiles of the data (Eubank 1999). We obtain $p = q = 3$. To stabilize the variance

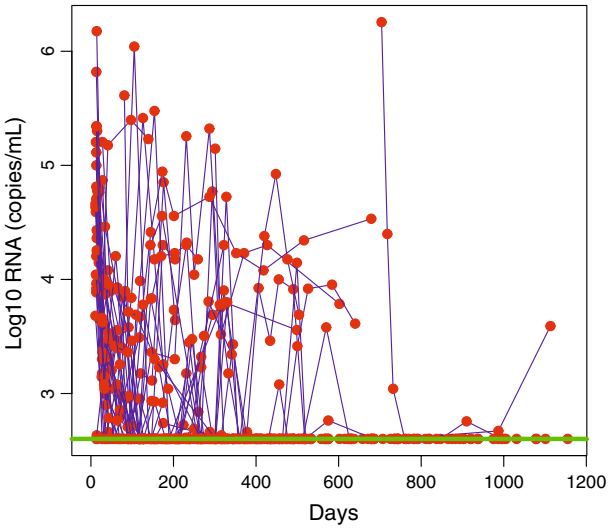

Fig. 1 Individual viral load measurements of plasma HID RNA concentration in the PACTG 345 study. The detection limit of 400 copies of HID RNA per mL of plasma is indicated by the *horizontal line*

Table 2 Estimates of the parameters in scenarios (i)–(iv), based on PACTG 345 data

	(i)	(ii)	(iii)	(iv)
$\widehat{\beta}$	-0.272	-0.358	-0.692	-1.046
SE	0.10	0.143	0.134	0.223
P value	0.007	0.0014	$< 10^{-4}$	$< 10^{-4}$

and computational algorithms, we centered the covariate CD4+ cell counts and took a log-transformation for time in the model fitting. We fitted the model by using the LME function of Splus.

The population estimated values of the parameter β for these four scenarios are presented in Table 2. Comparing the estimates of β under (i) scenarios to (iii) and (ii) to (iv), we saw that models (13) and (14) derive remarkably different results. We also compared the estimates obtained with or without considering measurement error in the CD4+ cell counts. The estimates attenuated toward zero regardless of the models (13) or (14) used when the measurement error in the covariate is not considered. This finding is similar to that in standard linear or nonlinear regression models with measurement error (Carroll et al. 1995).

The population estimates of $\theta(z)$ for scenarios (iii) and (iv) are shown in Fig. 2. Two population curves have similar patterns; that is, the curves rise rapidly at first and then maintain a steady decline from day 56 to the end of treatment. However, a consideration of measurement error reflects a sharper increase at the beginning.

Recalling the viral load of each subject shown in Fig. 1, we note that the between-subject variation can not be ignored and the individual curves, $\alpha_i(z) = \theta(z) + c_i(z)$, may not follow the pattern of the population curve. We present the results for four

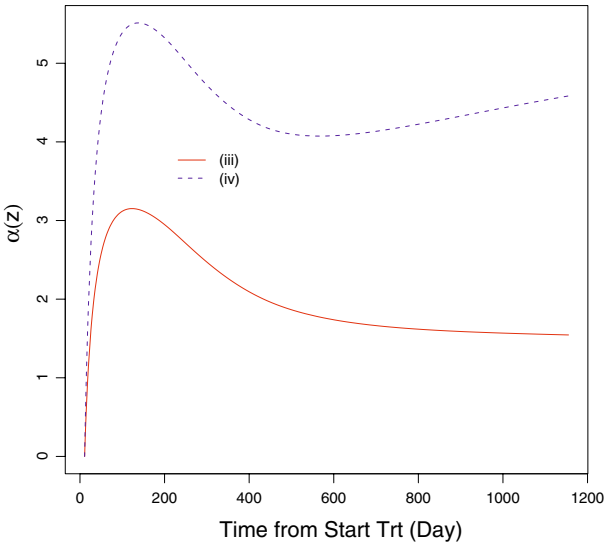

Fig. 2 Population curve estimates of $\theta(z)$ considering measurement errors (*solid line*) and ignoring measurement errors (*dashed line*)

individual subjects, and these results illustrate the principal advantage of the proposed model and developed methods in that the estimates can be obtained for both population and individuals. Because the patterns for the same individual under scenarios (iii) and (iv) are similar, we report in detail only the results for the case in which measurement errors in CD4+ cell counts are ignored. Figure 3 shows the individual estimates from four patients. For comparison, the corresponding population estimate is also plotted. The population and individual estimates are different not only in magnitude but also in patterns of change. The individual curve for subject 1 almost follows the population curve. The individual curves of the other three subjects are totally different from the population curve. Curves of subjects 6 and 24 decline sharply at first and then rebound later, to the end of treatment, but have different rebounding times. In contrast, the curve of subject 17 follows a convex shape. Given this large between-subject variation, the estimated trajectories of the individual curves become very important for individualizing treatment management and care for AIDS patients.

4 Discussion

In order to study the relation between binary virologic variables and immunologic variables, repeatedly measured indices of success of suppression of viral load (the virologic variable) and CD4+ cell counts (the immunologic variable) in AIDS clinical trials, we proposed a semiparametric mixed-effects model, which can parsimoniously reflect both population and individual relationships between the two longitudinal variables. Similar models have been reported in the literature. However, most published studies focused only on the estimation of the population characteristic, although

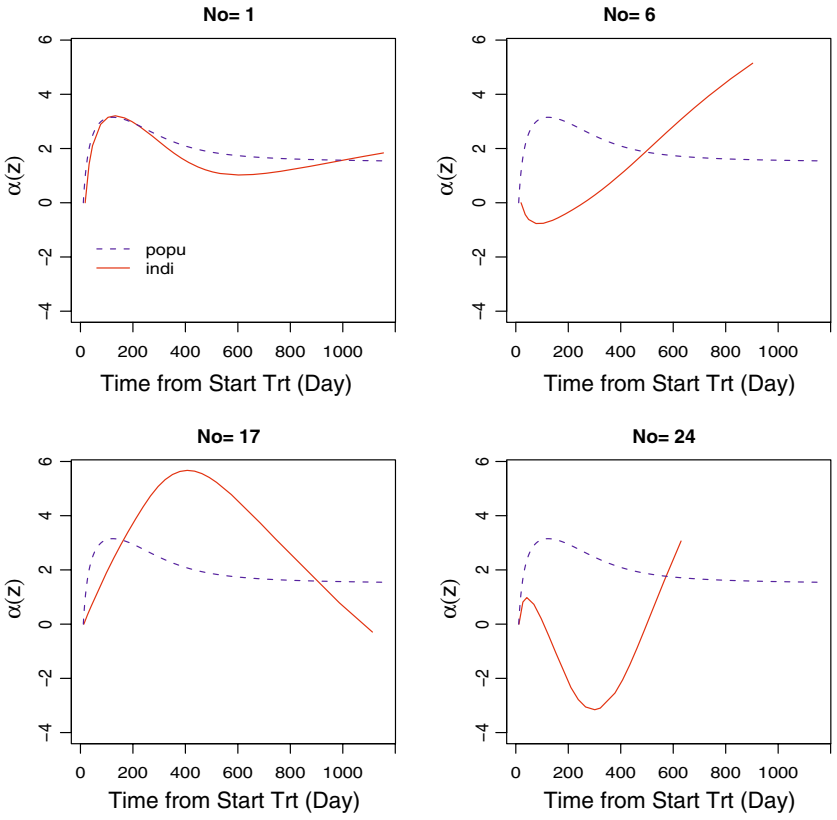

Fig. 3 Selected individual curve estimates (*solid line*), with the population curve estimate (*dashed line*) provided for comparison

between-subject and within-subject variations have been incorporated. The method proposed in this article extends existing methods and allows us to estimate the population profile and individual diversities.

In step 1 we minimized the objective function with respect to $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\mathbf{b}_i$ and $\boldsymbol{\nu}_i$. Alternatively, we may specify initial values $\boldsymbol{\beta}^{(0)}$ and $\mathbf{b}_i^{(0)}$, and maximize the objective function (a similar form to (2) but replacing $\boldsymbol{\beta}$ and $\mathbf{b}_i$ with $\boldsymbol{\beta}^{(0)}$ and $\mathbf{b}_i^{(0)}$) with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\nu}_i$. This alternative may increase efficiency (see Carroll et al. 1997 for a related discussion), but it increases computation efforts.

When considering measurement errors in CD4 cell counts, we used a mixed-effects based approach to calibrate the effect of measurement error. This calibration may also be achieved by using a SIMEX-based procedure such as Liang and Ren (2005). However, additional information is required to estimate the covariance matrix $\text{cov}(X)$ in implementing the SIMEX procedure. This investigation also increases computation efforts.

In this article, we develop a method using local linear regression. There are a lot of alternative ways for local linear approximation in step 1, including higher degree

local polynomial kernel methods and smoothing and regression splines. The details for these methods require a further investigation in our setting. We chose the local linear smoother because theoretical results can be derived and the estimators of nonparametric components do not suffer from boundary effects (Fan and Gijbels 1996).

Model (1) may be extended to a generalized additive partially linear mixed-effect model, of the form

$$g(\mu_{ij}) = X_{ij}^{\mathrm{T}}\boldsymbol{\beta} + A_{ij}^{\mathrm{T}}\mathbf{b}_i + \sum_{k=1}^{K}\theta_k(Z_{ij}^{(k)}) + \sum_{k=1}^{K}c_{ik}(Z_{ij}^{(k)}).$$

The study of this model is interesting and requires tedious efforts, but is beyond the scope of this paper.

Acknowledgment The author appreciates the Editor, Associate Editor, and two referees for their valuable comments and suggestions. This research was partially supported by NIH/NIAID grants AI62247 and AI59773.

Appendix

A.1 Proof of Theorem 1

Let $c_n = (Nh)^{-1/2}$,

$$\boldsymbol{\Xi}_{ij}^{*} = \begin{pmatrix} 1 \\ (Z_{ij} - z_0)/h \\ X_{ij} \end{pmatrix}, \quad \boldsymbol{\beta}^{*} = \begin{pmatrix} c_n^{-1}\{\alpha_0 - \theta(z_0)\} \\ c_n^{-1}h\{\alpha_1 - \theta'(z_0)\} \\ c_n^{-1}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \end{pmatrix},$$

$$\widehat{\boldsymbol{\beta}}^{*} = \begin{pmatrix} c_n^{-1}\{\widehat{\alpha}_0 - \theta(z_0)\} \\ c_n^{-1}h\{\widehat{\alpha}_1 - \theta'(z_0)\} \\ c_n^{-1}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \end{pmatrix},$$

and let $f(z)$ denote the density function of Z_{ij}. Denote further $\bar{\eta}_{ij} = \bar{\eta}_{ij}(z_0) = X_{ij}^{\mathrm{T}}\boldsymbol{\beta} + A_{ij}^{\mathrm{T}}\widehat{\mathbf{b}}_i + \theta(z_0) + \theta'(z_0)(Z_{ij} - z_0) + \widehat{c}_i(z_0) + \widehat{c}_i{}'(z_0)(Z_{ij} - z_0)$. Equation (3) implies that $\widehat{\boldsymbol{\beta}}^{*}$ maximizes

$$\ell_n(\boldsymbol{\beta}^{*}) = h\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left[Q\left\{g^{-1}(c_n\boldsymbol{\beta}^{*\mathrm{T}}\boldsymbol{\Xi}_{ij}^{*} + \bar{\eta}_{ij}), Y_{ij}\right\} - Q\left\{g^{-1}(\bar{\eta}_{ij}), Y_{ij}\right\}\right]K_h(Z_{ij} - z_0),$$

with respect to $\boldsymbol{\beta}^{*}$. By the concavity of the function $\ell_n(\boldsymbol{\beta}^{*})$ and Taylor expansion of the function $Q\{g^{-1}(\cdot), Y_{ij}\}$ with respect to $\boldsymbol{\beta}^{*}$, we obtain that

$$\ell_n(\boldsymbol{\beta}^{*}) = \mathbf{W}_n^{\mathrm{T}}\boldsymbol{\beta}^{*} + \tfrac{1}{2}\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{A}_n\boldsymbol{\beta}^{*}\{1 + o_P(1)\}, \tag{15}$$

where

$$\mathbf{W}_n = hc_n \sum_{i=1}^{n}\sum_{j=1}^{n_i} q_1(\bar{\eta}_{ij}, Y_{ij})\boldsymbol{\Xi}_{ij}^* K_h(Z_{ij} - z_0) \quad \text{and}$$
$$\mathbf{A}_n = hc_n^2 \sum_{i=1}^{n}\sum_{j=1}^{n_i} q_2(\bar{\eta}_{ij}, Y_{ij})\boldsymbol{\Xi}_{ij}^* \boldsymbol{\Xi}_{ij}^{*\mathrm{T}} K_h(Z_{ij} - z_0).$$

Furthermore, let $\eta_{ij}^0 = X_{ij}^{\mathrm{T}}\boldsymbol{\beta} + A_{ij}^{\mathrm{T}}\mathbf{b}_i + \theta(z_0) + \theta'(z_0)(Z_{ij} - z_0) + c_i(z_0) + c_i'(z_0)(Z_{ij} - z_0)$. Then

$$\begin{aligned}\mathbf{A}_n = {} & hc_n^2 \sum_{i=1}^{n}\sum_{j=1}^{n_i} q_2(\eta_{ij}^0, Y_{ij})\boldsymbol{\Xi}_{ij}^* \boldsymbol{\Xi}_{ij}^{*\mathrm{T}} K_h(Z_{ij} - z_0) \\ & + hc_n^2 \sum_{i=1}^{n}\sum_{j=1}^{n_i} q_3(\zeta_{ij}, Y_{ij})\boldsymbol{\Xi}_{ij}^* \boldsymbol{\Xi}_{ij}^{*\mathrm{T}} (\bar{\eta}_{ij} - \eta_{ij}^0) K_h(Z_{ij} - z_0), \end{aligned} \tag{16}$$

where ζ_{ij} is between η_{ij}^0 and $\bar{\eta}_{ij}$. Note that $(\bar{\eta}_{ij} - \eta_{ij}^0) = A_{ij}^{\mathrm{T}}(\widehat{\mathbf{b}}_i - \mathbf{b}_i) + \{\widehat{c}_i(z_0) - c_i(z_0)\} + \{\widehat{c}_i'(z_0) - c_i'(z_0)\}(Z_{ij} - z_0)$. Using an argument similar to the proof of (A.9) of Carroll et al. (1997) and condition (i) yield that the second term in the (16) is of order $o(h)$. We therefore have

$$\mathbf{A}_n = hc_n^2 \sum_{i=1}^{n}\sum_{j=1}^{n_i} q_2(\eta_{ij}^0, Y_{ij})\boldsymbol{\Xi}_{ij}^* \boldsymbol{\Xi}_{ij}^{*\mathrm{T}} K_h(Z_{ij} - z_0) + o(h) \triangleq \mathbf{A}_n^0 + o(1).$$

In a similar way, we obtain that

$$\mathbf{W}_n = hc_n \sum_{i=1}^{n}\sum_{j=1}^{n_i} q_1(\eta_{ij}^0, Y_{ij})\boldsymbol{\Xi}_{ij}^* K_h(Z_{ij} - z_0) + o(c_n^{-1}h^2) \triangleq \mathbf{W}_n^0 + o(c_n^{-1}h^2).$$

Let

$$\mathbf{A}(X) = \begin{pmatrix} 1 & 0 & X^{\mathrm{T}} \\ 0 & \kappa_2 & 0 \\ X & 0 & XX^{\mathrm{T}} \end{pmatrix}; \quad \mathbf{B}(X) = \begin{pmatrix} \mu_0 & 0 & \mu_0 X^{\mathrm{T}} \\ 0 & \mu_2 & 0 \\ \mu_0 X & 0 & \mu_0 XX^{\mathrm{T}} \end{pmatrix}.$$

We write $(\mathbf{A}_n^0)_{ij} = (E\mathbf{A}_n^0)_{ij} + O_P\left(\{\mathrm{var}(\mathbf{A}_n^0)_{ij}\}^{1/2}\right)$ and deal with the first two moments of $\mathbf{A}_n^0$. Note that $q_2(x, y) = \{y - g^{-1}(x)\}\rho_1'(x) - \rho_2(x)$. We find for the first moment,

with $\bar{\eta} = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta} + A^{\mathrm{T}}\mathbf{b} + \theta(z_0) + \theta(z_0)(Z - z_0) + c(z_0) + c'(z_0)(Z - z_0)$,

$$\begin{aligned} E\mathbf{A}_n^0 &= E\left[E\left\{q_2(\bar{\eta}, Y)\boldsymbol{\Xi}^*\boldsymbol{\Xi}^{*\mathrm{T}}K_h(Z - z_0)|\mathbf{X}, \mathbf{b}\right\}\right] \\ &= E\left[\{g^{-1}(R) - g^{-1}(\bar{\eta})\}\rho_1'(\bar{\eta})\boldsymbol{\Xi}^*\boldsymbol{\Xi}^{*\mathrm{T}}K_h(Z - z_0)\right] \\ &\quad -E\left\{\rho_2(\bar{\eta})\boldsymbol{\Xi}^*\boldsymbol{\Xi}^{*\mathrm{T}}K_h(Z - z_0)\right\}. \end{aligned} \tag{17}$$

The first term in (17) is of order $O(h^2)$. It follows that

$$\begin{aligned} E\mathbf{A}_n^0 &= -E\left\{\rho_2(\bar{\eta})\boldsymbol{\Xi}^*\boldsymbol{\Xi}^{*\mathrm{T}}K_h(Z - z_0)\right\} + O(h^2) \\ &= -E\left(E\left[\rho_2\{R + O(h)\}\boldsymbol{\Xi}^*\boldsymbol{\Xi}^{*\mathrm{T}}|Z = z_0\right]K_h(Z - z_0)\right) + O(h^2) \\ &= -f(u)E\left\{\rho_2(R)\mathbf{A}(X)|Z = z_0\right\} + o(1) \equiv \mathbf{A} + o(1). \end{aligned} \tag{18}$$

In an analogous way we can show that

$$\mathrm{var}(\mathbf{A}_n^0)_{ij} = O(c_n^2). \tag{19}$$

Combining (18) and (19), we obtain

$$\mathbf{A}_n^0 = -\mathbf{A} + o_P(1). \tag{20}$$

Therefore, by (15),

$$\ell_n(\boldsymbol{\beta}^*) = \mathbf{W}_n^{0\mathrm{T}}\boldsymbol{\beta}^* - \frac{1}{2}\boldsymbol{\beta}^{*\mathrm{T}}\mathbf{A}\boldsymbol{\beta}^* + o_P(1). \tag{21}$$

By applying the convexity lemma (Pollard 1991), we obtain that $\widehat{\boldsymbol{\beta}}^* = \mathbf{A}^{-1}\mathbf{W}_n^0 + o_P(1)$. Hence the asymptotic normality of $\widehat{\boldsymbol{\beta}}^*$ will follow from that of $\mathbf{W}_n^0$, which we will establish below. By the definition of $\mathbf{W}_n^0$ and using Taylor's expansion we get for the first moment,

$$\begin{aligned} E\mathbf{W}_n^0 &= c_n^{-1}E\left[E\left\{q_1(\bar{\eta}, Y)\boldsymbol{\Xi}^*K_h(Z - z_0)|\mathbf{X}, \mathbf{b}\right\}\right] \\ &= c_n^{-1}E\left[\left\{g^{-1}(R) - g^{-1}(\bar{\eta})\right\}\rho_1(\bar{\eta})\boldsymbol{\Xi}^*K_h(Z - z)\right] \\ &= \frac{1}{2}c_n^{-1}E\left[(g^{-1})'(R)\{c''(z_0) + \theta''(z_0)\}(Z - z_0)^2\rho_1(\bar{\eta})\boldsymbol{\Xi}^*K_h(Z - z_0)\right] \\ &\quad + o(c_n^{-1}h^2) \\ &= c_n^{-1}\frac{1}{2}\theta''(z_0)h^2 f(z_0)E\left\{\rho_2(R)(\kappa_2, 0, \kappa_2 X^{\mathrm{T}})^{\mathrm{T}}|Z = z_0\right\} \\ &\quad + c_n^{-1}\frac{1}{2}h^2 f(z_0)E\left\{c''(Z)\rho_2(R)(\kappa_2, 0, \kappa_2 X^{\mathrm{T}})^{\mathrm{T}}|Z = z_0\right\} + o(c_n^{-1}h^2). \end{aligned} \tag{22}$$

It follows from these statements and (20) that

$$\begin{aligned} & c_n^{-1}\{\widehat{\theta}(z_0) - \theta(z_0) - b_\theta(z_0) + o_p(h^2)\} \\ & \quad = [f(z_0)E\{\rho_2(R)|Z = z_0\}]^{-1} S_n\{1 + o_p(1)\}, \end{aligned}$$

where $S_n = c_n \sum_{i=1}^{n} \sum_{j=1}^{n_i} s_{ij}$ with $s_{ij} = q_1(\eta_{ij}, y_{ij}) K_h(Z_{ij} - z_0)$. We have

$$\mathrm{var}(S_n) = c_n^2 \sum_{i=1}^{n} \mathrm{var}\left(\sum_{j=1}^{n_i} s_{ij} \right) = J_{n1} + J_{n2},$$

where

$$J_{n1} = c_n^2 \sum_{i=1}^{n} \sum_{j=1}^{n_i} \mathrm{var}(s_{ij}) \quad \text{and} \quad J_{n2} = c_n^2 \sum_{i=1}^{n} \sum_{j_1 \neq j_2} \mathrm{cov}(s_{ij_1}, s_{ij_2}).$$

It is easy to see that

$$\mathrm{var}(s_{ij}) = h\mu_0 f(z_0) E\{\rho_2(R)|Z = z_0\} + o(1)$$

and

$$J_{n1} = \mu_0 f(z_0) E\{\rho_2(R)|Z = z_0\} + o(1).$$

Furthermore, for $j_1 \neq j_2$, by condition (h)

$$\begin{aligned} \mathrm{cov}(s_{ij_1}, s_{ij_2}) &= E\{\gamma_y(Z_{ij_1}, Z_{ij_2}) K_h(Z_{ij_1} - z_0) K_h(Z_{ij_2} - z_0)\} \\ &= h^2 f^2(z_0) \gamma_y(z_0, z_0)\{1 + o_p(1)\}, \end{aligned}$$

which, combined with condition (g), yields that

$$\begin{aligned} J_{n2} &= h^2 f^2(z_0) \gamma_y(z_0, z_0) c_n^2 \sum_{i=1}^{n} n_i(n_i - 1)\{1 + o(1)\} \\ &= \lambda f^2(z_0) \gamma_y(z_0, z_0)\{1 + o(1)\}. \end{aligned}$$

It follows that

$$\mathrm{var}(S_n) = \mu_0 f(z_0) E\{\rho_2(R)|Z = z_0\} + \lambda f^2(z_0) \gamma_y(z_0, z_0) + o(1).$$

Finally, using condition (f), it can be shown that Liapounov's condition is satisfied and hence Theorem 1 holds, as claimed.

A.2 Proof of Theorem 2

First of all, we note that under the specified condition, in a proof similar to that for Theorem 2 of Carroll et al. (1997), we can establish

$$\sup_{z\in D}\left|\widehat{\theta}(z_0)-\theta(z_0)-\frac{1}{Nf(z)E\{\rho_2(R)|Z=z_0\}}\sum_{i=1}^{n}\sum_{j=1}^{n_i}q_1(\eta_{ij},y_{ij})K_h(Z_{ij}-z_0)\right| = O_P\left\{h^2c_n+c_n^2\log^{1/2}(1/h)\right\}. \tag{23}$$

Let $\widehat{\zeta}=N^{1/2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$, $\widehat{m}_{ij}=\widehat{\theta}(Z_{ij})+X_{ij}^{\mathrm{T}}\boldsymbol{\beta}+\widehat{c}_i(Z_{ij})+A_{ij}^{\mathrm{T}}\mathbf{b}_i$, and $m_{ij}=\theta(Z_{ij})+X_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0+c_i(Z_{ij})+A_{ij}^{\mathrm{T}}\mathbf{b}_i$. Then, $\widehat{\zeta}$ maximizes

$$\ell_n(\zeta)=\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left[Q\left\{g^{-1}(\widehat{m}_{ij}+N^{-1/2}X_{ij}^{\mathrm{T}}\zeta),Y_{ij}\right\}-Q\left\{g^{-1}(\widehat{m}_{ij}),Y_{ij}\right\}\right] \tag{24}$$

with respect to ζ. By Taylor's expansion, we have

$$\ell_n(\zeta)=N^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{n_i}q_1(\widehat{m}_{ij},Y_{ij})X_{ij}^{\mathrm{T}}\zeta+\frac{1}{2}\zeta^{\mathrm{T}}\mathbf{B}_n\zeta, \tag{25}$$

where

$$\mathbf{B}_n=\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left[Y_{ij}\rho_1'\left\{g^{-1}(\widehat{m}_{ij}+\zeta_{nij})\right\}-\rho_3\left\{g^{-1}(\widehat{m}_{ij}+\zeta_{nij}')\right\}\right]X_{ij}X_{ij}^{\mathrm{T}},$$

with ζ_{nij} and ζ_{nij}' between 0 and $N^{-1/2}X_{ij}^{\mathrm{T}}\zeta$, independent of Y_{ij}, and with $\rho_3(x)=-g^{-1}(x)\rho_1'(x)-\rho_2(x)$. It can be shown that

$$\mathbf{B}_n=-E\{\rho_2(R)XX^{\mathrm{T}}\}+o_P(1)=-\mathbf{B}+o_P(1). \tag{26}$$

Using similar arguments as for obtaining (26), we get

$$\begin{aligned}&N^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{n_i}q_1(\widehat{m}_{ij},Y_{ij})X_{ij}\\&=N^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{n_i}q_1(m_{ij},Y_{ij})X_{ij}+N^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{n_i}q_2(m_{ij},Y_{ij})\\&\quad\times\left\{\widehat{\theta}(Z_{ij})-\theta(Z_{ij})\right\}X_{ij}+O_P\left\{N^{1/2}c_n^2\log^{1/2}(1/h)\right\}.\end{aligned}$$

By (23), the second term in the above expression can be expressed as

$$
\begin{aligned}
N^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{n_i} & q_2(m_{ij}, Y_{ij}) \frac{1}{Nhf(Z_{ij}) E\{\rho_2(R)|Z_{ij}\}} \\
& \times \left\{ \sum_{l=1}^{n} \sum_{s=1}^{n_i} q_1(\eta_{ls}, y_{ls}) K_h(Z_{ls} - Z_{ij}) \right\} X_{ij} \\
& + O_P \left\{ n^{1/2} c_n^2 \log^{1/2}(1/h) \right\} \\
& \equiv T_{n1} + O_P \left\{ N^{1/2} c_n^2 \log^{1/2}(1/h) \right\}.
\end{aligned}
$$

T_{n1} can further be simplified as

$$
T_{n1} = N^{-1/2} \sum_{l=1}^{n} \sum_{s=1}^{n_i} q_1(\eta_{ls}, y_{ls}) \frac{E\{\rho_2(R) X | Z_{ls}\}}{E\{\rho_2(R)|Z_{ls}\}} + O_P(N^{1/2} h^2). \tag{27}
$$

Combining (24)–(27) we obtain that

$$
\ell_n(\zeta) = N^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \omega(Z_{ij}, Y_{ij}, X_{ij}) - \frac{1}{2} \zeta^{\mathrm{T}} \mathbf{B} \zeta + o_P(1),
$$

where

$$
\Omega(Z_{ij}, Y_{ij}, X_{ij}) = q_1(m_{ij}, Y_{ij}) \left[X_{ij} - \frac{E\{\rho_2(R) X | Z_{ij}\}}{E\{\rho_2(R)|Z_{ij}\}} \right].
$$

By the convexity lemma (Pollard 1991) we find that $\widehat{\zeta} = \mathbf{B}^{-1} N^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \Omega(Z_{ij}, Y_{ij}, X_{ij}) + o_P(1)$, from which it follows that

$$
N^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N(0, \mathbf{B}^{-1} \Sigma_1 \mathbf{B}^{-1}),
$$

as claimed.

References

Breslow, N. E., Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association, 88*, 9–25.

Buonaccorsi, J. P., Demidenko, E., Tosteson, T. D. (2000). Estimation in longitudinal random effects models with measurement error. *Statistica Sinica, 10*, 885–904.

Carroll, R. J., Fan, J., Gijbels, I., Wand, M. P. (1997). Generalized partially single-index models. *Journal of the American Statistical Association, 92*, 477–489.

Carroll, R. J., Ruppert, D., Stefanski, L. A. (1995). *Nonlinear measurement error models*. New York: Chapman and Hall.

Davidian, M., Giltinan, D. (1995). *Nonlinear models for repeated measurement data.* New York: Chapman and Hall.

Eubank, R. L. (1999). *Nonparametric regression and spline smoothing*. New York: Marcel Dekker.
Fan, J., Gijbels, I. (1996). *Local polynomial modeling and its applications*. London: Chapman and Hall.
Fuller, W. A. (1987). *Measurement error models*. New York: Wiley.
Higgins, K. M., Davidian, M., Giltinan, D. M. (1997). A two-step approach to measurement error in time-dependent covariates in nonlinear mixed-effects models, with application to IGF-I pharmacokinetics. *Journal of the American Statistical Association, 92*, 436–448.
Ke, C. L., Wang, Y. D. (2001). Semiparametric nonlinear mixed-effects models and their applications (with discussions). *Journal of the American Statistical Association, 96*, 1272–1298.
Laird, N. M., Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics, 38*, 963–974.
Liang, H., Ren, H. B. (2005). Generalized partially linear measurement error models. *Journal of Computational and Graphical Statistics, 14*, 237–250.
Liang, H., Wu, H. L., Carroll, R. J. (2003). The relationship between virologic and immunologic responses in AIDS clinical research using mixed-effect varying-coefficient semiparametric models with measurement error. *Biostatistics, 4*, 297–312.
Lin, X. H., Carroll, R. J. (2001). Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association, 96*, 1045–1056.
Lin, X. H., Zhang, D. W. (1999). Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society, Series B, 61*, 381–400.
Pinheiro, J. C., Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.
Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory, 7*, 186–199.
Ruppert, D. (1997). Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation. *Journal of the American Statistical Association, 92*, 1049–1062.
Rice, J. A., Wu, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curve. *Biometrics, 57*, 253–259.
Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika, 78*, 717–727.
Scott, Z. A., Chadwick, E. G., Gibson, L. L. et al. (2001). Infrequent detection of HIV-1-specific, but not cytomegalovirus-specific, CD8+T cell responses in young HIV-1-infected infants. *Journal of Immunology, 167*, 7134–7140.
Severini, T. A., Staniswalis, J. G. (1994). Quasilikelihood estimation in semiparametric models. *Journal of the American Statistical Association, 89*, 501–511.
Shi, M., Weiss R. E., Taylor, J. M. G. (1996). An analysis of pediatric CD4+ counts for acquired immune deficiency syndrome using flexible random curves. *Applied Statistics, 45*, 151–163.
Stefanski, L. A., Cook, J. R. (1995). Simulation-extrapolation: the measurement error jackknife. *Journal of the American Statistical Association, 90*, 1247–1256.
Wang, N. S., Lin, X. H., Gutierrez, R. G., Carroll, R. J. (1998). Bias analysis and SIMEX approach in generalized linear mixed measurement error models. *Journal of the American Statistical Association, 93*, 249–261.
Wu, H. L., Kuritzkes, D. R., Clair, M. S., Spear, G., Connick, E., Landay, A., Lederman, M. M. (1999). Characterizing individual and population viral dynamics in HIV-1-infected patients with potent antiretroviral therapy: correlations with host-specific factors and virological endpoints. *Journal of Infectious Disease, 179*, 799–897.
Wu, H. L., Zhang, J. T. (2002). Semiparametric nonlinear mixed-effects models for longitudinal data application to an AIDS clinical study. *Statistics in Medicine, 21*, 3655–3675.
Wu, L. (2002). A joint model for nonlinear mixed-effects models with censoring and covariates measured with error, with application to AIDS studies. *Journal of the American Statistical Association, 97*, 955–964.
Zeger, S. L., Karim, M. R. (1991). Generalized linear models with random effects: a gibbs sampling approach. *Journal of the American Statistical Association, 6*, 79–86.
Zhang, D. (2004). Generalized linear mixed models with varying coefficients for longitudinal data. *Biometrics, 60*, 8–15.
Zhang, D., Lin, X., Raz, J., Sowers, M. (1998). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association, 93*, 710–719.