Robust Element-Wise Empirical Likelihood Estimation Method for Longitudinal Data

Tianyu Huang, Yali Fan, Zongren Sun

College of Science, University of Shanghai for Science and Technology, Shanghai, China
Email: yalifan@usst.edu.cn

Abstract

For the regression model about longitudinal data, we combine the robust estimation equation with the elemental empirical likelihood method, and propose an efficient robust estimator, where the robust estimation equation is based on bounded scoring function and the covariate depended weight function. This method reduces the influence of outliers in response variables and covariates on parameter estimation, takes into account the correlation between data, and improves the efficiency of estimation. The simulation results show that the proposed method is robust and efficient.

Keywords

Longitudinal Data, Element-Wise Empirical Likelihood, Robust Estimation Equation

1. Introduction

Longitudinal data is a dataset obtained by repeatedly measuring multiple times for each individual over a period of time. The longitudinal data is equivalent to the combination of cross section and time series data, and is composed of a plurality of short time series. For a fixed time point, the observation data of different individuals is similar to the cross-sectional data; for fixed individuals, different time points observation data is similar to time series. Therefore, longitudinal data can make full use of the information inside the individual while distinguishing individual differences. In the fields of medicine and finance, the frequency of longitudinal data appears to be higher and higher, so the research on longitudinal data is of great significance.

Longitudinal data is a hot topic in statistical research in recent years. So far, significant progress has been made in the field of theoretical research. Liang *et al.* (1986) [1] extended generalized linear model research to longitudinal data,

proposed a generalized estimating equation (GEE) method, introduced correlation matrices in estimating equations, and gave corresponding estimates of regression parameters and their variances. It is proved that the consistent estimation of regression coefficients can be obtained by using GEE method even if the work correlation matrix is misspecified (See Diggle *et al.* (2002) [2] for more details). However, the principle of the GEE developed from the generalized linear model is similar to the principle of the weighted least squares method, and is sensitive to outliers. In the longitudinal data, because of repeated measurements, there are abnormal values in individual measurements, which will lead to a series of abnormal values in samples. In order to reduce the interference of outliers, Fan *et al.* (2012) [3] introduced a generalized estimating equation method based on the bounded scoring function of Huber function to achieve robustness. For the definition of Huber function, (see Huber (1964) [4]), Wang *et al.* (2013) [5] and Lv *et al.* (2015) [6] applied the bounded exponential score function to the generalized estimating equation. There are many articles (Qin *et al.* (2005) [7]; Wang *et al.* (2005) [8]; Qin *et al.* (2009) [9]; Zheng *et al.* (2013) [10]) about generalized estimation equation and robustness research.

In the field of longitudinal data research, empirical likelihood methods are also one of the frequently used methods. The empirical likelihood (EL) method was originally applied by Owen (1988) [11] to the estimation of the population mean of completely independent and identically distributed data. The method has the characteristics of asymmetric confidence intervals, transformation-preserving and better coverage probability. Azzalini (2017) [12] comprehensively introduced the application of empirical likelihood method in statistical inference. Qin and Lawless (1994) [13] first linked the empirical likelihood method with the estimation equation. They proved that the empirical likelihood estimation is effective when the moment conditions are correctly specified in the estimation equation. Bondell and Stefansk (2013) [14] proposed a robust estimator in linear regression which has relatively high efficiency compared to other robust estimators by using generalized EL methods. Bai *et al.* (2010) [15] introduced the EL method into longitudinal data research, and proposed a weighted empirical likelihood (WEL) inference method for generalized linear models of longitudinal data. Wang *et al.* (2010) [16] established two methods based on generalized empirical likelihood: elemental empirical likelihood and object empirical likelihood, where the element-wise empirical likelihood method can give slightly better coverage probabilities for small or medium samples.

Based on the existing research results of longitudinal data, this paper combines the bounded scoring function based on Huber function and the robust estimation equation of weight function based on covariate with elemental empirical likelihood method, and proposes an effective robust estimation method. The effectiveness of the proposed method is derived from the advantages of the empirical likelihood method, while robustness is obtained by means of robust estimation equations. We carried out the simulation and the results showed that

whether the correlation matrix used is consistent with the real correlation matrix or not, the estimation method in this paper can reduce the impact of outliers on the estimation and improve the estimation efficiency.

The following content is divided into four subsections. In Section 1, we give the linear regression model of the longitudinal data and the estimation method used in this paper. The iterative algorithm of this paper is introduced in Section 2. Section 3 is the simulation experiment part and Section 4 is the summary and outlook.

2. Proposed Method

2.1. Models

Linear models are often used in longitudinal data research. Their structure is simple for analysis and the basis of many models. We will consider the following continuous response variable longitudinal regression model

$$y_{ij} = x_{ij}^{\mathrm{T}} \beta + \varepsilon_{ij}, \quad i = 1, \cdots, n; \quad j = 1, \cdots, m_i \tag{2.1}$$

where y_{ij} is the jth observation on the ith subject, x_{ij} is a p-vector of covariance values and β_0 is a p-vector of unknown regression coefficients, $x_{ij}^{\mathrm{T}} = \left(x_{ij}^{(1)}, x_{ij}^{(2)}, \cdots, x_{ij}^{(p)} \right)$, $\beta^{\mathrm{T}} = \left(\beta^{(1)}, \beta^{(2)}, \cdots, \beta^{(p)} \right)$. n is the number of subjects participating in the study, m_i is the number of repeated measurements for the ith subject, $m_1, \cdots, m_n$ are bounded positive integers. Denote the total sample size by $N = m_1 + \cdots + m_n$. ε_{ij} is the random error term, satisfying $E(\varepsilon_{ij}) = 0$, $Var(\varepsilon_{ij}) = \sigma^2$. We set $Y_i^{\mathrm{T}} = (y_{i1}, y_{i2}, \cdots, y_{im_i})$, $X_i = (x_{i1}, x_{i2}, \cdots, x_{im_i})$ and $\varepsilon_i^{\mathrm{T}} = (\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{im_i})$. Then model (2.1) can be rewritten as

$$Y_i = X_i^{\mathrm{T}} \beta + \varepsilon_i \tag{2.2}$$

For the longitudinal data model, it is usually assumed that the variables between different individuals are independent of each other, and the different measurements of the same individual are related. The covariance matrix of the random vector ε_i is $V_i = \sigma^2 R(\rho) = A_i^{\frac{1}{2}} R(\rho) A_i^{\frac{1}{2}}$, where V_i is an $m_i \times m_i$ invertible matrix, $R(\rho)$ is a correlation matrix, ρ is a correlation coefficient vector, $A_i^{\frac{1}{2}} = \sigma I_{m_i}$. Exchangeable structure (Exch), work-independent structure (Ind) and first-order autoregressive structure (AR(1)) are common related structures in practice.

$$R_{Exch} = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}, \quad R_{Ind} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}, \quad R_{AR(1)} = \begin{pmatrix} 1 & \rho & \cdots & \rho^{m_i-1} \\ \rho & 1 & \cdots & \rho^{m_i-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{m_i-1} & \rho^{m_i-2} & \cdots & 1 \end{pmatrix}.$$

Let

$$Z_i(\beta) = \left\{ z_{i1}(\beta), z_{i2}(\beta), \cdots, z_{im_i}(\beta) \right\}^{\mathrm{T}} = V_i^{-1} \left(Y_i - X_i^{\mathrm{T}} \beta \right) \tag{2.3}$$

2.2. Proposed Estimator

More generally, we can define an estimating equation

$$\sum_{i=1}^{n} X_i Z_i(\beta) = \sum_{i=1}^{n}\sum_{j=1}^{m_i} x_{ij} z_{ij}(\beta) = 0 \qquad (2.4)$$

Such estimating equation is susceptible to the influence of outliers. Bounded scoring function of Huber function and weight function depending on covariates are introduced in formula (3)

$$Z_i^R(\beta) = \left\{ z_{i1}^R(\beta), z_{i2}^R(\beta), \cdots, z_{im_i}^R(\beta) \right\} = \left(A_i^{\frac{1}{2}} R_i \right)^{-1} W_i \psi_c \left(A_i^{-\frac{1}{2}} \left(Y_i - X_i^{\mathrm{T}} \beta \right) \right) \quad (2.5)$$

Consequently, a robust estimation equation is obtained

$$\sum_{i=1}^{n} X_i Z_i^R(\beta) = \sum_{i=1}^{n}\sum_{j=1}^{m_i} x_{ij} z_{ij}^R(\beta) = 0 \qquad (2.6)$$

where $\psi_c(x) = \min(c, \max(-c, x))$ is the bounded scoring function, it is used to limit the influence on outliers in response. Because it is applied to the standardized residuals, the value of c are generally between 1 and 2. $W_i = diag\left(w_{i1}, w_{i2}, \cdots, w_{im_i} \right)$ is the weight function. There are many ways to select the weight function, similar to the reference [14], we consider a function of the Mahalanobis distance

$$w_{ij} = \min\left\{ 1, \left(\frac{b_0}{\left(x_{ij} - m_x \right)^{\mathrm{T}} S_x^{-1} \left(x_{ij} - m_x \right)} \right)^{\frac{r}{2}} \right\} \qquad (2.7)$$

where $r \geq 1$, b_0 is the 0.95 quantile of the chi-square distribution with the degree of freedom equal to the dimension of x_{ij}, m_x and S_x are some robust estimates of location and scatter of x_{ij}. If x_{ij} deviates from the whole, $\left(x_{ij} - m_x \right)^{\mathrm{T}} S_x^{-1} \left(x_{ij} - m_x \right)$ will be larger and

$$\left(\frac{b_0}{\left(x_{ij} - m_x \right)^{\mathrm{T}} S_x^{-1} \left(x_{ij} - m_x \right)} \right)^{\frac{r}{2}}$$

will be smaller. Then the corresponding weight w_{ij} is less than 1. On the contrary, the corresponding weight w_{ij} is 1. Therefore, the influence of the outliers on the estimation can be controlled by the Mahalanobis distance function. For data without outlying points, we can set $\psi_c(x) = x$ and $w_{ij} = 1$ in the robust estimating Equation (2.6), and this will lead to the classical nonrobust estimating equation.

Empirical likelihood method is a non-parametric statistical method, which has many good properties. The empirical likelihood and the estimated equation were first associated by Qin and Lawless (1994) [13]. Wang *et al.* (2010) [16] proposed two methods for combining empirical likelihood methods with estimation equations. Subject-wise empirical likelihood is assigning a probability p_i to each subject y_i. Li *et al.* (2018) [17] introduced a robust estimation equation based

on the exponential score function and prove that a better estimate can be obtained with outliers. Element-wise empirical likelihood is assigning a probability mass p_{ij} to each observation y_{ij}. This paper combines the element-wise empirical likelihood method with the robust estimation Equation (2.6) to obtain the following empirical likelihood ratio function

$$L_1(\beta) = \sup\left\{\prod_{i=1}^{n}\prod_{j=1}^{m_i} Np_{ij} \mid p_{ij} \ge 0, \sum_{i=1}^{n}\sum_{j=1}^{m_i} p_{ij} = 1, \sum_{i=1}^{n}\sum_{j=1}^{m_i} p_{ij}x_{ij}z_{ij}^R(\beta) = 0\right\} \quad (2.8)$$

Similar to Owen (1988) [11] classic empirical likelihood, by using the Lagrangian multiplier method, we obtain that $L_1(\beta)$ is maximized at

$$p_{ij} = \frac{1}{N}\left(1 + \lambda'x_{ij}z_{ij}^R(\beta)\right)^{-1}, i = 1,\cdots,n; j = 1,\cdots,m_i \quad (2.9)$$

where the vector $\lambda = (\lambda_1,\lambda_2,\cdots,\lambda_p)'$ satisfies $\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{x_{ij}z_{ij}^R(\beta)}{1+\lambda'x_{ij}z_{ij}^R(\beta)} = 0$.

From this we can get

$$-2\log L_1(\beta) = 2\sum_{i=1}^{n}\sum_{j=1}^{m_i}\log\left(1+\lambda'x_{ij}z_{ij}^R(\beta)\right) \quad (2.10)$$

The estimate proposed in this paper is the maximum point of Equation (2.8), which is equivalently the minimum point of Equation (2.10). Recorded as

$$\hat{\beta}_{\text{RELGEE}} = \arg\max L_1(\beta) = \arg\min\left(-2\log L_1(\beta)\right) \quad (2.11)$$

$\hat{\beta}_{\text{RELGEE}}$ is the RELEGE estimate of β. Under some regular conditions, the estimate can be proved to be consistent and subject to an asymptotic normal distribution. $-2\log L_1(\beta)$ subject to an asymptotic linear combination of independent chi-Square distribution. The proof process is similar to Wang *et al.* (2010) [16].

The parameter of interest in this paper is β. But the parameters σ and ρ are unknown. Reference to He *et al.* (2005) [18], we can use a median of absolute deviations to give σ a robust estimate

$$\hat{\sigma} = \left\{1.483\text{median}\left\{\left|\left(y_{ij} - x_{ij}^{\text{T}}\beta^*\right) - \text{median}\left(y_{ij} - x_{ij}^{\text{T}}\beta^*\right)\right|\right\}\right\} \quad (2.12)$$

where β^* is the current estimate of β.

According to the above description of the correlation matrix, the specific robust estimation of the parameter ρ depends on the selection of the relevant structure. Simply consider, when $m_i = m$, for Exchange-able structure (Exch)

$$\hat{\rho} = \frac{\sum_{i=1}^{n}\sum_{j>j'}\hat{r}_{ij}^*\hat{r}_{ij'}^*}{0.5nm(m-1)-p} \quad (2.13)$$

where $\hat{r}_{ij}^* = \psi_c\left(\dfrac{y_{ij} - x_{ij}^{\text{T}}\beta^*}{\hat{\sigma}}\right)$. For first-order autoregressive structure (AR(1)),

We use robust moment estimators

$$\hat{\rho} = \frac{a}{b} - \sqrt{\frac{a}{b}-1} \quad (2.14)$$

where $a = \sum_{i=1}^{n} r_i^{*'} T_1 r_i^{*}$, $b = \sum_{i=1}^{n} r_i^{*'} T_2 r_i^{*}$, $r_i^{*} = \left(r_{i1}^{*}, r_{i2}^{*}, \cdots, r_{im}^{*} \right)^{\mathrm{T}}$, T_1 is m-order diagonal matrix $diag(1,1,1,1,\cdots,1,1)$ and T_2 is m-order diagonal matrix $diag(0,1,1,1,\cdots,1,0)$.

3. Algorithm

Since the maximal estimation of computational empirical likelihood will encounter numerical calculation problems, when solving $\hat{\beta}_{\mathrm{RELGEE}}$, we refer to the Newton-type algorithm of Lagrange multiplier for constrained optimization problems proposed by Özdemir (2018) [19]. In order to make the calculation simple and without loss of generality, we pull p_{ij}, y_{ij}, x_{ij} and $z_{ij}^{R}(\beta)$ into a column vector. Like $(p_1, p_2, \cdots, p_N) = \left(p_{11}, \cdots, p_{1m_1}, \cdots, p_{n1}, \cdots, p_{nm_n} \right)$, $(y_1, y_2, \cdots, y_N)$, $\left(z_1^{R}(\beta), z_2^{R}(\beta), \cdots, z_N^{R}(\beta) \right)$ and $(x_1, x_2, \cdots, x_N)$ are the same. We get

$$L_2(\beta) = \sup\left\{ \prod_{i=1}^{N} p_i \mid \sum_{i=1}^{N} p_i = 1, p_i > 0, \sum_{i=1}^{N} p_i x_i z_i^{R}(\beta) \right\} \tag{3.1}$$

This problem can also be defined as follows

$$J(p, \beta) = \min_{p_i \in (0,1)} \left[-\sum_{i=1}^{N} \ln p_i \right] \tag{3.2}$$

under the constraints $\sum_{i=1}^{N} p_i = 1$ and $\sum_{i=1}^{N} p_i x_i z_i^{R}(\beta) = 0$.

Lagrangian multiplier method is commonly used to find the extreme value of a general constrained function. We can get

$$J\left(p, \beta, \lambda_0, \lambda_1^{\mathrm{T}} \right) = -\sum_{i=1}^{N} \ln p_i + \lambda_0 \left(\sum_{i=1}^{N} p_i - 1 \right) - \lambda_1^{\mathrm{T}} \sum_{i=1}^{N} p_i x_i z_i^{R}(\beta) \tag{3.3}$$

where $\lambda_0 \in R^1$ and $\lambda_1^{\mathrm{T}} \in R^p$ are Lagrange multipliers

$$A = \begin{bmatrix} p \\ \beta \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \lambda_0 \\ \lambda_1^{\mathrm{T}} \end{bmatrix} \tag{3.4}$$

The first order gradient of Equation (3.3) is

$$\nabla J(A, \Lambda) = \begin{bmatrix} \nabla R(A) + \left(DG(A) \right)^{\mathrm{T}} \Lambda \\ G(A)^{\mathrm{T}} \end{bmatrix} \tag{3.5}$$

where

$$\nabla J(A) = \begin{bmatrix} -\dfrac{1}{p_1} - \dfrac{1}{p_2} \cdots - \dfrac{1}{p_N} & 0_{1 \times p} \end{bmatrix}^{\mathrm{T}}, \quad G(A) = \left[\sum_{i=1}^{N} p_i - 1, \sum_{i=1}^{N} p_i x_i z_i^{R}(\beta) \right]$$

$G(A)$ Jacobi matrix about A can be written

$$DG(A) = \begin{bmatrix} 1 & 1 & \cdots & 1 & 0_{l \times p} \\ x_1 z_1^{R}(\beta) & x_2 z_2^{R}(\beta) & \cdots & x_N z_N^{R}(\beta) & -\left[\sum_{i=1}^{N} p_i x_i \phi_i \right]^{\mathrm{T}} \end{bmatrix} \tag{3.6}$$

where $\phi_i = \dfrac{\partial z_i^{R}(\beta)}{\partial \beta}$.

So the Hessian matrix of Equation (3.3) is

$$HL(A, \Lambda) = \begin{bmatrix} B(A, \Lambda) & DG(A)^{\mathrm{T}} \\ DG(A) & 0_{p \times p} \end{bmatrix} \tag{3.7}$$

where $B(A, \Lambda) = \begin{bmatrix} \frac{1}{p_1^2} & 0 & \cdots & 0 & -\lambda_1^{\mathrm{T}} x_1 \phi_1 \\ 0 & \frac{1}{p_2^2} & \cdots & 0 & -\lambda_1^{\mathrm{T}} x_2 \phi_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{1}{p_N^2} & -\lambda_1^{\mathrm{T}} x_N \phi_N \\ -[\lambda_1^{\mathrm{T}} x_1 \phi_1]^{\mathrm{T}} & -[\lambda_1^{\mathrm{T}} x_2 \phi_2]^{\mathrm{T}} & \cdots & -[\lambda_1^{\mathrm{T}} x_N \phi_N]^{\mathrm{T}} & 0_{p \times p} \end{bmatrix}$,

it's a $(N+k) \times (N+k)$ matrix.

By Newton iteration, we can get

$$HL(A_t, \Lambda_t) \begin{bmatrix} A_{t+1} - A_t \\ \Lambda_{t+1} - \Lambda_t \end{bmatrix} + \nabla L(A_t, \Lambda_t) = 0 \tag{3.8}$$

where the value of A_t, Λ_t represents the result of iteration A, Λ at the tth time, the value of A_{t+1}, Λ_{t+1} is calculated by A_t, Λ_t.

We can get the iterative expression of A_t, Λ_t

$$\begin{bmatrix} A_{t+1} \\ \Lambda_{t+1} \end{bmatrix} = \begin{bmatrix} A_t \\ \Lambda_t \end{bmatrix} - HL(A_t, \Lambda_t)^{-1} \nabla L(A_t, \Lambda_t) \tag{9}$$

Summarize the algorithm for estimating the parameter $\hat{\beta}_{\mathrm{RELGEE}}$ as follows:

Step 1. Set the initial value of $p, \beta, \lambda_0, \lambda_1$ and the threshold ξ, where the initial value of β can be set to the estimated value obtained by the least squares method to speed up the convergence.

Step 2. Calculate $HL(A_t, \Lambda_t)$ and $\nabla L(A_t, \Lambda_t)$ at the tth time.

Step 3. Calculate A_{t+1}, Λ_{t+1} according to Iterative Formula (3.9).

Step 4. Let $t \leftarrow t+1$, repeat Step 2 and Step 3 until $\|\beta_{t+1} - \beta_t\| < \xi$ is satisfied and obtain a convergent $\hat{\beta}_{\mathrm{RELGEE}}$. At this point, the solution of the robust empirical likelihood estimate is completed.

4. Simulation Study

In this section, we present a simulation study. The estimators obtained by the RELGEE method proposed in this paper are compared with the estimators obtained by the common element empirical likelihood method (ELGEE). The finite sample properties of the estimators are explored. The main research contents are as follows:

1) Estimated relative efficiency when there is no pollution in the data;

2) Estimated robustness when the data is contaminated;

3) The effect on the estimation efficiency when the work correlation matrix is correctly or incorrectly specified.

The model is set to

$$y_{ij} = x_{ij}^{\mathrm{T}} \beta + \varepsilon_{ij}, i = 1, 2, \cdots, n; j = 1, 2, 3 \tag{4.1}$$

where the number of repeated observations is 3, $x_{ij}^{\mathrm{T}} = \left(x_{ij}^{(1)}, x_{ij}^{(2)}, x_{ij}^{(3)} \right)$, $x_{ij} \sim \left((0,0,0)^{\mathrm{T}}, I_3 \right)$, $\beta^{\mathrm{T}} = (7, 1.7, -1.5)$ and

$$\left(\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3} \right)^{\mathrm{T}} \sim N\left((0,0,0)^{\mathrm{T}}, R(\rho) \right).$$

Considering sample size $n = 50$ and $n = 100$. Let $R(\rho)$ take exchangeable structure (Exch) and first-order autoregressive structure (AR(1)), where the parameter ρ is taken as 0.3 and 0.7 respectively. Because of the different values of parameter ρ and the different settings of real correlation matrix and work correlation matrix, We repeat the simulation 1000 times for different settings to calculate MSE ($\times 100$) that represents 100 times the mean square error of the sample under different conditions. Since there are three parameters, we find the average of the mean square error of the three parameters.

In order to study the problem (1), we compared the mean squared error of the estimating method (RELGEE) and the ordinary element empirical likelihood method (ELGEE) in the case of no pollution. The simulation results are shown in Table 1.

When processing non-polluting data, due to the robust processing of the longitudinal data to some extent, resulting in the loss of part of the information, the efficiency of robust estimation is usually lower than the non-stable estimate when there is no pollution. Table 1 shows that the mean square error of RELGEE estimator is only slightly larger than that of ELGEE estimator, which shows that this method is efficient even in the case of no pollution.

In order to explore questions (2) and (3), we have designed three ways of pollution:

Table 1. Comparison of Two Estimators under Non-pollution Conditions.

	Non-pollution			TR		
n	ρ	WR	Exch		AR(1)	
			ELGEE	RELGEE	ELGEE	RELGEE
	0.3	Exch	0.57386	0.58100	0.57386	0.58100
		AR(1)	0.58885	0.60091	0.54477	0.56010
50		Ind	0.72633	0.73693	0.72633	0.73693
	0.7	Exch	0.28785	0.29751	0.36407	0.37805
		AR(1)	0.36407	0.37805	0.28785	0.29751
		Ind	0.66757	0.67960	0.66757	0.67960
	0.3	Exch	0.27770	0.27829	0.39119	0.41201
		AR(1)	0.34903	0.35454	0.29119	0.29492
100		Ind	0.44306	0.43724	0.40696	0.41015
	0.7	Exch	0.17467	0.17678	0.16708	0.16904
		AR(1)	0.19500	0.19504	0.14110	0.14566
		Ind	0.42543	0.42879	0.40011	0.42267

WR: Woring R, TR: True R.

(C1). Pollution of the response variable Y: randomly turn $S\%$ of y_{ij} into $y_{ij}+b$, $b \sim N(2,1)$.

(C2). Pollution only for covariate X: randomly turn $S\%$ of x_{ij} into $x_{ij}+a$, $a \sim N\left((1,1,1)^{\mathrm{T}}, I_3\right)$.

(C3). Simultaneous contamination of X and Y: randomly turn $S\%/3$ of y_{ij} into $y_{ij}+b$, $b \sim N(2,1)$, randomly turn $S\%/3$ of x_{ij} into $x_{ij}+a$, $a \sim N\left((1,1,1)^{\mathrm{T}}, I_3\right)$.

Where $S\%$ is pollution rate. In this paper, 0.06 and 0.1 are selected. Some simulation results are shown in Tables 2-4.

Table 2 is the simulation results under C1. By comparison, in most cases, the

Table 2. Comparison of Two Estimators under C1.

C1				TR			
n	ρ	YP	WR	Exch		AR-1	
				ELGEE	RELGEE	ELGEE	RELGEE
			Exch	1.04790	0.79504	0.99345	0.74981
		0.06	AR-1	1.07018	0.81152	0.97866	0.73729
	0.3		Ind	1.14262	0.87243	0.98603	0.76988
			Exch	1.23468	0.88941	1.50948	1.05796
		0.1	AR-1	1.29277	0.92924	1.43798	1.02313
50			Ind	1.33828	0.98621	1.46806	1.03712
			Exch	0.80301	0.56517	0.91464	0.62653
		0.06	AR-1	0.89749	0.65623	0.86469	0.62672
	0.7		Ind	1.18518	0.99555	1.02135	0.84933
			Exch	1.15356	0.78475	1.30655	0.79852
		0.1	AR-1	1.28685	0.90425	1.24750	0.79105
			Ind	1.46320	1.12052	1.39281	1.01290
			Exch	0.42925	0.36534	0.45132	0.41615
		0.06	AR-1	0.44041	0.37885	0.44808	0.40805
	0.3		Ind	0.47229	0.41564	0.45741	0.42950
			Exch	0.49700	0.44274	0.42543	0.38569
		0.1	AR-1	0.50975	0.45279	0.41175	0.37517
100			Ind	0.53253	0.48246	0.43528	0.38958
			Exch	0.26753	0.23767	0.28060	0.24274
		0.06	AR-1	0.30773	0.27131	0.28723	0.23802
	0.7		Ind	0.42582	0.39690	0.47786	0.43469
			Exch	0.33180	0.25774	0.36312	0.30250
		0.1	AR-1	0.39815	0.31492	0.38295	0.30005
			Ind	0.51711	0.46137	0.50765	0.46074

YP: Pollution probability of the response variable *Y*.

Table 3. Comparison of Two Estimators under C2.

C2				TR			
n	*ρ*	*XP*	WR	Exch		AR-1	
				ELGEE	RELGEE	ELGEE	RELGEE
			Exch	12.91886	0.99187	13.97928	1.02373
		0.06	AR(1)	12.82331	0.99716	14.12307	1.02851
	0.3		Ind	13.06875	1.03024	13.80362	1.00403
			Exch	42.76916	3.00972	39.56463	2.88551
		0.1	AR(1)	43.36231	3.05236	39.30508	2.88421
50			Ind	43.31341	3.05817	40.03126	2.89432
			Exch	14.44630	0.88005	13.3182	0.89685
		0.06	AR(1)	14.57675	0.93714	13.72404	0.87365
	0.7		Ind	14.76241	1.09283	13.10561	0.93932
			Exch	39.52561	2.89129	37.91395	2.73270
		0.1	AR(1)	40.71029	3.08226	37.78436	2.74435
			Ind	40.76240	3.18213	39.59294	2.85362

XP: Pollution probability of covariate X.

Table 4. Comparison of Two Estimators under C3.

C3				TR			
n	*ρ*	*YXP*	WR	Exch		AR-1	
				ELGEE	RELGEE	ELGEE	RELGEE
			Exch	17.02696	1.67541	13.75782	1.83632
		0.06	AR(1)	17.02611	1.68535	13.80805	1.77938
	0.3		Ind	16.81146	1.69671	13.69043	1.70931
			Exch	17.74346	1.62392	15.18327	1.75757
		0.1	AR(1)	18.27163	1.59380	15.07867	1.73827
50			Ind	19.71323	1.63598	15.07310	1.73339
			Exch	19.71323	1.66511	16.65348	1.49746
		0.06	AR(1)	20.38681	1.82246	16.67835	1.47826
	0.7		Ind	19.18665	1.92724	15.74532	1.71987
			Exch	15.62539	1.63831	15.67169	1.71610
		0.1	AR(1)	17.59703	1.83022	16.35533	1.61183
			Ind	14.57003	2.01667	16.47950	1.73774

YXP: Pollution probability of the response variable *Y* when Pollution probability of covariate *X* is 0.06.

mean square error of the estimator in this paper is smaller than that of the estimator in ELGEE method. Comparing different pollution intensity, we can see that the greater the pollution intensity, the more obvious the superiority of robust estimation efficiency, which shows that the estimation method in this paper

has a strong robustness. It can also be seen that when the working matrix is set incorrectly, the difference of estimator is relatively small; when the working correlation matrix is a real matrix, the estimation efficiency is the highest; when the working matrix is independent structure (Ind), the estimation efficiency is the lowest without considering the correlation between data.

Table 3 and Table 4 are part of the results under C2 and pollution mode C3. Similar to Table 2, in most cases, the mean square error of this estimator is smaller than that of ELGEE estimator. Compared with Table 2, the mean square error of ELGEE estimator is only one-tenth of the mean square error of ELGEE estimator when the pollution intensity is significantly increased. The estimation method in this paper can significantly reduce the impact of outliers on the estimator. Similarly, the estimation efficiency is the highest when the working correlation matrix is a real matrix. When there is intra-group correlation in the model, the estimation efficiency is the lowest when the correlation is neglected in the estimation, which reflects the necessity of considering the longitudinal data model.

It is worth adding that because the method in this paper is a non-parametric method, the distribution of random errors does not necessarily follow the normal distribution. We simulate the distribution of random errors satisfying $t(3)$ again. The simulation results are shown in Table 5. Bias (×100) represents 100 times the deviation.

In Table 5, in all cases, the mean square error of the estimator in this paper is less than that of ELGEE estimator, which shows that the estimation method constructed in this paper can also effectively and robustly estimate the data of heavy-tailed distribution.

5. Summary

We introduce the generalized estimating equations commonly used in longitudinal data, and derive robust estimation functions. Then we combine the robust estimation equations with the elemental empirical likelihood method to obtain the empirical likelihood ratio function of the estimated parameters. We show a

Table 5. The error distribution of data satisfies $t(3)$.

			$t(3)$			
n	ρ	WR	Bias (×100)		MSE (×100)	
			ELGEE	RELGEE	ELGEE	RELGEE
	0.3	Exch	−0.03467	−0.05221	1.81182	1.01774
		AR(1)	0.17968	0.30765	1.75744	1.15462
50		Ind	0.26076	0.11889	2.14819	1.30291
	0.7	Exch	0.55148	0.64514	0.84950	0.58737
		AR(1)	0.18788	0.18082	1.25479	0.67416
		Ind	−0.19433	−0.07454	2.04687	1.18954

relatively optimized algorithm that can improve the efficiency and computational time of operation. We do a systematic simulation study. The simulation results show that our method maintains high estimation efficiency when the data is not polluted; in the case of data pollution, the estimator of this paper is obviously better than the non-robust estimator. With the increase of pollution intensity, the robustness of our method is more significant, and it has a significant resistance to outliers. When the working matrix is set incorrectly, the difference of estimator is relatively small; when the working correlation matrix is a real matrix, the estimation efficiency is the highest; when the working matrix is independent structure (Ind), the estimation efficiency is the lowest without considering the correlation between data. At the same time, since the estimator in this paper is based on empirical likelihood method, it is suitable for the longitudinal data of thick-tailed distribution. There are still many problems worth further study in this paper, such as the application of estimation methods to partial linear models and variable selection based on robust estimation.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Liang, K.Y. and Zeger, S.L. (1986) Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, **73**, 13-22. https://doi.org/10.1093/biomet/73.1.13

[2] Diggle, P., Diggle, P.J., Heagerty, P., *et al.* (2002) Analysis of Longitudinal Data. Oxford University Press, Oxford.

[3] Fan, Y.L, Qin, G.Y. and Zhu, Z.Y. (2012) Variable Selection in Robust Regression Models for Longitudinal Data. *Journal of Multivariate Analysis*, **109**, 156-167. https://doi.org/10.1016/j.jmva.2012.03.007

[4] Huber, P.J. (1964) Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, **35**, 73-101. https://doi.org/10.1214/aoms/1177703732

[5] Wang, X., Jiang, Y., Huang, M. and Zhang, H.P. (2013) Robust Variable Selection with Exponential Squared Loss. *Journal of the American Statistical Association*, **108**, 632-643. https://doi.org/10.1080/01621459.2013.766613

[6] Lv, J., Yang, H. and Guo, C. (2015) An Efficient and Robust Variable Selection Method for Longitudinal Generalized Linear Models. *Computational Statistics Data Analysis*, **82**, 74-88. https://doi.org/10.1016/j.csda.2014.08.006

[7] Qin, G.Y., Zhu, Z.Y. and Fung, W.K. (2009) Robust Estimation of Covariance Parameters in Partial Linear Model for Longitudinal Data. *Journal of Statistical Planning and Inference*, **139**, 558-570. https://doi.org/10.1016/j.jspi.2008.03.042

[8] Wang, Y.G., Lin, X. and Zhu, M. (2005) Robust Estimating Functions and Bias Correction for Longitudinal Data Analysis. *Biometrics*, **61**, 684-691. https://doi.org/10.1111/j.1541-0420.2005.00354.x

[9] Qin, G.Y. and Zhu, Z.Y. (2007) Robust Estimation in Generalized Semiparametric Mixed Models for Longitudinal Data. *Journal of Multivariate Analysis*, **98**, 1658-1683. https://doi.org/10.1016/j.jmva.2007.01.006

[10] Zheng, X., Fung, W.K. and Zhu, Z. (2013) Robust Estimation in Joint Mean? Covariance Regression Model for Longitudinal Data. *Annals of the Institute of Statistical Mathematics*, **65**, 617-638. https://doi.org/10.1007/s10463-012-0383-8

[11] Owen, A.B. (1988) Empirical Likelihood Ratio Confidence Intervals for a Single Functional. *Biometrika*, **75**, 237-249. https://doi.org/10.1093/biomet/75.2.237

[12] Azzalini, A. (2017) Statistical Inference Based on the Likelihood. Routledge, Abingdon-on-Thames.

[13] Qin, J. and Lawless, J. (1994) Empirical Likelihood and General Estimating Equations. *The Annals of Statistics*, **100**, 300-325.
https://doi.org/10.1214/aos/1176325370

[14] Bondell, H.D. and Stefansk, L.A. (1994) Empirical Robust Regression via Two-Stage Generalized Empirical Likelihood. *Journal of the American Statistical Association*, **108**, 644-655. https://doi.org/10.1080/01621459.2013.779847

[15] Bai, Y., Fung, W.K. and Zhu, Z. (2010) Weighted Empirical Likelihood for Generalized Linear Models with Longitudinal Data. *Journal of Statistical Planning and Inference*, **140**, 3446-3456. https://doi.org/10.1016/j.jspi.2010.05.007

[16] Wang, S., Qian, L. and Carroll, R.J. (2010) Generalized Empirical Likelihood Methods for Analyzing Longitudinal Data. *Biometrika*, **97**, 79-93.
https://doi.org/10.1093/biomet/asp073

[17] Li, S.M., Ren, Y.Y. and Lu, G. (2018) A Robust Estimation and Empirical Likelihood Inference for Longitudinal Data Models. *Journal of Applied Statistics and Management*, **4**, 631-638.

[18] He, X., Fung, W.K. and Zhu, Z. (2005) Robust Estimation in Generalized Partial Linear Models for Clustered Data. *Journal of the American Statistical Association*, **100**, 1176-1184. https://doi.org/10.1198/016214505000000277

[19] Özdemir, S. and Arslan, O. (2015) An Alternative Algorithm of the Empirical Likelihood Estimation for the Parameter of a Linear Regression Model. *Communications in Statistics-Simulation and Computation*, **42**, 1-9.
https://doi.org/10.1080/03610918.2018.1435801