Journal of Multivariate Analysis 171 (2019) 397–411

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

# Robust maximum $L_q$-likelihood estimation of joint mean–covariance models for longitudinal data

Lin Xu [a], Sijia Xiang [a], Weixin Yao [b,*]

[a] School of Data Sciences, Zhejiang University of Finance and Economics, China
[b] Department of Statistics, University of California, Riverside, CA, United States

ARTICLE INFO

*Article history:*
Received 4 September 2017
Available online 22 January 2019

*Keywords:*
Joint mean–covariance models
Longitudinal daâĂŚta analysis
Maximum $L_q$-likelihood
Modified Cholesky decomposition

*AMS subject classifications:*
62F35
62H12

ABSTRACT

A comprehensive longitudinal data analysis requires screening for unusual observations. Outliers or measurement errors might lead to considerable efficiency loss or even misleading results in longitudinal data inference. Via joint mean–covariance modelings (Pourahmadi, 2000; Zhang et al., 2015) and $q$-order entropy theory (Ferrari, 2010), we propose a maximum $L_q$-likelihood estimation for longitudinal data, which can yield robust and consistent estimators of the mean regression coefficients. An EM type algorithm is introduced to achieve both efficient and stable computation. The asymptotic properties of the proposed estimators are provided. Simulation studies and an application to Turkish anesthesiology data are used to show the effectiveness of the new approach.

## 1. Introduction

As data collection techniques improve in finance, biomedical sciences, environmental sciences, and linguistics, among others, an increasing number of high-dimensional data sets are collected and stored. One prominent feature of these data sets is that they may contain outliers, due to heavy-tailed error distributions and/or inevitable errors in the data collection process. These unusual observations, if not treated properly, may result in efficiency loss in statistical inference, or even biased and incorrect conclusions.

When repeated measurements are taken on a subject, the correlation between them is typically a function of time difference or location distance. When the within-subject correlation changes dynamically according to time or location, traditional static covariation patterns (e.g., AR, MA, or exchangeable structure) can hardly depict the dynamic dependence structure flexibly [2,7,16].

To overcome the limitations mentioned above, dynamic covariance modeling can be used to allow the within-subject correlation to change dynamically according to time or location. Joint mean–covariance modeling approaches such as MCDF and HSCF are commonly used for dynamic covariance modeling to depict the dynamic within-subject covariations; see, e.g., [13,25]. They also provide many parsimonious unconstrained parameterizations by interpreting the dependence structure and innovation variance in a time series context; see, e.g., [3,5].

Usually, the variances and correlations of observations depend on their own real-time characters, which can be depicted by the corresponding observed times and covariates in regression. A joint mean–covariance model can capture the real-time variant information of covariation, by modeling the logarithms of prediction error variances as a linear structure of the covariates and the entries in the correlation structure as a polynomial of the observed time difference. These parsimonious

* Corresponding author.
*E-mail address:* weixin.yao@ucr.edu (W. Yao).

https://doi.org/10.1016/j.jmva.2019.01.001
0047-259X/© 2019 Elsevier Inc. All rights reserved.

parameterizations are more flexible and adaptive for characterizing the dynamic correlation mechanism and yield more efficient maximum likelihood or GEE estimators of the mean regression coefficients [13,22]. However, these modeling approaches are all based on the traditional maximum likelihood estimation (MLE) technique, which is well-known not to be a robust estimation methodology. Thus almost all joint mean–covariance models studied thus far are sensitive to outliers, contamination, or heavy-tailed distributions.

In this paper, we propose a robust estimation of joint mean–covariance models by combining ideas from maximum $L_q$-likelihood estimation [4,15] and joint mean–covariance modeling. The new approach combines the flexibility of existing joint mean–covariance models and the robustness of $L_q$-likelihood. It yields comparable performance to the traditional MLE when there is no outlier but much better estimators when outliers are present, as illustrated by simulation studies and an application to spinal anesthesiology data.

The remainder of the paper is organized as follows. In Section 2, we propose a maximum $L_q$-likelihood estimation based on a joint mean–covariance modeling framework. The asymptotic properties of the estimators are given in Section 3, and simulation studies are conducted in Section 4. An application to spinal anesthesiology data is described in Section 5. We conclude the article with a brief discussion in Section 6 and defer the proofs to Appendices A and B.

## 2. New estimation procedure

For each $i \in \{1,\ldots,n\}$, let $\mathbf{y}_i = (y_{i1},\ldots,y_{im_i})^\top$ be the repeated measurements of the $i$th subject which are observed at irregular time points $\mathbf{t}_i = (t_{i1},\ldots,t_{im_i})^\top$ where $n$ is the total number subjects and $m_i$ is the number of repeated measures for $i$th subject. The design matrix of each subject is denoted by $\mathbf{x}_i$, with size $m_i \times p$, and could have a column of 1s if an intercept term is desired. By allowing $m_i$ to be subject specific, our framework is valid for unbalanced longitudinal data. Assume that for each $i \in \{1,\ldots,n\}$, the response vector $\mathbf{y}_i$ follows a multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ with

$$\boldsymbol{\mu}_i = \mathrm{E}(\mathbf{y}_i|\mathbf{x}_i) = \mathbf{x}_i\boldsymbol{\beta}, \tag{1}$$

where $\boldsymbol{\mu} = (\mu_{i1},\ldots,\mu_{im_i})^\top$ and $\boldsymbol{\Sigma}_i$ is an $m_i \times m_i$ positive definite covariance matrix.

A commonly seen joint mean–covariance model would set the mean structure as (1), and model the corresponding covariance matrix as $\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})$, where $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ are the dependence parameter and the innovation variance parameter, with dimension $q$ and $d$, respectively. The distinction among different joint mean–covariance models mainly lies in the decomposition methods of covariance matrices and the interpretation of the within-subject correlation mechanisms. We will give more detailed discussion about the joint mean–covariance modeling of $\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})$ in Section 2.2.

Parameterized by the natural parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top,\boldsymbol{\gamma}^\top,\boldsymbol{\lambda}^\top)^\top \in \Theta \subset \mathbb{R}^{p+q+d}$, the density function of the $i$th subject can be written as

$$f(\mathbf{y}_i;\boldsymbol{\theta}) = \frac{1}{(2\pi)^{m_i/2}|\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})|^{1/2}} \exp\{-(\mathbf{y}_i-\mathbf{x}_i\boldsymbol{\beta})^\top\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})^{-1}(\mathbf{y}_i-\mathbf{x}_i\boldsymbol{\beta})/2\},$$

and so the log-likelihood of a joint mean–covariance model is

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \ln f(\mathbf{y}_i;\boldsymbol{\theta}) \propto -\frac{1}{2}\sum_{i=1}^n \{\ln|\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})| + (\mathbf{y}_i-\mathbf{x}_i\boldsymbol{\beta})^\top\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})^{-1}(\mathbf{y}_i-\mathbf{x}_i\boldsymbol{\beta})\},$$

and the MLE of a joint mean–covariance model is defined as

$$\hat{\boldsymbol{\theta}}_{\mathrm{MLE}} = \arg\max_{\boldsymbol{\theta}\in\Theta} \ell(\boldsymbol{\theta}).$$

However, since the classic MLE assigns equal weights to all data points, MLE-based joint mean–covariance model is sensitive to outliers, contaminations, and heavy-tailed distributions.

To overcome the drawback of MLE, we propose the following $L_q$-likelihood [4] for a joint mean–covariance model to provide a robust and flexible estimation methodology in longitudinal data framework. Let

$$\mathcal{L}q(\boldsymbol{\theta}) = \sum_{i=1}^n Lq\{f(\mathbf{y}_i;\boldsymbol{\theta})\} \propto \sum_{i=1}^n \frac{1}{|\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})|^{(1-q)/2}} \exp\{-(1-q)(\mathbf{y}_i-\mathbf{x}_i\boldsymbol{\beta})^\top\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})^{-1}(\mathbf{y}_i-\mathbf{x}_i\boldsymbol{\beta})/2\}, \tag{2}$$

where $Lq(u) = (u^{1-q}-1)/(1-q)$ and $q \in (0,1)$ is a tuning parameter that goes to 1 when $n \to \infty$. For the simplicity of notation, we omit the dependence of $q$ on $n$ when there is no confusion. Therefore, given a pre-chosen tuning parameter $q$, the maximum $L_q$-likelihood estimator (MLqE) of the model parameters $\boldsymbol{\theta}$ is defined as

$$\tilde{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\Theta} \mathcal{L}q(\boldsymbol{\theta}) \tag{3}$$

and the score equation based on $L_q$-likelihood $\mathbf{U}^*_{\boldsymbol{\theta}}(\mathbf{y}_i;\boldsymbol{\theta})$ can be obtained as

$$\mathbf{U}^*_{\boldsymbol{\theta}}(\mathbf{y}_i;\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}\,[Lq\{f(\mathbf{y}_i;\boldsymbol{\theta})\}] = f(\mathbf{y}_i;\boldsymbol{\theta})^{-q}\nabla_{\boldsymbol{\theta}} f(\mathbf{y}_i;\boldsymbol{\theta}) = f(\mathbf{y}_i;\boldsymbol{\theta})^{1-q}\mathbf{U}_{\boldsymbol{\theta}}(\mathbf{y}_i;\boldsymbol{\theta}), \tag{4}$$

where $\mathbf{U}_{\boldsymbol{\theta}}(\mathbf{y}_i;\boldsymbol{\theta})$ is the score equation of the MLE for the $i$th subject, viz.

$$\mathbf{U}_{\boldsymbol{\theta}}(\mathbf{y}_i;\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \ln f(\mathbf{y}_i;\boldsymbol{\theta}) = f(\mathbf{y}_i;\boldsymbol{\theta})^{-1}\nabla_{\boldsymbol{\theta}} f(\mathbf{y}_i;\boldsymbol{\theta}).$$

Specifically, the score equation for the regression coefficients $\boldsymbol{\beta}$ is

$$\mathbf{U}_{\beta}(\mathbf{y}_i;\boldsymbol{\theta}) = \nabla_{\beta} \ln f(\mathbf{y}_i;\boldsymbol{\theta}) = \mathbf{x}_i^{\top} \boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})^{-1}(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\beta}),$$

and the score equations for $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ depend on the specific joint mean–covariance modeling of $\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})$.

The robustness of the proposed maximum $L_q$-likelihood estimation stems from the following three points:

(i) Based on (4), the $L_q$-likelihood score equation can be considered as a weighted version of the MLE score equation. Since the weight $w_i = f(\mathbf{y}_i;\boldsymbol{\theta})^{1-q}$ of the $i$th subject is proportional to its density, it can reduce the impact of outliers with low density, which makes the new estimation procedure robust.

(ii) The tuning parameter $q$ governs the sensitivity of the estimator against outliers [15]. The smaller $q$ is, the more robust the MLqE is to outliers or heavy-tailed distributions. Please see [15] for more details.

(iii) The conventional maximum likelihood method can be considered as a special case of the MLqE if taking $q = 1$. Therefore, the conventional likelihood assigns a weight $w_i = 1$ for each observation and thus cannot decrease the effects of outliers.

Please refer to the comments after Algorithm 1 for more explanations.

### 2.1. Computation algorithm

Note that there is no explicit solution to (3). Maximizing (3) is equivalent to maximizing $\ln\{\sum_{i=1}^{n} f^{1-q}(\mathbf{y}_i;\boldsymbol{\theta})\}$, which has a similar structure to a log-density of a mixture model with $n$ components and $f^{1-q}(\mathbf{y}_i;\boldsymbol{\theta})$ mimicking the $i$th component density. Therefore, we can adapt the EM algorithm for mixture models to simplify the computation of (3). Inspired by [15,21], we propose the following modified modal EM algorithm (MMEM) to compute the maximum $L_q$-likelihood estimator.

**Algorithm 1.** Given an initial value $\boldsymbol{\theta}^{(0)}$, start with $k = 0$.

E-Step: Update $P_i(\boldsymbol{\theta}^{(k)})$, viz.

$$P_i(\boldsymbol{\theta}^{(k)}) = \frac{f(\mathbf{y}_i;\boldsymbol{\theta}^{(k)})^{(1-q)}}{\sum_{i=1}^{n} f(\mathbf{y}_i;\boldsymbol{\theta}^{(k)})^{(1-q)}} \propto |\boldsymbol{\Sigma}_i(\boldsymbol{\gamma}^{(k)},\boldsymbol{\lambda}^{(k)})|^{-(1-q)/2} \exp\{-(1-q)(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\beta}^{(k)})^{\top}\boldsymbol{\Sigma}_i(\boldsymbol{\gamma}^{(k)},\boldsymbol{\lambda}^{(k)})^{-1}(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\beta}^{(k)})/2\}.$$

M-Step: Update $\boldsymbol{\theta}^{(k+1)}$, viz.

$$\begin{aligned}(\boldsymbol{\gamma}^{(k+1)},\boldsymbol{\lambda}^{(k+1)}) &= \arg\max_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{n} \left\{P_i(\boldsymbol{\theta}^{(k)}) \ln f(\mathbf{y}_i;\boldsymbol{\theta})\right\} \qquad (5)\\ &= \arg\min_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{n} P_i(\boldsymbol{\theta}^{(k)})\{\ln|\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})| + (\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\beta}^{(k)})^{\top}\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})^{-1}(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\beta}^{(k)})\},\\ \tilde{\boldsymbol{\beta}}^{(k+1)} &= \left\{\sum_{i=1}^{n} P_i(\boldsymbol{\theta}^{(k+1)})\mathbf{x}_i^{\top}\tilde{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{\gamma}^{(k+1)},\boldsymbol{\lambda}^{(k+1)})\mathbf{x}_i\right\}^{-1} \sum_{i=1}^{n} P_i(\boldsymbol{\theta}^{(k+1)})\mathbf{x}_i^{\top}\tilde{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{\gamma}^{(k+1)},\boldsymbol{\lambda}^{(k+1)})y_i.\end{aligned}$$

The above algorithm is in fact a variant of the generalized modal EM algorithm proposed by [8] and further extended by [20]. The proposed MMEM achieves robustness in the regression by the weights $P_i$ calculated in the E-step. Since $P_i$ is proportional to the density $f(\mathbf{y}_i;\boldsymbol{\theta})$, the weighted log-likelihood (5) ensures that observations in the tails of those heavy-tailed distributions or outliers have less impact on the objective function than the majority of the data, thereby guaranteeing the robustness of the MLqE. Note that in the M-step, there are no explicit solutions to $\boldsymbol{\gamma}^{(k+1)}$ and $\boldsymbol{\lambda}^{(k+1)}$, and so some numerical methods need to be applied. Once $\tilde{\boldsymbol{\Sigma}}_i(\boldsymbol{\gamma}^{(k+1)},\boldsymbol{\lambda}^{(k+1)})$ is updated, the estimator of $\boldsymbol{\beta}$ can be updated by a weighted least squares with weights proportional to $P_i$'s.

The following theorem presents the monotonicity of Algorithm 1.

**Theorem 1.** *The objective function (2) is non-decreasing after each iteration of the modified modal EM algorithm, i.e.,*

$$\sum_{i=1}^{n} f\{\mathbf{y}_i;\boldsymbol{\theta}^{(k+1)}\}^{1-q} \geq \sum_{i=1}^{n} f\{\mathbf{y}_i;\boldsymbol{\theta}^{(k)}\}^{1-q}$$

*until a fixed point is reached.*

**Remark 1.** Similar to the usual EM algorithm, the value to which the algorithm converges may depend on the starting values, and there is no guarantee that the algorithm converges to the global optimum. Thus, initiating the algorithm from different starting values and then choosing the best local optimal solution is advised.

### 2.2. Joint mean–covariance modeling approaches

In order to compute (2), we need to specify a working covariance model for $\boldsymbol{\Sigma}_i(\boldsymbol{\gamma}, \boldsymbol{\lambda})$. Note that the working covariance model of $\boldsymbol{\Sigma}_i(\boldsymbol{\gamma}, \boldsymbol{\lambda})$ will not affect the consistency of the estimation of regression parameters, but a better specified covariance model could improve the efficiency of the regression parameter estimates. One simple and traditional way is to use the static covariance models such as AR, MA, or exchangeable structures. However, when the within-subject correlation changes dynamically according to time or location, traditional static covariation patterns can hardly satisfy the positive-definiteness of a covariance matrix for high-dimensional data or flexibly depict the dynamic dependence structure.

In the last decade, parsimonious models for characterizing the dependence structure among repeated measurements have attracted increasing attention to better reveal how the within-subject correlations depend on time and other predictors. Pourahmadi [12] proposed to model dynamically the covariance matrices by using a modified Cholesky decomposition; see, e.g., [1,6,11,14]. An attractive aspect of such a decomposition is that the entries in the decomposition have autoregressive and log innovation interpretations [24]. By applying hyperspherical coordinates, Zhang et al. [25] proposed a novel dynamic variance-correlation, in which the parsimonious covariance models are more flexible and adaptive than those that only specify the correlation structure, e.g., an AR, MA, or exchangeable structure. Thus, it is expected that more efficient maximum likelihood or GEE estimators of the mean regression parameters could be obtained by using these estimated dynamical covariance matrices than using the estimated ones under specified correlation structure [13,22].

The modified Cholesky decomposition factor (MCDF) based covariance model, originated from [12,13], is a commonly used joint mean–covariance modeling approach, which provides an unconstrained parameterization for the covariance matrix via modeling the autoregressive parameters and the innovation variances through covariates. Let $\mathbf{T}_i\boldsymbol{\Sigma}_i\mathbf{T}_i^\top = \mathbf{D}_i$, where $\mathbf{T}_i$ is a lower unitriangular matrix with main diagonal 1s and the $(j,k)$th below diagonal entry being $-\phi_{ijk}$, and $\phi_{ijk}$ is the autoregressive coefficients in the autoregressive model defined, for all $j \in \{1,\ldots,m_i\}$, by

$$y_{ij} - \mu_{ij} = \sum_{k=1}^{j-1} \phi_{ijk}(y_{ik} - \mu_{ik}) + \epsilon_{ij}.$$

Here, $\mathbf{D}_i = \mathrm{diag}(\sigma_{i1}^2, \ldots, \sigma_{im_i}^2)$, where $\sigma_{ij}^2$ is the innovation variance $\sigma_{ij}^2 = \mathrm{var}(\epsilon_{ij})$. Throughout the article, the summation $\Sigma_{k=1}^{0}$ is defined to be zero. Define

$$\phi_{ijk} = \mathbf{z}_{ijk}^\top \boldsymbol{\gamma}, \quad \ln(\sigma_{ij}^2) = \mathbf{h}_{ij}^\top \boldsymbol{\lambda}, \tag{6}$$

where $\mathbf{z}_{ijk}$ and $\mathbf{h}_{ij}$ are $b \times 1$ and $d \times 1$ vectors of covariates, and $\mathbf{z}_{ijk}$ is commonly assumed to be a polynomial function of time differences $t_{ik} - t_{ij}$ with $k > j$. Refer to [13] for the algorithms to compute the MLEs $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\lambda}}$, and then $\hat{\mathbf{T}}_i$ and $\hat{\mathbf{D}}_i$ could be computed accordingly. As a result, the estimator of $\boldsymbol{\Sigma}_i$ could be obtained simply as $\hat{\boldsymbol{\Sigma}}_i^{MCDF} = \hat{\mathbf{T}}_i^{-1}\hat{\mathbf{D}}_i(\hat{\mathbf{T}}_i^{-1})^\top$. It is noteworthy that traditional static covariation patterns such as AR, MA, or exchangeable structures are just some special cases of MCDF. For example, the error has the independence correlation structure if the $\phi_{ijk}$s are all zero, and has an AR(1) correlation structure if $\phi_{ijk}$ is zero for $j - k \geq 2$.

Zhang and Leng [24] transformed MCDF modeling into the moving average Cholesky factor modeling (MACF) by parameterizing covariance structures as $\boldsymbol{\Sigma}_i = \mathbf{L}_i\mathbf{D}_i\mathbf{L}_i^\top$. Here $\mathbf{L}_i = (\phi_{ijk})$ is a lower unitriangular matrix, where $\phi_{ijk}$ is the moving average parameter in

$$y_{ij} - \mu_{ij} = \sum_{k=1}^{j-1} \phi_{ijk}\varepsilon_{ik} + \varepsilon_{ij},$$

where $j \in \{1,\ldots,m_i\}$ with $\varepsilon_{i1} = y_{i1} - \mu_{i1}$ and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{im_i})^\top \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_i)$. Here $\mathbf{D}_i = \mathrm{diag}(\sigma_{i1}^2, \ldots, \sigma_{im_i}^2)$, where $\sigma_{ij}^2$ is the innovation variance $\sigma_{ij}^2 = \mathrm{var}(\varepsilon_{ij})$. Similar to MCDF, Zhang and Leng [24] also suggested to model $\phi_{ijk}$ and $\sigma_{ij}^2$ as (6). Due to the similarity in both decomposition methods, we omit MACF henceforth.

Zhang et al. [25] innovated by proposing a new parameterization of the correlation matrix of MACF to a hyperspherical coordinates factor modeling. Specifically, write $\boldsymbol{\Sigma}_i = \boldsymbol{\Gamma}_i\mathbf{R}_i\boldsymbol{\Gamma}_i$, where $\boldsymbol{\Gamma}_i = \mathrm{diag}(\sigma_{i1}, \ldots, \sigma_{im_i})$ with $\sigma_{ij}$ being the standard deviation of $y_{ij}$ and $\mathbf{R}_i = (\rho_{ijk})_{j,k=1}^{m_i}$ is the correlation matrix of $\mathbf{y}_i$. Applying hyperspherical coordinates and trigonometric functions, the correlation matrix $\mathbf{R}_i$ is parameterized as $\mathbf{R}_i = \mathbf{C}_i\mathbf{C}_i^\top$, where $\mathbf{C}_i$ is a lower triangular matrix, viz.

$$\mathbf{C}_i = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ c_{i21} & s_{i21} & 0 & \cdots & 0 \\ c_{i31} & c_{i32}s_{i31} & s_{i32}s_{i31} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{im_i1} & c_{im_i2}s_{im_i1} & c_{im_i3}s_{m_i2}s_{im_i1} & \cdots & \prod_{l=1}^{m_i-1} s_{im_il} \end{pmatrix}_{m_i \times m_i},$$

with $c_{ijk} = \cos(\phi_{ijk})$ and $s_{ijk} = \sin(\phi_{ijk})$. Here, $\phi_{ijk}$ is a function of the correlation parameters and could be interpreted as an angle; for details, see [25]. Furthermore, $\phi_{ijk}$ and $\sigma_{ij}^2$ are also modeled as (6). The detailed algorithm for finding the MLE $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\lambda}}$, and therefore $\hat{\boldsymbol{\Sigma}}_i$ could be found in [25]. Hereinafter, we call this hyperspherical coordinates factor modeling and the estimating procedure as HSCF.

## 3. Asymptotics

In this section, we investigate the asymptotic properties of the MLqE estimators for joint mean–covariance models. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{\lambda}^\top)^\top \in \boldsymbol{\Theta} \subset \mathbb{R}^{p+q+d}$ and $\mathbf{I}(\boldsymbol{\theta}) = -\mathrm{E}\partial^2\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$. To establish formally the theoretical properties of the MLqE, we impose the following assumptions.

(i) The dimensions of $\mathbf{x}_{ij}$, $\mathbf{z}_{ij}$, $\mathbf{h}_{ij}$, namely $p$, $q$ and $d$, are fixed and $\max(m_1, \ldots, m_n)$ is bounded. The parametric space $\boldsymbol{\Theta}$ is a compact subset of $\mathbb{R}^{p+q+d}$, and the true parameter value $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0^\top, \boldsymbol{\gamma}_0^\top, \boldsymbol{\lambda}_0^\top)^\top$ is in the interior of $\boldsymbol{\Theta}$. Furthermore, $\mathbf{I}(\boldsymbol{\theta}_0)/n$ converges to a positive definite matrix $\mathcal{I}(\boldsymbol{\theta}_0)$ as $n \to \infty$.

(ii) $\mathrm{E}_{\boldsymbol{\theta}_0} \sup_{\boldsymbol{\theta}\in\Theta} \|\mathbf{U}_{\boldsymbol{\beta}}(\mathbf{y}_i; \boldsymbol{\theta}, q_n)\|^2 < \infty$ for all $i \in \{1, \ldots, n\}$. Furthermore, $\mathrm{E}_{\boldsymbol{\theta}_0} \sup_{\boldsymbol{\theta}\in\Theta}[\{\sum_{i=1}^n f(\mathbf{y}_i; \boldsymbol{\theta})^\delta - 1\}^2/n] \to 0$ as $\delta \to 0$ and the distortion parameter $q_n > 0$ is a sequence such that $q_n \to 1$ as $n \to \infty$.

(iii) As $q_n \to 1$, the second-order partial derivatives $\sum_{i=1}^n \nabla_{\boldsymbol{\beta}}^2 \mathbf{U}_{\boldsymbol{\beta}}^*(\mathbf{y}_i; \boldsymbol{\theta}, q_n)/n$ are dominated by an integrable function in a neighborhood of $\boldsymbol{\beta}_0$. The smallest eigenvalue of $\sum_{i=1}^n \mathrm{E}_{\boldsymbol{\theta}_0} \mathbf{U}_{\boldsymbol{\beta}}^*(\mathbf{y}_i; \boldsymbol{\theta}, q_n)\mathbf{U}_{\boldsymbol{\beta}}^*(\mathbf{y}_i; \boldsymbol{\theta}, q_n)^\top/n$ is bounded from zero. Furthermore, for all $k, \ell \in \{1, \ldots, p\}$,

$$\left\{\frac{1}{n}\sum_{i=1}^n \mathrm{E}_{\boldsymbol{\theta}_0}\mathbf{I}_{\boldsymbol{\beta}}^*(\mathbf{y}_i; \boldsymbol{\theta}, q_n)\right\}_{k\ell}^2 = \left\{\frac{1}{n}\sum_{i=1}^n \mathrm{E}_{\boldsymbol{\theta}_0} \nabla_{\boldsymbol{\beta}}^2 Lq_n(f(\mathbf{y}_i; \boldsymbol{\theta}))\right\}_{k\ell}^2$$

are bounded from above by a constant.

**Remark 2.** Assumption (i) is conventional for the theoretical analysis of the MLE approach in the longitudinal data analysis framework; see Chapter 2 in [25]. Assumptions (ii) and (iii) are natural requirements for the maximum $L_q$-likelihood estimation within longitudinal data modeling; see Chapters 3–4 in [4].

**Theorem 2.** *Suppose that Assumptions (i)–(ii) are satisfied. Then, as $n \to \infty$, the maximum $L_q$-likelihood estimator based on a joint mean–covariance modeling, denoted by $\tilde{\boldsymbol{\beta}}$, is consistent. That is, $\tilde{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$, and*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \mathrm{E}_{\boldsymbol{\theta}_0} \nabla_{\boldsymbol{\beta}} \ln(f(\mathbf{y}_i; \boldsymbol{\theta}))\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \mathrm{E}_{\boldsymbol{\theta}_0} \mathbf{U}_{\boldsymbol{\beta}}(\mathbf{y}_i; \boldsymbol{\theta})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = \mathbf{0}_p. \tag{7}$$

**Theorem 3.** *Suppose that Assumptions (i)–(iii) are satisfied, and $\boldsymbol{\beta}^* \to \boldsymbol{\beta}_0$ as $n \to \infty$, where $\boldsymbol{\beta}^*$ is the vector such that $\mathrm{E}_{\boldsymbol{\theta}_0}\mathbf{U}_{\boldsymbol{\beta}}^*(\mathbf{y}_i; \boldsymbol{\theta}, q_n)|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*} = \mathbf{0}_p$. Then the solution of the MLqE equation $\tilde{\boldsymbol{\beta}}$ is asymptotically normally distributed as*

$$\sqrt{n}\mathbf{V}_n^{-1/2}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) \rightsquigarrow \mathcal{N}(\mathbf{0}_p, \mathbf{I}_p),$$

*where*

$$\mathbf{V}_n = \left\{\frac{1}{n}\sum_{i=1}^n \mathrm{E}_{\boldsymbol{\theta}_0}\mathbf{I}_{\boldsymbol{\beta}}^*(\mathbf{y}_i; \boldsymbol{\theta}, q_n)\right\}^{-1} \left\{\frac{1}{n}\sum_{i=1}^n \mathrm{E}_{\boldsymbol{\theta}_0}\mathbf{U}_{\boldsymbol{\beta}}^*(\mathbf{y}_i; \boldsymbol{\theta}, q_n)\mathbf{U}_{\boldsymbol{\beta}}^*(\mathbf{y}_i; \boldsymbol{\theta}, q_n)^\top\right\} \left\{\frac{1}{n}\sum_{i=1}^n \mathrm{E}_{\boldsymbol{\theta}_0}\mathbf{I}_{\boldsymbol{\beta}}^*(\mathbf{y}_i; \boldsymbol{\theta}, q_n)\right\}^{-1}\Bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}$$

Note that the score equation (4) is in general biased for each fixed $q < 1$ due to the weight function $f(\mathbf{y}_i; \boldsymbol{\theta})^{1-q}$. Instead, the solution of the MLqE equation $\tilde{\boldsymbol{\beta}}$ is an unbiased estimate for $\boldsymbol{\beta}^*$, which converges to the true value $\boldsymbol{\beta}_0$ when $q \to 1$. In addition, the above result holds even when the covariance matrix is misspecified. If the covariance matrix is correctly specified, based on results in Appendices A and B, the asymptotic variance of $\sqrt{n}\,\tilde{\boldsymbol{\beta}}$ is $\mathbf{V}_n \xrightarrow{p} \mathbf{V}$, where

$$\mathbf{V} = \left\{\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \mathrm{E}_{\boldsymbol{\theta}_0}\mathbf{x}_i^\top \boldsymbol{\Sigma}_i(\boldsymbol{\gamma}, \boldsymbol{\lambda})^{-1}\mathbf{x}_i\right\}^{-1},$$

which is the same as the asymptotic variance of MLE.

## 4. Simulation studies

In this section, we conduct simulation studies to investigate the performance of the MLqE, and compare it with the MLE in joint mean–covariance modeling. The main objectives are to:

**Table 1**
The accuracy of the estimated coefficients in terms of MAB, SD and MSE for 1000 parameter estimates for Example 1. All the results are multiplied by a factor of $10^3$.

| Sample size | | $n = 100$ | | | $n = 200$ | | | $n = 300$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Method | | *MAB* | *SD* | *MSE* | *MAB* | *SD* | *MSE* | *MAB* | *SD* | *MSE* |
| | | Covariance structure: MCDF model | | | | | | | | |
| MLE<br>(MCDF) | $\beta_0$ | 43.75 | 54.04 | 2.93 | 30.01 | 37.86 | 1.44 | 24.83 | 31.05 | 0.97 |
| | $\beta_1$ | 39.68 | 49.20 | 2.45 | 26.03 | 32.66 | 1.08 | 21.97 | 27.71 | 0.77 |
| | $\beta_2$ | 38.59 | 48.46 | 2.37 | 26.41 | 32.90 | 1.10 | 21.45 | 26.87 | 0.72 |
| MLqE<br>(MCDF) | $\beta_0$ | 43.75 | 54.05 | 2.93 | 30.00 | 37.86 | 1.44 | 24.83 | 31.05 | 0.97 |
| | $\beta_1$ | 39.65 | 49.19 | 2.45 | 26.02 | 32.66 | 1.08 | 21.97 | 27.71 | 0.77 |
| | $\beta_2$ | 38.58 | 48.46 | 2.37 | 26.41 | 32.90 | 1.10 | 21.45 | 26.88 | 0.72 |
| | | Covariance structure: HSCF model | | | | | | | | |
| MLE<br>(HSCF) | $\beta_0$ | 1.12 | 1.45 | 0.0021 | 0.85 | 1.10 | 0.0012 | 0.75 | 0.98 | 0.0009 |
| | $\beta_1$ | 0.41 | 0.55 | 0.0003 | 0.31 | 0.43 | 0.0002 | 0.28 | 0.38 | 0.0001 |
| | $\beta_2$ | 0.41 | 0.54 | 0.0003 | 0.32 | 0.43 | 0.0002 | 0.28 | 0.37 | 0.0001 |
| MLqE<br>(HSCF) | $\beta_0$ | 1.42 | 8.31 | 0.0690 | 0.85 | 1.30 | 0.0017 | 0.77 | 1.56 | 0.0024 |
| | $\beta_1$ | 0.56 | 3.68 | 0.0136 | 0.38 | 1.78 | 0.0032 | 0.28 | 0.45 | 0.0002 |
| | $\beta_2$ | 0.50 | 2.71 | 0.0073 | 0.38 | 1.85 | 0.0034 | 0.28 | 0.48 | 0.0002 |

(i) compare the estimation efficiency of the MLqE and the MLE when data sets follow multivariate normal distributions and the assumptions of joint mean–covariance models;

(ii) evaluate the robustness of the MLqE when data sets are contaminated by outliers;

(iii) demonstrate the model fitting ability of the MLqE and the MLE for joint mean–covariance modeling when data sets are generated from some heavy-tailed distributions.

Throughout the study, $\boldsymbol{\beta} = (1, -0.5, 0.5)$ is used as the vector of regression coefficients and the corresponding covariate is $\mathbf{x}_{ij} = (1, x_{ij1}, x_{ij2})^\top$, where $(x_{ij1}, x_{ij2})^\top$ is generated from a multivariate normal distribution with mean zero, marginal variance 1 and bivariate exchange correlation structure with $\rho = 0.5$. Each subject is measured $m_i$ times with $m_i = \max(1, \mathcal{B}(10, 0.8))$; measurement times $t_{i1}, \ldots, t_{im_i}$ are generated from $\mathcal{U}(0, 1)$, where $\mathcal{B}(n, p)$ is a binomial distribution with $n$ experiments and a success probability $p$.

Note that the number of repeated measurements is subject specific and therefore, the measurements are allowed to be observed at irregular times. Thus the data structure is unbalanced. Quadratic polynomials are applied to time differences as $\mathbf{z}_{ijk} = (1, (t_{ij} - t_{ik}), (t_{ij} - t_{ik})^2)^\top$, and $\boldsymbol{\gamma} = (0.3, -0.2, 0.3)$ and $\boldsymbol{\lambda} = (1, 0.5, 0.25)$ are used. The covariate for the log innovation structure is taken to be $\mathbf{h}_{ij} = (1, h_{ij1}, h_{ij2})^\top$, where $(h_{ij1}, h_{ij2})^\top$ is generated from the same multivariate normal distribution as $(x_{ij1}, x_{ij2})^\top$. We consider the covariance structure of $\boldsymbol{\Sigma}_i(\boldsymbol{\gamma}, \boldsymbol{\lambda})$ based on both MCDF and HSCF.

For each generated data, MCDF and HSCF are applied to model the covariance structure, and both the MLE and the MLqE are used to estimate $\boldsymbol{\theta}$. In our study, we use linear regression estimates as initial values. To assess the accuracy of regression parameter estimates $\tilde{\boldsymbol{\beta}}$, sample standard deviation (SD), mean absolute bias (MAB $=$ average$(|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0|)$) and mean square errors (MSE $=$ average$\{(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^2\}$) are reported. Sample sizes $n \in \{100, 200, 300\}$ are used over 1,000 repetitions.

The tuning parameter $q_n$ governs the sensitivity of the estimator against outliers/heavy-tailed distributions. The smaller $q_n$, the less sensitive the MLqE is to outliers/heavy-tailed distributions. So far, there is no good universal way to choose the tuning parameter $q_n$. Usually, when the unusual observations become a concern (i.e., extreme outliers in the data set or many observations from the tails of a heavy-tailed distribution), a small $q$ should be used to protect against the outliers. In our numerical studies, we propose to select the tuning parameter $q_n$ empirically by a five-fold cross validation method. The tuning parameter $q_n$ ranges from 0.7 to 1 and we take the one that minimizes the average prediction errors. Our numerical studies demonstrate the effectiveness of such selection method.

### 4.1. Models

**Example 1** (*Multivariate Normal Distribution*)**.** We first consider the performance of the MLqE when the errors are multivariate normal. We conduct a simulation study with the model defined, for $i \in \{1, \ldots, n\}$ and $n_j \in \{1, \ldots, m_i\}$, by

$$y_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + e_{ij}, \tag{8}$$

where $e_i = (e_{i1}, \ldots, e_{im_i})^\top$ has mean $(0, \ldots, 0)^\top$ and MCDF or HSCF is used as the covariance structure. Both the MLE and the MLqE are employed to estimate the regression parameters, and Table 1 reports the MAB, SD and MSE of both estimates.

**Example 2** (*Outliers*)**.** In this example, we demonstrate the advantage of the maximum $L_q$-likelihood estimator over the MLE in data sets with outliers. Outliers could be some extreme values in the population or bad data points caused by measurement

**Table 2**
The accuracy of the estimated coefficients in terms of MAB, SD and MSE for 1000 parameter estimates for Example 2. All the results are multiplied by a factor of $10^3$.

| Sample size | | $n = 100$ | | | $n = 200$ | | | $n = 300$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Method | | *MAB* | *SD* | *MSE* | *MAB* | *SD* | *MSE* | *MAB* | *SD* | *MSE* |
| | | Covariance structure: MCDF model | | | | | | | | |
| MLE<br>(MCDF) | $\beta_0$ | 526.46 | 163.73 | 303.94 | 595.87 | 113.91 | 368.03 | 617.31 | 97.46 | 390.56 |
| | $\beta_1$ | 66.81 | 84.76 | 7.18 | 48.06 | 61.13 | 3.74 | 39.00 | 48.51 | 2.35 |
| | $\beta_2$ | 65.18 | 82.15 | 6.75 | 49.06 | 61.81 | 3.82 | 40.31 | 50.62 | 2.56 |
| MLqE<br>(MCDF) | $\beta_0$ | 80.52 | 118.67 | 14.20 | 292.12 | 271.47 | 139.97 | 514.40 | 160.69 | 285.93 |
| | $\beta_1$ | 44.97 | 56.96 | 3.25 | 42.32 | 53.31 | 2.85 | 36.77 | 45.77 | 2.09 |
| | $\beta_2$ | 45.68 | 57.69 | 3.33 | 41.63 | 52.61 | 2.77 | 38.22 | 48.05 | 2.31 |
| | | Covariance structure: HSCF model | | | | | | | | |
| MLE<br>(HSCF) | $\beta_0$ | 565.78 | 235.55 | 372.05 | 614.58 | 157.06 | 402.34 | 632.60 | 119.70 | 414.50 |
| | $\beta_1$ | 22.94 | 29.30 | 0.86 | 17.55 | 22.01 | 0.48 | 15.33 | 19.21 | 0.37 |
| | $\beta_2$ | 24.58 | 31.20 | 0.97 | 17.35 | 22.18 | 0.49 | 14.59 | 18.50 | 0.34 |
| MLqE<br>(HSCF) | $\beta_0$ | 1.13 | 1.54 | 0.0024 | 0.86 | 1.15 | 0.0013 | 1.61 | 21.19 | 0.4493 |
| | $\beta_1$ | 0.44 | 0.58 | 0.0003 | 0.31 | 0.41 | 0.0002 | 0.44 | 4.25 | 0.0181 |
| | $\beta_2$ | 0.42 | 0.56 | 0.0003 | 0.31 | 0.41 | 0.0002 | 0.48 | 5.60 | 0.0313 |

**Table 3**
The accuracy of the estimated coefficients in terms of MAB, SD and MSE for 1000 parameter estimates for Example 3. All the results are multiplied by a factor of $10^3$.

| Method | | *MAB* | *SD* | *MSE* | *MAB* | *SD* | *MSE* | *MAB* | *SD* | *MSE* |
|---|---|---|---|---|---|---|---|---|---|---|
| Sample size | | $n = 100$ | | | $n = 200$ | | | $n = 300$ | | |
| | | Multivariate Laplace distribution $\Sigma_L = \Sigma^{(MCDF)}$ | | | | | | | | |
| MLE<br>(MCDF) | $\beta_0$ | 42.57 | 53.71 | 2.88 | 30.27 | 38.06 | 1.45 | 23.72 | 30.23 | 0.91 |
| | $\beta_1$ | 39.31 | 49.19 | 2.46 | 27.30 | 34.21 | 1.17 | 21.77 | 27.28 | 0.75 |
| | $\beta_2$ | 37.73 | 47.21 | 2.26 | 26.79 | 34.04 | 1.16 | 21.78 | 27.14 | 0.75 |
| MLqE<br>(MCDF) | $\beta_0$ | 41.61 | 52.48 | 2.75 | 29.92 | 37.64 | 1.42 | 23.53 | 30.00 | 0.90 |
| | $\beta_1$ | 38.49 | 48.18 | 2.36 | 26.95 | 33.77 | 1.14 | 21.61 | 27.07 | 0.74 |
| | $\beta_2$ | 36.87 | 46.17 | 2.16 | 26.48 | 33.65 | 1.13 | 21.62 | 26.94 | 0.74 |
| | | Multivariate Laplace distribution $\Sigma_L = \Sigma^{(HSCF)}$ | | | | | | | | |
| MLE<br>(HSCF) | $\beta_0$ | 8.74 | 16.44 | 0.2701 | 5.27 | 9.96 | 0.0992 | 3.61 | 6.90 | 0.0476 |
| | $\beta_1$ | 3.33 | 5.07 | 0.0312 | 1.91 | 3.14 | 0.0118 | 1.33 | 2.24 | 0.0059 |
| | $\beta_2$ | 3.50 | 5.56 | 0.0375 | 1.94 | 3.41 | 0.0135 | 1.26 | 2.04 | 0.0048 |
| MLqE<br>(HSCF) | $\beta_0$ | 0.94 | 1.29 | 0.0017 | 0.70 | 0.93 | 0.0009 | 0.64 | 0.86 | 0.0007 |
| | $\beta_1$ | 0.35 | 0.47 | 0.0002 | 0.27 | 0.36 | 0.0001 | 0.26 | 0.71 | 0.0005 |
| | $\beta_2$ | 0.36 | 0.48 | 0.0002 | 0.28 | 0.37 | 0.0001 | 0.24 | 0.33 | 0.0001 |

errors in the data collection process. A few outliers might cause biased estimates and even misleading inference for the MLE approach when our interest is to build the model/relationship for the majority of the population.

To be more specific, 97% data are generated from model (8), and the other 3% are generated from the gross error model defined, for all $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m_i\}$, by

$$y_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + e_{ij} + \xi_{ij},$$

where $\xi_i = (\xi_{i1}, \ldots, \xi_{im_i})^\top$ are sampled from $\mathcal{U}(20, 25)$. Table 2 summarizes the simulation results.

**Example 3** (*Heavy-Tailed Distribution*)**.** In this example, we compare the performance of the new estimate with the MLE when the error has a heavy-tailed distribution. With an appropriate pre-determined distortion parameter $q < 1$, the MLqE could give less weight to data points in the tail of the multivariate distribution, and more to data points near the center of the distribution [15].

For illustration purposes, we consider the multivariate Laplace distribution with MCDF and HSCF covariance structure as the error distributions. The results are given in Table 3.

### 4.2. *Summary of simulation results*

Based on the simulation results listed in Tables 1–3, we can draw the following conclusions:

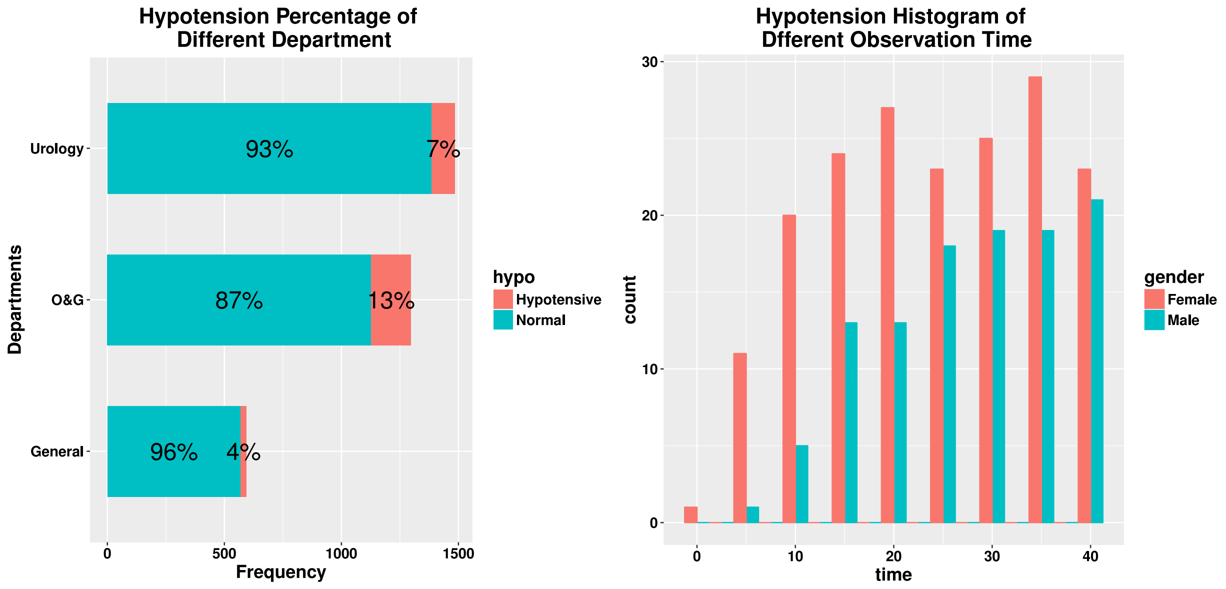

**Fig. 1.** Frequency description of hypotension.

**Table 4**
Comparison between the classic MLE and the proposed MLqE regression methods based on MCDF and HSCF for the Spinal anesthesia data set.

| Covariate | Method: Bootstrap without replacement ($N_{training} = 300$, $N_{testing} = 75$) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $MLE_{MCDF}$ | | $MLqE_{MCDF}$ | | $MLE_{HSCF}$ | | $MLqE_{HSCF}$ | |
| | $\hat{\beta}$ | SD | $\hat{\beta}$ | SD | $\hat{\beta}$ | SD | $\hat{\beta}$ | SD |
| Intercept | 72.13 | 1.05 | 71.43 | 1.16 | 73.03 | 2.64 | 69.78 | 1.85 |
| Age | 0.06 | 0.02 | 0.05 | 0.02 | 0.05 | 0.03 | 0.05 | 0.03 |
| Gender | 3.67 | 1.13 | 3.97 | 1.15 | 3.78 | 1.24 | 3.64 | 1.57 |
| Time | −0.17 | 0.01 | −0.15 | 0.01 | −0.14 | 0.04 | −0.12 | 0.02 |
| Hypotension | −4.40 | 0.38 | −4.07 | 0.35 | −8.86 | 1.19 | −8.98 | 0.72 |
| Surgical department | 2.23 | 0.71 | 2.32 | 0.72 | 1.73 | 1.04 | 2.93 | 0.89 |
| Marcain-heavy | −0.28 | 0.08 | −0.25 | 0.09 | −0.24 | 0.14 | −0.19 | 0.14 |
| Midazolam | −0.73 | 0.31 | −0.87 | 0.33 | −0.65 | 0.35 | −1.39 | 0.44 |
| Chirocaine | −0.08 | 0.06 | −0.09 | 0.07 | −0.08 | 0.11 | −0.07 | 0.13 |
| MSE | 152.61 | | 152.23 | | 151.82 | | 149.57 | |
| MSPE | 136.21 | | 135.60 | | 135.41 | | 132.46 | |

(i) When data sets are generated from a pure joint mean–covariance model without outliers, the proposed MLqE approach is comparable to the MLE when the sample size is moderate or large, but slightly worse when the sample size is small.

(ii) The MLqE provides more robust estimates than the MLE when outliers are present in the data set.

(iii) The MLqE outperforms the MLE for heavy-tailed error distributions, and the superiority is quite significant for HSCF covariance structure.

## 5. Real data analysis

As a common clinic anesthetic technique used in surgery, spinal anesthesia may cause hypotension during operations. Therefore, investigating the relationship between patients' diastolic blood pressure (DBP) and particular risk factors (such as age, gender, anesthesia drug doses, etc.) is valuable for anesthesiology studies [9,17,18]. The data set in our study is from the Department of Anesthesiology and Reanimation, Akdeniz University Hospital, Antalya, Turkey. The data are from 375 patients (210 males and 165 females) who had spinal anesthesia from January 2008 to January 2011, and were recorded from three surgical departments (General, O&G and Urology). The response variable is DBP, and the explanatory variables are age, gender, operation time, surgical department and the dosages of Marcain-heavy, midazolam and chirocaine. There are no missing observations in either outcome or covariates.

Fig. 1 displays the percentage and the frequency of hypotension occurrence. According to the left picture in Fig. 1, the percentage of hypotension occurrence in the Department of Obstetrics and Gynecology (O&G) is higher than in the other two

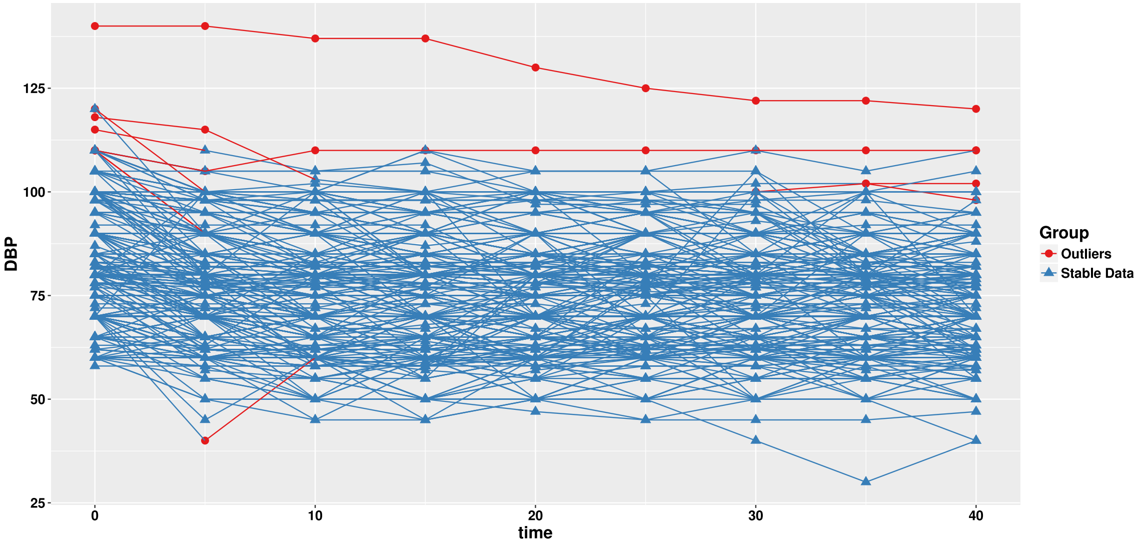

**Fig. 2.** Outliers in the anesthesia data set. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

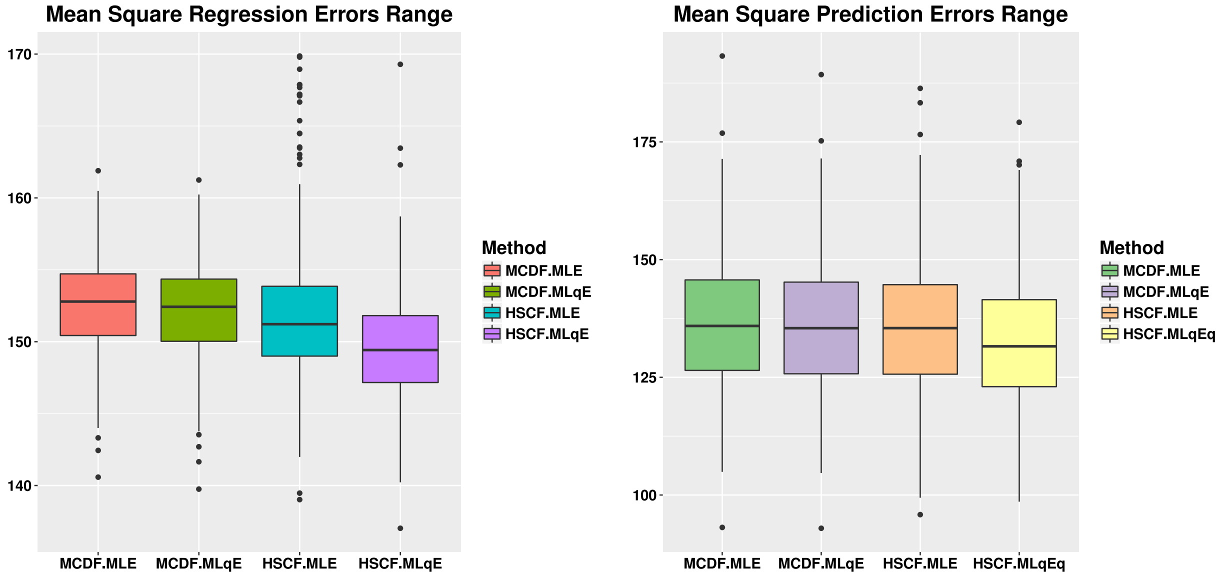

**Fig. 3.** MSE and MSPE box plots for different regression methods.

departments. Based on the histogram in the right of Fig. 1, female patients are more likely to experience hypotension than male patients in the surgery. In addition, a female patient tends to have hypotension in the first half stage of an operation, but a man often does in the second half stage, which means the gender is a noteworthy factor in anesthesia data analysis.

Next, we compare the performance of the MLE and the MLqE for different joint mean–covariance models, i.e., MCDF and HSCF. Let MCDF and HSCF share the same covariates for covariance modeling, i.e., $\mathbf{h}_{ij} = \mathbf{x}_{ij}$ and $\mathbf{z}_{ijk} = (1, (t_{ij} - t_{ik}), (t_{ij} - t_{ik})^2)^\top$. First, using the MLE with HSCF model, we obtain the coefficients estimates from the whole data; these are listed in Table 4. Then, we find the outliers (11 subjects) by the residuals criterion proposed in [10] and plot them in red in Fig. 2. Next, 75 out of the 364 stable subjects are randomly selected as the testing set and the remaining subjects are used as a training set. The standard deviations, average values of mean square regression errors (MSE) and the mean square prediction errors (MSPE) based on 400 bootstrap replications are calculated and reported in Table 4. Fig. 3 also displays the MSE and MSPE for different modeling methods.

The coefficient estimates listed in Table 4 indicate that all three kinds of anesthetics have negative effects and are significant on blood pressure responses. This implies that excessive use may lead to hypotension during the surgery and the dose of anesthetics should be selected cautiously. Our analysis result is consistent with the clinical phenomena observed in most hospital departments. In addition, both Table 4 and Fig. 3 show that the MSE and the MSPE of the MLqE$_{\text{(HSCF)}}$ are smallest, and therefore, MLqE$_{\text{(HSCF)}}$ is the most preferred method for modeling this data set.

## 6. Discussion

We have proposed a unified robust maximum $L_q$-likelihood estimation of joint mean–covariance models for longitudinal studies. The new method provides comparable performance to the traditional MLE when there is no outlier but provides much better estimates when there are outliers or the errors are heavy-tailed in the longitudinal data.

We mainly focused on the joint mean–covariance models based on MCDF and HSCF, but the procedure and asymptotic results can also be extended to other joint mean–covariance models for longitudinal data. In the future, it would be of interest to combine the ideas of maximum $L_q$-likelihood estimation and nonparametric covariance modeling [1,21,23]. In addition, similar to [4,15], all the theoretical results provided in this article only cover cases without outliers. It requires further research to study the properties of the proposed estimates when there are outliers in the data set.

## Acknowledgments

The authors are grateful to the editor, associate editor and two referees for numerous helpful comments during the preparation of the article. Yao's research is supported by NSF, China grant DMS-1461677. Xu's research is supported by Zhejiang Provincial NSF of China grant LY19A010006, Zhejiang Federation of Humanities and Social Sciences Circles, China grant 16NDJC154YB and the First Class Discipline of Zhejiang - A (Zhejiang University of Finance and Economics-Statistics), China. Xiang's research is supported by NSF of China grant 11601477 and the First Class Discipline of Zhejiang - A (Zhejiang University of Finance and Economics-Statistics), China.

## Appendix A. Monotonicity of MMEM

The following proof of Theorem 1 indicates that the objective function (2) is nondecreasing in each iteration of the MMEM algorithm.

**Proof of Theorem 1.** Let $Z_i^{(k+1)}$ be a discrete random variable such that

$$\Pr\left\{Z_i^{(k+1)} = \frac{f(\mathbf{y}_i;\boldsymbol{\theta}^{(k+1)})^{1-q}}{f(\mathbf{y}_i;\boldsymbol{\theta}^{(k)})^{1-q}}\right\} = \frac{f(\mathbf{y}_i;\boldsymbol{\theta}^{(k)})^{1-q}}{\sum_{i=1}^n f(\mathbf{y}_i;\boldsymbol{\theta}^{(k)})^{1-q}} \triangleq P_i.$$

Then

$$\begin{aligned}\ln\left\{\frac{\sum_{i=1}^n f(\mathbf{y}_i;\boldsymbol{\theta}^{(k+1)})^{1-q}}{\sum_{i=1}^n f(\mathbf{y}_i;\boldsymbol{\theta}^{(k)})^{1-q}}\right\} &= \ln\left\{\sum_{i=1}^n \frac{f(\mathbf{y}_i;\boldsymbol{\theta}^{(k)})^{1-q}}{\sum_{i=1}^n f(\mathbf{y}_i;\boldsymbol{\theta}^{(k)})^{1-q}} \times \frac{f(\mathbf{y}_i;\boldsymbol{\theta}^{(k+1)})^{1-q}}{f(\mathbf{y}_i;\boldsymbol{\theta}^{(k)})^{1-q}}\right\}\\ &= \ln\left\{\sum_{i=1}^n P_i \times \frac{f(\mathbf{y}_i;\boldsymbol{\theta}^{(k+1)})^{1-q}}{f(\mathbf{y}_i;\boldsymbol{\theta}^{(k)})^{1-q}}\right\} = \ln\{\mathrm{E}(Z_i^{(k+1)})\}.\end{aligned}$$

By Jensen's inequality,

$$\ln\{\mathrm{E}(Z_i^{(k+1)})\} \geq \mathrm{E}\{\ln Z_i^{(k+1)}\} = \sum_{i=1}^n \frac{f(\mathbf{y}_i;\boldsymbol{\theta}^{(k)})^{1-q}}{\sum_{i=1}^n f(\mathbf{y}_i;\boldsymbol{\theta}^{(k)})^{1-q}} \times \ln\left\{\frac{f(\mathbf{y}_i;\boldsymbol{\theta}^{(k+1)})^{1-q}}{f(\mathbf{y}_i;\boldsymbol{\theta}^{(k)})^{1-q}}\right\} = \sum_{i=1}^n P_i \times \ln\left\{\frac{f(\mathbf{y}_i;\boldsymbol{\theta}^{(k+1)})^{1-q}}{f(\mathbf{y}_i;\boldsymbol{\theta}^{(k)})^{1-q}}\right\}.$$

By the M-step property (5),

$$\sum_{i=1}^n P_i \times \ln f\{\mathbf{y}_i;\boldsymbol{\theta}^{(k+1)}\} \geq \sum_{i=1}^n P_i \times \ln f\{\mathbf{y}_i;\boldsymbol{\theta}^{(k)}\}.$$

Then we have

$$\ln\left\{\frac{\sum_{i=1}^n f(\mathbf{y}_i;\boldsymbol{\theta}^{(k+1)})^{1-q}}{\sum_{i=1}^n f(\mathbf{y}_i;\boldsymbol{\theta}^{(k)})^{1-q}}\right\} \geq \mathrm{E}\{\ln Z_i^{(k+1)}\} = \sum_{i=1}^n P_i \times \ln\left[\frac{f\{\mathbf{y}_i;\boldsymbol{\theta}^{(k+1)}\}^{1-q}}{f\{\mathbf{y}_i;\boldsymbol{\theta}^{(k)}\}^{1-q}}\right] = (1-q)\sum_{i=1}^n P_i \times \ln\left[\frac{f\{\mathbf{y}_i;\boldsymbol{\theta}^{(k+1)}\}^{1-q}}{f\{\mathbf{y}_i;\boldsymbol{\theta}^{(k)}\}}\right] \geq 0.$$

Therefore,

$$\sum_{i=1}^n f\{\mathbf{y}_i;\boldsymbol{\theta}^{(k+1)}\}^{1-q} \geq \sum_{i=1}^n f\{\mathbf{y}_i;\boldsymbol{\theta}^{(k)}\}^{1-q}.$$

## Appendix B. Consistency and asymptotic normality

Based on joint mean–covariance models and Assumptions (i)–(iii), this Appendix gives a sketch of the proof of the consistency and asymptotic normality results of the mean regression coefficients estimate $\hat{\boldsymbol{\beta}}$ by fixing the covariance parameters. Without loss of generality, the identity function $\mu(\mathbf{x}_i\boldsymbol{\beta}) = \mathbf{x}_i\boldsymbol{\beta}$ is used as the link function in this context. The derivations for other link functions are similar.

**Proofs of Theorem 2.** Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ be $n$ mutually independent random vectors with multivariate normal probability distributions $\mathcal{N}[\mathbf{x}_i\boldsymbol{\beta}, \boldsymbol{\Sigma}_i(\boldsymbol{\gamma}, \boldsymbol{\lambda})]$, respectively. The logarithm of $\mathbf{Y}_i$'s density is given by

$$\ln\{f(\mathbf{y}_i;\boldsymbol{\theta})\} = -\frac{m_i}{2}\ln(2\pi) - \frac{1}{2}(\ln|\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda}))| - \frac{1}{2}(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\beta})^\top \boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})^{-1}(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\beta}).$$

The deformed ln pdf of $\mathbf{Y}_i$ is defined as

$$Lq\{f(\mathbf{y}_i;\boldsymbol{\theta})\} = \{f(\mathbf{y}_i;\boldsymbol{\theta})^{1-q} - 1\}/(1-q),$$

where

$$f(\mathbf{y}_i;\boldsymbol{\theta}) = \exp\left\{-\frac{m_i}{2}\ln(2\pi) - \frac{1}{2}(\ln|\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda}))| - \frac{1}{2}(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\beta})^\top \boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})^{-1}(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\beta})\right\}.$$

Next, we set

$$\mathbf{K}_n(\boldsymbol{\beta};\boldsymbol{\gamma},\boldsymbol{\lambda}) = \frac{1}{n}\sum_{i=1}^n \mathrm{E}_{\boldsymbol{\theta}_0}\nabla_{\boldsymbol{\beta}}\ln\{f(\mathbf{y}_i;\boldsymbol{\theta})\} = \frac{1}{n}\sum_{i=1}^n \mathrm{E}_{\boldsymbol{\theta}_0}\mathbf{x}_i^\top\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})^{-1}(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i^\top\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})^{-1}\mathbf{x}_i(\boldsymbol{\beta}_0 - \boldsymbol{\beta}),$$

and

$$\mathbf{S}_n(\boldsymbol{\beta};\boldsymbol{\gamma},\boldsymbol{\lambda}) = \frac{1}{n}\sum_{i=1}^n \nabla_{\boldsymbol{\beta}}Lq_n\{f(\mathbf{y}_i;\boldsymbol{\theta})\} = \frac{1}{n}\sum_{i=1}^n f(\mathbf{y}_i;\boldsymbol{\theta})^{1-q_n}\cdot\mathbf{U}(\mathbf{y}_i;\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n f(\mathbf{y}_i;\boldsymbol{\theta})^{1-q_n}\cdot\mathbf{x}_i^\top\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})^{-1}(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\beta}).$$

Let $\|\cdot\|_1$ denote the $L_1$-norm. Then

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\Theta}\|\mathbf{S}_n(\boldsymbol{\beta};\boldsymbol{\gamma},\boldsymbol{\lambda}) - \mathbf{K}_n(\boldsymbol{\beta};\boldsymbol{\gamma},\boldsymbol{\lambda})\|_1 &= \sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^n \nabla_{\boldsymbol{\beta}}Lq_n\{f(\mathbf{y}_i;\boldsymbol{\theta})\} - \frac{1}{n}\sum_{i=1}^n \mathrm{E}_{\boldsymbol{\theta}_0}\nabla_{\boldsymbol{\beta}}\ln\{f(\mathbf{y}_i;\boldsymbol{\theta})\}\right\|_1 \\ &\le \sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^n \nabla_{\boldsymbol{\beta}}Lq_n\{f(\mathbf{y}_i;\boldsymbol{\theta})\} - \frac{1}{n}\sum_{i=1}^n \nabla_{\boldsymbol{\beta}}\ln\{f(\mathbf{y}_i;\boldsymbol{\theta})\}\right\|_1 \\ &\quad + \sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^n \nabla_{\boldsymbol{\beta}}\ln\{f(\mathbf{y}_i;\boldsymbol{\theta})\} - \frac{1}{n}\sum_{i=1}^n \mathrm{E}_{\boldsymbol{\theta}_0}\nabla_{\boldsymbol{\beta}}\ln\{f(\mathbf{y}_i;\boldsymbol{\theta})\}\right\|_1.\end{aligned} \tag{9}$$

By Assumption (ii) and the multivariate Law of Large Numbers, we find

$$\frac{1}{n}\sum_{i=1}^n \nabla_{\boldsymbol{\beta}}\ln\{f(\mathbf{y}_i;\boldsymbol{\theta})\} - \frac{1}{n}\sum_{i=1}^n \mathrm{E}_{\boldsymbol{\theta}_0}\nabla_{\boldsymbol{\beta}}\ln\{f(\mathbf{y}_i;\boldsymbol{\theta})\} \xrightarrow{p} \mathbf{0}_p,$$

which is equivalent to saying that the second summand in (9) satisfies

$$\sup_{\boldsymbol{\theta}\in\Theta}\|\ \frac{1}{n}\sum_{i=1}^n \nabla_{\boldsymbol{\beta}}\ln\{f(\mathbf{y}_i;\boldsymbol{\theta})\} - \frac{1}{n}\sum_{i=1}^n \mathrm{E}_{\boldsymbol{\theta}_0}\nabla_{\boldsymbol{\beta}}\ln\{f(\mathbf{y}_i;\boldsymbol{\theta})\}\ \|_1 \xrightarrow{p} 0.$$

By the Cauchy–Schwarz inequality, the first summand in (9) satisfies

$$\begin{aligned}&\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^n \nabla_{\boldsymbol{\beta}}Lq_n\{f(\mathbf{y}_i;\boldsymbol{\theta})\} - \frac{1}{n}\sum_{i=1}^n \nabla_{\boldsymbol{\beta}}\ln\{f(\mathbf{y}_i;\boldsymbol{\theta})\}\right\|_1 \\ &= \sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^n f(\mathbf{y}_i;\boldsymbol{\theta})^{1-q_n}\cdot\mathbf{x}_i^\top\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})^{-1}(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\beta}) - \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i^\top\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})^{-1}(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\beta})\right\|_1 \\ &= \sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^n \{f(\mathbf{y}_i;\boldsymbol{\theta})^{(1-q_n)} - 1\}\{\mathbf{x}_i^\top\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})^{-1}(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\beta})\}\right\|_1 \\ &\le \sup_{\boldsymbol{\theta}\in\Theta}\left[\sum_{j=1}^p\sqrt{\frac{1}{n}\sum_{i=1}^n\{f(\mathbf{y}_i;\boldsymbol{\theta})^{(1-q_n)} - 1\}^2} \times \sqrt{\frac{1}{n}\sum_{i=1}^n\{\mathbf{x}_i^\top\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})^{-1}\mathbf{r}_i\}_{(j)}^2}\right],\end{aligned}$$

where $\mathbf{r}_i = \mathbf{y}_i - \mathbf{x}_i\boldsymbol{\beta}$ and for $j \in \{1, \dots, p\}$, $\{\mathbf{x}_i^\top \boldsymbol{\Sigma}_i(\boldsymbol{\gamma}, \boldsymbol{\lambda})^{-1}\mathbf{r}_i\}_{(j)}$ is the $j$th element of the vector $\mathbf{x}_i^\top \boldsymbol{\Sigma}_i(\boldsymbol{\gamma}, \boldsymbol{\lambda})^{-1}\mathbf{r}_i$. Based on Assumption (ii), $\mathrm{E}_{\boldsymbol{\theta}_0} \sup_{\boldsymbol{\theta}\in\Theta} \|\mathbf{U}_\beta(\mathbf{y}_i; \boldsymbol{\theta}, q_n)\|^2 < \infty$ for all $i \in \{1, \dots, n\}$, and

$$\frac{1}{n}\sum_{i=1}^n \sup_{\boldsymbol{\theta}\in\Theta}\{\mathbf{x}_i^\top \boldsymbol{\Sigma}_i(\boldsymbol{\gamma}, \boldsymbol{\lambda})^{-1}\mathbf{r}_i\}_{(j)}^2 = O_p(1).$$

Next, by Markov's inequality,

$$P\left[\sup_{\boldsymbol{\theta}\in\Theta} \frac{1}{n}\sum_{i=1}^n \{f(\mathbf{y}_i; \boldsymbol{\theta})^{(1-q_n)} - 1\}^2 > \epsilon\right] \le \frac{1}{\epsilon}\sup_{\boldsymbol{\theta}\in\Theta} \frac{1}{n}\sum_{i=1}^n \mathrm{E}_{\boldsymbol{\theta}_0}\{f(\mathbf{y}_i; \boldsymbol{\theta})^{(1-q_n)} - 1\}^2.$$

By Assumption (ii), $\mathrm{E}_{\boldsymbol{\theta}_0}\{f(\mathbf{y}_i; \boldsymbol{\theta})^{1-q_n} - 1\}^2 \to 0$ as $q_n \to 1$. Then,

$$\sup_{\boldsymbol{\theta}\in\Theta}\left\| \frac{1}{n}\sum_{i=1}^n \{f(\mathbf{y}_i; \boldsymbol{\theta})^{(1-q_n)} - 1\}\{\mathbf{x}_i^\top \boldsymbol{\Sigma}_i(\boldsymbol{\gamma}, \boldsymbol{\lambda})^{-1}(\mathbf{y}_i - \mathbf{x}_i\boldsymbol{\beta})\}\right\|_1 \xrightarrow{p} 0.$$

With the above derivations, we conclude that

$$\sup_{\boldsymbol{\theta}\in\Theta}\left\| \frac{1}{n}\sum_{i=1}^n \nabla_\beta Lq_n\{f(\mathbf{y}_i; \boldsymbol{\theta})\} - \frac{1}{n}\sum_{i=1}^n \nabla_\beta \ln\{f(\mathbf{y}_i; \boldsymbol{\theta})\}\right\|_1 \xrightarrow{p} 0,$$

and therefore,

$$\sup_{\boldsymbol{\theta}\in\Theta} \| \mathbf{S}_n(\boldsymbol{\beta}; \boldsymbol{\gamma}, \boldsymbol{\lambda}) - \mathbf{K}_n(\boldsymbol{\beta}; \boldsymbol{\gamma}, \boldsymbol{\lambda}) \|_1 \xrightarrow{p} 0. \tag{10}$$

Obviously,

$$\mathbf{K}_n(\boldsymbol{\beta}; \boldsymbol{\gamma}, \boldsymbol{\lambda}) = \frac{1}{n}\sum_{i=1}^n \mathrm{E}_{\boldsymbol{\theta}_0}\nabla_\beta \ln\{f(\mathbf{y}_i; \boldsymbol{\theta})\}$$

is equicontinuous in $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{\lambda}^\top)^\top \in \boldsymbol{\Theta} \subset \mathbb{R}^{p+q+d}$. Since the parameter space $\boldsymbol{\Theta}$ is compact, and by Assumption (i), it is easy to see that $\mathbf{K}_n(\boldsymbol{\beta}; \boldsymbol{\gamma}, \boldsymbol{\lambda})$ converges to a finite limit, namely (7). Indeed,

$$\mathbf{K}(\boldsymbol{\beta}; \boldsymbol{\gamma}, \boldsymbol{\lambda}) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \mathrm{E}_{\boldsymbol{\theta}_0}\nabla_\beta \ln\{f(\mathbf{y}_i; \boldsymbol{\theta})\} = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \mathrm{E}_{\boldsymbol{\theta}_0}\mathbf{U}_\beta(\mathbf{y}_i; \boldsymbol{\theta}).$$

The foregoing convergence is uniform in $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{\lambda}^\top)^\top$ and the limit $\mathbf{K}(\boldsymbol{\beta}; \boldsymbol{\gamma}, \boldsymbol{\lambda})$ is continuous in $\boldsymbol{\Theta}$. Thus by Eq. (10),

$$\frac{1}{n}\sum_{i=1}^n \nabla_\beta Lq_n\{f(\mathbf{y}_i; \boldsymbol{\theta})\} \xrightarrow{p} \mathbf{K}(\boldsymbol{\beta}; \boldsymbol{\gamma}, \boldsymbol{\lambda})$$

uniformly in $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{\lambda}^\top)^\top$. Here $\mathbf{K}(\boldsymbol{\beta}; \boldsymbol{\gamma}, \boldsymbol{\lambda})$ is uniformly continuous in $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{\lambda}^\top)^\top$ since $\boldsymbol{\Theta}$ is compact.

Based on Theorem 5.9 in [19], the consistency of mean regression coefficients estimate is justified, i.e., $\tilde{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$. Similarly, we can establish the consistency of $\tilde{\boldsymbol{\gamma}}$ and $\boldsymbol{\lambda}$. We omit the proof here. □

**Proof of Theorem 3.** Denote $\mathbf{U}_\beta^*(\mathbf{y}_i; \boldsymbol{\theta}, q_n) = \nabla_\beta Lq_n\{f(\mathbf{y}_i; \boldsymbol{\theta})\} = f(\mathbf{y}_i; \boldsymbol{\theta})^{1-q_n}\mathbf{U}_\beta(\mathbf{y}_i; \boldsymbol{\theta})$. Suppose $\boldsymbol{\beta}^*$ is a vector such that $\mathrm{E}_{\boldsymbol{\theta}_0}\mathbf{U}_\beta^*(\mathbf{y}_i; \boldsymbol{\theta}, q_n) = \mathbf{0}_p$. By Taylor's expansion, for a solution $\tilde{\boldsymbol{\beta}}$ of the maximum $L_q$-likelihood equation, there exists a random vector $\boldsymbol{\beta}^{**}$ between $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^*$ such that

$$\begin{aligned}\mathbf{0}_p = \frac{1}{n}\sum_{i=1}^n \mathbf{U}_\beta^*(\mathbf{y}_i; \boldsymbol{\theta}, q_n)\Big|_{\beta=\tilde{\beta}} &= \frac{1}{n}\sum_{i=1}^n \mathbf{U}_\beta^*(\mathbf{y}_i; \boldsymbol{\theta}, q_n)\Big|_{\beta=\beta^*} + \frac{1}{n}\sum_{i=1}^n \mathbf{I}_\beta^*(\mathbf{y}_i; \boldsymbol{\theta}, q_n)\Big|_{\beta=\beta^*}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\\ &\quad + \frac{1}{2}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^\top\left\{\frac{1}{n}\sum_{i=1}^n \nabla_\beta^2\mathbf{U}_\beta^*(\mathbf{y}_i; \boldsymbol{\theta}, q_n)\Big|_{\beta=\beta^{**}}\right\}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*).\end{aligned} \tag{11}$$

Here, $\mathbf{I}_\beta^*(\mathbf{y}_i; \boldsymbol{\theta}, q_n) = \nabla_\beta\mathbf{U}_\beta^*(\mathbf{y}_i; \boldsymbol{\theta}, q_n) = \nabla_\beta^2 Lq_n\{f(\mathbf{y}_i; \boldsymbol{\theta})\}$ is a $p \times p$ matrix of first order derivatives and $\nabla_\beta^2\mathbf{U}_\beta^*(\mathbf{y}_i; \boldsymbol{\theta}, q_n)$ is a $p \times p \times p$ array of partial second-order derivatives. Then expression (11) can be rewritten as

$$\begin{aligned}-\sqrt{n}A_n\left\{\frac{1}{n}\sum_{i=1}^n \mathbf{U}_\beta^*(\mathbf{y}_i; \boldsymbol{\theta}, q_n)\Big|_{\beta=\beta^*}\right\} &= A_n\left\{\frac{1}{n}\sum_{i=1}^n \mathbf{I}_\beta^*(\mathbf{y}_i; \boldsymbol{\theta}, q_n)\Big|_{\beta=\beta^*}\right\}\sqrt{n}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\\ &\quad + A_n\frac{\sqrt{n}}{2}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^\top\left\{\frac{1}{n}\sum_{i=1}^n \nabla_\beta^2\mathbf{U}_\beta^*(\mathbf{y}_i; \boldsymbol{\theta}, q_n)\Big|_{\beta=\beta^{**}}\right\}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*),\end{aligned}$$

where

$$A_n = \left\{ \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}_{\boldsymbol{\theta}_0} \mathbf{U}_{\beta}^{*}(\mathbf{y}_i; \boldsymbol{\theta}, q_n) \mathbf{U}_{\beta}^{*}(\mathbf{y}_i; \boldsymbol{\theta}, q_n)^{\top} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*} \right\}^{-1/2}.$$

First, we want to show that $-\sqrt{n} A_n \sum_{i=1}^{n} \mathbf{U}_{\beta}^{*}(\mathbf{y}_i; \boldsymbol{\theta}, q_n)|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}/n$ converges in distribution to $\mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$. Consider an arbitrary vector $a \in \mathbb{R}^p$ such that $\|a\| > 0$. Define $\mathbf{W}_{n,i} = \mathbf{a}^{\top} \mathbf{U}_{\beta}^{*}(\mathbf{y}_i; \boldsymbol{\theta}, q_n)|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}$ and $\bar{\mathbf{W}}_n = (\mathbf{W}_{n,1} + \cdots + \mathbf{W}_{n,n})/n$. Since the $\mathbf{W}_{n,i}$s form a triangular array where the $\mathbf{W}_{n,i}$s are row-wise independent, we check the Lyapunov condition, which reads

$$n^{1/3} (\mathrm{E}\mathbf{W}_{n,i}^2)^{-1} \{\mathrm{E}(\mathbf{W}_{n,i}^3)^{2/3}\} \to 0$$

as $n \to \infty$. By the Lindeberg–Feller Central Limit Theorem, we have

$$\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}\mathbf{W}_{n,i}^2 \right)^{-1/2} \bar{\mathbf{W}}_n \rightsquigarrow \mathcal{N}(0, 1).$$

Hence, by the Cramér–Wold theorem, we have

$$-\sqrt{n} \left\{ \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}_{\boldsymbol{\theta}_0} \mathbf{U}_{\beta}^{*}(\mathbf{y}_i; \boldsymbol{\theta}, q_n) \mathbf{U}_{\beta}^{*}(\mathbf{y}_i; \boldsymbol{\theta}, q_n)^{\top} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*} \right\}^{-1/2} \times \left\{ \frac{1}{n} \sum_{i=1}^{n} \mathbf{U}_{\beta}^{*}(\mathbf{y}_i; \boldsymbol{\theta}, q_n) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*} \right\} \rightsquigarrow \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p).$$

Second, we show

$$\left\{ \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}_{\boldsymbol{\theta}_0} \mathbf{I}_{\beta}^{*}(\mathbf{y}_i; \boldsymbol{\theta}, q_n) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*} \right\}^{-1} \times \left\{ \frac{1}{n} \sum_{i=1}^{n} \mathbf{I}_{\beta}^{*}(\mathbf{y}_i; \boldsymbol{\theta}, q_n) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*} \right\} \xrightarrow{p} \mathbf{I}_p.$$

For fixed $k, \ell \in \{1, \ldots, p\}$, and given $\varepsilon > 0$,

$$\Pr_{\boldsymbol{\theta}_0} \left[ \left| \left\{ \frac{1}{n} \sum_{i=1}^{n} \mathbf{I}_{\beta}^{*}(\mathbf{y}_i; \boldsymbol{\theta}, q_n) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*} \right\}_{k,\ell} - \left\{ \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}_{\boldsymbol{\theta}_0} \mathbf{I}_{\beta}^{*}(\mathbf{y}_i; \boldsymbol{\theta}, q_n) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*} \right\}_{k,\ell} \right| > \varepsilon \right] \leq \frac{1}{n\varepsilon^2} \mathrm{E}_{\boldsymbol{\theta}_0} \left[ \frac{\partial}{\partial \boldsymbol{\beta}_k \partial \boldsymbol{\beta}_\ell} L q_n \{\ln f(\mathbf{y}_i; \boldsymbol{\theta})\} \right]^2.$$

Since $\boldsymbol{\Theta}$ is compact, the $\partial L q_n \{\ln f(\mathbf{y}_i; \boldsymbol{\theta})\}/(\partial \boldsymbol{\beta}_k \partial \boldsymbol{\beta}_\ell)$ is bounded from above by a constant. Hence,

$$\mathrm{E}_{\boldsymbol{\theta}_0} \left[ \partial L q_n \{\ln f(\mathbf{y}_i; \boldsymbol{\theta})\}/\partial \boldsymbol{\beta}_k \partial \boldsymbol{\beta}_\ell \right]^2 / (n\varepsilon^2) \to 0$$

as $n \to \infty$. Since convergence in probability is ensured for each $k$, $\ell$ and $p < \infty$, we conclude that

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{I}_{\beta}^{*}(\mathbf{y}_i; \boldsymbol{\theta}, q_n) - \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}_{\boldsymbol{\theta}_0} \mathbf{I}_{\beta}^{*}(\mathbf{y}_i; \boldsymbol{\theta}, q_n)$$

converges in probability to a $p \times p$ zero matrix. Therefore,

$$\left\{ \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}_{\boldsymbol{\theta}_0} \mathbf{I}_{\beta}^{*}(\mathbf{y}_i; \boldsymbol{\theta}, q_n) \right\}^{-1} \times \left\{ \frac{1}{n} \sum_{i=1}^{n} \mathbf{I}_{\beta}^{*}(\mathbf{y}_i; \boldsymbol{\theta}, q_n) \right\} \Bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*} \xrightarrow{p} \mathbf{I}_p.$$

Finally, we show that

$$\left\{ \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}_{\boldsymbol{\theta}_0} \mathbf{U}_{\beta}^{*}(\mathbf{y}_i; \boldsymbol{\theta}, q_n) \mathbf{U}_{\beta}^{*}(\mathbf{y}_i; \boldsymbol{\theta}, q_n)^{\top} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*} \right\}^{-1/2} \frac{\sqrt{n}}{2} (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)^{\top} \left\{ \frac{1}{n} \sum_{i=1}^{n} \nabla_{\beta}^2 \mathbf{U}_{\beta}^{*}(\mathbf{y}_i; \boldsymbol{\theta}, q_n) \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{**}} \right\} (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)$$

is negligible. The second-order derivatives $\sum_{i=1}^{n} \nabla_{\beta}^2 \mathbf{U}_{\beta}^{*}(\mathbf{y}_i; \boldsymbol{\beta}^{**}, q_n)/n$ is a $p \times p \times p$ array. By Assumption (iv), there is a neighborhood of $\boldsymbol{\beta}_0$ such that $\sum_{i=1}^{n} \nabla_{\beta}^2 \mathbf{U}_{\beta}^{*}(\mathbf{y}_i; \boldsymbol{\beta}^{**}, q_n)/n$ is dominated by $g_0(y)$ for some $g_0(y) \geq 0$. With probability tending to 1,

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \nabla_{\beta}^2 \mathbf{U}_{\beta}^{*}(\mathbf{y}_i; \boldsymbol{\beta}^{**}, q_n) \right\| \leq p^3 \frac{1}{n} \sum_{i=1}^{n} |g_0(\mathbf{y}_i)|.$$

Therefore, each entry of $\sum_{i=1}^{n} \nabla_{\beta}^2 \mathbf{U}_{\beta}^{*}(\mathbf{y}_i; \boldsymbol{\beta}^{**}, q_n)/n$ is bounded in probability. Recall Assumption (v), that

$$\left\{ \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}_{\boldsymbol{\theta}_0} \mathbf{U}_{\beta}^{*}(\mathbf{y}_i; \boldsymbol{\theta}, q_n) \mathbf{U}_{\beta}^{*}(\mathbf{y}_i; \boldsymbol{\theta}, q_n)^{\top} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*} \right\}^{-1/2}$$

is a bounded matrix. Hence,

$$\left\{\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_{\theta_0}\mathbf{U}_\beta^*(\mathbf{y}_i;\boldsymbol{\theta},q_n)\mathbf{U}_\beta^*(\mathbf{y}_i;\boldsymbol{\theta},q_n)^\top\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}\right\}^{-1/2}\frac{\sqrt{n}}{2}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)^\top\left\{\frac{1}{n}\sum_{i=1}^{n}\nabla_\beta^2\mathbf{U}_\beta^*(\mathbf{y}_i;\boldsymbol{\theta},q_n)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{**}}\right\}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)$$

is bounded in probability and it is of higher order than the second term.

By combining the above and applying Slutsky's lemma, we obtain the following asymptotic normality result:

$$\sqrt{n}\,\mathbf{V}_n^{-1/2}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\rightsquigarrow\mathcal{N}[\mathbf{0}_p,\mathbf{I}_p],$$

where

$$\mathbf{V}_n=\left\{\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_{\theta_0}\mathbf{I}_\beta^*(\mathbf{y}_i;\boldsymbol{\theta},q_n)\right\}^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_{\theta_0}\mathbf{U}_\beta^*(\mathbf{y}_i;\boldsymbol{\theta},q_n)\mathbf{U}_\beta^*(\mathbf{y}_i;\boldsymbol{\theta},q_n)^\top\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_{\theta_0}\mathbf{I}_\beta^*(\mathbf{y}_i;\boldsymbol{\theta},q_n)\right\}^{-1}\Bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}.$$

So the asymptotic normality result of Theorem 3 is established. Note that the above result holds even when the covariance matrix is misspecified.

Calling on Assumption (iii), we have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_{\theta_0}\mathbf{U}_\beta^*(\mathbf{y}_i;\boldsymbol{\theta},q_n)\mathbf{U}_\beta^*(\mathbf{y}_i;\boldsymbol{\theta},q_n)^\top\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}&=\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_{\theta_0}f(\mathbf{y}_i;\boldsymbol{\theta})^{2(1-q_n)}\cdot\mathbf{U}_\beta(\mathbf{y}_i;\boldsymbol{\theta})\mathbf{U}_\beta(\mathbf{y}_i;\boldsymbol{\theta})^\top\Bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}\\&=\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_{\theta_0}\mathbf{x}_i^\top\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})^{-1}\mathbf{x}_i\{1+O_p(1)\}\end{aligned}$$

and

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\mathbf{I}_\beta^*(\mathbf{y}_i;\boldsymbol{\theta},q_n)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}&=\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_{\theta_0}\mathbf{I}_\beta^*(\mathbf{y}_i;\boldsymbol{\theta},q_n)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}\{1+o_p(1)\}\\&=\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_{\theta_0}\left\{-(1-q_n)f(\mathbf{y}_i;\boldsymbol{\theta})^{1-q_n}\cdot\mathbf{U}_\beta(\mathbf{y}_i;\boldsymbol{\theta})\mathbf{U}_\beta(\mathbf{y}_i;\boldsymbol{\theta})^\top+f(\mathbf{y}_i;\boldsymbol{\theta})^{(1-q_n)}\cdot\mathbf{I}_\beta(\mathbf{y}_i;\boldsymbol{\theta})\right\}\Bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}\\&=\left\{\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_{\theta_0}\mathbf{x}_i^\top\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})^{-1}\mathbf{x}_i\right\}\Bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}\{1+o_p(1)\}.\end{aligned}$$

So

$$\begin{aligned}\mathbf{V}_n&=\left\{\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_{\theta_0}\mathbf{x}_i^\top\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})^{-1}\mathbf{x}_i\right\}^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i^\top\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})^{-1}\mathbf{x}_i\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_{\theta_0}\mathbf{x}_i^\top\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})^{-1}\mathbf{x}_i\right\}^{-1}\{1+o_p(1)\}\\&=\left\{\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_{\theta_0}\mathbf{x}_i^\top\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})^{-1}\mathbf{x}_i\right\}^{-1}\{1+o_p(1)\}\xrightarrow{p}\mathbf{V},\end{aligned}$$

where

$$\mathbf{V}=\left\{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_{\theta_0}\mathbf{x}_i^\top\boldsymbol{\Sigma}_i(\boldsymbol{\gamma},\boldsymbol{\lambda})^{-1}\mathbf{x}_i\right\}^{-1}.$$

## References

[1] Z. Chen, C. Leng, Local linear estimation of covariance matrices via Cholesky decomposition, Statist. Sinica 25 (2015) 1249–1263.

[2] P. Diggle, P. Heagerty, K. Liang, S. Zeger, Analysis of Longitudinal Data, Oxford University Press, 2002.

[3] J. Fan, Y. Wu, Semiparametric estimation of covariance matrices for longitudinal data, J. Amer. Statist. Assoc. 105 (2008) 1520–1533.

[4] D. Ferrari, Y. Yang, Maximum $L_q$-likelihood estimation, Ann. Statist. 38 (2010) 753–783.

[5] C.F. Kou, J.X. Pan, Variable Selection for Joint Mean and Covariance Models Via Penalized Likelihood, Technical Report, School of Mathematics, The University of Manchester, England, 2010.

[6] C. Leng, W. Zhang, J. Pan, Semiparametric mean-covariance regression analysis for longitudinal data, J. Amer. Statist. Assoc. 105 (2010) 181–193.

[7] D. Leung, Y.-G. Wang, M. Zhu, Efficient parameter estimation in longitudinal data analysis using a hybrid GEE method, Biostatistics 10 (2009) 436–445.

[8] J. Li, S. Ray, B.G. Lindsay, A nonparametric statistical approach to clustering via mode identification, J. Mach. Learn. Res. 8 (2007) 1687–1723.

[9] C.S. Lin, J.S. Chiu, M.H. Hsieh, M.S. Mok, Y.C. Li, H.W. Chiu, Predicting hypotensive episodes during spinal anesthesia with the application of artificial neural networks, Comput. Methods Programs Biomed. 92 (2008) 193–197.

[10] W. Mendenhall, R.J. Beaver, B.M. Beaver, An Brief Introduction to Probability and Statistics, first ed. Duxbury, CA, 2002.

[11] J. Pan, G. Mackenzie, Model selection for joint mean-covariance structures in longitudinal studies, Biometrika 90 (2003) 239–244.

[12] M. Pourahmadi, Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation, Biometrika 86 (1999), 677–690.
[13] M. Pourahmadi, Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix, Biometrika 87 (2000), 425–435.
[14] M. Pourahmadi, Cholesky decompositions and estimation of a covariance matrix: Orthogonality of variance-correlation parameters, Biometrika 94 (2007) 1006–1013.
[15] Y. Qin, C.E. Priebe, Maximum $L_q$-likelihood estimation via the expectation–maximization algorithm: A robust estimation of mixture models, J. Amer. Statist. Assoc. 108 (2013) 914–928.
[16] A. Qu, B.G. Lindsay, B. Li, Improving estimating equations using quadratic inference functions, Biometrika 87 (2000) 823–836.
[17] A.A. Samur, N. Coskunfirat, O. Saka, Comparison of predictor approaches for longitudinal binary outcomes application to anesthesiology data, PeerJ 2 (2014) http://dx.doi.org/10.7717/peerj.648.
[18] S.K. Sharma, N.M. Gajraj, J.E. Sidawi, Prevention of hypotension during spinal anesthesia: A comparison of intravascular administration of hetastarch versus lactated Ringer's solution, Anesth. Analg. 84 (1997) 111–114.
[19] A.W. van der Vaart, Asymptotic Statistics, Cambridge University Press, 1998.
[20] W. Yao, A note on EM algorithm for mixture models, Statist. Probab. Lett. 83 (2013) 519–526.
[21] W. Yao, R. Li, New local estimation procedure for nonparametric regression function of longitudinal data, J. R. Stat. Soc. Ser. B Stat. Methodol. 75 (2013) 123–138.
[22] H. Ye, J. Pan, Modelling covariance structures in generalized estimating equations for longitudinal data, Biometrika 93 (2006) 927–941.
[23] J. Yin, Z. Geng, R. Li, H. Wang, Nonparametric covariance model, Statist. Sinica 20 (2010) 469–479.
[24] W. Zhang, C. Leng, A moving average cholesky factor model in covariance modeling for longitudinal data, Biometrika 99 (2012) 141–150.
[25] W. Zhang, C. Leng, C.Y. Tang, A joint modeling approach for longitudinal studies, J. R. Stat. Soc. Ser. B Stat. Methodol. 77 (2015) 219–238.