Ann Inst Stat Math (2016) 68:977–1000
DOI 10.1007/s10463-015-0519-8

Robust estimation of generalized partially linear model for longitudinal data with dropouts

Guoyou Qin[1] · Zhongyi Zhu[2] · Wing K. Fung[3]

Received: 24 October 2013 / Revised: 28 September 2014 / Published online: 23 April 2015

Abstract In this paper, we study the robust estimation of generalized partially linear models (GPLMs) for longitudinal data with dropouts. We aim at achieving robustness against outliers. To this end, a weighted likelihood method is first proposed to obtain the robust estimation of the parameters involved in the dropout model for describing the missing process. Then, a robust inverse probability-weighted generalized estimating equation is developed to achieve robust estimation of the mean model. To approximate the nonparametric function in the GPLM, a regression spline smoothing method is adopted which can linearize the nonparametric function such that statistical inference can be conducted operationally as if a generalized linear model was used. The asymptotic properties of the proposed estimator are established under some regularity conditions, and simulation studies show the robustness of the proposed estimator. In the end, the proposed method is applied to analyze a real data set.

Keywords Dropouts · Partially linear models · Regression splines · Robustness

1 Introduction

Longitudinal studies arise commonly in many research areas including medicine, public health and social science. Usually, longitudinal studies are designed to collect

✉ Zhongyi Zhu
zhuzy@fudan.edu.cn

1 Department of Biostatistics, School of Public Health and Key Laboratory of Public Health Safety, Fudan University, Shanghai 200032, China

2 Department of Statistics, Fudan University, Shanghai 200433, China

3 Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong, China

data for each subject in the sample at different time, and incomplete data often arise during the period of study due to various reasons. Sometimes, subjects may drop out before the end of the follow-up. A simple idea to handle the missing data is to exclude those observations and directly make analysis based on the remaining complete data. However, this way can lead to invalid inference when the data are not missing completely at random [see Little and Rubin (2002)].

There are extensive literatures on the approaches for dealing with incomplete longitudinal data. The inverse probability-weighted generalized estimating equation (IPW-GEE) proposed by Robins et al. (1995) is one of the popular methods which can provide consistent estimators when the data are missing at random (MAR). This approach is attractive and has been widely used since it does not require specification of the joint distribution of the correlated longitudinal responses but only need to specify the first two moments. However, to our knowledge, most of the existing work focused on the generalized linear models (GLMs) while few devoted to the generalized partially linear models (GPLMs) with longitudinal incomplete data possibly due to the complexity resulted from the nonparametric component in the GPLM. The GPLM can be viewed as a combination of generalized linear models and fully nonparametric models. It can be used to model nonnormally distributed response such as binary and Poisson data, and allows to model the covariates nonparametrically when the assumption of linearity may not be suitable. Due to its great flexibility, the GPLM has attracted considerable attention in theoretical study and practical application. An incomplete list of recent literatures on the GPLM with longitudinal data includes Qin and Zhu (2009), Lian et al. (2014) and Chen and Zhou (2013). Particularly, based on the IPW-GEE, Chen and Zhou (2013) studied the estimation of the GPLM for longitudinal data with dropouts by incorporating the population-level information. The local linear approximation method is adopted to estimate the nonparametric function.

Generalized estimating equations (GEEs) are very popular and have been widely used. However, it is well known that the GEE is sensitive to outliers in the data. For the complete longitudinal data, there are substantial studies on the robust GEE approaches (e.g., Cantoni and Ronchetti 2001; Sinha 2004; He et al. 2005; Qin and Zhu 2007). Compared with the robust methods for complete longitudinal data, the study on robust approaches for incomplete longitudinal data received limited attention although valuable. Recently, Yi and He (2009) studied median regression for longitudinal data with dropouts through the IPW-GEE approach. Sinha (2012) discussed the robust analysis of longitudinal data with missing responses based on likelihood methods. However, these works are developed based on the linear models.

In this paper, we focus on robust estimation of the GPLM for longitudinal data with dropouts based on the IPW-GEE method. First, we propose a weighted likelihood function for the missing process to obtain the robust estimates of the probability of observation whose inverse is incorporated into the GEE for the mean model to correct the bias induced by the missingness. Then, to construct a robust GEE for the mean model, we make use of a bounded score function of Pearson residuals to limit the influence of outliers in the response, and adopt a function of the covariates to down weight the impact of outliers in the covariates. Finally, we utilize the regression spline to approximate the nonparametric function in the GPLM which is different from the local linear method adopted in Chen and Zhou (2013). The regression spline is easy to implement since it linearizes the nonparametric function as a linear combination of a set of basis functions. Thus, any computational algorithm developed for the GLM can be directly applied to the GPLM. However, the use of regression spline brings some difficulties in establishing the asymptotic properties of the proposed estimator because an infinite-dimensional problem has to be solved, which arises from the number of knots increasing to infinity with the sample size. This is a significant difference with other smoothing method such as kernel and local polynomial methods. Incorporating all these techniques, we propose a robust IPW-GEE for the GPLM with dropouts.

The rest of this paper is organized as follows. The models and proposed method are given in Sect. 2. Some regularity conditions and the asymptotic properties of the proposed estimator are shown in Sect. 3. Simulation studies are conducted to investigate the performance of the proposed estimator in Sect. 4. In Sect. 5, the proposed method is applied to a real data analysis for illustration. The technical details of the proof are presented in the Appendix.

2 Model and proposed method

2.1 Mean model and dropout model

We consider a longitudinal study consisting of n subjects with m observations over time for each subject. Let us denote $\{Y_{ij}, X_{ij}, T_{ij}\}$, $i = 1, \ldots, n$, $j = 1, \ldots, m$ as the observed data set. We model the longitudinal data using a GPLM and specify the first two moments of the response y_{ij} as $E(Y_{ij}) = \mu_{0,ij}$, $\text{var}(Y_{ij}) = \phi v(\mu_{0,ij})$ where ϕ is a scale parameter and $v(\cdot)$ is a known variance function. In this paper, we model the marginal mean as:

$$\eta_{0,ij} = g(\mu_{0,ij}) = X_{ij}^T \beta_0 + f_0(T_{ij}), \tag{1}$$

where β_0 is a p-dimensional vector of regression parameter, X_{ij} is the associated $p \times 1$ covariate vector, $f_0(\cdot)$ is an unknown smoothing function and $g(\cdot)$ is a given link function. Without loss of generality, we assume that T_{ij} are all scaled into the interval [0, 1]. We are interested in estimation of β_0 and f_0. In the following, let $Y_i = (Y_{i1}, \ldots, Y_{im})^T$ denote the response vector for the ith subject. We define X_i and T_i in a similar fashion.

Following He et al. (2002), we approximate the nonparametric function through regression spline. Let $0 = a_0 < a_1, \ldots, a_{k_n} < a_{k_n+1} = 1$ be a partition of the interval [0, 1]. Taking these $\{a_i\}$ as knots, we can get $N_k = k_n + l$ normalized B-spline basis functions of order l, denoted by $\{B_1(t), \ldots, B_{N_k}(t)\}$. Then, let $f_0(t)$ be approximated by $d(t)^T \alpha_0$ where $d(t) = (B_1(t), \ldots, B_{N_k}(t))^T$ and $\alpha_0 \in R^{N_k}$ is the vector of spline coefficient. This linearizes regression model (1) so that our regression problem becomes

$$\eta(\theta_0) = g(\mu(\theta_0)) = X_{ij}^T \beta_0 + d_{ij}^T \alpha_0 = D_{ij}^T \theta_0, \tag{2}$$

where $D_{ij} = (X_{ij}^T, d_{ij}^T)^T$, $\theta_0 = (\beta_0^T, \alpha_0^T)^T$ is the combined regression parameters. In this paper, we use cubic spline of order 4 and select the sample quantiles of $\{T_{ij}\}$

as knots. The number of the internal knots k_n is taken to be the integer part of $F_n^{1/5}$ where F_n is the number of distinct values of $\{T_{ij}\}$. This choice is consistent with the asymptotic results given in Sect. 3. As indicated in He et al. (2005), the regression spline can provide good approximation with small number of knots. And this approach linearizes the nonparametric function so that any algorithm designed for the linear models can be directly applied to the partially linear models.

Next, we will consider the model for the dropouts. The issue of dropouts may be particularly acute in epidemiological cohort studies where interests lie in estimating trends over time and where subjects are followed prospectively over several years. We suppose that the covariates $\{X_{ij}, T_{ij}\}$ can be completely observed while the response $\{Y_{ij}\}$ can be missing. Assuming that the corresponding covariate X_{ij} is observed when the response Y_{ij} is missing due to dropouts may be not realistic for some real problems and need to be improved in the future study. Let R_{ij} be 1 if Y_{ij} is observed, and 0 otherwise. In this paper, we consider dropouts or monotone missing data pattern, i.e., $R_{ij} = 0$ indicates $R_{ik} = 0$ for all $k > j$. Without loss of generality, we assume $R_{i1} = 1$ for each subject. Furthermore, we consider a missing at random (MAR) mechanism for the dropout process. Here, the MAR mechanism means that for given covariates, the conditional distribution of the missing data indicator $R_i = (R_{i1}, \ldots, R_{im})^T$, $f_{R_i}(r_i|X_i, T_i, Y_i)$, only depends on the observed response components Y_i^{obs}.

Let $\lambda_{ij} = P(R_{ij} = 1|R_{i,j-1} = 1, X_i, Y_i)$, and $\pi_{ij} = P(R_{ij} = 1|X_i, T_i, Y_i)$. It is obvious that $\pi_{ij} = \prod_{k=2}^{j} \lambda_{ik}$. We denote the response history up to (but not including) time point j by $\tilde{Y}_i = \{y_{i1}, \ldots, y_{i,j-1}\}$. A common logistic regression model used in modeling the dropout process is

$$\ln \frac{\lambda_{ij}}{1-\lambda_{ij}} = Z_{ij}^T \gamma_0, \tag{3}$$

where Z_{ij} is the vector consisting of the information of the covariates X_i, T_i and the observed responses $\tilde{Y}_i$, and γ_0 is the q-vector of regression parameters. Let L_i denote the random dropout time for subject i, and l_i be its observed value, $i = 1, \ldots, n$. The likelihood function for this missing process can be defined as: $L_i(\gamma) = (1 - \lambda_{il_i}) \prod_{k=2}^{l_i-1} (\lambda_{ik})$, if $l_i < m$; otherwise, $L_i(\gamma) = \prod_{k=2}^{m} \lambda_{ik}$, where λ_{ik} is determined by model (3). Then, the classical maximum likelihood method can be applied to obtain the estimate of γ_0.

2.2 The proposed method

It has been well recognized that the complete-case methods which exclude the missing data and directly analyze the remaining completely observed data can lead to invalid inference when data are not missing at random like the missing mechanism considered here. To deal with the missingness, we consider to use the IPW method. As addressed in Sect. 2.1, classical maximum likelihood method can be used to estimate γ_0 and then obtain the estimates of the probability of observation π_{ij} whose inverse can be incorporated into the GEE to get the IPW-GEE. However, the maximum likelihood method is sensitive to the possible outliers in the data, which means that the outliers may result in seriously biased estimates of γ_0 and thus the π_{ij} cannot be consistently estimated. Therefore, we first propose a robust method for estimation of γ_0. For the above logistics model for dropout process (3), the outliers usually arise in the covariates Z_{ij} which generally involve Y_{ij} and some components of X_{ij} and T_{ij}. Thus, we propose a robust approach to downweight the impact of outliers in the covariates through a weighted likelihood function which is defined as $L_i(\gamma) = (1-\lambda_{il_i})^{w_{il_i}^{(DP)}} \prod_{k=2}^{l_i-1}(\lambda_{ik})^{w_{ik}^{(DP)}}$, if $l_i < m$; otherwise, $L_i(\gamma) = \prod_{k=2}^{m} \lambda_{ik}^{w_{ik}^{(DP)}}$, where λ_{ik} is determined by model (3) and $w_{ik}^{(DP)} = w(Z_{ij}^0)$ with Z_{ij}^0 being some components of Z_{ij} which may be contaminated by outliers. The weights $w_{ik}^{(DP)}$ are used to reduce the impact of outliers in the covariates. Small weights will be assigned to the outliers, then the contribution of the observations involving outliers to the likelihood function is reduced. Thus, the impact of the outliers on the final estimates will be limited. The weighting function $w(\cdot)$ is a function of Z_{ij}^0 which will be presented later. Then, the robust estimator $\hat{\gamma}$ of γ can be obtained by solving

$$G_{\gamma,n}(\gamma) = \sum_{i=1}^{n} G_{\gamma,i}(\gamma) = 0, \tag{4}$$

where $G_{\gamma,i}(\gamma) = \partial \log L_i(\gamma)/\partial\gamma$. Note that the covariates can also create large standardized residuals. However, using similar weighting method to control the outliers in the residuals does not work since the residuals are functions of the responses which will result in inconsistent estimate for the parameters in the dropout model. How to further control the effect of outliers in the residuals on this likelihood function deserves further studies in the future.

In the following, we construct a robust IPW-GEE for the mean model to deal with both the missing response and the outliers. Specifically, we adopt the inverse probability-weighted method to deal with the missingness, and make use of the bounded score functions and weighting functions to reduce the impacts of outliers in the response and covariates, respectively. Incorporating these techniques, we propose a robust IPW-GEE as:

$$U_{\theta,n} = \sum_{i=1}^{n} U_{\theta,i} = \sum_{i=1}^{n} \{D_i \Delta_i^T(\mu_i(\theta)) A_i^{-1/2}(\mu_i(\theta)) C_i^{-1}(\rho) J_i(\mu_i(\theta), \hat{\gamma})\} = 0, \tag{5}$$

where $D_i = (D_{i1}^T, \ldots, D_{im}^T)^T$, $\Delta_i(\mu_i(\theta)) = diag\{\dot{\mu}_{i1}(\theta), \ldots, \dot{\mu}_{im}(\theta)\}$, $\dot{\mu}$ denotes the first derivative of $\mu(\theta)$ evaluated at $D_i\theta$, $A_i(\mu_i(\theta)) = \phi \text{diag}\{v(\mu_{i1}(\theta)), \ldots, v(\mu_{im}(\theta))\}$, $C_i(\rho)$ is a working correlation matrix with ρ which is a vector of correlation parameters, $J_i(\mu_i(\theta), \hat{\gamma}) = S_i(\hat{\gamma}) W_i h_i(\mu_i(\theta))$ with $S_i(\hat{\gamma}) = \text{diag}\{\frac{R_{i1}}{\pi_{i1}(\hat{\gamma})}, \ldots, \frac{R_{im}}{\pi_{im}(\hat{\gamma})}\}$ which are used to correct the bias induced by the missingness, $W_i = \text{diag}\{w_{i1}^{(\text{MR})}, \ldots, w_{im}^{(\text{MR})}\}$ are weighted matrices used to down weight the influence of outliers in the covariates of the mean regression model, $w_{ij}^{(\text{MR})} = w(x_{ij})$ with the weighting function $w(\cdot)$ to be specified later, $h_i(\mu_i(\theta)) = \psi(A_i^{-1/2}(\mu_i(\theta))(Y_i - \mu_i(\theta))) - E_{Y_i|X_i,T_i}\psi(A_i^{-1/2}(\mu_i(\theta))(Y_i - \mu_i(\theta)))$, $\psi(\cdot)$ is a bounded function used to limit the impact of the outliers in the response and chosen to be Huber function here which is $\psi(x) = \min\{c, \max\{-c, x\}\}$. The tuning constant c is usually chosen to balance the estimation efficiency and robustness and normally selected between 1 and 2. Specifically, if larger c is chosen, the corresponding robust estimator would have more estimation efficiency, but weak robustness against outliers in the residuals. Otherwise, the robust estimator would have better robustness, but lower efficiency. According to our empirical experience and following He et al. (2005), we choose $c = 1.5$. The proposed robust estimator shows desirable robustness. Although the choice of different c will have some influence on the resulting estimators, the conclusions of robustness for the corresponding estimators are consistent. Wang et al. (2007) discuss this issue and we refer to this paper for more details on this choice. Following Sinha (2004), the weighting function $w(x_{ij})$ is chosen to be $\min\left[1, b_0/((x_{ij} - m_x)^T S_x^{-1}(x_{ij} - m_x))^{\tau/2}\right]$ with $\tau \geq 1$ which is taken to be 1 in the latter simulation studies and real data analysis, where b_0 is chosen as the 95th percentile of Chi-square distribution with degrees of freedom equal to the dimension of x_{ij}, and m_x and S_x are some robust estimates of location and scale of x_{ij}, such as minimum volume ellipsoid (MVE) estimates of Rousseeuw and van Zomeren (1990). Similar to the tuning parameter c in the Huber function, the choice of τ may also influence the robustness and estimation efficiency of the resulting estimator. How to choose a suitable τ is an interesting topic and deserves future study.

Remark The IPW method only uses the observed measurements. It is just for brevity and uniform expression to define (2.5) for all m measurements corresponding to m time-points. In fact, the diagonal weight matrix $S_i(\hat{\gamma}) = \text{diag}\{\frac{R_{i1}}{\pi_{i1}(\hat{\gamma})}, \ldots, \frac{R_{im}}{\pi_{im}(\hat{\gamma})}\}$ can naturally identify the observed measurements and missing measurements. The missing measurements are set to be zeros since their corresponding R_{ij} are zeros. Thus, the missing measurements do not contribute to the estimating equation any more and only the observed measurements are used for estimation.

The robust IPW-GEE (5) involves nuisance parameters ϕ and ρ which are also required to be estimated. For the scale parameter ϕ, to achieve robust estimation and avoid the influence of missingness, we estimate it through the median absolute deviation based on the Pearson residuals of the first observation from each subject $\hat{e}_{i1} = (Y_{i1} - \mu_{i1}(D_{i1}^T\hat{\theta}))$ and $\hat{\theta}$ is the current estimate of θ_0. Since the first observation for the individuals is always observed and independent, the median absolute deviation estimate based on these residuals is consistent for the error variance in a Gaussian model. For the correlation parameters ρ, to correct the bias induced by the missingness, we consider the bias-corrected robust Wang–Carey estimating equation approach proposed in Qin et al. (2008) for estimation of the correlation parameter. The basic idea of their method is to estimate the bias of estimates solved from biased estimating equations and then correct the bias. Their method can be also applied to our case since the general estimating equations developed for the complete data are biased due to the missing data.

The procedure for estimation of θ_0 is described as follows:

Step 1 Solving (4) to get the robust estimate of $\hat{\gamma}$, and then calculate the estimated $\pi_{ij}(\hat{\gamma})$.

Step 2

(a) Choose initial value $\theta^{(0)}$, $\phi^{(0)}$ and $\rho^{(0)}$ which can be obtained using the complete-case robust method like He et al. (2005). Set $i = 0$.

(b) With initial value of $\theta^{(i)}$, $\phi^{(i)}$ and $\rho^{(i)}$, calculate

$$\theta^{(i+1)} = \theta^{(i)} - \left[\sum_{i=1}^{n} \frac{\partial}{\partial \theta} U_{\theta,i}|_{\theta=\theta^{(i)}} \right]^{-1} \sum_{i=1}^{n} U_{\theta,i}|_{\theta=\theta^{(i)}}.$$

(c) With $\theta^{(i+1)}$, calculate $\phi^{(i+1)}$ and $\rho^{(i+1)}$ using the above addressed method. Set $i = i + 1$.

Iterate (b) and (c) until convergence.

The solution of (5) is denoted as $\hat{\theta} = (\hat{\beta}^T, \hat{\alpha}^T)^T$. Then, the estimators of the regression coefficient and nonparametric function are $\hat{\beta}$ and $\hat{f}_0(t) = d^T(t)\hat{\alpha}$.

In this section, we not only propose a robust IPW-GEE for the mean regression model but also develop a weighted likelihood method for the robust estimation of the dropout model. With the guarantee of these robust approaches, we have finally achieved robust estimation of the GPLM for longitudinal data with dropouts. The latter simulation studies show the good performance of the proposed method in dealing with both outliers and missingness.

3 Asymptotic properties

In this section, we will investigate the asymptotic properties of the proposed estimator. To this end, we first give some regularity conditions. We use $||\cdot||$ to denote Euclidean norm. Let $e_i = (A_i^{-1/2}(\mu_{0,i})(Y_i - \mu_{0,i}))$ be the standardized responses, and $J_{0,i}(e_i) = S_i(\gamma_0)W_i h_i(e_i)$. Note that $J_{0,i}(e_i)$ is similar to $J_i(\mu_i(\theta), \gamma)$, but the former is evaluated at the true $\mu_{0,i}$ and γ_0 whereas the latter is at $\mu_i(\theta)$ and γ. If the estimating Eq. (5) has multiple solutions, only a sequence of consistent estimator $\hat{\theta}$ is considered. A sequence of $\hat{\theta}$ is said to be consistent if both $\hat{\beta} - \beta_0$ and $\sup_t|(d^T(t)\hat{\alpha} - f_0(t))|$ converge to zero in probability. The basic conditions we assumed are as follows:

(C.1) The parameter vector γ_0 is an interior point of the parameter space Γ which is a compact set.

(C.2) $\lambda_{ij}(\gamma) > c_1 > 0$ for all $\gamma \in \Gamma$, for some constant c_1.

(C.3) $\frac{1}{n}\frac{\partial}{\partial\gamma}G_{\gamma,n}(\gamma_0)$ and $\frac{1}{n}\sum_{i=1}^{n} G_{\gamma,i}(\gamma_0)G_{\gamma,i}^T(\gamma_0)$ converges to Σ_γ and V_γ, respectively, in probability for some positive-definite matrix Σ_γ and V_γ.

(C.4) The rth derivative of f_0 is bounded for some $r \geq 2$.

(C.5) The distinct values of $\{t_{ij}\}$ form a quasi-uniform sequence that grows dense on $[0, 1]$.

(C.6) There exists positive constant c_2 such that $0 < c_2 \leq v(\cdot) < \infty$, $v(\cdot)$ and $g^{-1}(\cdot)$ have bounded second derivatives and third derivatives, respectively.

To establish the asymptotic normality for the estimator $\hat{\beta}$, some conditions on the covariates X and T are required. One complexity with the partial linear models is from the dependence between X and T. To this purpose, we assume the following relationship as Rice (1986): $X_{ijk} = m_k(T_{ij}) + \varrho_{ijk}$, $1 \le i \le n$, $1 \le j \le m$, $1 \le k \le p$ where the $m_k(\cdot)$ are functions with bounded rth derivatives and the ϱ_{ijk} are independent random variables with mean 0 and are independent of e_{ij}.

Let Λ_n denote as a $N(= n \times m)$ by p matrix whose lth column is $\varrho_l = (\varrho_{11l}, \ldots, \varrho_{1ml}, \ldots, \varrho_{nml})^T$. For brevity, we denote $A_i = A_i(\mu_i(\theta))$, $\Delta_i = \Delta_i(\mu_i(\theta))$, $J_i = J_i(\mu_i(\theta), \gamma)$ and $C_i = C_i(\rho)$. We further denote $M = (d_1^T, \ldots, d_n^T)^T$, $X = (X_1^T, \ldots, X_n^T)^T$, $\Omega = \text{diag}\{\Omega_1, \ldots, \Omega_n\}$, $\Omega_i = \Delta_i^T A_i^{-1/2} C_i^{-1} E\{\frac{\partial}{\partial \mu_i} J_i\}\Delta_i$, $P = M(M^T \Omega M)^{-1} M^T \Omega$, $X^* = (X_1^{*T}, \cdots, X_n^{*T})^T = (I - P)X$, $Q_i = X_i^{*T} \Delta_i^T A_i^{-1/2} C_i^{-1} J_i - [\sum_{i=1}^n X_i^{*T} \Delta_i^T A_i^{-1/2} C_i^{-1} \frac{\partial}{\partial \gamma} J_i(\mu_i, \gamma)] \cdot [\frac{\partial}{\partial \gamma} G_{\gamma,n}(\gamma)]^{-1} G_{\gamma,i}(\gamma)$ and the notations $\Omega_{0,i}$ represent Ω_i evaluated at the true $\mu_{0,i}$ and γ_0. Notations $Q_{0,i}$, $A_{0,i}^{-1/2}$, $\Delta_{0,i}$ and $X_{0,i}^*$ are defined in similar fashion. We further assume:

(C.7) For sufficiently large n, $k_n(M^T \Omega_0 M)$ is nonsingular, and the eigenvalues of $(k_n/n) M^T \Omega_0 M$ are bounded away from zero and infinity in probability, where $\Omega_0 = \text{diag}\{\Omega_{0,1}, \ldots, \Omega_{0,n}\}$.

(C.8) $E\Lambda_n = 0$ and $\sup_n \frac{1}{n} E||\Lambda_n||^2 < \infty$, and $\frac{1}{n} K_n \to K$, $\frac{1}{n} B_n \to B$ in probability for some positive definite matrix K and B, where $K_n = \sum_{i=1}^n X_{0,i}^{*T} \Omega_{0,i} X_{0,i}^*$ and $B_n = \sum_{i=1}^n Q_{0,i} Q_{0,i}^T$.

Conditions (C.1)–(C.3) are assumed to ensure the existence and asymptotic normality of $\hat{\gamma}$ and Conditions (C.4)–(C.8) are usually assumed in the study of the asymptotic normality of $\hat{\beta}$ and convergence rate of $\hat{f}_0$ in the context of GPLM when regression spline is utilized to approximate the nonparametric function; see He et al. (2005).

Under the regularity conditions (C.1)–(C.8), the asymptotic properties of the proposed robust IPW estimator can be established. Specifically, Theorem 1 gives the asymptotic normality of the robust estimator $\hat{\gamma}$. Theorem 2 shows the asymptotic normality of the proposed estimator for regression coefficients $\hat{\beta}$, and shows that the proposed estimator of the nonparametric function can achieve the optimal rate of convergence under the smoothing condition (C.4).

Theorem 1 *Assume that Conditions* (C.1)–(C.3) *hold. Then*

$$\sqrt{n}(\hat{\gamma} - \gamma_0) \to N(0, \Sigma_\gamma^{-1} V_\gamma \Sigma_\gamma^{-1}). \tag{6}$$

Using the standard techniques, the conclusion of Theorem 1 can be easily obtained and its proof is omitted here.

Theorem 2 *Assume that Conditions* (C.1)–(C.8) *hold. If the number of knots* $k_n \approx n^{1/(2r+1)}$, *then*

$$\frac{1}{n} \sum_{i=1}^n (\hat{f}(t_i) - f_0(t_i))^2 = O_p\left(n^{-\frac{2r}{2r+1}}\right), \tag{7}$$

and

$$\sqrt{n}(\hat{\beta} - \beta_0) \to N(0, K^{-1} B K^{-1}). \tag{8}$$

The proof of Theorem 2 is given in the Appendix.

To make statistical inference of the regression coefficients β_0, the covariance matrix of $\hat{\beta}$ needs to be consistently estimated. Following Theorem 2 in Qin and Zhu (2007), the covariance matrix can be consistently estimated by $\hat{K}^{-1}\hat{B}\hat{K}^{-1}$ where

$$\hat{K} = \frac{1}{n}\sum_{i=1}^{n} X_i^{*T}\Omega_i X_i^*, \tag{9}$$

$$\hat{B} = \frac{1}{n}\sum_{i=1}^{n} Q_i Q_i^T, \tag{10}$$

where all the quantities involved are evaluated at $\hat{\theta}$ and $\hat{\gamma}$.

Remark The asymptotic properties of the robust estimator are only established under the case of no contamination. For the proposed robust estimator, bounded score function is used to reduce the impact of outliers in the response. Furthermore, leverage-based weights are adopted to limit the influence of outliers in the covariates. Therefore, the proposed robust estimator would be insensitive to small deviations from the assumed model and useful to deal with the outliers, which are also demonstrated by the simulation study. The asymptotic properties established on the assumed model with no outliers and the insensitivity to small deviation from the model assumptions are some desirable features which a robust procedure should achieve (Huber 1981).

4 Simulation studies

To evaluate the performance of the proposed method, we conduct simulation studies in this section. We compare the proposed robust inverse probability-weighted (R-IPW) method with the robust complete-case (R-CC) method (He et al. 2005) as well as their nonrobust version. The robust methods reduce their corresponding nonrobust versions when $w(x) = 1$ and $\psi(x) = x$. We calculated the bias, standard error (SE), and mean squared error (MSE) of $\hat{\beta}$, as well as the integrated mean squared error (IMSE) of $\hat{f}(\cdot)$ through Monte Carlo simulation in each study. Here, the IMSE is defined as $\int(\hat{f}(t) - f_0(t))^2 \mathrm{d}t$.
Study 1 In this study, we consider a binary partial linear model with dropouts. The marginal mean model is considered to be

$$\ln(\mu_{ij}/(1-\mu_{ij})) = X_{ij}\beta_0+0.5\cos(\pi T_{ij}), \quad i = 1, \ldots, 400, \quad m = 1, \ldots, 6, \tag{11}$$

where $\beta_0 = 0.5$, X_{ij} and T_{ij} are independently drawn from uniform distributions on $(-0.8, 0.8)$ and $(-0.5, 0.5)$, respectively, $R_i(\rho)$ is the correlation matrix of Y_i considered to be first-order autoregressive (AR1) structure with correlation parameter $\rho = 0.6$. The correlated binary data are generated using the method proposed in Preisser et al. (2002). The working AR1 and working independence structures are chosen for GEE.

To investigate the case of missingness, the values of the indicators R_{ij} are generated from the model

$$\ln \frac{\lambda_{ij}}{1-\lambda_{ij}} = \gamma_0 + \gamma_1 Y_{ij-1} + \gamma_2 X_{ij}, \tag{12}$$

where $(\gamma_0, \gamma_1, \gamma_2)^T$ is taken to be $(1.5, 1, -1)^T$ which leads to about 25 % missingness.

To assess the robustness of the proposed method, we consider two ways to create outliers in the data as follows:

C1: Twelve completely observed points are randomly selected. Their covariates X_{ij} are replaced by $X_{ij} + 2$, but their response Y_{ij} is set to be 1 regardless of their original values before contamination.

C2: Twenty-four completely observed points are randomly selected. The contamination of the covariates and response is the same as that in C1.

We conduct 500 replications in each study. The number of internal knots is chosen to be 4, the integer part of $2400^{1/5}$.

As addressed previously, the bias-corrected robust Wang–Carey estimating equation approach proposed in Qin et al. (2008) is used by the proposed method to estimate the correlation parameter. For the nonrobust CC method, we consider the following commonly used estimator proposed in Diggle et al. (2002) to estimate the correlation parameter ρ which is defined as:

$$\hat{\rho} = \frac{2\left[\sum_{i=1}^n b_i - \left\{|(\sum_{i=1}^n b_i)^2 - (\sum_{i=1}^n a_i)^2|\right\}^{1/2}\right]}{\sum_{i=1}^n a_i},$$

where $a_i = 2\sum_{1\leq j<k\leq m} e_{ij}e_{ik}$, $b_i = 2\sum_{j=1}^m e_{ij}^2 - e_{i1}^2 - e_{im}^2$, $e_{ij} = \frac{R_{ij}(Y_{ij}-\hat{\mu}_{ij})}{(\hat{\phi} v(\hat{\mu}_{ij}))^{1/2}}$, $i = 1, \ldots, n$, $j = 1, \ldots, m$. For the robust CC method, we consider its robust version with e_{ij} replaced by $e_{ij}^R = \psi(e_{ij})$ to estimate ρ and for the nonrobust IPW method, we consider its IPW version with e_{ij} replaced by $e_{ij}^{\mathrm{IPW}} = \frac{e_{ij}}{\pi_{ij}(\hat{\gamma})}$.

Table 1 summarizes the simulation results for the mean model in Study 1. It can be observed that the proposed robust IPW method shows its bias and efficiency gain over the other three estimates in comparison. Specifically, the proposed robust IPW method shows smaller bias, MSE and IMSE than the robust CC method in all the cases. For example, in the case of no outliers, the bias of the estimate by the robust CC method is 0.0588, but the estimate by the proposed robust IPW method is only 0.0203. Moreover, the MSE of the estimate for the regression parameter and IMSE of the estimate for the nonparametric function are 0.0077 and 0.0191, respectively, by the proposed robust IPW method, which are much smaller than 0.0102 and 0.0329 by the robust CC method. Similar findings can be obtained in the case of contamination where the proposed robust IPW method further shows its strength in dealing with outliers and missing data.

Moreover, we consider the effects of working correlation on the compared estimators. It is further demonstrated that the methods which ignore the correlation structures among the observations within the same subject can lose estimation efficiency, and this is more obvious in the estimation of the nonparametric function $f_0(\cdot)$.

Table 1 Simulation results for the mean model in Study 1

	$\hat{\beta}$				$\hat{f}$
	BIAS	SE	ESE	MSE	IMSE
Working AR1 structure					
NC					
R-IPW	0.0203	0.0853	0.0851	0.0077	0.0191
R-CC	0.0588	0.0818	0.0831	0.0102	0.0329
NR-IPW	0.0201	0.0852	0.0854	0.0077	0.0183
NR-CC	0.0608	0.0823	0.0829	0.0105	0.0330
C1					
R-IPW	0.0508	0.0898	0.0877	0.0106	0.0192
R-CC	0.0748	0.0833	0.0837	0.0125	0.0337
NR-IPW	0.0916	0.0857	0.0839	0.0157	0.0185
NR-CC	0.1016	0.0821	0.0806	0.0171	0.0342
C2					
R-IPW	0.0818	0.0876	0.0884	0.0144	0.0195
R-CC	0.0929	0.0819	0.0841	0.0153	0.0341
NR-IPW	0.1400	0.0826	0.0815	0.0264	0.0185
NR-CC	0.1383	0.0794	0.0788	0.0254	0.0352
Working independence structure					
NC					
R-IPW	0.0173	0.1104	0.1128	0.0125	0.0408
R-CC	0.0640	0.1042	0.1083	0.0150	0.0576
NR-IPW	0.0180	0.1108	0.1128	0.0126	0.0411
NR-CC	0.0653	0.1046	0.1082	0.0152	0.0582
C1					
R-IPW	0.0573	0.1074	0.1086	0.0148	0.0395
R-CC	0.0858	0.1021	0.1053	0.0178	0.0591
NR-IPW	0.0951	0.1051	0.1041	0.0201	0.0396
NR-CC	0.1115	0.1000	0.1018	0.0224	0.0610
C2					
R-IPW	0.0904	0.1022	0.1055	0.0186	0.0381
R-CC	0.1067	0.0982	0.1030	0.0210	0.0597
NR-IPW	0.1475	0.0992	0.0987	0.0316	0.0385
NR-CC	0.1506	0.0950	0.0972	0.0317	0.0628

SE standard error, *ESE* estimated standard error obtained from the asymptotic theory, *MSE* mean squared error, *IMSE* integrated MSE, NC no contamination, *C1* and *C2* refer to Contaminations 1 and 2, *R-IPW* the robust *IPW-GEE; R-CC* robust complete case *GEE*, *NR-IPW* and NR-CC are the corresponding nonrobust versions

We also calculate the estimated standard errors for the estimators of $\hat{\beta}_0$ using the large sample approximation (8)–(10). The average estimated standard errors based on 500 replications are also provided in Table 1. It can be found that the empirical standard errors are close to the estimated ones, indicating that the large-sample estimate of the variance for the proposed estimator is quite acceptable.

Table 3 presents the estimates of the parameters in the dropout model. It can be observed that Contaminations C1 and C2 produce notable impacts on the estimation of the parameters in the dropout model. And the proposed WML method shows its robustness against these contaminations with smaller biases than the CML method. Note that the results for WML and CML are identical under the case of no outliers. The reason is that in this case, the weight $w_{ik}^{(DP)}$ used to down weight the impact of the outliers in the covariate is all 1 which indicates that there are no possible outliers in the covariate generated from the model in Study 1.

Study 2 We consider a normal partial linear model with dropouts. The marginal mean model is taken as

$$\mu_{ij} = X_{ij}^T\beta_0 + 0.5\sin(2T_{ij}), \quad i = 1, \ldots, 400, \quad j = 1, \ldots, 6, \tag{13}$$

where $\beta_0 = 0.5$. The covariates are generated as follows: $X_{ij} = u_{ij} + b_{1,ij}$, $T_{ij} = u_{ij} + b_{2,ij}$ where u_{ij}, $b_{1,ij}$ and $b_{2,ij}$ are independently drawn from a uniform distribution on $(-0.5, 0.5)$. The random error $e_i = (e_{i1}, \ldots, e_{im})^T$ of the ith subject follows a multivariate normal distribution with mean zero and covariance matrix $R_i(\rho)\sigma^2$ where $R_i(\rho)$ is the correlation matrix, also chosen to be AR1 structure with $\rho = 0.6$, and σ^2 is taken to be 1.

The values of the indicators R_{ij} are generated from the model similar to model (12) in Study 1 except that the parameter vector $(\gamma_0, \gamma_1, \gamma_2)^T$ is taken to be $(3, 1, -1)^T$ which yields about 17 % missingness.

Similar to Study 1, two ways to create outliers are considered. The difference is that here the selected observed points have their covariates X_{ij} replaced by $X_{ij} - 1$ and their response Y_{ij} replaced by $Y_{ij} - 3$.

The simulation results for the mean model in Study 2 are presented in Table 2. Similar findings to Study 1 can be obtained. The proposed robust IPW method generally outperforms the other three methods with smaller bias, MSE and IMSE.

Table 3 also gives the estimates of the parameters in the dropout model in Study 2. Similar to Study 1, the proposed WML method shows its robustness against these contaminations.

Study 3 We consider another normal partial linear model with different nonparametric function from that in Study 2, which is taken as:

$$\mu_{ij} = X_{ij}^T\beta_0 - 0.5\exp(T_{ij}), \quad i = 1, \ldots, 200, \quad j = 1, \ldots, 6, \tag{14}$$

where $\beta_0 = 1.2$. The covariates are generated as follows: $X_{ij} = u_{ij} + b_{1,ij}$, $T_{ij} = u_{ij} + b_{2,ij}$ where $b_{1,ij}$ are generated from normal distribution with mean 2 and standard deviation 0.5, u_{ij}, and $b_{2,ij}$ are independently drawn from a uniform distribution on $(-0.25, 0.25)$. The random error $e_i = (e_{i1}, \ldots, e_{im})^T$ follows a multivariate normal distribution with mean zero and covariance matrix $R_i(\rho)\sigma^2$ where $R_i(\rho)$ is chosen to be AR1 structure with $\rho = 0.6$ and σ^2 is taken to be 0.6.

The values of the indicators R_{ij} are generated from the same model as that in Study 2 which yields about 18 % missingness in this study.

Table 2 Simulation results for the mean model in Study 2

	$\hat{\beta}$				$\hat{f}$
	BIAS	SE	ESE	MSE	IMSE
Working AR1 structure					
NC					
R-IPW	0.0026	0.0463	0.0514	0.0021	0.0038
R-CC	0.0269	0.0441	0.0494	0.0027	0.0083
NR-IPW	0.0024	0.0466	0.0502	0.0022	0.0045
NR-CC	0.0262	0.0433	0.0480	0.0026	0.0097
C1					
R-IPW	0.0521	0.0697	0.0625	0.0076	0.0028
R-CC	0.0946	0.0526	0.0537	0.0117	0.0065
NR-IPW	0.1128	0.0603	0.0626	0.0164	0.0037
NR-CC	0.1512	0.0509	0.0588	0.0255	0.0073
C2					
R-IPW	0.1097	0.0730	0.0679	0.0174	0.0024
R-CC	0.1577	0.0533	0.0573	0.0277	0.0053
NR-IPW	0.2132	0.0600	0.0678	0.0490	0.0019
NR-CC	0.2544	0.0514	0.0643	0.0674	0.0060
Working independence structure					
NC					
R-IPW	0.0027	0.0632	0.0674	0.0040	0.0044
R-CC	0.0242	0.0595	0.0640	0.0041	0.0090
NR-IPW	0.0019	0.0630	0.0664	0.0040	0.0049
NR-CC	0.0235	0.0584	0.0627	0.0040	0.0101
C1					
R-IPW	0.0586	0.0766	0.0720	0.0093	0.0044
R-CC	0.0867	0.0651	0.0663	0.0118	0.0077
NR-IPW	0.1153	0.0733	0.0742	0.0187	0.0061
NR-CC	0.1437	0.0645	0.0701	0.0248	0.0087
C2					
R-IPW	0.1185	0.0765	0.0742	0.0199	0.0069
R-CC	0.1483	0.0651	0.0683	0.0262	0.0068
NR-IPW	0.2155	0.0713	0.0775	0.0515	0.0057
NR-CC	0.2460	0.0648	0.0738	0.0647	0.0075

See Table 1

Similar to Study 2, two ways to create outliers are considered. Twelve and twenty-four completely observed points are randomly selected, respectively, and the selected observed points have their covariates X_{ij} replaced by $X_{ij} - 1$ and their response Y_{ij} replaced by $Y_{ij} - 3.5$.

In this study, the estimation of the correlation parameter and dispersion parameter is also investigated along with the mean model. The simulation results for the mean model are presented in Table 4 and the results for the correlation and dispersion

Table 3 The simulation results for the dropout model in Studies 1 and 2

	$\gamma_0 = 1.5$			$\gamma_1 = 1$			$\gamma_2 = -1$		
	BIAS	SE	MSE	BIAS	SE	MSE	BIAS	SE	MSE
Study 1									
NC									
WML	0.0083	0.1067	0.0115	−0.0055	0.1600	0.0256	0.0043	0.1828	0.0334
CML	0.0083	0.1067	0.0115	−0.0055	0.1600	0.0256	0.0043	0.1828	0.0334
C1									
WML	−0.0063	0.1030	0.0106	−0.0059	0.1603	0.0257	0.1381	0.1493	0.0414
CML	−0.0085	0.1024	0.0106	−0.0067	0.1605	0.0258	0.2212	0.1384	0.0681
C2									
WML	−0.0141	0.1030	0.0108	−0.0134	0.1618	0.0264	0.2398	0.1340	0.0754
CML	−0.0151	0.1025	0.0107	−0.0155	0.1622	0.0265	0.3618	0.1204	0.1454
	$\gamma_0 = 3$			$\gamma_1 = 1$			$\gamma_2 = -1$		
	BIAS	SE	MSE	BIAS	SE	MSE	BIAS	SE	MSE
Study 2									
NC									
WML	0.0166	0.1295	0.0170	0.0103	0.1042	0.0110	0.0032	0.2348	0.0551
CML	0.0168	0.1287	0.0168	0.0113	0.1018	0.0105	0.0026	0.2332	0.0544
C1									
WML	−0.0042	0.1285	0.0165	−0.0543	0.0996	0.0129	−0.0249	0.2428	0.0596
CML	−0.0200	0.1263	0.0163	−0.0890	0.0973	0.0174	−0.0272	0.2412	0.0589
C2									
WML	−0.0247	0.1248	0.0162	−0.1087	0.0929	0.0205	−0.0352	0.2401	0.0589
CML	−0.0478	0.1210	0.0169	−0.1638	0.0887	0.0347	−0.0381	0.2382	0.0582
	$\gamma_0 = 3$			$\gamma_1 = 1$			$\gamma_2 = -1$		
	BIAS	SE	MSE	BIAS	SE	MSE	BIAS	SE	MSE
Study 3									
NC									
WML	−0.0050	0.5992	0.3591	0.0051	0.1383	0.0192	0.0082	0.2712	0.0736
CML	−0.0046	0.5967	0.3561	0.0049	0.1386	0.0192	0.0081	0.2704	0.0732
C1									
WML	0.3783	0.5850	0.4854	−0.2383	0.1045	0.0677	0.0015	0.2671	0.0714
CML	0.4193	0.5829	0.5157	−0.2649	0.1077	0.0818	0.0018	0.2660	0.0708
C2									
WML	0.5897	0.5721	0.6751	−0.3659	0.0893	0.1419	−0.0019	0.2593	0.0672
CML	0.6379	0.5672	0.7286	−0.3956	0.0937	0.1653	−0.0021	0.2573	0.0662

WML weighted maximum likelihood method, *CML* classical maximum likelihood method

Table 4 Simulation results for the mean model in Study 3

	$\hat{\beta}$				$\hat{f}$
	BIAS	SE	ESE	MSE	IMSE
Working AR1 structure					
NC					
R-IPW	0.0056	0.0391	0.0388	0.0016	0.0027
R-CC	0.0269	0.0364	0.0376	0.0020	0.0044
NR-IPW	0.0055	0.0372	0.0372	0.0014	0.0027
NR-CC	0.0267	0.0344	0.0360	0.0019	0.0052
C1					
R-IPW	0.0645	0.0441	0.0461	0.0061	0.0173
R-CC	0.0854	0.0389	0.0428	0.0088	0.0436
NR-IPW	0.1151	0.0414	0.0495	0.0150	0.0791
NR-CC	0.1359	0.0383	0.0484	0.0199	0.1163
C2					
R-IPW	0.1178	0.0462	0.0498	0.0160	0.0689
R-CC	0.1397	0.0411	0.0469	0.0212	0.1026
NR-IPW	0.2052	0.0450	0.0555	0.0441	0.2065
NR-CC	0.2272	0.0413	0.0549	0.0533	0.2493

See Table 1

parameters are shown in Table 5. For the mean model, similar findings to previous studies can be obtained. The proposed robust IPW method generally shows its gain of bias and efficiency over the other three methods with smaller bias, MSE and IMSE. For the estimation of correlation and dispersion parameters, the proposed estimators also show their gain of bias and efficiency particularly in the case of outliers. For example, in the case of outliers (C2), the bias and MSE of the proposed estimate for the correlation parameter are -0.0124 and 0.0015, respectively, which are much smaller than -0.0741 and 0.0066 by the robust CC method.

We also conduct additional simulation studies to investigate the performance of the proposed estimator for the mean regression model in some case of nonmonotonic missingness and similar findings to the case of monotonic missingness can be observed which are, therefore, omitted for brevity.

5 Application to a real data

We apply the proposed method to a real data set from a clinical trial study (Drake et al. 1998) for illustration. The purpose of this study is to assess an assertive community treatment for patients with occurring severe mental illness and substance abuse disorder. This study involved 203 patients who were enrolled to be measured every half year for a total of seven visits.

The outcome variable is the score for general quality of life (SQL). The interested covariate variables include the Substance Abuse Treatment Scale (SATS), the Treat-

Table 5 Simulation results for the correlation parameter ρ and dispersion parameter σ^2 in Study 3

	$\rho = 0.6$		$\sigma^2 = 0.3$	
	BIAS	SE	BIAS	SE
NC				
R-IPW	0.0229	0.0015	0.0004	0.0090
R-CC	−0.0054	0.0013	−0.0025	0.0090
NR-IPW	0.0209	0.0029	0.0248	0.0051
NR-CC	0.0072	0.0013	0.0240	0.0050
C1				
R-IPW	0.0107	0.0013	0.0205	0.0106
R-CC	−0.0424	0.0030	0.0178	0.0100
NR-IPW	−0.0525	0.0061	0.0945	0.0147
NR-CC	−0.0658	0.0057	0.0934	0.0145
C2				
R-IPW	−0.0124	0.0015	0.0390	0.0117
R-CC	−0.0741	0.0066	0.0410	0.0122
NR-IPW	−0.0963	0.0118	0.1578	0.0327
NR-CC	−0.1145	0.0144	0.1562	0.0320

See Table 1

ment (Trt) (1 if the patient is in the treatment group, 0 otherwise), Education (Edu) (1 if the patient took higher school education or above, 0 otherwise), Diagnosis (Diag) (1 if the patient had schizophrenia/schizoaffective disorder, 0 if the patient had bipolar disorder), gender and age. In our analysis, the standardized age $[(\text{Age (in year)} - 18)/45]$ is used.

Usually, the relationship between the outcome and the covariate Age is nonlinear; thus, we adopt the following partial linear model to fit the data set:

$$\begin{aligned} &E[\text{SQL}|\text{Age, SATS, Trt, Edu, Diag, Gender}] \\ &\quad = f(\text{Age}) + \beta_{1\,\text{SATS}} + \beta_{2\,\text{Trt}} + \beta_{3\,\text{Edu}} + \beta_{4\,\text{Diag}} + \beta_{5\,\text{Gender}}, \end{aligned} \tag{15}$$

where $f(\cdot)$ is an unknown smooth function. We adopt a four-order regression spline with one internal knot to approximate $f(\cdot)$.

During the period of this study, some scheduled measurements were missing for some patients, and the monotonic missing mechanism sometimes may not hold. For illustration, we assume the dropouts occur when the first missingness appears during the following six visits after the first baseline visit. Then, the total missing rate is about 15 %. Moreover, some observations in the data set missed both values of outcome and covariate SATS (the proportion is less than 2.3 %). However, the discussion of missingness in both the response and covariates is beyond the scope of this paper. For simplicity, we use only the baseline values of SATS. Note that if the working independence matrix is adopted, the missingness in both response and covariates will not influence the implementation of the proposed method. However, if a correlated

Table 6 Regression coefficient estimates in the analysis of the real data

	R-IPW	R-CC	NR-IPW	NR-CC
Working AR1 structure				
Gender	0.2320 (0.1840)	0.2428 (0.1815)	0.2632 (0.1741)	0.2705 (0.1709)
Edu	0.0951 (0.1868)	0.0949 (0.1839)	0.1010 (0.1752)	0.0993 (0.1727)
Diag	0.1955 (0.1769)	0.2368 (0.1748)	0.1565 (0.1686)	0.1950 (0.1664)
SATS	0.1317 (0.0292)	0.1385 (0.0292)	0.1321 (0.0275)	0.1380 (0.0274)
Trt	0.0350 (0.1689)	−0.0052 (0.1654)	0.0440 (0.1585)	0.0042 (0.1560)
Working independence structure				
Gender	0.2090 (0.1846)	0.2190 (0.1812)	0.2419 (0.1738)	0.2490 (0.1706)
Edu	0.0961 (0.1892)	0.0987 (0.1863)	0.0970 (0.1763)	0.0986 (0.1740)
Diag	0.1875 (0.1819)	0.2135 (0.1786)	0.1593 (0.1722)	0.1863 (0.1694)
SATS	0.1171 (0.0321)	0.1226 (0.0314)	0.1193 (0.0299)	0.1249 (0.0294)
Trt	0.0456 (0.1720)	0.0132 (0.1686)	0.0608 (0.1604)	0.0283 (0.1577)

The figures in the parenthesis are standard errors

structure is used, e.g., the AR1 structure used here, the covariates are required to be completely observed.

The dropout model taken in this analysis is

$$\ln \frac{\lambda_{ij}}{1-\lambda_{ij}} = \gamma_0 + \gamma_1 \mathrm{SQL}_{ij-1} + \gamma_2 \mathrm{SATS}_{ij}. \tag{16}$$

Table 6 presents the results of estimates for the linear component of model (15) and Fig. 1 shows the estimate of the nonparametric function $f(\cdot)$. From Fig. 1, an obvious nonlinear relation between the outcome SQL and the covariate Age can be observed. Although the mean of the outcome generally decreases with Age, the trend varies among different intervals of the Age. An interesting feature is that the Age effect appears to be constant within the interval of years (27, 36), while in the other part before 54 years old, SQL decreases with Age obviously. Particularly, the curves estimated by the robust methods appear to drop faster than the nonrobust ones. For the estimates of the regression coefficients, it is found that only the effect of SATS is significantly positively related with SQL at the 0.05 level by all the methods, which is also consistent with the results based on the linear regression model in Wang et al. (2011). Although the conclusions are consistent to each other, there exists some obvious numerical differences among these methods. The robust-CC method which does not consider the influence of missingness even gives negative estimate of the Trt effect. All of these indicate the influence of possible outliers and missingness.

To explore possible outliers in the data set, we calculate the weights based on the proposed method

$$\varpi_{ij} = w(x_{ij})\psi(r_i)/r_i, \tag{17}$$

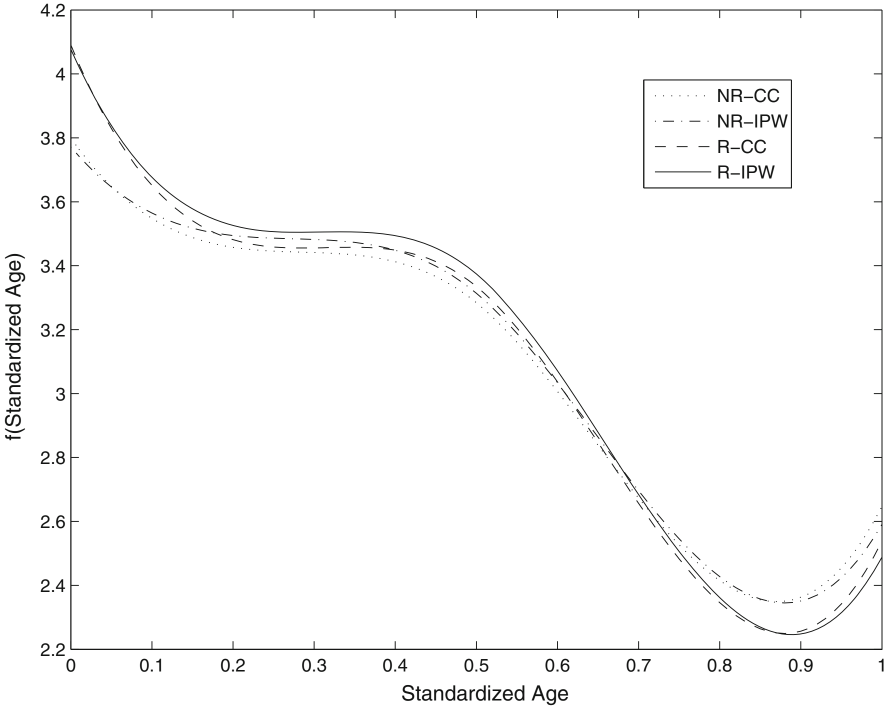

Fig. 1 The estimated function f on standardized age

where $r_i = (Y_{ij} - D_{ij}^T\hat{\theta})/(v_i(\hat{\theta})\hat{\phi})^{1/2}$, $w(\cdot)$ and ψ are defined in Sect. 2. Small weights indicate that the corresponding observations are possible outliers. There are totally 18 observations with weights less than 0.7. We investigate two observations with smallest weights less than 0.5 which are from Patients 51 and 40. The interesting findings are that both observations have small SQL values of 1.5 and 1, respectively, but have very large SATS values of 8. These findings prompt us to pay more attention to these patients and explore the possible reasons.

In summary, the GPLM enables us to gain insight into the relationship inherited in the data, and the robust and inverse probability-weighted approaches are able to provide more accurate and reliable results.

Appendix

Lemma 1 *Under conditions* (C.4) and (C.5)*, there exist* $\alpha_0 \in R^{N_k}$ *depending on* f_0*, and a constant* C_4 *depending only on* l *and* C_0 *such that*

$$\sup_{t\in[0,1]} |f_0(t) - \pi^T(t)\alpha_0| \le C_4 k_n^{-r}.$$

The proof of Lemma 1 *follows easily from Theorem 12.7 in* Schumaker (1981).

Proof of Theorem 2 We first introduce some notations. Let

$$\xi(\beta,\alpha)=\begin{bmatrix}\xi_1\\ \xi_2\end{bmatrix}=\begin{bmatrix}K_n^{1/2}(\beta-\beta_0)\\ k_n^{-1/2}H_n(\alpha-\alpha_0)+k_n^{1/2}H_n^{-1}M^T\Omega_0X(\beta-\beta_0)\end{bmatrix},$$

and $\hat{\xi}=(\hat{\xi}_1^T,\hat{\xi}_2^T)^T=\xi(\hat{\beta},\hat{\alpha})$, where $H_n^2=k_nM^T\Omega_0M$. Then, the robust-weighted estimating equation can be expressed as:

$$U_{\xi,n}(\mu(\xi))=\sum_{i=1}^{n}D_i\Delta_i(\mu_i(\xi))A_i^{-1/2}(\mu_i(\xi))C_i^{-1}J_i(\mu_i(\xi),\hat{\gamma}). \tag{18}$$

We denote $\widetilde{X}_i=K_n^{-1/2}X_i^*,\ \widetilde{M}_i=k_n^{1/2}H_n^{-1}d_i,\ R_{ni}=d_i^T\alpha_0-f_0(t_i)$, and $\zeta_i=\widetilde{X}_i^T\xi_1+\widetilde{M}_i^T\xi_2+R_{ni}$, then $\eta_i(\theta)=D_i^T\theta=\eta_{0,i}+\zeta_i, i=1,\ldots,n$, where $\eta_{0,i}=X_i^T\beta_0+f_0(t_i)$. We further denote

$$N=\begin{bmatrix}K_n^{1/2} & -K_n^{-1/2}X^T\Omega_0M(M^T\Omega_0M)^{-1}\\ 0 & k_n^{-1/2}H_n^{-1}\end{bmatrix}.$$

Then, (18) can be written as:

$$\begin{aligned}\Psi(\mu(\xi),\hat{\gamma})&=\begin{pmatrix}\Psi_1(\mu(\xi),\hat{\gamma})\\ \Psi_2(\mu(\xi),\hat{\gamma})\end{pmatrix}=NU_{\xi}(\mu(\xi))\\ &=\begin{pmatrix}\sum_{i=1}^{n}K_n^{-1/2}X_i^*\Delta_i^T(\mu_i(\xi))A_i^{-1/2}(\mu_i(\xi))C_i^{-1}J_i(\mu_i(\xi),\hat{\gamma})\\ \sum_{i=1}^{n}k_n^{1/2}H_n^{-1}d_i\Delta_i^T(\mu_i(\xi))A_i^{-1/2}(\mu_i(\xi))C_i^{-1}J_i(\mu_i(\xi),\hat{\gamma})\end{pmatrix}\\ &=\sum_{i=1}^{n}\tilde{D}_i\Delta_i^T(\mu_i(\xi))A_i^{-1/2}(\mu_i(\xi))C_i^{-1}J_i(\mu_i(\xi),\hat{\gamma}),\end{aligned} \tag{19}$$

where $\tilde{D}_i=(X_i^{*T}K_n^{-1/2},\pi_i^TH_n^{-1}k_n^{1/2})^T$.

Combining (C.7) and (C.8), both Eqs. (18) and (19) give the same root for ξ as our estimator. We denote

$$\begin{aligned}\Phi(\xi)=\begin{pmatrix}\Phi_1(\xi)\\ \Phi_2(\xi)\end{pmatrix}=\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}+\sum_{i=1}^{n}\Bigg\{&\tilde{D}_i\Delta_{0,i}^TA_{0,i}^{-1/2}C_i^{-1}J_{0,i}\\ &-\left[\sum_{i=1}^{n}\tilde{D}_i\Delta_{0,i}^TA_{0,i}^{-1/2}C_i^{-1}\frac{\partial}{\partial\gamma}J_{0,i}\right]\Sigma_{\gamma}^{-1}G_{\gamma,i}(\gamma_0)\Bigg\}.\end{aligned} \tag{20}$$

The zero $\tilde{\xi}$ of $\Phi(\xi)$,

$$\tilde{\xi} = \begin{pmatrix} \tilde{\xi}_1 \\ \tilde{\xi}_2 \end{pmatrix} = -\sum_{i=1}^{n} \Bigg\{ \tilde{D}_i \Delta_{0,i}^T A_{0,i}^{-1/2} C_i^{-1} J_{0,i} + \left[\sum_{i=1}^{n} \tilde{D}_i \Delta_{0,i}^T A_{0,i}^{-1/2} C_i^{-1} \frac{\partial}{\partial \gamma} J_{0,i} \right] \Sigma_\gamma^{-1} G_{\gamma,i}(\gamma_0) \Bigg\}. \qquad (21)$$

is not an estimator. In the following, we will show that the difference between $\tilde{\xi}$ and $\hat{\xi}$ is small.

Let $a \in R^{p+N_k}$, satisfying $a^T a = 1$. We expand $a^T \Psi(\mu(\xi), \hat{\gamma})$ in a Taylor series and have

$$\begin{aligned} a^T \Psi(\mu(\xi), \hat{\gamma}) &= a^T \Psi(\mu(\eta_0 + \zeta), \hat{\gamma}) \\ &= \sum_{i=1}^{n} a^T \tilde{D}_i \Delta_i^T \mu_i(\eta_0 + \zeta) A_i^{-1/2}(\mu_i(\eta_0 + \zeta)) C_i^{-1} J_i(\mu_i(\eta_0 + \zeta), \hat{\gamma}) \\ &= \sum_{i=1}^{n} a^T \tilde{D}_i \Delta_{0,i}^T A_{0,i}^{-1/2} C_i^{-1} J_i(\mu_{0,i}, \hat{\gamma}) \\ &\quad + \sum_{i=1}^{n} a^T \tilde{D}_i \Delta_{0,i}^T A_{0,i}^{-1/2} C_i^{-1} \frac{\partial}{\partial \mu_i} J_i(\mu_{0,i}, \hat{\gamma}) \Delta_{0,i} \zeta_i \\ &\quad + \sum_{i=1}^{n} \frac{\partial}{\partial \mu_i} (a^T \tilde{D}_i \Delta_{0,i}^T A_{0,i}^{-1/2}) C_i^{-1} J_i(\mu_{0,i}, \hat{\gamma}) \Delta_{0,i} \zeta_i + R_n^{**}(\mu^*, \hat{\gamma}) \\ &=: A_1 + A_2 + A_3 + A_4, \end{aligned} \qquad (22)$$

where $R_n^{**}(\mu^*, \hat{\gamma}) = \sum_{i=1}^{n} R_{ni}^{**}(\mu_i^*, \hat{\gamma})$ and $R_{ni}^{**}(\mu_i^*, \hat{\gamma}) = \frac{1}{2} \zeta_i^T \Delta_i^T \frac{\partial^2}{\partial \mu_i \mu_i^T} (a^T \tilde{D}_i \Delta_i (\mu_i) A_i^{-1/2}(\mu_i) C_i^{-1} \cdot J_i(\mu^*, \hat{\gamma})) \Delta_i \zeta_i$ evaluated at $\mu_i^* = g^{-1}(\mu_{0,i} + \tau_i \zeta_i)$ for $i = 1, \ldots, n$ with $0 < \tau_i < 1$.

We first consider A_1, and expand A_1 with respect to γ, then we have

$$\begin{aligned} A_1 &= \sum_{i=1}^{n} a^T \tilde{D}_i \Delta_{0,i}^T A_{0,i}^{-1/2} C_i^{-1} J_{0,i} \\ &\quad + \sum_{i=1}^{n} a^T \tilde{D}_i \Delta_{0,i}^T A_{0,i}^{-1/2} C_i^{-1} \frac{\partial}{\partial \gamma} J_{0,i} (\hat{\gamma} - \gamma_0) \\ &\quad + \frac{1}{2} \sum_{i=1}^{n} (\hat{\gamma} - \gamma_0)^T \frac{\partial}{\partial \gamma \partial \gamma^T} \left[a^T \tilde{D}_i \Delta_{0,i}^T A_{0,i}^{-1/2} C_i^{-1} J_i(\mu_{0,i}, \gamma^*) \right] (\hat{\gamma} - \gamma_0) \\ &=: A_{1,1} + A_{1,2} + A_{1,3}, \end{aligned} \qquad (23)$$

where γ^* is the point on the line between $\hat{\gamma}$ and γ_0.

For $A_{1,2}$, by conditions (C.1)–(C.3), it is not difficult to obtain $n^{1/2}(\hat{\gamma} - \gamma_0) = -\Sigma_{\gamma}^{-1}(n^{-1/2}G_{\gamma_0,n}) + o_p(1)$. Combining conditions (C.6) and (C.7), we have

$$A_{1,2} = -\left[\sum_{i=1}^{n} a^T \tilde{D}_i \Delta_{0,i} A_{0,i}^{-1/2} C_i^{-1} \frac{\partial}{\partial \gamma} J_{0,i}\right] \Sigma_{\gamma}^{-1} \sum_{i=1}^{n} G_{\gamma,i} + o_p(k_n^{1/2}).$$

Note that $||\hat{\gamma} - \gamma_0|| = O_p(n^{-1/2})$, it is not difficult to show $A_{1,3} = o_p(k_n^{1/2})$. Combining all these results, we can get

$$\begin{aligned} A_1 = \sum_{i=1}^{n} \Bigg[& a^T \tilde{D}_i \Delta_{0,i}^T A_{0,i}^{-1/2} C_i^{-1} J_{0,i} \\ & - \left\{\sum_{i=1}^{n} a^T \tilde{D}_i \Delta_{0,i} A_{0,i}^{-1/2} C_i^{-1} \frac{\partial}{\partial \gamma} J_{0,i}\right\} \Sigma^{-1} G_{\gamma,i} \Bigg] + o_p(k_n^{1/2}). \end{aligned} \quad (24)$$

Then, let us turn to consider A_2. Similarly, applying Taylor expansion with respect to γ we have

$$\begin{aligned} A_2 &= \sum_{i=1}^{n} a^T \tilde{D}_i \Delta_{0,i}^T A_{0,i}^{-1/2} \frac{\partial}{\partial \mu_i} J_i(\mu_{0,i}, \hat{\gamma}) \Delta_{0,i} \zeta_i \\ &= \sum_{i=1}^{n} a^T \tilde{D}_i \Delta_{0,i}^T A_{0,i}^{-1/2} \frac{\partial}{\partial \mu_i} J_{0,i} \Delta_{0,i} \zeta_i \\ &\quad + \sum_{i=1}^{n} \frac{\partial}{\partial \gamma} \left[a^T \tilde{D}_i \Delta_{0,i}^T A_{0,i}^{-1/2} \frac{\partial}{\partial \mu_i} J_i(\mu_{0,i}, \gamma^*) \Delta_{0,i} \zeta_i \right] (\hat{\gamma} - \gamma_0) \\ &= \sum_{i=1}^{n} a^T \tilde{D}_i \Delta_{0,i}^T A_{0,i}^{-1/2} \frac{\partial}{\partial \mu_i} J_{0,i} \Delta_{0,i} \zeta_i \\ &\quad + \sum_{i=1}^{n} \frac{\partial}{\partial \gamma} \left[a^T \tilde{D}_i \Delta_{0,i}^T A_{0,i}^{-1/2} \frac{\partial}{\partial \mu_i} J_i(\mu_{0,i}, \gamma^*) \Delta_{0,i} \tilde{D}_i \xi \right] (\hat{\gamma} - \gamma_0) \\ &\quad + \sum_{i=1}^{n} \frac{\partial}{\partial \gamma} \left[a^T \tilde{D}_i \Delta_{0,i}^T A_{0,i}^{-1/2} \frac{\partial}{\partial \mu_i} J_i(\mu_{0,i}, \gamma^*) \Delta_{0,i} R_{ni} \right] (\hat{\gamma} - \gamma_0) \\ &=: A_{2,1} + A_{2,2} + A_{2,3}, \end{aligned} \quad (25)$$

where γ^* is the point on the line between $\hat{\gamma}$ and γ_0.

According to conditions (C.6)–(C.8), the result that $||\hat{\gamma} - \gamma_0|| = O_p(n^{-1/2})$ and Lemma 1, it can be obtained that $A_{2,2} = O_p(n^{-1/2} k_n ||\xi||) = o_p(||\xi||)$ and $A_{2,3} = O_p(k_n^{1/2-r})$. Then, we have

$$A_2 = \sum_{i=1}^{n} a^T \tilde{D}_i \Delta_{0,i}^T A_{0,i}^{-1/2} \frac{\partial}{\partial \mu_i} J_{0,i} \Delta_{0,i} \zeta_i + o_p(||\xi||) + o_p(k_n^{1/2}). \quad (26)$$

By tedious and similar derivation, we can show

$$A_3 = \sum_{i=1}^{n} \frac{\partial}{\partial \mu_i}(a^T \tilde{D}_i \Delta_{0,i}^T A_{0,i}^{-1/2}) C_i^{-1} J_{0,i} \Delta_{0,i} \zeta_i + o_p(||\xi||) + o_p(k_n^{1/2}), \quad (27)$$

and

$$\begin{aligned} A_4 &= R_n^{**}(\mu^*, \gamma_0) + o_p(||\xi||) + o_p(k_n^{1/2}) \quad (28) \\ &= \sum_{i=1}^{n} R_{ni}^*(\mu_i^*, \gamma_0) + o_p(||\xi||) + o_p(k_n^{1/2}). \end{aligned}$$

Combining (24) and (26)–(28), we have

$$\begin{aligned} a^T \Psi(\mu(\xi), \hat{\gamma}) &= a^T \Psi(\mu(\eta_0 + \zeta), \hat{\gamma}) \\ &= \sum_{i=1}^{n} \Bigg[a^T \tilde{D}_i \Delta_{0,i}^T A_{0,i}^{-1/2} C_i^{-1} J_{0,i} \\ &\quad - \left\{ \sum_{i=1}^{n} a^T \tilde{D}_i \Delta_{0,i} A_{0,i}^{-1/2} C_i^{-1} \frac{\partial}{\partial \gamma} J_{0,i} \right\} \Sigma^{-1} G_{\gamma,i} \Bigg] \\ &\quad + \sum_{i=1}^{n} a^T \tilde{D}_i \Delta_{0,i}^T A_{0,i}^{-1/2} \frac{\partial}{\partial \mu_i} J_{0,i} \Delta_{0,i} \zeta_i \\ &\quad + \sum_{i=1}^{n} \frac{\partial}{\partial \mu_i}(a^T \tilde{D}_i \Delta_{0,i}^T A_{0,i}^{-1/2}) C_i^{-1} J_{0,i} \Delta_{0,i} \zeta_i \\ &\quad + R_n^{**}(\mu^*, \gamma_0) + o_p(||\xi||) + o_p(k_n^{1/2}). \quad (29) \end{aligned}$$

Then

$$\begin{aligned} a^T (\Psi(\mu(\xi), \hat{\gamma}) - \Phi(\xi)) &= \sum_{i=1}^{n} a^T \tilde{D}_i \Delta_{0,i}^T A_{0,i}^{-1/2} \left\{ \frac{\partial}{\partial \mu_i} J_{0,i} - E \frac{\partial}{\partial \mu_i} J_{0,i} \right\} \Delta_{0,i} \zeta_i \\ &\quad + \sum_{i=1}^{n} a^T \tilde{D}_i \Delta_{0,i}^T A_{0,i}^{-1/2} \left\{ E \frac{\partial}{\partial \mu_i} J_{0,i} \right\} \Delta_{0,i} \zeta_i \\ &\quad + \sum_{i=1}^{n} \frac{\partial}{\partial \mu_i}(a^T \tilde{D}_i \Delta_{0,i}^T A_{0,i}^{-1/2}) C_i^{-1} J_{0,i} \Delta_{0,i} \zeta_i \\ &\quad + R_n^{**}(\mu^*, \gamma_0) + o_p(||\xi||) + o_p(k_n^{1/2}). \quad (30) \end{aligned}$$

Then, using the same arguments as He et al. (2005), it can be shown that

$$\sup_{||\xi|| \le L k_n^{1/2}} ||\Psi(\mu(\xi), \hat{\gamma}) - \Phi(\xi)|| = O_p(k_n^{1/2}), \quad (31)$$

for sufficiently large constant L.

By direct calculation,

$$E||\tilde{\xi}||^2 = O(k_n). \tag{32}$$

Then we have

$$\sup_{||\xi||\leq Lk_n^{1/2}} ||\Psi(\xi) - \xi|| \leq \sup_{||\xi||\leq Lk_n^{1/2}} ||\Psi(\xi) - \Phi(\xi)|| + ||\tilde{\xi}|| = O_p(k_n^{1/2}), \tag{33}$$

which indicates that $\sup_{||\xi||\leq Lk_n^{1/2}} ||\Psi(\xi) - \xi|| \leq Lk_n^{1/2}$ in probability, for sufficiently large L. Thus, Brouwer's fixed point theorem guarantees that there exists a zero $\hat{\xi}$ of $\Psi(\xi)$ with $||\hat{\xi}|| = O_p(k_n^{1/2})$ and hence the optimal convergence rate of the estimator of the nonparametric function can be achieved. Applying the central limit theorem on $\tilde{\xi}_1$, the asymptotic normality of the estimator of $\hat{\beta}$ can be established similarly. □

Acknowledgments The authors are grateful to the Editor and two referees for their constructive suggestions that largely improve the presentation of the paper. This work was partially supported by the National Natural Science Foundation of China (11371100, 11271080) and Shanghai Leading Academic Discipline Project, Project Number: B118.

References

Cantoni, E., Ronchetti, E. (2001). Robust inference for generalized linear models. *Journal of the American Statistical Association*, *96*, 1022–1030.

Chen, B., Zhou, X.H. (2013). Generalized partially linear models for incomplete longitudinal data in the presence of population-level information. *Biometrics*, *69*, 386–395.

Diggle, P. J., Heagerty, P., Liang, K. Y., Zeger, S. L. (2002). *Analysis of longitudinal data* (2nd ed.). Oxford: Oxford University Press.

Drake, R. E., McHugo, G. J., Clark, R. E., Teague, G. B., Xie, H., Miles, K., et al. (1998). Assertive community treatment for patients with co-occurring severe mental illness and substance use disorder: a clinical trial. *American Journal of Orthopsychiatry*, *68*, 201–215.

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

He, X., Fung, W.K., Zhu, Z.Y. (2005). Robust estimation in generalized partial linear models for clustered data. *Journal of the American Statistical Association*, *100*, 1176–1184.

Huber, P. J. (1981). *Robust statistics*. New York: Wiley.

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

Little, R. J. A., Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). NewYork: Wiley.

Preisser, J. S., Lohman, K. K., Rathouz, P. J. (2002). Performance of weighted estimating equations for longitudinal binary data with drop-outs missing at random. *Statistics in Medicine*, *21*, 3035–3054.

Qin, G.Y., Zhu, Z.Y. (2007). Robust estimation in generalized semiparametric mixed models for longitudinal data. *Journal of Multivariate Analysis*, *98*, 1658–1683.

Qin, G.Y., Zhu, Z.Y. (2009). Robustified maximum likelihood estimation in generalized partial linear mixed model for longitudinal data. *Biometrics*, *65*, 52–59.

Qin, G.Y., Zhu, Z.Y., Fung, W.K. (2008). Robust estimating equations and bias correction of correlation parameters for longitudinal data. *Computational Statistics and Data Analysis*, *52*, 4745–4753.

Rice, J. (1986). Convergence rates for partially splined models. *Statistics and Probability Letters*, *4*, 203–208.

Robins, J. M., Rontnitzky, A., Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, *90*, 106–121.

Rousseeuw, P. J., van Zomeren, B. C. (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*, *85*, 633–639.

Schumaker, L. L. (1981). *Spline functions*. New York: Wiley.

Sinha, S.K. (2004). Robust analysis of generalized linear mixed models. *Journal of the American Statistical Association*, *99*, 451–460.

Sinha, S.K. (2012). Robust analysis of longitudinal data with nonignorable missing responses. *Metrika*, *75*, 913–938.

Wang, Y. G., Lin, X., Zhu, M., Bai, Z. (2007). Robust estimation using the Huber function with a data-Dependent tuning constant. *Journal of Computational and Graphical Statistics*, *16*, 468–481.

Wang, J., Xie, H., Fisher, J. H. (2011). *Multilevel models: applications using SAS*. Beijing: China Higher Education Press.

Yi, G.Y., He, W. (2009). Median regression models for longitudinal data with dropouts. *Biometrics*, *65*, 618–625.