

# Profile Estimation of Generalized Semiparametric Varying-Coefficient Additive Models for Longitudinal Data with Within-Subject Correlations

Yanqing Sun and Fang Fang

**Abstract** In this paper, we study several profile estimation methods for the generalized semiparametric varying-coefficient additive model for longitudinal data by utilizing the within-subject correlations. The model is flexible in allowing time-varying effects for some covariates and constant effects for others, and in having the option to choose different link functions which can be used to analyze both discrete and continuous longitudinal responses. We investigated the profile generalized estimating equation (GEE) approaches and the profile quadratic inference function (QIF) approach. The profile estimations are assisted with the local linear smoothing technique to estimate the time-varying effects. Several approaches that incorporate the within-subject correlations are investigated including the quasi-likelihood (QL), the minimum generalized variance (MGV), the quadratic inference function and the weighted least squares (WLS). The proposed estimation procedures can accommodate flexible sampling schemes. These methods provide a unified approach that work well for discrete longitudinal responses as well as for continuous longitudinal responses. Finite sample performances of these methods are examined through Monte Carlo simulations under various correlation structures for both discrete and continuous longitudinal responses. The simulation results show efficiency improvement over the working independence approach by utilizing the within-subject correlations as well as comparative performances of different approaches.

## 1 Introduction

The repeated measurements on same individuals over time are common in medical and public health researches. In AIDS clinical trials, for example, the viral load

---

Yanqing Sun

University of North Carolina at Charlotte, Charlotte, NC 28223, e-mail: yasun@uncc.edu

Fang Fang

Corporate Model Risk at Wells Fargo, Charlotte, NC 28202 e-mail: Fang.Fang2@wellsfargo.com

and CD4 cell counts, which are considered as surrogate endpoints for HIV disease progression and HIV transmission to others, are measured repeatedly during the course of studies for trial participants. The repeated measurements in the longitudinal follow-up often display temporal effects and are correlated. We investigate several estimation methods for analyzing longitudinal data under the generalized semiparametric varying-coefficient additive models by incorporating the within subject correlations.

Suppose that there is a random sample of  $n$  subjects. For the  $i$ th subject, let  $Y_i(t)$  be the response at time  $t$  and let  $X_i(t)$  and  $Z_i(t)$  be the possibly time-dependent covariates of dimensions  $p + 1$  and  $q$ , respectively, over the time interval  $[0, \tau]$ , where  $\tau$  is the end of follow-up. Let  $\mu_i(t) = E\{Y_i(t)|X_i(t), Z_i(t)\}$  be the conditional expectation of  $Y_i(t)$  given  $X_i(t)$  and  $Z_i(t)$  at time  $t$ . The generalized semiparametric regression model speculates that

$$\mu_i(t) = g^{-1}\{\alpha^T(t)X_i(t) + \beta^T Z_i(t)\}, \quad i = 1, \dots, n, \quad (1)$$

for  $0 \leq t \leq \tau$ , where  $g(\cdot)$  is a known link function,  $\alpha(t)$  is a  $(p + 1)$ -dimensional vector of unspecified functions and  $\beta$  is a  $q$ -dimensional vector of unknown parameters. The notation  $\theta^T$  represents transpose of a vector or matrix  $\theta$ . When the link function  $g(\cdot)$  is the identity function, model (1) is known as the semiparametric additive model. When the link function is the natural logarithm function and  $X_i(t) = 1$ , model (1) is known as the proportional means model. Setting the first component of  $X_i(t)$  as 1 gives a nonparametric baseline function. Under model (1), the effects of some covariates are constant while others are time-varying. Model (1) is more flexible than the parametric regression model where all the regression coefficients are time-independent and more desirable for model building than the nonparametric regression model where every covariate effect is an unspecified function of time. Different link functions can be selected to provide a rich family of models for longitudinal data. Both the categorical and continuous longitudinal responses can be modelled with appropriately chosen link functions. For example, the identity and logarithm link functions can be used for the continuous response variables while the logit link function can be used for the binary responses.

The semiparametric additive model for longitudinal data has been studied extensively for decades. These approaches include the nonparametric kernel smoothing by Hoover et al. (1998), the joint modelling of longitudinal responses and sampling times by Martinussen and Scheike (1999) and Lin and Ying (2001), the backfitting method by Zeger and Diggle (1994) and Wu and Liang (2004), and the profile kernel smoothing approach by Sun and Wu (2005). Fan and Li (2004) considered the profile local linear approach and the joint modelling for partially linear models. Hu et al. (2004) showed that for partially linear models, the backfitting is less efficient than the profile kernel method. Sun et al. (2013) investigated the generalized semiparametric additive model (1) using the local linear profile estimation method. The aforementioned estimation and inference procedures are derived without considering the correlations of longitudinal responses within subjects known as the working

independence approach. The estimation methods under the working independence are valid and yield asymptotically unbiased estimators.

Correlation among repeated measurements on the same subject often exists for longitudinal data or clustered data. Incorporating such within-subject correlation into estimation procedure can lead to improved efficiency. Liang and Zeger (1986) introduced the idea of using a working correlation matrix with a small set of nuisance parameters to avoid specification of correlation between measurements within the cluster. Severini and Staniswalis (1994) and Lin and Carroll (2001a,b) estimated  $\alpha(t)$  using the kernel method by ignoring the within-subject correlation while estimating  $\beta$  using weighted least squares by accounting for the within-subject correlation when  $X_i(t) \equiv 1$ . Chen and Jin (2006) studied the method of generalized estimating equations by modeling the within-cluster correlation. Using piecewise local polynomial approximation of  $\alpha(t)$ , Chen and Jin (2006) showed that the weighted least square estimator of  $\beta$  achieves the semiparametric efficiency. Fan et al. (2007) proposed a profile local linear approach by imposing certain correlation structure for the longitudinal data for improved efficiency. Fan et al. (2007) proposed two methods to estimate for the within-subject correlation by optimizing the quasi-likelihood (QL) and by minimizing the generalized variance of the estimator of  $\beta$  (MGV). Following the generalized method of moments of Hansen (1982), Qu et al. (2000) proposed the quadratic inference function method (QIF) by representing the inverse of working correlation matrix by a linear combination of basis matrices. Song et al. (2009), Madsen et al. (2011) and Tang et al. (2019) studied a mean-correlation parametric regression method for a family of discrete longitudinal responses by assuming that the marginal distributions of longitudinal responses follow an exponential family distribution and the joint distributions of the discrete responses from the same subject are modeled by a copula model. These approaches have a limitation of not allowing for time-varying covariate effects.

Semiparametric statistical modeling of discrete longitudinal responses beyond the marginal approach has been understudied. We investigate several profile estimation methods for the generalized semiparametric varying-coefficient additive model (1) by incorporating the within-subject correlations including the profile generalized estimating equation (GEE) approaches and the quadratic inference function approach. These methods provide a unified approach that work well for discrete longitudinal responses as well as for continuous longitudinal responses. Different methods for estimating the within-subject correlations such as the QL and MGV methods as well as a newly proposed profile weighted least square (WLS) approach fall under the umbrella of profile GEE approaches. The performances of these different methods are examined through extensive simulation studies under a variety of models and the within-subject correlation structures. The proposed semiparametric methods utilizing the within-subject correlations work well for discrete longitudinal responses as well as for continuous longitudinal responses.

The rest of the paper is organized as follows. The profile GEE estimation using fixed working covariance matrices is presented in Section 2.1. The methods for estimating the correlations are described in Section 2.2. An alternative profile estimation of model (1) via quadratic inference function is proposed in Section

2.3. The computational algorithms of the proposed procedures are summarized in Section 2.4. Section 3 presents the results of simulation studies for evaluating the finite sample performances of different methods. The results of simulation studies for continuous longitudinal responses are presented in Section 3.1 and the results of simulation studies for discrete longitudinal responses are given in Section 3.2. Some concluding remarks are given in Section 4.

## 2 Profile GEE Estimation Procedures

This section presents several profile estimation methods for the generalized semiparametric varying-coefficient additive model (1) by incorporating the within-subject correlations and the approaches for estimating the within-subject correlations. Choices of kernel function, bandwidth and link function are also discussed.

### 2.1 Model estimation using fixed working covariance matrices

Suppose that the longitudinal response  $Y_i(t)$  and the possibly time-dependent covariates  $X_i(t)$  and  $Z_i(t)$  are observed at the sampling times  $T_{i1} < T_{i2} < \dots < T_{iJ_i}$ , where  $J_i$  is the total number of observations on the  $i$ th subject. Let  $Y_{ij} = Y_i(T_{ij})$ ,  $X_{ij} = X_i(T_{ij})$  and  $Z_{ij} = Z_i(T_{ij})$ . Let  $Y_i = (Y_{i1}, \dots, Y_{iJ_i})^T$  be the vector of responses for individual  $i$ . Similarly, define  $X_i = (X_{i1}, \dots, X_{iJ_i})^T$ ,  $Z_i = (Z_{i1}, \dots, Z_{iJ_i})^T$  and  $T_i = (T_{i1}, \dots, T_{iJ_i})$ . The sampling times  $\{T_{ij}, j = 1, \dots, J_i\}$  varies among individuals under random designs, while they are not depend on  $i$  under fixed designs. We propose the kernel assisted profile method to estimate the nonparametric functions  $\alpha(t)$  and parametric coefficients  $\beta$  under model (1) by taking into consideration of the within-subject correlations.

For given  $\beta$ , let  $\alpha(t) = \alpha(t_0) + \dot{\alpha}(t_0)(t - t_0) + O((t - t_0)^2)$  be the first order Taylor expansion of  $\alpha(\cdot)$  for  $t \in \mathcal{N}_{t_0}$ , a neighborhood of  $t_0$ , where  $\dot{\alpha}(t_0)$  is the derivative of  $\alpha(t)$  at  $t = t_0$ . Denote  $\alpha^*(t_0) = (\alpha^T(t_0), \dot{\alpha}^T(t_0))^T$  and  $X_i^*(t, t - t_0) = X_i(t) \otimes (1, t - t_0)^T$ , where  $\otimes$  is the Kronecker product. Then for  $t \in \mathcal{N}_{t_0}$ , model (1) can be approximated by

$$\tilde{\mu}(t, t_0, \alpha^*(t_0), \beta | X_i(t), Z_i(t)) = g^{-1}\{\alpha^{*T}(t_0)X_i^*(t, t - t_0) + \beta^T Z_i(t)\}. \quad (2)$$

Let  $X_{ij}^*(t_0) = X_{ij} \otimes (1, T_{ij} - t_0)^T$ ,  $j = 1, \dots, J_i$ . The approximated conditional expectation of  $Y_{ij}$  for  $T_{ij} \in \mathcal{N}_{t_0}$  is given by  $\mu_{ij}^*(t_0) = \mu\{\alpha^{*T}(t_0)X_{ij}^*(t_0) + \beta^T Z_{ij}\}$ , where  $\mu(\cdot) = g^{-1}(\cdot)$ . Denote  $\dot{\mu}_{ij}^*(t_0) = \dot{\mu}\{\alpha^{*T}(t_0)X_{ij}^*(t_0) + \beta^T Z_{ij}\}$  where  $\dot{\mu}(\cdot)$  is the first derivative of  $\mu(\cdot)$ . Let  $\mu_i^*(t_0) = (\mu_{i1}^*(t_0), \dots, \mu_{iJ_i}^*(t_0))^T$ . Let  $X_i^*(t_0)$  denote a  $2(p+1) \times J_i$  matrix with the  $j$ th column vector being the  $X_{ij}^*(t_0)$ ,  $j = 1, \dots, J_i$ .

Let  $K(\cdot)$  be a nonnegative kernel function and  $h = h_n > 0$  a bandwidth parameter. Let  $K_{ih}(t_0) = \text{diag}\{K_h(T_{ij} - t_0), j = 1, \dots, J_i\}$  be the  $J_i \times J_i$  diagonal matrix with  $\{K_h(T_{ij} - t_0), j = 1, \dots, J_i\}$ , on the diagonal and zero elsewhere, where  $K_h(\cdot) =$

$K(\cdot/h)/h$ . At each  $t_0$  and for fixed  $\beta$ , we consider the following local linear estimating function for  $\alpha^*(t_0)$ :

$$U_\alpha(\alpha^*; \beta, t_0) = \sum_{i=1}^n X_i^*(t_0) \Delta_i^*(t_0) K_{ih}^{1/2}(t_0) V_{li}^{-1} K_{ih}^{1/2}(t_0) [Y_i - \mu_i^*(t_0)], \quad (3)$$

where  $\Delta_i^*(t_0) = \text{diag}\{\dot{\mu}_{ij}^*(t_0), j = 1, \dots, J_i\}$  and  $V_{li}^{-1}$  is the inverse of the working covariance matrix for estimating  $\alpha^*(t_0)$ . The solution to the equation  $U_\alpha(\alpha^*; \beta, t_0) = 0$  is denoted by  $\tilde{\alpha}^*(t_0, \beta)$ . We denote the first  $p+1$  components of  $\tilde{\alpha}^*(t_0, \beta)$  by  $\tilde{\alpha}(t_0, \beta)$ .

Let  $\tilde{\mu}_{ij}(\beta) = \mu\{\tilde{\alpha}^T(T_{ij}, \beta)X_{ij} + \beta^T Z_{ij}\}$  and  $\tilde{\mu}_i(\beta) = (\tilde{\mu}_{i1}(\beta), \dots, \tilde{\mu}_{iJ_i}(\beta))^T$ . The profile weighted least-squares estimator  $\hat{\beta}$  is obtained by minimizing the following profile least-squares function:

$$\ell_\beta(\beta) = \frac{1}{n} \sum_{i=1}^n [Y_i - \tilde{\mu}_i(\beta)]^T V_{2i}^{-1} [Y_i - \tilde{\mu}_i(\beta)], \quad (4)$$

where  $V_{2i}^{-1}$  is the inverse of the working covariance matrix for estimating  $\beta$ .

Let  $A_{ij} = \partial \tilde{\alpha}^T(T_{ij}, \beta) / \partial \beta$  be the derivative of  $\tilde{\alpha}^T(T_{ij}, \beta)$  with respect to  $\beta$ , which is a  $q \times (p+1)$  matrix with the  $k$ th row having the partial derivative of  $\tilde{\alpha}^T(T_{ij}, \beta)$  with respect to the  $k$ th component of  $\beta_k$ ,  $1 \leq k \leq q$ . Let  $\frac{\partial \tilde{\alpha}^T(T_i, \beta)}{\partial \beta} = (\frac{\partial \tilde{\alpha}^T(T_{i1}, \beta)}{\partial \beta}, \dots, \frac{\partial \tilde{\alpha}^T(T_{iJ_i}, \beta)}{\partial \beta})$  and  $\tilde{X}_i = \text{diag}\{X_{ij}, j = 1, \dots, J_i\}$ . Then  $\frac{\partial \tilde{\alpha}^T(T_i, \beta)}{\partial \beta} \tilde{X}_i = (\frac{\partial \tilde{\alpha}^T(T_{i1}, \beta)}{\partial \beta} X_{i1}, \dots, \frac{\partial \tilde{\alpha}^T(T_{iJ_i}, \beta)}{\partial \beta} X_{iJ_i})$  is a  $q \times J_i$  matrix.

Taking the derivative of  $\ell_\beta(\beta)$  with respect to  $\beta$ , we have the score function

$$U_\beta(\beta) = \sum_{i=1}^n \left\{ \frac{\partial \tilde{\alpha}^T(T_i, \beta)}{\partial \beta} \tilde{X}_i + \tilde{Z}_i \right\} \tilde{\Delta}_i V_{2i}^{-1} [Y_i - \tilde{\mu}_i(\beta)], \quad (5)$$

where  $\tilde{\Delta}_i = \text{diag}\{\dot{\mu}_{ij}, j = 1, \dots, J_i\}$ ,  $\dot{\mu}_{ij} = \dot{\mu}\{\tilde{\alpha}^T(T_{ij}, \beta)X_{ij} + \beta^T Z_{ij}\}$ , and  $\tilde{Z}_i = (Z_{i1}, \dots, Z_{iJ_i})$  is a  $q \times J_i$  matrix.

For given working covariance matrices  $V_{li}$  and  $V_{2i}$ , the profile GEE estimator  $\hat{\beta}$  of  $\beta$  is obtained by solving the estimating equation  $U_\beta(\beta) = 0$ . The profile GEE estimator for  $\alpha(t)$  is given by  $\hat{\alpha}(t) = \tilde{\alpha}(t, \hat{\beta})$ .

Note that  $\partial \tilde{\alpha}^T(t, \beta) / \partial \beta$  is the first  $p+1$  columns of  $\partial \alpha^{*T}(t, \beta) / \partial \beta$ . Next we show that  $\partial \alpha^{*T}(t, \beta) / \partial \beta$  can be expressed in terms of the partial derivatives of  $U_\alpha(\alpha^*; \beta, t)$  at  $\alpha^* = \tilde{\alpha}^*(t, \beta)$ . Specifically, since  $U_\alpha(\tilde{\alpha}^*(t, \beta); \beta, t) \equiv \mathbf{0}_{2(p+1)}$  by (3), it follows that  $\tilde{\alpha}^*(t, \beta)$  satisfies

$$\left\{ \frac{\partial U_\alpha(\alpha^*; \beta, t)}{\partial \alpha^*} \frac{\partial \tilde{\alpha}^{*T}(t, \beta)}{\partial \beta} + \frac{\partial U_\alpha(\alpha^*; \beta, t)}{\partial \beta} \right\} \Big|_{\alpha^* = \tilde{\alpha}^*(t, \beta)} = \mathbf{0}_{2(p+1)}.$$

Therefore,

$$\frac{\partial \tilde{\alpha}^{*T}(t, \beta)}{\partial \beta} = - \left\{ \frac{\partial U_{\alpha}(\alpha^*; \beta, t)}{\partial \alpha^*} \right\}^{-1} \frac{\partial U_{\alpha}(\alpha^*; \beta, t)}{\partial \beta} \Big|_{\alpha^* = \tilde{\alpha}^*(t, \beta)}, \quad (6)$$

where

$$\frac{\partial U_{\alpha}(\alpha^*; \beta, t)}{\partial \alpha^*} = - \sum_{i=1}^n X_i^*(t) \Delta_i^*(t) K_{ih}^{1/2}(t) V_{li}^{-1} K_{ih}^{1/2}(t) \Delta_i^*(t) X_i^{*T}(t), \quad (7)$$

and

$$\frac{\partial U_{\alpha}(\alpha^*; \beta, t)}{\partial \beta} = - \sum_{i=1}^n X_i^*(t) \Delta_i^*(t) K_{ih}^{1/2}(t) V_{li}^{-1} K_{ih}^{1/2}(t) \Delta_i^*(t) \tilde{Z}_i^T. \quad (8)$$

Under the identity link in model (1),  $\tilde{\alpha}^*(t, \beta)$  and  $\hat{\beta}$  can be solved explicitly as the roots of the score functions (3) and (5), respectively. When there are no explicit solutions, the Newton-Raphson iterative algorithm can be used to solve the equations. The estimation procedure iteratively updates estimates of the nonparametric component  $\tilde{\alpha}^*(t, \beta)$  and the parametric component  $\hat{\beta}$  until convergence. We denote the first  $p + 1$  components of the convergent  $\tilde{\alpha}^*(t, \beta)$  as  $\hat{\alpha}(t)$ .

Let  $\hat{\mu}_{ij} = \hat{\mu}\{\hat{\alpha}^T(T_{ij})X_{ij} + \beta^T Z_{ij}\}$  and  $\hat{\Delta}_i = \text{diag}\{\hat{\mu}_{ij}\}$ . Define  $\hat{E}_{11}(t) = n^{-1} \sum_{i=1}^n X_i \hat{\Delta}_i K_{ih}^{1/2}(t) V_{li}^{-1} K_{ih}^{1/2}(t) \hat{\Delta}_i X_i^T$  and  $\hat{E}_{12}(t) = n^{-1} \sum_{i=1}^n X_i \hat{\Delta}_i K_{ih}^{1/2}(t) V_{li}^{-1} K_{ih}^{1/2}(t) \hat{\Delta}_i Z_i^T$ . Let  $\hat{B}_{ij} = -\hat{E}_{12}^T(T_{ij}) \hat{E}_{11}^{-1}(T_{ij}) X_{ij} + Z_{ij}$  and  $\hat{B}_i = (\hat{B}_{i1}, \dots, \hat{B}_{iJ_i})^T$ . Following the derivations in Fan et al. (2007), we estimate the variance of  $\hat{\beta}$  by  $\hat{P}^{-1} \hat{D} \hat{P}^{-1}$  for given covariance matrices  $V_{1i}$  and  $V_{2i}$ , where

$$\hat{P} = n^{-1} \sum_{i=1}^n [\hat{B}_i^T \hat{\Delta}_i V_{2i}^{-1} \hat{\Delta}_i \hat{B}_i],$$

and

$$\hat{D} = n^{-1} \sum_{i=1}^n [\hat{B}_i^T \hat{\Delta}_i V_{2i}^{-1} (Y_i - \hat{\mu}_i)(Y_i - \hat{\mu}_i)^T V_{2i}^{-1} \hat{\Delta}_i \hat{B}_i].$$

## 2.2 Estimation of the within-subject covariance matrix

The conditional within-subject correlation of longitudinal responses  $Y_i(\cdot)$  at times  $s, t \in [0, \tau]$  can be measured by the Pearson correlation coefficient  $\rho_i(s, t) = \text{Corr}(Y_i(s), Y_i(t) | X_i(\cdot), Z_i(\cdot)) = \text{Cov}(Y_i(s), Y_i(t) | X_i(\cdot), Z_i(\cdot)) / (\sigma_i(s) \sigma_i(t))$ , where  $\sigma_i(t)$  be the conditional standard deviation of  $Y_i(t)$  given  $X_i(t)$  and  $Z_i(t)$ ,  $0 \leq t \leq \tau$ . For simplicity, we assume that both  $\sigma_i(t)$  and  $\rho_i(s, t)$  do not depend on the covariates  $X_i(\cdot)$  and  $Z_i(\cdot)$ . Thus we use the notations  $\sigma(t)$  and  $\rho(s, t)$  in place of  $\sigma_i(t)$  and  $\rho_i(s, t)$ , respectively. In practice, the correlation structure  $\rho(s, t)$  is often unknown or complex, and a working correlation is employed by assuming a correlation model for  $\rho(s, t)$ .

The working independence corresponds to assuming  $\rho(s, t) = 0$  for  $s \neq t$ . Other commonly used correlation models include the compound symmetry or exchangeable structure (Exchangeable) with  $\rho(s, t) = \theta$ ,  $|\theta| < 1$ ; a generalized the first-order autoregressive (AR(1)) with  $\rho(s, t) = \theta^{|s-t|}$ ,  $0 < \theta < 1$ , which is a generalization of AR(1) model in time series to allow the possibility of unequally-spaced times; and a generalization of the first-order autoregressive moving-average (ARMA(1,1)) with  $\rho(s, t) = pq^{|s-t|}$ , where  $|p| < 1$  and  $q > 0$ . Fan et al. (2007) considered more complex correlation structure by embedding the working correlation into a collection of the correlation families  $\rho_0(s, t, \theta_0), \dots, \rho_m(s, t, \theta_m)$ :

$$\rho(s, t, \theta) = b_0\rho_0(s, t, \theta_0) + b_1\rho_1(s, t, \theta_1) + \dots + b_m\rho_m(s, t, \theta_m), \quad (9)$$

where  $\theta = (\theta_0, b_0, \theta_1, b_1, \dots, b_m, \theta_m)$  and  $b_0 + \dots + b_m = 1$  with all  $b_i \geq 0$ .

Let  $\rho_k(s, t, \theta)$ ,  $\theta \in \Theta$ , be the working correlation function for  $Y_i(t)$ ,  $0 \leq t \leq \tau$ , for  $k = 1, 2$ . We consider decomposition of the working covariance  $V_{ki}$  of  $(Y_{i1}, \dots, Y_{iJ_i})$  into

$$V_{ki} = A_i R_{ki}(\theta) A_i, \quad (10)$$

where  $A_i = \text{diag}\{\sigma(T_{ij}), j = 1, \dots, J_i\}$ , and  $R_{ki}(\theta)$  is the working correlation matrix of  $(Y_{i1}, \dots, Y_{iJ_i})$  under the working correlation model  $\rho_k(s, t, \theta)$  for  $k = 1, 2$ .

The examples of correlation matrices  $R_i(\theta)$  of  $(Y_{i1}, \dots, Y_{iJ_i})$  at the measurement times  $t_1, \dots, t_{J_i}$  for  $J_i = 4$  for Exchangeable, AR(1) and ARMA(1,1) correlations are shown in the following:

$$\begin{array}{ccc} \begin{bmatrix} 1 & \theta & \theta & \theta \\ & 1 & \theta & \theta \\ & & 1 & \theta \\ & & & 1 \end{bmatrix}, & \begin{bmatrix} 1 & \theta^{|t_1-t_2|} & \theta^{|t_1-t_3|} & \theta^{|t_1-t_4|} \\ & 1 & \theta^{|t_2-t_3|} & \theta^{|t_2-t_4|} \\ & & 1 & \theta^{|t_3-t_4|} \\ & & & 1 \end{bmatrix}, & \begin{bmatrix} 1 & pq^{|t_1-t_2|} & pq^{|t_1-t_3|} & pq^{|t_1-t_4|} \\ & 1 & pq^{|t_2-t_3|} & pq^{|t_2-t_4|} \\ & & 1 & pq^{|t_3-t_4|} \\ & & & 1 \end{bmatrix}. \\ \text{(a) Exchangeable} & \text{(b) AR(1)} & \text{(c) ARMA(1, 1)} \end{array}$$

The GEE estimation of the regression coefficients are consistent even when the true correlation matrix is not an element of the class of working correlation matrices, and are efficient when the working correlation is correctly specified (Liang and Zeger, 1986). Lin and Carroll (2000) showed that the most efficient estimation of the nonparametric component  $\alpha(t)$  can be achieved by ignoring the within-subject correlation. However, more efficient estimation for the parametric component  $\beta$  is obtained by letting  $V_{2i}$  in (5) to be the inverse of true covariance matrix of  $Y_i$ ; see Lin and Carroll (2001a,b), Wang et al. (2005), and Fan et al. (2007). Thus we set  $R_{1i}(\theta)$  to be the identity matrix and focus on discussing the approaches for estimating  $A_i$  and  $R_{2i}(\theta)$ . For convenience, we use the notation  $\rho(s, t, \theta)$  for  $\rho_2(s, t, \theta)$  and  $R_i(\theta)$  for  $R_{2i}(\theta)$ .

### 2.2.1 Estimation of marginal variance

Let  $\hat{\alpha}_0(t)$  and  $\hat{\beta}_0$  be the marginal estimators of  $\alpha(t)$  and  $\beta$  in Section 2 by setting  $V_{ki}$  to the identity matrix for  $k = 1, 2$ . Define the residual  $\hat{r}_{ij} = Y_{ij} - \hat{\mu}_{ij}$ , where  $\hat{\mu}_{ij} = g^{-1}\{\hat{\alpha}_0^T(T_{ij})X_{ij} + \hat{\beta}_0^T Z_{ij}\}$ . Following Fan et al. (2007), we estimate the marginal variance of response  $Y_i(t)$  when it is continuous using kernel smoothing:

$$\hat{\sigma}^2(t) = \frac{\sum_{i=1}^n \sum_{j=1}^{J_i} \hat{r}_{ij}^2 K_h^*(t - T_{ij})}{\sum_{i=1}^n \sum_{j=1}^{J_i} K_h^*(t - T_{ij})}, \quad (11)$$

where  $K_h^*(\cdot) = K^*(\cdot/h)/h$ ,  $K^*(\cdot)$  is a nonnegative kernel function and  $h = h_n > 0$  a bandwidth parameter.

When the response  $Y_i(t)$  is a discrete random variable, the variance estimation can take different form to account for the model structure of the particular distribution family. For example,  $\hat{r}_{ij}^2$  is replaced by  $\hat{\mu}_{ij}(1 - \hat{\mu}_{ij})$  if the response  $Y_i(t)$  is a Bernoulli random variable, and by  $\hat{\mu}_{ij}$  if  $Y_i(t)$  is a Poisson random variable. We refer to Liang and Zeger (1986) for the relationship between variance and the model parameters when marginal distribution of  $Y_i(t)$  belongs to an exponential family.

### 2.2.2 Estimation of correlation coefficients

We study different approaches to estimate  $\theta$  of the correlation matrix  $R(\theta)$ . Two of the methods, the quasi-likelihood approach and the minimum generalized variance approach, were adopted from Fan et al. (2007) for model (1) with the identity link function. We also propose the minimum weighted least squares approach to estimate  $\theta$ .

The QL estimation of  $\theta$  is obtained by maximizing the quasi-likelihood function:

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \left( -\frac{1}{2} \sum_{i=1}^n \{ \log |R_i(\theta)| + \hat{r}_i^T \hat{A}_i^{-1} R_i^{-1}(\theta) \hat{A}_i^{-1} \hat{r}_i \} \right), \quad (12)$$

where  $R_i(\theta)$  and  $\hat{A}_i = \text{diag}\{\hat{\sigma}(T_{ij}), j = 1, \dots, J_i\}$  are defined the same as in equation (10),  $\hat{r}_i = \{\hat{r}_{i1}, \dots, \hat{r}_{iJ_i}\}$  is the estimator for vector  $\epsilon_i$  and  $\hat{r}_{ij}$  are defined above.

Let  $\Sigma_{\hat{\beta}}(\hat{\sigma}^2, \theta)$  be the estimated covariance matrix of  $\hat{\beta}$  under the working correlation model  $\rho_k(s, t, \theta)$ , which depends on the estimated marginal variance  $\hat{\sigma}^2$  and the correlation parameter vector  $\theta$ . Defining the generalized variance of  $\hat{\beta}$  as the determinant  $|\Sigma_{\hat{\beta}}(\hat{\sigma}^2, \theta)|$  of  $\Sigma_{\hat{\beta}}(\hat{\sigma}^2, \theta)$ . By (Dempster, 1969, Section 3.5), the volume of the ellipsoid of  $(\hat{\beta} - \beta)^T \Sigma_{\hat{\beta}}^{-1}(\hat{\sigma}^2, \theta) (\hat{\beta} - \beta) < c$  for any positive constant  $c$  equals  $\pi^{q/2} c^{1/2} |\Sigma_{\hat{\beta}}(\hat{\sigma}^2, \theta)|^{1/2} / \Gamma(\frac{q}{2} + 1)$ , where  $\Gamma(\cdot)$  is the gamma function. It follows that minimizing the volume of the confidence ellipsoid of  $(\hat{\beta} - \beta)^T \Sigma_{\hat{\beta}}^{-1}(\hat{\sigma}^2, \theta) (\hat{\beta} - \beta) < c$  over  $\theta \in \Theta$  is equivalent to minimizing  $|\Sigma_{\hat{\beta}}(\hat{\sigma}^2, \theta)|$  for  $\theta \in \Theta$  and that the minimizer of the volume of the confidence ellipsoid over  $\theta \in \Theta$  is not affected by  $c$ . Here  $c$  can



be viewed as a constant associated with a confidence level. The MGCV estimation of  $\theta$  by Fan et al. (2007) is obtained by minimizing the generalized variance of  $\hat{\beta}$ :

$$\hat{\theta} = \arg \min_{\theta \in \Theta} |\Sigma_{\hat{\beta}}(\hat{\sigma}^2, \theta)|. \quad (13)$$

Following the idea of the quasi-likelihood approach of Fan et al. (2007), we also study estimation of  $\theta$  obtained by minimizing the weighted least squares:

$$\hat{\theta} = \arg \min_{\theta} (\hat{r}_i^T \hat{A}_i^{-1} R_i^{-1}(\theta) \hat{A}_i^{-1} \hat{r}_i), \quad (14)$$

### 2.3 Profile estimation via quadratic inference function

Qu et al. (2000) proposed the method of quadratic inference functions that does not involve direct estimation of the correlation parameter. The idea is to represent the inverse of the working correlation matrix by the linear combination of basis matrices:

$$R^{-1} \approx a_1 M_1 + a_2 M_2 + \cdots + a_K M_K, \quad (15)$$

where  $M_1$  is the identity matrix, and  $M_2, \dots, M_K$  are symmetric matrices, and  $a_1, \dots, a_K$  are constant coefficients. The representation is applicable to many commonly used working correlations (Qu et al., 2000). For example, if the correlation structure exchangeable, then  $R(\theta)$  has 1's on the diagonal, and  $\theta$ 's everywhere off the diagonal. The inversion  $R^{-1}$  can be written as  $a_1 M_1 + a_2 M_2$ , where  $M_1$  is the identity matrix, and  $M_2$  is a matrix with 0 on the diagonal and 1 off the diagonal. For the AR(1) correlation with  $\rho(s, t) = \theta^{|s-t|}$ , the inversion  $R^{-1}$  of a  $J \times J$  correlation matrix can be written as a linear combination of three basis matrices, where  $M_1$  is the identity matrix, and  $M_2$  has 1 on the two main off-diagonals and 0 elsewhere, and  $M_3$  has 1 on the corners (1, 1) and (J, J), and 0 elsewhere.

Applying the QIF approach, we propose an alternative profile estimation of model (1). We replace the GEE estimator of  $\beta$  that solves  $U_{\beta}(\beta) = 0$  in Section 2 by the estimator that minimizes the quadratic inference function while keep the estimation for  $\tilde{\alpha}(t, \beta)$  as the root of (3) unchanged. Applying idea of the QIF, we define the 'extended score' function:

$$g_n(\beta) = \frac{1}{n} \sum_{i=1}^n g_i(\beta) = \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} \left\{ \frac{\partial \tilde{\alpha}^T(T_i, \beta)}{\partial \beta} \tilde{X}_i + \tilde{Z}_i \right\} \Delta_i \hat{A}_i^{-1/2} M_1 \hat{A}_i^{-1/2} [Y_i - \tilde{\mu}_i(\beta)] \\ \vdots \\ \left\{ \frac{\partial \tilde{\alpha}^T(T_i, \beta)}{\partial \beta} \tilde{X}_i + \tilde{Z}_i \right\} \Delta_i \hat{A}_i^{-1/2} M_K \hat{A}_i^{-1/2} [Y_i - \tilde{\mu}_i(\beta)] \end{pmatrix}. \quad (16)$$

The quadratic inference function is defined as  $Q_n(\beta) = g_n^T(\beta) C_n^{-1}(\beta) g_n(\beta)$ , where  $C_n(\beta) = (1/n^2) \sum_{i=1}^n g_i(\beta) g_i^T(\beta)$ . The profile QIF estimator is the minimizer of  $Q_n(\beta)$ :

$$\hat{\beta} = \arg \min_{\beta} Q_n(\beta) \quad (17)$$

Following the derivations of the asymptotic properties shown in Qu et al. (2000), we estimate the variance of the QIF estimator  $\hat{\beta}$  by  $\{\dot{g}_n(\hat{\beta})C_n^{-1}(\hat{\beta})\dot{g}_n^T(\hat{\beta})\}^{-1}$ , where

$$\dot{g}_n(\beta) = \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} \left\{ \frac{\partial \tilde{\alpha}^T(T_i, \beta)}{\partial \beta} \tilde{X}_i + \tilde{Z}_i \right\} \hat{\Delta}_i \hat{A}_i^{-1/2} M_1 \hat{A}_i^{-1/2} \hat{\Delta}_i \left\{ \frac{\partial \tilde{\alpha}(T_i, \beta)}{\partial \beta} \tilde{X}_i + \tilde{Z}_i \right\}^T \\ \vdots \\ \left\{ \frac{\partial \tilde{\alpha}^T(T_i, \beta)}{\partial \beta} \tilde{X}_i + \tilde{Z}_i \right\} \hat{\Delta}_i \hat{A}_i^{-1/2} M_K \hat{A}_i^{-1/2} \hat{\Delta}_i \left\{ \frac{\partial \tilde{\alpha}(T_i, \beta)}{\partial \beta} \tilde{X}_i + \tilde{Z}_i \right\}^T \end{pmatrix}. \quad (18)$$

## 2.4 Computational algorithms

The iterative algorithms of the procedures using the QL, MGW, WLS and QIF approaches for estimating  $\alpha(t)$  and  $\beta$  under model (1) are outlined in the following.

1. Calculate the estimates of  $\alpha(t)$  and  $\beta$  using the working independence approach and use them as the initial estimates  $\hat{\alpha}^{(0)}(t)$  and  $\hat{\beta}^{(0)}$ ;
2. Given the  $m$ -step estimates  $\hat{\alpha}^{(m)}(t)$  and  $\hat{\beta}^{(m)}$ , calculate  $\hat{r}_{ij} = Y_{ij} - g^{-1}\{(\hat{\alpha}^{(m)}(T_{ij}))^T X_{ij} + (\hat{\beta}^{(m)})^T Z_{ij}\}$  and obtain the matrix  $\hat{A}_i^{(m)}$  whose diagonal elements are estimated by (11);
3. For the QL, MGW and WLS approaches for estimating the correlation matrix, obtain the estimate  $\hat{\theta}^{(m)}$  using one of the QL, MGW and WLS methods described in Section 2.2.2; Set  $\hat{V}_{2i}^{(m)} = \hat{A}_i^{(m)} R_i(\hat{\theta}^{(m)}) \hat{A}_i^{(m)}$  as in (10); Then update the estimate of  $\beta$  to  $\hat{\beta}^{(m+1)}$  by solving (5) and the estimate of  $\alpha(t)$  to  $\hat{\alpha}^{(m+1)}(t) = \tilde{\alpha}(t, \hat{\beta}^{(m+1)})$ ;
4. For the QIF approach, update the estimate of  $\beta$  to  $\hat{\beta}^{(m+1)}$  obtained by minimizing  $Q_n(\beta) = g_n^T(\beta) C_n^{-1}(\beta) g_n(\beta)$  where and  $\hat{A}_i$  in  $g_n(\beta)$  is replaced by  $\hat{A}_i^{(m)}$ , and then update the estimate of  $\alpha(t)$  to  $\hat{\alpha}^{(m+1)}(t) = \tilde{\alpha}(t, \hat{\beta}^{(m+1)})$ ;
5. Repeating steps 2 to 4 until convergence, which is usually achieved within a few iterations.

## 2.5 Choices of kernel function, bandwidth and link function

We employ local linear techniques to estimate the nonparametric time-varying effects  $\alpha(t)$ . The kernel function is designed to give greater weight to observations with sampling time near  $t$  than those further away. In kernel density estimation, the Epanechnikov kernel function  $K(x) = \frac{3}{4}(1 - x^2)_+$  is asymptotically optimal with the smallest mean integrated squared error among probability density functions. (Silverman, 1986, p.43) showed that there is not much variation in the efficiency in the choice of kernel function: the asymptotic relative efficiency of the Tukey kernel function  $K(x) = \frac{15}{16}(1 - x^2)_+^2$  compared to the optimal Epanechnikov kernel is 99%, the Gaussian kernel has a relative efficiency of 95% and the rectangular kernel has

a relative efficiency about 93%. We expect that the choice of kernel function has little effect on the performance of the proposed estimators for model (1) as well. It is common to assume compact support for technical simplicity. This assumption can be relaxed to include the Gaussian kernel (Silverman, 1986, p.38).

The bandwidth, on the other hand, is much more of a concern. The cross-validation bandwidth selection is widely used to choose the bandwidth. Rice and Silverman (1991) suggested a leave-one-subject-out cross-validation approach. We recommend the  $K$ -fold cross-validation bandwidth selection considered by Sun et al. (2013) in the marginal estimation approach for the generalized semiparametric regression model (1). Specifically, subjects are divided into  $K$  approximately equal-sized groups. Let  $D_k$  denote the  $k$ th subgroup of data, then the  $k$ th prediction error is given by

$$PE_k(h) = \sum_{i \in D_k} \sum_{t_1 \leq T_{ij} \leq t_2} \left[ Y_{ij} - g^{-1}\{(\hat{\alpha}_{(-k)}(T_{ij}))^T X_{ij} + \hat{\beta}_{(-k)}^T Z_{ij}\} \right]^2, \quad (19)$$

for  $k = 1, \dots, K$ , where  $\hat{\alpha}_{(-k)}(t)$  and  $\hat{\beta}_{(-k)}$  are the estimators of  $\alpha(t)$  and  $\beta$  based on the data without the subgroup  $D_k$ , and  $[t_1, t_2] \subset (0, \tau)$ . The subset  $[t_1, t_2]$  is considered to avoid possible instability in estimating  $\alpha(t)$  near the boundary. In practice, this interval can be taken to be close to  $[0, \tau]$ . The data-driven bandwidth selection based on the  $K$ -fold cross-validation is to choose the bandwidth  $h$  that minimizes the total prediction error  $PE(h) = \sum_{k=1}^K PE_k(h)$ . The  $K$ -fold cross-validation bandwidth selection provides a working tool for locating an appropriate bandwidth.

The proposed estimation procedure holds for a wide class of link functions under model (1). A link function needs to be selected for a particular data application. The choice may be clear for some applications based on prior knowledge, but more often one needs to choose a link function that gives the “best fit” of the data. One criterion proposed by Sun et al. (2013) is to assess the model fit by the regression deviation defined as

$$RD(g(\cdot), h_{cv}) = \sum_{i=1}^n \sum_{t_1 \leq T_{ij} \leq t_2} \left[ Y_{ij} - g^{-1}\{(\hat{\alpha}_g(T_{ij}))^T X_{ij} + \hat{\beta}_g^T Z_{ij}\} \right]^2, \quad (20)$$

where  $h_{cv}$  is the bandwidth selected based on the  $K$ -fold cross-validation method for the given link function  $g(\cdot)$  described above, and  $\hat{\alpha}_g(t)$  and  $\hat{\beta}_g$  are the estimators of  $\alpha(t)$  and  $\beta$  under model (1) with the bandwidth  $h_{cv}$ . In practice, the link function  $g(\cdot)$  can be selected to minimize the regression deviation. Further examination of model fitness should be accompanied by model assessment tools such as the residual plots and formal goodness-of-fit tests.

### 3 Simulation Studies

In this section, we conduct a simulation study to assess the performances of the profile estimation methods using the QL, MGV, WLS and QIF approaches presented

in Section 2 under various models for longitudinal responses, different types of the within-subject correlation structures and different models for the measurement times. For convenience, we refer to the profile estimators resulted from these approaches as the the QL, MGW, WLS and QIF estimators. Section 3.1 presents a study of model (1) for continuous longitudinal responses and Section 3.2 shows the performances of these approaches for discrete longitudinal responses.

### 3.1 Continuous longitudinal responses

We study the performances of the proposed methods for continuous longitudinal responses under model (1) with the identity link function:  $Y_i(t) = \alpha^T(t)X_i(t) + \beta^T Z_i(t) + \epsilon_i(t)$ . We consider two simulation settings. In the first simulation setting (C1), the true correlation structure of the longitudinal responses is ARMA(1,1) and the measurement times are independent of covariates. In the second simulation setting (C2), the true correlation structure of the longitudinal responses is Exchangeable and the measurement times are dependent of covariates.

**Simulation setting (C1).** Similar to Fan et al. (2007), for each subject  $i$ , we consider time independent covariates  $X_i(t) = (X_{1i}(t), X_{2i}(t))^T$  and  $Z_i(t) = (Z_{1i}(t), Z_{2i}(t))^T$ , where  $X_{1i} = 1$ ,  $(X_{2i}(t), Z_{1i}(t))$  are time-varying covariates having a bivariate normal distribution with mean 0, variance 1 and correlation coefficient of 0.5 at each time  $t$ , and  $Z_{2i}$  is a time-independent covariate from Bernoulli distribution with success probability 0.5. We take  $\alpha(t) = (\sqrt{t/12}, \sin(2\pi t/12))^T$  and  $\beta = (1, 2)^T$ . The error  $\epsilon_i(t)$  is a Gaussian process with mean 0, variance varying with time  $\sigma^2(t) = 0.5 \exp(t/12)$  and of the ARMA(1,1) correlation structure, i.e.,  $\text{Corr}(Y_i(s), Y_i(t)) = \gamma \rho^{|t-s|}$  for  $s \neq t$ . We take  $(\gamma, \rho) = (0.85, 0.9)$  and  $(0.85, 0.6)$  for strong and moderate, respectively. All subjects have the same scheduled measurement time points,  $\{0, 1, 2, \dots, 12\}$ , but each of the scheduled time points has a 20% probability of being skipped except for the time 0. A random perturbation generated from the uniform distribution on  $[0, 1]$  is added to the non-skipped scheduled time points. Every subject has approximately 7 to 13 observations with an average of 11.

**Simulation setting (C2).** Similar to Sun et al. (2013), for each subject  $i$ , we let  $Z_{1i}(t)$  be a time-varying covariate from a uniform distribution on  $[0, 1]$ ,  $Z_{2i}$  a time-independent Bernoulli random variable with the success probability of 0.5,  $X_{1i} = 1$ , and  $X_{2i}(t)$  a time-varying Bernoulli random variable with the success probability of 0.5 at each time  $t$ . Let  $\alpha(t) = (0.5\sqrt{t}, 0.5 \sin(2t))^T$  and  $\beta = (0.5, 1)^T$ . The error  $\epsilon_i(t)$  has a normal distribution with mean  $\phi_i$  and variance  $\nu^2$ , where  $\phi_i$  is a random variable from  $N(0, 1)$ . Thus  $\epsilon_i(t)$  has an Exchangeable correlation structure with the correlation coefficients equal to  $\theta = 0.8$  and  $\theta = 0.5$  for  $\nu = 0.5$  and  $\nu = 1$ , respectively. The measurement times  $T_{ij}$  for each subject  $i$  follow a Poisson process with the intensity  $h_i(t) = 0.6 \exp(0.7Z_{2i})$ , for  $0 \leq t \leq \tau$  with  $\tau = 3.5$ . The censoring times  $C_i$  are generated from a uniform distribution on  $[1.5, 8]$ . There are

approximately three observations per subject on  $[0, \tau]$  and about 30% of subjects are censored before  $\tau = 3.5$ .

The performances of the profile GEE estimators using the QL, MGW and WLS approaches for estimating the correlation parameter  $\theta$  as well as the profile estimators via the QIF approach are examined under the settings (C1) and (C2). We let  $V_{1i}$  to be the identity matrix for all the estimators while using different correlation structures are assumed for  $V_{2i}$ . The working independence estimator (WI) is obtained by letting  $V_{2i}^{-1}$  be the identity matrix. The Epanechnikov kernel function  $K(x) = \frac{3}{4}(1 - x^2)_+$  is used in the study.

The simulation results for estimating  $\beta$  under the setting (C1) and ARMA(1,1) correlation with strong and moderate correlations are shown in Tables 1 and 2, respectively. The simulation results for estimating  $\beta$  under the setting (C2) and Exchangeable correlation with strong and moderate correlations are shown in Tables 3 and 4, respectively. The tables summarize the estimation bias (Bias), the sample standard error of estimates (SEE), the sample mean of the estimated standard errors (ESE), and the 95% empirical coverage probability (CP) for  $n = 200$ . Each entry of the table is calculated based on 1000 repetitions. The bandwidth used for each table is selected based on the 10-fold cross-validation of a single simulation that minimizes the total prediction error  $PE(h)$  for  $h$  in  $[0.7, 1.3]$  and carried it over for all 1000 repetitions.

**Table 1** Summary of Bias, SEE, ESE and CP under different estimation methods for  $\beta_1$  and  $\beta_2$  with  $n = 200$ ,  $h = 0.8$  based on 1000 simulations under the model setting (C1) and the strong ARMA(1,1) correlation with  $(\gamma, \rho) = (0.85, 0.9)$ .

Method	$\beta_1 = 1$				$\beta_2 = 2$			
	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP
Working Independence								
WI	0.0013	0.0244	0.0235	0.938	0.0059	0.1016	0.1011	0.944
Assuming Exchangeable Correlation (Misspecification)								
QL	0.0001	0.0166	0.0160	0.937	0.0041	0.1096	0.1020	0.938
MGV	0.0001	0.0166	0.0160	0.936	0.0041	0.1061	0.0992	0.940
QIF	0.0003	0.0168	0.0160	0.930	0.0053	0.0954	0.0915	0.940
WLS	0.0004	0.0171	0.0164	0.934	0.0052	0.0946	0.0912	0.941
Assuming ARMA(1,1) Correlation (True)								
QL	0.0005	0.0130	0.0127	0.936	0.0030	0.0955	0.0848	0.914
MGV	0.0008	0.0147	0.0140	0.929	0.0052	0.0936	0.0918	0.943
QIF	0.0005	0.0143	0.0136	0.931	0.0051	0.0926	0.0877	0.939
WLS	0.0006	0.0142	0.0136	0.935	0.0050	0.0926	0.0897	0.942
Assuming Mixed Correlation								
QL	0.0005	0.0131	0.0127	0.937	0.0030	0.0956	0.0847	0.916
MGV	0.0005	0.0145	0.0138	0.930	0.0047	0.0934	0.0895	0.942
QIF	0.0004	0.0142	0.0133	0.928	0.0053	0.0925	0.0864	0.932
WLS	0.0006	0.0142	0.0136	0.934	0.0051	0.0927	0.0897	0.942

**Table 2** Summary of Bias, SEE, ESE and CP under different estimation methods for  $\beta_1$  and  $\beta_2$  with  $n = 200$ ,  $h = 0.8$  based on 1000 simulations under the model setting (C1) and the moderate ARMA(1,1) correlation with  $(\gamma, \rho) = (0.85, 0.6)$ .

Method	$\beta_1 = 1$				$\beta_2 = 2$			
	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP
Working Independence								
WI	0.0009	0.0242	0.0236	0.942	0.004	0.0672	0.0666	0.946
Assuming Exchangeable Correlation (Misspecification)								
QL	0.0003	0.0220	0.0210	0.937	0.0033	0.0665	0.0634	0.941
MGV	0.0001	0.0221	0.0212	0.939	0.0024	0.0937	0.0895	0.949
QIF	0.0004	0.0221	0.0209	0.931	0.0035	0.0647	0.0618	0.943
WLS	0.0005	0.0222	0.0212	0.937	0.0035	0.0645	0.0623	0.943
Assuming ARMA(1,1) Correlation (True)								
QL	0.0008	0.0182	0.0174	0.935	0.0028	0.0618	0.0601	0.938
MGV	0.0007	0.0187	0.0179	0.933	0.0031	0.0618	0.0604	0.942
QIF	0.0007	0.0195	0.0182	0.928	0.0036	0.0629	0.0606	0.941
WLS	0.0008	0.0194	0.0184	0.933	0.0033	0.0622	0.0609	0.941
Assuming Mixed Correlation								
QL	0.0008	0.0182	0.0174	0.935	0.0028	0.0619	0.0601	0.937
MGV	0.0005	0.0192	0.0183	0.942	0.0027	0.0659	0.0630	0.945
QIF	0.0007	0.0196	0.0181	0.924	0.0036	0.0633	0.0600	0.944
WLS	0.0008	0.0193	0.0184	0.933	0.0033	0.0622	0.0609	0.941

**Table 3** Summary of Bias, SEE, ESE and CP under different estimation methods for  $\beta_1$  and  $\beta_2$  with  $n = 200$ ,  $h = 1.2$  based on 1000 simulations under the model setting (C2) and the strong Exchangeable correlation with  $\theta = 0.8$ .

Method	$\beta_1 = 1$				$\beta_2 = 2$			
	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP
Working Independence								
WI	0.0023	0.1535	0.1544	0.949	0.0048	0.1691	0.1629	0.934
Assuming ARMA(1,1) Correlation								
QL	0.0064	0.0884	0.0878	0.949	-0.0084	0.1605	0.1427	0.915
MGV	0.0047	0.0974	0.0973	0.948	0.0008	0.1574	0.1466	0.933
QIF	0.0048	0.1064	0.1019	0.935	-0.0001	0.1576	0.1434	0.917
WLS	0.0051	0.0945	0.0948	0.953	-0.0004	0.1564	0.1451	0.929
Assuming Exchangeable Correlation (True)								
QL	0.0069	0.0853	0.0855	0.956	-0.0145	0.1625	0.1423	0.910
MGV	0.0044	0.1113	0.1119	0.950	0.0022	0.1588	0.1493	0.928
QIF	0.0066	0.1069	0.1037	0.950	0.0021	0.1584	0.1473	0.922
WLS	0.0053	0.0942	0.0945	0.956	-0.0009	0.1563	0.1449	0.926
Assuming Mixed Correlation								
QL	0.0068	0.0855	0.0853	0.953	-0.0134	0.1622	0.1422	0.911
MGV	0.0031	0.1246	0.1249	0.949	0.0033	0.1616	0.1532	0.929
QIF	0.0060	0.1044	0.0976	0.930	-0.0003	0.1573	0.1421	0.914
WLS	0.0052	0.0942	0.0944	0.954	-0.0008	0.1563	0.1448	0.927

The results for WI is obtained by assuming working independence case. The performances of the estimators QL, MGV, WLS and QIF are examined under both

**Table 4** Summary of Bias, SEE, ESE and CP under different estimation methods for  $\beta_1$  and  $\beta_2$  with  $n = 200$ ,  $h = 1.2$  based on 1000 simulations under the model setting (C2) and the moderate Exchangeable correlation with  $\theta = 0.5$ .

Method	$\beta_1 = 1$				$\beta_2 = 2$			
	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP
Working Independence								
WI	0.006	0.1952	0.196	0.949	0.0062	0.1863	0.1787	0.94
Assuming ARMA(1,1) Correlation								
QL	0.0112	0.1611	0.1597	0.953	0.0006	0.1774	0.1647	0.932
MGV	0.0113	0.1723	0.1667	0.946	0.0020	0.1785	0.1665	0.929
QIF	0.0091	0.1740	0.1684	0.939	0.0033	0.1815	0.1658	0.917
WLS	0.0090	0.1661	0.1655	0.952	0.0039	0.1778	0.1667	0.927
Assuming Exchangeable Correlation (True)								
QL	0.0118	0.1602	0.1591	0.954	-0.0001	0.1773	0.1646	0.935
MGV	0.0083	0.1723	0.1731	0.952	0.0048	0.1795	0.1695	0.933
QIF	0.0114	0.1676	0.1637	0.951	0.0045	0.1787	0.1659	0.928
WLS	0.0092	0.1659	0.1654	0.956	0.0037	0.1777	0.1666	0.927
Assuming Mixed Correlation								
QL	0.0116	0.1602	0.1588	0.951	0.0001	0.1773	0.1646	0.933
MGV	0.0042	0.1837	0.1777	0.949	0.0055	0.1806	0.1710	0.936
QIF	0.0106	0.1692	0.1611	0.946	0.0039	0.1793	0.1632	0.921
WLS	0.0092	0.1658	0.1652	0.954	0.0037	0.1777	0.1665	0.927

the correctly specified correlation model and the misspecified correlation models. The results under "Assuming Exchangeable Correlation" are obtained by assuming exchangeable correlation in the estimation, the results under "Assuming ARMA(1,1) Correlation" are obtained by assuming ARMA(1,1) correlation in the estimation, while the results under "Assuming Mixed Correlation" are obtained by assuming the correlation to be the mix of the exchangeable and AR(1) correlation in the estimation. The basis matrices for the QIF estimator are taken as a combination the basis matrices for Exchangeable and AR(1) when ARMA(1,1) and Mixed Correlation Structures are assumed.

The simulation study shows that all estimators are consistent with small estimation bias. The WLS, QL, MGV and QIF estimators all perform well and improve the estimation efficiency compared with the working independence (WI) method. The methods utilizing the within-subject correlations show reduced estimation standard errors in SEE and ESE. More efficiency is gained by assuming the true or mixed correlation structures than the scenarios where correlation structures are misspecified. More efficiency gain is also observed in the settings with stronger within-subject correlations than with moderate within-subject correlations. For example, compared with the WI estimator, the sample standard errors of the QL, MGV, WLS and QIF estimators for  $\beta_1$  reduced between 30% to 46% in Table 1 for strong within-subject correlation and the sample standard errors reduced between 8% to 25% in Table 2 for moderate within-subject correlations under the true ARMA(1,1) correlation. Similarly, compared with the WI estimator, the sample standard errors of the QL, MGV, WLS and QIF estimators for  $\beta_1$  reduced between 18% to 44% in Table 3 for strong

within-subject correlation and the sample standard errors reduced between 6% to 18% in Table 4 for moderate within-subject correlations under the true Exchangeable correlation. The efficiency improved is more evident in estimating the effect of time-varying covariate than for the time-invariant covariate. This phenomenon also appeared in the simulation studies in Lin and Carroll (2001b) and Wang et al. (2005).

The performances of the estimators by assuming ARMA(1,1) working correlation and those under the mixed working correlation are close. The QL estimator appeared to achieve most efficiency gain out of these estimators in most scenarios. The above observations hold for both covariate-independent and covariate-dependent measurement times.

### 3.2 Discrete longitudinal responses

In this section we examine the performance of the proposed methods for model (1) for discrete longitudinal responses. We consider binary longitudinal responses in the simulation setting (D1), and Poisson count responses in the simulation setting (D2). Both settings have an Exchangeable correlation structure.

**Simulation setting (D1) – The Bernoulli model.** For binary longitudinal responses, we let  $g(\mu) = \log\{\mu/(1 - \mu)\}$  be the logistic link function. The observation times are generated similarly to the simulation setting (C1). All subjects have the same scheduled observation time points,  $\{0, 1, 2, \dots, 8\}$ , but each of the scheduled time points has a 20% probability of being skipped except for the time 0. A random perturbation generated from the uniform distribution on  $[0, 1]$  is added to the non-skipped scheduled time points. The number of observations,  $J_i$ , ranges from 4 to 9. At each observation time  $T_{ij}$ ,  $j = 1, \dots, J_i$ ,  $X_{ij} = 1$ ,  $Z_{1ij}$  and  $Z_{2ij}$  are independent standard normal random variables that don't vary with time. Let  $\alpha(t) = \sin(\pi t/30) - 0.5$  and  $\beta = (0.01, 0.01)^T$  and  $\mu_{ij} = P(Y_{ij} = 1 | X_{ij}, Z_{ij})$ . The binary longitudinal responses  $Y_{ij} = Y_i(T_{ij})$ ,  $j = 1, \dots, J_i$ , are generated with the marginal means following the logit model  $\text{logit}(\mu_{ij}) = \alpha(T_{ij})X_{ij} + \beta^T Z_{ij}$  and with constant correlation coefficient  $\text{Corr}(Y_i(s), Y_i(t)) = 0.5$  for  $s \neq t$ . We refer to Macke et al. (2009) for the techniques for simulating correlated binary responses. Our simulation used the Matlab code provided in the paper to generate the correlated binary variables with the specified mean and covariance.

**Simulation setting (D2) – The Poisson model.** Suppose that  $T_{ij}$ ,  $X_{ij}$  and  $Z_{ij}$  are the same as in the simulation setting (D1). We also use  $\alpha(t) = \sin(\pi t/30) - 0.5$  and  $\beta = (0.01, 0.01)^T$ . Let  $\mu_{ij} = E(Y_{ij} | X_{ij}, Z_{ij})$ . Using the method of Macke et al. (2009), we generate Poisson longitudinal process  $Y_{ij} = Y_i(T_{ij})$ ,  $j = 1, \dots, J_i$ , with the conditional marginal mean model  $\log(\mu_{ij}) = \alpha(T_{ij})X_{ij} + \beta^T Z_{ij}$  and with constant correlation coefficient  $\text{Corr}(Y_i(s), Y_i(t)) = 0.5$  for  $s \neq t$ .

The estimation results under the simulation settings (D1) and (D2) are summarized in Table 5 and Table 6, respectively. The simulation shows that estimation



bias is small for all estimators. The QL, MGW, WLS and QIF estimators that utilize the within-subject correlations show improved efficiency compared with using the working independence (WI) method with the sample standard errors reduced between 17% to 24% in Table 5 and between 10% to 30% in Table 6. The QL estimator achieved most efficiency gain out of these estimators. Efficiency gains are slightly higher when the true or mixed correlation structures are assumed compared to assuming the ARMA(1,1) correlation structures.

**Table 5** Summary of Bias, SEE, ESE and CP under different estimation methods for  $\beta_1$  and  $\beta_2$  with  $n = 200$ ,  $h = 1.2$  based on 1000 simulations under the Bernoulli model (D1) and the moderate Exchangeable correlation with  $\theta = 0.5$ .

Method	$\beta_1 = 1$				$\beta_2 = 2$			
	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP
Working Independence								
WI	0.0002	0.0524	0.0525	0.954	0.0014	0.0516	0.0525	0.962
Assuming ARMA(1,1) Correlation								
QL	-0.0002	0.0406	0.0404	0.945	-0.0012	0.0407	0.0405	0.948
MGV	-0.0004	0.0429	0.0428	0.937	-0.0005	0.0429	0.0429	0.950
QIF	-0.0002	0.0445	0.0434	0.932	-0.0008	0.0440	0.0435	0.942
WLS	0.0000	0.0411	0.0410	0.944	-0.0005	0.0410	0.0410	0.955
Assuming Exchangeable Correlation (True)								
QL	0.0000	0.0396	0.0395	0.947	-0.0013	0.0396	0.0395	0.949
MGV	0.0000	0.0396	0.0395	0.947	-0.0013	0.0396	0.0395	0.948
QIF	0.0003	0.0404	0.0397	0.944	-0.0012	0.0404	0.0397	0.943
WLS	0.0001	0.0408	0.0407	0.950	-0.0005	0.0406	0.0407	0.956
Assuming Mixed Correlation								
QL	0.0000	0.0396	0.0394	0.947	-0.0012	0.0396	0.0394	0.949
MGV	0.0003	0.0434	0.0432	0.946	0.0000	0.0428	0.0432	0.951
QIF	0.0001	0.0406	0.0393	0.941	-0.0014	0.0406	0.0393	0.936
WLS	0.0001	0.0408	0.0406	0.948	-0.0005	0.0406	0.0406	0.956

## 4 Concluding Remarks

The generalized semiparametric varying-coefficient additive model (1) specifies a model for the conditional mean of longitudinal responses. The model allows time-varying effects for some covariates and constant effects for others, and is an umbrella for many different models with selections of the link function. The intensively studied semiparametric additive model obtained by using the identity link function is popular for modeling continuous longitudinal responses. Semiparametric statistical modeling of discrete longitudinal responses has been understudied. With selection of link functions, model (1) can be used to model both continuous and discrete responses. Sun et al. (2013) investigated the local linear profile marginal estimation method for model (1) under the working independence. The estimation methods

**Table 6** Summary of Bias, SEE, ESE and CP under different estimation methods for  $\beta_1$  and  $\beta_2$  with  $n = 200$ ,  $h = 1.2$  based on 1000 simulations under the Poisson model (D2) and the moderate Exchangeable correlation with  $\theta = 0.5$ .

Method	$\beta_1 = 1$				$\beta_2 = 2$			
	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP
Working Independence								
WI	-0.0008	0.0271	0.0268	0.944	-0.0005	0.0271	0.0269	0.943
Assuming ARMA(1,1) Correlation								
QL	-0.0005	0.0206	0.0202	0.954	-0.0008	0.0199	0.0203	0.955
MGV	-0.0008	0.0243	0.0241	0.943	-0.0011	0.0245	0.0240	0.949
QIF	-0.0005	0.0220	0.0216	0.952	-0.0010	0.0224	0.0217	0.940
WLS	-0.0006	0.0208	0.0205	0.954	-0.0007	0.0202	0.0205	0.954
Assuming Exchangeable Correlation (True)								
QL	-0.0006	0.0199	0.0197	0.951	-0.0007	0.0191	0.0197	0.955
MGV	-0.0007	0.0219	0.0216	0.947	-0.0006	0.0215	0.0217	0.950
QIF	-0.0005	0.0201	0.0197	0.950	-0.0006	0.0193	0.0197	0.953
WLS	-0.0006	0.0206	0.0203	0.952	-0.0007	0.0199	0.0203	0.954
Assuming Mixed Correlation								
QL	-0.0006	0.0199	0.0197	0.948	-0.0007	0.0192	0.0197	0.956
MGV	-0.0006	0.0222	0.0218	0.956	-0.0008	0.0219	0.0219	0.945
QIF	-0.0006	0.0202	0.0195	0.949	-0.0007	0.0194	0.0195	0.954
WLS	-0.0006	0.0206	0.0202	0.954	-0.0007	0.0200	0.0203	0.955

under working independence that ignore the within-subject correlation are valid and yield asymptotically unbiased estimators.

In this paper, we studied several profile estimation methods for model (1) that utilize the within-subject correlations to improve estimation efficiency. Several profile estimation methods that utilize the within-subject correlations including the profile GEE approaches and the profile QIF approach were investigated. The profile estimations are assisted with the local linear smoothing technique by approximating the time-varying effects with linear functions in the neighborhood of each time. The profile GEE approaches include the quasi-likelihood, the minimum generalized variance, and the weighted least squares. These methods differ by different procedures used in estimating the within-subject correlations. The proposed profile estimation methods for the generalized semiparametric varying-coefficient additive model (1) provide a unified approach that work well for discrete longitudinal responses as well as for continuous longitudinal responses.

Finite sample performances of these different methods are examined through Monto Carlo simulations under various correlation structures for both discrete and continuous longitudinal responses. Our study showed significant efficiency improvement of all the estimators, the QL, WLS, WLS and QIF estimators, over the working independence approach. The QL estimator appeared to achieve most efficiency gain out of all estimators in most scenarios. The efficiency improved is more evident in estimating the effects of time-varying covariates than for the time-invariant covariates. Efficiency gains are higher when the true or mixed correlation structures are assumed

compared to the mis-specified correlation structures. The above observations hold for both covariate-independent and covariate-dependent measurement times.

**Acknowledgements** We thank the editors for your handling of our manuscript. We also thank the reviewer's valuable comments and suggestions that have improved the paper. This research was partially supported by the NIAID NIH award number R37AI054165 and the National Science Foundation grant DMS1915829. The authors would like to thank professor Runzi Li for providing the program code for Fan, Huang and Li (2007).

## References

- Chen, K. and Jin, Z. (2006). Partial linear regression models for clustered data. *Journal of the American Statistical Association*, 101(473):195–204.
- Dempster, A. P. (1969). *Elements of continuous multivariate analysis*. Addison-Wesley, Reading, MA.
- 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(478):632–641.
- Fan, J. and Li, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association*, 99(467):710–723.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pages 1029–1054.
- Hoover, D. R., Rice, J. A., Wu, C. O., and Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, 85(4):809–822.
- Hu, Z., Wang, N., and Carroll, R. J. (2004). Profile-kernel versus backfitting in the partially linear models for longitudinal/clustered data. *Biometrika*, 91(2):251–262.
- Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, pages 13–22.
- Lin, D. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association*, 96(453):103–126.
- Lin, X. and Carroll, R. J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American statistical Association*, 95(450):520–534.
- Lin, X. and Carroll, R. J. (2001a). Semiparametric regression for clustered data. *Biometrika*, pages 1179–1185.
- Lin, X. and Carroll, R. J. (2001b). Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association*, 96(455):1045–1056.

- Macke, J. H., Berens, P., Ecker, A. S., Tolias, A. S., and Bethge, M. (2009). Generating spike trains with specified correlation coefficients. *Neural computation*, 21(2):397–423.
- Madsen, L., Fang, Y., Song, P. X.-K., Li, M., and Yuan, Y. (2011). Joint regression analysis for discrete longitudinal data. *Biometrics*, 67(3):1171–1176.
- Martinussen, T. and Scheike, T. H. (1999). A semiparametric additive regression model for longitudinal data. *Biometrika*, 86:691–702.
- Qu, A., Lindsay, B. G., and Li, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika*, 87(4):823–836.
- Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(1):233–243.
- Severini, T. A. and Staniswalis, J. G. (1994). Quasi-likelihood estimation in semiparametric models. *Journal of the American statistical Association*, 89(426):501–511.
- Silverman, B. (1986). Density estimation for statistics and data analysis, publisher chapman and hall. *London, UK*.
- Song, P. X.-K., Li, M., and Yuan, Y. (2009). Joint regression analysis of correlated data using gaussian copulas. *Biometrics*, 65(1):60–68.
- Sun, Y., Sun, L., and Zhou, J. (2013). Profile local linear estimation of generalized semiparametric regression model for longitudinal data. *Lifetime data analysis*, 19(3):317–349. PMID: PMC3710313.
- Sun, Y. and Wu, H. (2005). Semiparametric time-varying coefficients regression model for longitudinal data. *Scandinavian Journal of Statistics*, 32(1):21–47.
- Tang, C. Y., Zhang, W., and Leng, C. (2019). Discrete longitudinal data modeling with a mean-correlation regression approach. *Statistica Sinica*, 29(2):853–876.
- Wang, N., Carroll, R. J., and Lin, X. (2005). Efficient semiparametric marginal estimation for longitudinal/clustered data. *Journal of the American Statistical Association*, 100(469):147–157.
- Wu, H. and Liang, H. (2004). Backfitting random varying-coefficient models with time-dependent smoothing covariates. *Scandinavian Journal of Statistics*, 31(1):3–19.
- Zeger, S. L. and Diggle, P. J. (1994). Semiparametric models for longitudinal data with application to cd4 cell numbers in hiv seroconverters. *Biometrics*, pages 689–699.