Biometrics 74, 220–228
March 2018

DOI: 10.1111/biom.12712

A Profile Likelihood Approach for Longitudinal Data Analysis

Ziqi Chen,[1,*] Man-Lai Tang,[2] and Wei Gao[3,]**

[1]School of Mathematics and Statistics, Central South University, Changsha, 410083, China
[2]Department of Mathematics and Statistics, Hang Seng Management College, HongKong
[3]Key Laboratory for Applied Statistics of MOE, School of Mathematics and Statistics Northeast Normal University, Changchun, 130024, China
*email: chenzq453@gmail.com
**email: gaow@nenu.edu.cn

Summary. Inappropriate choice of working correlation structure in generalized estimating equations (GEE) could lead to inefficient parameter estimation while impractical normality assumption in likelihood approach would limit its applicability in longitudinal data analysis. In this article, we propose a profile likelihood method for estimating parameters in longitudinal data analysis via maximizing the estimated likelihood. The proposed method yields consistent and efficient estimates without specifications of the working correlation structure nor the underlying error distribution. Both theoretical and simulation results confirm the satisfactory performance of the proposed method. We illustrate our methodology with a diastolic blood pressure data set.

Key words: Generalized estimating equations; Kernel estimation; Modified Cholesky decomposition; Profile likelihood.

1. Introduction

Longitudinal data arise frequently in biomedical and health studies in which repeated measurements from the same subject are correlated. Consistency and efficiency of estimators for the regression parameters are important for longitudinal data analysis. Liang and Zeger (1986)developed the generalized estimating equation (GEE) for longitudinal data analysis. GEE approach takes advantage of the built-in robustness since no specification of the full likelihood is required. It is well known that GEE estimators are efficient when the working correlation structure is correctly specified. However, misspecification of the working correlation structure may lead to a great loss of efficiency even though the consistency may remain valid (Wang and Carey, 2003). The quadratic inference function (QIF) method proposed by Qu, Lindsay, and Li (2000)does not involve direct estimation of the correlation matrix and remains optimal even if the working correlation matrix is misspecified. Ye and Pan (2006)proposed the simultaneous GEE equations to estimate both the mean regression coefficients and the covariance structure parameters. Leung, Wang, and Zhu (2009)proposed a hybrid method that combines multiple GEEs based on different working correlation models and obtains parameter estimates by maximizing the empirical likelihood (Qin and Lawless, 1994). Nonetheless, all aforementioned articles require the specification of working correlation models. Again, correct specification of the correlation structure is necessary in order to increase the efficiency.

Since the correlation structure plays a crucial role in mean structure estimation (e.g., efficiency), the estimation of the covariance matrix is important in longitudinal analysis.Motivated by the modified Cholesky decomposition, Pourahmadi (1999, 2000)proposed to simultaneously estimate the mean regression coefficients and covariance matrices by assuming that the errors follow the multivariate normal distribution. Although the corresponding estimates can be shown to be efficient, it is impractical to specify the full likelihood function (in particular, multivariate normal distribution) due to the correlated nature of longitudinal data. To overcome the latter issue, we propose to regress the error on its predecessors (Pourahmadi, 1999), treat the prediction error density as an unknown nonparametric function, and estimate it via kernel smoothing. With the estimated prediction error density, we obtain the estimates of the regression parameters by maximizing the so-called profile likelihood function. Our proposed method performs well without specification of the likelihood nor the correlation structure. Most importantly, we show that the proposed estimates are efficient in both theory and practice.

We organize our article as follows. In Section 2, we first introduce the independence maximum profile likelihood method and derive the efficient maximum profile likelihood estimators. The asymptotic properties of the estimators are investigated. In Section 3, simulation studies are conducted to evaluate the performance of our proposed method and its competitors. A real data set in diastolic blood pressure study is analyzed to illustrate our methodologies. A brief discussion is presented in Section 4. Some simulation results, the technical conditions, some lemmas, and the proof of Theorem 2 are presented in the Supplementary Materials.

2. The Maximum Profile Likelihood Methods

We intend to record the response variable for the i-th subject m times, denoted as $y_i^* = (y_{i1}, \ldots, y_{im})^T$. Here, y_i^*'s can be safely assumed to be independently distributed, $i = 1, \ldots, n$, with n being the sample size. The covariate corresponding to y_{ij} is x_{ij}, which is a p-dimensional vector. Denote $X_i^* = (x_{i1}, \ldots, x_{im})^T$, which is the $m \times p$-dimensional design matrix for the i-th subject. Let $\mu_{ij} = E(y_{ij}|X_i) = x_{ij}^T\beta$, where β is a p-dimensional parameter vector with true value being β^*. Denote $\epsilon_{ij} = y_{ij} - x_{ij}^T\beta^*$ and $\epsilon_i^* = (\epsilon_{i1}, \ldots, \epsilon_{im})^T$. We assume that the error vector ϵ_i^* is independent of the covariate matrix X_i^*, and ϵ_i^* $(i = 1, \ldots, n)$ are independently and identically distributed. In practice, missing responses are not uncommon in longitudinal studies. Let r_{ij} take value 1 if y_{ij} is observed, and 0 otherwise. Here, we assume that the missing data patterns are monotone (Rubin 1987; Robins and Rotnitzky, 1995), that is, $r_{i1} \geq r_{i2} \geq \ldots \geq r_{i(m-1)} \geq r_{im}$, and assume the data are missing completely at random (MCAR), that is, R_i^* is independent of y_i^* and X_i^*, where $R_i^* = (r_{i1}, \ldots, r_{im})^T$. Assume R_i^* $(i = 1, \ldots, n)$ are independently and identically distributed. We assume $P(r_{ij} = 1) = \alpha_j$ with $1 = \alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_{m-1} \geq \alpha_m > 0$. Suppose the responses observed on the i-th subject are $y_i = (y_{i1}, \ldots, y_{im_i})^T$ with the corresponding observed covariates $X_i = (x_{i1}, \ldots, x_{im_i})^T$, where $m_i = \sum_{j=1}^m r_{ij}$. Let $n_j = \sum_{i=1}^n r_{ij}$ and we have $\lim_{n\to\infty} n_j/n = \alpha_j > 0$.

2.1. *The Independence Estimating Procedure*

We first assume that the responses of the i-th subject are independent of each other, for $i = 1, \ldots, n$. We know that the cross-sectional errors $\epsilon_{1j}, \ldots, \epsilon_{nj}$ are i.i.d. with common density function, denoted as $f_{\epsilon_{1j}}$ $(j = 1, \ldots, m)$. The maximum likelihood estimator of β could be readily obtained if the density functions, that is, $f_{\epsilon_{1j}}$ $(j = 1, \ldots, m)$, are known. However, these density functions are usually unknown in practice. Let $\epsilon_{ij}(\beta) = y_{ij} - x_{ij}^T\beta$. We propose to estimate the density function $f_{\epsilon_{1j}(\beta)}$ for any β by kernel smoothing as

$$\hat{f}_{\epsilon_{1j}(\beta)}(u) = \frac{1}{(\sum_{i=1}^n r_{ij})h_j} \sum_{i=1}^n r_{ij} K\left(\frac{\epsilon_{ij}(\beta) - u}{h_j}\right),$$

where K is a scalar kernel and h_j is any appropriate bandwidth, for $j = 1, \ldots, m$. If the responses of the same subject are independent of each other (i.e., $\epsilon_{11}, \ldots, \epsilon_{1m}, \ldots, \epsilon_{n1}, \ldots, \epsilon_{nm}$ are all independent of each other), we could simply propose to obtain the estimator of β by maximizing the profile likelihood equation (Chen et al., 2014):

$$\hat{\beta}_{MPL}^I = \arg\max_\beta \sum_{j=1}^m \sum_{i=1}^n r_{ij} \log \hat{f}_{\epsilon_{1j}(\beta)}(\epsilon_{ij}(\beta)). \tag{1}$$

Since x_{ij} is independent of ϵ_{ij}, by Lemma 1 of Chen et al. (2014), we know

$$\sum_{j=1}^m \alpha_j \int f_{\epsilon_{1j}(\beta)}(u) \log f_{\epsilon_{1j}(\beta)}(u) du$$
$$< \sum_{j=1}^m \alpha_j \int f_{\epsilon_{1j}(\beta^*)}(u) \log f_{\epsilon_{1j}(\beta^*)}(u) du,$$

for any $\beta \neq \beta^*$. Given the MCAR mechanism and $n_j = O(n)$, using similar arguments in Lemma 3 (see Supplementary Materials), we have, for given $r_{1j}, \ldots, r_{nj}$, $\sup_\beta |\hat{f}_{\epsilon_{1j}(\beta)}(\epsilon_{ij}(\beta)) - f_{\epsilon_{1j}(\beta)}(\epsilon_{ij}(\beta))| = O_p\{\sqrt{\frac{\log n_j}{n_j h_j}} + h_j^2\}$, for $j = 1, \ldots, m$. Thus,

$$\sum_{j=1}^m \frac{1}{n} \sum_{i=1}^n r_{ij} \log \hat{f}_{\epsilon_{1j}(\beta)}(\epsilon_{ij}(\beta))$$
$$\to_P \sum_{j=1}^m \alpha_j \int f_{\epsilon_{1j}(\beta)}(u) \log f_{\epsilon_{1j}(\beta)}(u) du$$

holds uniformly in β. By Theorem 2.1 of Newey and McFadden (1994), we have $\hat{\beta}_{MPL}^I \to_P \beta^*$.

Even though the consistency of $\hat{\beta}_{MPL}^I$ holds, it may not be an efficient estimator since $\hat{\beta}_{MPL}^I$ is obtained based on the independence assumption (i.e., the within-subject correlation is not yet taken into account). This is consistent with the property of GEE estimator that one could not get the fully efficient estimator when independence working correlation structure is adopted (Liang and Zeger, 1986; Diggle, et al., 2002). The asymptotic normality theory in the next subsection and numerical studies in Section 3 further verify this fact.

REMARK 1. *Although we only consider the monotone missing pattern in this article, the independence estimating procedure could be readily modified to deal with non-monotone missing pattern.*

2.2. *The Efficient Estimation Approach*

In practice, it is more realistic to assume that the m complete responses from each subject are correlated. For this purpose, let the complete errors $\epsilon_i^* = (\epsilon_{i1}, \ldots, \epsilon_{im})^T$ and assume that $\mathrm{Cov}(\epsilon_i^*) = \Sigma$ $(i = 1, \ldots, n)$ with Σ being a $m \times m$ positive definite matrix. By the modified Cholesky decomposition (Pourahmadi, 1999), there exists a lower triangular matrix P with ones as diagonal entries and $-\phi_{jl}$ as the (j, l)th element and a diagonal matrix $D = \mathrm{diag}(\sigma_{i1}^2, \ldots, \sigma_{im}^2)$ such that $P\Sigma P^T = D$. Based on this decomposition, one can regress ϵ_{ij} on its predecessors $\epsilon_{i1}, \ldots, \epsilon_{i(j-1)}$ with the corresponding regression coefficients being $\phi_{j1}, \ldots, \phi_{j(j-1)}$ and denotes the corresponding successive prediction error as η_{ij}, that is,

$$\eta_{ij} = \epsilon_{ij} - \sum_{l=1}^{j-1} \phi_{jl}\epsilon_{il}, \text{ for } j = 2, \ldots, m, \ i = 1, \ldots, n.$$

It is noteworthy that $\mathrm{Cov}(\eta_i^*) = D$ where $\eta_i^* = (\eta_{i1}, \ldots, \eta_{im})^T$ for $i = 1, \ldots, n$. Hence, $\eta_{i1}, \ldots, \eta_{im}$ are uncorrelated random variables. The specification of ϕ_{jl}'s determines the correlation structure of the error ϵ_i. For example, the error has the independence correlation structure if ϕ_{jl}'s are all zero, and has the AR-1 correlation structure if ϕ_{jl} is zero for $j - l \geq 2$.

Define $\eta_{ij}(\beta, \phi_j) = y_{ij} - x_{ij}^T\beta - \sum_{l=1}^{j-1}(y_{il} - x_{il}^T\beta)\phi_{jl}$, where $\sum_{l=1}^0 = 0$ and $\phi_j = (\phi_{j1}, \ldots, \phi_{j(j-1)})$. We assume η_{1j} follows some completely nonparametric distribution for $j = 1, \dots, m$ and use the kernel nonparametric technique to estimate the density function of η_{1j}. Due to the presence of missing responses, we could only observe the first m_i responses and the corresponding covariates for subject i, $i = 1, \dots, n$. If ϕ_j ($j = 2, \dots, m$) are known, we can estimate the density function of $\eta_{1j}(\beta, \phi_j)$ using the following kernel smoothing

$$\hat{f}_{\eta_{1j}(\beta,\phi_j)}(u) = \frac{1}{(\sum_{i=1}^{n} r_{ij})h_j} \sum_{i=1}^{n} r_{ij} K\left(\frac{\eta_{ij}(\beta, \phi_j) - u}{h_j}\right), \quad \text{for } j = 1, \dots, m,$$

where K is a scalar kernel and h_j is an appropriate bandwidth. With the estimated prediction error densities, we propose to get the estimate of β through the maximum profile likelihood (MPL) method considered in the last sub-section. That is, $\hat{\beta}_{MPL}$ is obtained through maximizing the following profile likelihood equation:

$$\sum_{j=1}^{m} \sum_{i=1}^{n} r_{ij} \log \hat{f}_{\eta_{1j}(\beta,\phi_j)}(\eta_{ij}(\beta, \phi_j)). \tag{2}$$

It should be noticed that the estimation procedure proposed here is consistent with the profile likelihood approach for estimating regression parameters in semiparametric longitudinal models (see also, Wang, Carroll, and Lin, 2005; Lin and Carroll, 2006; Fan, Huang, and Li, 2007; Lombardia and Sperlich, 2008). The basic idea of profile likelihood approach is to replace the unknown function by its nonparametric (kernel) estimate for given parametric components (Linton, Sperlich, and Van Keilegom, 2008). We conclude that the choice of ϕ_{jl} and thus the correlation structure being used has no impact on the consistency property of the estimated β. This coincides with that in the GEE approach in the literature. Specifically, for any ϕ_j $(j = 2, \dots, m)$ and noting that ϵ_i is independent of X_i, we know by Lemma 1 of Chen et al. (2014) that

$$\begin{aligned} &\sum_{j=1}^{m} \alpha_j \int f_{\eta_{1j}(\beta,\phi_j)}(u) \log f_{\eta_{1j}(\beta,\phi_j)}(u) du \\ &< \sum_{j=1}^{m} \alpha_j \int f_{\eta_{1j}(\beta^*,\phi_j)}(u) \log f_{\eta_{1j}(\beta^*,\phi_j)}(u) du \end{aligned}$$

since

$$\begin{aligned} &\sum_{j=1}^{m} \frac{1}{n} \sum_{i=1}^{n} r_{ij} \log \hat{f}_{\eta_{1j}(\beta)}(\eta_{ij}(\beta, \phi_j)) \\ &\to_P \sum_{j=1}^{m} \alpha_j \int f_{\eta_{1j}(\beta,\phi_j)}(u) \log f_{\eta_{1j}(\beta,\phi_j)}(u) du \end{aligned}$$

holds uniformly in β. Hence, $\hat{\beta}_{MPL}$ is a consistent estimator (Newey and McFadden, 1994).

Pourahmadi (1999, 2000)assumed ϵ_i^* follows the multivariate normal distribution and η_{ij}'s are hence normally and independently distributed. Ones could then use the maximum likelihood (ML) method to obtain efficient estimator of β by using the observed responses and the corresponding covariates. If ϵ_i^* are not normally distributed (e.g., multivariate T distribution or mixture of multivariate normal distributions), the ML method in Pourahmadi (1999, 2000)may not work satisfactorily. In real applications, ones usually do not know the distribution of ϵ_i^*. In these cases, the proposed MPL approach would be a good alternative for estimating β consistently and efficiently, since it estimates the prediction error density via kernel smoothing and thus does not require specification of the underlying error distribution.

REMARK 2. *The proposed MPL estimator is motivated by Cholesky decomposition of the covariance matrix Σ. However, the consistency property of the proposed estimator does not rely on existence of the covariance matrix of the error. Specifically, we regress the error on its predecessors motivated by the Cholesky decomposition and estimate the prediction error density. The proposed approach is based on maximizing the estimated likelihood and does not directly include the covariance of error ϵ_i or operate on its empirical counterpart. On the contrary, the GEE approach requires the covariance matrix of the error directly. As shown in Study 6 in our simulation study, when the covariance matrix diverges, our proposed method still performs very well while the GEE method breaks down.*

REMARK 3. *Our proposed method is robust against outliers. Intuitively, by the property of commonly used kernel functions (e.g., gaussian and Epanechnikov kernels), an outlier being distant from other observations would have much smaller estimated density value and thus has little impact on the estimation of β. The robust performance is further illustrated via the subsequent simulation studies.*

Let $\Phi = (\phi_2, \dots, \phi_m)$. Given R_i^*, $m_i = \sum_{j=1}^{m} r_{ij}$, for $i = 1, \dots, n$. Define

$$\begin{aligned} W_i(\Phi) = \Bigg(&\frac{f'_{\eta_{11}(\beta^*)}(\eta_{i1}(\beta^*))}{f_{\eta_{11}(\beta^*)}(\eta_{i1}(\beta^*))}, \frac{f'_{\eta_{12}(\beta^*,\phi_2)}(\eta_{i2}(\beta^*, \phi_2))}{f_{\eta_{12}(\beta^*,\phi_2)}(\eta_{i2}(\beta^*, \phi_2))}, \\ &\dots, \frac{f'_{\eta_{1m_i}(\beta^*,\phi_{m_i})}(\eta_{im_i}(\beta^*, \phi_{m_i}))}{f_{\eta_{1m_i}(\beta^*,\phi_{m_i})}(\eta_{im_i}(\beta^*, \phi_{m_i}))} \Bigg)^T, \end{aligned}$$

$$\begin{aligned} Q_i(\Phi) = \operatorname{diag} \Bigg\{ &\left\{ \frac{f'_{\eta_{11}(\beta^*)}(\eta_{i1}(\beta^*))}{f_{\eta_{11}(\beta^*)}(\eta_{i1}(\beta^*))} \right\}^2, \left\{ \frac{f'_{\eta_{12}(\beta^*,\phi_2)}(\eta_{i2}(\beta^*,\phi_2))}{f_{\eta_{12}(\beta^*,\phi_2)}(\eta_{i2}(\beta^*,\phi_2))} \right\}^2, \\ &\dots, \left\{ \frac{f'_{\eta_{1m_i}(\beta^*,\phi_{m_i})}(\eta_{im_i}(\beta^*,\phi_{m_i}))}{f_{\eta_{1m_i}(\beta^*,\phi_{m_i})}(\eta_{im_i}(\beta^*,\phi_{m_i}))} \right\}^2 \Bigg\}, \end{aligned}$$

and P_i be an $m_i \times m_i$ lower unitriangular matrix with the (j, l)-th below diagonal entry being $-\phi_{jl}$. We present the asymptotic normality of our proposed MPL estimator as follow.

Theorem 1. *Under Conditions (a)–(f) in the Supplementary Materials, for any given* Φ,

$$\sqrt{n}(\hat{\beta}_{MPL}-\beta^*) = -\Omega^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - EX_i)^T P_i^T W_i(\Phi) + o_p(1)$$

holds for given R_i^* $(i=1,\ldots,n)$, *where* $\Omega = \frac{1}{n}\sum_{i=1}^{n}(X_i - EX_i)^T P_i^T Q_i(\Phi) P_i (X_i - EX_i)$.

This theorem could be proved by similar arguments given for Theorem 2 in Chen et al. (2014)and we thus omit its proof. The asymptotic normality of $\hat{\beta}^I_{MPL}$ can be readily obtained by setting each element in Φ to be zero.

We show the efficiency property of $\hat{\beta}_{MPL}$ under the assumption that the errors are normally distributed. Without loss of generality, assume that $EX = 0$ and $m_i = m$. When the true value of Φ (i.e., Φ^*) is used when maximizing (2), since

$$W_i(\Phi^*) = \left(-\frac{\eta_{i1}(\beta^*)}{\sigma_{11}^{*2}}, -\frac{\eta_{i2}(\beta^*,\phi_2^*)}{\sigma_{12}^{*2}}, \ldots, -\frac{\eta_{im}(\beta^*,\phi_m^*)}{\sigma_{1m}^{*2}}\right)^T$$

with σ_{1j}^{*2} being the variance of $\eta_{ij}(\beta^*,\phi_j^*)$, we can readily show that

$$\sqrt{n}(\hat{\beta}_{MPL}-\beta^*) \to_L N\left(\mathbf{0}, \{E(X_1^T\Sigma^{*-1}X_1)\}^{-1}\right),$$

where Σ^* is the true value of Σ. It is easily seen that $\hat{\beta}_{MPL}$ is asymptotically as efficient as the maximum likelihood estimator based on the true distribution of the error. We conclude that the choice of Φ does not affect the asymptotical consistency as well as normality of the MPL estimator; however, determines the efficiency of the resulted MPL estimator. This result is consistent with the properties of GEE approach in the literature. Most importantly, as long as $\eta_{i1}(\beta^*), \eta_{i2}(\beta^*,\phi_2^*), \ldots, \eta_{im}(\beta^*,\phi_m^*)$ are independent of each other, we can obtain the fully efficient estimator $\hat{\beta}_{MPL}$ when Φ is set to Φ^* regardless of the error distribution.

In practice, ones usually do not know the true value of Φ. Let $\hat{\beta}$ be a $\sqrt{n}$−consistent estimator of β, $\hat{\mu}_{ij} = x_{ij}^T\hat{\beta}$ and $\hat{\eta}_{ij}(\phi_j) = y_{ij} - \hat{\mu}_{ij} - \sum_{l=1}^{j-1}(y_{il}-\hat{\mu}_{il})\phi_{jl}$, $j = 2,\ldots,m$, $i = 1,\ldots,n$. We propose to obtain the estimator of ϕ_j (i.e., $\hat{\phi}_j$) by maximizing

$$\sum_{i=1}^{n} r_{ij}\log \hat{f}_{\hat{\eta}_{1j}(\phi_j)}(\hat{\eta}_{ij}(\phi_j)) \quad \text{for} \quad j=2,\ldots,m,$$

where

$$\hat{f}_{\hat{\eta}_{1j}(\phi_j)}(u) = \frac{1}{(\sum_{i=1}^{n} r_{ij})h_j}\sum_{i=1}^{n} r_{ij}K\left(\frac{\hat{\eta}_{ij}(\phi_j)-u}{h_j}\right). \qquad (3)$$

The asymptotical consistency of $\hat{\phi}_j$ is presented in the following theorem with the proof being reported in the Supplementary Materials. To facilitate the proof, we assume that $\eta_{i1}(\beta^*), \eta_{i2}(\beta^*,\phi_2^*), \ldots, \eta_{im_i}(\beta^*,\phi_{m_i}^*)$ are independent of each other, for $i = 1,\ldots,n$.

Theorem 2. *Under Conditions (a)–(f) in the Supplementary Materials,* $\hat{\phi}_j$ *converges in probability to* ϕ_j^*, *for* $j = 2,\ldots,m$.

Due to the consistency property of $\hat{\phi}_j$ ($j = 2,\ldots,m$), when ϕ_j in (2) is replaced by $\hat{\phi}_j$, the resulted MPL estimator of β, that is, $\hat{\beta}^*_{MPL}$, has the asymptotic normality property in Theorem 1 by letting $\Phi = \Phi^*$, that is,

$$\sqrt{n}(\hat{\beta}^*_{MPL}-\beta^*) \to_L N(\mathbf{0}, \Omega^{*-1}\Theta^*\Omega^{*-1}),$$

where $\Omega^* = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\{(X_i - EX_i)^T P_i^{*T} Q_i(\Phi^*) P_i^*(X_i - EX_i)\}$ with P_i^* being the true value of P_i and $\Theta^* = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}[(X_i - EX_i)^T P_i^{*T} W_i(\Phi^*) W_i(\Phi^*)^T P_i^*(X_i - EX_i)]$. We note that $\hat{\beta}^*_{MPL}$ is an efficient estimator. We estimate $Q_i(\Phi^*)$ as

$$\begin{aligned}\hat{Q}_i(\Phi^*) := \operatorname{diag}\Bigg\{&\left\{\frac{\hat{f}'_{\eta_{11}(\hat{\beta}^*_{MPL})}(\eta_{i1}(\hat{\beta}^*_{MPL}))}{\hat{f}_{\eta_{11}(\hat{\beta}^*_{MPL})}(\eta_{i1}(\hat{\beta}^*_{MPL}))}\right\}^2, \\ &\left\{\frac{\hat{f}'_{\eta_{12}(\hat{\beta}^*_{MPL},\hat{\phi}_2)}(\eta_{i2}(\hat{\beta}^*_{MPL},\hat{\phi}_2))}{\hat{f}_{\eta_{12}(\hat{\beta}^*_{MPL},\hat{\phi}_2)}(\eta_{i2}(\hat{\beta}^*_{MPL},\hat{\phi}_2))}\right\}^2, \\ &\ldots, \left\{\frac{\hat{f}'_{\eta_{1m_i}(\hat{\beta}^*_{MPL},\hat{\phi}_{m_i})}(\eta_{im_i}(\hat{\beta}^*_{MPL},\hat{\phi}_{m_i}))}{\hat{f}_{\eta_{1m_i}(\hat{\beta}^*_{MPL},\hat{\phi}_{m_i})}(\eta_{im_i}(\hat{\beta}^*_{MPL},\hat{\phi}_{m_i}))}\right\}^2\Bigg\},\end{aligned}$$

where

$$\hat{f}'_{\eta_{1j}(\beta,\phi_j)}(u) = -\frac{1}{(\sum_{l=1}^{n} r_{lj})h_j^2}\sum_{l=1}^{n} r_{lj}K'\left(\frac{\eta_{lj}(\beta,\phi_j)-u}{h_j}\right), \quad \text{for } j=1,\ldots,m_i.$$

As a result, we estimate Ω^* by

$$\hat{\Omega} := \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_i)^T\hat{P}_i^T\hat{Q}_i(\Phi^*)\hat{P}_i(X_i - \bar{X}_i),$$

where $\bar{X}_i = (\frac{1}{\sum_{l=1}^{n} r_{l1}}\sum_{l=1}^{n} r_{l1}x_{l1}, \ldots, \frac{1}{\sum_{l=1}^{n} r_{lm_i}}\sum_{l=1}^{n} r_{lm_i}x_{lm_i})^T$, and $\hat{P}_i$, the estimator of P_i, is the lower triangular matrix with 1′s as diagonal entries and $-\hat{\phi}_{jl}$ as the (j,l)th element. Under some mild conditions, $\hat{\Omega}^{-1}$ converges in probability to Ω^{*-1} under the Frobenius norm (using Lemma 5). We could similarly obtain a consistent estimator of Θ^* (i.e., $\hat{\Theta}$). Thus, the covariance matrix of $\hat{\beta}^*_{MPL}$ could be estimated by $\dfrac{\hat{\Omega}^{-1}\hat{\Theta}\hat{\Omega}^{-1}}{n}$.

Remark 4. *The assumption* $\eta_{i1},\ldots,\eta_{im_i}$ *are independent are used to facilitate the proof of Theorem 2. However, the consistency and the asymptotical normality (Theorem 1) properties of the MPL estimator does not depend on the assumption that* $\eta_{i1},\ldots,\eta_{im_i}$ *are independent. In practice, even if* $\eta_{i1},\ldots,\eta_{im_i}$ *are uncorrelated but not independent, our proposed MPL estimator performs very well (see, Studies 5 and 6). Note that, the* $\eta_{i1},\ldots,\eta_{im_i}$ *corresponding to the multivariate T distribution are uncorrelated but not independent.*

2.3. *Bandwidth Selection: Maximum-Likelihood Cross-Validation*

The proposed MPL approach is based on maximizing the estimated likelihood and its performance depends on the choice of bandwidth parameters h_j $(j = 1, \ldots, m)$ for kernel smoothing. In this article, we propose to select the bandwidth via the so-called maximum-likelihood cross-validation procedure (Duin, 1976). Here, we only state the selection of the bandwidth parameters for the estimated density functions used in equation (2). The bandwidth selection regarding equations (1) and (3) could be conducted accordingly. Specifically, the maximum-likelihood cross-validation bandwidth for estimating $f_{\eta_{1j}(\beta,\phi_j)}(\cdot)$ is given by

$$\hat{h}_j = \arg\max_{h>0} \sum_{i=1}^{n} r_{ij} \log\left\{ \frac{1}{(\sum_{i=1}^{n} r_{ij} - 1)h} \sum_{k\neq i} r_{kj} \times K\left(\frac{\eta_{kj}(\hat{\beta}, \phi_j) - \eta_{ij}(\hat{\beta}, \phi_j)}{h} \right) \right\}, \text{ for } j = 1, \ldots, m.$$

3. Numerical Studies

3.1. *Simulation Studies*

In this section, we investigate the finite sample performances of the proposed estimation via Monte Carlo simulation studies. We consider the following model:

$$y_{ij} = x_{ij}^T \beta_0 + \epsilon_{ij}, \quad \text{for} \quad i = 1, \ldots, n; \quad j = 1, \ldots, m_i,$$

where β_0 is a p-dimensional vector of parameters. We generate 200 data sets, respectively, for all the studies. We use the simulated average mean square error (SAMSE), which is obtained by averaging $||\hat{\beta} - \beta_0||^2/p$ over 200 simulated samples, to measure the accuracy of estimators (Wang, 2011). We compare our proposed MPL estimator $\hat{\beta}^*_{MPL}$ with GEE and QIF estimators using three different working correlation structures: independence, exchangeable, and the AR-1 working correlation matrices (Zhou and Qu, 2012). Note, the GEE and QIF estimators using the independence working correlation structure are equal. In order to explain that MPL approach accounting for within-subject correlation using our proposed estimating procedure via regressing the error on its predecessors could improve accuracy of the estimators, we also compute the estimators of the parameter (i.e., $\hat{\beta}^I_{MPL}$) using the independence MPL estimating method described in Section 2.1. For all the numerical demonstrations considered in this section, the Gaussian kernel function $K(u) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2)$ is used.

Study 1. The sample size n is taken to be 100, the dimension of the covariates is set to be $p = 3$, and m_i is set to be 5 for $i = 1, \ldots, n$. The dimension of the parameter is low and fixed in this study. The covariate x_{ij} is set to be $(x_{ij1}, x_{ij2}, x_{ij3})^T$, which follows the multivariate normal distribution with mean being zero and covariance matrix being $\Sigma^{(1)}$ with $(\Sigma)^{(1)}_{i,j} = 0.5^{|i-j|}$ for $1 \leq i, j \leq 3$. Let $\beta_0 = (1, 0.5, -0.5)^T$. The errors are set as follows: $\epsilon_{i1} = \eta_{i1}$, $\epsilon_{i2} = \eta_{i2} + 0.5\epsilon_{i1}$, $\epsilon_{i3} = \eta_{i3} + 0.4\epsilon_{i1} + 0.4\epsilon_{i2}$, $\epsilon_{i4} = \eta_{i4} + 0.3\epsilon_{i1} + 0.3\epsilon_{i2} + 0.3\epsilon_{i3}$, $\epsilon_{i5} = \eta_{i5} + 0.1\epsilon_{i1} + 0.1\epsilon_{i2} + 0.1\epsilon_{i3} + 0.1\epsilon_{i4}$, where η_{i1}, η_{i2}, and η_{i3} follow the standard normal distribution, η_{i4} and η_{i5} follow the T-distribution with 3 degrees of freedom, and η_{ij} $(j = 1, \ldots, 5)$ are all independent of each other.

Study 2. We consider the sample size $n = 50, 100, 200, 500$ and the dimension of the parameter $p_n = [2.5n^{1/3}]$, where $[q]$ is the the largest integer not greater than q. In this study, $\beta_0 = (0.4\mathbf{1}_k^T, -0.3\mathbf{1}_k^T, 0.2\mathbf{1}_k^T, -0.1\mathbf{1}_{p_n-3k}^T)^T$, where $\mathbf{1}_k$ denotes a k-dimensional vector of ones and $k = [p_n/4]$. We take the covariate $x_{ij} = (x_{ij1}, \ldots, x_{ijp_n})^T$, which follows the multivariate normal distribution with mean being zero and covariance matrix being $\Sigma^{(2)}$ with $(\Sigma)^{(2)}_{i,j} = 0.5^{|i-j|}$ for $1 \leq i, j \leq p_n$. The error used in this study is identical to that of Study 1. This study is to investigate how the proposed approach performs when the dimension of the parameter p_n grows with the sample size n.

Study 3. The sample size in this study is set to be $n = 100$. For $i = 1, \ldots, 50$, m_i equals 5; for $i = 51, \ldots, 100$, m_i are i.i.d. from a discrete uniform distribution on the set $\{2, 3, 4, 5\}$. We consider $p = 3$ and $p = 11$. The covariates for the case of $p = 3$ and $p = 11$ are generated via the methods of Study 1 and Study 2, respectively. The error $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{im_i})^T$ is generated from the multivariate normal distribution with mean being zero and covariance matrix being $\Sigma^{(3)}_{m_i}$, which is an $m_i \times m_i$ AR-1 correlation matrix with autocorrelation coefficient 0.8.

Study 4. The error $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{im_i})^T$ follows the multivariate normal distribution with mean zero and covariance matrix $\Sigma^{(4)}_{m_i}$, which is an $m_i \times m_i$ exchangeable correlation matrix with all pairs of observations having the common correlation 0.8. The data sets are generated following the procedure in Study 3. The main purpose of Studies 3 and 4 is to investigate the robust performances of the new method to the correlation structures.

Study 5. The error $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{im_i})^T$ is generated from the multivariate T distribution with location parameter being zero, scale matrix being $\Sigma^{(4)}_{m_i}$ and degree of freedom being 3. We generate the data sets following the procedure in Study 3. This study is to investigate the performance of the proposed approach when the error is heavy-tailed.

Study 6. The error $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{im_i})^T$ is generated from the multivariate Cauchy distribution with location parameter being zero and scale matrix being $\Sigma^{(4)}_{m_i}$. We generate the data sets following the procedure in Study 3. This study is designed to investigate how the proposed approach performs when covariance matrix of the error is undefined.

Study 7. The error ϵ_i follows the mixture of multivariate normal distributions, that is, 0.9 $N(\mathbf{0}, \Sigma^{(4)}_{m_i})$ + $0.1N(\mathbf{0}, 100\Sigma^{(4)}_{m_i})$ and the data sets are sampled using the procedure in Study 3. The main purpose of this study is to investigate robustness of the proposed method in terms of resistance to outliers. Studies 3–7 could also serve to investigate how the proposed method works when the true covariance matrix does not

Table 1

The simulated average mean squared errors for the estimated regression parameters using different methods. The independence MPL estimating method is abbreviated as MPL.I

					GEE			QIF	
	n	p	MPL	MPL.I	Ind	Excha	AR-1	Excha	AR-1
Study 1	100	3	0.0037	0.0056	0.0072	0.0060	0.0060	0.0055	0.0056
Study 2	50	9	0.0099	0.0125	0.0132	0.0109	0.0119	0.0113	0.0115
	100	11	0.0046	0.0060	0.0067	0.0057	0.0062	0.0054	0.0057
	200	14	0.0021	0.0027	0.0033	0.0029	0.0030	0.0029	0.0029
	500	19	0.0008	0.0010	0.0015	0.0013	0.0014	0.0011	0.0012
Study 3	100	3	0.0013	0.0034	0.0034	0.0015	0.0009	0.0016	0.0011
	100	11	0.0016	0.0041	0.0039	0.0016	0.0011	0.0020	0.0014
Study 4	100	3	0.0013	0.0032	0.0031	0.0009	0.0012	0.0011	0.0013
	100	11	0.0016	0.0042	0.0040	0.0010	0.0013	0.0014	0.0016
Study 5	100	3	0.0019	0.0064	0.0086	0.0023	0.0032	0.0023	0.0028
	100	11	0.0020	0.0065	0.0115	0.0029	0.0036	0.0033	0.0034
Study 6	100	3	0.0062	0.0143	54.741	31.411	31.302	0.1711	0.1293
	100	11	0.0059	0.0119	182.62	70.859	57.323	0.0401	0.0314
Study 7	100	3	0.0022	0.0059	0.0354	0.0106	0.0136	0.0092	0.0109
	100	11	0.0019	0.0054	0.0435	0.0118	0.0150	0.0071	0.0065

admit an explicit Cholesky decomposition for different error distributions.

The results for the above seven simulation studies are presented in Table 1. First, we see clearly that the proposed MPL method has generally smaller SAMSEs than the GEE and QIF methods for the errors under study. Second, the results based on multivariate normal errors in Studies 3 and 4 seem to be in favor of the GEE and QIF approaches using AR-1 and exchangeable correlation structures, respectively. In these cases, our proposed method produces sightly larger SAMSEs than the GEE and QIF methods using the correct correlation structures. The new method works well without the specification of the correlation matrix and is thus robust regardless of the correlation structure. Third, consistent with the general knowledge in longitudinal data analysis, our proposed method has taken into account the within-subject correlation and thus has smaller SAMSEs than the MPL method based on the independence estimating procedure. Fourth, even if the covariance matrix being studied does not admit a clear Cholesky decomposition, our proposed method continues to perform well in Studies 3–7. Fifth, Study 2 shows that our proposed approach performs satisfactorily when the dimension of the parameter p_n grows with the sample size n. Sixth, the proposed method is resistant to heavy-tailed errors being considered in Study 5. Seventh, when the covariance matrix of the error diverges (e.g., Cauchy distribution in Study 6), our method works well while the GEE method breaks down. Finally, the proposed method is robust against outliers as is seen in Study 7. Our simulation results generally demonstrate excellent robust performance of our proposed approach for longitudinal data analysis.

Next, we examine the accuracy of the estimates of the covariance matrix of $\hat{\beta}^*_{MPL}$. "SD" represents the sample standard deviation over the 200 estimates and is regarded as the true standard error. "SE" represents the sample average of 200 standard errors using the covariance estimating method described at the end of Section 2.2. Table 2 compares SD with SE for all models used above for the case of $n = 100$ and $p = 3$. We observe that the covariance estimating method works remarkably well. Similar conclusions can be drawn for other sample sizes and dimension of covariates being studied and the results are not reported here in order to save space.

Finally, we conduct the following simulation studies to investigate the performance of our proposed method when the covariance matrix of the responses is covariate-dependent, although it is assumed that the covariance matrix is static throughout the article.

Studies 3*. The settings here is identical to that in Studies 3 except that the covariance matrix is set to be $S_i^{1/2}\Sigma_{m_i}^{(3)}S_i^{1/2}$, where $S_i = \mathrm{diag}\Big(\exp(\sum_{l=1}^{p} x_{i1l}/p + z_{i1}), \ldots, \exp(\sum_{l=1}^{p} x_{im_il}/p + z_{im_i})\Big)$, where z_{ij} are identically

Table 2

Comparisons of standard deviation (SD) and estimated standard error (SE) of the MPL estimators for $n = 100$ and $p = 3$

	$\hat{\beta}_1^*$		$\hat{\beta}_2^*$		$\hat{\beta}_3^*$	
	SD	SE	SD	SE	SD	SE
Study 1	0.0536	0.0519	0.0635	0.0671	0.0639	0.0673
Study 3	0.0309	0.0298	0.0370	0.0376	0.0394	0.0382
Study 4	0.0338	0.0324	0.0406	0.0393	0.0402	0.0385
Study 5	0.0371	0.0389	0.0443	0.0461	0.0500	0.0507
Study 6	0.0689	0.0679	0.0799	0.0826	0.0875	0.0876
Study 7	0.0399	0.0398	0.0500	0.0514	0.0495	0.0510

Table 3
The simulated average mean squared errors for the estimated regression parameters for different methods when covariance matrices are covariate-dependent. The independence MPL estimating method is abbreviated as MPL.I

					GEE			QIF	
	n	p	MPL	MPL.I	Ind	Excha	AR-1	Excha	AR-1
Study 3*	100	3	0.0026	0.0048	0.0084	0.0056	0.0055	0.0057	0.0054
	100	11	0.0029	0.0051	0.0079	0.0047	0.0043	0.0047	0.0041
Study 4*	100	3	0.0028	0.0065	0.0083	0.0049	0.0065	0.0050	0.0056
	100	11	0.0025	0.0050	0.0079	0.0038	0.0047	0.0038	0.0042
Study 5*	100	3	0.0037	0.0075	0.0242	0.0151	0.0205	0.0124	0.0135
	100	11	0.0037	0.0067	0.0204	0.0105	0.0129	0.0076	0.0087
Study 6*	100	3	0.0060	0.0107	298.41	166.61	149.47	0.3367	0.2505
	100	11	0.0091	0.0125	5046.7	763.08	1127.5	0.0708	0.0874
Study 7*	100	3	0.0032	0.0066	0.0948	0.0579	0.0663	0.0327	0.0355
	100	11	0.0041	0.0069	0.0877	0.0472	0.0602	0.0148	0.0147

and independently distributed from $N(0, 1)$, for $j = 1, \ldots, m_i$ and $i = 1, \ldots, n$.

The setings in Studies 4*–7* are identical to those in Studies 4–7, except that the covariance matrix is covariate-dependent as in Study 3*. The results are reported in Table 3. In general, the good performances observed in Studies 3–7 still hold when the covariance matrices are covariate-dependent. The results for performances of the covariance estimating method here are reported in the Supplementary Materials.

3.2. *The Diastolic Blood Pressure Data*

The data were collected by the Akdeniz University Hospital Anesthesiology and Reanimation Department during the period of January 2008 to January 2011. There are 375 patients and the diastolic blood pressures (DBP) were observed 9 times for each individual, which were measured every 5 minutes during the surgery. Hypertension is a common clinical problem and a major risk factor for cardiovascular disease and stroke. Due to the lack of evidence supporting heart rate lowering as a therapeutic strategy in hypertension, heart rate is generally not a major consideration in choosing antihypertensive medications (Reule and Drawz, 2012). We are interested in the relationship between the DBP and the pulse, which may help to investigate the chronotropic therapy in hypertension.

The slope coefficient estimates based on the proposed MPL approach, the independence MPL estimating approach, the GEE and QIF approaches using independence, exchangeable, and the AR-1 working correlation structures and their corresponding standard errors are reported in Table 4. We also report the results in Table 4 of MLE estimates by assuming the responses from the i-th subject follows the multivariate normal distribution with mean $(\beta_0 + \beta_1 * pulse_{i1}, \ldots, \beta_0 + \beta_1 * pulse_{i9})$ and covariance $\Sigma_{9\times 9}$, where $pulse_{ij}$ is the pulse value of the j-th observation for subject i, for $i = 1, \ldots, n$. It is interesting to see that an opposite conclusion (i.e., negative slope estimate) may be produced if the dependence nature of the longitudinal data is totally ignored. We also note that the slope estimates from MPL and GEE are very different. This may be due to the fact that the 9-dimensional responses from the 375 subjects are not multivariate-normally distributed, since both p-values of skewness and kurtosis statistics is less than 10^{-5} when the Mardia's multivariate normality test is conducted (Mardia, 1974). Again, the standard errors of the MPL estimates are generally the smallest among all methods being studied. The test for the hypothesis that the DBP and pulse have no relationship yields a p-value of 0.0323, indicating that there is no relationship between DBP and pulse.

We next evaluate the efficiency of the five methods via a bootstrapping method. We randomly choose a total of 200 individuals as training data to fit the model and various methods are used to estimate the regression coefficients and the corresponding standard errors. This procedure is repeated 100 times under the sampling-with-replacement scheme. The results are reported in Table 4. Consistent with our simulation results, our proposed MPL method yields the most efficient estimates. Program codes prepared in R have been developed to implement the methodologies developed in this article and are available from the first author upon request.

4. Discussion

We proposed a novel profile likelihood based method for longitudinal data analysis. The proposed method takes the within-subject correlation into account and works well without specifications of the likelihood as well as the correlation structure. Our theoretical and numerical results show that our proposed methods produce consistent and efficient estimates.

In this article, we only considered monotone missing data patterns and the covariance matrix is assumed to be constant for different subjects in this article. We are currently studying a new profile likelihood based method which could deal with non-monotone missing data patterns, and incorporate the time and covariate information into the covariance structure of the longitudinal observations.

There are several possible directions for future study. First, it is interesting to investigate the theory of the proposed method under the "large n, diverging p" asymptotic framework (Wang, 2011; Lian, Liang, and Wang, 2014; Zhang and Wang, 2016). Second, a large number of time points in lon-

Table 4
The estimates and the corresponding standard errors (SE), and the standard deviations (SD) and the means of standard errors (MSE) over 100 replications for the real data using different methods. The independence MPL estimating method is abbreviated as MPL.I

			GEE			QIF		
	MPL	MPL.I	Ind	Excha	AR-1	Excha	AR-1	MLE
Estimate	0.0274	−0.0008	0.1261	0.3992	0.3149	0.3381	0.3786	0.2265
SE	0.0128	0.0244	0.0464	0.0363	0.0377	0.0353	0.0342	0.0192
Bootstrap								
SD	0.0259	0.0439	0.0473	0.0336	0.0373	0.0435	0.0420	0.0306
MSE	0.0220	0.0464	0.0624	0.0492	0.0490	0.0476	0.0459	0.0259

gitudinal setting may be observed for each subject (Xie and Yang, 2003). It would be of great research interest to consider "large n, diverging m" asymptotic properties for our proposed approach. Third, one could adopt the within-subject correlation via alternative decompositions (Rothman, Levina, and Zhu, 2010; Zhang and Leng, 2012). These topics are beyond the scope of the current article and will be pursued elsewhere.

5. Supplementary Materials

Web Appendices referenced in Sections 2–3, the R code for the proposed procedures are available with this article at the *Biometrics* website on Wiley Online Library.

Acknowledgements

We are grateful to the editor, an associate editor, and two reviewers for their constructive comments that greatly improved the article. Chen's research is supported in part by National Nature Science Foundation of China (Project No. 11401593), Specialized Research Fund for the Doctoral Program of Higher Education of China (Project No. 20130162120086) and Hunan Provincial Natural Science Foundation of China (Project No. 2016JJ3138). Tang's research is fully supported by a grant from the Research Grant Council of the Hong Kong Special Administration Region (Project No. UGC/FDS14/P01/16). Gao's research is supported in part by National Nature Science Foundation of China(No. 11471068).

References

Chen, Z., Tang, M. L., Gao, W., and Shi, N.-Z. (2014). New robust variable selection methods for linear regression models. *Scandinavion Journal of Statistics* **41**, 725–741.

Diggle, P. J., Heagerty, P. J., Liang, K. Y., and Zeger, S. L. (2002). *Analyis of Longitudinal Data*. New York: Oxford University Press.

Duin, R. P. W. (1976). On the choice of smoothing parameters of Parzen estimators of probability density functions. *IEEE Transactions on Computers* **25**, 1175–1179.

Fan, J., Huang, T., and Li, R. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function. *Journal of the American Statistical Association* **102**, 632–641.

Leung, D. H. Y., Wang, Y. G., and Zhu, M. (2009). Efficient parameter estimation in longitudinal data analysis using a hybrid GEE method. *Biostatistics* **10**, 436–445.

Lian, H., Liang, H., and Wang, L. (2014). Generalized additive partial linear models for clustered data with diverging number of covariates using GEE. *Statistic Sinica* **24**, 173–196.

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

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.

Linton, O., Sperlich, S., and Van Keilegom, I. (2008). Estimation of a semiparametric transformation model. *Annals of Statistics* **36**, 686–718.

Lombardia, M. J. and Sperlich, S. (2008). Semiparametric inference in generalized mixed effects models. *Journal of the Royal Statistical Society, Series B* **70**, 913–930.

Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis for testing normality and robustness studies. *Sankhy A* **36**, 115–128.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, R. F. Engle and D. McFadden (eds), **4**, 2111–2245. Amsterdam: Elsevier Science.

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.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Annals of Statistics* **22**, 300–325.

Qu, A., Lindsay, B., and Li, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika* **87**, 823–836.

Reule, S., Drawz, P. E. (2012). Heart rate and blood pressure: Any possible implications for management of hypertension? *Current Hypertension Reports* **14**, 478–484.

Robins, J. M. and Rotnitzky, A. (1995). Semiparanetric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association* **90**, 122–129.

Rothman, A., Levina, L., and Zhu, J. (2010). A new approach to Cholesky-based covariance regularization in high dimensions. *Biometrika* **97**, 539–550.

Rubin, D. B. (1987). *Multiple Imputation for Survey Nonresponse*. New York: Wiley.

Wang, L. (2011). GEE analysis of clustered binary data with diverging number of covariates. *Annals of Statistics* **39**, 389–417.

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.

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.

Xie, M. and Yang, Y. (2003). Asymptotics for generalized estimating equations with large cluster sizes. *The Annals of Statistics* **31**, 310–347.

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

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

Zhang, X. and Wang, J. (2016). From sparse to sense functional data and beyond. *The Annals of Statistics* **44**, 2281–2321.

Zhou, J. and Qu, A. (2012). Informative estimation and selection of correlation structure for longitudinal data. *Journal of the American Statistical Association* **107**, 701–710.

Received January 2016. *Revised April* 2017.
Accepted April 2017.