Journal of Multivariate Analysis 105 (2012) 32–44

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

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

# Robust empirical likelihood inference for generalized partial linear models with longitudinal data

Guoyou Qin [a,b], Yang Bai [c,*], Zhongyi Zhu [d]

[a] Department of Biostatistics, Fudan University, Shanghai, 200032, China
[b] Key Laboratory of Public Health Safety, Ministry of Education (Fudan University), Shanghai 200032, China
[c] School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, 200433, China
[d] Department of Statistics, Fudan University, Shanghai, 200433, China

ARTICLE INFO

*Article history:*
Received 23 July 2010
Available online 18 August 2011

*AMS 2000 subject classification:*
46N30

*Keywords:*
*B*-spline
Efficiency
Empirical likelihood
Generalized estimating equations
Generalized partial linear models
Longitudinal data
Robustness

ABSTRACT

In this paper, we propose a robust empirical likelihood (REL) inference for the parametric component in a generalized partial linear model (GPLM) with longitudinal data. We make use of bounded scores and leverage-based weights in the auxiliary random vectors to achieve robustness against outliers in both the response and covariates. Simulation studies demonstrate the good performance of our proposed REL method, which is more accurate and efficient than the robust generalized estimating equation (GEE) method (X. He, W.K. Fung, Z.Y. Zhu, Robust estimation in generalized partial linear models for clustered data, Journal of the American Statistical Association 100 (2005) 1176–1184). The proposed robust method is also illustrated by analyzing a real data set.



## 1. Introduction

Generalized partial linear models (GPLMs) can be regarded as an integration of generalized linear models (GLMs) [10] and fully nonparametric models. By involving both parametric and nonparametric components, GPLMs have great flexibility in modeling real data, and therefore have attracted many research interests and are widely used in practice.

The inference for GPLMs is usually based on maximum likelihood method and generalized estimating equation (GEE) method [8]. However, both the classical maximum likelihood and GEE methods are sensitive to outliers. In longitudinal studies, an outlier in a subject-level measurement can result in multiple outliers in the sample. So many robust methods have been developed to limit the impact of the outliers, e.g., [14,3,2,23]. Particularly, for the GPLMs with longitudinal data, He et al. [4] proposed a robust GEE (RGEE) method by using *B*-spline to approximate the nonparametric function. The commonly used sandwich method was adopted to obtain the variance estimation of their proposed RGEE estimator for the parametric component. However, it is well known that the sandwich method usually underestimates the variance of the GEE estimator, which possibly leads to biased statistical inference. For more detail and systematic introductions about the robust statistical methods we can refer to the book of Heritier et al. [6].

The empirical likelihood (EL) method, first developed by Owen [11], is a popular statistical inference method and has attracted a great deal of interests [7,16]. Many advantages of the EL over the normal approximation-based method have

* Corresponding author.
*E-mail addresses:* gyqin@fudan.edu.cn (G. Qin), statbyang@mail.shufe.edu.cn (Y. Bai), zhuzy@fudan.edu.cn (Z. Zhu).

0047-259X/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2011.08.003

been shown in the literature. For example, the shape of confidence regions based on EL is determined totally by data itself. The EL is Bartlett correctable, hence comparable to the bootstrap method. In particular, another attractive feature of EL method is that the statistical inference based on EL does not require variance estimation. Recently, the EL method has been further developed in the areas of partial linear model and longitudinal data analysis. Wang and Jin [25] and Shi and Lau [22] considered EL inference for a partial linear model. Xue and Zhu [26] considered the EL method for a partial linear model with longitudinal data, but they ignored the correlations within subjects, which may lead to the loss of efficiency. Bai et al. [1] constructed a new weighted EL inference by taking into account the correlations within subjects. The EL method may also be influenced by the outliers due to its close relationship with the maximum likelihood and the GEE methods [16]. Owen [13] pointed out that the EL confidence regions may be greatly lengthened in the direction of the outliers. Shi and Lau [21] proposed REL for linear models under median constraints. Qin et al. [15] proposed REL for generalized linear models with longitudinal data by constructing robust auxiliary random vectors.

To the best of our knowledge, it is lack of literature on the REL inference for GPLMs with longitudinal data, possibly due to the complexity of the nonparametric component in the GPLMs. In this article, we develop an REL method for inference of the parametric component in GPLMs with longitudinal data through constructing robust auxiliary random vectors, which utilizes a weight function to downweight the effect of leverage points and a bounded score function on the Pearson residuals to limit the influence of outliers in the responses. We extend our work in [15] to deal with GPLMs with longitudinal data, and the $B$-spline method is adopted to approximate the nonparametric component in the model. Moreover, our method incorporates working correlation matrix into the robust auxiliary random vectors to interpret the correlations within the subjects. Our proposed REL method does not require the estimation of the variance of the proposed estimator, and provides more accurate inference than the existing RGEE method.

The rest of this article is organized as follows. In Section 2, the REL method for GPLMs is proposed. The asymptotic normality of the proposed REL estimator and the asymptotic chi-square distribution of the proposed REL ratio are derived under some regularity assumptions in Section 3. In Section 4, simulation studies are conducted to investigate the performance of the proposed method. A real data set is modeled by a GPLM and analyzed by the proposed method in Section 5. The details of the proof is given in Appendix.

## 2. Model and robust empirical likelihood

### 2.1. GPLMs for longitudinal data

In this article, we consider a longitudinal study with $m$ subjects and $n_i$ observations over time for the $i$th subject ($i = 1, \ldots, m,\ j = 1, \ldots, n_i$) for a total of $n = \sum_{i=1}^m n_i$ observations. Let us denote $\{(x_{ij}, y_{ij}, t_{ij}),\ i = 1, \ldots, m,\ j = 1, \ldots, n_i\}$ as the observed data set, and $E(y_{ij}) = \mu_{0,ij}$ and $\mathrm{var}(y_{ij}) = \phi v(\mu_{0,ij})$, where $\phi$ is a scale parameter and $v(\cdot)$ is a known variance function. We model the longitudinal data with a GPLM, and specify a marginal model on the first two moments of $y_{ij}$. Especially, the marginal mean $\mu_{0,ij}$ is modeled as

$$\eta_{0,ij} = g(\mu_{0,ij}) = x_{ij}^T\beta_0 + f_0(t_{ij}), \qquad \mu_{0,ij} = \mu(\eta_{0,ij}) = g^{-1}(\eta_{0,ij}), \tag{2.1}$$

where $\beta_0$ is a $p$-dimensional vector of regression coefficient with covariates $x_{ij}$, $f_0(\cdot)$ is an unknown smooth function and $g(\cdot)$ is a given link function. Furthermore, we assume that the observations from different subjects are independent. Without loss of generality, we also assume that $t_{ij}$ are all scaled into the interval [0, 1].

Following He et al. [4], we approximate $f_0$ by a regression spline. Let $0 = s_0 < s_1 < \cdots < s_{k_n+1} = 1$ be a partition of the interval [0, 1]. Using the $s_i$ as knots, we have $N_k = k_n + l$ normalized $B$-spline basis functions of order $l - 1$ that form a basis for the linear spline space. Just as pointed out in [4], regression splines have some desirable properties in approximating a smooth function. It often provides good approximations with a small number of knots. The spline approach also treats a nonparametric function as a linear function with the basis functions as pseudo-design variables, and thus any computational algorithm developed for the generalized linear models can be used for the GPLM.

Let $f_0(t)$ be approximated by $\pi(t)^{\mathrm{T}}\alpha_0$, where $\pi(t) = (B_1(t), \ldots, B_{N_k}(t))^{\mathrm{T}}$ is the vector of basis functions, and $\alpha_0 \in R^{N_k}$ is the vector of spline coefficient. This linearizes our regression model so that our regression problem becomes

$$\eta_{ij}(\theta_0) = g(\mu_{ij}(\theta_0)) = x_{ij}^{\mathrm{T}}\beta_0 + \pi(t_{ij})^{\mathrm{T}}\alpha_0 = D_{ij}^T\theta_0, \tag{2.2}$$

where $D_{ij} = (x_{ij}^T, \pi_{ij}^T)^T$, and $\theta_0^{\mathrm{T}} = (\beta_0^{\mathrm{T}}, \alpha_0^{\mathrm{T}})$ is the vector of combined regression parameters. In matrix notations, we let $\mu_i = (\mu_{i1}, \ldots, \mu_{in_i})^T$, $Y_i = (y_{i1}, \ldots, y_{in_i})^T$, where $\mu_{ij} = g^{-1}(D_{ij}^T\theta)$, $D_{ij} = (x_{ij}^T, \pi^T(t_{ij}))^T$, and define $X_i$ and $\pi_i$ in a similar fashion for $i = 1, \ldots, m,\ j = 1, \ldots, n_i$.

**Remark 1.** Selection of knots is generally an important aspect of spline smoothing. Usually, knot selection is less critical for the estimate of $\beta_0$ than for the estimate of $f_0(t)$ (see, [4]). In this article, for the reason that our main focus is inference on the parameter $\beta_0$, and for the sake of simplicity, we use the sample quantiles of $\{t_{ij}, i = 1, \ldots, m,\ j = 1, \ldots, n_i\}$ as knots. Moreover, we use cubic splines (i.e., splines of order 4) and take the number of internal knots to be the integer around $n^{1/5}$. This particular choice is consistent with the asymptotic theory of Section 3, and also performs well in the simulations of Section 4.

### 2.2. Robust empirical likelihood for $\beta_0$

In most applications of GPLMs, the primary research interest is to make statistical inferences on the regression coefficient $\beta_0$, along with the understanding of some basic feature of $f_0(t)$. For a partially linear model with independent data, Shi and Lau [22] conducted a EL inference on $\beta_0$ by taking nonparametric part as nuisance. Thus, we regard the nonparametric function $f_0(t)$, i.e. $\alpha_0$, as nuisance, and conduct a suitable estimator of it to make sure the efficient statistical inference on $\beta_0$. In this paper, we let $\hat{\alpha}$ denote the robust estimator of $\alpha_0$, which is proposed in [4]. Motivated by the idea of Cantoni [2] and He et al. [4], we adopt a weight function to downweight the effect of leverage points and a bounded score function on the Pearson residuals to limit the influence of outliers in the response. This is slightly different from the weights on the response in [2] where they are defined through dividing the Huber function by the Pearson residual; see formula (10) in [2] or formula (6.9) in [6]. Meanwhile, a working correlation matrix is incorporated into the following auxiliary random vectors as:

$$Z_i(\beta;\hat{\alpha}) = X_i^*(\mu_i(\beta;\hat{\alpha}))\Delta_i^T(\mu_i(\beta;\hat{\alpha}))V_i^{-1}(\mu_i(\beta;\hat{\alpha}),\gamma)h_i(\mu_i(\beta;\hat{\alpha})), \quad (2.3)$$

where

$$(X_1^{*T}(\mu_1(\beta;\hat{\alpha})),\ldots,X_m^{*T}(\mu_m(\beta;\hat{\alpha})))^T = (I-P(\mu(\beta;\hat{\alpha})))(X_1^T,\ldots,X_m^T)^T$$

with

$$P(\mu(\beta;\hat{\alpha})) = M(M^T\Sigma(\mu(\beta;\hat{\alpha}))M)^{-1}M^T\Sigma(\mu(\beta;\hat{\alpha})),$$
$$M = (\pi_1^T,\ldots,\pi_m^T)^T,$$
$$\Sigma(\mu(\beta;\hat{\alpha})) = \mathrm{diag}\{\Sigma_1,\ldots,\Sigma_m\},$$
$$\Sigma_i = \Delta_i^T(\mu_i(\beta;\hat{\alpha}))V_i^{-1}(\mu_i(\beta;\hat{\alpha}),\gamma)\Gamma_i(\mu_i(\beta;\hat{\alpha}))\Delta_i(\mu_i(\beta;\hat{\alpha})),$$
$$\Gamma_i(\mu_i(\beta;\hat{\alpha})) = E\dot{h}_i(\mu_i(\beta;\hat{\alpha})) = E\frac{\partial h_i(\mu_i(\beta;\hat{\alpha}))}{\partial\mu_i},$$
$$\Delta_i(\mu_i(\beta;\hat{\alpha})) = \mathrm{diag}\{\dot{\mu}_{i1}(\beta;\hat{\alpha}),\ldots,\dot{\mu}_{in_i}(\beta;\hat{\alpha})\}$$

with $\dot{\mu}(\cdot)$ denoting the first derivative of $\mu(\cdot)$ evaluated at $X_i\beta+\pi_i\hat{\alpha}$,

$$V_i(\mu_i(\beta;\hat{\alpha}),\gamma) = R_i(\gamma)A_i^{1/2}(\mu_i(\beta;\hat{\alpha}))$$

with $A_i(\mu_i(\beta;\hat{\alpha})) = \phi\cdot\mathrm{diag}\{v(\mu_{i1}(\beta;\hat{\alpha})),\ldots,v(\mu_{in_i}(\beta;\hat{\alpha}))\}$, $R_i(\gamma)$ is a working correlation matrix, and

$$h_i(\mu_i(\beta;\hat{\alpha})) = W_i(\psi(\mu_i(\beta;\hat{\alpha})) - E\psi(\mu_i(\beta;\hat{\alpha})))$$

with $\psi(\mu_i) = \psi(A_i^{-1/2}(Y_i-\mu_i))$ where function $\psi$ is considered to be Huber's psi function $\psi(x) = \min\{c,\max\{-c,x\}\}$, the tuning constant $c$ is typically chosen to give a certain level of asymptotic efficiency at the underlying distribution and selected to be 1.5 in this article, and the weighting matrix $W_i = \mathrm{diag}\{w_{i1},\ldots,w_{in_i}\}$ is a diagonal matrix. Similar to Sinha [23], we choose the weight function $w_{ij}$ as a function of the Mahalanobis distance in the form

$$w_{ij} = w(x_{ij}) = \min\left[1,\left\{\frac{b_0}{(x_{ij}-m_x)^TS_x^{-1}(x_{ij}-m_x)}\right\}^{\gamma_w/2}\right],$$

with $\gamma_w\ge 1$; $b_0$ 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 Zomerem [19].

Before we move to construct the EL procedure for $\beta$, we would like to highlight the goodness of our proposed auxiliary random vector (2.3). We approximate the nonparametric part through $B$-spline and absorb it in the auxiliary random vector by projection to help improve the inferences on $\beta$. The auxiliary random vector discriminates from the one in [26] by considering working correlation to improve the efficiency.

Using the standard empirical likelihood procedure with the proposed auxiliary random vectors $Z_i(\beta;\hat{\alpha})$, we can get the REL function of $\beta$ as

$$L(\beta) = \sup\left\{\prod_{i=1}^m p_i\,\middle|\,\sum_{i=1}^m p_i = 1, p_i\ge 0, \sum_{i=1}^m p_iZ_i(\beta;\hat{\alpha}) = 0\right\}.$$

Note that $\prod_{i=1}^m p_i$ attains its maximum at all $p_i = 1/m$. Thus, the REL ratio is defined as

$$R(\beta,\hat{\alpha}) = \sup\left\{\prod_{i=1}^m (mp_i)\,\middle|\,\sum_{i=1}^m p_i = 1, p_i\ge 0, \sum_{i=1}^m p_iZ_i(\beta;\hat{\alpha}) = 0\right\}.$$

By using the Lagrange multiplier method, we obtain that $R(\beta)$ is maximized at

$$p_i = \frac{1}{m}(1+\lambda'Z_i(\beta;\hat{\alpha}))^{-1}, \quad i = 1,\ldots,m,$$

where the vector $\lambda = (\lambda_1, \ldots, \lambda_p)^T$ satisfies the following equation

$$q(\lambda) = \frac{1}{m}\sum_{i=1}^{m} \frac{Z_i(\beta; \hat{\alpha})}{1 + \lambda^T Z_i(\beta; \hat{\alpha})} = 0. \quad (2.4)$$

Combining the above equations, we have

$$-2\log R(\beta; \hat{\alpha}) = -2\log \prod_{i=1}^{m}(1 + \lambda^T Z_i(\beta; \hat{\alpha}))^{-1} = 2\sum_{i=1}^{m} \log(1 + \lambda^T Z_i(\beta; \hat{\alpha})). \quad (2.5)$$

Note that for the choice of $\psi(x) = x$ and $w_{ij} = 1$, we have $E\psi(\mu_i) = 0$; then, the proposed robust auxiliary random vectors reduce to ordinary auxiliary random vector and the REL ratio reduces to the ordinary EL ratio.

As discussed in [1], the correlation parameter $\gamma$ and the scale parameter $\phi$ involved in (2.5) can be replaced by their corresponding $\sqrt{n}$-consistent estimators. Here, we replace them with some robust estimates $\hat{\gamma}$ and $\hat{\phi}$, such as the robust estimating equation estimators proposed by Cantoni [2] and He et al. [4]. Then, we can obtain the robust maximum empirical likelihood estimator (RMELE) of the parameter $\beta_0$, $\tilde{\beta}$, by minimizing the $-2\log R(\beta; \hat{\alpha})$ under equation constraints (2.4).

Finally, we make a summary on the process of deriving the RMELE $\tilde{\beta}$ as follows to finish this section.

1. Get the robust estimator $\hat{\alpha}$ of $\alpha$ by adopting the RGEE method of He et al. [4].
2. Construct the robust auxiliary score (2.3) for $\beta$.
3. Substitute the $\sqrt{n}$-consistent estimators $\hat{\gamma}$ and $\hat{\phi}$ for nuisance parameters $\gamma$ and $\phi$ in (2.3).
4. Obtain the RMELE $\tilde{\beta}$ by minimizing the negative log empirical likelihood ratio (2.5).

## 3. Asymptotic properties

We use $-2\log R(\beta; \hat{\alpha}, \hat{\gamma}, \hat{\phi})$ to denote $-2\log R(\beta; \hat{\alpha})$ with $\gamma$ and $\phi$ replaced by $\hat{\gamma}$ and $\hat{\phi}$, and use a similar way to define $Z_i(\beta; \hat{\alpha}, \hat{\gamma}, \hat{\phi})$. We use $\|\cdot\|$ to denote Euclidean norm. Moreover, let $e_i = (\phi v(\mu_{0,i}))^{-1/2}(Y_i - \mu_{0,i})$ be the vector of standardized responses, and $h_{0,i}(e_i) = W_i(\psi(e_i) - E\psi(e_i))$. Note that $h_{0,i}(e_i)$ is similar to $h_i(\mu_i(\beta; \hat{\alpha}))$, but the former centers $Y_i$ by its true mean $\mu_{0,i}$, whereas the latter involves centering by $\mu_i(\beta; \hat{\alpha})$.

To derive the asymptotic normality of the RMELE $\tilde{\beta}$ and the asymptotic chi-square distribution of the proposed REL ratio $-2\log_R(\beta_0; \hat{\alpha}, \hat{\gamma}, \hat{\phi})$, some regularity conditions similar to He et al. [4] are assumed as follows.

(A.1) The sequence $\{n_i\}$ is a bounded sequence of positive integers, and the distinct values of $t_{ij}$ form a quasi-uniform sequence that grows dense on the internal [0, 1].
(A.2) The $r$th derivative of $f_0$ is bounded for any $r \geq 2$.
(A.3) The regression parameter $\beta_0$ is identifiable, i.e., there is a unique $\beta_0 \in B$ satisfying the mean model assumption which guarantees $E(Z_i(\mu_{0,i})) = 0$, where $B$ is a compact parameter space.
(A.4) There exists positive constant $C_1$ such that $0 < C_1 \leq v(\cdot) < \infty$, $v(\cdot)$ has bounded second derivatives and $g^{-1}(\cdot)$ has bounded third derivatives,
$\sup_{i\geq 1} E\|X_i\|^3 < \infty$ and $Z_i(\beta; \hat{\alpha}, \hat{\gamma}, \hat{\phi})$ is assumed to be continuously differentiable.

To obtain the asymptotic properties of the RMELE $\tilde{\beta}$, some assumptions on the covariates $X$ and $t$ are required. One complicated issue for the GPLMs comes from the dependence between $X$ and $t$. To this end, we denote $x_{ij} = (x_{ij1}, \ldots, x_{ijn_i})^T$ and assume the following relationship as Rice [18]:

$$x_{ijk} = g_k(t_{ij}) + \delta_{ijk}, \quad 1 \leq i \leq m,\ 1 \leq j \leq n_i,\ 1 \leq k \leq p, \quad (3.1)$$

where $g_k(t)$ are functions with bounded $r$th derivatives, and $\delta_{ijk}$ are mean zero random variables independent of $e_i$ and of one another. This is a common used and sufficient condition which had been also used in [5,24]. We also assume the following.

(A.5) For sufficiently large $n$, $k_n(M^T\Sigma_0 M)$ is non-singular, and the eigenvalues of $(k_n/n)M^T\Sigma_0 M$ are bounded away from zero and infinity, where $\Sigma_0 = \mathrm{diag}\{\Sigma_{0,i}\}$, $\Sigma_{0,i} = \Delta_{0,i}V_{0,i}^{-1}\Gamma_{0,i}\Delta_{0,i}$, $\Delta_{0,i}$, $V_{0,i}^{-1}$ and $\Gamma_{0,i}$ denote $\Delta_i(\mu_i)$, $V_i^{-1}(\mu_i)$ and $\Gamma_i(\mu_i)$ evaluated at $\mu_{0,i}$, respectively.
(A.6) (a) $E\Lambda_n = 0$ and $\sup_{n\geq 1}\|\Lambda_n\|^2 \leq \infty$, and
(b) $\frac{1}{n}K_n \to K$, $\frac{1}{n}S_n \to S$ in probability for some positive definite matrix $K$ and $S$, where $\Lambda_n$ is $n$ by $p$ matrix, whose $s$th column is $\delta_s = (\delta_{11s}, \ldots, \delta_{mn_m})^T$, $S_n = \sum_{i=1}^m X_{0,i}^{*T}\Delta_{0,i}V_{0,i}^{-1}\mathrm{cov}(h_{0,i}(e_i))V_{0,i}\Delta_{0,i}X_{0,i}^*$, $K_n = \sum_{i=1}^m X_{0,i}^{*T}\Sigma_{0,i}X_{0,i}^*$, and $X_{0,i}^*$ denote $X_i^*(\mu_i(\beta; \hat{\alpha}))$ evaluated at $\mu_{0,i}$.
(A.7) We assume that some estimated correlation parameter vector $\tilde{\gamma}$ is consistent for some $\gamma_0$; that is $\|\tilde{\gamma} - \gamma_0\| = O_p(n^{-1/2})$ for some $\gamma_0$.

A similar condition to (A.4) that $Z_i(\beta; \hat{\alpha}, \hat{\gamma}, \hat{\phi})$ is assumed to be continuously differentiable is also used by Sinha [23]. As indicated by Sinha [23], the derivative of Huber's function used here is not continuous at the points $\pm c$, and as a remedy, it is possible to smooth the psi function that leads to a continuous derivative. However, there would not be any dramatic change to the estimate for using such a smoothed psi function.

**Theorem 1.** *Suppose that conditions* (A.1)–(A.7) *hold. If the number of knots* $k_n \approx n^{1/(2r+1)}$, *then* $\sqrt{n}(\tilde{\beta} - \beta) \to N(0, V)$, *where* $V = K^{-1}SK^{-1}$.

In the following numerical studies, we use the method proposed by He et al. [4] to obtain consistent estimates of $K$ and $S$.

**Theorem 2.** *Under the conditions of Theorem 1,* $-2\log R(\beta_0; \hat{\alpha}, \hat{\gamma}, \hat{\phi})$ *converges to* $\chi_p^2$ *in distribution when* $m \to \infty$, *where* $\chi_p^2$ *is a chi-squared distribution with the degrees of freedom p.*

Now an asymptotic empirical likelihood-based $100(1-\tau)\%$ confidence region for $\beta_0$ can be given by

$$\Upsilon = \{\beta : -2\log R(\beta; \hat{\alpha}, \hat{\gamma}, \hat{\phi}) \le \chi_p^2(\tau)\}, \tag{3.2}$$

where $\chi_p^2(\tau)$ is the upper $\tau$-quantile of the chi-squared distribution with the degrees of freedom $p$.

**Remark 2.** As pointed out in [1], the point estimators of EL and GEE methods are asymptotically equivalent according to the discussion in [16]. However, we still prefer the proposed REL method which is more accurate and efficient in the case of statistical inferences, since it avoids to estimate the asymptotic variances of estimators which make the inferences unstable in most cases, especially in the case of outliers. These is also demonstrated numerically by the simulations in following Section 4.

## 4. Simulation studies

In this section, some simulations are conducted to investigate the performance of our proposed REL method. The commonly used normal, binary and Poisson partial linear models are considered. We compare the proposed REL method with the RGEE method provided by He et al. [4] as well as the non-robust EL and GEE methods in terms of robustness, accuracy and efficiency.

We adopt the robust estimate $\hat{\gamma}$ and $\hat{\phi}$ proposed in [4] to estimate the nuisance parameters $\gamma$ and $\phi$, and denote $\hat{\beta}$ as the RGEE estimator proposed by He et al. [4]. The robust methods, including the EL and GEE methods, reduce to the corresponding non-robust ones by setting $w_{ij} = 1$, $\psi(x) = x$. In our simulations, the $\gamma_w$ in the weight function $w_{ij}$ was chosen to be 1 and the constant $c$ in the Huber function was chosen to be 1.5. A total of 1000 replicates were generated from each of the considered models.

*Study* 1: First, we consider a multiple linear model

$$y_{ij} = x_{1,ij}\beta_1 + x_{2,ij}\beta_2 + x_{3,ij}\beta_3 + x_{4,i}\beta_4 + \sin(2t_{ij}) + \epsilon_{ij}, \quad i = 1, \ldots, m,\ j = 1, 2, 3, \tag{4.1}$$

where $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 1$, and the covariates are generated as follows: $x_{4,i}$ is a subject-level covariate taking values 0 or 1 with equal probability, and $x_{1,ij} = b_i + u_{1,ij}$, $x_{2,ij} = z_i + u_{2,ij}$, $x_{3,ij} = td_i + u_{3,ij}$ and $t_{ij} = b_i + u_{4,ij}$ with $b_i, z_i, td_i$ and all $u_{k,ij}$, $k = 1, \ldots, 4$ as independent draws from the uniform distribution on $[-1, 1]$, standard norm distribution, $t$ distribution with 5 degrees of freedom and the uniform distribution on $[-1, 1]$, respectively. $(\epsilon_{i1}, \epsilon_{i2}, \epsilon_{i3})^T$ are multivariate normal with mean 0 and variance 1, and exchangeable correlation matrix $R_i(\gamma_0) = 1_i 1_i^T \gamma_0 + I_i(1-\gamma_0)$, where $1_i$ denotes a $n_i$-dimensional vector with 1 as its components, $I_i$ denotes a $n_i$-dimensional identity matrix and the correlation parameter $\gamma_0$ is taken to be 0.5. To study robustness, we contaminated the data by the following two ways.

C1. Creating outliers in the covariates by adding 2 to $x_{1,ij}$, $x_{2,ij}$ and $x_{3,ij}$ for any randomly chosen observations with 2% probability. We do not contaminate the subject-level covariate $x_4$.

C2. Besides perturbing the covariates by C1, we also contaminate the response by replacing each $y_{ij}$ values with $y_{ij} + 5$ for other randomly chosen observations with 2% probability.

**Remark 3.** In practice, our proposed method actually can handle larger level of contaminations. Here, we just present the results under 2% contamination because most of the time, those non-robust methods under more than 2% contamination do not converge and are not comparable to our proposed method.

To compare the accuracy of statistical inference based on our proposed REL method and the RGEE method, 95% confidence regions for $\beta$ are computed based on both the EL and GEE methods. Table 1 reports the coverage probabilities (CPs) of the two types of confidence regions for sample sizes $m = 50$, 100, respectively. The entries demonstrate that when the sample size increase from $m = 50$ to 100, the accuracies of both the EL and GEE methods are satisfactory. Table 2 reports the empirical biases and mean squares errors (MSEs) of the proposed REL estimators of parameters as well as the corresponding non-robust estimators under $m = 100$. In the case of no outliers, the proposed REL method performs almost equally well with other non-robust or robust methods, since all the CPs are very close to the nominal confidence level 95%. However, in the case of contaminations, both the two non-robust EL and GEE methods are seriously influenced by the outliers (their CPs are far away from 95% and the estimating biases are very large), while the robust ones look still acceptable in terms of both CPs and estimating accuracies. What is worth mentioning, when the data are perturbed, the CPs improve a lot from the EL method to the proposed REL method, but it is not the case of GEE methods. We believe that this is because the sandwich

**Table 1**
CPs of confidence regions in Study 1.

| | | $m=50$ | | | |
|---|---|---|---|---|---|
| | | $EL_{Corr}$ | $GEE_{Corr}$ | $EL_{Ind}$ | $GEE_{Ind}$ |
| NC | NR | 0.886 | 0.870 | 0.886 | 0.861 |
| | R | 0.894 | 0.888 | 0.875 | 0.872 |
| C1 | NR | 0.378 | 0.506 | 0.478 | 0.592 |
| | R | 0.748 | 0.674 | 0.814 | 0.740 |
| C2 | NR | 0.454 | 0.556 | 0.498 | 0.592 |
| | R | 0.776 | 0.686 | 0.814 | 0.776 |
| | | $m=100$ | | | |
| NC | NR | 0.964 | 0.952 | 0.934 | 0.926 |
| | R | 0.946 | 0.952 | 0.944 | 0.912 |
| C1 | NR | 0.610 | 0.814 | 0.746 | 0.860 |
| | R | 0.934 | 0.902 | 0.942 | 0.918 |
| C2 | NR | 0.626 | 0.794 | 0.742 | 0.836 |
| | R | 0.926 | 0.892 | 0.918 | 0.902 |

NOTE: NC = No contamination; C1 or C2 = contamination 1 or 2.
NR or R = Non-robust method or Robust method.
GEE and EL denote used different estimating method.
Corr or Ind denote considering correlation or not.

**Table 2**
Empirical biases (MSEs) of $\tilde{\beta}$ in Study 1.

| | | | $m=100$ | | | |
|---|---|---|---|---|---|---|
| | | | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ |
| NC | NR | $EL_{Corr}$ | −0.0064 (0.0057) | −0.0009 (0.0035) | 0.0027 (0.0026) | −0.0020 (0.0091) |
| | | $EL_{Ind}$ | −0.0097 (0.0089) | −0.0022 (0.0043) | 0.0065 (0.0035) | −0.0059 (0.0128) |
| | R | $EL_{Corr}$ | −0.0080 (0.0061) | −0.0009 (0.0037) | 0.0036 (0.0028) | −0.0028 (0.0095) |
| | | $EL_{Ind}$ | −0.0099 (0.0092) | −0.0018 (0.0045) | 0.0062 (0.0035) | −0.0059 (0.013) |
| C1 | NR | $EL_{Corr}$ | −0.2220 (0.0672) | −0.1313 (0.0256) | −0.0966 (0.0150) | 0.0022 (0.0126) |
| | | $EL_{Ind}$ | −0.2061 (0.0616) | −0.0804 (0.0133) | −0.0520 (0.0071) | 0.0054 (0.0156) |
| | R | $EL_{Corr}$ | −0.0643 (0.0117) | −0.0393 (0.0060) | −0.0271 (0.0038) | 0.0043 (0.0092) |
| | | $EL_{Ind}$ | −0.0573 (0.0127) | −0.0257 (0.0057) | −0.0139 (0.0035) | 0.0046 (0.0125) |
| C2 | NR | $EL_{Corr}$ | −0.2202 (0.0696) | −0.1282 (0.0253) | −0.951 (0.0145) | 0.0074 (0.0139) |
| | | $EL_{Ind}$ | −0.2006 (0.0607) | −0.0783 (0.0133) | 0.0516 (0.0071) | 0.0108 (0.0173) |
| | R | $EL_{Corr}$ | −0.0638 (0.0131) | −0.0356 (0.0059) | −0.0264 (0.0040) | 0.0072 (0.0108) |
| | | $EL_{Ind}$ | −0.0550 (0.0131) | −0.0224 (0.0055) | −0.0136 (0.0038) | 0.0087 (0.0142) |

NOTE: NC = No contamination; C1 or C2 = contamination 1 or 2.
NR or R = Non-robust method or Robust method.
Corr or Ind denote considering correlation or not.

estimation of the variance of the GEE estimator is badly interfered by the perturbed data. For this moment, we think that the proposed REL method are more robust than the RGEE method. We will explore this finding more based on other aspects in the two subsequent discrete studies.

Before we move on, we would like to mention some simulation results that we do not present here for reasons of brevity. We try different correlation structures in the very beginning of the simulation setting and find that the similar results can be obtained. Moreover, we do not present the empirical biases and mean squares errors (MSEs) of the GEE or RGEE estimators, since they are asymptotic equivalent to the corresponding EL estimators ([16], see). For all these reasons, we will just present the results under the same simulation structure with $m = 100$ as above in the following two simulation studies.

*Study* 2. We considered a logistic partial linear model with $m$ subjects and 3 observations within each subject:

$$y_{ij} \sim \text{Binomial}(1, \mu_{ij}), \quad i = 1, \ldots, m,\ j = 1, \ldots, 3,$$

and $\mu_{ij}$ satisfies that

$$\text{logit}(\mu_{ij}) = x_{ij}\beta + \sin(2t_{ij}), \tag{4.2}$$

where $\beta = 1$, and $x_{ij}$ and $t_{ij}$ are drawn independently from the uniform distribution on $[-1, 1]$. Here, we just consider one single covariate, since it is easy to present the mean lengths (MLs) of confidence intervals as well as CPs. We use the Madsen and Dalthorp [9] MATLAB package at http://www.stat.oregonstate.edu/people/lmadsen to generate clustered correlated binary responses with exchangeable correlation structure $R_i(\gamma_0) = 1_i 1_i^T \gamma_0 + I_i(1 - \gamma_0)$ for each subject, and the correlation parameter $\gamma_0$ is also taken to be 0.5. Since that the response is binary, we just consider to perturb the covariate part as follow.

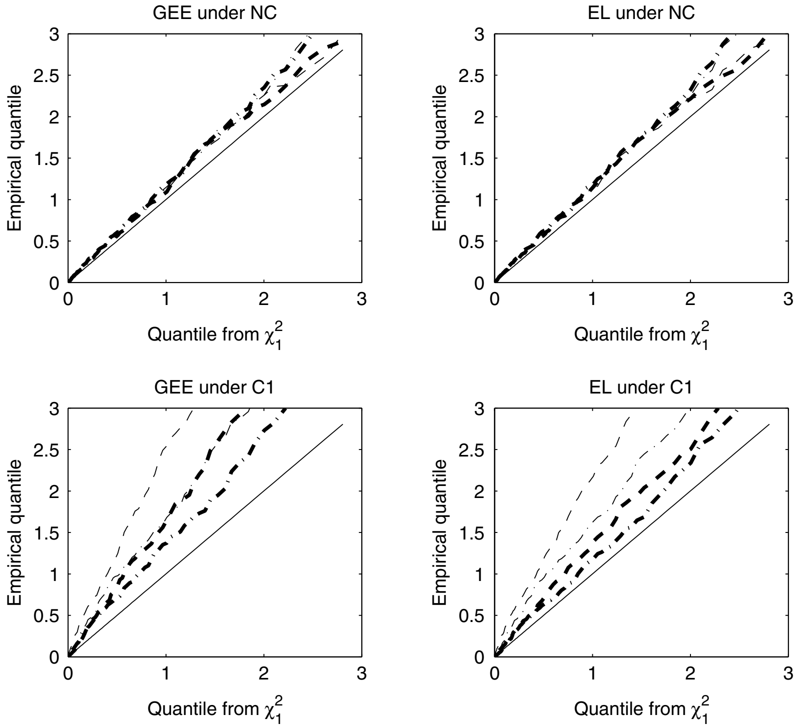


**Fig. 1.** Q–Q plots in Study 2 (left panel is based on the GEE method and right panel is based on the EL method; upper panel is under no contaminations and lower panel is under contamination 1. In every figure, dash lines are related to considering correlation while dot dash lines are related to not considering correlation; thick lines are related to the robust method while thin lines are related to the non-robust method).

**Table 3**
CPs(MLs) of confidence intervals in Study 2.

| | | $m = 100$ | | | |
|---|---|---|---|---|---|
| | | $EL_{Corr}$ | $GEE_{Corr}$ | $EL_{Ind}$ | $GEE_{Ind}$ |
| NC | NR | 0.940 (0.647) | 0.930 (0.824) | 0.932 (0.735) | 0.914 (1.026) |
| | R | 0.936 (0.648) | 0.931 (0.826) | 0.924 (0.732) | 0.916 (1.029) |
| C1 | NR | 0.884 (0.694) | 0.798 (0.816) | 0.832 (0.760) | 0.876 (0.984) |
| | R | 0.910 (0.721) | 0.874 (0.800) | 0.926 (0.780) | 0.902 (0.981) |

NOTE: NC = No contamination; C1 = contamination 1.
NR or R = Non-robust method or Robust method.
GEE and EL denote used different estimating methods.
Corr or Ind denote considering correlation or not.

C1. Creating outliers in the covariate $x_{ij}$ by replacing each $x_{ij}$ values by $x_{ij} + 2$ for any randomly chosen observations with 2% probability.

Table 3 reports CPs(MLs) of the corresponding confidence intervals at nominal confidence level of 95% based on different methods in study 2. Similar to the finding in study 1, CPs of both EL and GEE methods have been close to the nominal level under $m = 100$. Moreover, although the choice of the working correlation does not affect CPs, choosing the correct correlation does improve the efficiency of statistical inference a lot (shorter MLs). Which matches the findings in [1]. We also find that the proposed REL method are more efficient and accurate than the RGEE method in terms of closer CPs and shorter MLs. The values of empirical biases (MSEs) of the RMELE of $\beta$ in Table 4 also tell us the same story as Table 2. Fig. 1 gives the chi-square with 1 degree of freedom quantile–quantile (Q–Q) plots for the REL ratio statistics, $-2\log R(\beta_0; \hat{\alpha}; \hat{\gamma})$, and the Wald statistics based on the normal approximation $T_{NC} = (\hat{\beta} - \beta_0)^T \hat{V}^{-1}(\hat{\beta} - \beta_0)$. In the case of no outliers, the Q–Q plots obtained by all the methods are close to the 45° line, which means that the distributions of the EL ratio statistics and the statistics based on normal approximation are close to the chi-square distribution with 1 degree of freedom, and show the validation of the asymptotic results of these statistics. Meanwhile, in the case of contaminations, the Q–Q plots by the REL method are closer to the 45° line, which shows that the outliers have smaller impact on the asymptotic distribution of the proposed REL ratio statistic. This also shows that the proposed REL method can make more accurate inference than the RGEE method.

**Table 4**
Empirical biases (MSEs) of $\tilde{\beta}$ in Study 2.

| | $EL_{Corr}$ | | $EL_{Ind}$ | |
|---|---|---|---|---|
| | NR | R | NR | R |
| NC | 0.0672 (0.0374) | 0.0673 (0.0377) | 0.0755 (0.0649) | 0.0790 (0.0673) |
| C1 | −0.1740 (0.0775) | −0.0681 (0.0562) | −0.1736 (0.0864) | −0.0548 (0.0855) |

NOTE: NC = No contamination; C1 = contamination 1.
NR or R = Non-robust method or Robust method.
Corr or Ind denote considering correlation or not.

**Table 5**
CPs in Study 3.

| | | $m = 100$ | | | |
|---|---|---|---|---|---|
| | | $EL_{Corr}$ | $GEE_{Corr}$ | $EL_{Ind}$ | $GEE_{Ind}$ |
| NC | NR | 0.926 | 0.907 | 0.931 | 0.925 |
| | R | 0.934 | 0.927 | 0.933 | 0.945 |
| C1 | NR | 0.872 | 0.937 | 0.686 | 0.885 |
| | R | 0.940 | 0.951 | 0.961 | 0.935 |
| C2 | NR | 0.876 | 0.961 | 0.846 | 0.916 |
| | R | 0.951 | 0.943 | 0.963 | 0.927 |

NOTE: NC = No contamination; C1 or C2 = contamination 1 or 2.
NR or R = Non-robust method or Robust method.
GEE and EL denote used different estimating methods.
Corr or Ind denote considering correlation or not.

*Study* 3: Finally, we considered a Poisson partial linear model with $m$ subjects and 3 observations within each subject:

$$y_{ij} \sim \text{Poisson}(\mu_{ij}), \quad i = 1, \ldots, m,\ j = 1, \ldots, 3,$$
$$\eta_{ij} = \log(\mu_{ij}) = x_{1,ij}\beta_1 + x_{2,ij}\beta_2 + \sin(3t_{ij}), \tag{4.3}$$

where $\beta_1 = \beta_2 = 1$, and the covariates are generated as follows: $x_{1,ij}$ is generated from a standard normal distribution, and $x_{2,ij} = b_i + u_{1,ij}$ and $t_{ij} = b_i + u_{2,ij}$ with $b_i$, $u_{1,ij}$ and $u_{2,ij}$ as independent draws from the uniform distribution on $[-1, 1]$. We use the same MATLAB package as above used to generate clustered correlated counting responses with exchangeable correlation structure $R_i(\gamma_0) = 1_i 1_i^T \gamma_0 + I_i(1 - \gamma_0)$ for each subject, and the correlation parameter $\gamma_0$ is also taken to be 0.5. We adopted a little different method from above studies to create outliers. We let all the observations in the same subject are perturbed at the same time which is common in longitudinal studies and details are as follows.

C1. Creating outliers in the covariate $x_{1,ij}$ by replacing each $x_{1,ij}$ values by $x_{1,ij} - 5$ for two randomly chosen subject from $m = 100$ subject.
C2. Creating outliers both in the covariate and in the response by replacing each $x_{1,ij}$ values with $x_{1,ij} - 5$ for one randomly chosen subject from $m = 100$ subject, and replacing each $y_{ij}$ values with $y_{ij} + 10$ for another randomly chosen subject from $m = 100$ subject.

Similar to the above studies, Table 5 reports the CPs of the corresponding confidence regions at nominal confidence level of 95% under $m = 100$ in Study 3. The entries show that CPs based on REL and RGEE are comparable. Fig. 2 represents the confidence regions based on the non-robust EL (or REL) and the non-robust GEE (or RGEE) methods when considering the true correlation structure. In the case of no outliers, the left panel shows that confidence regions based on EL and REL are comparable and the areas of the corresponding confidence regions are smaller than those based on GEE or RGEE method. But in the case of contaminations, the right panel shows that the center of the confidence regions obtained by the non-robust methods obviously deviate from the true values (1, 1) and the contaminations appear to have much smaller influence on the confidence region obtained by the robust methods, especially the proposed REL method. Fig. 3 also represents the similar confidence regions in Study 3, but here we compare the accuracy of the confidence regions under different working correlations. Generally, one who consider the true exchangeable correlation have smaller areas. The findings from the Q–Q plots and the empirical biases (MSEs) are similar to the case in Study 2 and omitted here for saving space.

We end this section by making a summary as follows.

1. In the case of no outliers, both the non-robust and robust methods perform almost equally well. However, in the cases of contaminations, statistical inferences based on the non-robust EL methods are seriously biased (bad CPs) while the CPs of the proposed REL method still keep close to the nominal 95% level, which shows the robustness of the proposed REL method. Furthermore, compared with the RGEE method, the CPs obtained by the REL method are stable and much closer to the nominal confidence level 95%.
2. Although the choice of the working correlation does not impact the CPs [1], the MLs or confidence regions obtained by the REL method considering the correlation structure are much shorter or smaller than those considering working

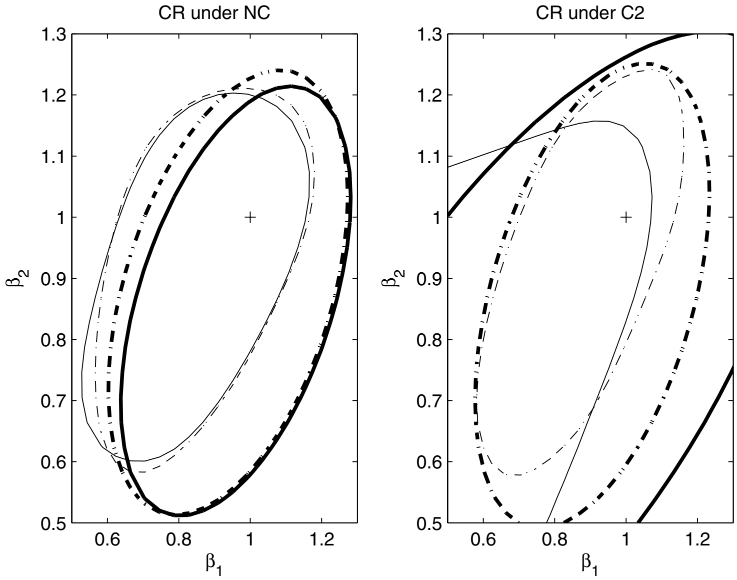


**Fig. 2.** Confidence regions based on both GEE and EL methods with considering correlation in Study 3 (left panel includes 4 ellipses without outliers (NC) while right panel includes 4 ellipses under contamination (C2): thin solid lines for the EL method, thin dot dash lines for the REL method, thick solid line for the GEE method and thick dot dash line for the RGEE method).

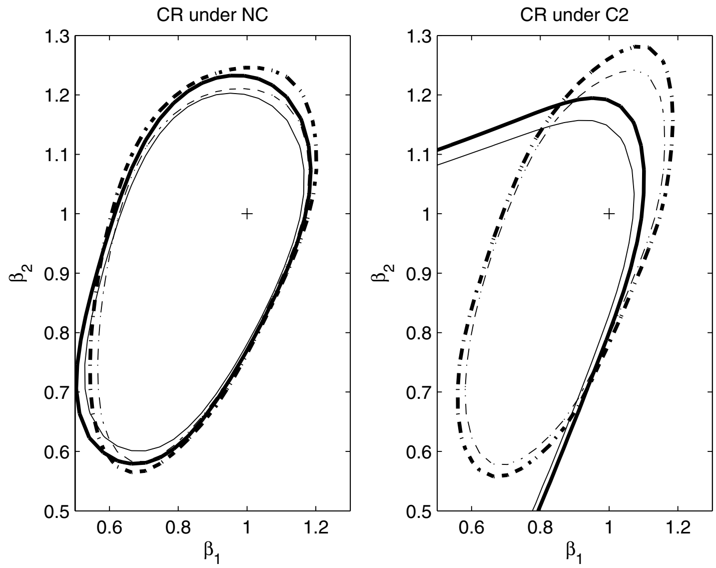


**Fig. 3.** Confidence regions based on both non-robust EL and proposed REL methods in Study 3 (left panel includes 4 ellipses without outliers (NC) while right panel includes 4 ellipses under contamination (C2): solid line for the EL method, dot dash line for the REL method and thick lines for those under working independent correlation while thin lines for those under working exchangeable correlation).

independence structure, which shows that the REL method considering the correlation structure is more efficient than that ignoring the correlation.

3. The Q–Q plots tell us that the likelihood ratio test based on the proposed REL method is more accurate than the Wald test based on the normal approximation, especially in the case of contaminations.

## 5. Application to GUIDE data of Preisser and Qaqish [14]

We apply the proposed method to analyze the data collected from a study on Guidelines for Urinary Incontinence Discussion and Evaluation. In the study, there are a total of 137 patients age 76 or older, who had experienced accidental loss of urine and had been using some of 38 medical practices, were asked whether they were bothered by the problem. The outcome $y_{ij}$ is the coded answer, 1 for "bothered" and 0 otherwise, where the subject $i$ refers to the patient and $j$ refers to

**Table 6**
Regression coefficients estimates in the analysis of the GUIDE data.

| | Semiparametric model | | | Parametric model |
|---|---|---|---|---|
| | $\text{RMELE}_{\text{Ind}}$ | $\text{RMELE}_{\text{Corr}}$ | He et al. | Sinha |
| Intercept | – | – | – | −3.593 (0.952) |
| Age | – | – | – | −1.298 (0.632) |
| GENDER | −1.63 (0.612) | −1.59 (0.603) | −1.57 (0.61) | −1.297 (0.632) |
| DAYACC | 0.50 (0.112) | 0.53 (0.131) | 0.59 (0.14) | 0.506 (0.116) |
| SEVERE | 0.72 (0.410) | 0.70 (0.345) | 0.67 (0.40) | 0.827 (0.373) |
| TOILET | 0.31 (0.092) | 0.33 (0.125) | 0.27 (0.10) | 0.240 (0.110) |

NOTE: the entries in the parenthesis are their standard errors.

the $j$th medical practice. The data has been well analyzed by many authors. Preisser and Qaqish [14] fitted the data through a GLM and provided robust GEE to estimate the regression coefficients. More recently, Sinha [23] analyzed the data by a generalized linear mixed model and He et al. [4] conducted a GPLM to analyze the GUIDE data by allowing one of the covariates to enter the model as a nonlinear curve. Here, we use the same GPLM as He et al. did:

$$\text{Log}\frac{E[y_{ij}]}{1-E[y_{ij}]} = \text{GENDER}_i\beta_1 + \text{DAYACC}_{ij}\beta_2 + \text{SEVERE}_i\beta_3 + \text{TOILET}_{ij}\beta_4 + f(\text{AGE}_{ij}),$$

where the five covariates are explained as follows: GENDER (1 for female, 0 otherwise), the number of leaking accidents per day (DAYCC), severity of leaking (SEVERE) on a scales of 1–4, the number of times during the day they usually go to the toilet to urinate (TOILET), and the standardized age (AGE, (age in year-76)/10).

As pointed by Sinha [23], the potential perturbed observations may include the patient 7, 10, 27, 56, 59, 97 and 131. Particularly, the patient 97 appears to be the most extreme one. We calculate the values of the weights $w_{ij}$ used in our proposed REL. The heavily downweighted points (with weights less than 0.10) include the patients 10, 45, 47, 56, 59, 97, 98 and 131. Specially, patients 97 reports SEVERE = 3, DAYACC = 16.7, and TOILET = 8 and appears to be the most extreme one with the smallest weight 0.005 which is consistent with that in [23]. So the REL method can downweight those subjects and more accurately reflect the relationship in the majority of patients.

Although there are very weak within-patient correlations (refer to [4]), we analyze the data under both working exchangeable and working independence. Table 6 gives our proposed RMELE estimators of the regression coefficients in comparison with the estimates of both He et al. [4] and Sinha [23]. Basically, we are quite in agreement with He et al. [4], except that the variable SEVERE could be significant with our proposed REL under considering the weak correlation. To make sure of this, we conduct the profile EL ratio test for the significance of the variable SEVERE and obtain the $P$-values 0.049 for working independent and 0.032 for working exchangeable. For this moment, we agree with Sinha [23] that we should pay attentions on the variable SEVERE in the GUIDE study.

## Acknowledgments

This work was partially supported by the Natural Science Foundation of China (11001162, 10801039, 10931002, 1091120386), Shanghai University of Finance and Economics through Project 211 Phase III and Shanghai Leading Academic Discipline Project, Project number: B803 and B118. The authors also would like to thank the Associate Editor and two referees for their insightful comments that improve this manuscript significantly.

## Appendix. Proofs of Theorems 1 and 2

In this section, we consider the case where the correlation parameter $\gamma$ is known. The same asymptotic results hold when $\gamma$ is estimated under condition (A.7), but we omit the details. Let

$$\xi(\beta,\alpha) = \begin{bmatrix}\xi_1\\ \xi_2\end{bmatrix} = \begin{bmatrix} K_n^{1/2}(\beta-\beta_0) \\ k_n^{-1/2}H_n(\alpha-\alpha_0) + k_n^{1/2}H_n^{-1}M^T\Sigma_0 X(\beta-\beta_0)\end{bmatrix},$$

where $H_n^2 = k_n M^T \Sigma_0 M$. We denote $\widetilde{X}_{0,i}^T = K_n^{-1/2}X_{0,i}^{*T}$, $\widetilde{M}_{0,i}^T = k_n^{1/2}H_n^{-1}\pi_i^T$, $R_{ni} = \pi_i\alpha_0 - f_0(t_i)$, where $f_0(t_i) = (f(t_{i1}), \ldots, f(t_{in_i}))^T$, and $\zeta_i = \widetilde{X}_{0,i}\xi_1 + \widetilde{M}_{0,i}\xi_2 + R_{ni}$, then $\eta_i(\theta) = D_i\theta = \eta_{0,i} + \zeta_i,\ i = 1, \ldots, m$, where $\eta_{0,i} = X_i\beta_0 + f_0(t_i)$.

In order to prove Theorems 1 and 2, we first introduce the following lemmas.

**Lemma 1.** *Under condition* (A.1) *and* (A.2), *there exist* $\alpha_0 \in R^{N_k}$ *depending on* $f_0$, *and a constant* $C_3$ *depending only on* $l$ *and* $C_0$ *such that*

$$\sup_{t\in[0,1]} |f_0(t) - \pi^T(t)\alpha_0| \le C_3 k_n^{-r}. \tag{A.1}$$

*The proof of this lemma follows easily from Theorem 12.7 in* [20].

**Lemma 2.** *Under the conditions* (A.1)–(A.6), *we have*

$$n^{-1/2}\sum_{i=1}^{m} Z_i(\beta_0;\hat{\alpha}) \to N(0,S). \tag{A.2}$$

**Proof.** We multiply $n^{1/2}K_n^{-1/2}$ to $n^{-1/2}\sum_{i=1}^{m} Z_i(\beta_0;\hat{\alpha})$. If we show $K_n^{-1/2}\sum_{i=1}^{m} Z_i(\beta_0;\hat{\alpha}) \to_D N(0, K^{-1/2}SK^{-1/2})$, then Lemma 2 holds since $n^{1/2}K_n^{-1/2}$ is a consistent estimate of $K^{-1/2}$. Next, we shall show $K_n^{-1/2}\sum_{i=1}^{m} Z_i(\beta_0;\hat{\alpha}) \to_D N(0, K^{-1/2}S K^{-1/2})$.

We have

$$K_n^{-1/2}\sum_{i=1}^{m} Z_i(\beta_0;\hat{\alpha}) = \sum_{i=1}^{m}\tilde{X}_i(\beta_0;\hat{\alpha})\Delta_i(\beta_0;\hat{\alpha})^T V_i^{-1}(\beta_0;\hat{\alpha})h_i(\beta_0;\hat{\alpha}), \tag{A.3}$$

where $\tilde{X}_i(\beta_0;\hat{\alpha}) = K_n^{-1/2}X_i^*(\beta_0;\hat{\alpha})$. Take $a$ to be any $p$-dimensional vector which satisfies that $\|a\| = 1$. Then we can expand $a^T\sum_{i=1}^{m}\tilde{X}_i(\beta_0;\hat{\alpha})\Delta_i(\beta_0;\hat{\alpha})^T V_i^{-1}(\beta_0;\hat{\alpha})h_i(\beta_0;\hat{\alpha})$ into a Taylor series, and get

$$\begin{aligned} a^T\sum_{i=1}^{m}\tilde{X}_i(\beta_0;\hat{\alpha})\Delta_i(\beta_0;\hat{\alpha})^T V_i^{-1}(\beta_0;\hat{\alpha})h_i(\beta_0;\hat{\alpha}) &= a^T\sum_{i=1}^{m}\tilde{X}_{0,i}\Delta_{0,i}^T V_{0,i}^{-1}h_{0,i} + a^T\sum_{i=1}^{m}\tilde{X}_{0,i}\Delta_{0,i}^T V_{0,i}^{-1}\dot{h}_{0,i}\Delta_{0,i}\zeta_i(\beta_0;\hat{\alpha}) \\ &\quad + \sum_{i=1}^{m}\zeta_i(\beta_0;\hat{\alpha})\Delta_{0,i}^T\left(\left.\frac{\partial a^T\tilde{X}_i\Delta_i V_i^{-1}}{\partial\mu_i}\right|_{\mu_i=\mu_{0,i}}\right)h_{0,i} + R_m^*(\mu^*) \\ &\doteq I_1 + I_2 + I_3 + I_4, \end{aligned} \tag{A.4}$$

where $R_m^*(\mu^*) = \sum_{i=1}^{m} R_i^*(\mu_i^*)$ and $R_i^*(\mu_i^*) = \frac{1}{2}\zeta_i(\beta_0;\hat{\alpha})^T\Delta_i^T\left(\frac{\partial^2 a^T\tilde{X}_i\Delta_i V_i^{-1}h_i}{\partial\mu_i\partial\mu_i^T}\right)\Delta_i\zeta_i(\beta_0;\hat{\alpha})$ evaluated at $\mu_i^* = g^{-1}(\eta_{0,i} + \tau_i\zeta_i(\beta_0;\hat{\alpha}))$, $i = 1,\ldots,m$ with $0 < \tau_i < 1$.

By the central limit theorem,

$$I_1 \to N(0, a^T K^{-1/2}SK^{-1/2}a). \tag{A.5}$$

For $I_2$, considering that $\xi_1(\beta_0) = 0$, we have

$$\begin{aligned} I_2 &= a^T\sum_{i=1}^{m}\tilde{X}_{0,i}\Delta_{0,i}^T V_{0,i}^{-1}\dot{h}_{0,i}\Delta_{0,i}\zeta_i(\beta_0;\hat{\alpha}) \\ &= a^T\sum_{i=1}^{m}\tilde{X}_{0,i}\Delta_{0,i}^T V_{0,i}^{-1}\dot{h}_{0,i}\Delta_{0,i}(\tilde{X}_{0,i}\xi_1(\beta_0) + \tilde{M}_{0,i}\xi_2(\beta_0;\hat{\alpha}) + R_{ni}) \\ &= a^T\sum_{i=1}^{m}\tilde{X}_{0,i}\Delta_{0,i}^T V_{0,i}^{-1}\dot{h}_{0,i}\Delta_{0,i}(\tilde{M}_{0,i}\xi_2(\beta_0;\hat{\alpha}) + R_{ni}) \\ &= a^T\sum_{i=1}^{m}\tilde{X}_{0,i}\Delta_{0,i}^T V_{0,i}^{-1}(\dot{h}_{0,i} - \Gamma_{0,i})\Delta_{0,i}\tilde{M}_{0,i}\xi_2(\beta_0;\hat{\alpha}) + a^T\sum_{i=1}^{m}\tilde{X}_{0,i}\Delta_{0,i}^T V_{0,i}^{-1}\dot{h}_{0,i}\Delta_{0,i}R_{ni} \\ &\doteq I_{21} + I_{22}. \end{aligned} \tag{A.6}$$

For $I_{21}$, applying the Cauchy–Schwarz inequality, we have

$$\begin{aligned} I_{21}^2 &= \left(a^T\sum_{i=1}^{m}\tilde{X}_{0,i}\Delta_{0,i}^T V_{0,i}^{-1}(\dot{h}_{0,i} - \Gamma_{0,i})\Delta_{0,i}\tilde{M}_i\xi_2(\beta_0;\hat{\alpha})\right)^2 \\ &= \left(\sum_{k=1}^{p} a_k 1_k\sum_{i=1}^{m}\tilde{X}_{0,i}\Delta_{0,i}^T V_{0,i}^{-1}(\dot{h}_{0,i} - \Gamma_{0,i})\Delta_{0,i}\tilde{M}_{0,i}\xi_2(\beta_0;\hat{\alpha})\right)^2 \\ &\le \sum_{k=1}^{p}\left(\sum_{i=1}^{m} 1_k^T\tilde{X}_{0,i}\Delta_{0,i}^T V_{0,i}^{-1}(\dot{h}_{0,i} - \Gamma_{0,i})\Delta_{0,i}\tilde{M}_{0,i}\xi_2(\beta_0;\hat{\alpha})\right)^2 \\ &\le \sum_{k}^{p}\sum_{j}^{\bar{p}}\left(\sum_{i=1}^{m} 1_k^T\tilde{X}_{0,i}\Delta_{0,i}^T V_{0,i}^{-1}(\dot{h}_{0,i} - \Gamma_{0,i})\Delta_{0,i}\tilde{M}_{0,i}1_j\right)^2\|\xi_2(\beta_0;\hat{\alpha})\|^2, \end{aligned} \tag{A.7}$$

where $\bar{p} = N_k$, then by conditions (A.4)–(A.6) and $\|\xi_2(\beta_0; \hat{\alpha})\| = O_p(k_n^{1/2})$, we can obtain

$$\begin{aligned} E(I_{21}^2) &\le \sum_k^p \sum_j^{\bar{p}} \sum_{i=1}^m E\left(1_k^T \tilde{X}_{0,i} \Delta_{0,i}^T V_{0,i}^{-1} (\dot{h}_{0,i} - \Gamma_{0,i}) \Delta_{0,i} \tilde{M}_{0,i} 1_j\right) \|\xi_2(\beta_0; \hat{\alpha})\|^2 \\ &\le C \sum_k^p \sum_j^{\bar{p}} \sum_{i=1}^m 1_k^T \tilde{X}_{0,i}^T \tilde{X}_{0,i} 1_k E \|\Delta_{0,i}^T V_{0,i}^{-1} (\dot{h}_{0,i} - \Gamma_{0,i}) \Delta_{0,i} \tilde{M}_{0,i} 1_j\|^2 \|\xi_2(\beta_0; \hat{\alpha})\|^2 \\ &\le C \sup_i \sum_{k=1}^p 1_k^T \tilde{X}_{0,i}^T \tilde{X}_{0,i} 1_k \sum_{i=1}^m \sum_{j=1}^{\bar{p}} 1_j^T \tilde{M}_{0,i}^T \tilde{M}_{0,i} 1_j \|\xi_2(\beta_0; \hat{\alpha})\|^2 \\ &= C \sup_i \operatorname{trace}(\tilde{X}_{0,i} \tilde{X}_{0,i}^T) \operatorname{trace}\left(\sum_{i=1}^m \tilde{M}_{0,i} \tilde{M}_{0,i}^T\right) \|\xi_2(\beta_0; \hat{\alpha})\|^2 \\ &\le C k_n \sup_i \operatorname{trace}(\tilde{X}_{0,i} \tilde{X}_{0,i}^T) \|\xi_2(\beta_0; \hat{\alpha})\|^2 \\ &= O(k_n^2 n^{-1}), \end{aligned} \tag{A.8}$$

where the finite constant $C$, independent of $m$, may vary from line to line. Consequently, we have

$$\sup_{a^T a=1} |I_{21}| = O_p(k_n n^{-1/2}) = o_p(1). \tag{A.9}$$

It is not difficult to show

$$\sup_{a^T a=1} |I_{22}| = O_p(k_n^{-r}) = o_p(1). \tag{A.10}$$

Combining (A.9) and (A.10), we have

$$\sup_{a^T a=1} |I_2| = o_p(1). \tag{A.11}$$

Using the similar arguments to proof of (A.11), we can obtain $\sup_{a^T a=1} |I_3| = o_p(1)$ and $\sup_{a^T a=1} |I_4| = o_p(1)$.

So far, we prove $K_n^{-1/2} \sum_{i=1}^m Z_i(\beta_0; \hat{\alpha}) \to_D N(0, K^{-1/2} S K^{-1/2})$, hence Lemma 2 is proved. □

**Lemma 3.** *Under the conditions* (A.1)–(A.6), *we have*

$$1/n \sum_{i=1}^m Z_i(\beta_0; \hat{\alpha}) Z_i(\beta_0; \hat{\alpha})^T \to S. \tag{A.12}$$

Using Taylor expansion and similar arguments to the proof of Theorem 2 in [17], Lemma 3 can be obtained.

**Lemma 4.** *Under the conditions* (A.4)–(A.6), *we have*

$$\sup_i \|Z_i(\beta_0; \hat{\alpha})\| = o_p(n^{1/2}), \tag{A.13}$$

*and*

$$\|\lambda\| = O_p(n^{-1/2}). \tag{A.14}$$

**Proof.** By conditions (A.4)–(A.6), we have $\|X_i^*(\beta_0; \hat{\alpha})\| = O_p(n^{-1/2})$. Note that function $\psi(\cdot)$ is a bounded function, then we have

$$\sup_i \|Z_i(\beta_0; \hat{\alpha})\| = \sup_i \|X_i^*(\beta_0; \hat{\alpha}) \Delta_i^T(\beta_0; \hat{\alpha}) V_i^{-1}(\beta_0; \hat{\alpha}) h_i(\beta_0; \hat{\alpha})\| = o_p(n^{1/2}).$$

By Lemmas 2 and 3, Taylor expansion and using the same arguments as those used in the proof of (2.14) in [12], we can prove (A.14). □

**Proof of Theorem 1.** We first define bivariate functions $Q_1(\beta, \lambda; \hat{\alpha})$ and $Q_2(\beta, \lambda; \hat{\alpha})$ respectively as

$$Q_1(\beta, \lambda; \hat{\alpha}) = \frac{1}{m} \sum_{i=1}^m \frac{Z_i(\beta; \hat{\alpha})}{1 + \lambda^T Z_i(\beta; \hat{\alpha})}, \tag{A.15}$$

and

$$Q_2(\beta,\lambda;\hat{\alpha}) = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{1+\lambda^T Z_i(\beta;\hat{\alpha})}\left(\frac{\partial Z_i(\beta;\hat{\alpha})}{\partial\beta}\right)^T\lambda. \tag{A.16}$$

Under the conditions (A.3)–(A.4) and (A.6), if $\tilde{\beta}$ is the RMELE of $\beta$, and $\tilde{\lambda} = \lambda(\tilde{\beta})$ is the root of (2.4), following Lemma 1 in [16], we have

$$Q_1(\tilde{\beta},\tilde{\lambda};\hat{\alpha}) = 0, \qquad Q_2(\tilde{\beta},\tilde{\lambda};\hat{\alpha}) = 0.$$

By Taylor expansion at $(\beta_0, 0)$, conditions (A.4)–(A.6) and Lemmas 2 and 3, we have

$$\begin{aligned}\sqrt{n}(\tilde{\beta}-\beta_0) &= (K^TS^{-1}K)^{-1}K^TS^{-1}\sqrt{n}Q_1(\beta_0;\hat{\alpha}) + o_p(1)\\ &= (K^TS^{-1}K)^{-1}K^TS^{-1}\sqrt{n}Q_1(\beta_0;\hat{\alpha}) + o_p(1) \to N(0,V). \quad \square\end{aligned}$$

**Proof of Theorem 2.** Applying Taylor's expansion to (2.5), we have

$$-2\log R(\beta_0;\hat{\alpha}) = 2\sum_{i=1}^{m}\left\{\lambda^T Z_i(\beta_0;\hat{\alpha}) - \frac{1}{2}(\lambda^T Z_i(\beta_0;\hat{\alpha}))^2\right\} + o_p(1). \tag{A.17}$$

By Lemmas 2–4, we have the following representations:

$$\lambda = \left(\sum_{i=1}^{m} Z_i(\beta_0;\hat{\alpha})Z_i^T(\beta_0;\hat{\alpha})\right)^{-1}\sum_{i=1}^{m} Z_i(\beta_0;\hat{\alpha}) + o_p(n^{-1/2}); \tag{A.18}$$

$$\sum_{i=1}^{m}(\lambda^T Z_i(\beta_0;\hat{\alpha}))^2 = \sum_{i=1}^{m}(\lambda^T Z_i(\beta_0;\hat{\alpha})) + o_p(1). \tag{A.19}$$

By (A.17)–(A.19) and Lemmas 2 and 3, we have

$$\begin{aligned}-2\log R(\beta_0;\hat{\alpha}) &= \sum_{i=1}^{m}(\lambda^T Z_i(\beta_0;\hat{\alpha})) + o_p(1)\\ &\to \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{m} Z_i(\beta_0;\hat{\alpha})\right)^T\left(n^{-1}\sum_{i=1}^{m} Z_i(\beta_0;\hat{\alpha})Z_i^T(\beta_0;\hat{\alpha})\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{m} Z_i(\beta_0;\hat{\alpha})\right) + o_p(1)\\ &\to \chi_p^2. \quad \square\end{aligned}$$

## References


[1] Y. Bai, Z.Y. Zhu, W.K. Fung, Empirical likelihood inference for longitudinal generalized linear models, Journal of Statistical Planning and Inference 140 (2010) 3446–3456.
[2] E. Cantoni, A robust approach to longitudinal data analysis, Canadian Journal of Statistics 32 (2004) 169–180.
[3] E. Cantoni, E. Ronchetti, Robust inference for generalized linear models, Journal of the American Statistical Association 96 (2001) 1022–1030.
[4] X. He, W.K. Fung, Z.Y. Zhu, Robust estimation in generalized partial linear models for clustered data, Journal of the American Statistical Association 100 (2005) 1176–1184.
[5] N.E. Heckman, Spline smoothing in partly linear models, Journal of the Royal Statistical Society. Series B 48 (1986) 244–248.
[6] S. Heritier, E. Contoni, S. Copt, M.P. Victoria-Feser, Robust Methods in Biostatistics, John Wiley & Sons Ltd., 2009
[7] E.D. Kolaczyk, Empirical likelihood for generalized linear models, Statistica Sinica 4 (1994) 199–218.
[8] K.L. Liang, S.L. Zeger, Longitudinal data analysis using generalised estimating equations, Biometrika 73 (1986) 13–22.
[9] L. Madsen, D. Dalthorp, Simulating correlated count data, Environmental and Ecological Statistics 14 (2007) 129–148.
[10] P. McCullagh, J.A. Nelder, Generalized Linear Models, 2nd ed., Chapman and Hall, London, 1989
[11] A.B. Owen, Empirical likelihood ratio confidence intervals for a single functional, Biometrika 74 (1988) 237–249.
[12] A.B. Owen, Empirical likelihood ratio confidence regins, The Annals of Statistics 18 (1990) 90–120.
[13] A.B. Owen, Empirical Likelihood, Chapman and Hall, CRC, New York, 2001.
[14] J.S. Preisser, B.F. Qaqish, Robust regression for clustered data with application to binary responses, Biometrics 55 (1999) 574–579.
[15] G.Y. Qin, Y. Bai, Z.Y. Zhu, Robust empirical likelihood inference for longitudinal data, Statistics and Probability Letters 79 (2009) 2101–2108.
[16] J. Qin, J.F. Lawless, Empirical likelihood and genral estimating equations, The Annals of Statistics 22 (1994) 300–325.
[17] G.Y. Qin, Z.Y. Zhu, Robust estimation in generalized semiparametric mixed models for longitudinal data, Journal of Multivariate Analysis 98 (2007) 1658–1683.
[18] J. Rice, Convergence rates for patially splined models, Statistics and Probability Letters 4 (1986) 203–208.
[19] P.J. Rousseeuw, B.C. van Zomerem, Unmasking multivariate outliers and leverage points, Journal of the American Statistical Association 85 (1990) 633–639.
[20] L.L. Schumaker, Spline Functions, Wiley, New York, 1981.
[21] J. Shi, T.S. Lau, Robust empirical likelihood for linear model under median constrait, Communications in Statistics—Theory and Methods 28 (1999) 2465–2476.
[22] J. Shi, T.S. Lau, Empirical likelihood for partially linear models, Journal of Multivariate Analysis 72 (2000) 132–148.
[23] S.K. Sinha, Robust analysis of generalized linear mixed models, Journal of the American Statistical Association 99 (2004) 451–460.
[24] P. Speckman, Kernel smoothing in partial linear models, Journal of Multivariate Analysis 50 (1988) 413–436.
[25] Q.H. Wang, B.Y. Jin, Empirical likeilihoood for partial linear model with fixed designs, Statistics and Probability Letters 41 (1999) 425–433.
[26] L.G. Xue, L.X. Zhu, Empirical likelihood semiparametric regression analysis for longitudinal data, Biometrika 94 (2007) 921–937.