Computational Statistics and Data Analysis 55 (2011) 3344–3354

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Efficient semiparametric estimation via Cholesky decomposition for longitudinal data

Ziqi Chen[a], Ning-Zhong Shi[a], Wei Gao[a,*], Man-Lai Tang[b]

[a] Key Laboratory for Applied Statistics of MOE, School of Mathematics and Statistics, Northeast Normal University, Changchun, 130024, China
[b] Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong

ARTICLE INFO

Article history:
Received 26 October 2010
Received in revised form 29 March 2011
Accepted 20 June 2011
Available online 26 June 2011

Keywords:
Efficient semiparametric estimation
Longitudinal data
Modified Cholesky decomposition
Profile likelihood estimator
Within-subject correlation

ABSTRACT

Semiparametric methods for longitudinal data with dependence within subjects have recently received considerable attention. Existing approaches that focus on modeling the mean structure require a correct specification of the covariance structure as misspecified covariance structures may lead to inefficient or biased mean parameter estimates. Besides, computation and estimation problems arise when the repeated measurements are taken at irregular and possibly subject-specific time points, the dimension of the covariance matrix is large, and the positive definiteness of the covariance matrix is required. In this article, we propose a profile kernel approach based on semiparametric partially linear regression models for the mean and model covariance structures simultaneously, motivated by the modified Cholesky decomposition. We also study the large-sample properties of the parameter estimates. The proposed method is evaluated through simulation and applied to a real dataset. Both theoretical and empirical results indicate that properly taking into account the within-subject correlation among the responses using our method can substantially improve efficiency.

1. Introduction

In longitudinal studies, repeated measurements are made on subjects over time and responses from a subject are very likely to be correlated (Liang and Zeger, 1986; Diggle et al., 2002). Semiparametric methods for analyzing longitudinal data with dependence within subjects have received a lot of attention in recent years (see Lin and Carroll, 2001, 2006; Wang et al., 2005; Fan et al., 2007; Fan and Wu, 2008; Lombardia and Sperlich, 2008). When modeling longitudinal data semiparametrically, the within-subject correlation must be taken into account; otherwise, the estimators for both parametric and nonparametric components may be inefficient (see Daniels and Zhao, 2003; Li et al., 2009) or biased (when missing values present, see Wang and Carey, 2003). Therefore, the proper estimation of covariance matrices could feature prominently in forecasting the trajectory of an individual's responses over time and substantially improving the efficiency for parameter estimates (Fan et al., 2007).

Wu and Pourahmadi (2003) proposed a nonparametric method for estimating covariance matrix. However, their proposed method could not deal with irregular observed measurements. To address the irregular time points issue, Fan et al. (2007) considered semiparametric varying-coefficient partially linear models with a parametric correlation structure while a nonparametric variance function was allowed. Fan and Wu (2008) proposed a difference-based technique to estimate the parametric regression coefficients and provided a robust estimator for the variance function. Consistency and asymptotic normality for the quasi-maximum likelihood estimators of the parameters in the correlation function were developed.

* Corresponding author. Tel.: +86 431 5099589.
E-mail address: gaow@nenu.edu.cn (W. Gao).

doi:10.1016/j.csda.2011.06.025

Unfortunately, the estimates of covariance matrices obtained by both methods cannot be guaranteed to be positive definite. Moreover, the estimation of mean and covariance structure has to be done in a 2-step fashion. By parametrically modeling means and covariances, Ye and Pan (2006) proposed the generalized estimating equation (GEE) method to simultaneously estimate both the mean regression coefficients and covariance structure parameters. Motivated by the modified Cholesky decomposition, Leng et al. (2010) extended the results to the partially linear model and the GEE method based on regression spline was developed. One disadvantage of this approach is the knots selection in B-spline smoothing, which could lead to tedious computational time. Based on the conditional likelihood, the general profile kernel method proposed by Lin and Carroll (2006) can get efficient estimates; however, the computation could be complicated.

Based on kernel regressions, a new approach for estimating the mean and covariance structure parameters is proposed in this paper. The proposed method can handle irregular and possibly subject-specific time points data, and the estimated covariance matrices are positive definite. For the model given in the paper, our method is more computationally efficient than the general profile likelihood approach proposed by Lin and Carroll (2006). We show in theory that our regression coefficient estimates for the mean and covariance structure are semiparametrically efficient, i.e., their asymptotic variances achieve the semiparametric efficiency bounds. Furthermore, our approach produces more accurate covariance matrix estimate than the method of Leng et al. (2010) and is robust to misspecification of the covariance structure.

We organize our article as follows. In Section 2, we consider the semiparametric partially linear model for the mean and model covariance structures simultaneously, motivated by the modified Cholesky decomposition. We describe the profile likelihood approach for estimating the parameters. The asymptotic properties of the estimators are given in Section 3. In Section 4, simulation studies are conducted to evaluate the performance of the proposed method. We analyze a longitudinal dataset of CD4 cell count of human immunodeficiency virus seroconverters to illustrate our methodology in Section 5. A brief discussion is presented in Section 6. Technical proofs and a detailed algorithm are relegated to the Appendix.

2. Modeling and estimation procedures

Suppose that a random sample of n subject is coming from the following semiparametric partially linear model

$$y_{ij} = z_{ij}'\beta + m(t_{ij}) + \epsilon_{ij}, \tag{1}$$

where y_{ij} is a response for ith subject at time t_{ij}, z_{ij} is a p-dimensional covariate for $j = 1, \ldots, J_i$ and $i = 1, \ldots, n$. Here, β is a $p \times 1$ vector of regression coefficient associated with z_{ij}, and $m(\cdot)$ is an unknown smooth function of t_{ij}. In this model, the mean response is linearly related to z_{ij}, while its relation with t_{ij} is not specified up to any finite number of parameters. This model combines the flexibility of nonparametric regression and parsimony of linear regression. When the relation between y_{ij} and z_{ij} is of main interest and can be approximated by a linear function, it offers more interpretability than a purely nonparametric model. Let $Z_i = (z_{i1}, \ldots, z_{iJ_i})'$, and define T_i and ϵ_i similarly.

Another important feature of the semiparametric partially linear model in (1) is that it allows different subjects to be observed at different time points (i.e., t_{ij} with $j = 1, \ldots, J_i$). Here, we assume that ϵ_i is normally distributed with mean zero and covariance matrix Σ_i given the covariates, observed times and J_i. Lin and Carroll (2001) demonstrated that in order to reach the semiparametric information bound, the within-subject correlation must be properly taken into account in both the parametric and nonparametric estimation procedures. Unfortunately, even when the true correlation structure is used, the standard kernel estimator of the nonparametric function proposed by Lin and Carroll (2001) will not be a suitable choice. In fact, unless either the 'working independence' assumption or the undersmoothing step is employed, the parameter estimators could be $\sqrt{n}$-inconsistent (see also Lin and Carroll, 2001). Hence, the way that the within-subject correlation being properly incorporated into both parametric and nonparametric estimations becomes another challenge to the semiparametric partially linear model in (1) for longitudinal data analysis. To overcome the above problems, we model the covariance matrix of Σ_i's in (1) via the modified Cholesky decomposition approach considered in Pourahmadi (1999). Briefly, for each Σ_i, there exist a unique lower triangular matrix P_i with 1's as diagonal entries and $-\phi_{ijl}$ as the (j, l)th element and a unique diagonal matrix $D_i = \mathrm{diag}\{\sigma_{i1}^2, \ldots, \sigma_{iJ_i}^2\}$ with $\sigma_{ij}^2 > 0$ such that

$$P_i \Sigma_i P_i' = D_i. \tag{2}$$

This diagonalization allows modeling P_i and D_i instead of the covariance matrix Σ_i. Two immediate advantages of (2) are noteworthy. First, it guarantees the positive definiteness of Σ_i. That is, if we can find estimates $\hat{P}_i$ and $\hat{D}_i$ of P_i and D_i, respectively, then an estimator of Σ_i can be simply obtained as $\hat{\Sigma}_i = \hat{P}_i^{-1}\hat{D}_i\hat{P}_i'^{-1}$, which is positive definite. Second, it is possible to model P_i and D_i since their nonredundant entries have statistical interpretation. That is, the subdiagonal entries of P_i are the regression coefficients when each ϵ_{ij} $(j = 1, \ldots, J_i)$ is regressed on its predecessors $\epsilon_{i,j-1}, \ldots, \epsilon_{i,1}$ and the entries of D_i are the corresponding prediction error variances. More precisely, for $j = 2, \ldots, J_i$,

$$\epsilon_{ij} = \sum_{l=1}^{j-1} \phi_{ijl}\epsilon_{il} + \delta_{ij},$$

where $-\phi_{ijl}$ is the (j, l)th element (for $l < j$) of P_i, and $\sigma_{ij}^2 = \mathrm{var}(\delta_{ij})$ is the jth diagonal element of D_i. Here, ϕ_{ijl}'s and σ_{ij}^2's are referred as generalized autoregressive parameters (GARPs) and innovation variances (IVs), respectively. As a result, modeling the covariance matrix, through P_i and D_i, is equivalent to modeling a sequence of regressions. In this article, we model the GARP/IV parameters using the following linear and log link functions

$$\phi_{ijl} = w_{ijl}'\gamma, \qquad \log\sigma_{ij}^2 = u_{ij}'\lambda, \tag{3}$$

where w_{ijl} and u_{ij} are $q \times 1$ and $d \times 1$ vectors of covariates, respectively. These design vectors are used to model the GARP/IV parameters as functions of subject-specific covariates and/or to model structure on the ϕ_{ijl} and σ_{ij}^2 within a subject. For some examples of structured GARP/IV models, see Pourahmadi and Daniels (2002).

Under ϵ_i being normally distributed with mean zero and covariance matrix Σ_i given covariates z_{ij}, observed times t_{ij} and J_i, model (1) can be rewritten, via (2), as follows

$$y_{ij} = z_{ij}'\beta + m(t_{ij}) + \sum_{l=1}^{j-1}(y_{il} - z_{il}'\beta - m(t_{il}))\phi_{ijl} + \delta_{ij} \tag{4}$$

where $(\delta_{i1}, \dots, \delta_{iJ_i})'$ is normally distributed with mean zero and covariance $D_i = \mathrm{diag}\{\sigma_{i1}^2, \dots, \sigma_{iJ_i}^2\}$. It is noted that (4) is an autoregression-type equation of the longitudinal observations on each individual.

Here, our main focus is to improve efficiency of the estimates of β, γ and λ given in (1) and (3). The smooth function $m(\cdot)$ given in (1) can also be estimated. For any given symmetric standard kernel function $K(t)$, we propose the following so-called Cholesky decomposition based semiparametric estimation (CDSE) procedure to estimate the parameters. This procedure consists of two sub-procedures, namely Procedures A and B. Briefly, Procedure A provides initial/starting values for parameters β and $m(\cdot)$ while Procedure B obtains efficient semiparametric estimators for all parameters (i.e., β, γ, λ via the profile likelihood estimation approach based on (4)).

Procedure A (*Initial Estimators of β and $m(\cdot)$*).

A1. For given β, let $\hat{m}_I(t;\beta) = \{\sum_{i=1}^n\sum_{j=1}^{J_i} K_h(t_{ij}-t)(y_{ij} - z_{ij}'\beta)\}/\{\sum_{i=1}^n\sum_{j=1}^{J_i} K_h(t_{ij}-t)\}$, where $K_h(t) = h^{-1}K(t/h)$ and h is any appropriate bandwidth.

A2. Let $S(\beta) = \sum_{i=1}^n\sum_{j=1}^{J_i}(y_{ij} - z_{ij}'\beta - \hat{m}_I(t_{ij},\beta))^2 = \sum_{i=1}^n\sum_{j=1}^{J_i}(\hat{y}_{ij} - \hat{z}_{ij}'\beta)^2$, where $\hat{y}_{ij} = y_{ij} - \{\sum_{u=1}^n\sum_{v=1}^{J_i} K_h(t_{uv} - t_{ij})y_{uv}\}/\{\sum_{u=1}^n\sum_{v=1}^{J_i} K_h(t_{uv}-t_{ij})\}$, and $\hat{z}_{ij} = z_{ij} - \{\sum_{u=1}^n\sum_{v=1}^{J_i} K_h(t_{uv}-t_{ij})z_{uv}\}/\{\sum_{u=1}^n\sum_{v=1}^{J_i} K_h(t_{uv}-t_{ij})\}$.

An initial value for β is $\hat{\beta}_I = \arg\min_\beta S(\beta)$, which can readily be obtained by the ordinary least square regression method.

A3. An initial estimate of $m(\cdot)$ is then computed by $\hat{m}_I(t) = \hat{m}_I(t;\hat{\beta}_I)$.

Procedure B (*Efficient Semiparametric Estimators*). Let $r_{ij}(\beta, m) = y_{ij} - z_{ij}'\beta - m$.

B1. For any given β and γ in (1) and (3), let

$$\hat{m}(t;\beta,\gamma) = \frac{\sum_{i=1}^n\sum_{j=1}^{J_i} K_h(t_{ij}-t)\left\{y_{ij} - z_{ij}'\beta - \sum_{l=1}^{j-1} r_{il}(\hat{\beta}_I, \hat{m}_I(t_{il}))\phi_{ijl}\right\}}{\sum_{i=1}^n\sum_{j=1}^{J_i} K_h(t_{ij}-t)}.$$

B2. Obtain the profile likelihood estimators of β, γ, λ as follows

$$(\hat{\beta}', \hat{\gamma}', \hat{\lambda}')' = \arg\max_{\beta,\gamma,\lambda}\left\{-\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^{J_i}\log\sigma_{ij}^2 - \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^{J_i}\frac{\left\{r_{ij}(\beta, \hat{m}(t_{ij};\beta,\gamma)) - \sum_{l=1}^{j-1} r_{il}(\beta, \hat{m}(t_{il};\beta,\gamma))\phi_{ijl}\right\}^2}{\sigma_{ij}^2}\right\}, \tag{5}$$

where σ_{ij}^2 and ϕ_{ijl} are given in (3). The parameter estimates can be computed by the algorithm given in the Appendix.

B3. Improve nonparametric component by

$$\hat{m}(t) = \hat{m}(t;\hat{\beta},\hat{\gamma}).$$

$\hat{\beta}$, $\hat{\gamma}$ and $\hat{\lambda}$ obtained via our proposed CDSE approach are efficient semiparametric estimators for models (1) and (3).

We summarize our CDSE approach as follows. Firstly, ignoring the correlation among errors in (1), we obtain an initial estimator, which is essentially the 'working independence estimator' (Lin and Carroll, 2001, WI estimator), and is the most efficient conventional local kernel estimator. Secondly, in order to obtain the efficient estimators for regression coefficients in Procedure B, correlation has been incorporated into estimation by the autoregression-type equation (4) and initial estimators given in Procedure A.

Remark 1. The estimation procedure here is analogous to that proposed by Severini and Wong (1992) and Lombardia and Sperlich (2008).

Remark 2. The general profile likelihood approach proposed by Lin and Carroll (2006) can achieve efficiency for estimates of the parameters. However, getting solutions to both the kernel estimating equation (i.e., the equation for getting the nonparametric estimate) and the profile kernel estimating equation (i.e., the equation for obtaining the parameter estimates) requires intensive iterative steps. In practice, a 2-stratum loop (the outside loop is the iterative steps for getting solution to the profile kernel estimating equation, and the inside loop is the iterative steps for getting solution to the kernel estimating equation for the given parametric component) is needed to get the estimates of the parameters. Even worse, when the total number of observations $\sum J_i$ is large, $\sum J_i$ parallel inside loops must be accomplished in order to get $\hat{m}(t_{ij}; \beta, \gamma, \lambda)$ $(j = 1, \ldots, J_i;\ i = 1, \ldots, n)$. Compared with the general method by Lin and Carroll (2006), the nonparametric estimate suggested by our method is explicit in step B1, which makes the solution to (5) simpler. Therefore, for the model given in this paper, our proposed method requires shorter computing time. Unlike Fan et al. (2007) and Fan and Wu (2008) s' two-step fashion methods (i.e., the estimation of covariance structure must adopt the estimate of the mean from the previous step), our proposed method can estimate the covariance structure and mean simultaneously as is indicated in step B2.

3. Asymptotic properties of the estimators

In this section, we investigate the large-sample properties of the estimators given in Section 2. For this purpose, we give the following regular conditions which are mild and can be found in Speckman (1988), Pepe and Couper (1997), Lin and Carroll (2001), Wang et al. (2005), You and Zhou (2006), and Fan et al. (2007).

Assumptions:

1. $J_1, \ldots, J_n$ are independently and identically distributed with $E(J_1) < \infty$, and $t_{i1}, \ldots, t_{iJ_i}$ are independently and identically sampled from a density $f(t)$ given J_i.

2. Given J_i and T_i,

$$z_{ijs_1} = g_{1s_1}(t_{ij}) + \eta^{(1)}_{ijs_1}, \qquad u_{ijs_2} = g_{2s_2}(t_{ij}) + \eta^{(2)}_{ijs_2}, \qquad w_{ijls_3} = g_{3s_3}(t_{ij}, t_{il}) + \eta^{(3)}_{ijls_3},$$

where $\{\eta^{(1)}_{ijs_1}\}$, $\{\eta^{(2)}_{ijs_2}\}$ and $\{\eta^{(3)}_{ijls_3}\}$ are mutually independent variables with means being zero, and are all independent of ϵ_i, for $l = 1, \ldots, j-1$; $j = 1, \ldots, J_i$; $i = 1, \ldots, n$; $s_1 = 1, \ldots, p$; $s_2 = 1, \ldots, d$ and $s_3 = 1, \ldots, q$. Moreover, $\sum_{s=1}^{p} \sup_{t\in\Omega} |g_{1s}(t)| < \infty$, $\sum_{s=1}^{d} \sup_{t\in\Omega} |g_{2s}(t)| < \infty$ and $\sum_{s=1}^{q} \sup_{(t_1,t_2)\in\Omega\times\Omega} |g_{3s}(t_1, t_2)| < \infty$.

3. $E(y_{ij}|z_{ij}, t_{ij}) = E(y_{ij}|z_{ij}, t_{ij}, (z_{il}, t_{il})_{l\neq j})$, which is called the PA condition.

Let $A_1 = E(Z_1 - E(Z_1 \mid T_1))'\Sigma_1^{-1}(Z_1 - E(Z_1 \mid T_1))$, $A_2 = E\{\sum_{j=1}^{J_1}\sum_{l=1}^{j-1}\sum_{k=1}^{j-1} \sigma_{1lk} w_{1jl} w'_{1jk}/\sigma^2_{1j}\}$, and $A_3 = EU_1'U_1/2$, where $E(Z_1 \mid T_1) = (E(z_{11} \mid t_{11}), \ldots, E(z_{1J_1} \mid t_{1J_1}))'$, $\sigma_{1jl} = \Sigma_1(j, l)$, and $U_1 = (u_{11}, \ldots, u_{1J_1})'$.

We present the main results in the following theorem and leave the proof in the Appendix.

Theorem 3.1. *Let β_0, γ_0, λ_0, and $m_0(t)$ be the true values of β, γ, λ, and $m(t)$, respectively. Suppose that the assumptions given above and the conditions in the Appendix hold. We have the following results:*

(a) *If the matrices A_1, A_2 and A_3 are positive definite and as $n \to \infty$, then*

$$\sqrt{n}\begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{\gamma} - \gamma_0 \\ \hat{\lambda} - \lambda_0 \end{pmatrix} \xrightarrow{L} N\Big(0, \operatorname{diag}(A_1, A_2, A_3)^{-1}\Big), \tag{6}$$

where n is the total number of subjects. Moreover, the estimators $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\lambda}$ are efficient semiparametric estimators of β_0, γ_0, and λ_0, respectively.

(b) *The nonparametric component estimator $\hat{m}(t)$ satisfies*

$$\begin{aligned}\hat{m}(t) - m_0(t) = {} & \{E(J_1)f(t)\}^{-1}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{J_i} K_h(t_{ij} - t)\left\{r_{ij}(\beta_0, m_0(t_{ij})) - \sum_{l=1}^{j-1} r_{il}(\beta_0, m_0(t_{il}))\phi_{ijl}\right\} \\ & + \frac{1}{2}h^2 C_2(K)\mathbf{A}(t) + o_p(c_n),\end{aligned} \tag{7}$$

where $C_2(K) = \int t^2 K(t)\mathrm{d}t$, $b(t) = m_0''(t) + 2f'(t)m_0'(t)/f(t)$, $\mathbf{A}(t) = b(t) + \{2E(J_1)\}^{-1}E\{(J_1 - 1)J_1\}E\{b(t_{1l_1})\phi_{1j_1l_1} \mid t_{1j_1} = t\}$ $(l_1 < j_1)$, and $c_n = \{\log(1/h)/(nh)\}^{1/2} + h^2$.

Remark 3. The semiparametric efficiency of $\hat{\beta}$, $\hat{\gamma}$, $\hat{\lambda}$ indicates that the estimation procedure proposed in this article is reasonable (see Chamberlain, 1987; Kitamura et al., 2004; Liang et al., 2004).

Remark 4. From the orthogonality between the mean regression coefficients and error parameters, and the orthogonality between γ and λ by the statistical interpretation of Cholesky decomposition of covariance matrices, the mutual asymptotic independence among $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\lambda}$ is natural and intuitive, which coincides with the results given in Pourahmadi (2000) and Ye and Pan (2006).

Remark 5. The first term of (7) dominates the asymptotic variance of $\hat{m}(t)$ while the second term dominates the asymptotic bias. We have

$$\hat{m}_I(t) - m_0(t) = \{E(J_1)f(t)\}^{-1}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{J_i}K_h(t_{ij}-t)r_{ij}(\beta_0, m_0(t_{ij})) + \frac{1}{2}h^2C_2(K)b(t) + o_p(c_n).$$

From the statistical interpretation of Cholesky decomposition of covariance matrices, we can conclude that our proposed $\hat{m}(t)$ uniformly outperforms the WI estimator $\hat{m}_I(t)$, if the asymptotic variance is adopted as the criterion. This can also be seen in the analysis of CD4 count dataset. The asymptotic bias of our proposed estimate differs from that of $\hat{m}_I(t)$.

4. Simulation study

In this section, we investigate the finite sample performance of the proposed estimators in Section 2. We generate 1000 datasets, each consisting of $n = 150$ subjects from models given in (1) and (3). We sample the time points following the method described in Fan et al. (2007). That is, each individual has a set of scheduled time points, $\{0, 1, 2, 3, 4, 5, 6, 7, 8\}$, and each scheduled time, except time 0 and 1, has a 20% probability of being passed. The actual observation time is a random perturbation of a scheduled time: a uniform $[-0.5, 0.5]$ random variable is added to a non-skipped scheduled time. This results in different observed time points t_{ij} per subject, i.e., leading to subject-dependent or irregular time points.

Let $z_{ij} = (z_{ij1}, z_{ij2})'$. We generate the covariates as follows: $z_{ij1} = \arctan(t_{ij}) + \xi_{ij}$, where ξ_{ij} follows the standard normal distribution, z_{ij2} follows a Bernoulli distribution with success probability 0.5, $u_{ij} = (1, \arctan(t_{ij}) + \tau_{ij})'$, where τ_{ij} follows the standard normal distribution, and $w_{ijl} = (1, \theta_{ijl})'$, where $\theta_{ijl} \sim N(0, 0.1)$, for $l = 1, 2, \ldots, j-1, j = 1, 2, \ldots, J_i$, and $i = 1, 2, \ldots, n$. In this simulation, we take $\beta = (1, 0.5)'$, $\lambda = (0.5, 0.5)'$, $\gamma_2 = 0.5$, and the nonparametric function $m(t) = \sin(2\pi t/8)$. We set $\gamma_1 = 0.2, 0.3$, and 0.5 to represent weakly, moderately, and highly correlated random errors within the same individual, respectively. Note that too strong a correlation between z_{ij} and t_{ij} may result in asymptotically biased estimates of β. In our simulation studies, we compare our proposed method with the WI method (Lin and Carroll, 2001) and the "spline" method proposed by Leng et al. (2010) (with $\log\sigma_{ij}^2$ being modeled in (3), and with knots selected following their method). We also estimate β using the true correlation structure (i.e., the parameter vector $(\gamma', \lambda')'$ is evaluated at its true value instead of its estimated value).

Table 1 summarizes the results for the estimates of β, based on 1000 simulated datasets. Here, "Bias" represents the sample average over 1000 estimates subtracting the true value of β, "SD" represents the sample standard deviation over 1000 estimates, "Median" represents the median of the 1000 estimates subtracting the true value, and "MAD" represents the median absolute deviation of the 1000 estimates divided by a factor of 0.6745 (see Fan et al., 2007). It can be seen that our proposed method is more efficient than the WI method. For instance, the relative efficiency of $\hat{\beta}_1$, $[\text{MAD(WI)}/\text{MAD(CDSE)}]^2$ of $\hat{\beta}_1$, is about 27.23 for highly correlated random error, 3.279 for moderately correlated error, and 1.942 for weakly correlated error, respectively (see, Table 2). Moreover, the MADs of our CDSE method are almost the same regardless of the strength of the correlation among random errors. In other words, the strength of the correlation among random errors does not affect the efficiency of our proposed method. On the other hand, the MAD of the WI method changes greatly, i.e., MAD increases with the strength of correlation. When the true correlation structure is used, we observe that the relative efficiency $[\text{MAD(CDSE)}/\text{MAD(True)}]^2$ is nearly identical to 1 (see, Table 2), regardless of the strength of the correlation among random errors, which indicates that the CDSE estimator is as efficient as the estimator when the true covariance matrix is known.

The proposed CDSE method and "spline" method possess almost the same efficiency for $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\lambda}$ (see, Tables 1 and 3). However, it is noteworthy that $\hat{\lambda}_{1\text{Spline}}$ yields larger bias (or median) than $\hat{\lambda}_{1\text{CDSE}}$ (see, Table 3). This may be due to the fact that we estimate covariance structure and mean parameters simultaneously as is indicated in step B2. As suggested by an anonymous reviewer, we can use the entropy loss function, $L(\Sigma, \hat{\Sigma}) = n^{-1}\sum_{i=1}^{n}\{\text{trace}(\Sigma_i\hat{\Sigma}_i^{-1}) - \log|\Sigma_i\hat{\Sigma}_i^{-1}| - J_i\}$, to compare accuracy in estimating the covariance matrix, where Σ_i is the true covariance matrix and $\hat{\Sigma}_i$ is its estimate. We summarize the results in Table 4. Our approach performs better than the "spline" method in terms of smaller SDs and entropy losses.

For the proposed CDSE method and the "spline" method (see, Table 3), their MAD's for $\hat{\gamma}$ become lower when the random errors are highly correlated while their MADs for $\hat{\lambda}$ are nearly the same regardless of the strength of the correlation among errors. This phenomenon is due to the fact that γ describes the strength of dependence among responses and thus more information for γ will be reflected in the stronger correlation situation while λ contains little information from the correlation among errors. Also, $\hat{\beta}$ becomes more efficient in highly correlated cases for both methods (see, Table 1).

Table 1
Performance of $\hat{\beta}$. (All values are multiplied by a factor of 1000.)

Methods	$\hat{\beta}_1$				$\hat{\beta}_2$			
	Bias	Median	SD	MAD	Bias	Median	SD	MAD
$\gamma_1 = 0.2$								
WI	−5.428	−1.665	62.159	62.242	1.233	−2.598	124.43	119.70
True	4.068	1.533	43.324	44.410	0.728	0.581	88.792	94.709
CDSE	4.052	1.119	43.416	44.659	0.800	0.445	88.844	93.132
Spline	−1.005	−3.230	43.333	44.132	1.332	3.249	88.926	94.332
$\gamma_1 = 0.3$								
WI	4.700	3.422	78.092	76.542	−15.87	−11.43	155.22	153.84
True	3.788	3.193	41.741	42.437	−1.593	1.703	82.107	82.051
CDSE	3.712	3.455	41.841	42.270	−1.735	3.356	82.095	82.319
Spline	−2.289	1.773	42.404	42.402	−1.943	1.996	82.188	81.989
$\gamma_1 = 0.5$								
WI	−3.428	−3.418	183.69	185.14	6.294	1.627	350.71	350.04
True	3.081	2.228	35.421	35.054	−1.580	1.930	71.940	72.955
CDSE	3.063	2.108	35.418	35.480	−1.989	0.486	72.026	72.679
Spline	−1.005	−1.746	35.133	35.391	−2.103	−1.197	72.826	73.299

Table 2
The relative efficiency: $[\mathrm{MAD(WI)}/\mathrm{MAD(CDSE)}]^2$ (WI:CDSE) and $[\mathrm{MAD(CDSE)}/\mathrm{MAD(True)}]^2$ (CDSE:True).

	$\hat{\beta}_1$			$\hat{\beta}_2$		
	$\gamma_1 = 0.2$	$\gamma_1 = 0.3$	$\gamma_1 = 0.5$	$\gamma_1 = 0.2$	$\gamma_1 = 0.3$	$\gamma_1 = 0.5$
WI:CDSE	1.942	3.279	27.23	1.652	3.493	23.20
CDSE:True	1.011	0.992	1.024	0.967	1.007	0.993

Table 3
Performance of $(\hat{\gamma}', \hat{\lambda}')'$ estimated by the CDSE method and the Spline method (All values are multiplied by a factor of 1000.)

Parameters	CDSE method				Spline method			
	Bias	Median	SD	MAD	Bias	Median	SD	MAD
$\gamma_1 = 0.2$								
$\hat{\gamma}_1$	−0.060	0.280	10.578	10.553	−0.090	0.411	10.572	10.648
$\hat{\gamma}_2$	3.853	0.074	153.90	149.31	6.461	4.782	153.73	149.53
$\hat{\lambda}_1$	−9.545	−8.190	56.426	58.374	−20.54	−19.04	57.779	57.509
$\hat{\lambda}_2$	−3.156	−3.841	37.237	37.586	6.286	5.712	37.843	38.576
$\gamma_1 = 0.3$								
$\hat{\gamma}_1$	−0.056	0.134	8.0890	8.2217	−0.057	0.176	8.0614	8.0381
$\hat{\gamma}_2$	−5.423	−5.735	133.48	129.47	−3.250	−2.061	133.58	129.31
$\hat{\lambda}_1$	−6.379	−5.746	56.006	57.646	−17.36	−16.41	56.642	56.720
$\hat{\lambda}_2$	−5.737	−5.895	38.523	38.264	3.133	2.299	39.054	38.769
$\gamma_1 = 0.5$								
$\hat{\gamma}_1$	0.224	0.178	4.4040	4.4111	0.207	0.307	4.3442	4.2499
$\hat{\gamma}_2$	−1.052	2.471	88.675	86.876	1.856	2.577	87.801	86.616
$\hat{\lambda}_1$	−0.585	−1.001	59.312	57.664	−19.46	−20.78	59.098	57.349
$\hat{\lambda}_2$	−3.325	−3.651	38.296	37.663	6.067	6.802	38.891	40.561

Table 4
The means and the SDs (in parentheses) for $L(\Sigma, \hat{\Sigma})$. (All values are multiplied by a factor of 1000.)

Method	$\gamma_1 = 0.2$	$\gamma_1 = 0.3$	$\gamma_1 = 0.5$
CDSE	178.13(52.063)	209.09(54.542)	463.99(95.477)
Spline	179.52(52.549)	211.02(54.890)	469.79(99.027)

Next, we study the robustness of our proposed method. For this purpose, the simulated datasets are generated as before except that a disturbance $\Delta\Sigma_{iJ_i\times J_i} = \sigma^2[(1-\rho)I_{J_i\times J_i} + \rho J_{J_i\times J_i}]$ is added to Σ_i with $I_{J_i\times J_i}$ being the identity matrix and $J_{J_i\times J_i}$ being a matrix with all elements equal to 1. Let $\rho = 0.25$, $\sigma^2 = 0.2$, and Y_i is generated from $N(Z_i\beta + m(T_i), \Sigma_i^* = \Sigma_i + \Delta\Sigma_i)$ for $i = 1, 2, \ldots, n$. We summarize the results in Table 5. We can see that our CDSE estimates of β here are much more efficient than the non-disturbance-case WI estimates in Table 1. Moreover, the efficiency of our CDSE estimates changes little due to this disturbance. It is noted that the "spline" method and our proposed CDSE method demonstrate similar robust behavior in this covariance structure disturbance investigation.

Table 5
Impacts on $\hat{\beta}$ when response data are generated from disturbed covariance matrices. (All values are multiplied by a factor of 1000.)

Methods	$\hat{\beta}_1$				$\hat{\beta}_2$			
	Bias	Median	SD	MAD	Bias	Median	SD	MAD
$\gamma_1 = 0.2$								
CDSE	4.080	5.457	44.223	39.857	−1.653	0.863	92.141	90.047
Spline	2.005	1.705	44.299	40.958	−1.944	−2.003	91.984	90.633
$\gamma_1 = 0.3$								
CDSE	1.376	2.417	45.910	45.964	2.540	7.569	88.071	88.658
Spline	−2.729	−2.413	46.065	44.936	2.518	8.233	87.929	88.536
$\gamma_1 = 0.5$								
CDSE	3.719	5.083	39.335	40.307	−5.011	−6.319	78.166	75.557
Spline	3.227	1.512	39.062	40.233	−5.103	−5.675	77.201	75.624

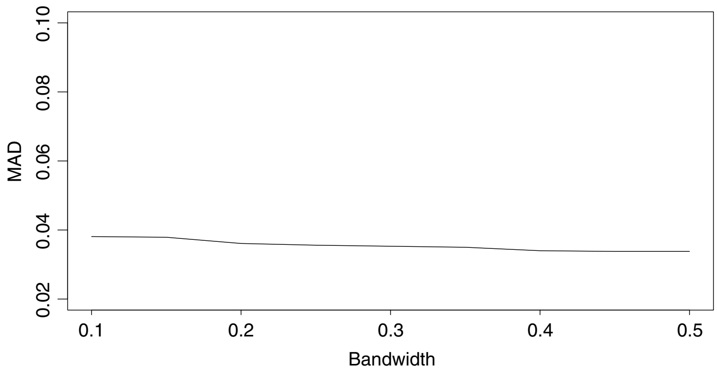

Fig. 1. MADs of the CDSE estimates of β_1 based on 1000 datasets when the bandwidth varies from 0.1 to 0.5, here $\gamma_1 = 0.5$.

Bandwidth selection

In this article, we focus on statistical inference for the parameters $(\beta', \gamma', \lambda')'$ and are interested in the efficiency of the parameter estimates. We find that bandwidth selection is less critical for the parameter estimates than for the estimate of $m(t)$. For example, for $\gamma_1 = 0.5$, when the bandwidth parameter varies from 0.1 to 0.5, the change of MAD for $\hat{\beta}_1$ is negligible (see, Fig. 1). For this reason and for the sake of simplicity, we fix the bandwidth parameter at 0.4 in our simulation study; however, this particular bandwidth choice does not represent the optimal choice. It is noted that for high frequency functions such as $m(t) = \sin(2\pi t)$, the bandwidth parameter for our proposed CDSE method is suggested to be less than 0.4 while the number of knots for the "spline" method proposed by Leng et al. (2010) must be larger than the integer part of $M^{1/5}$, where M is the number of distinct values in $\{t_{ij}, j = 1, \ldots, J_i; i = 1, \ldots, n\}$ in order that the parameter estimates for both methods can achieve the expected efficiency. For the "spline" method, it is tedious and thus time consuming to determine the number of knots, compute the internal knots and select the locations of knots. However, it is relatively simpler and thus requires less computations to determine bandwidth parameter for our proposed CDSE method.

5. Application to the longitudinal CD4 cell count data

We apply the semiparametric model given in (1) to the longitudinal CD4 cell count data among HIV seroconverters (Zeger and Diggle, 1994). For comparison purpose, we analyze this dataset by the WI method (Lin and Carroll, 2001), our proposed CDSE method, and the "spline" method. Following the model adopted in Leng et al. (2010), the number of internal knots is taken to be 7, which is also the optimal number of knots according to the leave-one-subject-out cross-validation. This study involved 369 subjects whose CD4 counts were measured during a period ranging from 3 years before to 6 years after seroconversion. A total of 2376 CD4 measurements were available, and the number of CD4 observations per subject varied from 1 to 12, with most of the subjects having between 4 and 10 observations. It is of interest to estimate the average time course of CD4 counts and the effects of other covariates including age (z_{ij1}), smoking status measured by packs of cigarettes (z_{ij2}), drug use (yes, 1; no, 0) (z_{ij3}), number of sex partners (z_{ij4}) and depression status measured by the CESD Scale (the Center for Epidemiologic Studies Depression Scale, large values indicating more depression symptoms) (z_{ij5}). Let t_{ij} be the year since seroconversion. We apply the Cholesky decomposition of the covariance matrix of measures coming from each subject. Here, we let w_{ijl} (in (3)) to be $(1, t_{ij} - t_{il}, (t_{ij} - t_{il})^2, (t_{ij} - t_{il})^3)'$ following the arguments in Ye and Pan (2006), and u_{ij} (in (3)) to be $(z_{ij1}, z_{ij2}, z_{ij3}, z_{ij4}, z_{ij5}, 1, t_{ij}, t_{ij}^2, t_{ij}^3)'$. All analyses are conducted on the square root transformed CD4 counts as this transformation can reduce the skewness of the original CD4 measurements, as indicated by Zeger and Diggle (1994). We simply adopt a leave-one-subject-out cross-validation method to choose the optimal bandwidth of 0.25.

Table 6
Regression coefficients and the corresponding standard errors in the CD4 cell counts study in HIV seroconverters using the WI method, the CDSE method, and the "spline" method.

Methods	Age	Smoking	Drug	Sexual partners	Depression
WI	0.0148 (0.0380)	0.973 (0.177)	1.084 (0.554)	−0.0702 (0.0634)	−0.0323 (0.0254)
CDSE	0.0090 (0.0256)	0.760 (0.113)	0.862 (0.300)	0.0332 (0.0361)	−0.0333 (0.0128)
Spline	0.0094 (0.0257)	0.769 (0.114)	0.907 (0.299)	0.0381 (0.0360)	−0.0315 (0.0127)

Table 7
Estimates of γ and λ, and the corresponding standard errors in the CD4 cell counts study in HIV seroconverters using the CDSE method and the "spline" method.

Methods	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\lambda}_4$	$\hat{\lambda}_5$
	(0.048)	(0.077)	(0.033)	(0.0039)	(0.0039)	(0.021)	(0.070)	(0.0087)	(0.0031)
CDSE	0.690	−0.583	0.177	−0.0178	−0.0023	0.094	−0.031	0.0065	−0.0053
	(0.048)	(0.077)	(0.032)	(0.0039)	(0.0040)	(0.021)	(0.071)	(0.0087)	(0.0031)
Spline	0.688	−0.577	0.174	−0.0175	−0.0024	0.092	−0.022	0.0048	−0.0049

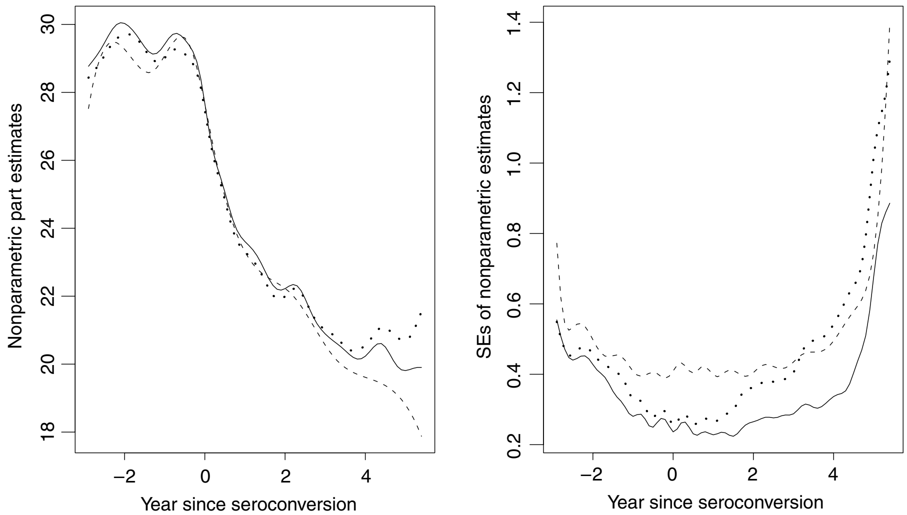

Fig. 2. Estimates of $m(t)$ (average time course of squared-root CD4 number) (left) and their estimated pointwise SEs (right). The dotted, solid, and short dashes curves correspond to the WI estimates, the CDSE estimates, and the "spline" method estimates, respectively.

Table 6 gives the estimates of β based on the WI method, the proposed CDSE method, and the "spline" method. The values in parentheses are standard errors (SEs) of the estimates. Based on our proposed method, we can conclude that (i) age and the number of sexual partners have no significant effects on the CD4 cell counts, (ii) smoking and recreational drug use are significantly positively associated with the CD4 cell numbers, and (iii) depression symptom is significantly negatively associated with the CD4 counts. It is noted that there is a sign difference for number of sexual partners between the WI method and our proposed CDSE method. In this case, our proposed method seems to suggest a more sensible conclusion that the more the sexual partners the higher the CD4 cell counts, which is also consistent with the conclusion drawn in Zeger and Diggle (1994). Consistent with the theorem and simulation results, our proposed CDSE method yields overall smaller SEs than the WI method. From Tables 6 and 7, our proposed method reaches similar conclusions as those from the "spline" method.

Finally, the nonparametric curve estimates based on the WI method (i.e., dotted line), our proposed CDSE method (i.e., solid line), and the "spline" method (i.e., short dashes line) are plotted in Fig. 2 (left). The CD4 counts were stable before seroconversion and sharply decreased after seroconversion. By taking the correlation into account, our proposed method and the "spline" method both suggest that the decreasing trend remained even 3 years after seroconversion. The estimated SEs are given in Fig. 2 (right). The SE of our proposed curve estimate is uniformly smaller than that of the WI estimate, which agrees with the theory. Moreover, the SE of our proposed curve estimate is uniformly smaller than that of the "spline" estimate.

Similar to Cook and Weisberg (1982), we define the PRESS (predicted residual sum of squares) as $\sum_i \sum_j r_{ij}^2(\hat{\beta}_{(-i)}, \hat{m}_{(-i)}(t_{ij}))$ with $r_{ij}(\cdot, \cdot)$ being defined in Section 2, where $\hat{\beta}_{(-i)}$ and $\hat{m}_{(-i)}(\cdot)$ are the estimates with the ith subject excluded. The PRESS value of our proposed method is 86 917, which is smaller than that of the "spline" method (the PRESS value of the "spline" method is 87 743). Based on the PRESS measure, our CDSE method works with less prediction error and is thus appealing.

6. Conclusion

Based on the modified Cholesky decomposition, we propose a two-step-procedure to obtain efficient semiparametric estimates for β, γ and λ via profile likelihood approach for the longitudinal partially linear model. Here, γ and λ are used to characterize the within-subject correlation through GARP/IV parameters. Our proposed method is more computationally efficient than the general method proposed by Lin and Carroll (2006). Our approach can be applied to unbalanced data and guarantees the positive definiteness of the estimated covariance matrices. Theorem and simulation results show that our proposed method performs better than the WI method (Lin and Carroll, 2001). Specifically, our proposed method produces less bias for $\hat{\lambda}_1$ and more accuracy in estimating the covariance matrix than the "spline" method by Leng et al. (2010). In the real data analysis, the SE of our proposed nonparametric curve estimate for the mean is uniformly smaller than that of Leng et al. (2010). Moreover, our CDSE method outperforms the "spline" method when the PRESS is used as the criterion.

From the results of our simulated studies given in Table 5, our method is robust to disturbance of the covariance structure. Hence, our proposed model and method could be widely applicable in practice. It would be of future research interest to consider semiparametric models for the mean and the covariance structure simultaneously as pointed out by one referee.

Acknowledgments

We thank two anonymous referees and an associate editor very much for their helpful and constructive comments on this paper. This work has been partly supported by Program for New Century Excellent Talents in University, National Nature Science Foundation of China (No. 11071035), National Nature Science Foundation of China (No. 10931002) and the Fundamental Research Funds for the Central Universities (No. 09SSXT116).

Appendix

A.1. Conditions and proof

The following technical conditions are imposed. They may not be the weakest possible conditions; but, they are imposed to facilitate the proofs.

1. The density function $f(t)$ has compact support Ω, and is Lipschitz-continuous and bounded away from 0. The function $K(\cdot)$ is a symmetric density function with a compact support.
2. As is common in longitudinal data analyses, there exists an integer K_0 such that $J_i \le K_0$, for $i = 1, 2, \ldots, n$.
3. $nh^8 \to 0$ and $nh^2/(\log n)^2 \to \infty$.
4. $m(\cdot)$ has a continuous second order derivative.

Detailed proof of Theorem 3.1 is given in a longer version of this paper available from the authors.

Lemma 1. *Suppose that the assumptions given in Section 3 and the conditions above hold. We have*

$$\hat{m}(t;\beta,\gamma) = m(t;\beta,\gamma) + O_p(c_n),$$

holds uniformly in t, where $m(t;\beta,\gamma) = m_0(t) - g_1(t)'(\beta-\beta_0)$ *with* $g_1 = (g_{11},\ldots,g_{1p})'$, *and* $c_n = \{\log(1/h)/(nh)\}^{1/2} + h^2$.

We define

$$Q(\beta,\gamma,\lambda,m(\cdot)) = -\frac{1}{2}\left\{\sum_{i=1}^{n}\sum_{j=1}^{J_i} u_{ij}'\lambda + \sum_{i=1}^{n}\sum_{j=1}^{J_i}\frac{\left\{r_{ij}(\beta,m(t_{ij})) - \sum_{l=1}^{j-1} r_{il}(\beta,m(t_{il}))w_{ijl}'\gamma\right\}^2}{\exp(u_{ij}'\lambda)}\right\},$$

and $\alpha = (\beta',\gamma',\lambda')'$. α_0 denotes the true value of α. From (5), we have

$$\begin{aligned} 0 &= \frac{1}{\sqrt{n}}\frac{\partial}{\partial\alpha}Q(\hat{\beta},\hat{\gamma},\hat{\lambda},\hat{m}(\cdot;\hat{\beta},\hat{\gamma})) \\ &= \frac{1}{\sqrt{n}}\frac{\partial}{\partial\alpha}Q(\beta_0,\gamma_0,\lambda_0,\hat{m}(\cdot;\beta_0,\gamma_0)) + \frac{1}{n}\frac{\partial^2}{\partial\alpha\partial\alpha'}Q(\beta,\gamma,\lambda,\hat{m}(\cdot;\beta,\gamma))\,|_{\alpha=\alpha^*}\sqrt{n}(\hat{\alpha}-\alpha_0), \end{aligned} \tag{8}$$

where $\alpha^* = (\beta^{*'},\gamma^{*'},\lambda^{*'})'$ lies between $\hat{\alpha}$ and α_0. By Taylor's expansion and Lemma 1, following the same lines as the proof of Lemma 2 in Severini and Wong (1992), tedious calculations show:

Lemma 2. *Suppose that the assumptions given in Section 3 and the conditions above hold.*

$$\frac{1}{\sqrt{n}}\frac{\partial}{\partial\alpha}Q(\beta_0,\gamma_0,\lambda_0,\hat{m}(\cdot;\beta_0,\gamma_0)) = \frac{1}{\sqrt{n}}\frac{\partial}{\partial\alpha}Q(\beta_0,\gamma_0,\lambda_0,m(\cdot;\beta_0,\gamma_0)) + o_p(1).$$

Following the same arguments of Carroll et al. (1997), we can obtain:

Lemma 3. *Suppose that the assumptions given in Section 3 and the conditions above hold.*

$$\hat{m}(t) - m_0(t) = \left\{\sum_{i=1}^{n}\sum_{j=1}^{J_i} K_h(t_{ij} - t)\right\}^{-1} \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{J_i} K_h(t_{ij} - t)\left\{r_{ij}(\beta_0, m_0(t_{ij})) - \sum_{l=1}^{j-1} r_{il}(\beta_0, m_0(t_{il}))\phi_{ijl}\right\}$$
$$+ \frac{1}{2}h^2 C_2(K)b(t) + \frac{1}{2}h^2 C_2(K)\left\{\sum_{i=1}^{n}\sum_{j=1}^{J_i} K_h(t_{ij} - t)\right\}^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J_i}\sum_{l=1}^{j-1} K_h(t_{ij} - t)b(t_{il})\phi_{ijl} + o_p(c_n),$$

where $C_2(K) = \int t^2 K(t)\mathrm{d}t$ *and* $b(t) = m_0''(t) + 2f'(t)m_0'(t)/f(t)$.

Proof of Theorem 3.1. By (8) and Lemma 2, we can obtain (6). The semiparametric asymptotic efficiency of the estimators $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\lambda}$ can be obtained following the same arguments of Begun et al. (1983). From Lemma 3, (7) comes immediately. □

A.2. The main algorithm

The solutions for β, γ, and λ of (5) can be obtained iteratively by the following iterative procedure, and the procedure is natural from Verbyla (1993) and Daniels and Pourahmadi (2002) s' algorithms.

Step 0. Given a initial value $(\beta_{(0)}', \gamma_{(0)}', \lambda_{(0)}')'$, use models (3) to form the lower triangular matrices $P_{i(0)}$ and diagonal matrices $D_{i(0)}$, then $\Sigma_{i(0)}$ are obtained as the starting values of Σ_i.

Step $(n+1, 1)$. β is updated through

$$\beta_{(n+1)} = (\hat{Z}'\Sigma^{-1}\hat{Z})^{-1}\hat{Z}'\Sigma^{-1}\check{Y}\,|_{\lambda=\lambda_{(n)},\gamma=\gamma_{(n)}},$$

where $\hat{Z} = (\hat{z}_{11}, \ldots, \hat{z}_{1J_1}, \ldots, \hat{z}_{nJ_n})'$, $\Sigma = \operatorname{diag}\{\Sigma_i\}_{i=1}^{n}$, and $\check{Y} = (\check{y}_{11}, \ldots, \check{y}_{1J_1}, \ldots, \check{y}_{nJ_n})'$ with $\check{y}_{ij} = y_{ij} - \{\sum_{u=1}^{n}\sum_{v=1}^{J_i} K_h(t_{uv} - t_{ij})(y_{uv} - \sum_{k=1}^{v-1} r_{uk}(\hat{\beta}_I, \hat{m}_I(t_{uk}))w_{uvk}'\gamma)\}/\{\sum_{u=1}^{n}\sum_{v=1}^{J_i} K_h(t_{uv} - t_{ij})\}$;

Step $(n+1, 2)$. γ is updated as

$$\gamma_{(n+1)} = (\tilde{Z}'D^{-1}\tilde{Z})^{-1}\tilde{Z}'D^{-1}(Y - Z\beta - \hat{m})\,|_{\beta=\beta_{(n+1)},\gamma=\gamma_{(n)},\lambda=\lambda_{(n)}},$$

where $D = \operatorname{diag}\{\sigma_{ij}^2\}_{i=1 j=1}^{n\ J_i}$, $Y = (y_{11}, \ldots, y_{1J_1}, \ldots, y_{nJ_n})'$, $Z = (z_{11}, \ldots, z_{1J_1}, \ldots, z_{nJ_n})'$, $\hat{m} = (\hat{m}(t_{11}; \beta, \gamma), \ldots, \hat{m}(t_{1J_1}; \beta, \gamma), \ldots, \hat{m}(t_{nJ_n}; \beta, \gamma))'$, and $\tilde{Z} = (Z(1)', \ldots, Z(n)')'$ with $Z(i) = (z(i, 1), \ldots, z(i, J_i))'$, here $z(i, j) = \sum_{v=1}^{j-1}(y_{iv} - z_{iv}'\beta - \hat{m}(t_{iv}; \beta, \gamma))w_{ijv}$;

Step $(n+1, 3)$. λ can be updated by

$$\lambda_{(n+1)} = \lambda_{(n)} + (U'U)^{-1}U'(D^{-1}d - \mathbf{1})\,|_{\beta=\beta_{(n+1)},\gamma=\gamma_{(n+1)},\lambda=\lambda_{(n)}},$$

where $U = (u_{11}, \ldots, u_{1J_1}, \ldots, u_{nJ_n})'$, $d = (d_{11}, \ldots, d_{1J_1}, \ldots, d_{nJ_n})'$ with $d_{ij} = (y_{ij} - z_{ij}'\beta - \hat{m}(t_{ij}; \beta, \gamma) - \sum_{v=1}^{j-1}(y_{iv} - z_{iv}'\beta - \hat{m}(t_{iv}; \beta, \gamma))w_{ijv}'\gamma)^2$, and $\mathbf{1}$ denotes the $\sum_i^n J_i \times 1$ vector of unit elements.

Until convergence, we get the estimators $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\lambda}$ on (5).

A convenient initial value for $(\gamma', \lambda')'$ is $\gamma_{(0)} = 0$ and $\lambda_{(0)} = 0$. In other words, the $(J_i \times J_i)$ identity matrix may be chosen as the starting value for the covariance matrix Σ_i.

References

Begun, J.H., Hall, W.J., Huang, W.M., Wellner, J.A., 1983. Information and asymptotic efficiency in parametric–nonparametric models. Ann. Statist. 11, 432–452.

Carroll, R.J., Fan, J., Gilbels, I., Wand, M.P., 1997. Generalized partially linear single-index models. J. Amer. Statist. Assoc. 92, 477–489.

Chamberlain, G., 1987. Asymptotic efficiency in estimation with conditional moment restrictions. J. Econometrics 34, 305–334.

Cook, R.D., Weisberg, S., 1982. Residuals and Influnce in Regression. Chapman and Hall Press.

Daniels, M.J., Pourahmadi, M., 2002. Bayesian analysis of covariance matrices and dynamic models for longitudinal data. Biometrika 89 (3), 553–566.

Daniels, M.J., Zhao, Y.D., 2003. Modelling the random effects covariance matrix in longitudinal data. Stat. Med. 22, 1631–1647.

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

Fan, J., Huang, T., Li, R., 2007. Analysis of longitudinal data with semiparametric estimation of covariance function. J. Amer. Statist. Assoc. 102, 632–641.

Fan, J., Wu, Y., 2008. Semiparametric estimation of covariance matrices for longitudinal data. J. Amer. Statist. Assoc. 103, 1520–1533.

Kitamura, Y., Tripathi, G., Ahn, H., 2004. Empirical likelihood-based inference in conditional moment restriction models. Econometrica 72, 1667–1714.

Leng, C., Zhang, W., Pan, J., 2010. Semiparametric mean–covariance regression analysis for longitudinal data. J. Amer. Statist. Assoc. 105, 181–193.

Li, J., Xia, Y., Palta, M., Shankar, A., 2009. Impact of unknown covariance structures in semiparametric models for longitudinal data: an application to Wisconsin diabetes data. Comput. Statist. Data Anal. 53, 4186–4197.

Liang, H., Wang, S., Robins, J.M., Carroll, R.J., 2004. Estimation in partially linear models with missing covariates. J. Amer. Statist. Assoc. 99, 357–367.

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

Lin, X., Carroll, R.J., 2001. Semiparametric regression for clustered data using generalized estimating equations. J. Amer. Statist. Assoc. 96, 1045–1056.

Lin, X., Carroll, R.J., 2006. Semiparametric estimation in general repeated measures problems. J. Roy. Statist. Soc. Ser. B 68 (Part 1), 69–88.

Lombardia, M.J., Sperlich, S., 2008. Semiparametric inference in generalized mixed effects models. J. Roy. Statist. Soc. Ser. B 70 (Part 5), 913–930.

Pepe, M.S., Couper, D., 1997. Modeling partly conditional means with longitudinal data. J. Amer. Statist. Assoc. 92, 991–998.
Pourahmadi, M., 1999. Joint mean–covariance models with applications to longitudinal data: unconstrained parameterisation. Biometrika 86, 677–690.
Pourahmadi, M., 2000. Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. Biometrika 87, 425–435.
Pourahmadi, M., Daniels, M.J., 2002. Dynamic conditional linear mixed models for longitudinal data. Biometrics 58, 225–231.
Severini, T.A., Wong, W.H., 1992. Profile likelihood and conditionally parametric models. Ann. Statist. 20, 1768–1802.
Speckman, P., 1988. Kernel smoothing in partial linear models. J. Roy. Statist. Soc. Ser. B 50, 413–436.
Verbyla, A.P., 1993. Modeling variance heterogeneity: residual maximum likelihood and diagnostics. J. Roy. Statist. Soc. Ser. B 55, 493–508.
Wang, N., Carroll, R.J., Lin, X., 2005. Efficient semiparametric marginal estimation for longitudinal/clustered data. J. Amer. Statist. Assoc. 100, 147–157.
Wang, Y.G., Carey, V., 2003. Working correlation structure misspecification, estimation and covariate design: implications for generalised estimating equations performance. Biometrika 90, 29–41.
Wu, W., Pourahmadi, M., 2003. Nonparametric estimation of large covariance matrices of longitudinal data. Biometrika 90, 831–844.
Ye, H., Pan, J., 2006. Modelling of covariance structures in generalised estimating equations for longitudinal data. Biometrika 93, 927–941.
You, J., Zhou, X., 2006. Statistical inference in a panel data semiparametric regression model with serially correlated errors. J. Multivariate Anal. 97, 844–873.
Zeger, S.L., Diggle, P.J., 1994. Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. Biometrics 50, 689–699.