Semiparametric Mean-Covariance Regression Analysis for Longitudinal Data

Chenlei Leng, Weiping Zhang and Jianxin Pan *

August 30, 2008

Abstract

Efficient estimation of the regression coefficients in longitudinal data analysis requires a correct specification of the covariance structure. Existing approaches usually focus on modeling the mean with specification of certain covariance structures, which may lead to inefficient or biased estimators of parameters in the mean if misspecification occurs. In this paper, we propose a data-driven approach based on semiparametric regression models for the mean and the covariance simultaneously, motivated by the modified Cholesky decomposition. A regression spline based approach using generalized estimating

*Leng is with Department of Statistics and Applied Probability, National University of Singapore. Leng's research is supported in part by NUS faculty research grant R155-000-080-112 and a grant from NUS Risk Management Institute. (Email: stalc@nus.edu.sg). Zhang is with Department of Statistics and Finance, University of Science and Technology of China. Zhang's research is supported by NSF of China (No. 10671189, 10771204) (Email: zwp@ustc.edu.cn). Pan is with School of Mathematics, The University of Manchester, UK. Pan's research is supported by a grant from MRC (No. 80209) (Email: jianxin.pan@manchester.ac.uk).

equations is developed to estimate the parameters in the mean and the covariance. The resulting estimators for the regression coefficients in both the mean and the covariance are shown to be consistent and asymptotically normally distributed. In addition, the nonparametric functions in these two structures are estimated at their optimal rate of convergence. Simulation studies and a real data analysis show that the proposed approach yields highly efficient estimators for the parameters in the mean, and provides parsimonious estimation for the covariance structure.

Some Keywords: Covariance misspecification; Efficiency; Generalized estimating equation; Longitudinal data; Modified Cholesky decomposition; Semiparametric models.

1 Background

Longitudinal data arise frequently in the biomedical, epidemiological, social and economical fields. A salient feature of longitudinal studies is that subjects are measured repeatedly over time. Thus, observations for the same subject are intrinsically correlated, though observations from different subjects may be independent. Regression methods for such data sets accounting for within-subject correlation is abundant in the literature. See, for example, book-length expositions by Diggle et al. (2002) and Wu and Zhang (2006), and references therein. Within the framework of generalized linear models (GLM), the technique of generalized estimating equations (GEE, Liang and Zeger, 1986) is widely used for dealing with longitudinal data. GEE makes the use of a working correlation model to estimate the mean parameters in the marginal specification of the regression. Although consistency of the mean parameter estimators is not affected, misspecification of the correlation may result in a great loss of efficiency (Wang and Carey, 2003). On the other hand, the correlation structure itself may be of scientific interest (Diggle and Verbyla, 1998). Therefore, there is a great need to model the covariance structure. However, modeling the correlation matrix is more challenging than modeling the mean as there are usually much more parameters in the former and the positive definiteness of the covariance matrix has to be assured. This calls for effective methods for modeling the covariance matrices. Prentice and Zhao (1991) proposed a moment based approach to parametrize the correlation matrix. More recently, in a series of important papers, Pourahmadi (1999, 2000) proposed a modified Cholesky decomposition to decompose the covariance matrix. This decomposition is attractive as it leads automatically to positive definite covariance matrices. In addition, the decomposition is appealing as the parameters in it are related to well founded statistical concepts, as can be seen later. Thereafter, the parameters in this decomposition can be modeled via regression techniques, enabling model based inference for the parameters in the mean and the covariances. See also Pan and MacKenzie (2003) for a related discussion. More recently, Ye and Pan (2006) further proposed to use GEE to model the parameters in this decomposition. By formulating several sets of parametric estimating equations, they showed that the approach yields efficient estimators for both the mean and the covariance parameters.

There is a clear need to relax the parametric assumption posed in Ye and Pan (2006), as model misspecification may result in biased estimation, a problem even more severe than misspecification of the covariance. Fully nonparametric models are desirable for low dimensional covariates, but may suffer from the curse of dimensionality when the dimensionality is high. As a compromise, the semiparametric regression model, or the partly linear model (PLM, Härdle, Liang and Gao, 2000) is more attractive, as it retains the flexibility of the nonparametric model and avoids the need to model a fully nonparametric model. In this approach, the main covariates of interest may be modeled in a parametric form while the other scalar covariate (usually time) is treated nonparametrically. At other times, the semiparametric model arises naturally due to categorical covariates (e.g., treatment effects). Existing applications of PLM to longitudinal data accounting for within-subject dependence mainly focus on regression analysis of the mean. The covariance is usually assumed known up to a few parameters. For example, Lin and Carroll (2001) considered the generalized partial linear model in a general case with the profile-kernel estimating equations. Wang (2003) and Wang, Carroll and Lin (2005) further proposed a semiparametric estimation method in marginal partly linear models. This estimation achieves the semiparametric efficiency if the covariance is correctly specified. He, Zhu and Fung (2002) considered a robust semiparametric model for the mean by ignoring the dependence structure. Furthermore, He, Fung and Zhu (2005) proposed a robust estimation method accounting for correlation using the techniques of GEE and regression splines. Welsh et al. (2002) showed that regression splines can also result in better efficiency over kernel methods in nonparametric regression models with longitudinal data. Lin and Carroll (2006) considered a wide class of semiparametric problems for the mean structure.

Compared to the models for the mean in longitudinal data analysis, model based analysis for the covariance is much less studied. To address this issue, we propose semiparametric models for the mean and the covariance structure for longitudinal data. Our formulation builds on the modified Cholesky decomposition advocated by Pourahmadi (1999) such that the entries in this decomposition can be modeled by semiparametric regression models. We adopt the regression spline method for the nonparametric part as it is theoretically sound and computationally convenient (He and Shi, 1996). On the one hand, our model retains the flexibility of the nonparametric approach, compared to that in Ye and Pan (2006). On the other hand, it avoids the problem of dimensionality due to the partly linear form of the regression functions. At the same time, the positive definiteness of the variance matrices is assured. There are also related works in this regard. Wu and Pourahmadi (2003) proposed nonparametric estimates of the covariance matrix, but their method does not deal with irregular observed measurements. Fan, Huang and Li (2007) and Fan and Wu (2008) studied a different semiparametric model for the covariance structure. They estimated the marginal variance via kernel smoothing and proposed a parametric model for the correlation matrix. Similar to the method in Fan, Huang and Li (2007), our approach can handle irregularly and possibly subject-specific times points. We show that the resulting estimators for the regression coefficients in the mean and the variance are consistent and asymptotically normally distributed. Furthermore, the nonparametric parts are estimated at the optimal convergence rate. An iterative algorithm, which is simple to implement, is also developed for computing the estimates.

The rest of the paper is organized as follows. Section 2 introduces the models and estimation methods. Theoretical properties of the proposed estimators are given in Section 3. Extensive simulations and data analysis are presented in Section 4. Section 5 gives some concluding remarks. All the proofs are relegated to the Appendix.

2 The Models and The Estimation Methods

2.1 The models

Let $y_i = (y_{i1}, \cdots, y_{in_i})'$ be the n_i repeated measurements at time points $t_i = (t_{i1}, \cdots, t_{in_i})'$ on the ith subject ($i = 1, \cdots, m$), for a total of $n = \sum_{i=1}^{m} n_i$ observations. Note that t_{ij} may be the time or any time-dependent covariate which is modeled nonparametrically. Without loss of generality, we assume that all $\{t_{ij}\}$ are scaled into the interval $[0, 1]$. Furthermore, we assume that the first two moments of the response satisfy $E(y_{ij}|x_{ij}, t_{ij}) = \mu_{ij}^0$ and $V(y_i|x_i, t_i) = \Sigma_{0i}$, where x_{ij} is a p-vector covariate and $x_i = (x_{i1}, ..., x_{in_i})'$ is the covariate matrix for the ith subject. To guarantee the positive definiteness of the matrices Σ_{0i}, an explicit way of modeling Σ_{0i} is via its modified Cholesky decomposition as $\Phi_i \Sigma_{0i} \Phi_i' = D_{0i}$, where Φ_i is a lower triangular matrix with 1's on its diagonal, and D_{0i} is a diagonal matrix. As indicated by Pourahmadi (1999), this decomposition has a clear statistical interpretation. The below-diagonal entries of Φ_i are the negatives of the autoregressive coefficients ϕ_{ijk} defined in

$$\hat{y}_{ij} = \mu_{ij} + \sum_{k=1}^{j-1} \phi_{ijk}(y_{ik} - \mu_{ik}). \tag{1}$$

That is, the autoregressive coefficients are the population regression coefficients of the linear regression of y_{ij} on its predecessors $y_{i(j-1)}, \cdots, y_{i1}$. The diagonal entries σ_{0ij}^2 of D_{0i} can be seen as the innovation variance $\sigma_{0ij}^2 = var(\epsilon_{ij})$, for $\epsilon_{ij} = y_{ij} - \hat{y}_{ij}$. Clearly, the modified Cholesky decomposition is advantageous in that ϕ and $\log(\sigma^2)$ are unconstrained, rather than a constrained parameter Σ_{0i} that must be positive definite. To use the semiparametric regression tools, we postulate three sets of models for μ, ϕ and σ as follows

$$g(\mu_{ij}^0) = x_{ij}'\beta_0 + f_0(t_{ij}), \qquad \phi_{ijk} = w_{ijk}'\gamma_0, \qquad \log(\sigma_{0ij}^2) = z_{ij}'\lambda_0 + f_1(t_{ij}), \tag{2}$$

where x_{ij}, w_{ijk} and z_{ij} are the $p \times 1$, $q \times 1$ and $d \times 1$ vectors of covariates respectively; β_0 is the regression coefficients in the marginal mean; $f_0(\cdot)$ and $f_1(\cdot)$ are unknown smooth functions. The known link function $g(\cdot)$ is assumed to be monotone and differentiable. The covariates x_{ij}, w_{ijk} and z_{ij} may contain the baseline covariates, the time and the associated interactions. etc. The idea of using (2) reflects the belief that regression models for the autoregressive coefficients and innovation variances are as important as that for the mean. Furthermore, model based analysis of these parameters permit more accessible statistical inference. This technique was also used by Ye and Pan (2006), but they assumed parametric models for $g(\mu_{ij}^0)$ and $\log(\sigma_{0ij}^2)$. Thus, the parametric models in Ye and Pan (2006) can be seen as a special case of the model studied in this paper. As discussed earlier, the semiparametric estimating equations are more flexible and can be less biased if the parametric assumption is violated.

2.2 The estimating equations

The two nonparametric functions f_0 and f_1 are parametrized by regression splines, as splines can provide optimal rates of convergence for both the parametric and the nonparametric components in PLM with a small number of knots (Heckman 1986; He and Shi 1996). Additionally, any computational algorithm developed for GLM can be used for fitting a semiparametric extension of GLM, since they treat a nonparametric function as a linear function with the basis functions as covariates. We follow He et al. (2002) and He and Shi (1996) by approximating f_0 and f_1 using the following regression splines representation. For simplicity, we assume that f_0 and f_1 have the same smoothness property. Let $0 = s_0 < s_1 < \cdots < s_{k_n} < s_{k_n+1} = 1$ be a partition of the interval $[0,1]$. Using $\{s_i\}$ as the internal knots, we have $K = k_n + l$ normalized B-spline basis functions of order l that form a basis for the linear spline space. We use the B-spline basis functions because they have bounded support and are numerically stable (Schumaker, 1981). Thus $f_0(t)$ and $f_1(t)$ are approximated by $\pi'(t)\alpha$ and $\pi'(t)\tilde{\alpha}$ respectively, where $\pi(t) = (B_1(t), \cdots, B_K(t))'$ is the vector of basis functions and $\alpha, \tilde{\alpha} \in \mathbb{R}^K$. Let $\pi_{ij} = \pi(t_{ij})$. With this notation, the nonlinear regression models in (2) can be linearized as following:

$$g(\mu_{ij}) = x'_{ij}\beta + \pi'(t_{ij})\alpha = b'_{ij}\theta, \qquad \log(\sigma^2_{ij}) = z'_{ij}\lambda + \pi'(t_{ij})\tilde{\alpha} := h'_{ij}\rho, \tag{3}$$

where $b'_{ij} = (x'_{ij}, \pi'_{ij})$, $h'_{ij} = (z'_{ij}, \pi'_{ij})$, $\theta = (\beta', \alpha')'$ and $\rho = (\lambda', \tilde{\alpha}')'$. We then let $\mu_i = (\mu_{i1}, \cdots, \mu_{in_i})'$, $B_i = (b_{i1}, \cdots, b_{in_i})'$ and define x_i, π_i, z_i and H_i in a similar fashion. Throughout this paper, a scalar function acting on a vector is set to be the vector of the function on each component, for example, $g(\mu_i) = (g(\mu_{i1}), \cdots, g(\mu_{in_i}))'$. Using the GEE method from Liang and Zeger (1986), we construct the estimating equations for θ, γ and ρ as follows

$$\begin{aligned} S_1(\theta) &= \sum_{i=1}^m B'_i \Delta_i \Sigma_i^{-1}(y_i - \mu_i(B_i\theta)) = 0, \\ S_2(\gamma) &= \sum_{i=1}^m V'_i D_i^{-1}(r_i - \hat{r}_i) = 0, \\ S_3(\rho) &= \sum_{i=1}^m H'_i D_i(H_i\rho) W_i^{-1}(\epsilon_i^2 - \sigma_i^2(H_i\rho)) = 0, \end{aligned} \tag{4}$$

where $\Delta_i = \Delta_i(B_i\theta) = diag\{\dot{g}^{-1}(b'_{ij}\theta), \cdots, \dot{g}^{-1}(b'_{in_i}\theta)\}$ and $\dot{g}^{-1}(\cdot)$ is the derivative of the inverse function $g^{-1}(\cdot)$; r_i and $\hat{r}_i$ are the $n_i \times 1$ vectors with jth components $r_{ij} = y_{ij} - \mu_{ij}$ and $\hat{r}_{ij} = E(r_{ij}|r_{i1}, \cdots, r_{i(j-1)}) = \sum_{k=1}^{j-1} \phi_{ijk} r_{ik}$ $(j = 1, \cdots, n_i)$. Note that when $j = 1$ the notation $\sum_{k=1}^{0}$ means zero throughout this paper. It can be shown that $D_i = diag\{\sigma^2_{i1}, \cdots, \sigma^2_{in_i}\}$ in $S_2(\gamma)$ is actually the covariance matrix of $r_i - \hat{r}_i$ and that $V'_i = \partial \hat{r}'_i / \partial\gamma$ is the $q \times n_i$ matrix with jth column $\partial\hat{r}_{ij}/\partial\gamma = \sum_{k=1}^{j-1} r_{ik} w_{ijk}$.

On the other hand, ϵ_i^2 and σ_i^2 in $S_3(\lambda)$ are the $n_i \times 1$ vectors with jth components ϵ_{ij}^2 and σ_{ij}^2 $(j = 1, \cdots, n_i)$, respectively, where $\epsilon_{ij} = y_{ij} - \hat{y}_{ij}$ and $\hat{y}_{ij}$ are given in (1). Obviously, we have $E(\epsilon_i^2) = \sigma_i^2$. In addition, W_i is the covariance matrix of ϵ_i^2, that is, $W_i = var(\epsilon_i^2)$. The solutions of these generalized estimating equations, $\hat{\theta}, \hat{\gamma}$ and $\hat{\rho}$ say, are termed the GEE estimators of θ, γ and ρ. As suggested by Ye and Pan (2006), a sandwich 'working' covariance structure $W_i = A_i^{1/2} R_i(\delta) A_i^{1/2}$ can be used to approximate the true W_i's, where $A_i = 2diag\{\sigma_{i1}^4, \cdots, \sigma_{in_i}^4\}$ and $R_i(\delta)$ mimic the correlation between ϵ_{ij}^2 and ϵ_{ik}^2 $(i \neq k)$ by introducing a new parameter δ. Typical structures for $R_i(\delta)$ include compound symmetry (exchangeable) and $AR(1)$. As with the conventional generalized estimating equations for the mean, the parameter δ may have very little effect on the estimators of γ and ρ. Our real data analysis and simulation studies reported in later sections confirm this point very well.

The three GEE equations in (4) can be seen as a generalization of the conventional GEE for the mean parameters. If we use a working covariance structure for Σ_i in $S_1(\theta)$ and ignore $S_2(\gamma)$ and $S_3(\rho)$, we have the PLM for the mean. The modified Cholesky decomposition allows us to proceed a step further to impose (unconstrained) partly linear structures for the variance components as well.

2.3 The main algorithm

The solutions of θ, γ and ρ satisfy the equations in (4). These parameters can be solved iteratively by fixing the other parameters. For example, for fixed values of θ and γ, ρ can be computed via the third equation in (4). An application of the quasi-Fisher scoring algorithm on equation (4) directly yields the numerical solutions for these parameters. More specifically, given Σ_i, θ can be updated by the iterative procedure

$$\theta^{(k+1)} = \theta^{(k)} + \left\{ \left[\sum_{i=1}^{m} B_i' \Delta_i \Sigma_i^{-1} \Delta_i B_i \right]^{-1} \sum_{i=1}^{m} B_i' \Delta_i \Sigma_i^{-1} (y_i - \mu_i(B_i \theta)) \right\} \Big|_{\theta=\theta^{(k)}}. \quad (5)$$

On the other hand, given θ and ρ, the generalized autoregressive parameters γ can be updated approximately through

$$\gamma^{(k+1)} = \left\{ \left[E \sum_{i=1}^{m} V_i' D_i^{-1} V_i \right]^{-1} \sum_{i=1}^{m} V_i' D_i^{-1} r_i \right\} \Big|_{\gamma=\gamma^{(k)}}. \quad (6)$$

Finally, given θ and γ, the innovation variance parameters ρ can be updated using

$$\rho^{(k+1)} = \rho^{(k)} + \left\{ \left[\sum_{i=1}^{m} H_i' D_i W_i^{-1} D_i H_i \right]^{-1} \sum_{i=1}^{m} H_i' D_i W_i^{-1} (\epsilon_i^2 - \sigma_i^2) \right\} \Big|_{\rho=\rho^{(k)}}, \quad (7)$$

Equation (5)-(7) indicate that, iteratively, the parameters can be estimated using weighted generalized least squares. We summarize the algorithm as follows

1. Initialization step: given a starting value $\zeta^{(0)} = (\theta^{(0)'}, \gamma^{(0)'}, \rho^{(0)'})'$, use the model (2) to form the lower triangular matrices $T_i^{(0)}$ and diagonal matrices $D_i^{(0)}$. Set $k = 0$ to obtain $\Sigma_i^{(0)}$, the starting values of Σ_i;

2. Iteration step: use equation $(5) - (7)$ to calculate the estimators $\theta^{(k+1)}, \gamma^{(k+1)}$ and $\rho^{(k+1)}$;

3. Updating step: replace $\theta^{(k)}, \gamma^{(k)}$ and $\rho^{(k)}$ with the estimators $\theta^{(k+1)}, \gamma^{(k+1)}$ and $\rho^{(k+1)}$. Repeat Steps 2-3 until a desired convergence criterion is met.

A good starting value of $\Sigma_i^{(0)}$ can be simply chosen as I_i, the identity matrix for the ith subject. This initial value of Σ_i guarantees the consistency of the initial estimators in the mean, which in return guarantees consistency of the autoregressive parameters and innovative parameters after the first iteration. In the analysis presented in this paper, the convergence criterion is met as long as the successive difference in the Euclidean norm is less than 10^{-6}. Our numerical experience shows that this iterative algorithm converges very quickly, usually in fewer than 5 iterations.

2.4 Knot selection

Knot selection is an important issue in spline smoothing. The number of knots plays the same role as the smoothing parameter in the smoothing spline models and the bandwidth parameter in kernel smoothing. Intuitively, the number of distinct knots k_n has to increase with $n = \sum_{i=1}^{m} n_i$ for asymptotic consistency. On the other hand, too many knots would increase the variance of estimators. Thus, an objective choice on the optimal number of the knots is needed. In this article, we follow the spline literature (He, et al. 2005) and use the sample quartiles of $\{t_{ij}, i = 1, \cdots, m, j = 1, \cdots, n_i\}$ as knots. For example, if we use three internal knots, they are taken to be the three quartiles of the observed $\{t_{ij}\}$. We use cubic splines (splines of order 4) in the numerical simulation section, and the number of internal knots is taken to be the integer part of $n^{1/5}$, where n is the sample size. This particular choice is consistent with the asymptotic theory of Section 3. According to our empirical experience, this choice works well in a wide variety of problems. A data adaptive procedure is to use the leave-one-subject-out cross validation method, which is usually computationally demanding. Theoretical justification of the leave-one-subject-out cross validation is possible but is already beyond the scope of the paper. We will address this issue in a follow-up work.

3 Asymptotic Properties

Here and throughout, $\|\cdot\|$ for a vector denotes its Euclidean norm, and for any square matrix A, $\|A\|$ denotes its modulus of the largest singular value of A. To study the rates of convergence for $\hat{\beta}, \hat{\gamma}, \hat{\lambda}$ and $\hat{f}_0, \hat{f}_1$, we first give a set of regularity conditions. If the estimating equation (4) has multiple solutions, then only a sequence of consistent estimator $(\hat{\theta}, \hat{\gamma}, \hat{\rho})$ is considered in this section. A sequence $(\hat{\theta}, \hat{\gamma}, \hat{\rho})$ is said to be a consistent sequence if $(\hat{\beta}', \hat{\gamma}', \hat{\lambda}')' \to (\beta_0', \gamma_0', \lambda_0')'$ and $\sup_t |\pi'(t)\hat{\alpha} - f_0(t)| \to 0$, $\sup_t |\tilde{\pi}'(t)\hat{\tilde{\alpha}} - f_1(t)| \to 0$ in probability as $m \to \infty$. The fact that our iterative algorithm starts from consistent estimators of the parameters ensures that the final estimators are also consistent. Our basic conditions are as follows:

(A1) The dimensions p, q and d of covariates x_{ij}, w_{ijk} and z_{ij} are fixed; $m \to \infty$ and $\max_i\{n_i\}$ is bounded, and the distinct values of t_{ij} form a quasi-uniform sequence that grows dense on $[0, 1]$. We also assume that the first four moments of the response exist.

(A2) The sth derivative of f_0 and f_1 are bounded for some $s \geq 2$.

(A3) The covariates w_{ijk} and the matrices W_i^{-1} are all bounded, which means that all the elements of the vectors and matrices are bounded. The function $g^{-1}(\cdot)$ has bounded second derivatives.

(A4) The parametric space Θ is a compact subset of R^{p+q+d}, and the parameter value $\vartheta_0 = (\beta_0', \gamma_0', \lambda_0')'$ is in the interior of the parameter space Θ.

We can see that $n = O(m)$ from (A1) where $n = \sum_{i=1}^m n_i$. The existence of the first four moments of the response is needed for consistently estimating the parameters in the variance. The smoothness conditions on f_0 and f_1 given by Condition (A2) determine the rate of convergence of the spline estimates. Condition (A3) is satisfied as t is bounded. Assumption (A4) is routinely made in linear models.

To study the asymptotic properties of estimators, some assumptions on the covariates x, t and z are needed. The dependence between x_{ij} and t_{ij} is the common issue in semiparametric inference. We assume that

$$x_{ijk} = g_k(t_{ij}) + \delta_{ijk}, \quad k = 1, \cdots, p. \tag{8}$$

$$z_{ijl} = \tilde{g}_l(t_{ij}) + \tilde{\delta}_{ijl}, \quad l = 1, \cdots, d; i = 1, \cdots, m; j = 1, \cdots, n_i; \tag{9}$$

where δ_{ijk}'s and $\tilde{\delta}_{ijl}$'s are mean zero random variables independent of the corresponding random errors and of one another. We let Λ_n and $\tilde{\Lambda}_n$ be the $n \times p$ and $n \times d$ matrices whose kth column are $\delta_k = (\delta_{11k}, \cdots, \delta_{1n_1k}, \cdots, \delta_{mn_mk})'$ and $\tilde{\delta}_k = (\tilde{\delta}_{11k}, \cdots, \tilde{\delta}_{1n_1k}, \cdots, \tilde{\delta}_{mn_mk})'$ respectively. We also make the following assumption:

(A5) (1) $E\Lambda_n$=0, $\sup_n \frac{1}{n} E\|\Lambda_n\|^2 < \infty$; $E\tilde{\Lambda}_n$=0, $\sup_n \frac{1}{n} E\|\tilde{\Lambda}_n\|^2 < \infty$;

(2) $k_n(M'\Sigma^0 M)$ and $k_n M'W^0 M$ are nonsingular for sufficiently large n, and the eigenvalues of $M'\Sigma^0 M k_n/n$ and $M'W^0 M k_n/n$ are bounded away from 0 and infinity, where $M = (\pi_1', \cdots, \pi_m')'$, $\Sigma^0 = diag\{\Sigma_1^0, \cdots, \Sigma_m^0\}$ with $\Sigma_i^0 = \Delta_{0i}\Sigma_{0i}^{-1}\Delta_{0i} = \Delta_i(\eta_i^0)\Sigma_{0i}^{-1}\Delta_i(\eta_i^0)$, and W^0 are defined in a similar fashion respectively.

Condition (2) of (A5) is a property of the B-spline basis functions and is expected to hold under rather general design conditions. The dimension of the approximating B-spline space must increase with n for asymptotic consistency. The number of knots must be chosen properly to balance between the bias and variance. For the optimal rate of convergence, we choose $k_n = O(n^{1/(2s+1)})$.

The asymptotic properties of $(\hat{\beta}, \hat{\gamma}, \hat{\lambda})$ involve computation of the covariance matrix $\Delta_m = (\delta_m^{kl})_{k,l=1,2,3}$ of $(\tilde{S}_1', \tilde{S}_2', \tilde{S}_3')'/\sqrt{m}$, where $\tilde{S}_1, \tilde{S}_2$ and $\tilde{S}_3$ are defined by

$$\begin{aligned}\tilde{S}_1 &= \sum_{i=1}^{m} X_i^{*\prime} \Delta_{0i} \Sigma_{0i}^{-1} (y_i - \mu_{0i}),\\ \tilde{S}_2 &= \sum_{i=1}^{m} V_i^{0\prime} D_{0i}^{-1} (r_{0i} - \hat{r}_{0i}),\\ \tilde{S}_3 &= \sum_{i=1}^{m} Z_i^{*\prime} D_{0i} W_{0i}^{-1} (\epsilon_{0i}^2 - \sigma_{0i}^2).\end{aligned} \tag{10}$$

Here $X^* = (I-P)X$ with $P = M(M'\Sigma^0 M)^{-1} M'\Sigma^0$; $r_{0i} = y_i - \mu_{0i}$, $\hat{r}_{0i} = (\hat{r}_{0i1}, \cdots, \hat{r}_{0ij}, \cdots, \hat{r}_{0in_i})'$ with $\hat{r}_{0ij} = \sum_{k=1}^{j-1} r_{0ik} w_{ijk}' \gamma_0$; $V_i^0 = (0, r_{0i1} w_{i21}', \cdots, \sum_{k=1}^{j-1} r_{0ik} w_{ijk}')'$; and $Z^* = (I-P)Z$. We make the following assumption, similar to the assumption (11) in Ye and Pan (2006).

(A6) The covariance matrix Δ_m is positive definite, and

$$\Delta_m = \begin{pmatrix} \delta_m^{11} & \delta_m^{12} & \delta_m^{13} \\ \delta_m^{21} & \delta_m^{22} & \delta_m^{23} \\ \delta_m^{31} & \delta_m^{32} & \delta_m^{33} \end{pmatrix} \to \Delta = \begin{pmatrix} \delta^{11} & \delta^{12} & \delta^{13} \\ \delta^{21} & \delta^{22} & \delta^{23} \\ \delta^{31} & \delta^{32} & \delta^{33} \end{pmatrix}, \tag{11}$$

as $m \to \infty$. Here Δ is a positive definite matrix.

Theorem 1. *If Assumptions (A1) to (A6) hold and the number of knots satisfies $k_n = O(n^{1/(2s+1)})$, then*

$$\frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{n_i} \left\{ \hat{f}_0(t_{ij}) - f_0(t_{ij}) \right\}^2 = O_p(n^{-2s/(2s+1)}) \tag{12}$$

$$\frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{n_i} \left\{ \hat{f}_1(t_{ij}) - f_1(t_{ij}) \right\}^2 = O_p(n^{-2s/(2s+1)}) \tag{13}$$

where $\hat{f}_0(t) = \pi'(t)\hat{\alpha}$ and $\hat{f}_1(t) = \pi'(t)\widehat{\tilde{\alpha}}$.

As pointed out by He *et al.* (2005), (12) and (13) imply that

$$\int (\hat{f}_i(t) - f_i(t))^2 dt = O_p(n^{-2s/(2s+1)}) \qquad (i = 0, 1)$$

under some general conditions (see, e.g., Lemmas 8 and 9 in Stone 1985). This is the optimal rate of convergence for estimating f_0 and f_1 under the smoothness condition (A2) above. For the parametric parameters, we have the following asymptotic normality results.

Theorem 2. *Under conditions (A1)-(A6), the generalized estimating equation estimator* $(\hat{\beta}'_m, \hat{\gamma}'_m, \hat{\lambda}'_m)'$ *is* $\sqrt{m}$*-consistent and asymptotically normal, that is*

$$\sqrt{m}\begin{pmatrix} \hat{\beta}_m - \beta_0 \\ \hat{\gamma}_m - \gamma_0 \\ \hat{\lambda}_m - \lambda_0 \end{pmatrix} \to N\left\{0, \begin{pmatrix} \delta^{11} & 0 & 0 \\ 0 & \delta^{22} & 0 \\ 0 & 0 & \delta^{33} \end{pmatrix}^{-1} \begin{pmatrix} \delta^{11} & \delta^{12} & \delta^{13} \\ \delta^{21} & \delta^{22} & \delta^{23} \\ \delta^{31} & \delta^{32} & \delta^{33} \end{pmatrix} \begin{pmatrix} \delta^{11} & 0 & 0 \\ 0 & \delta^{22} & 0 \\ 0 & 0 & \delta^{33} \end{pmatrix}^{-1} \right\}$$

in distribution as $m \to \infty$.

The asymptotic variance reduces to a diagonal matrix when the response variable is normally distributed. This is due to the fact that $\delta_{kl} = 0$ when $k \neq l$. This is the semiparametric analog to Theorem 2 by Ye and Pan (2006).

For statistical inference, we use a robust estimator of the covariance matrix of $\hat{\beta}_m$, i.e.,

$$V(\hat{\beta}_m) = M_0^{-1} M_1 M_0^{-1}, \tag{14}$$

where

$$M_0 = \sum_{i=1}^{m} X_i^{*\prime} \hat{\Delta}_i \hat{\Sigma}_i \hat{\Delta}_i X_i^*, \qquad M_1 = \sum_{i=1}^{m} X_i^{*\prime} \hat{\Delta}_i \hat{\Sigma}_i (y_i - \hat{\mu}_i)(y_i - \hat{\mu}_i)' \hat{\Delta}_i X_i^*.$$

The estimated covariance matrices of $\hat{\gamma}_m$ and $\hat{\lambda}_m$ can be obtained in a similar way, and the covariances $\delta^{kl} (k \neq l)$ can also be estimated by their sample versions.

4 Numerical Study

4.1 Real data analysis

We apply the proposed estimation method to the CD4 cell study. This data set was analyzed by many authors, see Ye and Pan (2006), Zeger and Diggle (1994) for example. This data set comprises CD4 cell counts of 369 HIV-infected men with six covariates including time since seroconversion (t_{ij}), age (relative to arbitrary origin,

x_{ij1}), packs of cigarettes smoked per day (x_{ij2}), recreation drug use (x_{ij3}), number of sexual partners (x_{ij4}), cesd (mental illness score, x_{ij5}). Altogether there are 2,376 values of CD4 cell counts, with multiple repeated measurements taken for each individual at different times, covering a period of approximately eight and a half years. The number of measurements for each individual varies from 1 to 12 and the time are not equally spaced. Thus, the CD4 cell data are highly unbalanced. We use square root transformation on the response by the suggestion in Zeger and Diggle (1994), where further details about the design and the medical implications of the study can be found.

The object of our analysis is to model jointly the mean and covariance structures for the CD4 cell data. For that, we propose to use the following mean model

$$y_{ij} = x_{ij1}\beta_1 + x_{ij2}\beta_2 + x_{ij3}\beta_3 + x_{ij4}\beta_4 + x_{ij5}\beta_5 + f(t_{ij}) + e_{ij},$$

where $i = 1, \cdots, 369$; $j = 1, \cdots, n_i$, $\sum_{i=1}^{369} n_i = 2376$. We take covariates for the autoregressive components as $w_{ijk} = (1, t_{ij} - t_{ik}, (t_{ij} - t_{ik})^2, (t_{ij} - t_{ik})^3)$ following the arguments in Ye and Pan (2006), and for the innovation variances as $z_{ij} = x_{ij}$. The latter specification allows us to examine whether the innovations are dependent on the covariates. Finally the number of knots is taken to be $[(2356)^{1/5}] = 7$, which is also the optimal number of knots according to the leave-one-subject-out cross validation. Table 1 lists the results for β by our modified Cholesky decomposition method, where a normal working AR(1) model with $\delta = 0.2$ is used for the innovation. For comparison, we also list the conventional GEE method for the mean using different working correlations, including independent, AR(1) and exchangeable structures. The results show that our method gives estimators with generally smaller standard errors. For our approach, smoking and drug use are highly significant variables, while mental illness score is marginally significant. The significance of smoking is missed by GEE using AR(1) covariance structure, while that of drug use is missed by GEE using either AR(1) or exchangeable variance structure. Finally, GEE using the independent working correlation indicates that mental score is not significant at all, which contradicts with the GEE results using other working correlations.

Figure 1 displays the three fitted curves for f_0, ϕ as a function of w and f_1, when $R_i(\delta)$ is specified by $AR(1)$ with $\delta = 0.2$. The asymptotic pointwise 95% confidence intervals are also provided. The trajectory of the mean curve is consistent with that in Zeger and Diggle (2002). The autoregressive curve is decreasing with the time lag and the innovation curve seems to fluctuate around a constant. These observations basically agree with those in Ye and Pan (2006).

Table 1 is about here.

Figure 1 is about here.

4.2 Simulation study

We conduct extensive numerical studies to assess the finite sample performance of the proposed method. We also test the asymptotic covariance formula in Theorem 2 and compare the proposed approach with conventional GEE using a working correlation matrix.

Study 1. We first consider the following model

$$y_{ij} = x_{ij1}\beta_1 + x_{ij2}\beta_2 + f_0(t_{ij}) + \epsilon_{ij},\ i = 1, \cdots, m;\ j = 1, \cdots, n_i,$$

for $m = 100$. We use the sample scheme similar to that in Fan, Huang and Li (2007) such that the observation times are regularly scheduled but may be randomly missed in practice. More precisely, we generate the observation times in the following way.

Each individual has a set of scheduled time points $\{0, 1, 2, ..., 12\}$, and each scheduled time, except time 0, has a 20% probability of being skipped. The actual observation time is a random perturbation of a scheduled time: a uniform [0, 1] random variable is added to a nonskipped scheduled time. This results in different observed time points t_{ij} per subject, and then t_{ij} is transformed onto $[0, 1]$.

We take $x_{ij1} = t_{ij} + \delta_{ij}$, where δ_{ij} follows the standard normal distribution and let x_{ij2} follows a Bernoulli distribution with success probability 0.5. Thus x_{ij1} is time-varying. For the nonparametric function in the mean, we take $f_0(t) = cos(t\pi)$. The error $(\epsilon_{in1}, \cdots, \epsilon_{in_i})'$ is generated to follow a multivariate normal distribution with mean 0 and covariance Σ_i satisfying $\Phi_i \Sigma_i \Phi_i' = D_i$, where Φ_i and D_i are described in Section 2.1 with $w_{ijk} = (1, t_{ij} - t_{ik})'$, $z_{ij} = x_{ij}$, and $f_1(t) = sin(\pi t)$. We consider two kinds of correlation structures: compound symmetry (exchangeable) and $AR(1)$ for $R_i(\delta)$ in $W_i = A_i^{1/2} R_i(\delta) A_i^{1/2}$ the working covariance structure of ϵ_i^2. In each case the parameter δ, measuring the correlation between $\epsilon_{ij}^2 = (y_{ij} - \hat{y}_{ij})^2$ and $\epsilon_{ik}^2 = (y_{ik} - \hat{y}_{ik})^2$, takes four different values, $\delta = 0, 0.2, 0.5, 0.8$, so that the effect of misspecification of $R_i(\delta)$ on the GEE estimators β, γ, and λ can be studied. We take two specifications of the parametric coefficients as (1) $\beta = (1, 0.5)'$, $\gamma = (0.2, 1.5)'$ and $\lambda = (-0.5, 1.5)'$; (2) $\beta = (1, 0)', \gamma = (0.2, 0)'$, and $\lambda = (-0.5, 0)'$. For each setting, one hundred data sets are simulated such that the expected sample size is about 1060. The number of the knots is taken to be $4 \approx 1060^{1/5}$. Numerical experiments show that the results are not very sensitive to the number of the knots.

Table 2 shows that our semiparametric methods literally yield unbiased estimates for the parameters. Additionally, the parameter δ used in the working covariance structure for the innovations has little effect on the estimation of β, γ and λ, and the estimated mean square error for f_0 and f_1, when the structure for $R_i(\delta)$ is based on $AR(1)$. The results obtained based on the compound symmetry structure are very similar and thus are omitted. These results imply that the semiparametric GEE estimators are robust against misspecification of the structure of $R_i(\delta)$. Figure 2 displays the true and fitted curves for nonparametric function f_0 and f_1 when $R_i(\delta)$ is specified by $AR(1)$ with $\delta = 0.2$. The three curves $\hat{f}_5, \hat{f}_{50}$ and $\hat{f}_{95}$ represent the fits which are 5%, 50% and 95% best in terms of the mean squred errors in 100 runs, respectively. They show a close agreement with the true functions. Note that the longitudinal observations are highly irregular and some of $\{n_i\}$ are less than the number of the parameters in the same subject.

Table 2 is about here.

Figure 2 is about here.

Study 2. We use this example to illustrate the performance of the asymptotic covariance formula in Theorem 2. Here the simulation setup is the case (1) in Study 1 but the number of runs is increased to 1,000. In Table 3, "SD" represents the sample standard deviation of 1,000 estimates of β, γ and λ, which can be viewed as the true standard deviation of the resulting estimates. "SE" represents the sample average of 1,000 estimated standard errors using formula (14), and "Std" represents the standard deviation of these 1,000 standard errors. Table 3 demonstrates that the standard error formula works well for different AR(1) correlation structures.

Table 3 is about here.

Study 3. In this example, we study the effect of misspecification of the working covariance structure Σ_i on the estimation of β. Again, we repeat the experiment 100 times. For comparison, we apply GEE with independent, exchangeable and AR(1) working correlation. The results are summarized in Table 4, in which the third column is based on the estimation of Σ_i by the modified Cholesky decomposition method proposed in this paper, under AR(1) structure with $\delta = 0.2$. Other choices of δ are tried and they yield similar results. The fourth column to the last column is based on GEE using conventional working covariance structure of Σ_i and they represent the results by using the working independent, exchangeable and AR(1) working correlation matrix, respectively. Not surprisingly, all methods give almost unbiased estimates for β. However, the standard error of our semiparametric method is much smaller than those of the other methods, implying that our estimator is more efficient. Furthermore, for the nonparametric part f_0 in the mean, our semiparametric approach gives estimates with significantly smaller mean square errors. Taken together, the study shows that the semiparametric approach is more accurate in estimating the mean.

Table 4 is about here.

5 Discussion

We have proposed semiparametric mean-covariance models for longitudinal data analysis. The modified Cholesky decomposition is adopted such that partly linear regression models can be applied to the autoregressive coefficients and log innovation variances. On the one hand, our approach extends the semiparametric model for the mean in longitudinal analysis. On the other hand, our approach relaxes the parametric assumption made by Ye and Pan (2006) on the innovation variances. For future research, it would be interesting to extend the semiparametric approach to nonnormal longitudinal data analysis.

Appendix: sketch of proofs

The following lemma, which follows easily from Theorem 12.7 of Schumaker (1981), is stated for easy reference.

Lemma 1. *Under Assumptions (A1) and (A2), there exists constants C_0 and C_1 such that*

$$\sup_{t\in[0,1]} |f_0(t) - \pi'(t)\alpha_0| \le C_0 k_n^{-s}, \qquad \sup_{t\in[0,1]} |f_1(t) - \pi'(t)\tilde{\alpha}_0| \le C_1 k_n{}^{-s}.$$

Proof of Theorem 1

Equation (12) can be obtained directly from He et al (2005). Here we only give a proof of equation (13) when all W_i are known, denoted by W_{0i}. Similar asymptotic results hold when all W_{0i} are replaced by consistent estimates. Let

$$T_m = \begin{pmatrix} A_m^{-1/2} & -A_m^{-1/2} H' W^0 M (M' W^0 M)^{-1} \\ 0 & k_n^{1/2} Q_m^{-1} \end{pmatrix},$$

where $A_m = H^{*\prime} W^0 H^* = \sum_{i=1}^m H_i^{*\prime} W_i^0 H_i^*$, $Q_m^2 = k_n M' W^0 M$. Obviously, condition (A6) implies that $A_m/m \to A > 0$ in probability for some positive matrix A as $m \to \infty$. From the definition of T_m, it is easy to know that $T_m \sum_{i=1}^m H_i' D_{0i} W_{0i}^{-1} D_{0i} H_i T_m' = I_{d+K}$, where I_{d+K} is $(d+K)\times(d+K)$ identity matrix.

We further let

$$\zeta(\rho) = \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix} = (T_m')^{-1}(\rho - \rho_0) = \begin{pmatrix} A_m^{1/2}(\lambda - \lambda_0) \\ k_n^{-1/2} Q_m(\tilde{\alpha} - \tilde{\alpha}_0) + k_n^{1/2} Q_m^{-1} M' W^0 H (\lambda - \lambda_0) \end{pmatrix} \tag{A.1}$$

and

$$\hat{\zeta} = \begin{pmatrix} \hat{\zeta}_1 \\ \hat{\zeta}_2 \end{pmatrix} = \zeta(\hat{\lambda}, \widehat{\tilde{\alpha}}). \tag{A.2}$$

From Lemma 1, it is easy to know that for sufficiently large n,

$$\frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left\{\hat{f}_1(t_{ij})-f_1(t_{ij})\right\}^2 \;\le\; \frac{2}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\pi'_{ij}(\widehat{\tilde{\alpha}}-\tilde{\alpha}_0))^2+2C_0k_n^{-2s}, \quad \text{(A.3)}$$

and $\|A_m^{-1/2}(\hat{\lambda}-\lambda_0)\|\le\|\hat{\zeta}\|$,

$$\begin{aligned}\Big[\frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\pi'_{ij}(\widehat{\tilde{\alpha}}-\tilde{\alpha}_0))^2\Big]^{1/2} &= n^{-1/2}\|M(\widehat{\tilde{\alpha}}-\tilde{\alpha}_0)\|\le Cn^{-1/2}\|k_n^{-1/2}Q_n(\widehat{\tilde{\alpha}}-\tilde{\alpha}_0)\| \\ &\le\; Cn^{-1/2}\|\hat{\zeta}\|+C\lambda_n^{-1/2}\|\hat{\lambda}-\lambda_0\|\sup_{\|a\|=1,\|b\|=1}|n^{-1}a'M'W^0Hbk_n^{1/2}|,\end{aligned}$$

where λ_n is the minimum eigenvalue of $k_nM'W^0M/n$. Then by lemma 6.2 of He and Shi (1996) it suffices to show that $\|\hat{\zeta}\|=O_p(k_n^{1/2})$. To do so, let $R_{mi}=\pi_i\tilde{\alpha}_0-f_1(t_i)$, $\eta_i^0=H_i\lambda_0+f_1(t_i)$, and $\varsigma_i=\tilde{H}_i\zeta+R_{mi}$, where $\tilde{H}_i=H_iT'_m=(H_i^*A_m^{-1/2},\pi_iQ_m^{-1}k_n^{1/2})$. Then it's easy to see that

$$H_i\zeta=\eta_i^0+\varsigma_i, \quad \sigma_i^2=\exp(\eta_i^0+\varsigma_i),$$

and the third estimating equation of (4) can be rewritten as

$$S_\zeta(\zeta)=\sum_{i=1}^{m}H'_iD_i(\eta_i^0+\varsigma_i)W_{0i}^{-1}(\epsilon_i^2-\exp(\eta_i^0+\varsigma_i))=0. \quad \text{(A.4)}$$

Multiply T_m to equation (A.4) and we get

$$\Psi(\zeta)=T_mS_\zeta(\zeta)=\sum_{i=1}^{m}\tilde{H}'_iD_i(\eta_i^0+\varsigma_i)W_{0i}^{-1}(\epsilon_i^2-\exp(\eta_i^0+\varsigma_i))=0. \quad \text{(A.5)}$$

It is easy to known that both (A.4) and (A.5) give the same root for ζ by conditions (A4) and (A5). Let $a\in\mathbb{R}^{d+\tilde{K}}$, satisfying $a'a=1$. We expand $a'\Psi(\zeta)$ in a Taylor series,

$$\begin{aligned}a'\Psi(\zeta) &= \sum_{i=1}^{m}a'\tilde{H}'_iD_i(\eta_i^0+\varsigma_i)W_{0i}^{-1}(\epsilon_i^2-\exp(\eta_i^0+\varsigma_i)) \\ &= \sum_{i=1}^{m}a'\tilde{H}'_iD_{0i}W_{0i}^{-1}(\epsilon_i^2-\sigma_{0i}^2)-\sum_{i=1}^{m}a'\tilde{H}'_iD_{0i}W_{0i}^{-1}D_{0i}\varsigma_i\end{aligned}$$

$$+ \quad \sum_{i=1}^{m} \varsigma_i' \frac{\partial a' \tilde{H}_i' D_i}{\partial \varsigma_i}\Big|_{\varsigma_i=0} W_{0i}^{-1}(\epsilon_i^2 - \sigma_{0i}^2) + R_m^*(\varsigma^*), \tag{A.6}$$

where $R_m^*(\varsigma^*) = \sum_{i=1}^m R_{mi}^*(\varsigma_i^*)$ and $R_{mi}^*(\varsigma_i^*) = \frac{1}{2}\varsigma_i'[\partial^2 a' \tilde{H}_i' D_i W_{0i}^{-1}(\epsilon_i^2 - \sigma_i^2)/\partial\varsigma_i\partial\varsigma_i'|_{\varsigma_i=\varsigma_i^*}]\varsigma_i$ for $\varsigma_i^* = \eta_i^0 + \tau_i\varsigma_i$ $(i = 1, \cdots, m)$ with $0 < \tau_i < 1$.

Further, let

$$\Phi(\zeta) = \sum_{i=1}^{m} \tilde{H}_i' D_{0i} W_{0i}^{-1}(\epsilon_{0i}^2 - \sigma_{0i}^2) - \zeta, \tag{A.7}$$

where $\epsilon_{0i} = y_i - \mu_{0i}$. Denote the solution of Φ as $\tilde{\zeta}$, that is,

$$\tilde{\zeta} = \begin{pmatrix} \tilde{\zeta}_1 \\ \tilde{\zeta}_2 \end{pmatrix} = \sum_{i=1}^{m} \tilde{H}_i' D_{0i} W_{0i}^{-1}(\epsilon_{0i}^2 - \sigma_{0i}^2). \tag{A.8}$$

From (A.6) and (A.7) the difference between $a'\Psi(\zeta)$ and $a'\Phi(\zeta)$ can be expressed as

$$\begin{aligned} a'(\Psi(\zeta) - \Phi(\zeta)) &= \sum_{i=1}^{m} a' \tilde{H}_i' D_{0i} W_{0i}^{-1}(\epsilon_i^2 - \epsilon_{0i}^2) - \sum_{i=1}^{m} a' \tilde{H}_i' D_{0i} W_{0i}^{-1} D_{0i} R_{mi} \\ &+ \quad \sum_{i=1}^{m} \varsigma_i' \frac{\partial a' \tilde{H}_i' D_i}{\partial \varsigma_i}\Big|_{\varsigma_i=0} W_{0i}^{-1}(\epsilon_i^2 - \sigma_{0i}^2) + R_m^*(\varsigma^*) \\ &=: \quad I_{n0} - I_{n1} + I_{n2}(\zeta) + R_m^*(\varsigma^*). \end{aligned} \tag{A.9}$$

By Cauchy-Schwarz inequality, the definition of $\tilde{H}$ and $k_n = O(n^{1/(2s+1)})$, we have

$$\begin{aligned} E(I_{n0})^2 &\leq \sum_{i=1}^{m} a' \tilde{H}_i' D_{0i} W_{0i}^{-1} D_{0i} \tilde{H}_i a \sum_{i=1}^{m} E(\epsilon_i^2 - \epsilon_{0i}^2)' W_{0i}^{-1}(\epsilon_i^2 - \epsilon_{0i}^2) \\ &= \sum_{i=1}^{m} E(\tilde{D}(\nabla\mu_i)\epsilon_{0i} + (\nabla\mu_i)^2)' W_{0i}^{-1}(\tilde{D}(\nabla\mu_i)\epsilon_{0i} + (\nabla\mu_i)^2) \\ &= \sum_{i=1}^{m} trace\{\tilde{D}(\nabla\mu_i) W_{0i}^{-1} \tilde{D}(\nabla\mu_i)\Sigma_{0i}\} + \sum_{i=1}^{m} \{(\nabla\mu_i)^2\}' W_{0i}^{-1}(\nabla\mu_i)^2 \\ &\leq C\sum_{i=1}^{m}[\|\nabla\mu_i\|^2 + \sum_{i=1}^{m}\sum_{j=1}^{n_i}[\nabla\mu_{ij}]^4 = O_p(k_n), \end{aligned} \tag{A.10}$$

where $\nabla\mu_i = (\mu_{i1} - \mu_{0i1}, \cdots, \mu_{in_i} - \mu_{0in_i})'$ and $\tilde{D}(\nabla\mu_i) = diag\{\mu_{i1} - \mu_{0i1}, \cdots, \mu_{in_i} - \mu_{0in_i}\}$ with $\mu_i = \mu_i(B_i\hat{\theta}) = g^{-1}(B_i\hat{\theta})$. The last inequality in (A.10) can be obtained easily by He et al (2005). Thus

$$|I_{n0}| = O_p(k_n^{1/2}). \tag{A.11}$$

For I_{n1}, Obviously,

$$\begin{aligned} |I_{n1}| &= \Big|\sum_{i=1}^{m} a'\tilde{H}_i' D_{0i} W_{0i}^{-1} D_{0i} R_{mi}\Big| = |a'\tilde{H}W^0 R_m| \\ &= \{a'\tilde{H}'W^0\tilde{H}a\}^{1/2}\{R_m'\Sigma^0 R_m\}^{1/2} = O(n^{1/2}k_n^{-s}) = O(k_n^{1/2}), \qquad \text{(A.12)} \end{aligned}$$

where $R_m = (R_{m1}, \cdots, R_{mn_m})'$.

For $I_{n2}(\zeta)$, write

$$I_{n2}(\zeta) = \sum_{i=1}^{m} \zeta'\tilde{H}_i' G_{0,i} + \sum_{i=1}^{m} R_{mi}' G_{0,i} =: I_{n2}^{(1)}(\zeta) + I_{n2}^{(2)}, \qquad \text{(A.13)}$$

where $G_{0,i} = \frac{\partial a'\tilde{H}_i' D_i}{\partial \varsigma_i}\Big|_{\varsigma_i=0} W_{0i}^{-1}(\epsilon_i^2 - \sigma_{0i}^2)$. Let $\tilde{e}_i = W_{0i}^{-1}(\epsilon_i^2 - \sigma_{0i}^2) = (\tilde{e}_{i1}, \cdots, \tilde{e}_{in_i})'$. It is easy to know that $G_{0i} = diag\{\sigma_{0i1}^2\tilde{e}_{1i}, \cdots, \sigma_{0in_i}^2\tilde{e}_{in_i}\}\tilde{H}_i a =: A(\tilde{e}_i)\tilde{H}_i a$. Then by Cauchy-Schwarz inequality, we have

$$\begin{aligned} \left(I_{n2}^{(1)}\right)^2 &= \left(\sum_{i=1}^{m} \zeta'\tilde{H}_i' A(\tilde{e}_i)\tilde{H}_i a\right)^2 = \left(\sum_{k=1}^{\bar{d}} \xi_k \sum_{i=1}^{m} \mathbf{1}_k'\tilde{H}_i' A(\tilde{e}_i)\tilde{H}_i a\right)^2 \\ &\le \|\xi\|^2 \sum_{k=1}^{\bar{d}} \left(\sum_{i=1}^{m} \mathbf{1}_k'\tilde{H}_i' A(\tilde{e}_i)\tilde{H}_i a\right)^2 \le \|\xi\|^2 \sum_{k,j=1}^{\bar{d}} \left(\sum_{i=1}^{m} \mathbf{1}_k'\tilde{H}_i' A(\tilde{e}_i)\tilde{H}_i \mathbf{1}_j\right)^2, \end{aligned}$$

where $\bar{d} = d + K$ and $\mathbf{1}_k = (0, \cdots, 0, 1, 0, \cdots, 0)'$ is a $\bar{d}$ vector with 1 as its kth element and 0 elsewhere. By conditions (A1),(A3), (A5) and (A6), we have

$$\begin{aligned} E(I_{n2}^{(1)})^2 &\le C\|\zeta\|^2 \sum_{k,j=1}^{\bar{d}} \sum_{i=1}^{m} E\left(\mathbf{1}_k'\tilde{H}_i' A(\tilde{e}_i)\tilde{H}_i \mathbf{1}_j\right)^2 \\ &\le C\|\zeta\|^2 \sum_{k,j=1}^{\bar{d}} \sum_{i=1}^{m} \mathbf{1}_k'\tilde{H}_i'\tilde{H}_i \mathbf{1}_k E\|A(\tilde{e}_i)\tilde{H}_i \mathbf{1}_j\|^2 \\ &\le C\|\zeta\|^2 \sup_i \sum_{k=1}^{\bar{d}} \mathbf{1}_k'\tilde{H}_i'\tilde{H}_i \mathbf{1}_k \sum_{i=1}^{m}\sum_{k=1}^{\bar{d}} \mathbf{1}_k'\tilde{H}_i'\tilde{H}_i \mathbf{1}_k O(k_n) \\ &= C\|\zeta\|^2 \sup_i trace\{\tilde{H}_i\tilde{H}_i'\} trace\{\sum_{i=1}^{m} \tilde{H}_i\tilde{H}_i'\} O(k_n) \\ &\le C\|\zeta\|^2 k_n \sup_i trace\{H_i^* A_m^{-1} H_i^{*\prime} + k_n \pi_i Q_m^{-2}\pi_i'\} O(k_n) \\ &= O(k_n^3\|\zeta\|^2/n), \end{aligned}$$

where the constant C, independent of n, may vary from line to line. Therefore, for sufficiently large L, we have

$$\sup_{\|\zeta\|\leq Lk_n^{1/2}, a'a=1} |I_{n2}^{(1)}(\zeta)| = O_p(n^{-1/2}k_n^2). \tag{A.14}$$

Similarly,

$$\sup_{a'a=1} |I_{n2}^{(2)}| = O(k_n^{1-s}). \tag{A.15}$$

Combining (A.14) and (A.15) and $k_n = O(n^{1/(2s+1)})$, we obtain

$$\sup_{a'a=1} |I_{n2}| = O_p(k_n^{1/2}). \tag{A.16}$$

For $R_m^*(\varsigma^*)$, let $F_i^* = (\partial^2 a'\tilde{H}_i' D_i W_{0i}^{-1}(\epsilon_i^2 - \sigma_i^2)/\partial\varsigma_i\partial\varsigma_i')|\varsigma_i = \varsigma_i^*$, we see that

$$\begin{aligned} R_m^*(\varsigma^*) &= \frac{1}{2}\sum_{i=1}^m \zeta'\tilde{H}_i' F_i^* \tilde{H}_i\xi + \sum_{i=1}^m R_{mi}' F_i^* \tilde{H}_i\zeta + \frac{1}{2}\sum_{i=1}^m R_{mi}' F_i^* R_{mi} \\ &= I_{n3}^{(1)}(\zeta) + I_{n3}^{(2)}(\zeta) + I_{n3}^{(3)}(\zeta). \end{aligned}$$

By assumptions (A3), (A5) and (A6), we have that $\sup_{1\leq i\leq m, a'a=1} \|F_i^*\| = O_p(n^{-1/2}k_n^{1/2})$. Hence

$$\begin{aligned} \sup_{\|\zeta\|\leq Lk_n^{1/2}, a'a=1} |I_{n3}^{(1)}(\xi)| &= O_p(n^{-1/2}k_n^{5/2}), \\ \sup_{\|\zeta\|\leq Lk_n^{1/2}, a'a=1} |I_{n3}^{(2)}(\xi)| &= O_p(k_n^{3/2-s}), \\ \sup_{\|\zeta\|\leq Lk_n^{1/2}, a'a=1} |I_{n3}^{(3)}(\xi)| &= O_p(n^{1/2}k_n^{1/2-2s}), \\ \sup_{\|\zeta\|\leq Lk_n^{1/2}, a'a=1} |R_n^*(\varsigma^*)| &= O_p(k_n^{1/2}). \end{aligned}$$

Putting all the approximations together, we have

$$\sup_{\|\zeta\|\leq Lk_n^{1/2}} \|\Psi(\zeta) - \Phi(\zeta)\| = O_p(k_n^{1/2}),$$

and for sufficient large C, direct calculations give

$$E\|\tilde{\zeta}\|^2 = \sum_{i=1}^m E[(\epsilon_{0i}^2 - \sigma_{0i}^2)' W_{0i}^{-1} D_{0i} H_i^* A_m^{-1} H_i^{*\prime} D_{0i} W_{0i}^{-1}(\epsilon_{0i}^2 - \sigma_{0i}^2)$$

$$+k_n(\epsilon_{0i}^2-\sigma_{0i}^2)'W_{0i}^{-1}D_{0i}\pi_i Q_m^{-2}\pi_i' D_{0i}W_{0i}^{-1}(\epsilon_{0i}^2-\sigma_{0i}^2)]$$

$$\leq Ctrace\{H^* A_m^{-1} H^{*\prime} + k_n M Q_m^{-2} M'\} = O(k_n).$$

Therefore,

$$\sup_{\|\zeta\|\leq Lk_n^{1/2}} \|\Psi(\zeta)-\zeta\| \leq \sup_{\|\zeta\|\leq Lk_n^{1/2}} \|\Psi(\zeta)-\Phi(\zeta)\| + \|\tilde{\zeta}\| = LO_p(k_n^{1/2}) + O_p(k_n^{1/2}), \quad \text{(A.17)}$$

which implies that $\sup_{\|\zeta\|\leq Lk_n^{1/2}} \|\Psi(\zeta)-\zeta\| \leq Lk_n^{1/2}$ in probability for sufficiently large L. Thus Brouwer's fixed-point theorem ensures that the map $\zeta \mapsto \zeta - \Psi(\zeta)$ has a fixed point $\hat{\zeta}$ that is a zero of $\Psi(\zeta)$ with $\|\hat{\zeta}\| = O_p(k_n^{1/2})$.

Lemma 2. *Under conditions (A1)-(A6), Let* $(\hat{\beta}_m', \hat{\alpha}_m', \hat{\gamma}_m', \hat{\lambda}_m', \widehat{\tilde{\alpha}}_m')'$ *be the root of generalized estimating equation (4), then*

$$\|\hat{\xi}_1 - \tilde{\xi}_1\| = o_p(1), \quad \|\sqrt{m}(\hat{\gamma}_m - \gamma_0) - \tilde{\gamma}\| = o_p(1), \quad \|\hat{\zeta}_1 - \tilde{\zeta}_1\| = o_p(1). \qquad \text{(A.18)}$$

where $\hat{\xi}_1 = C_m^{1/2}(\hat{\beta}_m - \beta_0)$, $\tilde{\xi}_1 = C_m^{1/2}\tilde{S}_1$ *with* $C_m = X^{*\prime}\Sigma^0 X^*$; $\tilde{\gamma} = [\sum_{i=1}^m V_i^{0\prime} D_{0i}^{-1} V_i^0/m]^{-1}\frac{1}{\sqrt{m}}\tilde{S}_2$; $\hat{\zeta}_1$ *and* $\tilde{\zeta}_1$ *are given by (A.2)and (A.8) respectively.*

Proof. First, we show that $\|\hat{\zeta}_1 - \tilde{\zeta}_1\| = o_p(1)$. Similar to the arguments of the proof of Theorem 1, we have that

$$\sup_{\|\zeta_1\|\leq L, \|\zeta_2\|\leq k_n^{1/2}} \|\Psi_1(\zeta_1,\zeta_2) - \Phi(\zeta_1,\zeta_2)\| = o_p(1), \quad \|\hat{\zeta}_1\| = O_p(1),$$

thus,

$$\|\tilde{\zeta}_1 - \hat{\zeta}_1\| = o_p(1).$$

Similarly, $\|\hat{\xi}_1 - \tilde{\xi}_1\| = o_p(1)$. To prove $\|\sqrt{m}(\hat{\gamma}_m - \gamma_0) - \tilde{\gamma}\| = o_p(1)$, it suffices to prove this claim when all D_i are known to be D_{0i}. Let $\tilde{B}_m = \sum_{i=1}^m V_i' D_{0i}^{-1} V_i/m$. By (4) we have

$$\sqrt{m}(\hat{\gamma}_m - \gamma_0) = \tilde{B}_m^{-1} \frac{1}{\sqrt{m}} \sum_{i=1}^m V_i' D_{0i}^{-1}(r_i - \hat{r}_i).$$

Obviously, $\tilde{B}_m$ can be written as follows by Kronecker product,

$$\begin{aligned}\tilde{B}_m &= \frac{1}{m}\sum_{i=1}^{m}(r'_{0i}\otimes I_q+(\nabla\mu_i)'\otimes I_q)\Omega'_i D_{0i}^{-1}\Omega_i(r_{0i}\otimes I_q+\nabla\mu_i\otimes I_q)\\ &= \frac{1}{m}\sum_{i=1}^{m}(r'_{0i}\otimes I_q)\Omega'_i D_{0i}^{-1}\Omega_i(r_{0i}\otimes I_q)+\frac{1}{m}\sum_{i=1}^{m}(r'_{0i}\otimes I_q)\Omega'_i D_{0i}^{-1}\Omega_i(\nabla\mu_i\otimes I_q)\\ &+ \frac{1}{m}\sum_{i=1}^{m}(\nabla\mu_i)'\otimes I_q)\Omega'_i D_{0i}^{-1}\Omega_i(r_{0i}\otimes I_q)+\frac{1}{m}\sum_{i=1}^{m}((\nabla\mu_i)'\otimes I_q)\Omega'_i D_{0i}^{-1}\Omega_i(\nabla\mu_i\otimes I_q)\\ &=: \frac{1}{m}\sum_{i=1}^{m}(r'_{0i}\otimes I_q)\Omega'_i D_{0i}^{-1}\Omega_i(r_{0i}\otimes I_q)+J_1+J_2+J_3, \qquad \text{(A.19)}\end{aligned}$$

where Ω_i is a lower triangular matrix with 0's on its diagonal and the j row is $(\omega'_{ij1},\cdots,\omega'_{ij(j-1)},0,\cdots,0)$. Since for any $a\in R^q$ satisfying $\|a\|=1$, by the proof of theorem 1, we have

$$\begin{aligned}&(Ea'J_1a)^2=\left\{\frac{1}{m}E\sum_{i=1}^{m}a'(r'_{0i}\otimes I_q)\Omega'_i D_{0i}^{-1}\Omega_i(\nabla\mu_i\otimes I_q)a\right\}^2\\ =\ & \frac{1}{m}\sum_{i=1}^{m}Ea'(r'_{0i}\otimes I_q)\Omega'_i D_{0i}^{-1}\Omega_i(r_{0i}\otimes I_q)a\cdot\frac{1}{m}\sum_{i=1}^{m}Ea'((\nabla\mu_i)'\otimes I_q)\Omega'_i D_{0i}^{-1}\Omega_i(\nabla\mu_i\otimes I_q)a\\ \leq\ & C\frac{1}{m}\sum_{i=1}^{m}\|\nabla\mu_i\|^2=o(n^{-(2s-1)/(2s+1)}).\end{aligned}$$

That is $J_1\to 0$ in probability. Similarly we can show that $J_2\to 0$ and that $J_3\to 0$ in probability. Thus $B_m-\frac{1}{m}\sum_{i=1}^m V_i^{0\prime}D_{0i}^{-1}V_i^0\to 0$ in probability. Similarly, we can prove that $\frac{1}{\sqrt{m}}\sum_{i=1}^m V_i'D_{0i}^{-1}(r_i-\hat{r}_i)-\frac{1}{\sqrt{m}}\tilde{S}_2\to 0$ in probability. The proof is completed by an application of Slutsky theorem.

Proof of Theorem 2

By Lemma 2, we only need to show the asymptotic normality of $(\tilde{\xi}'_1,\tilde{\gamma}',\tilde{\zeta}'_1)'/\sqrt{m}$. This is equivalent to the asymptotic normality of $(\tilde{S}'_1,\tilde{S}'_2,\tilde{S}'_3)'/\sqrt{m}$. Note that Conditions (A1), (A3), (A4) and (A6) imply that

$$E_0\left[\psi'\left\{X_i^{*\prime}\Delta_{0i}\Sigma_{0i}^{-1}(y_i-\mu_{0i})\right\}+\omega'\left\{V_{0i}D_{0i}^{-1}(r_{0i}-\hat{r}_{0i})\right\}+\varphi'\left(Z_i^{*\prime}D_{0i}W_{0i}^{-1}(\varepsilon_{0i}^2-\sigma_{0i}^2)\right\}\right]^3<\kappa,$$

for any $\psi \in \mathbb{R}^{p+K}$, $\omega \in \mathbb{R}^q$ and $\varphi \in \mathbb{R}^{d+K'}$, where κ is a constant independent of i. Furthermore, we have

$$\frac{1}{m}\sum_{i=1}^{m} Var\left[\psi'\left\{X_i^{*\prime}\Delta_{0i}\Sigma_{0i}^{-1}(y_i-\mu_{0i})\right\}+\omega'\left\{V_{0i}D_{0i}^{-1}(r_{0i}-\hat{r}_{0i})\right\}+\varphi'\left(Z_i^{*\prime}D_{0i}W_{0i}^{-1}(\varepsilon_{0i}^2-\sigma_{0i}^2)\right\}\right]$$

$$=(\psi',\omega',\varphi')\frac{1}{n}\Delta_n(\psi',\omega',\varphi')' \to (\psi',\omega',\varphi')\Delta(\psi',\omega',\varphi')' > 0.$$

Therefore the asymptotic normality of $(\tilde{S}_1',\tilde{S}_2',\tilde{S}_3')'/\sqrt{m}$ is easily proved by multivariate Liapounov central limit theorem. Therefore,

$$\sqrt{m}\begin{pmatrix}\hat{\beta}_m-\beta_0\\ \hat{\gamma}_m-\gamma_0\\ \hat{\lambda}_m-\lambda_0\end{pmatrix}=\begin{pmatrix}(A_m/m)^{-1} & 0 & 0\\ 0 & (B_m/m)^{-1} & 0\\ 0 & 0 & (C_m/m)^{-1}\end{pmatrix}\begin{pmatrix}\tilde{S}_1/\sqrt{m}\\ \tilde{S}_2/\sqrt{m}\\ \tilde{S}_3/\sqrt{m}\end{pmatrix}$$

$$\to N\left\{0,\begin{pmatrix}\delta^{11} & 0 & 0\\ 0 & \delta^{22} & 0\\ 0 & 0 & \delta^{33}\end{pmatrix}^{-1}\begin{pmatrix}\delta^{11} & \delta^{12} & \delta^{13}\\ \delta^{21} & \delta^{22} & \delta^{23}\\ \delta^{31} & \delta^{32} & \delta^{33}\end{pmatrix}\begin{pmatrix}\delta^{11} & 0 & 0\\ 0 & \delta^{22} & 0\\ 0 & 0 & \delta^{33}\end{pmatrix}^{-1}\right\}.$$

in distribution as $m \to \infty$. The proof of Theorem 2 is completed.

References

Diggle, P. J., Heagerty P. J., Liang, K. Y., Zeger, S. L. (2002) Analysis of Longitudinal Data. Oxford University Press.

Diggle, P. J. and Verbyla, A. P. (1998). Nonparametric estimation of covariance structure in longitudinal data. *Biometrics*, 54, 403–15.

Fan, J., Huang, T. and Li, R. Z. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function. *Journal of American Statistical Association*, 35, 632–641.

Fan, J. and Wu, Y. (2008). Semiparametric estimation of covariance matrices for longitudinal data. *Journal of American Statistical Association*, to appear.

Härdle, W., Liang, H. and Gao, J. (2000). Partially Linear Models. Springer -Verlag, Germany.

He, X., Zhu, Z. Y. and Fung, W. K. (2002). Estimating in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika*, 89, 579–590.

He, X., Fung, W.K., and Zhu, Z.Y. (2005). Robust estimation in generalized partial linear models for clustered data. *Journal of the American Statistical Association*, 472, 1176–1184.

He, X., and Shi, P. (1996). Bivariate tensor-product B-Splines in a partly linear model, *Journal of Multivariate Analysis*, 58, 162–181.

Heckman, N. E. (1986). Spline smoothing in a partly linear model, *Journal of the Royal Statistical Society, Ser. B*, 48, 244–248.

Liang, K. Y., Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13–22.

Lin, X., and Carroll, R. J. (2001). Semiparametric regression for clustered data using generalized estimating equations, *Journal of the American Statistical Association*, 96, 1045–1056.

Lin, X. and Carroll, R. J. (2006). Semiparametric estimation in general repeated measures problems, *Journal of the Royal Statistical Society, Series B*, 68, 69-88.

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

Prentice, R. L. and Zhao, L. P. (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics*, 47, 825–839.

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. *Biometrika*, 86, 677–90.

Pourahmadi, M. (2000). Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika*, 87, 425–35.

Schumaker, L. L. (1981). Spline Functions, New York, Wiley.

Stone, C. (1985). Additive regression and other nonparametric models. *The Annals of Statistics*, 13, 689–705.

Wang, N. (2003). marginal nonparametric kernel regression accounting within-subject correlation. *Biometrika*, 90, 29–42.

Wang, N., Carroll, R. J. and Lin, X. (2005). Efficient semiparametric marginal estimation for longitudinal/clustered data. *Journal of the American Statistical Association*, 100, 147-157.

Wang Y. G. and Carey, V. (2003). Working correlation structure misspecification, estimation and covariate design: Implications for generalised estimating equations performance. *Biometrika*, 90, 29–41.

Welsh, A. H., Lin, X., and Carroll, R. J. (2002). Marginal longitudinal nonparametric regression: locality and efficiency of spline and kernel methods. *Journal of the American Statistical Association*, 97, 482–493.

Wu, H. and Zhang, J. (2006). Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modeling Approaches, John Wiley and Sons, New York.

Wu, W. and Pourahmadi, M. (2003). Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika*, 831–844

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

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

Table 1: CD4 cell data. The estimates of parameters based on AR(1) structure and square root CD4 cell numbers, with standard errors in parentheses.

	Cholesky ($\delta = 0.2$)	Generalized Estamating Equations		
	Normal	Independence	AR(1)	Exchangable
β_1	0.005(0.030)	0.015(0.035)	0.016(0.034)	0.002(0.032)
β_2	0.768(0.130)	0.981(0.184)	0.262(0.190)	0.596(0.136)
β_3	0.821(0.345)	1.075(0.528)	0.471(0.350)	0.494(0.358)
β_4	0.044(0.038)	-0.064(0.059)	0.050(0.041)	0.060(0.043)
β_5	-0.030(0.014)	-0.031(0.021)	-0.046(0.014)	-0.048(0.015)

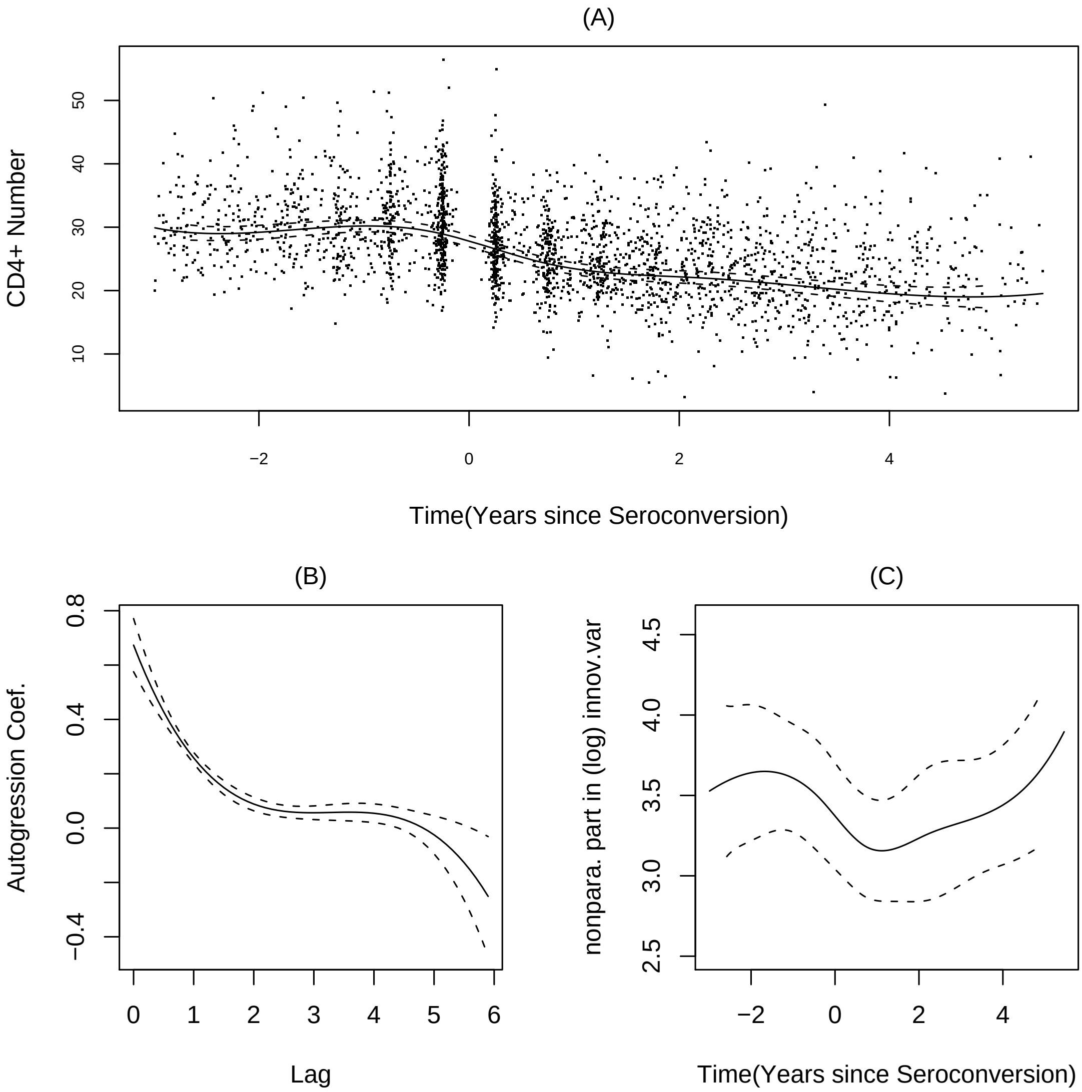

Figure 1: The CD4 cell data. The fitted curves of (A) nonparametric part in mean against time, (B) the generalized autoregressive parameters against lag and (C) the nonparametric part in (log) innovation variances against time based on AR(1) structure with $\delta = 0.2$ and square root CD+ cell numbers. Dashed curves represent asymptotic 95% confidence intervals.

Table 2: Simulation results for Study 1 over 100 replications with standard errors in parentheses.

	True	$\delta = 0$	$\delta = 0.2$	$\delta = 0.5$	$\delta = 0.8$
β_1	1.0	1.00	1.00	1.00	1.00
		(2.27×10^{-3})	(2.27×10^{-3})	(2.27×10^{-3})	(2.27×10^{-3})
β_2	0.5	0.51	0.51	0.51	0.51
		(4.83×10^{-3})	(4.83×10^{-3})	(4.83×10^{-3})	(4.84×10^{-3})
γ_1	0.2	0.20	0.20	0.20	0.20
		(7.68×10^{-4})	(7.68×10^{-4})	(7.69×10^{-4})	(7.71×10^{-4})
γ_2	1.0	1.0	1.00	1.00	1.00
		(2.85×10^{-3})	(2.86×10^{-3})	(2.86×10^{-3})	(2.86×10^{-3})
λ_1	-0.5	-0.50	-0.50	-0.50	-0.50
		(1.30×10^{-2})	(1.28×10^{-2})	(1.23×10^{-2})	(1.18×10^{-2})
λ_2	1.5	1.51	1.51	1.51	1.51
		(1.72×10^{-2})	(1.71×10^{-2})	(1.76×10^{-2})	(1.78×10^{-2})
$MSE(\hat{f}_0)$		0.0457	0.0457	0.0459	0.0462
		(0.0641)	(0.0646)	(0.0654)	(0.0666)
$MSE(\hat{f}_1)$		0.0119	0.0126	0.0150	0.0218
		(0.0080)	(0.0089)	(0.0115)	(0.0175)
β_1	1.0	1.00	1.00	1.00	1.00
		(2.98×10^{-3})	(2.98×10^{-3})	(2.98×10^{-3})	(2.99×10^{-3})
β_2	0	1.42×10^{-3}	1.42×10^{-3}	1.54×10^{-3}	1.95×10^{-3}
		(5.98×10^{-3})	(5.98×10^{-3})	(5.99×10^{-3})	(6.0×10^{-3})
γ_1	0.2	0.20	0.20	0.20	0.20
		(1.79×10^{-3})	(1.79×10^{-3})	(1.79×10^{-3})	(1.79×10^{-3})
γ_2	0	8.45×10^{-3}	8.43×10^{-3}	8.42×10^{-3}	8.46×10^{-3}
		(4.74×10^{-3})	(4.75×10^{-3})	(4.75×10^{-3})	(4.76×10^{-3})
λ_1	-0.5	-0.51	-0.50	-0.50	-0.50
		(2.03×10^{-2})	(2.02×10^{-2})	(1.96×10^{-2})	(1.83×10^{-2})
λ_2	0	7.38×10^{-3}	8.49×10^{-3}	9.57×10^{-3}	9.57×10^{-3}
		(1.70×10^{-2})	(1.71×10^{-2})	(1.77×10^{-2})	(1.81×10^{-2})
$MSE(\hat{f}_0)$		0.0187	0.0188	0.0189	0.0191
		(0.0228)	(0.0231)	(0.0233)	(0.0238)
$MSE(\hat{f}_1)$		0.0105	0.0103	0.0111	0.0146
		(0.0081)	(0.0076)	(0.0078)	(0.0105)

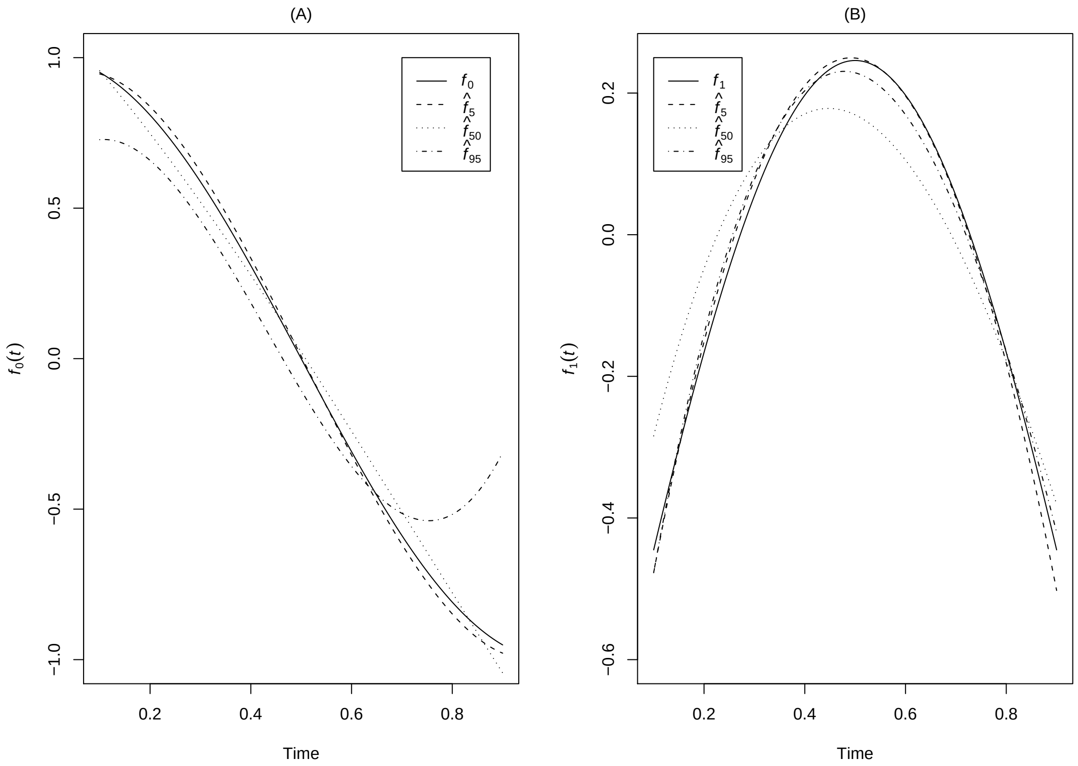

Figure 2: Nonparametric function f_0 and f_1 and their fitted curves $\hat{f}_5, \hat{f}_{50}, \hat{f}_{95}$, for AR(1) structure with $\delta = 0.2$.

Table 3: Assessment of the standard errors using formula (14).

		$\delta = 0$	$\delta = 0.2$	$\delta = 0.5$	$\delta = 0.8$
β_1	SD	0.0240	0.0240	0.0241	0.0241
	SE(Std)	0.0230(0.0030)	0.0230(0.0030)	0.0230(0.0030)	0.0230(0.0030)
β_2	SD	0.0518	0.0519	0.0519	0.0522
	SE(Std)	0.0481(0.0056)	0.0481(0.0056)	0.0481(0.0056)	0.0482(0.0057)

Table 4: Simulation results for Study 3 with standard errors in parentheses.

		Cholesky ($\delta = 0.2$)	Generalized Estamating Equations		
	True	Normal	Independence	Exchangeable	AR(1)
β_1	1.0	1.00	1.03	1.05	1.01
		(2.27×10^{-3})	(4.09×10^{-2})	(3.41×10^{-2})	(1.29×10^{-2})
β_2	0.5	0.51	0.49	0.54	0.54
		(4.83×10^{-3})	(8.29×10^{-2})	(6.53×10^{-2})	(2.55×10^{-2})
$MSE(f_0)$		0.0457(0.0646)	2.3851(3.8551)	2.1978(3.6879)	1.8847(3.5348)