Statistics and Its Interface Volume 12 (2019) 617–630

Subject-wise empirical likelihood inference for robust joint mean-covariance model with longitudinal data

Jing Lv, Chaohui Guo*, and Jibo Wu

In longitudinal studies, one of the biggest challenges is how to obtain a good estimator of covariance matrix to improve the estimation efficiency of the mean regression coefficients. Meanwhile, one outlier in a subject level may generate multiple outliers in the sample due to repeated measurements. To solve these problems, this paper develops a robust joint mean–covariance model using the bounded exponential score function and modified Cholesky decomposition. The motivation for this new procedure is that it enables us to achieve high effectiveness and robustness simultaneously by introducing an additional tuning parameter γ which can be automatically selected using a data-driven procedure. In addition, we propose a subject-wise empirical likelihood to construct the confidence intervals/regions for the mean regression coefficients. Furthermore, under some mild conditions, we have established asymptotic theories of the proposed procedures. Finally, simulation studies are constructed to evaluate the finite sample performance of the proposed methods. A practical progesterone example is used to demonstrate the superiority of our proposed method.

AMS 2000 subject classifications: 62H12.
Keywords and phrases: Empirical likelihood, Exponential score function, Longitudinal data, Modified Cholesky decomposition, Robustness and effectiveness.

1. INTRODUCTION

Longitudinal data arises frequently in the biomedical, epidemiological, social, and economical fields. The generalized estimating equation (GEE) proposed by Liang and Zeger [11] is a popular approach to deal with longitudinal data. However, the GEE is in principle very similar to the weighted least squares method, which is not robust for non-normal distributions. In the longitudinal data analysis, it is well known that one outlier in the subject level may produce a set of outliers in the sample due to repeated measurements. Hence, robustness against outliers is a very important issue in longitudinal studies. Recently, a robust Huber's estimation for longitudinal data has attracted much attention. Wang et al. [30] proposed a distribution-free bias correction method for robust estimating equations. He et al. [7] constructed robust generalized estimating equations for longitudinal generalized partial linear models. Fan et al. [5] developed a robust variable selection procedure for longitudinal linear models. Other related literature can refer to Croux et al. [4], Qin and Zhu [19] and Qin et al. [20]. However, the mentioned references only considered the mean regression model with some specified correlation structures, which will result in a loss of efficiency when the true correlation structure is misspecified. Thus, it is essential to model the covariance structure to improve the estimation efficiency of mean regression coefficients. In recent years, many scholars focused on studying joint mean and covariance models by using a modified Cholesky decomposition which is a useful tool to parameterize the covariance matrix. There are two salient merits for this decomposition. On the one hand, it automatically guarantees the positive definiteness of the covariance matrix. On the other hand, the parameters of this decomposition are unconstrained and have well founded statistical concepts. Recently, Ye and Pan [32] developed a joint mean covariance model by utilizing the GEE method and modified Cholesky decomposition. In order to relax the parametric assumption, Guo et al. [6], Leng et al. [9], Mao et al. [13], Qin et al. [18] and Zheng et al. [35] studied joint semiparametric mean-covariance models for longitudinal data.

Although there have been a few research results on joint mean covariance models for longitudinal data, these works were built on either likelihood method or GEE method, which is very sensitive to outliers and many commonly using non-normal correlated errors. Thus it is practical interest to develop robust joint mean covariance approaches. Recently, Zheng et al. [33] proposed the robust joint mean–covariance regression model by combining the GEE and Huber's score function. Combined regularized method with Huber's score function, Zheng et al. [34] proposed three penalized robust generalized estimating equations to select significant variables both in the mean and covariance models. Lv et al. [12] developed an adaptive robust estimation method for bivariate longitudinal data and discussed the selection of the turning parameter c in Huber's score function to achieve better robustness and efficiency. Although the Huber's score function is a robust modeling tool, there is some disadvan-

*Corresponding author.

tages in terms of the estimation efficiency. Thus, this stimulates us to look for other bounded score functions to obtain better robustness and effectiveness. Wang et al. [29] proposed a new exponential squared loss for independent linear regression models and showed that their approach was near optimal and superior to some recently developed methods. A distinguishing characteristic of this new approach is that it introduces an additional tuning parameter to achieve both robustness and effectiveness for the resulting estimator. Wang and Lin [27] proposed a simultaneous model structure identification and variable selection method for partial linear varying coefficient models by utilizing the exponential squared loss. Song et al. [23] extended exponential squared loss to high dimensional single index varying coefficient models. However, as above discussed literature, this new robust loss function was only considered for independent data. In this paper, based on the modified Cholesky decomposition, we will develop joint mean covariance models for longitudinal data using the exponential score function to achieve robust and effective estimators.

Empirical likelihood (EL) introduced by Owen [16] is a nonparametric inference method based on likelihood ratio type statistics, and its main advantages are as follows. On the one hand, it does not involve the asymptotic covariance of the estimators. On the other hand, it does not impose prior constraints on the region shape, and the shape and orientation of confidence regions are determined completely by practical data. Recently, empirical likelihood based method had been widely used to analyze longitudinal data. For example, Bai et al. [1] proposed a weighted empirical likelihood inference and established the asymptotic distribution of the weighted empirical likelihood ratio. Li and Pan [10] proposed a new EL ratio function to deal with the within subject correlation without involving the estimation of nuisance parameters in the correlation matrix, which results in higher coverage probabilities and shorter confidence interval. To improve the robustness of parametric estimation, combined with the quadratic inference function (Qu et al. [22]) and empirical likelihood, Tang and Leng [24] constructed weighted quantile estimators by taking into account the within subject correlations. Based on Tang and Leng [24], Tang et al. [25] developed weighted composite quantile regression estimators. Other related references on empirical likelihood with longitudinal data include Qin et al. [17], Wang and Zhu [26] and Wang et al. [28]. However, the above mentioned articles only focused on some specific correlation structures, which led to a loss of efficiency when the true correlation structure is misspecified. This paper proposes a subject-wise empirical log-likelihood ratio function for the regression coefficients to improve the accuracy of interval estimation on the basis of the modified Cholesky decomposition and exponential score function.

The remainder of this paper is organized as follows. In Sect. 2, we apply the modified Cholesky decomposition and bounded exponential score function to construct three generalized estimating equations for the mean regression coefficients, autoregressive coefficients and innovation variances. Then, we investigate their theoretical properties and propose an efficient algorithm to implement the procedure. Furthermore, we discuss how to select the tuning parameter γ so that the resulting estimators are robust and efficient. In Sect. 3, we propose a subject-wise empirical likelihood ratio statistic and establish its asymptotic distribution. Furthermore, we construct the proper confidence regions and pointwise confidence intervals for the parameters and its components. In Sect. 4, we conduct extensive simulation studies to compare the finite sample performance of the proposed method with some existing methods. Sect. 5 applies the new method to a progesterone data set. Some concluding remarks are given in Sect. 6. The proofs of theorems are provided in the Appendix.

2. ROBUST JOINT MEAN-COVARIANCE MODEL

We consider the longitudinal linear model

$$y_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \varepsilon_{ij}, i = 1, ..., n, j = 1, ..., m_i, \tag{1}$$

where y_{ij} is the jth measurement on the ith subject, $\boldsymbol{x}_{ij}$ is p dimensional vector of covariates, $\boldsymbol{\beta}$ is p dimensional vector of parameters and ε_{ij} is random error. According to the characteristic of longitudinal data, we assume that $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, ..., \varepsilon_{im_i})^T$ are correlated in the same subject but independent across the subjects.

2.1 Estimating equations under the independent working model

For independent data, Wang et al. [29] proposed a new robust regression estimator based on the exponential squared loss and pointed that their proposed method is near optimal and superior to some recently developed methods. According to Wang et al. [29], under an independent working model, we can estimate the regression coefficient $\boldsymbol{\beta}$ by minimizing

$$Q_\gamma(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left\{1 - \varphi_\gamma\left(y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta}\right)\right\}, \tag{2}$$

where $\varphi_\gamma(t) = \exp\left(-t^2/\gamma\right)$, $\gamma > 0$ determines the degree of robustness of the estimation. If γ is large, we have $1 - \exp\left(-t^2/\gamma\right) \approx t^2/\gamma$. Thereby, the new estimators are similar to the least squares estimators. For observations with large absolute values of $t_{ij} = y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta}$, a smaller γ can be used to downweight the influence of an outlier on the estimators. More detailed discussions on the exponential squared loss can refer to Wang et al. [29], Wang and Lin [27] and Song et al. [23]. Obviously, minimizing the objective function (2) with respect to $\boldsymbol{\beta}$ is equivalent to solving the following estimating equations

$$\sum_{i=1}^{n} \boldsymbol{X}_i^T \psi_\gamma \left(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}\right) = 0, \tag{3}$$

where $\psi_\gamma(t) = \varphi'_\gamma(t) = -\frac{2t}{\gamma}\exp\left(-t^2/\gamma\right)$ is the exponential score function, $\varphi'_\gamma(t)$ is the first derivative of $\varphi_\gamma(t)$, $\boldsymbol{Y}_i = (y_{i1}, ..., y_{im_i})^T$, $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{im_i})^T$ and $\psi_\gamma(\boldsymbol{t}_i) = (\psi_\gamma(t_{i1}), ..., \psi_\gamma(t_{im_i}))^T$ for $\boldsymbol{t}_i = (t_{i1}, ..., t_{im_i})^T$. Note that $\psi_\gamma(t)$ is a bounded score function due to $\lim_{\gamma\to+\infty}\psi_\gamma(t) = 0$ and $\lim_{\gamma\to 0^+}\psi_\gamma(t) = 0$. Although we can obtain a consistent estimator $\bar{\boldsymbol{\beta}}_\gamma$ from (3) by ignoring the possible correlations between repeated measures, the efficiency of $\bar{\boldsymbol{\beta}}_\gamma$ may not be satisfactory. We will omit the subscript γ from $\bar{\boldsymbol{\beta}}_\gamma$ in the rest of this article for simplicity.

2.2 Estimating equations for joint mean and covariance model

Efficient parameter estimators could be obtained by incorporating an appropriate weighted function that accounts for the correlation and variation of repeated measurements for each subject. Based on the idea of GEE (Liang and Zeger [11]), we can use the estimating equations that take the form

$$\sum_{i=1}^{n} \boldsymbol{X}_i^T \boldsymbol{\Sigma}_{i\gamma}^{-1} \psi_\gamma \left(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}\right) = \boldsymbol{0}, \tag{4}$$

where $\boldsymbol{\Sigma}_{i\gamma} = Cov\left(\psi_\gamma\left(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}\right)\right)$. But we can not directly obtain the estimator of $\boldsymbol{\beta}$ by solving (4). The main reason is that the estimating equation (4) includes the unknown covariance matrix $\boldsymbol{\Sigma}_{i\gamma}$, which need to be determined by the observed data. To guarantee the positive definiteness of the matrices $\boldsymbol{\Sigma}_{i\gamma}(i = 1, ..., n)$, motivated by Ye and Pan [32], we apply the modified Cholesky decomposition to decompose $\boldsymbol{\Sigma}_{i\gamma}$,

$$\boldsymbol{\Phi}_{i\gamma}\boldsymbol{\Sigma}_{i\gamma}\boldsymbol{\Phi}_{i\gamma}^T = \boldsymbol{D}_{i\gamma}, \tag{5}$$

where $\boldsymbol{\Phi}_{i\gamma}$ is a unique unit lower triangular matrix with 1's on the diagonal and the below diagonal entries of $\boldsymbol{\Phi}_{i\gamma}$ are the negatives of the autoregressive coefficients $\phi_{\gamma,ijk}$ in the model

$$\psi_\gamma\left(y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta}\right) = \sum_{k=1}^{j-1} \phi_{\gamma,ijk}\psi_\gamma\left(y_{ik} - \boldsymbol{x}_{ik}^T\boldsymbol{\beta}\right) + e_{\gamma,ij}.$$

Note that when $j = 1$ the notation $\sum_{k=1}^{0}$ means zero throughout this paper. $\boldsymbol{D}_{i\gamma}$ is an $m_i \times m_i$ diagonal matrix with the jth diagonal element being $Var\left(e_{\gamma,ij}\right) = d_{\gamma,ij}^2$, which can be seen as the innovation variance for $j = 1, ..., m_i$. Similar to Ye and Pan [32], we adopt two generalized linear models for the autoregressive parameters and innovation variances

$$\phi_{\gamma,ijk} = \boldsymbol{w}_{ijk}^T\boldsymbol{\theta}_\gamma, \quad \log\left(d_{\gamma,ij}^2\right) = \boldsymbol{z}_{ij}^T\boldsymbol{\lambda}_\gamma, \tag{6}$$

where $\boldsymbol{\theta}_\gamma = (\theta_{\gamma,1}, ..., \theta_{\gamma,q})^T$ and $\boldsymbol{\lambda}_\gamma = (\lambda_{\gamma,1}, ..., \lambda_{\gamma,d})^T$. We should notice that $\boldsymbol{\theta}_\gamma$ and $\boldsymbol{\lambda}_\gamma$ are γ-specific since the covariance matrix $\boldsymbol{\Sigma}_{i\gamma}$ is related to γ, but we omit the subscript γ from $\boldsymbol{\theta}_\gamma$ and $\boldsymbol{\lambda}_\gamma$ in the rest of this article for simplicity. The covariates $\boldsymbol{z}_{ij}$ are those used in regression analysis, while $\boldsymbol{w}_{ijk}$ is usually taken as a polynomial of time difference $t_{ij} - t_{ik}$. A common choice for $\boldsymbol{w}_{ijk}$ and $\boldsymbol{z}_{ij}$ is $\boldsymbol{w}_{ijk} = \left(1, t_{ij} - t_{ik}, ..., (t_{ij} - t_{ik})^{q-1}\right)^T$ and $\boldsymbol{z}_{ij} = \left(1, t_{ij}, ..., t_{ij}^{d-1}\right)^T$.

Remark 2.1. *This modified Cholesky decomposition approach can guarantee the positive definiteness of $\boldsymbol{\Sigma}_{i\gamma}$, and the below diagonal elements of $\boldsymbol{\Phi}_{i\gamma}$ are unconstrained. To estimate the autoregressive parameters $\phi_{\gamma,ijk}$ and innovation variances $d_{\gamma,ij}^2$ in $\boldsymbol{\Phi}_{i\gamma}$ and $\boldsymbol{D}_{i\gamma}$, we adopt two generalized linear models (6). Of course, other regression models also can be used to estimate $\phi_{\gamma,ijk}$ and $d_{\gamma,ij}^2$, for example, semiparametric regression models (Leng et al. [9]). But linear models are simple and popular regression tools, and thus linear models are considered here.*

Now we propose three generalized estimating equations for the mean, autoregressive parameters and innovation variances as follows:

$$U_1(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{X}_i^T \boldsymbol{\Sigma}_{i\gamma}^{-1} \psi_\gamma\left(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}\right) = \boldsymbol{0}, \tag{7}$$

$$U_2(\boldsymbol{\theta}) = \sum_{i=1}^{n} \boldsymbol{T}_{i\gamma}^T \boldsymbol{D}_{i\gamma}^{-1} \boldsymbol{e}_{i\gamma} = \boldsymbol{0}, \tag{8}$$

$$U_3(\boldsymbol{\lambda}) = \sum_{i=1}^{n} \boldsymbol{Z}_i^T \boldsymbol{D}_{i\gamma} \boldsymbol{W}_{i\gamma}^{-1}\left(\boldsymbol{e}_{i\gamma}^2 - \boldsymbol{d}_{i\gamma}^2\right) = \boldsymbol{0}, \tag{9}$$

where $\boldsymbol{e}_{i\gamma} = (e_{\gamma,i1}, ..., e_{\gamma,im_i})^T$ with $e_{\gamma,ij} = \psi_\gamma\left(y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta}\right) - \sum_{k=1}^{j-1}\phi_{\gamma,ijk}\psi_\gamma\left(y_{ik} - \boldsymbol{x}_{ik}^T\boldsymbol{\beta}\right)$, $\boldsymbol{T}_{i\gamma}^T = -\partial\boldsymbol{e}_{i\gamma}^T/\partial\boldsymbol{\theta}$ is a $q \times m_i$ matrix with the first column zero and the jth $(j \geq 2)$ column $-\partial e_{\gamma,ij}/\partial\boldsymbol{\theta} = \sum_{k=1}^{j-1}\boldsymbol{w}_{ijk}\psi_\gamma\left(y_{ik} - \boldsymbol{x}_{ik}^T\boldsymbol{\beta}\right)$, $\boldsymbol{Z}_i = (\boldsymbol{z}_{i1}, ..., \boldsymbol{z}_{im_i})^T$, $\boldsymbol{d}_{i\gamma}^2 = \left(d_{\gamma,i1}^2, ..., d_{\gamma,im_i}^2\right)^T$ and $\boldsymbol{W}_{i\gamma} = Cov\left(\boldsymbol{e}_{i\gamma}^2\right)$. A sandwich "working" covariance structure $\tilde{\boldsymbol{W}}_{i\gamma} = \boldsymbol{A}_{i\gamma}^{1/2}\boldsymbol{R}_{i\gamma}(\varrho)\boldsymbol{A}_{i\gamma}^{1/2}$ can be used to approximate the true $\boldsymbol{W}_{i\gamma}$, where $\boldsymbol{A}_{i\gamma} = 2diag\left(d_{\gamma,i1}^4, ..., d_{\gamma,im_i}^4\right)$ and $\boldsymbol{R}_{i\gamma}(\varrho)$ stands for the correlation between $e_{\gamma,ij}^2$ and $e_{\gamma,ik}^2$ $(j \neq k)$ by introducing a parameter ϱ. Typical structures for $\boldsymbol{R}_{i\gamma}(\varrho)$ include the compound symmetry and AR(1). Ye and Pan [32] pointed that the parameter ϱ has little effect on the estimators of $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$. So we take $\varrho = 0$ in our simulations and real data analysis. Based on the discussions above together with (5), we can obtain $\hat{\boldsymbol{\Sigma}}_{i\gamma} = \hat{\boldsymbol{\Phi}}_{i\gamma}^{-1}\hat{\boldsymbol{D}}_{i\gamma}(\hat{\boldsymbol{\Phi}}_{i\gamma}^T)^{-1}$, where $\hat{\boldsymbol{\Phi}}_{i\gamma}$ is an $m_i \times m_i$ lower triangular matrix with 1's on its diagonal and the below diagonal entries of $\hat{\boldsymbol{\Phi}}_{i\gamma}$ are

$-\hat{\phi}_{\gamma,ijk} = -\boldsymbol{w}_{ijk}^T\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{D}}_{i\gamma} = diag\left(\hat{d}^2_{\gamma,i1}, ..., \hat{d}^2_{\gamma,im_i}\right)$ with $\hat{d}^2_{\gamma,ij} = \exp(\boldsymbol{z}_{ij}^T\hat{\boldsymbol{\lambda}})$. Suppose that $\hat{\boldsymbol{\eta}} = \left(\hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{\theta}}^T, \hat{\boldsymbol{\lambda}}^T\right)^T$ is the root of the generalized estimating equations (7)–(9). Please note that $\hat{\boldsymbol{\eta}}$ may depend on γ but we omit the subscript for simplicity.

2.3 Asymptotic properties

Theorem 2.1. *Under the conditions stated in the Appendix, the proposed estimator* $\hat{\boldsymbol{\eta}} = \left(\hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{\theta}}^T, \hat{\boldsymbol{\lambda}}^T\right)^T$ *is strongly consistent for the true value* $\boldsymbol{\eta}_0 = \left(\boldsymbol{\beta}_0^T, \boldsymbol{\theta}_0^T, \boldsymbol{\lambda}_0^T\right)^T$, *that is,* $\hat{\boldsymbol{\eta}} \to \boldsymbol{\eta}_0$ *almost surely as* $n \to \infty$.

Now we need to calculate the covariance matrix of $\left(\boldsymbol{U}_1(\boldsymbol{\beta}_0)^T, \boldsymbol{U}_2(\boldsymbol{\theta}_0)^T, \boldsymbol{U}_3(\boldsymbol{\lambda}_0)^T\right)^T \Big/ \sqrt{n}$, denoted by $\boldsymbol{V}_n = \left(\boldsymbol{v}_n^{kl}\right)_{k,l=1,2,3}$ to prove the asymptotic normality of $\hat{\boldsymbol{\eta}}$, where $\boldsymbol{v}_n^{kl} = n^{-1}Cov\left(\boldsymbol{U}_k, \boldsymbol{U}_l\right)$ for $k \neq l$ and $\boldsymbol{v}_n^{kk} = n^{-1}Var\left(\boldsymbol{U}_k\right)$, for $k, l = 1, 2, 3$. We further assume that the covariance matrix $\boldsymbol{V}_n$ and $\boldsymbol{c}_n^{11} = -n^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i^T\boldsymbol{\Sigma}_{i\gamma}^{-1}\boldsymbol{\Lambda}_{i\gamma}\boldsymbol{X}_i$ are positive definite at the true value $\boldsymbol{\eta}_0$ and

(10)
$$\boldsymbol{V}_n = \begin{pmatrix} \boldsymbol{v}_n^{11} & \boldsymbol{v}_n^{12} & \boldsymbol{v}_n^{13} \\ \boldsymbol{v}_n^{21} & \boldsymbol{v}_n^{22} & \boldsymbol{v}_n^{23} \\ \boldsymbol{v}_n^{31} & \boldsymbol{v}_n^{32} & \boldsymbol{v}_n^{33} \end{pmatrix} \xrightarrow{p} \boldsymbol{V} = \begin{pmatrix} \boldsymbol{v}^{11} & \boldsymbol{v}^{12} & \boldsymbol{v}^{13} \\ \boldsymbol{v}^{21} & \boldsymbol{v}^{22} & \boldsymbol{v}^{23} \\ \boldsymbol{v}^{31} & \boldsymbol{v}^{32} & \boldsymbol{v}^{33} \end{pmatrix}$$
$$\text{and} \quad \boldsymbol{c}_n^{11} \xrightarrow{p} \boldsymbol{c}^{11}$$

as $n \to \infty$, where $\boldsymbol{\Lambda}_{i\gamma} = diag\left\{\Lambda_{\gamma,i1}, ..., \Lambda_{\gamma,im_i}\right\}$ with $\Lambda_{\gamma,ij} = \exp\left(-\frac{\varepsilon_{ij}^2}{\gamma}\right)\left\{\left[\frac{2\varepsilon_{ij}}{\gamma}\right]^2 - \frac{2}{\gamma}\right\}$ and $\varepsilon_{ij} = y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta}$. The constant matrices $\boldsymbol{V}$ and $\boldsymbol{c}^{11}$ in (10) are also assumed to be positive definite.

Theorem 2.2. *Suppose that (10) above is true. Under the conditions stated in the Appendix, the proposed estimator* $\hat{\boldsymbol{\eta}}$ *is asymptotically normally distributed with*

$$\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \\ \hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0 \end{pmatrix} \xrightarrow{d} N\left(\boldsymbol{0}, \boldsymbol{A}^{-1}\boldsymbol{V}\boldsymbol{A}^{-1}\right)$$

as $n \to \infty$, *where* $\boldsymbol{A} = diag\left(\boldsymbol{c}^{11}, \boldsymbol{v}^{22}, \boldsymbol{v}^{33}\right)$, *the matrices* $\boldsymbol{v}^{kl}(k, l = 1, 2, 3)$ *and* $\boldsymbol{c}^{11}$ *in* $\boldsymbol{V}$ *and* $\boldsymbol{A}$ *are evaluated at* $\boldsymbol{\eta}_0$ *and* $\xrightarrow{d}$ *represents convergence in distribution.*

2.4 Algorithm and the choice of tuning parameter γ

This paper uses a quasi–Fisher scoring algorithm to solve $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ iteratively. We assume the starting values of $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ to be $\boldsymbol{\theta}^{(0)} = \boldsymbol{0}$ and $\boldsymbol{\lambda}^{(0)} = \boldsymbol{0}$, so we obtain $\boldsymbol{\Sigma}_{i\gamma}^{(0)} = \boldsymbol{I}_{m_i \times m_i}$ based on (5) and (6). Hence, an initial estimate $\boldsymbol{\beta}^{(0)}$ of $\boldsymbol{\beta}$ is the solution of (7) under the independent working covariance structure.

Given $\boldsymbol{\Sigma}_{i\gamma}$, we solve (7) to find the estimate of $\boldsymbol{\beta}$ using the iterative procedure

(11)
$$\boldsymbol{\beta}^{(k+1)} = \boldsymbol{\beta}^{(k)} + \left\{\left[\sum_{i=1}^{n}\boldsymbol{X}_i^T\boldsymbol{\Sigma}_{i\gamma}^{-1}\boldsymbol{\Lambda}_{i\gamma}\boldsymbol{X}_i\right]^{-1} \times \sum_{i=1}^{n}\boldsymbol{X}_i^T\boldsymbol{\Sigma}_{i\gamma}^{-1}\psi_\gamma\left(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}\right)\right\}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(k)}}.$$

Given $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$, $\boldsymbol{\theta}$ can be updated through

(12)
$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \left\{\left[\sum_{i=1}^{n}\boldsymbol{T}_{i\gamma}^T\boldsymbol{D}_{i\gamma}^{-1}\boldsymbol{T}_{i\gamma}\right]^{-1} \times \sum_{i=1}^{n}\boldsymbol{T}_{i\gamma}^T\boldsymbol{D}_{i\gamma}^{-1}\boldsymbol{e}_{i\gamma}\right\}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(k)}}.$$

Finally, given $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, $\boldsymbol{\lambda}$ can be updated through

(13)
$$\boldsymbol{\lambda}^{(k+1)} = \boldsymbol{\lambda}^{(k)} + \left\{\left[\sum_{i=1}^{n}\boldsymbol{Z}_i^T\boldsymbol{D}_{i\gamma}\boldsymbol{W}_{i\gamma}^{-1}\boldsymbol{D}_{i\gamma}\boldsymbol{Z}_i\right]^{-1} \times \sum_{i=1}^{n}\boldsymbol{Z}_i^T\boldsymbol{D}_{i\gamma}\boldsymbol{W}_{i\gamma}^{-1}\left(\boldsymbol{e}_{i\gamma}^2 - \boldsymbol{d}_{i\gamma}^2\right)\right\}\Big|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^{(k)}}.$$

In summary, the main iterative algorithm is as follows:

Step 1. Given a starting value $\left(\boldsymbol{\beta}^{(0)T}, \boldsymbol{\theta}^{(0)T}, \boldsymbol{\lambda}^{(0)T}\right)^T$, we use model (6) to form $\boldsymbol{\Phi}_{i\gamma}^{(0)}$ and $\boldsymbol{D}_{i\gamma}^{(0)}$. Then $\boldsymbol{\Sigma}_{i\gamma}^{(0)}$, the starting value of $\boldsymbol{\Sigma}_{i\gamma}$ is obtained by (5). Set $k = 0$.

Step 2. Using the iterative formulas (11)–(13) to calculate the estimators $\boldsymbol{\beta}^{(k+1)}$, $\boldsymbol{\theta}^{(k+1)}$ and $\boldsymbol{\lambda}^{(k+1)}$ of the parameters $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ respectively. Furthermore, based on (5) and (6), we obtain $\boldsymbol{\Sigma}_{i\gamma}^{(k+1)}$.

Step 3. Set $k \leftarrow k + 1$. Repeat Step 2 until convergence of the parameter estimators.

The tuning parameter γ controls the degree of robustness and efficiency of the proposed estimator $\hat{\boldsymbol{\beta}}$. Here we propose a data–driven procedure to select γ, which achieves high robustness and effectiveness. Motivated by Wang et al. [29], we apply the grid search method to obtain the optimal γ_{opt} by minimizing $det(\widehat{Cov}(\hat{\boldsymbol{\beta}}))$, where $det(\cdot)$ denotes the determinant operator and

(14)
$$\widehat{Cov}\left(\hat{\boldsymbol{\beta}}\right) = \hat{\boldsymbol{M}}_{\gamma 0}^{-1}\hat{\boldsymbol{M}}_{\gamma 1}\hat{\boldsymbol{M}}_{\gamma 0}^{-1},$$

where

$$\hat{\boldsymbol{M}}_{\gamma 0} = \sum_{i=1}^{n}\boldsymbol{X}_i^T\hat{\boldsymbol{\Sigma}}_{i\gamma}^{-1}\boldsymbol{\Lambda}_{i\gamma}\boldsymbol{X}_i\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}$$

and

$$\hat{\boldsymbol{M}}_{\gamma 1} = \sum_{i=1}^{n}\boldsymbol{X}_i^T\hat{\boldsymbol{\Sigma}}_{i\gamma}^{-1}\psi_\gamma\left(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}\right) \times \left\{\psi_\gamma\left(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}\right)\right\}^T\hat{\boldsymbol{\Sigma}}_{i\gamma}^{-1}\boldsymbol{X}_i\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}.$$

Remark 2.2. *Our numerical experiences indicate that this iterative algorithm converges very quickly, usually in a few iterations. In addition, the optimal* γ *is obtained by minimizing the determinant of estimated covariance matrix of*

$\hat{\boldsymbol{\beta}}$*, which may guarantee the resulting estimator attains high efficiency. In our simulations and real data analysis, we use a grid search method to obtain the optimal tuning parameter* γ_{opt}*. Based on our empirical experience, the possible grids points for* γ *is considered as an arithmetic sequence from 2 to 50 with the common difference being two.*

3. EMPIRICAL LIKELIHOOD INFERENCE

Confidence region construction is also an important aspect for statistical inference. Although we can construct confidence regions of $\boldsymbol{\beta}$ based on the sandwich formula (14), the conventional normal based approximation method and direct estimation of the covariance matrix are unstable under the finite samples. This paper employs an empirical likelihood approach to construct the confidence intervals/regions of the regression coefficients, which does not need to estimate the unknown covariance matrix of $\boldsymbol{\beta}$. Compared with the traditional normal based approximation method, the main advantage of empirical likelihood approach is that the shape and orientation of the confidence regions are automatically determined by the data.

To construct the empirical likelihood ratio function for $\boldsymbol{\beta}$, we refer to Owen [16] and Qin and Lawless [21] about empirical likelihood and estimating equations. Based on estimated covariance matrix $\hat{\boldsymbol{\Sigma}}_{i\gamma}$, we construct an auxiliary random vector $\boldsymbol{\xi}_i(\boldsymbol{\beta}) = \boldsymbol{X}_i^T \hat{\boldsymbol{\Sigma}}_{i\gamma}^{-1} \psi_\gamma (\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})$. Note that $E(\boldsymbol{\xi}_i(\boldsymbol{\beta})) = 0$ if $\boldsymbol{\beta}$ is the true parameter. Let $p_1, ..., p_n$ be nonnegative numbers satisfying $\sum_{i=1}^n p_i = 1$. Using such information, a natural subject–wise empirical log–likelihood ratio for $\boldsymbol{\beta}$ is defined as

$$l(\boldsymbol{\beta}) = -2\max\left\{\sum_{i=1}^n \log(np_i)\middle| p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \boldsymbol{\xi}_i(\boldsymbol{\beta}) = \mathbf{0}\right\}. \tag{15}$$

A unique value for $l(\boldsymbol{\beta})$ exists for a given $\boldsymbol{\beta}$, provided that 0 is inside the convex hull of the points $\{\boldsymbol{\xi}_i(\boldsymbol{\beta}), i = 1, ..., n\}$. By the Lagrange multiplier method, the optimal value for p_i is given by

$$p_i(\boldsymbol{\beta}) = n^{-1}\left\{1 + \boldsymbol{\rho}^T \boldsymbol{\xi}_i(\boldsymbol{\beta})\right\}^{-1}, \tag{16}$$

where $\boldsymbol{\rho}$ is a p-dimensional Lagrange multiplier satisfying

$$n^{-1}\sum_{i=1}^n \frac{\boldsymbol{\xi}_i(\boldsymbol{\beta})}{1 + \boldsymbol{\rho}^T \boldsymbol{\xi}_i(\boldsymbol{\beta})} = \mathbf{0}. \tag{17}$$

By (15) and (16), $l(\boldsymbol{\beta})$ can be represented as

$$l(\boldsymbol{\beta}) = 2\sum_{i=1}^n \log\left\{1 + \boldsymbol{\rho}^T \boldsymbol{\xi}_i(\boldsymbol{\beta})\right\} \tag{18}$$

with $\boldsymbol{\rho}$ satisfying (17). In our numerical studies, we solve (17) for $\boldsymbol{\rho}$ by employing the modified Newton–Raphson algorithm (Chen et al. [3]). We define the maximum empirical likelihood (EL) estimator of $\boldsymbol{\beta}$ as $\hat{\boldsymbol{\beta}}_{el} = \arg\min_{\boldsymbol{\beta}} l(\boldsymbol{\beta})$. Here we should realize that $\hat{\boldsymbol{\beta}}_{el}$ also relies on the tuning parameter γ. Based on Xue and Zhu [31] and Wang and Zhu [26], we can adopt similar strategy to prove that $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_{el}$ are asymptotically equivalent for point estimation. To decrease the computational burden, we adopt the same optimal tuning parameter γ_{opt} as that in $\hat{\boldsymbol{\beta}}$ when constructing the empirical log-likelihood ratio function. Now we state the asymptotic properties of the empirical likelihood ratio.

Theorem 3.1. *Suppose that the regularity conditions in the Appendix hold, if* $\boldsymbol{\beta}_0$ *is the true parameter, then* $l(\boldsymbol{\beta}_0) \xrightarrow{d} \chi^2(p)$*, where* $\chi^2(p)$ *means the chi-square distribution with* p *degrees of freedom.*

Remark 3.1. *Let* $\chi^2_{1-\alpha}(p)$ *be the* $(1-\alpha)$*th quantile of the* $\chi^2(p)$ *for* $0 < \alpha < 1$*. By Theorem 3.1, an approximate* $(1-\alpha)$ *confidence region for* $\boldsymbol{\beta}$ *is defined by* $l_0 = \left\{\boldsymbol{\beta} : l(\boldsymbol{\beta}) \leq \chi^2_{1-\alpha}(p)\right\}$*. Theorem 3.1 can also be used to test the hypothesis* $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$*. One could reject* H_0 *at level* α *if* $l(\boldsymbol{\beta}_0) > \chi^2_{1-\alpha}(p)$*.*

If we are interested in a subset of the regression coefficients $\boldsymbol{\beta}$, then profile empirical log likelihood ratio test statistic can be applied to achieve this goal. To be more specific, we assume $\boldsymbol{\beta}_0 = \left(\boldsymbol{\beta}_0^{(1)^T}, \boldsymbol{\beta}_0^{(2)^T}\right)^T$, where $\boldsymbol{\beta}_0^{(1)}$ and $\boldsymbol{\beta}_0^{(2)}$ are $p_1 \times 1$ and $(p - p_1) \times 1$ vectors, respectively. If we are interested in testing $H_0 : \boldsymbol{\beta}_0^{(1)} = \boldsymbol{b}_0$, where $\boldsymbol{b}_0$ is some known $p_1 \times 1$ vector. Then the profile log likelihood ratio test statistic is defined as

$$\bar{l}(\boldsymbol{b}_0) = l\left(\boldsymbol{b}_0, \tilde{\boldsymbol{\beta}}^{(2)}\right) - l\left(\hat{\boldsymbol{\beta}}^{(1)}, \hat{\boldsymbol{\beta}}^{(2)}\right),$$

where $\tilde{\boldsymbol{\beta}}^{(2)}$ minimizes $l\left(\boldsymbol{b}_0, \boldsymbol{\beta}^{(2)}\right)$ with respect to $\boldsymbol{\beta}^{(2)}$ and $\left(\hat{\boldsymbol{\beta}}^{(1)T}, \hat{\boldsymbol{\beta}}^{(2)T}\right)^T$ is EL estimator.

Corollary 3.1. *Under the same conditions as Theorem 3.1 and* $H_0 : \boldsymbol{\beta}_0^{(1)} = \boldsymbol{b}_0$*, we have* $\bar{l}(\boldsymbol{b}_0) \xrightarrow{d} \chi^2_{p_1}$*.*

Remark 3.2. *Corollary 3.1 not only can be used to test the hypothesis* $H_0 : \boldsymbol{\beta}_0^{(1)} = \boldsymbol{b}_0$ *for some known* $\boldsymbol{b}_0$ *but also can construct the confidence interval/region for* $\boldsymbol{\beta}^{(1)}$*,* $\left\{\boldsymbol{\beta}^{(1)} : \bar{l}\left(\boldsymbol{\beta}^{(1)}\right) \leq \chi^2_{1-\alpha}(p_1)\right\}$*.*

4. SIMULATION STUDIES

To investigate the finite sample performance of the proposed method, we carry out simulation studies.

Example 1. The data is generated from model (1), where $\boldsymbol{\beta} = (1, 0.5)^T$, $\boldsymbol{x}_{ij} = (x_{ij1}, x_{ij2})^T$ follows a multivariate normal distribution $N(0, \boldsymbol{\Sigma}_{\boldsymbol{x}})$ with $(\boldsymbol{\Sigma}_{\boldsymbol{x}})_{k,l} = 0.5^{|k-l|}$ for $1 \leq k, l \leq 2$. Each subject is supposed to be measured by m_i times with $m_i \sim Binomial(11, 0.8) + 1$, which leads to different numbers of repeated measurements for each subjects. In order to assess the robustness of the proposed method, we consider the following three cases for the random error $\boldsymbol{\varepsilon}_i$.

case 1 Correlated normal error, $\boldsymbol{\varepsilon}_i$ follows a multivariate normal distribution $N(\mathbf{0}, \Xi_i)$, where Ξ_i will be listed later.
case 2 Correlated normal error with outliers, $\boldsymbol{\varepsilon}_i$ follows the same multivariate normal distribution as that in case 1, and we randomly choose 2% of y_{ij} to be $y_{ij}+5$ and 2% of y_{ij} to be $y_{ij}-5$ simultaneously.
case 3 Correlated t-distribution with outliers, $\boldsymbol{\varepsilon}_i$ follows a multivariate t-distribution with the degree 3 and covariance matrix Ξ_i, where Ξ_i is the same as that in case 1. Meanwhile, we adopt the same strategy as that in case 2 to generate some outliers.

Let the covariance matrix Ξ_i of $\boldsymbol{\varepsilon}_i$ be $\Xi_i = \boldsymbol{\Delta}_i^{-1}\boldsymbol{B}_i\left(\boldsymbol{\Delta}_i^T\right)^{-1}$, where $\boldsymbol{B}_i$ is an $m_i \times m_i$ diagonal matrix with the jth element $\exp\left(-0.5+0.2u_{ij}\right)$, $u_{ij} \sim N(0,1)$, and $\boldsymbol{\Delta}_i$ is a unit lower triangular matrix with (j,k) element $-\delta_{ijk}$ $(k<j)$, $\delta_{ijk} = 0.2+0.3(t_{ij}-t_{ik})$. For the covariates $\boldsymbol{z}_{ij}$ and $\boldsymbol{w}_{ijk}$ in covariance model (6), we take $\boldsymbol{z}_{ij} = (1, z_{ij2}, z_{ij3}, z_{ij4})^T$ with $z_{ij2}, z_{ij3}, z_{ij4} \sim N(0,1)$ and $\boldsymbol{w}_{ijk} = \left\{1, t_{ij}-t_{ik}, (t_{ij}-t_{ik})^2, (t_{ij}-t_{ik})^3\right\}^T$ with $t_{ij} \sim U(0,1)$.

Example 2. We set $\boldsymbol{\beta} = (0.5, -0.5)^T$ and the covariance matrix as $\Xi_i = \boldsymbol{D}_i^{1/2}\boldsymbol{R}(\rho)\boldsymbol{D}_i^{1/2}$, where $\boldsymbol{D}_i = diag\left(\sigma_{i1}^2, ..., \sigma_{im_i}^2\right)$ with $\sigma_{ij}^2 = \exp\left(\boldsymbol{z}_{ij}^T\boldsymbol{\lambda}\right)$ and $\boldsymbol{\lambda} = (-0.5, 0.2, 0, 0)^T$, and $\boldsymbol{R}(\rho)$ is AR(1) or compound symmetry structure with correlation coefficient $\rho = 0.85$. Other settings are the same as that in example 1.

We compare the proposed estimators $\hat{\boldsymbol{\beta}}$ defined in subsection 2.2 (denoted as $\hat{\boldsymbol{\beta}}_{pr}$) and $\hat{\boldsymbol{\beta}}_{el}$ presented in section 3 with other existing four types of estimators. (i) The conventional least squares estimator without considering correlations, denoted as $\hat{\boldsymbol{\beta}}_{ls}$. (ii) The estimator proposed by Ye and Pan [32], denoted as $\hat{\boldsymbol{\beta}}_{ye}$. (iii) The estimator proposed by Zheng et al. [33] using robust Huber's score function $\psi_c(x) = \min\{c, \max(-c, x)\}$ with constant $c=2$, denoted as $\hat{\boldsymbol{\beta}}_{rb}$. (iv) The estimator obtained from (3) under the independent working model, denoted as $\hat{\boldsymbol{\beta}}_{in}$. Tables 1 and 2 give the bias, the sample standard deviation (SD), the mean absolute deviation (MAD) and the model error (ME) based on 200 replications, where MAD and ME are defined by MAD $= \left|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right|$, ME $= \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(\boldsymbol{x}_{ij}^T\hat{\boldsymbol{\beta}} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta}_0\right)^2$ and $N = \sum_{i=1}^{n} m_i$.

Eyeballing Tables 1 and 2, we can derive the following several observations. Firstly, all estimators are unbiased due to small biases, indicating that they are consistent estimators. Secondly, under case 1, our proposed estimator $\hat{\boldsymbol{\beta}}_{pr}$ performs equally as well as $\hat{\boldsymbol{\beta}}_{ye}$ in terms of SD, MAD and ME. This result indicates that $\hat{\boldsymbol{\beta}}_{pr}$ performs no worse than GEE estimator $\hat{\boldsymbol{\beta}}_{ye}$ under the correlated normal error, since we apply a data–driven procedure to obtain the optimal tuning parameter γ_{opt} that can guarantee the proposed estimation $\hat{\boldsymbol{\beta}}_{pr}$ to attain high efficiency. Meanwhile,

Table 1. Simulation results ($\times 10^{-2}$) of the bias, SD, MAD and ME for $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$ with $n = 100$ for example 1.

	method	β_1 bias	β_1 SD	β_1 MAD	β_2 bias	β_2 SD	β_2 MAD	ME
case 1	$\hat{\boldsymbol{\beta}}_{ls}$	0.061	4.166	3.297	0.021	3.867	3.085	0.232
	$\hat{\boldsymbol{\beta}}_{ye}$	0.120	2.543	2.074	0.136	2.440	1.955	0.097
	$\hat{\boldsymbol{\beta}}_{rb}$	0.101	2.617	2.103	0.178	2.499	2.017	0.101
	$\hat{\boldsymbol{\beta}}_{in}$	0.105	3.868	3.063	0.020	3.634	2.913	0.203
	$\hat{\boldsymbol{\beta}}_{pr}$	0.093	2.579	2.092	0.119	2.461	1.992	0.099
	$\hat{\boldsymbol{\beta}}_{el}$	0.093	2.579	2.093	0.119	2.462	1.992	0.099
case 2	$\hat{\boldsymbol{\beta}}_{ls}$	0.180	4.483	3.602	0.194	5.078	3.987	0.353
	$\hat{\boldsymbol{\beta}}_{ye}$	0.003	4.350	3.532	-0.002	4.343	3.423	0.294
	$\hat{\boldsymbol{\beta}}_{rb}$	0.174	3.112	2.552	-0.182	3.079	2.395	0.141
	$\hat{\boldsymbol{\beta}}_{in}$	0.419	3.602	2.976	-0.290	3.628	2.866	0.191
	$\hat{\boldsymbol{\beta}}_{pr}$	0.330	2.913	2.318	-0.330	2.991	2.359	0.127
	$\hat{\boldsymbol{\beta}}_{el}$	0.331	2.913	2.318	-0.331	2.991	2.359	0.127
case 3	$\hat{\boldsymbol{\beta}}_{ls}$	0.569	6.953	5.594	-1.197	7.403	6.027	0.775
	$\hat{\boldsymbol{\beta}}_{ye}$	0.848	5.952	4.710	-0.965	5.474	4.541	0.473
	$\hat{\boldsymbol{\beta}}_{rb}$	0.475	4.325	3.379	-0.692	4.085	3.411	0.256
	$\hat{\boldsymbol{\beta}}_{in}$	0.051	4.507	3.620	-0.275	4.188	3.398	0.282
	$\hat{\boldsymbol{\beta}}_{pr}$	0.262	3.834	3.055	-0.402	3.958	3.173	0.227
	$\hat{\boldsymbol{\beta}}_{el}$	0.260	3.833	3.056	-0.401	3.959	3.173	0.227

$\hat{\boldsymbol{\beta}}_{pr}$ performs slightly better than $\hat{\boldsymbol{\beta}}_{rb}$ in terms of SD, MAD and ME. Thirdly, $\hat{\boldsymbol{\beta}}_{pr}$ and $\hat{\boldsymbol{\beta}}_{ye}$ apparently outperform $\hat{\boldsymbol{\beta}}_{in}$ and $\hat{\boldsymbol{\beta}}_{ls}$, respectively, which indicates that it is necessary to take account of the within correlations for longitudinal data. Fourthly, $\hat{\boldsymbol{\beta}}_{ls}$ and $\hat{\boldsymbol{\beta}}_{ye}$ perform worse when data set contains outliers, which means that they are not robust approaches. Fifthly, $\hat{\boldsymbol{\beta}}_{pr}$ and $\hat{\boldsymbol{\beta}}_{el}$ have similar performances in terms of bias, SD, MAD and ME, indicating that $\hat{\boldsymbol{\beta}}_{pr}$ and $\hat{\boldsymbol{\beta}}_{el}$ are asymptotically equivalent for point estimation. Finally, the proposed $\hat{\boldsymbol{\beta}}_{pr}$ and $\hat{\boldsymbol{\beta}}_{el}$ perform best among all methods for cases 2 and 3. The main reasons include two aspects. On the one hand, we apply the modified Cholesky decomposition to deal with the within correlation. On the other hand, we select the optimal tuning parameter γ_{opt} by minimizing the determinant of estimated covariance matrix, which yields both high robustness and high efficiency simultaneously. Thus, we can conclude that the proposed estimation approach can achieve better robustness and efficiency than the existing approaches, especially for the non–normal error distributions.

In addition, the means and standard errors (in parentheses) of the selected averaged optimal parameter γ_{opt} are given in Table 3 for example 1. We can see that the optimal tuning parameter γ_{opt} is smaller for the contamination data as a small γ leads to a greater resistance to outliers, which is consistent with the theory. Furthermore, it is easy to observe that the selected optimal parameter γ_{opt} tends to be stable with the sample size n increasing due to decreasing standard deviations.

Table 2. Simulation results ($\times 10^{-2}$) of the bias, SD, MAD and ME for $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$ with $n = 100$ for example 2.

		cs							ar1						
	method	β_1			β_2				β_1			β_2			
		bias	SD	MAD	bias	SD	MAD	ME	bias	SD	MAD	bias	SD	MAD	ME
case 1	$\hat{\boldsymbol{\beta}}_{ls}$	0.332	2.789	2.265	-0.029	2.836	2.217	0.121	0.120	3.207	2.461	-0.071	2.836	2.349	0.137
	$\hat{\boldsymbol{\beta}}_{ye}$	0.077	1.629	1.328	0.107	1.681	1.330	0.042	0.086	2.323	1.873	0.069	2.197	1.793	0.077
	$\hat{\boldsymbol{\beta}}_{rb}$	0.077	1.633	1.332	0.044	1.609	1.269	0.040	0.130	2.359	1.909	0.016	2.274	1.839	0.081
	$\hat{\boldsymbol{\beta}}_{in}$	0.329	2.794	2.280	-0.039	2.848	2.239	0.122	0.097	3.161	2.428	-0.045	2.878	2.378	0.137
	$\hat{\boldsymbol{\beta}}_{pr}$	0.075	1.602	1.309	0.098	1.651	1.311	0.041	0.100	2.327	1.881	0.051	2.213	1.802	0.078
	$\hat{\boldsymbol{\beta}}_{el}$	0.075	1.602	1.309	0.098	1.650	1.311	0.041	0.100	2.327	1.881	0.051	2.213	1.802	0.078
case 2	$\hat{\boldsymbol{\beta}}_{ls}$	0.237	4.736	3.781	-0.183	4.505	3.597	0.317	-0.309	4.805	3.857	0.173	4.213	3.284	0.340
	$\hat{\boldsymbol{\beta}}_{ye}$	0.107	4.006	3.172	0.011	3.913	3.099	0.232	-0.296	4.306	3.397	0.013	3.875	3.078	0.282
	$\hat{\boldsymbol{\beta}}_{rb}$	0.171	2.062	1.603	-0.139	1.922	1.523	0.060	-0.145	2.520	2.039	0.071	2.362	1.852	0.103
	$\hat{\boldsymbol{\beta}}_{in}$	0.184	3.153	2.611	-0.235	3.169	2.538	0.152	-0.176	3.035	2.447	0.166	2.851	2.251	0.142
	$\hat{\boldsymbol{\beta}}_{pr}$	0.089	1.827	1.428	-0.107	1.703	1.363	0.048	-0.107	2.315	1.850	0.118	2.282	1.776	0.087
	$\hat{\boldsymbol{\beta}}_{el}$	0.088	1.827	1.428	-0.108	1.703	1.363	0.048	-0.108	2.315	1.850	0.118	2.282	1.776	0.087
case 3	$\hat{\boldsymbol{\beta}}_{ls}$	-0.915	6.454	5.024	0.555	6.565	5.288	0.603	0.826	5.658	4.552	-0.589	6.368	5.020	0.539
	$\hat{\boldsymbol{\beta}}_{ye}$	-0.470	4.646	3.754	0.323	4.632	3.701	0.304	0.588	4.739	3.757	-0.759	5.342	4.332	0.372
	$\hat{\boldsymbol{\beta}}_{rb}$	-0.136	2.671	2.128	0.038	2.571	2.050	0.100	0.304	3.346	2.698	-0.566	3.507	2.898	0.171
	$\hat{\boldsymbol{\beta}}_{in}$	-0.334	3.739	2.949	0.067	3.579	2.843	0.210	0.316	3.643	2.874	-0.423	3.764	3.074	0.205
	$\hat{\boldsymbol{\beta}}_{pr}$	-0.020	2.289	1.841	-0.023	2.217	1.756	0.078	0.110	3.031	2.475	-0.500	2.952	2.406	0.136
	$\hat{\boldsymbol{\beta}}_{el}$	-0.020	2.289	1.841	-0.023	2.217	1.756	0.078	0.110	3.031	2.475	-0.500	2.952	2.406	0.136

Table 3. The selected averaged optimal parameter γ_{opt} and its standard deviations (in parentheses) for example 1.

n	case 1	case 2	case 3
50	48.26(6.226)	11.55(3.921)	8.570(5.558)
100	49.57(2.471)	10.37(2.481)	8.640(4.662)
200	49.65(2.409)	10.45(1.798)	7.580(1.991)

The true values $\boldsymbol{\theta}_0$ and $\boldsymbol{\lambda}_0$ are unknown in simulations since the covariance matrix of $\boldsymbol{\varepsilon}_i$ is different from that of $\psi_\gamma(\boldsymbol{\varepsilon}_i)$. So we can not compute the biases and MADs of $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\lambda}}$. Meanwhile, $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\lambda}}$ rely on the turning parameter γ. From Table 3, we can see that the averaged optimal parameter $\gamma_{opt} = 10.37$ for case 2 and $n = 100$ in example 1. Thus we fix $\gamma = 10$ and draw the histograms and Q–Q plots of $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\lambda}}$ for case 2 and $n = 100$ in example 1, which is displayed in Fig. 1. From Fig. 1, we can see that $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\lambda}}$ are asymptotically normal, because the estimated curve of density is very close to the curve of normal density and the scattered points of Q–Q plots are very close to the line. Meanwhile, we consider the Shapiro–Wilk normality test for $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\lambda}}$, and the results also indicate they are asymptotically normal. These results agree with the theoretical result of Theorem 2.2.

Now we use the following methods to construct the confidence intervals/regions, namely, the proposed empirical likelihood method ($\hat{\boldsymbol{\beta}}_{el}$) in Section 3 and two normal approximation methods including the joint mean–covariance estimate $\hat{\boldsymbol{\beta}}_{ye}$ (Ye and Pan [32]) and the robust Huber estimate $\hat{\boldsymbol{\beta}}_{rb}$ (Zheng et al. [33]). We run simulation experiments with $n = 50, 100, 200$. The confidence intervals/regions and their coverage probabilities, with nominal level $1 - \alpha = 0.95$, are computed from 500 runs. We can derive the following several observations from Table 4 and Fig. 2. Firstly, EL method performs better than two normal approximation methods, $\hat{\boldsymbol{\beta}}_{ye}$ and $\hat{\boldsymbol{\beta}}_{rb}$, because its confidence intervals have uniformly shorter average lengths and higher coverage probabilities. Secondly, the empirical coverage probabilities tend to the nominal level 0.95 as n increases. Thirdly, we can see that EL-based regions have a slightly higher coverage probabilities and smaller area of confidence region than those of the normal approximation methods ($\hat{\boldsymbol{\beta}}_{ye}$ and $\hat{\boldsymbol{\beta}}_{rb}$). In addition, the distribution of the empirical likelihood ratio statistic $l\,(\boldsymbol{\beta}_0)$ is asymptotically $\chi^2(2)$ by Theorem 3.1. In order to prove this empirically, Fig. 3 plots the quantile of the 300 empirical likelihood ratio statistics against the quantile of $\chi^2(2)$ distribution. From Fig. 3, we can see that the scattered points of Q–Q plots are very close to the line, which is consistent with the theoretical result.

5. REAL APPLICATION

In this section, we applied the proposed robust estimation method to analyze the longitudinal progesterone data which includes a total of 492 observations. This longitudinal hormone study collects 34 women's urine samples in a menstrual cycle and has been studied by Fan et al. [5] and Zheng et al. [33]. We consider the response (y) as the log-transformed progesterone level, x_1 and x_2 are age (AGE)

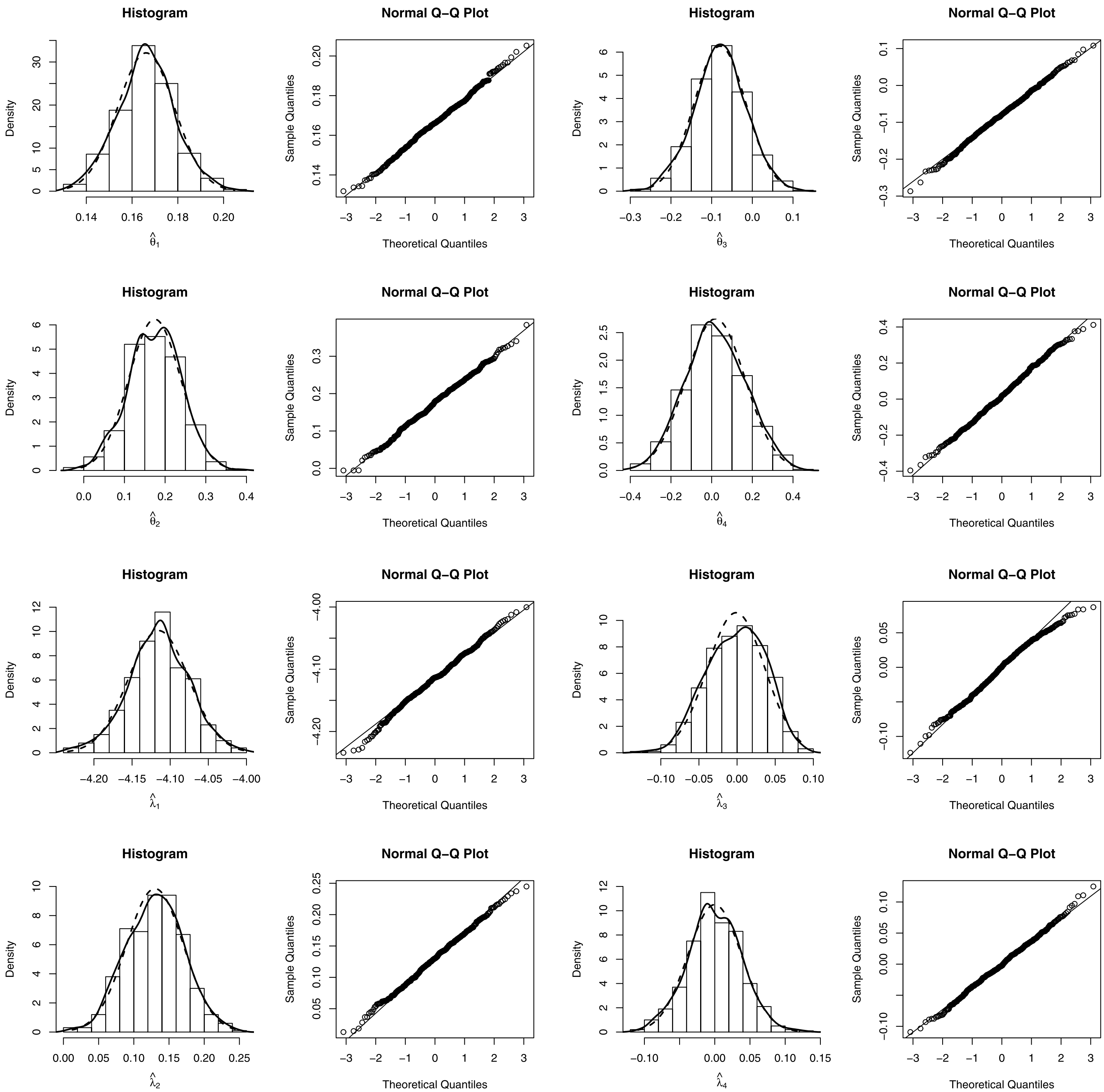

Figure 1. The histograms, the estimated curve of density (solid curve), the curve of normal density (dashed curve) and the Q-Q plots of the 300 estimates of $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)^T$ and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)^T$ for case 2 with $n = 100$ and $\gamma = 10$ in example 1.

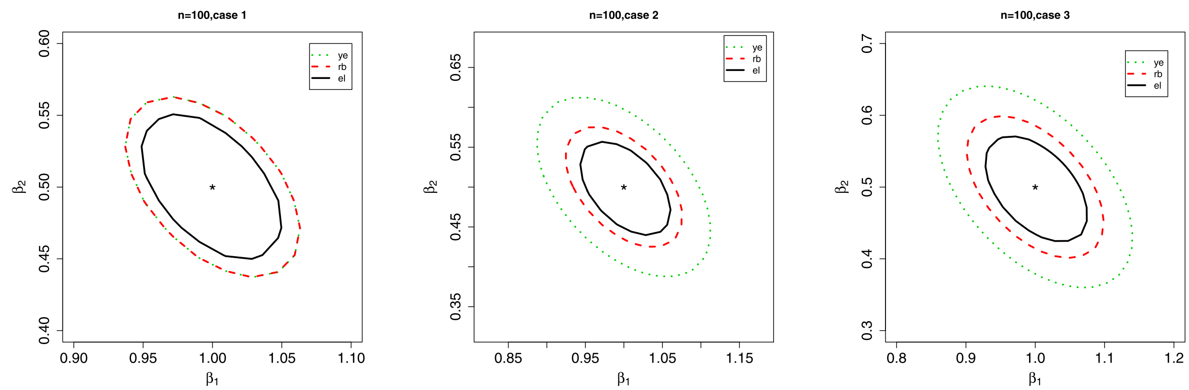

Figure 2. The 95% confidence regions for $\boldsymbol{\beta} = (1, 0.5)^T$ based on the three different methods with $n = 100$ for example 1.

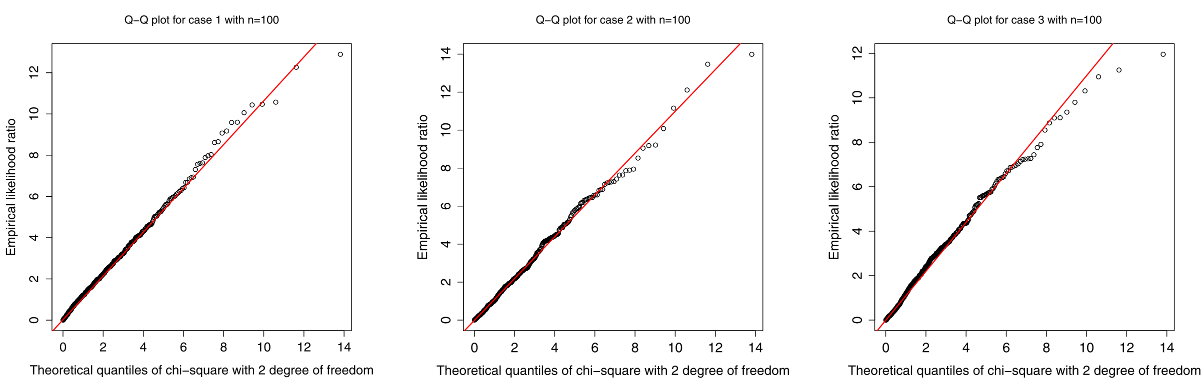

Figure 3. The Q-Q plot for example 1 with $n = 100$.

and body mass index (BMI). In this study, our main interest is to investigate whether AGE and BMI have significant influences on the progesterone level. To model jointly the mean and covariance structures for the data, we adopt the following three generalized linear models for the mean, autoregressive parameters and innovation variances:

$$\begin{aligned} &y_{ij} = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \varepsilon_{ij}, \\ &\phi_{\gamma,ijk} = \boldsymbol{w}_{ijk}^T \boldsymbol{\theta}, \\ &\log\left(d_{\gamma,ij}^2\right) = \boldsymbol{z}_{ij}^T \boldsymbol{\lambda}, \end{aligned}$$

where $\boldsymbol{w}_{ijk} = \left(1, t_{ij} - t_{ik}, ..., (t_{ij} - t_{ik})^{q-1}\right)^T$ and $\boldsymbol{z}_{ij} = \left(1, t_{ij}, ..., t_{ij}^{d-1}\right)^T$ and y_{ij} denotes the j-th log-transformed progesterone value of the ith woman, which is measured at t_{ij} with t_{ij} being rescaled into interval $[0, 1]$.

Obviously, our proposed estimators $\hat{\boldsymbol{\beta}}_{pr}$ and $\hat{\boldsymbol{\beta}}_{el}$ rely on the dimensions of covariates $\boldsymbol{w}_{ijk}$ and $\boldsymbol{z}_{ij}$ (q and d). To study the sensitivity of our approach to q and d for the finite sample, we consider different q and d for $\hat{\boldsymbol{\beta}}_{pr}$. Specifically, $\hat{\boldsymbol{\beta}}_{pr}^{22}$, $\hat{\boldsymbol{\beta}}_{pr}^{23}$, $\hat{\boldsymbol{\beta}}_{pr}^{24}$, $\hat{\boldsymbol{\beta}}_{pr}^{42}$, $\hat{\boldsymbol{\beta}}_{pr}^{43}$ and $\hat{\boldsymbol{\beta}}_{pr}^{44}$ represent the proposed estimators with $(q = 2, d = 2), (q = 2, d = 3), (q = 2, d = 4), (q = 4, d = 2), (q = 4, d = 3)$, and $(q = 4, d = 4)$. Please note that $\hat{\boldsymbol{\beta}}_{pr}$ and $\hat{\boldsymbol{\beta}}_{el}$ are asymptotically equivalent for point estimation, which has been demonstrated by simulation studies. Thus we only consider $(q = 2, d = 2)$ for $\hat{\boldsymbol{\beta}}_{el}$. For fair comparison, $\hat{\boldsymbol{\beta}}_{ls}$, $\hat{\boldsymbol{\beta}}_{ye}$, $\hat{\boldsymbol{\beta}}_{rb}$ and $\hat{\boldsymbol{\beta}}_{in}$ are also considered here. Table 5 lists the estimated regression coefficients and their 95% confidence intervals, as well as the confidence interval lengths for the intercept, AGE and BMI. Note that 95% confidence intervals of $\hat{\boldsymbol{\beta}}_{el}$ is constructed by empirical likelihood method and other approaches' 95% confidence intervals are constructed by the normal approximation method. First, it is easy to observe that AGE and BMI have insignificant influence on the progesterone level at 5% significance level for all methods since their confidence intervals contain zero. This result is consistent with that in Zheng et al. [33]. Second, $\hat{\boldsymbol{\beta}}_{pr}^{22}$, $\hat{\boldsymbol{\beta}}_{pr}^{23}$ and $\hat{\boldsymbol{\beta}}_{pr}^{24}$ have similar performances due to similar estimated values and 95% confidence intervals and perform better than $\hat{\boldsymbol{\beta}}_{pr}^{42}$, $\hat{\boldsymbol{\beta}}_{pr}^{43}$ and $\hat{\boldsymbol{\beta}}_{pr}^{44}$. The main reason may be as follows. More covariance parameters need to be estimated for large q and d and thus poorer covariance covariance matrix estimation may be obtained due to small finite sample size ($n = 34$), which leads to poor mean parameter estimators. Therefore, small q and d may be more suitable for this data set and we choose the covariates in the covariance model (6) as $\boldsymbol{z}_{ij} = (1, t_{ij})^T$ and

Table 4. Estimated coverage probabilities (CP) of confidence intervals (regions) for β_1, β_2 and $\boldsymbol{\beta}=(\beta_1,\beta_2)^T$, and averaged confidence interval lengths (Length) for β_1 and β_2 with $n=50,100,200$ for example 1.

n		method	β_1		β_2		$\boldsymbol{\beta}$
			Length	CP	Length	CP	CP
50	case 1	$\hat{\boldsymbol{\beta}}_{ye}$	0.1363	0.928	0.1376	0.952	0.946
		$\hat{\boldsymbol{\beta}}_{rb}$	0.1366	0.926	0.1376	0.950	0.934
		$\hat{\boldsymbol{\beta}}_{el}$	0.1382	0.920	0.1396	0.952	0.942
	case 2	$\hat{\boldsymbol{\beta}}_{ye}$	0.2422	0.936	0.2414	0.930	0.922
		$\hat{\boldsymbol{\beta}}_{rb}$	0.1651	0.922	0.1658	0.906	0.904
		$\hat{\boldsymbol{\beta}}_{el}$	0.1618	0.932	0.1618	0.922	0.910
	case 3	$\hat{\boldsymbol{\beta}}_{ye}$	0.3081	0.922	0.3024	0.934	0.926
		$\hat{\boldsymbol{\beta}}_{rb}$	0.2185	0.914	0.2186	0.928	0.894
		$\hat{\boldsymbol{\beta}}_{el}$	0.2093	0.924	0.2123	0.932	0.912
100	case 1	$\hat{\boldsymbol{\beta}}_{ye}$	0.0984	0.934	0.0989	0.942	0.932
		$\hat{\boldsymbol{\beta}}_{rb}$	0.0986	0.932	0.0990	0.944	0.936
		$\hat{\boldsymbol{\beta}}_{el}$	0.1003	0.940	0.1009	0.940	0.944
	case 2	$\hat{\boldsymbol{\beta}}_{ye}$	0.1766	0.952	0.1777	0.952	0.948
		$\hat{\boldsymbol{\beta}}_{rb}$	0.1190	0.938	0.1198	0.942	0.930
		$\hat{\boldsymbol{\beta}}_{el}$	0.1166	0.946	0.1179	0.942	0.938
	case 3	$\hat{\boldsymbol{\beta}}_{ye}$	0.2282	0.940	0.2288	0.952	0.934
		$\hat{\boldsymbol{\beta}}_{rb}$	0.1575	0.920	0.1565	0.920	0.890
		$\hat{\boldsymbol{\beta}}_{el}$	0.1514	0.942	0.1501	0.946	0.934
200	case 1	$\hat{\boldsymbol{\beta}}_{ye}$	0.0699	0.948	0.0698	0.946	0.936
		$\hat{\boldsymbol{\beta}}_{rb}$	0.0699	0.934	0.0699	0.930	0.942
		$\hat{\boldsymbol{\beta}}_{el}$	0.0700	0.946	0.0701	0.942	0.950
	case 2	$\hat{\boldsymbol{\beta}}_{ye}$	0.1284	0.954	0.1283	0.938	0.948
		$\hat{\boldsymbol{\beta}}_{rb}$	0.0851	0.930	0.0852	0.948	0.940
		$\hat{\boldsymbol{\beta}}_{el}$	0.0833	0.942	0.0835	0.950	0.950
	case 3	$\hat{\boldsymbol{\beta}}_{ye}$	0.1617	0.953	0.1626	0.952	0.945
		$\hat{\boldsymbol{\beta}}_{rb}$	0.1122	0.932	0.1116	0.936	0.918
		$\hat{\boldsymbol{\beta}}_{el}$	0.1067	0.955	0.1064	0.942	0.936

$\boldsymbol{w_{ijk}}=(1,t_{ij}-t_{ik})^T$. The estimated covariance parameters are $\hat{\lambda}_1=-5.8904(0.8538), \hat{\lambda}_2=1.2151(0.3482)$ and $\hat{\theta}_1=0.2551(0.0266), \hat{\theta}_2=-0.4875(0.0685)$, where the values in parentheses are standard errors obtained by the bootstrap method. Obviously, the estimated coefficients $\hat{\boldsymbol{\lambda}}$ and $\hat{\boldsymbol{\theta}}$ are significant at significance level 0.05, which means that the constructed covariance model is suitable for this data set. Third, $\hat{\boldsymbol{\beta}}_{pr}^{22}$ and $\hat{\boldsymbol{\beta}}_{el}$ have similar estimated values but $\hat{\boldsymbol{\beta}}_{el}$'s confidence interval is universally narrowest among all approaches, which indicates that $\hat{\boldsymbol{\beta}}_{el}$ has obvious superiority in confidence interval estimation.

In addition, a leave–one–out cross validation procedure is applied to investigate the predictive performance. Specifically, we assess the goodness of fit using the following criterion

$$\text{MSE}_{\text{CV}}=\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{Y}_i-\boldsymbol{X}_i\hat{\boldsymbol{\beta}}_{(-i)}\right\|,$$

where $n=34$ and $\hat{\boldsymbol{\beta}}_{(-i)}$ stands for a estimator that is obtained based on the data of the other 33 subjects except the ith subject. The MSEs of $\hat{\boldsymbol{\beta}}_{ls}$, $\hat{\boldsymbol{\beta}}_{ye}$, $\hat{\boldsymbol{\beta}}_{rb}$, $\hat{\boldsymbol{\beta}}_{in}$, $\hat{\boldsymbol{\beta}}_{pr}^{22}$, $\hat{\boldsymbol{\beta}}_{pr}^{23}$, $\hat{\boldsymbol{\beta}}_{pr}^{24}$, $\hat{\boldsymbol{\beta}}_{pr}^{42}$, $\hat{\boldsymbol{\beta}}_{pr}^{43}$, $\hat{\boldsymbol{\beta}}_{pr}^{44}$ and $\hat{\boldsymbol{\beta}}_{el}$ are 0.8627(0.3545), 0.8526(0.3499), 0.8484(0.3587), 0.8628(0.3551), 0.8478 (0.3500), 0.8477(0.3503), 0.8462(0.3567), 0.8636 (0.3600), 0.8631(0.3606), 0.8711(0.3727) and 0.8478(0.3501), where the values in parentheses are standard errors. We can clearly see that $\hat{\boldsymbol{\beta}}_{pr}^{22}$, $\hat{\boldsymbol{\beta}}_{pr}^{23}$, $\hat{\boldsymbol{\beta}}_{pr}^{23}$ and $\hat{\boldsymbol{\beta}}_{el}$ have better prediction performance than the other compared methods, which shows that it may more suitable to adopt small q and d for producing better prediction performance.

6. CONCLUDING REMARKS

This paper develops robust exponential joint mean-covariance models, in which the covariance matrix of the exponential score function is estimated via the modified Cholesky decomposition. We further develop a data–driven

Table 5. The estimates (EST), lower bound (LB), upper bound (UB) and confidence interval lengths (Length) of 95% confidence intervals for progesterone data.

method	EST			LB			UB			Length		
	Intercept	AGE	BMI	Intercept	AGE	BMI	Intercept	AGE	BMI	Intercept	AGE	BMI
$\hat{\boldsymbol{\beta}}_{ls}$	2.3275	0.0103	-0.0040	2.2794	-0.0475	-0.0719	2.3756	0.0682	0.0638	0.0962	0.1157	0.1357
$\hat{\boldsymbol{\beta}}_{ye}$	2.3432	-0.0031	0.0008	2.2913	-0.0645	-0.0600	2.3952	0.0583	0.0616	0.1039	0.1228	0.1216
$\hat{\boldsymbol{\beta}}_{rb}$	2.3454	0.0012	0.0105	2.2927	-0.0627	-0.0491	2.3981	0.0651	0.0701	0.1054	0.1277	0.1192
$\hat{\boldsymbol{\beta}}_{in}$	2.3280	0.0103	-0.0038	2.2793	-0.0480	-0.0726	2.3766	0.0686	0.0651	0.0973	0.1165	0.1377
$\hat{\boldsymbol{\beta}}_{pr}^{22}$	2.3428	-0.0021	0.0050	2.2909	-0.0629	-0.0538	2.3947	0.0586	0.0638	0.1038	0.1214	0.1177
$\hat{\boldsymbol{\beta}}_{pr}^{23}$	2.3468	0.0003	0.0049	2.2959	-0.0595	-0.0518	2.3978	0.0601	0.0617	0.1019	0.1196	0.1135
$\hat{\boldsymbol{\beta}}_{pr}^{24}$	2.3429	0.0028	0.0117	2.2899	-0.0590	-0.0453	2.3958	0.0646	0.0686	0.1059	0.1235	0.1140
$\hat{\boldsymbol{\beta}}_{pr}^{42}$	2.3361	0.0084	0.0007	2.2803	-0.0581	-0.0802	2.3919	0.0748	0.0815	0.1116	0.1329	0.1616
$\hat{\boldsymbol{\beta}}_{pr}^{43}$	2.3377	0.0095	0.0018	2.2806	-0.0571	-0.0783	2.3947	0.0761	0.0819	0.1141	0.1332	0.1602
$\hat{\boldsymbol{\beta}}_{pr}^{44}$	2.3309	0.0182	0.0046	2.2728	-0.0522	-0.0879	2.3890	0.0885	0.0972	0.1161	0.1407	0.1851
$\hat{\boldsymbol{\beta}}_{el}$	2.3428	-0.0021	0.0050	2.2987	-0.0504	-0.0507	2.3889	0.0507	0.0482	0.0902	0.1011	0.0989

procedure to select the optimal tuning parameter γ_{opt} by minimizing $det(\widehat{Cov}(\hat{\boldsymbol{\beta}}))$ to achieve better robustness and efficiency. The proposed parametric estimation method is easy to implement and more efficient than some traditional robust estimation methods. In addition, we utilize the subject–wise empirical likelihood method to construct the confidence regions/intervals of regression coefficients. Simulation studies and a real data analysis have confirmed that the proposed empirical likelihood method has better coverage and estimation accuracy than those of the normal approximation–based methods.

The focus of this article is the linear regression model. When nonlinearity is present, nonparametric or semiparametric models may be more useful. Leng et al. [9] constructed joint semiparametric mean–covariance models when analyzing longitudinal data. Lai et al. [8] developed variable selection procedure for longitudinal single index models based on smooth-threshold estimating equations, which is computationally simpler than traditional shrinkage penalty approaches. Chen et al. [2] studied longitudinal partially linear single index models under a general framework including both the sparse and dense longitudinal data cases. However, these approaches are not robust, and thus it is of great interest to extend the proposed method to these important areas of research.

ACKNOWLEDGEMENTS

The authors are very grateful to the editor and two anonymous referees for their detailed comments on the earlier version of the manuscript, which led to a much improved paper. This paper is supported by the National Social Science Fund of China (Grant No. 17CTJ015).

APPENDIX. PROOFS

To establish the asymptotic properties of proposed estimators, the following conditions are needed.

(C1) We assume that the dimensions p, q and d of the covariates $\boldsymbol{x}_{ij}$, $\boldsymbol{w}_{ijk}$ and $\boldsymbol{z}_{ij}$, $i = 1, ..., n, j = 1, ..., m_i, k = 1, ..., j-1$ are fixed and that $\{m_i\}$ is a bounded sequence of positive integers.

(C2) The parameter space Θ of $\boldsymbol{\eta} = \left(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T, \boldsymbol{\lambda}^T\right)^T$ is a compact subset of $\mathbb{R}^{p+q+d}$, and the true parameter value $\boldsymbol{\eta}_0 = \left(\boldsymbol{\beta}_0^T, \boldsymbol{\theta}_0^T, \boldsymbol{\lambda}_0^T\right)^T$ is in the interior of the parameter space Θ.

(C3) The covariates $\boldsymbol{x}_{ij}$, $\boldsymbol{w}_{ijk}$ and $\boldsymbol{z}_{ij}$, and the matrices $\boldsymbol{W}_{i\gamma}^{-1}$ are all bounded, meaning that all the elements of the vectors and matrices are bounded.

(C4) $E\left(\psi_\gamma\left(\varepsilon_{ij}\right)\right) = 0$, $E\left({\psi'}_\gamma\left(\varepsilon_{ij}\right)\right) < 0$ and $E\left(\psi_\gamma(\varepsilon_{ij})^2\right)$ is finite for any $\gamma > 0$.

Proof of Theorem 2.1. We only show that $\hat{\boldsymbol{\beta}} \to \boldsymbol{\beta}_0$ almost surely, because the proofs for $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\lambda}}$ are similar. According to McCullagh [14], we have

$$\begin{aligned}\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = & \left[\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i^T \boldsymbol{\Sigma}_{i\gamma}^{-1} \boldsymbol{\Lambda}_{i\gamma} \boldsymbol{X}_i\right]^{-1}_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \\ & \times \left\{\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i^T \boldsymbol{\Sigma}_{i\gamma}^{-1} \psi_\gamma\left(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}\right)\right\}_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \\ & + o_p\left(n^{-1/2}\right).\end{aligned} \tag{A.1}$$

Thus, by the condition (C4), the expectation and covariance matrix of $\boldsymbol{U}_{1i} = \boldsymbol{X}_i^T \boldsymbol{\Sigma}_{i\gamma}^{-1} \psi_\gamma\left(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}\right)$ are $E_0\left(\boldsymbol{U}_{1i}\right) = \mathbf{0}$ and $V_0\left(\boldsymbol{U}_{1i}\right) = \boldsymbol{X}_i^T \boldsymbol{\Sigma}_{i\gamma}^{-1} \boldsymbol{\Sigma}_{i\gamma} \boldsymbol{\Sigma}_{i\gamma}^{-1} \boldsymbol{X}_i = \boldsymbol{X}_i^T \boldsymbol{\Sigma}_{i\gamma}^{-1} \boldsymbol{X}_i$ at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, where $E_0(u)$ and $V_0(u)$ stand for the expectation and covariance of u at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. Since $\boldsymbol{\Sigma}_{i\gamma} = \boldsymbol{\Phi}_{i\gamma}^{-1} \boldsymbol{D}_{i\gamma} (\boldsymbol{\Phi}_{i\gamma}^T)^{-1}$ and $\boldsymbol{\Sigma}_{i\gamma}^{-1} = \boldsymbol{\Phi}_{i\gamma}^T \boldsymbol{D}_{i\gamma}^{-1} \boldsymbol{\Phi}_{i\gamma}$, the variance matrix $V_0\left(\boldsymbol{U}_{1i}\right)$ can be further written as

$$V_0\left(\boldsymbol{U}_{1i}\right) = \boldsymbol{X}_i^T \boldsymbol{\Phi}_{i\gamma}^T \boldsymbol{D}_{i\gamma}^{-1} \boldsymbol{\Phi}_{i\gamma} \boldsymbol{X}_i.$$

The condition (C3) means that there exists a constant κ_0 such that $V_0\left(\boldsymbol{U}_{1i}\right) \le \kappa_0 I_{p\times p}$ for any i and all $\boldsymbol{\eta} \in \Theta$, where $\boldsymbol{I}_{p\times p}$ is a $p \times p$ matrix with all elements being 1's. That is, all elements of $V_0\left(\boldsymbol{U}_{1i}\right)$ are bounded by κ_0. Hence we can obtain $\sum_{i=1}^{\infty} \frac{V_0(\boldsymbol{U}_{1i})}{i^2} < \infty$. Thus, by Kolmogorov's strong law of large numbers, together with $E_0\left(\boldsymbol{U}_{1i}\right) = \mathbf{0}$, we have

$$\left\{n^{-1}\sum_{i=1}^{n} \boldsymbol{X}_i^T \boldsymbol{\Sigma}_{i\gamma}^{-1} \psi_\gamma\left(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta}\right)\right\}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \overset{a.s.}{\to} \mathbf{0} \tag{A.2}$$

as $n \to \infty$. Similarly, it can be shown that $\left\{\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i^T \boldsymbol{\Sigma}_{i\gamma}^{-1} \boldsymbol{\Lambda}_{i\gamma} \boldsymbol{X}_i\right\}_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$ is a bounded matrix. Application of (A.2) to (A.1) leads to $\hat{\boldsymbol{\beta}} \to \boldsymbol{\beta}_0$ almost surely as $n \to \infty$. The proof is complete. □

Lemma 1. *Under conditions (C1)–(C4), let* $\left(\hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{\theta}}^T, \hat{\boldsymbol{\lambda}}^T\right)^T$ *be the root of the robust generalized estimating equations (7)–(9), then*

$$\left\|\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) - \tilde{\boldsymbol{\beta}}\right\| = o_p\left(1\right), \left\|\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) - \tilde{\boldsymbol{\theta}}\right\| = o_p\left(1\right),$$
$$\left\|\sqrt{n}\left(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0\right) - \tilde{\boldsymbol{\lambda}}\right\| = o_p\left(1\right),$$

where

$$\tilde{\boldsymbol{\beta}} = \left\{\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i^T \boldsymbol{\Sigma}_{i\gamma}^{-1} \boldsymbol{\Lambda}_{i\gamma} \boldsymbol{X}_i\right\}^{-1}_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \frac{1}{\sqrt{n}} \boldsymbol{U}_1\left(\boldsymbol{\beta}_0\right),$$

$$\tilde{\boldsymbol{\theta}} = \left\{\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{T}_{i\gamma}^T \boldsymbol{D}_{i\gamma}^{-1} \boldsymbol{T}_{i\gamma}\right\}^{-1}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \frac{1}{\sqrt{n}} \boldsymbol{U}_2\left(\boldsymbol{\theta}_0\right),$$

and

$$\tilde{\boldsymbol{\lambda}} = \left\{\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{Z}_i^T \boldsymbol{D}_{i\gamma} \boldsymbol{W}_{i\gamma}^{-1} \boldsymbol{D}_{i\gamma} \boldsymbol{Z}_i\right\}^{-1}_{\boldsymbol{\lambda}=\boldsymbol{\lambda}_0} \frac{1}{\sqrt{n}} \boldsymbol{U}_3\left(\boldsymbol{\lambda}_0\right).$$

Proof of Lemma 1. The proof is similar to Lemma 2 of Leng et al. [9] and Lemma in Zheng et al. [33], thus we omitted the details.

Proof of Theorem 2.2. According to Lemma 1, we only need to show the asymptotic normality of $\left(\tilde{\boldsymbol{\beta}}^T, \tilde{\boldsymbol{\theta}}^T, \tilde{\boldsymbol{\lambda}}^T\right)^T$. This is equivalent to the asymptotic normality of $\left(\boldsymbol{U}_1(\boldsymbol{\beta}_0)^T, \boldsymbol{U}_2(\boldsymbol{\theta}_0)^T \ \ , \boldsymbol{U}_3(\boldsymbol{\lambda}_0)^T\right)^T/\sqrt{n}$. Note that conditions (C1)–(C4) imply that

$$\begin{aligned} E_0 \Big[\big| & \boldsymbol{\varsigma}^T \left\{ \boldsymbol{X}_i^T \boldsymbol{\Sigma}_{i\gamma}^{-1} \psi_\gamma \left(\boldsymbol{Y}_i - \boldsymbol{X}_i \boldsymbol{\beta} \right) \right\} + \boldsymbol{\varrho}^T \left\{ \boldsymbol{T}_i^T \boldsymbol{D}_{i\gamma}^{-1} \boldsymbol{e}_{i\gamma} \right\} \\ & + \boldsymbol{\vartheta}^T \left\{ \boldsymbol{Z}_i^T \boldsymbol{D}_{i\gamma} \boldsymbol{W}_{i\gamma}^{-1} \left(\boldsymbol{e}_{i\gamma}^2 - \boldsymbol{\sigma}_{i\gamma}^2 \right) \right\} \big|^3 \Big] \leq \kappa \end{aligned}$$

for any $\boldsymbol{\varsigma} \in \mathbb{R}^p$, $\boldsymbol{\varrho} \in \mathbb{R}^q$ and $\boldsymbol{\vartheta} \in \mathbb{R}^d$, where κ is a constant independent i and E_0 represents the expectation at $\boldsymbol{\eta} = \boldsymbol{\eta}_0$. Furthermore, we have

$$\begin{aligned} & \tfrac{1}{n} \sum_{i=1}^n V_0 \Big[\big| \boldsymbol{\varsigma}^T \left\{ \boldsymbol{X}_i^T \boldsymbol{\Sigma}_{i\gamma}^{-1} \psi_\gamma \left(\boldsymbol{Y}_i - \boldsymbol{X}_i \boldsymbol{\beta} \right) \right\} \\ & \quad + \boldsymbol{\varrho}^T \left\{ \boldsymbol{T}_{i\gamma}^T \boldsymbol{D}_{i\gamma}^{-1} \boldsymbol{e}_{i\gamma} \right\} + \boldsymbol{\vartheta}^T \left\{ \boldsymbol{Z}_i^T \boldsymbol{D}_{i\gamma} \boldsymbol{W}_{i\gamma}^{-1} \left(\boldsymbol{e}_{i\gamma}^2 - \boldsymbol{\sigma}_{i\gamma}^2 \right) \right\} \big| \Big] \\ & = \left(\boldsymbol{\varsigma}^T, \boldsymbol{\varrho}^T, \boldsymbol{\vartheta}^T \right) \tfrac{1}{n} \boldsymbol{V}_n \left(\boldsymbol{\varsigma}^T, \boldsymbol{\varrho}^T, \boldsymbol{\vartheta}^T \right)^T \\ & \to \left(\boldsymbol{\varsigma}^T, \boldsymbol{\varrho}^T, \boldsymbol{\vartheta}^T \right) \boldsymbol{V} \left(\boldsymbol{\varsigma}^T, \boldsymbol{\varrho}^T, \boldsymbol{\vartheta}^T \right)^T > 0, \end{aligned}$$

where V_0 stands for the covariance at $\boldsymbol{\eta} = \boldsymbol{\eta}_0$. Thus, we can apply multivariate Liapounov central limit theorem to establish the asymptotic normality of $\left(\boldsymbol{U}_1(\boldsymbol{\beta}_0)^T, \boldsymbol{U}_2(\boldsymbol{\theta}_0)^T, \boldsymbol{U}_3(\boldsymbol{\lambda}_0)^T\right)^T/\sqrt{n}$. Let

$$\boldsymbol{K}_1 = \left\{ \frac{1}{n} \sum_{i=1}^n \boldsymbol{X}_i^T \boldsymbol{\Sigma}_{i\gamma}^{-1} \boldsymbol{\Lambda}_{i\gamma} \boldsymbol{X}_i \right\}_{\boldsymbol{\beta}=\boldsymbol{\beta}_0},$$

$$\boldsymbol{K}_2 = \left\{ \frac{1}{n} \sum_{i=1}^n \boldsymbol{T}_{i\gamma}^T \boldsymbol{D}_{i\gamma}^{-1} \boldsymbol{T}_{i\gamma} \right\}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0},$$

and

$$\boldsymbol{K}_3 = \left\{ \frac{1}{n} \sum_{i=1}^n \boldsymbol{Z}_i^T \boldsymbol{D}_{i\gamma} \boldsymbol{W}_{i\gamma}^{-1} \boldsymbol{D}_{i\gamma} \boldsymbol{Z}_i \right\}_{\boldsymbol{\lambda}=\boldsymbol{\lambda}_0}.$$

Therefore we have

$$\begin{aligned} & \sqrt{n} \begin{pmatrix} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \\ \hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0 \end{pmatrix} \\ & = \begin{pmatrix} \boldsymbol{K}_1 & 0 & 0 \\ 0 & \boldsymbol{K}_2 & 0 \\ 0 & 0 & \boldsymbol{K}_3 \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{U}_1 (\boldsymbol{\beta}_0)/\sqrt{n} \\ \boldsymbol{U}_2 (\boldsymbol{\theta}_0)/\sqrt{n} \\ \boldsymbol{U}_3 (\boldsymbol{\lambda}_0)/\sqrt{n} \end{pmatrix} + o_p(1) \\ & \xrightarrow{d} N\left(\boldsymbol{0}, \boldsymbol{A}^{-1} \boldsymbol{V} \boldsymbol{A}^{-1} \right). \end{aligned}$$

The proof of Theorem 2.2 is completed. □

Proof of Theorem 3.1. We first define $\boldsymbol{\Omega} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \boldsymbol{X}_i^T \hat{\boldsymbol{\Sigma}}_{i\gamma}^{-1} \boldsymbol{\Sigma}_{i\gamma} \hat{\boldsymbol{\Sigma}}_{i\gamma}^{-1} \boldsymbol{X}_i$. Applying the Lindeberg central limit theorem, together with condition (C4), we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{\xi}_i (\boldsymbol{\beta}_0) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Omega}), \tag{A.3}$$

and

$$\frac{1}{n} \sum_{i=1}^n \boldsymbol{\xi}_i (\boldsymbol{\beta}_0) \boldsymbol{\xi}_i^T (\boldsymbol{\beta}_0) \xrightarrow{p} \boldsymbol{\Omega}. \tag{A.4}$$

From (A.3), (A.4), and using the same arguments that are used in the proof of (2.14) in Owen [15], we can prove that

$$\boldsymbol{\rho} = O_p\left(n^{-1/2}\right), \tag{A.5}$$

where $\boldsymbol{\rho}$ is defined in section 3. Applying the Taylor expansion to (18) and invoking (A.3)–(A.5), we obtain

$$l(\boldsymbol{\beta}_0) = 2 \sum_{i=1}^n \left\{ \boldsymbol{\rho}^T \boldsymbol{\xi}_i (\boldsymbol{\beta}_0) - \left[\boldsymbol{\rho}^T \boldsymbol{\xi}_i (\boldsymbol{\beta}_0) \right]^2 \big/ 2 \right\} + o_p(1). \tag{A.6}$$

By (17), it follows that

$$\begin{aligned} \boldsymbol{0} &= \sum_{i=1}^n \frac{\boldsymbol{\xi}_i(\boldsymbol{\beta}_0)}{1+\boldsymbol{\rho}^T \boldsymbol{\xi}_i(\boldsymbol{\beta}_0)} \\ &= \sum_{i=1}^n \boldsymbol{\xi}_i (\boldsymbol{\beta}_0) - \sum_{i=1}^n \boldsymbol{\rho}^T \boldsymbol{\xi}_i (\boldsymbol{\beta}_0) \boldsymbol{\xi}_i (\boldsymbol{\beta}_0) \\ &\quad + \frac{\sum_{i=1}^n \boldsymbol{\xi}_i(\boldsymbol{\beta}_0) \left[\boldsymbol{\rho}^T \boldsymbol{\xi}_i(\boldsymbol{\beta}_0) \right]^2}{1+\boldsymbol{\rho}^T \boldsymbol{\xi}_i(\boldsymbol{\beta}_0)}. \end{aligned}$$

The application of (A.3)–(A.5) again yields

$$\boldsymbol{\rho} = \left[\sum_{i=1}^n \boldsymbol{\xi}_i (\boldsymbol{\beta}_0) \boldsymbol{\xi}_i^T (\boldsymbol{\beta}_0) \right]^{-1} \sum_{i=1}^n \boldsymbol{\xi}_i (\boldsymbol{\beta}_0) + o_p\left(n^{-1/2}\right), \tag{A.7}$$

and

$$\sum_{i=1}^n \left[\boldsymbol{\rho}^T \boldsymbol{\xi}_i (\boldsymbol{\beta}_0) \right]^2 = \sum_{i=1}^n \boldsymbol{\rho}^T \boldsymbol{\xi}_i (\boldsymbol{\beta}_0) + o_p(1), \tag{A.8}$$

Substituting (A.7) and (A.8) into (A.6), we obtain

$$\begin{aligned} l(\boldsymbol{\beta}_0) &= \left[\tfrac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{\xi}_i (\boldsymbol{\beta}_0) \right]^T \left[\tfrac{1}{n} \sum_{i=1}^n \boldsymbol{\xi}_i (\boldsymbol{\beta}_0) \boldsymbol{\xi}_i^T (\boldsymbol{\beta}_0) \right]^{-1} \\ &\quad \times \left[\tfrac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{\xi}_i (\boldsymbol{\beta}_0) \right] + o_p(1). \end{aligned} \tag{A.9}$$

Based on (A.3), (A.4) and (A.9), we can prove Theorem 3.1. □

Proof of Corollary 3.1. Let $\tilde{\boldsymbol{\rho}} = \boldsymbol{\rho}\left(\boldsymbol{b}_0, \tilde{\boldsymbol{\beta}}^{(2)}\right)$ and $\boldsymbol{\xi}_i^{(2)}\left(\boldsymbol{b}_0, \tilde{\boldsymbol{\beta}}^{(2)}\right) = \partial \boldsymbol{\xi}_i \left(\boldsymbol{b}_0, \tilde{\boldsymbol{\beta}}^{(2)}\right) \big/ \partial \tilde{\boldsymbol{\beta}}^{(2)}$, then $\tilde{\boldsymbol{\rho}}$ and $\tilde{\boldsymbol{\beta}}^{(2)}$ satisfy

$$Q_1\left(\boldsymbol{b}_0, \tilde{\boldsymbol{\beta}}^{(2)}, \tilde{\boldsymbol{\rho}}\right) = \sum_{i=1}^n \frac{\boldsymbol{\xi}_i\left(\boldsymbol{b}_0, \tilde{\boldsymbol{\beta}}^{(2)}\right)}{1 + \tilde{\boldsymbol{\rho}}^T \boldsymbol{\xi}_i\left(\boldsymbol{b}_0, \tilde{\boldsymbol{\beta}}^{(2)}\right)} = \boldsymbol{0},$$

and

$$Q_2\left(\boldsymbol{b}_0, \tilde{\boldsymbol{\beta}}^{(2)}, \tilde{\boldsymbol{\rho}}\right) = \sum_{i=1}^{n} \frac{\tilde{\boldsymbol{\rho}}^T \boldsymbol{\xi}_i^{(2)}\left(\boldsymbol{b}_0, \tilde{\boldsymbol{\beta}}^{(2)}\right)}{1 + \tilde{\boldsymbol{\rho}}^T \boldsymbol{\xi}_i\left(\boldsymbol{b}_0, \tilde{\boldsymbol{\beta}}^{(2)}\right)} = \mathbf{0}.$$

Expanding $Q_1\left(\boldsymbol{b}_0, \tilde{\boldsymbol{\beta}}^{(2)}, \tilde{\boldsymbol{\rho}}\right)$ and $Q_2\left(\boldsymbol{b}_0, \tilde{\boldsymbol{\beta}}^{(2)}, \tilde{\boldsymbol{\rho}}\right)$ at $\left(\boldsymbol{b}_0, \boldsymbol{\beta}_0^{(2)}, 0\right)$, we have

$$\tilde{\boldsymbol{\rho}} = (\boldsymbol{I} - \boldsymbol{P})\,\boldsymbol{\Sigma}_n^{-1}\tilde{\boldsymbol{\xi}} + o_p\left(n^{-1/2}\right),$$

and

$$\tilde{\boldsymbol{\beta}}^{(2)} - \boldsymbol{\beta}_0^{(2)} = -\left(\tilde{\boldsymbol{\xi}}^{(2)T}\boldsymbol{\Sigma}_n^{-1}\tilde{\boldsymbol{\xi}}^{(2)}\right)^{-1}\tilde{\boldsymbol{\xi}}^{(2)T}\boldsymbol{\Sigma}_n^{-1}\tilde{\boldsymbol{\xi}} + o_p\left(n^{-1/2}\right),$$

where

$$\tilde{\boldsymbol{\xi}} = \sum_{i=1}^{n} \boldsymbol{\xi}_i\left(\boldsymbol{b}_0, \boldsymbol{\beta}_0^{(2)}\right), \tilde{\boldsymbol{\xi}}^{(2)} = \sum_{i=1}^{n} \boldsymbol{\xi}_i^{(2)}\left(\boldsymbol{b}_0, \boldsymbol{\beta}_0^{(2)}\right),$$
$$\boldsymbol{P} = \boldsymbol{\Sigma}_n^{-1}\tilde{\boldsymbol{\xi}}^{(2)}\left(\tilde{\boldsymbol{\xi}}^{(2)T}\boldsymbol{\Sigma}_n^{-1}\tilde{\boldsymbol{\xi}}^{(2)}\right)^{-1}\tilde{\boldsymbol{\xi}}^{(2)T}$$

and

$$\boldsymbol{\Sigma}_n = \sum_{i=1}^{n} \boldsymbol{\xi}_i\left(\boldsymbol{b}_0, \boldsymbol{\beta}_0^{(2)}\right)\boldsymbol{\xi}_i^T\left(\boldsymbol{b}_0, \boldsymbol{\beta}_0^{(2)}\right).$$

Because

$$\begin{aligned} & l\left(\boldsymbol{b}_0, \tilde{\boldsymbol{\beta}}^{(2)}\right) \\ &= 2\sum_{i=1}^{n}\log\left\{1 + \tilde{\boldsymbol{\rho}}^T\boldsymbol{\xi}_i\left(\boldsymbol{b}_0, \tilde{\boldsymbol{\beta}}^{(2)}\right)\right\} \\ &= 2\sum_{i=1}^{n}\tilde{\boldsymbol{\rho}}^T\boldsymbol{\xi}_i\left(\boldsymbol{b}_0, \tilde{\boldsymbol{\beta}}^{(2)}\right) - \sum_{i=1}^{n}\left\{\tilde{\boldsymbol{\rho}}^T\boldsymbol{\xi}_i\left(\boldsymbol{b}_0, \tilde{\boldsymbol{\beta}}^{(2)}\right)\right\}^2 + o_p(1) \\ &= \tilde{\boldsymbol{\xi}}^T\boldsymbol{\Sigma}_n^{-1/2}\left(\boldsymbol{I} - \boldsymbol{\Sigma}_n^{1/2}\boldsymbol{P}\boldsymbol{\Sigma}_n^{-1/2}\right)\boldsymbol{\Sigma}_n^{-1/2}\tilde{\boldsymbol{\xi}} + o_p(1). \end{aligned}$$

Similar to the proof of Theorem 3.1, we have $\boldsymbol{\Sigma}_n^{-1/2}\tilde{\boldsymbol{\xi}} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{I})$ and $\boldsymbol{\Sigma}_n^{1/2}\boldsymbol{P}\boldsymbol{\Sigma}_n^{-1/2}$ is symmetric and idempotent, with trace equal to $p - p_1$. Because $l\left(\hat{\boldsymbol{\beta}}^{(1)}, \hat{\boldsymbol{\beta}}^{(2)}\right) = 0$. Hence the empirical likelihood ratio statistic $\bar{l}(\boldsymbol{b}_0)$ converges to $\chi^2_{p_1}$. □

Received 14 April 2018

REFERENCES

[1] 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. MR2659868

[2] Chen, J. Li, D. Liang, H. and Wang, S. (2015). Semiparametric GEE analysis in partially linear single-index models for longitudinal data. *The Annals of Statistics* **43** 1682–1715. MR3357875

[3] Chen, J. Sitter, R. R. and Wu, C. (2002). Using empirical likelihood methods to obtain range restricted weights in regression estimators for surveys. *Biometrika* **89** 230–237. MR1888365

[4] Croux, C. Gijbels, I. and Prosdocomo, I. (2012). Robust estimation of mean and dispersion functions in extended generalized additive models. *Biometrics* **68** 31–44. MR2909851

[5] Fan, Y. Qin, G. and Zhu, Z. (2012). Variable selection in robust regression models for longitudinal data. *Journal of Multivariate Analysis* **109** 156–167. MR2922861

[6] Guo, C. Yang, H. Lv, J. and Wu, J. (2016). Joint estimation for single index mean-covariance models with longitudinal data. *Journal of the Korean Statistical Society* **45** 526–543. MR3566160

[7] 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* **472** 1176–1184. MR2236433

[8] Lai, P. Wang, Q. and Liang, H. (2012). Bias-corrected GEE estimation and smooth-threshold GEE variable selection for single-index models with clustered data. *Journal of Multivariate Analysis* **105** 422–432. MR2877527

[9] Leng, C. Zhang, W. and Pan, J. (2010). Semiparametric mean-covariance regression analysis for longitudinal data. *Journal of the American Statistical Association* **105** 181–193. MR2656048

[10] Li, P. and Pan, J. (2013). Empirical likelihood for generalized linear models with longitudinal data. *Journal of Multivariate Analysis* **114** 63–73. MR2993874

[11] Liang, K.Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73** 13–22. MR0836430

[12] Lv, J. Guo, C. Li, T. Hao, Y. and Pan, X. (2018). Adaptive robust estimation in joint mean–covariance regression model for bivariate longitudinal data. *Statistics* **52** 64–83. MR3757189

[13] Mao, J. Zhu, Z. and Fung, W. K. (2011). Joint estimation of mean-covariance model for longitudinal data with basis function approximations. *Computational Statistics and Data Analysis* **55** 983–992. MR2736487

[14] McCullagh, P. (1983). Quasi-likelihood functions. *The Annals of Statistics* **11** 59–67. MR0684863

[15] Owen, A. B. (1990). Empirical Likelihood Ratio Confidence Regions. *The Annals of Statistics* **18** 90–120. MR1041387

[16] Owen, A.B. (2001). *Empirical likelihood*. Chapman and Hall-CRC, New York.

[17] Qin, G. Bai, Y. and Zhu, Z. 2012. Robust empirical likelihood inference for generalized partial linear models with longitudinal data. *Journal of Multivariate Analysis* **105** 32–44. MR2877501

[18] Qin, G. Mao, J. and Zhu, Z. 2016. Joint mean-covariance model in generalized partially linear varying coefficient models for longitudinal data. *Journal of Statistical Computation and Simulation* **86** 1166–1182. MR3441562

[19] Qin, G. and Zhu, Z. 2007. Robust estimation in generalized semiparametric mixed models for longitudinal data. *Journal of Multivariate Analysis* **98** 1658–1683. MR2370112

[20] Qin, G. Zhu, Z. and Fung, W.K. 2016. Robust estimation of generalized partially linear model for longitudinal data with dropouts. *Annals of the Institute of Statistical Mathematics* **68** 977–1000. MR3538582

[21] Qin, J. and Lawless, J. 1994. Empirical likelihood and general estimating equations. *The Annals of Statistics* **22** 300–325. MR1272085

[22] Qu, A. Lindsay, B. and Li, B. (2000). Improving estimating equations using quadratic inference functions. *Biometrika* **87** 823–836. MR1813977

[23] Song, Y. Jian, L. and Lin, L. (2016). Robust exponential squared loss-based variable selection for high-dimensional single-index varying-coefficient model. *Journal of Computational and Applied Mathematics* **308** 330–345. MR3523009

[24] Tang, C. Leng, C. (2011). Empirical likelihood and quantile regression in longitudinal data analysis. *Biometrika* **98** 1001–1006. MR2860340

[25] Tang, Y. Wang, Y. Li, J. and Qian, W. (2015). Improving estimation efficiency in quantile regression with longitudinal data. *Journal of Statistical Planning and Inference* **165** 38–55. MR3350257

[26] Wang, H. and Zhu, Z. (2011). Empirical likelihood for quantile regression models with longitudinal data. *Journal of Statistical Planning and Inference* **141** 1603–1615. MR2747929

[27] Wang, K. and Lin, L. (2016). Robust structure identification and variable selection in partial linear varying coefficient models. *Journal of Statistical Planning and Inference* **174** 153–168. MR3477705

[28] Wang, S. Qian, L. and Carroll, R.J. (2010). Generalized empirical likelihood methods for analyzing longitudinal data. *Biometrika* **97** 79–93. MR2594418

[29] Wang, X. Jiang, Y. Huang, M. and Zhang, H. (2013). Robust variable selection with exponential squared loss. *Journal of the American Statistical Association* **108** 632–643. MR3174647

[30] Wang, Y. G. Lin, X. and Zhu, M. (2005). Robust estimation functions and bias correction for longitudinal data analysis. *Biometrics* **61** 684–691. MR2196156

[31] Xue, L. and Zhu, L. (2007). Empirical likelihood for a varying coefficient model with longitudinal data, *Journal of the American Statistical Association* **102** 642–654. MR2370858

[32] Ye, H. and Pan, J. (2006). Modelling of covariance structures in generalised estimating equations for longitudinal data. *Biometrika* **93** 927–941. MR2285080

[33] Zheng, X. Fung, W. 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. MR3094949

[34] Zheng, X. Fung, W. and Zhu, Z. (2014). Variable selection in robust joint mean and covariance model for longitudinal data analysis. *Statistica Sinica* **24** 515–531. MR3235387

[35] Zheng, X. Qin, G. Tu, D. (2017). A generalized partially linear mean-covariance regression model for longitudinal proportional data, with applications to the analysis of quality of life data from cancer clinical trials. *Statistics in Medicine*, **36** 1884–1894. MR3648627

Jing Lv
School of Mathematics and Statistics
Southwest University
Chongqing, 400715
China
E-mail address: cqulv2015@126.com

Chaohui Guo
College of Mathematics Science
Chongqing Normal University
Chongqing, 401331
China
E-mail address: guochaohui2010@126.com

Jibo Wu
School of Mathematics and Finances
Chongqing University of Arts and Sciences
Chongqing, 402160
China
E-mail address: linfen52@126.com