# ESL-Based Robust Estimation for Mean-Covariance Regression with Longitudinal Data 

Fei Lu*, Liugen Xue, Xiong Cai<br>College of Applied Sciences, Beijing University of Technology, Beijing, China<br>Email: *lufeiby@163.com

How to cite this paper: Lu, F., Xue, L.G. and Cai, X. (2020) ESL-Based Robust Estimation for Mean-Covariance Regression with Longitudinal Data. Open Journal of Statistics, 10, 10-30.
https://doi.org/10.4236/ojs.2020.101002

Received: December 27, 2019
Accepted: January 10, 2020
Published: January 13, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/



#### Abstract

When longitudinal data contains outliers, the classical least-squares approach is known to be not robust. To solve this issue, the exponential squared loss (ESL) function with a tuning parameter has been investigated for longitudinal data. However, to our knowledge, there is no paper to investigate the robust estimation procedure against outliers within the framework of mean-covariance regression analysis for longitudinal data using the ESL function. In this paper, we propose a robust estimation approach for the model parameters of the mean and generalized autoregressive parameters with longitudinal data based on the ESL function. The proposed estimators can be shown to be asymptotically normal under certain conditions. Moreover, we develop an iteratively reweighted least squares (IRLS) algorithm to calculate the parameter estimates, and the balance between the robustness and efficiency can be achieved by choosing appropriate data adaptive tuning parameters. Simulation studies and real data analysis are carried out to illustrate the finite sample performance of the proposed approach.


## Keywords

Exponential Squared Loss Function, Within-Subject Correlation, Longitudinal Data, Modified Cholesky Decomposition, Robustness

## 1. Introduction

Longitudinal data arises frequently in many fields, such as biological research, social science and other fields. The observations on the same subject are measured repeatedly over time, and thus intrinsically correlated [1]. The covariance matrix of such data is important since ignoring the correlation structure may
lead to inefficient estimators of the mean parameter. Furthermore, the covariance matrix itself may be of scientific interest [2]. However, it is challenging to model the covariance matrix which suffers from the positive-definiteness constraint and includes many unknown parameters. To avoid this challenge, a common strategy is to specify the working correlation structure [3], which does not permit more general structures and cannot flexibly incorporate covariates that may be helpful to explain the covariations. To overcome this limitation, joint modelling for the mean and covariance of longitudinal data has received increasing interest recently; see, for example, [4]-[12]. Among these joint modeling approaches, the modified Cholesky decomposition (MCD) for the covariance matrix proposed in [4] [5] is attractive owing to the fact that it leads automatically to positive definite covariance matrix, and the parameters in it can be interpreted by suitable statistical concepts. As a result, the regression techniques and model based inference can be adopted to the parameters in this decomposition, see [7] for more details.

However, the aforementioned approaches are very sensitive to outliers or heavy-tailed distributions. [13] proposed a robust procedure for modeling the correlation matrix of longitudinal data based on an alternative Cholesky decomposition and heavy-tailed multivariate t-distributions with unknown degrees of freedom. It should be pointed out that the use of the multivariate $t$-distribution alone does not necessarily guarantee robustness. In addition, [14] [15] developed robust generalized estimating equations (GEE) for regression parameters in joint mean-covariance model by employing the Huber's function and leverage-based weights. [16] developed an efficient parameter estimation via MCD for quantile regression with longitudinal data. [17] further proposed a moving average Cholesky factor model, which is transformed from MCD, in covariance modeling for composite quantile regression with longitudinal data. Then, [18] carried out smoothed empirical likelihood inference via MCD for quantile varying coefficient models with longitudinal data. Later, [19] developed quantile estimation via MCD for longitudinal single-index models.

Although M-type regression and quantile regression procedures can overcome outliers and heavy-tail errors, they may lose efficiency under the normal distribution. To overcome this difficulty, [20] recently proposed a robust variable selection approach by adopting the exponential squared loss (ESL) function with a tuning parameter. They have showed that, with properly selected tuning parameter, the proposed approach not only achieves good robustness with respect to outliers in the dataset, but also is as asymptotically efficient as the least squares estimation without outliers under normal error. Later, some authors employed the ESL funtion for longitudinal data. For example, [21] propsed an efficient and robust variable selection method for longitudinal generalized linear models based on GEE. [22] proposed a robust and efficient estimation procedure for simultaneous model structure identification and variable selection in generalized partial linear varying coefficient models for longitudinal data. [23] developed GEE-based robust estimation and empirical likelihood inference approach with

ESL for panel data models. With a similar loss function, [24] proposed a robust variable selection method in modal varying-coefficient models with longitudinal data. [25] proposed modal regression statistical inference for longitudinal semivarying coefficient models, including GEE, empirical likelihood and variable selection. However, to our knowledge, there is no paper to investigate the robust estimation procedure against outliers within the framework of mean-covariance regression analysis for longitudinal data employing the ESL function.

In this paper, we propose a robust estimation approach for the model parameters of the mean and generalized autoregressive parameters in the within subject covariance matrices for longitudinal data based on the ESL function. We begin with the ESL-based estimation for the mean parameters pretending that the repeated measurements within a subject are independent. Then based on the roughly estimated mean parameters, the simultaneous estimation for the mean and generalized autoregressive parameters is carried out using the ESL function. The proposed estimators can be shown to be asymptotically normal under certain conditions. Moreover, we develop an iteratively reweighted least squares (IRLS) algorithm [26] to calculate the parameter estimates. Numerical studies are carried out to illustrate the finite sample performance of our approach.

The outline of this paper is organized as follows. Section 2 develops the robust estimation procedure. Section 3 establishes the asymptotic properties of the proposed ESL estimators. The IRLS algorithm and a data-driven method for the selection of tuning parameters are presented in Section 4. Simulations are carried out in Section 5. Section 6 analyses a real data set. A discussion is given in Section 7. The technical proofs are provided in the Appendix.

## 2. Robust Estimation Procedure

### 2.1. Initial Estimate for the Mean Parameters

Consider $n$ subjects, where each subject is measured repeatedly over time. For the $i$ th subject, suppose that $Y_{i j}$ is the observed scale response variable at time $t_{i j}$, and $X_{i j}$ is the corresponding $p \times 1$ covariate vector, $i=1, \cdots, n, j=1, \cdots, m_{i}$. Denote $N=\sum_{i=1}^{n} m_{i}$. Furthermore, let $Y_{i}=\left(Y_{i 1}, \cdots, Y_{i m_{i}}\right)^{\mathrm{T}}$, $X_{i}=\left(X_{i 1}, \cdots, X_{i m_{i}}\right)^{\mathrm{T}}, t_{i}=\left(t_{i 1}, \cdots, t_{i m_{i}}\right)^{\mathrm{T}=1}$. A longitudinal linear regression model has the form

$$
\begin{equation*}
Y_{i j}=X_{i j}^{\mathrm{T}} \beta+\varepsilon_{i j}, \quad i=1, \cdots, n, j=1, \cdots, m_{i} \tag{1}
\end{equation*}
$$

where $\beta$ is the $p \times 1$ vector of associated parameters, $\varepsilon_{i}=\left(\varepsilon_{i 1}, \cdots, \varepsilon_{i m_{i}}\right)^{\mathrm{T}}$ is the random error satisfying that $E\left(\varepsilon_{i j} \mid X_{i j}\right)=0$ and $\operatorname{var}\left(\varepsilon_{i} \mid X_{i}\right)=\Sigma_{i}$.

Define the ESL function $1-\phi_{\tau}(t)=1-\exp \left(-t^{2} / \tau\right)$. We first estimate $\beta$ pretending that the random error $\varepsilon_{i j}$ 's are independent. More specifically, we estimate $\beta$ by maximizing the objective function

$$
\begin{equation*}
L_{1}(\beta)=\sum_{i=1}^{n} \sum_{j=1}^{m_{i}} \phi_{\tau_{1}}\left(Y_{i j}-X_{i j}^{\mathrm{T}} \beta\right) \tag{2}
\end{equation*}
$$

where $\tau_{1}$ is a tuning parameter. The resulting initial estimate of $\beta$ is denoted
by $\tilde{\beta}$.

### 2.2. Simultaneous Estimate for the Mean and Generalized Autoregressive Parameters

Based on the Cholesky decomposition, there exists a lower triangle matrix $\Phi_{i}$ with diagonal ones such that

$$
\operatorname{var}\left(\Phi_{i} \varepsilon_{i} \mid X_{i}\right)=\Phi_{i} \Sigma_{i} \Phi_{i}^{\mathrm{T}}=D_{i},
$$

where $D_{i}=\operatorname{diag}\left(d_{i 1}^{2}, \cdots, d_{i m_{i}}^{2}\right)$. In other words, let $\epsilon_{i}=\left(\epsilon_{i 1}, \cdots, \epsilon_{i m_{i}}\right)^{\mathrm{T}}=\Phi_{i} \varepsilon_{i}$, we obtain that

$$
\begin{align*}
& \varepsilon_{i 1}=\epsilon_{i 1}, \\
& \varepsilon_{i j}=\phi_{j, 1}^{(i)} \varepsilon_{i 1}+\cdots+\phi_{j, j-1}^{(i)} \varepsilon_{i, j-1}+\epsilon_{i j}, \quad i=1, \cdots, n, j=2, \cdots, m_{i}, \tag{3}
\end{align*}
$$

where $\phi_{j, k}^{(i)}$ is the negative of the $(j, k)$ th element of $\Phi_{i}$. It's obvious that $\epsilon_{i j}$ 's are uncorrelated with $E\left(\epsilon_{i j}\right)=0$ and $\operatorname{var}\left(\epsilon_{i j}\right)=d_{i j}^{2}, j=1, \cdots, m_{i}$. If $\varepsilon_{i}$ 's were available, then (1) could be transformed as the following linear model with uncorrelated error $\epsilon_{i j}$ 's:

$$
\begin{align*}
& Y_{i 1}=X_{i 1}^{\mathrm{T}} \beta+\epsilon_{i 1}, \\
& Y_{i j}=X_{i j}^{\mathrm{T}} \beta+\phi_{j, 2}^{(i)} \varepsilon_{i 1}+\cdots+\phi_{j, j-1}^{(i)} \varepsilon_{i, j-1}+\epsilon_{i j}, \quad i=1, \cdots, n, j=2, \cdots, m_{i} . \tag{4}
\end{align*}
$$

[4] pointed out that the MCD has a well-founded statistical interpretation, and it has the advantage that the generalized autoregressive parameter $\phi_{j, k}^{(i)}$ 's and $\log$-innovation variance $\log \left(d_{i j}^{2}\right)$ 's are unconstrained. For simplicity, we assume that $d_{i j}^{2}=d^{2}$ for all $i=1, \cdots, n, j=2, \cdots, m_{i}$. Since $\Sigma_{i}=\operatorname{var}\left(\varepsilon_{i} \mid X_{i}\right)$ may depend on $X_{i}$, we adopt a more parsimonious structure,

$$
\begin{equation*}
\phi_{j, k}^{(i)}=Z_{j, k}^{(i) \mathrm{T}} \gamma, \tag{5}
\end{equation*}
$$

where $\gamma=\left(\gamma_{1}, \cdots, \gamma_{q}\right)^{\mathrm{T}}$ is the regression coefficient, and the covariates $Z_{j, k}^{(i)}=\left(Z_{j, k, 1}^{(i)}, \cdots, Z_{j, k, q}^{(i)}\right)^{\mathrm{T}}$ may contain the time, the baseline covariates $X_{i}$, the interactions and so on.

Based on $\tilde{\beta}$, we can obtain the estimated residuals

$$
\begin{equation*}
\tilde{\varepsilon}_{i j}=Y_{i j}-X_{i j}^{\mathrm{T}} \tilde{\beta}, \quad i=1, \cdots, n, j=1, \cdots, m_{i}-1 . \tag{6}
\end{equation*}
$$

From (4), (5) and (6), we can obtain the simultaneous estimate $\hat{\theta}=\left(\hat{\beta}^{\mathrm{T}}, \hat{\gamma}^{\mathrm{T}}\right)^{\mathrm{T}}$ of $\theta=\left(\beta^{\mathrm{T}}, \gamma^{\mathrm{T}}\right)^{\mathrm{T}}$ by maximizing the following objective function:

$$
\begin{align*}
L_{2}(\theta) & =\sum_{i=1}^{n} \sum_{j=2}^{m_{i}} \phi_{r_{2}}\left(Y_{i j}-X_{i j}^{\mathrm{T}} \beta-\tilde{\varepsilon}_{i 1} Z_{j, 1}^{(i) \mathrm{T}} \gamma-\cdots-\tilde{\varepsilon}_{i, j-1} Z_{j, j-1}^{(i \mathrm{~T}} \gamma\right) \\
& =\sum_{i=1}^{n} \sum_{j=2}^{m_{i}} \phi_{r_{2}}\left(Y_{i j}-X_{i j}^{\mathrm{T}} \beta-\sum_{k=1}^{j-1} \tilde{\varepsilon}_{i, k} Z_{j, k, 1}^{(i)} \gamma_{1}-\cdots-\sum_{k=1}^{i-1} \tilde{\varepsilon}_{i, k} Z_{j, k, q}^{(i)} \gamma_{q}\right)  \tag{7}\\
& =\sum_{i=1}^{n} \sum_{j=2}^{m_{i}} \phi_{r_{2}}\left(Y_{i j}-X_{i j}^{\mathrm{T}} \beta-\tilde{\zeta}_{i j}^{\mathrm{T}} \gamma\right) \\
& =\sum_{i=1}^{n} \sum_{j=2}^{m_{i}} \phi_{r_{2}}\left(Y_{i j}-\tilde{\delta}_{i j}^{\mathrm{T}} \theta\right),
\end{align*}
$$

where $\tilde{\delta}_{i j}=\left(X_{i j}^{\mathrm{T}}, \tilde{\zeta}_{i j}^{\mathrm{T}}\right)^{\mathrm{T}}$ with $\tilde{\zeta}_{i j}=\left(\sum_{k=1}^{j-1} \tilde{\varepsilon}_{i, k} Z_{j, k, 1}^{(i)} \cdots, \sum_{k=1}^{j-1} \tilde{\varepsilon}_{i, k} Z_{j, k, q}^{(i)}\right)^{\mathrm{T}}$. Then we can obtain the estimates of $\phi_{j, k}^{(i)}$ 's in model (3) by combining (5) and $\hat{\gamma}$.

## 3. Asymptotic Properties

Define $F_{1}\left(x, \tau_{1}\right)=E\left(\phi_{\tau_{1}}^{\prime \prime}\left(\varepsilon_{i j}\right) \mid X_{i j}=x\right), \quad G_{1}\left(x, \tau_{1}\right)=E\left(\phi_{\tau_{1}}^{\prime}\left(\varepsilon_{i j}\right)^{2} \mid X_{i j}=x\right)$, $F_{2}\left(x, \tau_{2}\right)=E\left(\phi_{\tau_{2}}^{\prime \prime}\left(\epsilon_{i j}\right) \mid X_{i j}=x\right), \quad G_{2}\left(x, \tau_{2}\right)=E\left(\phi_{\tau_{2}}^{\prime}\left(\epsilon_{i j}\right)^{2} \mid X_{i j}=x\right)$, $H\left(\bar{x}_{i j}\right)=E\left(\zeta_{i j} \zeta_{i j}^{\mathrm{T}} \mid \bar{X}_{i j}=\bar{x}_{i j}\right), \quad \bar{X}_{i j}=\left(X_{i 1}, \cdots, X_{i j}\right)^{\mathrm{T}}, \quad \bar{x}_{i j}=\left(x_{i 1}, \cdots, x_{i j}\right)^{\mathrm{T}}$, $\delta_{i j}=\left(X_{i j}^{\mathrm{T}}, \zeta_{i j}^{\mathrm{T}}\right)^{\mathrm{T}}, \quad \zeta_{i j}=\left(\sum_{k=1}^{j-1} \varepsilon_{i, k} Z_{j, k, 1}^{(i)}, \cdots, \sum_{k=1}^{j-1} \varepsilon_{i, k} Z_{j, k, q}^{(i)}\right)^{\mathrm{T}}$. To establish the asymptotic properties of the proposed ESL estimator, assume that the following regularity conditions hold:
(C1) There exists a positive integer $M$ such that $\max _{1 \leq i \leq n} m_{i} \leq M<\infty$. This means that $n$ and $N$ have the same order.
(C2) There exists a positive constant $C$ such that $\left|X_{i j}^{(k)}\right| \leq C$ for $1 \leq i \leq n$, $1 \leq j \leq m_{i}$ and $0 \leq k \leq p$. In addition, $\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_{i}} X_{i j} X_{i j}^{\mathrm{T}}$ converges to a finite positive definite matrix in probability.
(C3) $F_{1}\left(x, \tau_{1}\right)$ and $G_{1}\left(x, \tau_{1}\right)$ are continuous with respect to $x$. Moreover, for any $\tau_{1}>0, F_{1}\left(x, \tau_{1}\right)<0$.
(C4) $E\left(\phi_{\tau_{1}}^{\prime}\left(\varepsilon_{i j}\right) \mid X_{i j}=x\right)=0, E\left(\left|\phi_{\tau_{1}}^{\prime}\left(\varepsilon_{i j}\right)\right|^{3} \mid X_{i j}=x\right), E\left(\phi_{\tau_{1}}^{\prime \prime}\left(\varepsilon_{i j}\right)^{2} \mid X_{i j}=x\right)$, $E\left(\phi_{\tau_{1}}^{\prime \prime \prime}\left(\varepsilon_{i j}\right) \mid X_{i j}=x\right)$ and $E\left(\phi_{\tau_{1}^{\prime \prime \prime}}^{\prime \prime}\left(\varepsilon_{i j}\right)^{2} \mid X_{i j}=x\right)$ are continuous with respect to $x$.
(C5) $\frac{1}{n} \sum_{i=1}^{n} E\left(X_{i}^{\mathrm{T}} \phi_{\tau_{1}}^{\prime}\left(\varepsilon_{i}\right) \phi_{\tau_{1}}^{\prime}\left(\varepsilon_{i}\right)^{\mathrm{T}} X_{i}\right)$ converges to a finite positive definite matrix.
(C6) For $1 \leq i \leq n, 1 \leq j_{1}, \cdots, j_{6} \leq m_{i}-1, \quad E\left\{\left|\varepsilon_{i j_{1}} \varepsilon_{i j_{2}} \varepsilon_{i j_{3}}\right| \mid\left(X_{i j_{1}}, X_{i j_{2}}, X_{i j_{3}}\right)=x\right\}$ and $E\left\{\varepsilon_{i j_{1}} \cdots \varepsilon_{i j_{k}} \mid\left(X_{i j_{1}}, \cdots, X_{i j_{k}}\right)=x\right\}(k=2,3,4,6)$ are continuous with respect to $x$.
(C7) $\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_{i}} \zeta_{i j} \zeta_{i j}^{\mathrm{T}}$ converges to a finite positive definite matrix in probability.
(C8) $F_{2}\left(x, \tau_{2}\right)$ and $G_{2}\left(x, \tau_{2}\right)$ are continuous with respect to $x$. Moreover, for any $\tau_{2}>0, F_{2}\left(x, \tau_{2}\right)<0$.
(C9) $E\left(\phi_{\tau_{2}}^{\prime}\left(\epsilon_{i j}\right) \mid X_{i j}=x\right)=0, E\left(\left|\phi_{\tau_{2}}^{\prime}\left(\epsilon_{i j}\right)\right|^{3} \mid X_{i j}=x\right), E\left(\phi_{\tau_{2}}^{\prime \prime}\left(\epsilon_{i j}\right)^{2} \mid X_{i j}=x\right)$, $E\left(\phi_{\tau_{2}}^{\prime \prime \prime}\left(\epsilon_{i j}\right) \mid X_{i j}=x\right)$ and $E\left(\phi_{\tau_{2}}^{\prime \prime \prime}\left(\epsilon_{i j}\right)^{2} \mid X_{i j}=x\right)$ are continuous with respect to $x$.

Theorem 1 If regularity conditions ( $C 1)-(C 5)$ hold, then

$$
\sqrt{n}\left(\tilde{\beta}-\beta_{0}\right) \rightarrow N\left(0, \Sigma_{1}^{-1} \Sigma_{2} \Sigma_{1}^{-1}\right)
$$

in distribution as $n \rightarrow \infty$, where

$$
\begin{aligned}
& \Sigma_{1}=\lim _{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_{i}} E\left\{F_{1}\left(X_{i j}, \tau_{1}\right) X_{i j} X_{i j}^{\mathrm{T}}\right\}, \\
& \Sigma_{2}=\lim _{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^{n} E\left(X_{i}^{\mathrm{T}} \phi_{\tau_{1}}^{\prime}\left(\varepsilon_{i}\right) \phi_{\tau_{1}}^{\prime}\left(\varepsilon_{i}\right)^{\mathrm{T}} X_{i}\right) .
\end{aligned}
$$

Then if the random error $\varepsilon_{i j}$ 's are independent, we have the following corollary, which is similar to Corollary 2 and useful for the choice of tuning parameters in Section 4.2.

Corollary 1 If regularity conditions ( $C 1)-(C 4)$ hold and $\varepsilon_{i j}$ 's are independent, then

$$
\sqrt{n}\left(\tilde{\beta}-\beta_{0}\right) \rightarrow N\left(0, \Sigma_{1}^{-1} \Sigma_{2}^{*} \Sigma_{1}^{-1}\right)
$$

in distribution as $n \rightarrow \infty$, where

$$
\begin{aligned}
& \Sigma_{1}=\lim _{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_{i}} E\left\{F_{1}\left(X_{i j}, \tau_{1}\right) X_{i j} X_{i j}^{\mathrm{T}}\right\}, \\
& \Sigma_{2}^{*}=\lim _{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_{i}} E\left\{G_{1}\left(X_{i j}, \tau_{1}\right) X_{i j} X_{i j}^{\mathrm{T}}\right\} .
\end{aligned}
$$

Theorem 2 If regularity conditions (C1)-(C9) hold, then

$$
\sqrt{n}\left(\hat{\theta}-\theta_{0}\right) \rightarrow N\left(0, \Sigma_{3}^{-1} \Sigma_{4} \Sigma_{3}^{-1}\right)
$$

in distribution as $n \rightarrow \infty$, where

$$
\begin{aligned}
& \Sigma_{3}=\lim _{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=2}^{m_{i}} E\left\{F_{2}\left(X_{i j}, \tau_{2}\right) E\left(\delta_{i j} \delta_{i j}^{\mathrm{T}} \mid \bar{X}_{i j}\right)\right\}, \\
& \Sigma_{4}=\lim _{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=2}^{m_{i}} E\left\{G_{2}\left(X_{i j}, \tau_{2}\right) E\left(\delta_{i j} \delta_{i j}^{\mathrm{T}} \mid \bar{X}_{i j}\right)\right\} .
\end{aligned}
$$

In fact, we have

$$
E\left(\delta_{i j} \delta_{i j}^{\mathrm{T}} \mid \bar{X}_{i j}\right)=\left(\begin{array}{cc}
X_{i j} X_{i j}^{\mathrm{T}} & 0  \tag{8}\\
0 & H\left(\bar{X}_{i j}\right)
\end{array}\right) .
$$

Then it can be deduced that $\hat{\beta}$ and $\hat{\gamma}$ are asymptotically independent, that is, the following corollaries hold.

Corollary 2 If regularity conditions (C1)-(C9) hold, then

$$
\sqrt{n}\left(\hat{\beta}-\beta_{0}\right) \rightarrow N\left(0, \Sigma_{5}^{-1} \Sigma_{6} \Sigma_{5}^{-1}\right)
$$

in distribution as $n \rightarrow \infty$, where

$$
\begin{aligned}
& \Sigma_{5}=\lim _{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=2}^{m_{i}} E\left\{F_{2}\left(X_{i j}, \tau_{2}\right) X_{i j} X_{i j}^{\mathrm{T}}\right\}, \\
& \Sigma_{6}=\lim _{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=2}^{m_{i}} E\left\{G_{2}\left(X_{i j}, \tau_{2}\right) X_{i j} X_{i j}^{\mathrm{T}}\right\} .
\end{aligned}
$$

Corollary 3 If regularity conditions (C1)-(C9) hold, then

$$
\sqrt{n}\left(\hat{\gamma}-\gamma_{0}\right) \rightarrow N\left(0, \Sigma_{7}^{-1} \Sigma_{8} \Sigma_{7}^{-1}\right)
$$

in distribution as $n \rightarrow \infty$, where

$$
\begin{aligned}
& \Sigma_{7}=\lim _{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=2}^{m_{i}} E\left\{F_{2}\left(X_{i j}, \tau_{2}\right) H\left(\bar{X}_{i j}\right)\right\}, \\
& \Sigma_{8}=\lim _{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=2}^{m_{i}} E\left\{G_{2}\left(X_{i j}, \tau_{2}\right) H\left(\bar{X}_{i j}\right)\right\} .
\end{aligned}
$$

## 4. Implementation of the ESL Estimator

### 4.1. IRLS Algorithm

In this subsection, we develop an IRLS algorithm to calculate the parameter estimates. The IRLS algorithm has been commonly adopted for general M-estimators. Since the maximizers of (2) and (7) can be regarded as special M-estimators, the IRLS algorithm can be carried out to find $\tilde{\beta}$ and $\hat{\theta}$. In the following, we first develop the IRLS algorithm to find the maximizer of (2), and then we can calculate the maximizer of (7) in a similar way. Later, we summarize the algorithm in detail.

Because $\tilde{\beta}$ maximizes (2), we have the following normal equation:

$$
\sum_{i=1}^{n} \sum_{j=1}^{m_{i}} X_{i j} \phi_{\tau_{1}}^{\prime}\left(Y_{i j}-X_{i j}^{\mathrm{T}} \tilde{\beta}\right)=0
$$

or

$$
\begin{equation*}
\sum_{i=1}^{n} \sum_{j=1}^{m_{i}} X_{i j} \phi_{\tau_{1}}\left(Y_{i j}-X_{i j}^{\mathrm{T}} \tilde{\beta}\right)\left(Y_{i j}-X_{i j}^{\mathrm{T}} \tilde{\beta}\right)=0 \tag{9}
\end{equation*}
$$

Let $W_{i j}=\phi_{\tau_{1}}\left(Y_{i j}-X_{i j}^{\mathrm{T}} \tilde{\beta}\right)$, then (9) can be transformed as

$$
\sum_{i=1}^{n} \sum_{j=1}^{m_{i}} X_{i j} W_{i j}\left(Y_{i j}-X_{i j}^{\mathrm{T}} \tilde{\beta}\right)=0
$$

or $\quad X^{\mathrm{T}} W X \tilde{\beta}=X^{\mathrm{T}} W Y$, where $Y=\left(Y_{1}, \cdots, Y_{n}\right)^{\mathrm{T}}, \quad X=\left(X_{1}, \cdots, X_{n}\right)^{\mathrm{T}} \quad$ and $W=\operatorname{diag}\left(W_{11}, \cdots, W_{n, m_{n}}\right)$. Given the $k$-th approximation $\tilde{\beta}^{(k)}$, we can compute the corresponding weight matrix $W^{(k)}$ with $W_{i j}^{(k)}=\phi_{\tau_{1}}\left(Y_{i j}-X_{i j}^{\mathrm{T}} \tilde{\beta}^{(k)}\right)$. Then we have

$$
\begin{equation*}
\beta^{(k+1)}=\left(X^{\mathrm{T}} W^{(k)} X\right)^{-1} X^{\mathrm{T}} W^{(k)} Y \tag{10}
\end{equation*}
$$

This iteration of (10) will monotonically non-decrease the objective function (2), that is, $L_{1}\left(\tilde{\beta}^{(k+1)}\right)-L_{1}\left(\tilde{\beta}^{(k)}\right) \geq 0$. In fact,

$$
\begin{aligned}
& \log \left\{L_{1}\left(\tilde{\beta}^{(k+1)}\right)\right\}-\log \left\{L_{1}\left(\tilde{\beta}^{(k)}\right)\right\} \\
& =\log \left(\sum_{i=1}^{n} \sum_{j=1}^{m_{i}} W_{i j}^{(k+1)}\right)-\log \left(\sum_{i=1}^{n} \sum_{j=1}^{m_{i}} W_{i j}^{(k)}\right)=\log \left(\sum_{i=1}^{n} \sum_{j=1}^{m_{i}} \frac{W_{i j}^{(k+1)}}{\sum_{i=1}^{n} \sum_{j=1}^{m_{i}} W_{i j}^{(k)}}\right)
\end{aligned}
$$

$$
=\log \left(\sum_{i=1}^{n} \sum_{j=1}^{m_{i}} \frac{W_{i j}^{(k)}}{\sum_{i=1}^{n} \sum_{j=1}^{m_{i}} W_{i j}^{(k)}} \frac{W_{i j}^{(k+1)}}{W_{i j}^{(k)}}\right)=\log \left(\sum_{i=1}^{n} \sum_{j=1}^{m_{i}} \pi_{i j}^{(k)} \frac{W_{i j}^{(k+1)}}{W_{i j}^{(k)}}\right),
$$

where $\pi_{i j}^{(k)}=\frac{W_{i j}^{(k)}}{\sum_{i=1}^{n} \sum_{j=1}^{m_{i}} W_{i j}^{(k)}}$. Based on the Jensen's inequality, we have

$$
\begin{aligned}
& \log \left\{L_{1}\left(\tilde{\beta}^{(k+1)}\right)\right\}-\log \left\{L_{1}\left(\tilde{\beta}^{(k)}\right)\right\} \\
& \geq \sum_{i=1}^{n} \sum_{j=1}^{m_{i}} \pi_{i j}^{(k)} \log \left(\frac{W_{i j}^{(k+1)}}{W_{i j}^{(k)}}\right) \\
& =\frac{1}{\tau_{1}} \sum_{i=1}^{n} \sum_{j=1}^{m_{i}} \pi_{i j}^{(k)}\left\{\left(Y_{i j}-X_{i j}^{\mathrm{T}} \tilde{\beta}^{(k)}\right)^{2}-\left(Y_{i j}-X_{i j}^{\mathrm{T}} \tilde{\beta}^{(k+1)}\right)^{2}\right\} .
\end{aligned}
$$

From expression (10), we can find out that $\tilde{\beta}^{(k+1)}$ minimizes $(Y-X \beta)^{\mathrm{T}} W^{(k)}(Y-X \beta)$, or $\sum_{i=1}^{n} \sum_{j=1}^{m_{i}} W_{i j}^{(k)}\left(Y_{i j}-X_{i j}^{\mathrm{T}} \beta\right)^{2}$. Then we have $\log \left\{L_{1}\left(\tilde{\beta}^{(k+1)}\right)\right\}-\log \left\{L_{1}\left(\tilde{\beta}^{(k)}\right)\right\} \geq 0$.

The IRLS algorithm is summarized as follows:
Step 1. Computation of $\tilde{\beta}$ by maximizing (2). Take the initial value $\tilde{\beta}^{(0)}$ as the ordinary least squares (OLS) estimator. Given the $k$-th approximation $\tilde{\beta}^{(k)}$, the IRLS iteration updates $\tilde{\beta}^{(k+1)}$ through (10). Repeat this iteration until the convergence occurs. The resulting estimator is denoted as $\tilde{\beta}$.

Step 2. Computation of $\hat{\theta}=\left(\hat{\beta}^{\mathrm{T}}, \hat{\gamma}^{\mathrm{T}}\right)^{\mathrm{T}}$ by maximizing (7). Take $\hat{\theta}^{(0)}$ as the OLS estimator by minimizing $\sum_{i=1}^{n} \sum_{j=2}^{m_{i}}\left(Y_{i j}-\tilde{\delta}_{i j}^{\mathrm{T}} \theta\right)^{2}$. Similar to (10), given the $k$-th approximation $\hat{\theta}^{(k)}$, the IRLS iteration updates

$$
\hat{\theta}^{(k+1)}=\left(\tilde{\Delta}^{\mathrm{T}} V^{(k)} \tilde{\Delta}\right)^{-1} \tilde{\Delta}^{\mathrm{T}} V^{(k)} Y
$$

where $\quad \tilde{\Delta}=\left(\tilde{\delta}_{1}, \cdots, \tilde{\delta}_{n}\right)^{\mathrm{T}}, \quad \tilde{\delta}_{i}=\left(\tilde{\delta}_{i 2}, \cdots, \tilde{\delta}_{i m_{i}}\right)^{\mathrm{T}}, \quad V^{(k)}=\operatorname{diag}\left(V_{12}^{(k)}, \cdots, V_{n, m_{n}}^{(k)}\right)$, and $V_{i j}^{(k)}=\phi_{\tau_{2}}\left(Y_{i j}-\tilde{\delta}_{i j}^{\mathrm{T}} \hat{\theta}^{(k)}\right)$. Repeat this iteration until the convergence occurs. The resulting estimator is denoted as $\hat{\theta}$.

### 4.2. The Choice of Tuning Parameters

In this subsection, we give a data driving method to determine the tuning parameters $\left\{\tau_{1}, \tau_{2}\right\}$. In order to simplify the calculation with respect to $\tau_{1}$, we assume that the random error $\varepsilon_{i j}$ 's are independent of each other and $X_{i j}$ 's. Then from Corollary 1, we can obtain that the ratio between the asymptotic variance of the initial ESL estimator and that of the OLS estimator for $\beta$ is

$$
r\left(\tau_{1}\right) \triangleq \frac{G_{1}\left(\tau_{1}\right) F_{1}^{-2}\left(\tau_{1}\right)}{\sigma^{2}}
$$

where $\sigma^{2}=\operatorname{var}\left(\varepsilon_{i j}\right)$. Therefore, the ideal choice of $\tau_{1}$ is

$$
\tau_{1, \mathrm{opt}}=\arg \min _{\tau_{1}} r\left(\tau_{1}\right)=\arg \min _{\tau_{1}} G_{1}\left(\tau_{1}\right) F_{1}^{-2}\left(\tau_{1}\right) / \sigma^{2} .
$$

Then $r\left(\tau_{1}\right)$ can be estimated by $\tilde{G}_{1}\left(\tau_{1}\right) \tilde{F}_{1}^{-2}\left(\tau_{1}\right) / \tilde{\sigma}^{2}$, where

$$
\tilde{F}_{1}\left(\tau_{1}\right)=\frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{m_{i}} \phi_{\tau_{1}}^{\prime \prime}\left(\tilde{\varepsilon}_{i j}\right) \quad \text { and } \quad \tilde{G}_{1}\left(\tau_{1}\right)=\frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{m_{i}} \phi_{\tau_{1}}^{\prime}\left(\tilde{\varepsilon}_{i j}\right)^{2}
$$

with $\tilde{\varepsilon}_{i j}=Y_{i j}-X_{i j}^{\mathrm{T}} \tilde{\beta}$, and $\tilde{\sigma}$ is the standard deviation of $\tilde{\varepsilon}_{i j}$. Then $\tau_{1, \mathrm{opt}}$ can be easily obtained using the grid search approach. $\tau_{2}$ can also be chosen in a similar way.

## 5. Simulation Studies

In this section, we conduct some simulation studies to investigate the finite sample performance of the proposed approach. We generate 200 datasets and consider sample sizes $n=50,100$ and 200. In particular, the datasets are generated from the following model:

$$
Y_{i j}=X_{i j}^{\mathrm{T}} \beta+\varepsilon_{i j}, \quad i=1, \cdots, n, j=1, \cdots, m_{i}
$$

where $m_{i}-1 \sim \operatorname{binomial}(6,0.8), \quad X_{i j 1}, X_{i j 2} \sim N(0,1), \quad \beta=(1,2)^{\mathrm{T}}$, $t_{i j} \sim U(0,1), \quad Z_{j, k}^{(i)}=\left(1, t_{i j}-t_{i k}\right)^{\mathrm{T}}, \gamma=(0.5,0.2)^{\mathrm{T}}, \quad D_{i}=I_{m_{i}}$, and $\varepsilon_{i} \sim N\left(0, \Sigma_{i}\right)$ with $\Phi_{i} \Sigma_{i} \Phi_{i}^{\mathrm{T}}=D_{i}$.

To investigate robustness, we denote the above datasets as no contamination (NC) situation and consider the following four contaminations:

1) $\varepsilon_{i}$ follows the multivariate t -distribution with 2 degrees of freedom and covariance matrix $\Sigma_{i}$.
2) $\varepsilon_{i}$ follows the multivariate t-distribution with 2 degrees of freedom and covariance matrix $\Sigma_{i}$, and randomly choose $2 \%$ of $X_{i j}$ to be $X_{i j}+(6,6)^{\mathrm{T}}$.
3) $\varepsilon_{i} \sim N\left(0, \Sigma_{i}\right)$, and randomly choose $2 \%$ of $\varepsilon_{i j}$ to be $\varepsilon_{i j}+15$.
4) $\varepsilon_{i} \sim N\left(0, \Sigma_{i}\right)$, randomly choose $2 \%$ of $\varepsilon_{i j}$ to be $\varepsilon_{i j}+15$ and $2 \%$ of $X_{i j}$ to be $X_{i j}+(6,6)^{\mathrm{T}}$.

We compare the proposed ESL method with the OLS method, the M -estimator ( M ) in [26] and the quantile regression ( QR ) method in [27]. Note that the OLS, M and QR methods follow the estimation procedure similar to that of ESL, while the main difference is that the objective function is respectively replaced by their counterparts. To assess the finite sample performance, we calculate the mean and standard deviation (SD) for the estimators of $\beta$ and $\gamma$. The corresponding simulation results are displayed in Tables 1-5.

From Table 1, it can be observed that the performance of the M, QR and ESL methods is comparable to that of the OLS method, when the error follows a normal distribution and there are no outliers in the data. From Tables 2-5, it can be found out that the $\mathrm{M}, \mathrm{QR}$ and ESL methods outperform significantly the OLS method, particularly in terms of SD, in several contamination cases; moreover, the ESL method always perform best in these cases. More specifically, Ta-
ble 2 and Table 3 indicate that the $\mathrm{M}, \mathrm{QR}$ and ESL methods outperform significantly the OLS method with respect to both $\beta$ and $\gamma$, when the error follows a heavy-tailed error distribution. Table 4 and Table 5 indicate that the M, QR and ESL methods outperform significantly the OLS method with respect to $\beta$, when there are outliers in responses; the OLS, M and QR methods perform rather poorly with respect to $\gamma$, however, the ESL method sill performs well in this case.

## 6. Real Data Analysis

In this section, we analyse the CD4 cell study, which was previously analysed by [7] [28]. This dataset comprises CD4 cell counts of 369 HIV-infected men, and there are totally 2376 values collected at different times for each individual, over a period of approximately eight and a half years. The number of measurements for each individual varies from 1 to 12 and the time points are not equally spaced. We use square roots of the CD4 counts [28] to make the response variable closer to the normal distribution, and the related six covariates are respectively time since seroconversion $t_{i j}$, age relative to arbitrary origin $X_{i j 1}$, packs of cigarettes smoked per day $X_{i j 2}$, recreational drug use $X_{i j 3}$, number of sexual partners $X_{i j 4}$ and mental illness score $X_{i j 5}$. Note that Figure 1 displays the sample regressogram and local linear fitted curve for the square root of CD4 count over time, which reflects polynomial trend with respect to time. In the following, we use the mean model [1].

Table 1. Parameter estimates (with SD in parentheses) for the NC situation.

|  |  | $n=50$ | $n=100$ | $n=200$ |
| :---: | :---: | :---: | :---: | :---: |
| $\beta_{0}=1$ | OLS | $0.9968(0.0690)$ | $0.9973(0.0419)$ | $0.9997(0.0268)$ |
|  | M | $0.9972(0.0697)$ | $0.9968(0.0431)$ | $0.9997(0.0285)$ |
|  | QR | $0.9923(0.0845)$ | $0.9945(0.0493)$ | $0.9999(0.0357)$ |
| $\beta_{1}=2$ | ESL | $0.9972(0.0783)$ | $0.9956(0.0478)$ | $0.9997(0.0329)$ |
|  | OLS | $2.0025(0.0522)$ | $1.9987(0.0394)$ | $1.9961(0.0289)$ |
|  | QR | $2.0033(0.0535)$ | $1.9979(0.0414)$ | $1.9954(0.0289)$ |
|  | ESL | $2.0005(0.0642)$ | $1.9970(0.0483)$ | $1.9953(0.0340)$ |
|  | OLS | $0.0033(0.0601)$ | $1.9971(0.0466)$ | $1.9941(0.0315)$ |
|  | M | $0.4991(0.0333)$ | $0.4988(0.0227)$ | $0.5011(0.0167)$ |
|  | QR | $0.4975(0.0393)$ | $0.4977(0.0283)$ | $0.5010(0.0200)$ |
|  | ESL | $0.5008(0.0368)$ | $0.5001(0.0262)$ | $0.5010(0.0187)$ |
|  | OLS | $0.1922(0.1072)$ | $0.2057(0.0656)$ | $0.1956(0.0478)$ |
|  | M | $0.1961(0.1037)$ | $0.2043(0.0672)$ | $0.1962(0.0493)$ |
|  | QR | $0.2016(0.1191)$ | $0.2064(0.0881)$ | $0.1965(0.0603)$ |
|  |  | $0.1935(0.1121)$ | $0.2022(0.0802)$ | $0.1973(0.0560)$ |

Table 2. Parameter estimates (with SD in parentheses) for the first contamination situation.

|  |  | $n=50$ | $n=100$ | $n=200$ |
| :---: | :---: | :---: | :---: | :---: |
| $\beta_{0}=1$ | OLS | $0.9953(0.1707)$ | $1.0188(0.1392)$ | $1.0002(0.0819)$ |
|  | M | $0.9973(0.0878)$ | $1.0032(0.0585)$ | $1.0028(0.0387)$ |
|  | QR | $1.0034(0.0894)$ | $1.0011(0.0589)$ | $1.0040(0.0390)$ |
| $\beta_{1}=2$ | ESL | $1.0004(0.0869)$ | $1.0007(0.0513)$ | $1.0032(0.0379)$ |
|  | OLS | $1.9982(0.1207)$ | $2.0007(0.1092)$ | $1.9910(0.0844)$ |
|  | QR | $2.0041(0.0718)$ | $2.0018(0.0568)$ | $1.9958(0.0385)$ |
|  | ESL | $2.0048(0.0743)$ | $1.9997(0.0567)$ | $1.9985(0.0387)$ |
| $\gamma_{0}=0.5$ | MLS | $0.4821(0.0868)$ | $0.4680(0.1295)$ | $0.5081(0.0871)$ |
|  | QR | $0.4956(0.0508)$ | $0.4924(0.0465)$ | $0.5029(0.0367)$ |
|  | ESL | $0.4976(0.0491)$ | $0.4928(0.0445)$ | $0.5029(0.0345)$ |
|  | OLS | $0.4994(0.0546)$ | $0.4976(0.0452)$ | $0.5030(0.0321)$ |
|  | M | $0.2076(0.2766)$ | $0.2697(0.3493)$ | $0.1720(0.2637)$ |
|  | QR | $0.2027(0.1486)$ | $0.2202(0.1364)$ | $0.1930(0.1036)$ |
|  |  | $0.2010(0.1652)$ | $0.2028(0.1325)$ | $0.1929(0.1002)$ |

Table 3. Parameter estimates (with SD in parentheses) for the second contamination situation.

|  |  | $n=50$ | $n=100$ | $n=200$ |
| :---: | :---: | :---: | :---: | :---: |
| $\beta_{0}=1$ | OLS | $1.0063(0.1406)$ | $1.0106(0.1086)$ | $1.0041(0.0664)$ |
|  | M | $1.0003(0.0709)$ | $1.0007(0.0473)$ | $1.0018(0.0320)$ |
|  | QR | $1.0011(0.0786)$ | $1.0001(0.0474)$ | $1.0024(0.0324)$ |
| $\beta_{1}=2$ | ESL | $0.9997(0.0771)$ | $1.0009(0.0432)$ | $1.0014(0.0317)$ |
|  | OLS | $2.0048(0.1041)$ | $1.9993(0.0908)$ | $1.9970(0.0708)$ |
|  | M | $2.0051(0.0629)$ | $2.0004(0.0482)$ | $1.9966(0.0338)$ |
|  | ESL | $2.0032(0.0616)$ | $2.0001(0.0495)$ | $1.9980(0.0364)$ |
| $\gamma_{0}=0.5$ | MLS | $0.9973(0.0673)$ | $1.9980(0.0465)$ | $1.9974(0.0329)$ |
|  | QR | $0.4954(0.0522)$ | $0.4926(0.0465)$ | $0.5032(0.0368)$ |
|  | ESL | $0.4978(0.0501)$ | $0.4927(0.0447)$ | $0.5030(0.0336)$ |
|  | OLS | $0.4994(0.0543)$ | $0.4977(0.0446)$ | $0.5030(0.0323)$ |
|  | M | $0.2271(0.2820)$ | $0.2651(0.3357)$ | $0.1718(0.2648)$ |
|  | QR | $0.2022(0.1523)$ | $0.2213(0.1368)$ | $0.1935(0.1026)$ |
|  |  | $0.1988(0.1673)$ | $0.2043(0.1327)$ | $0.1929(0.1013)$ |

Table 4. Parameter estimates (with SD in parentheses) for the third contamination situation.

|  |  | $n=50$ | $n=100$ | $n=200$ |
| :---: | :---: | :---: | :---: | :---: |
| $\beta_{0}=1$ | OLS | $0.9969(0.1632)$ | $0.9955(0.1038)$ | $1.0023(0.0835)$ |
|  | M | $0.9945(0.0757)$ | $0.9934(0.0500)$ | $1.0013(0.0370)$ |
|  | QR | $0.9985(0.0830)$ | $0.9938(0.0552)$ | $0.9997(0.0418)$ |
| $\beta_{1}=2$ | ESL | $1.0025(0.0695)$ | $0.9949(0.0491)$ | $1.0014(0.0361)$ |
|  | OLS | $1.9826(0.1399)$ | $2.0107(0.1041)$ | $2.0099(0.0766)$ |
|  | QR | $1.9929(0.0631)$ | $2.0013(0.0473)$ | $2.0065(0.0365)$ |
|  | ESL | $1.9961(0.0756)$ | $2.0049(0.0532)$ | $2.0084(0.0448)$ |
|  | OLS | $0.0000(0.0573)$ | $2.0016(0.0457)$ | $2.0025(0.0343)$ |
|  | M | $0.1294)$ | $0.3651(0.0925)$ | $0.3519(0.0565)$ |
|  | QR | $0.4524(0.1412)$ | $0.4582(0.1125)$ | $0.4475(0.0709)$ |
|  | ESL | $0.4745(0.1266)$ | $0.4773(0.0971)$ | $0.4659(0.0576)$ |
|  | OLS | $0.00143(0.0784)$ | $0.5020(0.0371)$ | $0.4970(0.0206)$ |
|  | M | $0.0051(0.4775)$ | $0.0161(0.3990)$ | $0.0631(0.2578)$ |
|  | QR | $0.0067(0.4497)$ | $0.0363(0.3547)$ | $0.0879(0.2198)$ |
|  |  | $0.1263(0.3046)$ | $0.1938(0.1379)$ | $0.2034(0.0625)$ |

Table 5. Parameter estimates (with SD in parentheses) for the fourth contamination situation.

|  |  | $n=50$ | $n=100$ | $n=200$ |
| :---: | :---: | :---: | :---: | :---: |
| $\beta_{0}=1$ | OLS | $0.9921(0.1448)$ | $1.0105(0.0856)$ | $0.9985(0.0624)$ |
|  | M | $0.9911(0.0687)$ | $1.0018(0.0383)$ | $0.9959(0.0301)$ |
| $\beta_{1}=2$ | QR | $0.9920(0.0774)$ | $0.9996(0.0429)$ | $0.9970(0.0343)$ |
|  | ESL | $1.0001(0.0591)$ | $1.0018(0.0371)$ | $0.9992(0.0275)$ |
|  | MLS | $1.9993(0.1210)$ | $2.0014(0.0907)$ | $2.0010(0.0635)$ |
|  | QR | $1.9972(0.0541)$ | $1.9964(0.0423)$ | $1.9984(0.0304)$ |
| $\gamma_{0}=0.5$ | ESL | $1.9967(0.0613)$ | $1.9972(0.0477)$ | $1.9998(0.0339)$ |
|  | M | $0.3452(0.1183)$ | $0.3641(0.0792)$ | $0.3494(0.0570)$ |
|  | QR | $0.4481(0.1374)$ | $0.4661(0.0992)$ | $0.4468(0.0721)$ |
|  | ESL | $0.4736(0.1230)$ | $0.4819(0.0833)$ | $0.4642(0.0597)$ |
|  | OLS | $0.0138(0.4240)$ | $-0.0171(0.2704)$ | $0.0122(0.1909)$ |
|  | M | $-0.0174(0.4892)$ | $0.0004(0.3634)$ | $0.0664(0.2663)$ |
|  | QR | $-0.0161(0.4618)$ | $0.0302(0.3149)$ | $0.0936(0.2250)$ |
|  |  | $0.1324(0.3339)$ | $0.1962(0.1071)$ | $0.1964(0.0648)$ |



Figure 1. Sample regressogram and local linear fitted curve for the CD4 data: square root of CD4+ number versus time.

$$
Y_{i j}=\beta_{1} X_{i j 1}+\beta_{2} X_{i j 2}+\beta_{3} X_{i j 3}+\beta_{4} X_{i j 4}+\beta_{5} X_{i j 5}+f\left(t_{i j}\right)+\varepsilon_{i j}
$$

where $f(t)=\beta_{0}+\beta_{6} t_{+}+\beta_{7} t_{+}^{2}$ with $t_{+}=t I(t>0)$. We use cubic polynomial to model the generalized autoregressive parameters, that is,
$Z_{j, k}^{(i)}=\left(1, t_{i j}-t_{i k},\left(t_{i j}-t_{i k}\right)^{2},\left(t_{i j}-t_{i k}\right)^{3}\right)^{\mathrm{T}}$.
We apply the OLS, M, QR methods and the proposed ESL method to the CD4 cell study. To assess the prediction performance, we randomly split the data into three parts, each with $369 / 3=123$ subjects. We use the first two parts as the training dataset to fit the model, and then assess the out of sample performance on the testing dataset (defined as TD) which is left out. This process is repeated 200 times. We define the median absolute prediction error (MAPE) as median of $\left\{\left|Y_{i j}-\hat{Y}_{i j}\right|, i \in \mathrm{TD}, j=1, \cdots, m_{i}\right\}$. To illustrate the estimation robustness of the proposed ESL method compared with the other methods, we re-analyse the dataset by including 5\% outliers in the dataset, which are randomly generated by replacing $Y_{i j}$ with $Y_{i j}+300$. Moreover, we also re-analyse the dataset by including $5 \%$ outliers only in the training data set, in order to assess the robustness in terms of prediction performance. The results are displayed in Table 6. It can be observed that the estimates for $\gamma$ are very similar based on different methods, but the estimates for $\beta$ based on the ESL method are somewhat distinct with those of the OLS, M and QR methods in the no outlier case. Moreover, the MAPE of the proposed ESL method is slightly larger than the others, indicating that these methods possess comparative prediction performance in the case of no outlier. In the $5 \%$ outliers case, the MAPEs indicate that the ESL, M and QR methods perform similarly but much better than the OLS method in the prediction performance. Comparing the no outlier with $5 \%$ outliers case, we can see that the estimates based on the ESL method varies more slightly than the other methods, especially in terms of the estimates for $\gamma$; the ESL, M and QR methods are robust to the outliers, while the OLS method is adversely affected by the outliers.

## 7. Conclusions

In this paper, based on the ESL function, we proposed a robust estimation approach for the mean and generalized autoregressive parameters with longitudinal

Table 6. Parameter estimates and prediction results for the CD4 data.

| No outlier |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | OLS | M | QR | ESL | OLS | M | QR | ESL |
| $\beta_{0}$ | 28.8062 | 28.4158 | 28.0787 | 28.8102 | 40.6204 | 28.7359 | 28.1893 | 28.3733 |
| $\beta_{1}$ | -0.0057 | 0.0016 | 0.0129 | 0.0554 | 0.2781 | 0.0255 | 0.0458 | 0.0515 |
| $\beta_{2}$ | 0.8205 | 0.7354 | 0.5780 | 1.9846 | 1.8209 | 0.9624 | 0.8708 | 1.8776 |
| $\beta_{3}$ | 0.7355 | 0.9075 | 1.1222 | 0.1831 | 4.9434 | 1.1518 | 1.4320 | 0.6300 |
| $\beta_{4}$ | -0.0417 | -0.0375 | -0.0194 | 0.0712 | -0.2606 | -0.0716 | -0.0944 | 0.0781 |
| $\beta_{5}$ | -0.0370 | -0.0263 | -0.0095 | -0.0239 | 0.2315 | -0.0202 | -0.0150 | -0.0159 |
| $\beta_{6}$ | -4.6644 | -4.4641 | -4.2399 | -8.9039 | -10.1189 | -5.2498 | -5.2998 | -8.2283 |
| $\beta_{7}$ | 0.6175 | 0.5793 | 0.5339 | 3.1414 | 1.9313 | 0.8235 | 0.8420 | 2.7472 |
| $\gamma_{0}$ | 0.7108 | 0.7118 | 0.7240 | 0.7575 | -0.0061 | 0.0022 | 0.0006 | 0.6983 |
| $\gamma_{1}$ | -0.6340 | -0.6291 | -0.6446 | -0.6536 | 0.0302 | 0.0016 | 0.0044 | -0.5495 |
| $\gamma_{2}$ | 0.1994 | 0.1965 | 0.1999 | 0.1908 | -0.0162 | -0.0008 | -0.0022 | 0.1402 |
| $\gamma_{3}$ | -0.0204 | -0.0202 | -0.0202 | -0.0183 | 0.0020 | 0.0001 | 0.0002 | -0.0116 |
| MAPE | 3.9138 | 3.9052 | 3.9269 | 4.7627 | 15.2271 | 3.9738 | 3.9743 | 4.2478 |

data. The generalized autoregressive parameters that resulted from the MCD of the covariance matrix are unconstrained and can be well interpreted by statistical concepts in the framework of time series. Then the mean and generalized autoregressive parameters can be estimated via linear regression models using the ESL function. Moreover, the balance between the robustness and efficiency can be achieved by choosing appropriate data adaptive tuning parameters. Under certain conditions, we established the theoretical properties. Simulation studies and real data analysis were also carried out to illustrate the finite sample performance of our approach.

Several further problems need to be investigated. First, the dimension of the covariates in regression models is assumed to be fixed, thus it is interesting to extend our approach to the high-dimensional settings. Second, the models can be extended to nonparametric and semiparametric models. For more discussion along this line, references including [10] [29] may be helpful. Moreover, this paper targets the conditional mean of the response given covariates, which suffers from difficulties when the conditional distribution of the response is asymmetric. In this case, the conditional mode may be a more useful summary than the conditional mean, and thus modal linear regression [30] may be an interesting problem.

## Acknowledgements

We thank the editor and the referee for their comments. This work was supported by the National Natural Science Foundation of China (No. 11971001) and the Beijing Natural Science Foundation (No. 1182002).

## Conflicts of Interest

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

## References

[1] Diggle, P.J., Heagerty, P.J., Liang, K.-Y. and Zeger, S. (2002) Analysis of Longitudinal Data. Oxford University Press, Oxford.
[2] Diggle, P.J. and Verbyla, A.P. (1998) Nonparametric Estimation of Covariance Structure in Longitudinal Data. Biometrics, 91, 403-415. https://doi.org/10.2307/3109751
[3] Liang, K.-Y. and Zeger, S.L. (1986) Longitudinal Data Analysis Using Generalized Linear Models. Biometrika, 73, 13-22. https://doi.org/10.1093/biomet/73.1.13
[4] Pourahmadi, M. (1999) Joint Mean-Covariance Models with Applications to Longitudinal Data: Unconstrained Parameterisation. Biometrika, 86, 677-690. https://doi.org/10.1093/biomet/86.3.677
[5] Pourahmadi, M. (2000) Maximum Likelihood Estimation of Generalised Linear Models for Multivariate Normal Covariance Matrix. Biometrika, 87, 425-435. https://doi.org/10.1093/biomet/87.2.425
[6] Pan, J. and Mackenzie, G. (2003) On Modelling Mean-Covariance Structures in Longitudinal Studies. Biometrika, 90, 239-244. https://doi.org/10.1093/biomet/90.1.239
[7] Ye, H. and Pan, J. (2006) Modelling of Covariance Structures in Generalised Estimating Equations for Longitudinal Data. Biometrika, 93, 927-941. https://doi.org/10.1093/biomet/93.4.927
[8] Pourahmadi, M. (2007) Cholesky Decompositions and Estimation of a Covariance Matrix: Orthogonality of Variance-Correlation Parameters. Biometrika, 94, 1006-1013. https://doi.org/10.1093/biomet/asm073
[9] Daniels, M.J. and Pourahmadi, M. (2009) Modeling Covariance Matrices via Partial Autocorrelations. Journal of Multivariate Analysis, 100, 2352-2363. https://doi.org/10.1016/j.jmva.2009.04.015
[10] 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. https://doi.org/10.1198/jasa.2009.tm08485
[11] Zhang, W. and Leng, C. (2012) A Moving Average Cholesky Factor Model in Covariance Modelling for Longitudinal Data. Biometrika, 99, 141-150. https://doi.org/10.1093/biomet/asr068
[12] Zhang, W., Leng, C. and Tang, C.Y. (2015) A Joint Modelling Approach for Longitudinal Studies. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 77, 219-238. https://doi.org/10.1111/rssb. 12065
[13] Maadooliat, M., Pourahmadi, M. and Huang, J.Z. (2013) Robust Estimation of the Correlation Matrix of Longitudinal Data. Statistics and Computing, 23, 17-28. https://doi.org/10.1007/s11222-011-9284-6
[14] Zheng, X., Fung, W.K. and Zhu, Z. (2013) Robust Estimation in Joint Mean-Covariance Regression Model for Longitudinal Data. Annals of the Institute of Statistical Mathematics, 65, 617-638. https://doi.org/10.1007/s10463-012-0383-8
[15] Zheng, X., Fung, W.K. and Zhu, Z. (2014) Variable Selection in Robust Joint Mean and Covariance Model for Longitudinal Data Analysis. Statistica Sinica, 24,

515-531.
[16] Lv, J. and Guo, C. (2017) Efficient Parameter Estimation via Modified Cholesky Decomposition for Quantile Regression with Longitudinal Data. Computational Statistics, 32, 947-975. https://doi.org/10.1007/s00180-017-0714-6
[17] Lv, J., Guo, C., Yang, H. and Li, Y. (2017) A Moving Average Cholesky Factor Model in Covariance Modeling for Composite Quantile Regression with Longitudinal Data. Computational Statistics \& Data Analysis, 112, 129-144. https://doi.org/10.1016/j.csda.2017.02.015
[18] Lv, J., Guo, C. and Wu, J. (2019) Smoothed Empirical Likelihood Inference via the Modified Cholesky Decomposition for Quantile Varying Coefficient Models with Longitudinal Data. TEST, 28, 999-1032. https://doi.org/10.1007/s11749-018-0616-0
[19] Lv, J. and Guo, C. (2019) Quantile Estimations via Modified Cholesky Decomposition for Longitudinal Single-Index Models. Annals of the Institute of Statistical Mathematics, 71, 1163-1199. https://doi.org/10.1007/s10463-018-0673-x
[20] 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. https://doi.org/10.1080/01621459.2013.766613
[21] Lv, J., Yang, H. and Guo, C. (2015) An Efficient and Robust Variable Selection Method for Longitudinal Generalized Linear Models. Computational Statistics \& Data Analysis, 82, 74-88. https://doi.org/10.1016/j.csda.2014.08.006
[22] Wang, K. and Lin, L. (2016) Robust and Efficient Estimator for Simultaneous Model Structure Identification and Variable Selection in Generalized Partial Linear Varying Coefficient Models with Longitudinal Data. Statistical Papers, 1-28.
[23] Li, S., Wang, K. and Ren, Y. (2018) Robust Estimation and Empirical Likelihood Inference with Exponential Squared Loss for Panel Data Models. Economics Letters, 164, 19-23. https://doi.org/10.1016/j.econlet.2017.12.029
[24] Yang, H., Lv, J. and Guo, C. (2015) Robust Variable Selection in Modal Vary-ing-Coefficient Models with Longitudinal. Journal of Statistical Computation and Simulation, 85, 3064-3079. https://doi.org/10.1080/00949655.2014.949716
[25] Wang, K., Li, S., Sun, X. and Lin, L. (2019) Modal Regression Statistical Inference for Longitudinal Data Semivarying Coefficient Models: Generalized Estimating Equations, Empirical Likelihood and Variable Selection. Computational Statistics \& Data Analysis, 133, 257-276. https://doi.org/10.1016/j.csda.2018.10.010
[26] Huber, P.J. (2011) Robust Statistics. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04898-2 594
[27] Koenker, R. (2005) Quantile Regression (Econometric Society Monographs; No. 38). Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511754098
[28] Zeger, S.L. and Diggle, P.J. (1994) Semiparametric Models for Longitudinal Data with Application to CD4 Cell Numbers in HIV Seroconverters. Biometrics, 50, 689-699. https://doi.org/10.2307/2532783
[29] Yao, W. and Li, R. (2013) New Local Estimation Procedure for a Non-Parametric Regression Function for Longitudinal Data. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 75, 123-138. https://doi.org/10.1111/j.1467-9868.2012.01038.x
[30] Yao, W. and Li, L. (2014) A New Regression Model: Modal Linear Regression. Scandinavian Journal of Statistics, 41, 656-671. https://doi.org/10.1111/sjos. 12054

## Appendix

Lemma 1 If regularity conditions (C1)-(C5) hold, then with probability approaching to 1 , there exists a local maximizer of $(2)$, denoted as $\tilde{\beta}$, such that

$$
\left\|\tilde{\beta}-\beta_{0}\right\|=O_{P}\left(\frac{1}{\sqrt{n}}\right)
$$

Proof of Lemma 1. Let $R_{1}(\beta)=\frac{1}{n} L_{1}(\beta)=\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_{i}} \phi_{\tau_{1}}\left(Y_{i j}-X_{i j}^{\mathrm{T}} \beta\right)$. It suffices to show that for any given $\rho>0$, there exists a large constant $C>0$ such that

$$
\begin{equation*}
P\left\{\sup _{\|v\|=C} R_{1}\left(\beta_{0}+v / \sqrt{n}\right)<R_{1}\left(\beta_{0}\right)\right\} \geq 1-\rho, \tag{1}
\end{equation*}
$$

for any $p$-dimensional vector $v$ satisfying that $\|v\|=C$. Based on the Taylor expansion, we have

$$
\begin{aligned}
& R_{1}\left(\beta_{0}+v / \sqrt{n}\right)-R_{1}\left(\beta_{0}\right) \\
& =\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_{i}}\left\{\phi_{\tau_{1}}\left(Y_{i j}-X_{i j}^{\mathrm{T}}\left(\beta_{0}+v / \sqrt{n}\right)\right)-\phi_{\tau_{1}}\left(Y_{i j}-X_{i j}^{\mathrm{T}} \beta_{0}\right)\right\} \\
& =\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_{i}}\left\{\phi_{\tau_{1}}\left(\varepsilon_{i j}-v^{\mathrm{T}} X_{i j} / \sqrt{n}\right)-\phi_{\tau_{1}}\left(\varepsilon_{i j}\right)\right\} \\
& =\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_{i}}\left\{-\phi_{\tau_{1}}^{\prime}\left(\varepsilon_{i j}\right) v^{\mathrm{T}} X_{i j} / \sqrt{n}+\frac{1}{2 n} \phi_{\tau_{1}}^{\prime \prime}\left(\varepsilon_{i j}\right)\left(v^{\mathrm{T}} X_{i j}\right)^{2}-\frac{1}{6 n^{\frac{3}{2}}} \phi_{\tau_{1}}^{\prime \prime \prime}\left(\varepsilon_{i j}^{*}\right)\left(v^{\mathrm{T}} X_{i j}\right)^{3}\right\} \\
& \triangleq I_{1}+I_{2}+I_{3}
\end{aligned}
$$

where $\varepsilon_{i j}^{*}$ lies between $\varepsilon_{i j}-v^{\mathrm{T}} X_{i j} / \sqrt{n}$ and $\varepsilon_{i j}$. Then, for $I_{1}$, we have

$$
\begin{aligned}
E\left(I_{1}\right) & =-\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_{i}} E\left\{\phi_{\tau_{1}}^{\prime}\left(\varepsilon_{i j}\right) v^{\mathrm{T}} X_{i j} / \sqrt{n}\right\} \\
& =-\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_{i}} E\left\{E\left(\phi_{\tau_{1}}^{\prime}\left(\varepsilon_{i j}\right) \mid X_{i j}\right) v^{\mathrm{T}} X_{i j} / \sqrt{n}\right\}=0
\end{aligned}
$$

and

$$
\operatorname{var}\left(I_{1}\right)=E\left\{\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_{i}} \phi_{\tau_{1}}^{\prime}\left(\varepsilon_{i j}\right) v^{\mathrm{T}} X_{i j} / \sqrt{n}\right\}^{2}=\frac{1}{n^{3}} \sum_{i=1}^{n} E\left\{\sum_{j=1}^{m_{i}} \phi_{\tau_{1}}^{\prime}\left(\varepsilon_{i j}\right) v^{\mathrm{T}} X_{i j}\right\}^{2} .
$$

Using the $C_{r}$ inequality and condition (C1), we can obtain that

$$
E\left\{\sum_{j=1}^{m_{i}} \phi_{\tau_{1}}^{\prime}\left(\varepsilon_{i j}\right) v^{\mathrm{T}} X_{i j}\right\}^{2} \leq m_{i} \sum_{j=1}^{m_{i}} E\left\{\phi_{\tau_{1}}^{\prime}\left(\varepsilon_{i j}\right) v^{\mathrm{T}} X_{i j}\right\}^{2} \leq M \sum_{j=1}^{m_{i}} E\left\{\phi_{\tau_{1}}^{\prime}\left(\varepsilon_{i j}\right) v^{\mathrm{T}} X_{i j}\right\}^{2}
$$

for $i=1, \cdots, n$, then we have

$$
\begin{aligned}
\operatorname{var}\left(I_{1}\right) & \leq \frac{M}{n^{3}} \sum_{i=1}^{n} \sum_{j=1}^{m_{i}} v^{\mathrm{T}} E\left\{\phi_{\tau_{1}}^{\prime}\left(\varepsilon_{i j}\right)^{2} X_{i j} X_{i j}^{\mathrm{T}}\right\} v \\
& =\frac{M}{n^{3}} \sum_{i=1}^{n} \sum_{j=1}^{m_{i}} v^{\mathrm{T}} E\left\{G_{1}\left(X_{i j}, \tau_{1}\right) X_{i j} X_{i j}^{\mathrm{T}}\right\} v \\
& =O\left(\|v\|^{2} n^{-2}\right)=O\left(C^{2} n^{-2}\right) .
\end{aligned}
$$

Therefore, we can deduce that $I_{1}=E\left(I_{1}\right)+O_{P}\left(\sqrt{\operatorname{var}\left(I_{1}\right)}\right)=O_{P}\left(C n^{-1}\right)$. Similarly, we get $I_{3}=O_{P}\left(n^{-3 / 2}\right)$. For $I_{2}$, we have

$$
\begin{aligned}
I_{2} & =\frac{1}{2 n^{2}} \sum_{i=1}^{n} \sum_{j=1}^{m_{i}} v^{\mathrm{T}} E\left(F_{1}\left(X_{i j}, \tau_{1}\right) X_{i j} X_{i j}^{\mathrm{T}}\right) v\left(1+o_{P}(1)\right) \\
& =O\left(\|v\|^{2} n^{-1}\right)=O\left(C^{2} n^{-1}\right) .
\end{aligned}
$$

Note that $\|v\|=C$, we can choose a sufficiently large $C$ such that $I_{2}$ dominates both $I_{1}$ and $I_{3}$ with a probability of at least $1-\rho$. From condition (C3), we get $F_{1}\left(x, \tau_{1}\right)<0$, then (1) holds. The proof of Lemma 1 is finished.

Proof of Theorem 1. Let $\tilde{\varphi}_{i j}=X_{i j}^{\mathrm{T}}\left(\tilde{\beta}-\beta_{0}\right)$, then $\tilde{\beta}$ satisfies the following equation:

$$
\begin{aligned}
0 & =\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_{i}} X_{i j} \phi_{\tau_{1}}^{\prime}\left(\varepsilon_{i j}-\tilde{\varphi}_{i j}\right) \\
& =\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_{i}} X_{i j}\left\{\phi_{\tau_{1}}^{\prime}\left(\varepsilon_{i j}\right)-\phi_{\tau_{1}}^{\prime \prime}\left(\varepsilon_{i j}\right) \tilde{\varphi}_{i j}+\frac{1}{2} \phi_{\tau_{1}}^{\prime \prime \prime}\left(\varepsilon_{i j}^{*}\right) \tilde{\varphi}_{i j}^{2}\right\} \\
& \triangleq I_{4}+I_{5}+I_{6}
\end{aligned}
$$

where $\varepsilon_{i j}^{*}$ locates between $\varepsilon_{i j}-\tilde{\varphi}_{i j}$ and $\varepsilon_{i j}$. It can be easily shown that

$$
\begin{aligned}
I_{5} & =-\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_{i}} \phi_{\tau_{1}}^{\prime \prime}\left(\varepsilon_{i j}\right) X_{i j} X_{i j}^{\mathrm{T}}\left(\tilde{\beta}-\beta_{0}\right) \\
& =-\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_{i}} E\left(F_{1}\left(X_{i j}, \tau_{1}\right) X_{i j} X_{i j}^{\mathrm{T}}\right)\left(\tilde{\beta}-\beta_{0}\right)+o_{P}(1)
\end{aligned}
$$

Since $\left|\tilde{\varphi}_{i j}\right|=\left|X_{i j}^{\mathrm{T}}\left(\tilde{\beta}-\beta_{0}\right)\right|=O_{P}\left(\left\|\tilde{\beta}-\beta_{0}\right\|\right)$, then $\left|\tilde{\varphi}_{i j}\right|^{2}=O_{P}\left(\left\|\tilde{\beta}-\beta_{0}\right\|^{2}\right)$. Thus, $I_{6}=o_{P}\left(I_{5}\right)$. Then,

$$
\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_{i}} E\left(F_{1}\left(X_{i j}, \tau_{1}\right) X_{i j} X_{i j}^{\mathrm{T}}\right) \sqrt{n}\left(\tilde{\beta}-\beta_{0}\right)=\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{m_{i}} X_{i j} \phi_{\tau_{1}}^{\prime}\left(\varepsilon_{i j}\right)+o_{P}(1)
$$

Notice that $E\left(X_{i j} \phi_{\tau_{1}}^{\prime}\left(\varepsilon_{i j}\right)\right)=0$. For any $p \times 1$ vector $\kappa$, let $e_{i}=\kappa^{\mathrm{T}} U_{i}$, where $U_{i}=\sum_{j=1}^{m_{i}} X_{i j} \phi_{\tau_{1}}^{\prime}\left(\varepsilon_{i j}\right)=X_{i}^{\mathrm{T}} \phi_{\tau_{1}}^{\prime}\left(\varepsilon_{i}\right)$. By $\mathrm{C}_{\mathrm{r}}$ inequality and condition (C1), we have

$$
\begin{aligned}
E\left|e_{i}\right|^{3} & \leq M \sum_{j=1}^{m_{i}} E\left|\left(\kappa^{\mathrm{T}} X_{i j}\right) \phi_{\tau_{1}}^{\prime}\left(\varepsilon_{i j}\right)\right|^{3} \\
& =M \sum_{j=1}^{m_{i}} E\left\{E\left(\left|\phi_{\tau_{1}}^{\prime}\left(\varepsilon_{i j}\right)\right|^{3} \mid X_{i j}\right)\left|\kappa^{\mathrm{T}} X_{i j}\right|^{3}\right\} \\
& \leq K_{1},
\end{aligned}
$$

where $K_{1}$ is a constant independent of $i$. Then by condition (C5), $\frac{1}{n} \sum_{i=1}^{n} E\left(U_{i} U_{i}^{\mathrm{T}}\right)$ converges to a finite positive definite matrix, which is denoted by $\Sigma_{2}$ here. Thus, we obtain that

$$
\frac{1}{n} \sum_{i=1}^{n} \operatorname{var}\left(e_{i}\right)=\frac{1}{n} \sum_{i=1}^{n} E\left(e_{i}^{2}\right)=\frac{1}{n} \sum_{i=1}^{n} \kappa^{\mathrm{T}} E\left(U_{i} U_{i}^{\mathrm{T}}\right) \kappa \rightarrow \kappa^{\mathrm{T}} \Sigma_{2} \kappa>0 .
$$

So the proof of theorem 1 is completed by the Lyapunov central limit theorem and Slutsky's theorem.

Lemma 2 If regularity conditions (C1)-(C9) hold, then with probability approaching to 1 , there exists a local maximizer of (7), denoted as $\hat{\theta}$, such that

$$
\left\|\hat{\theta}-\theta_{0}\right\|=O_{P}\left(\frac{1}{\sqrt{n}}\right)
$$

Proof of Lemma 2. Since $\left\|\tilde{\beta}-\beta_{0}\right\|=O_{P}\left(\frac{1}{\sqrt{n}}\right)$, we can replace $\tilde{\beta}$ with $\beta_{0}$ during the proof from now on. Then $\tilde{\delta}_{i j}=\left(X_{i j}^{\mathrm{T}}, \tilde{\zeta}_{i j}^{\mathrm{T}}\right)^{\mathrm{T}}$ can be substituted by

$$
\begin{aligned}
& \delta_{i j}=\left(X_{i j}^{\mathrm{T}}, \zeta_{i j}^{\mathrm{T}}\right)^{\mathrm{T}} \text {, where } \zeta_{i j}=\left(\sum_{k=1}^{j-1} \varepsilon_{i, k} Z_{j, k, 1}^{(i)}, \cdots, \sum_{k=1}^{j-1} \varepsilon_{i, k} Z_{j, k, q}^{(i)}\right)^{\mathrm{T}} \text {. Let } \\
& \qquad R_{2}(\theta)=\frac{1}{n} \sum_{i=1}^{n} \sum_{j=2}^{m_{i}} \phi_{\tau_{2}}\left(Y_{i j}-X_{i j}^{\mathrm{T}} \beta-\zeta_{i j}^{\mathrm{T}} \gamma\right)=\frac{1}{n} \sum_{i=1}^{n} \sum_{j=2}^{m_{i}} \phi_{\tau_{2}}\left(Y_{i j}-\delta_{i j}^{\mathrm{T}} \theta\right) .
\end{aligned}
$$

It suffices to show that for any given $\rho>0$, there exists a large constant $C>0$ such that

$$
\begin{equation*}
P\left\{\lim _{\|v\|=C} R_{2}\left(\theta_{0}+v / \sqrt{n}\right)<R_{2}\left(\theta_{0}\right)\right\} \geq 1-\rho, \tag{2}
\end{equation*}
$$

for any $(p+q)$-dimensional vector $v$ satisfying that $\|v\|=C$. Based on the Taylor expansion, we have

$$
\begin{aligned}
& R_{2}\left(\theta_{0}+v / \sqrt{n}\right)-R_{2}\left(\theta_{0}\right) \\
& =\frac{1}{n} \sum_{i=1}^{n} \sum_{j=2}^{m_{i}}\left\{\phi_{r_{2}}\left(Y_{i j}-\delta_{i j}^{\mathrm{T}}\left(\theta_{0}+v / \sqrt{n}\right)\right)-\phi_{\tau_{2}}\left(Y_{i j}-\delta_{i j}^{\mathrm{T}} \theta_{0}\right)\right\} \\
& =\frac{1}{n} \sum_{i=1}^{n} \sum_{j=2}^{m_{i}}\left\{\phi_{\tau_{2}}\left(\epsilon_{i j}-v^{\mathrm{T}} \delta_{i j} / \sqrt{n}\right)-\phi_{\tau_{2}}\left(\epsilon_{i j}\right)\right\} \\
& =\frac{1}{n} \sum_{i=1}^{n} \sum_{j=2}^{m_{i}}\left\{-\phi_{\tau_{2}}^{\prime}\left(\epsilon_{i j}\right) v^{\mathrm{T}} \delta_{i j} / \sqrt{n}+\frac{1}{2 n} \phi_{\tau_{2}}^{\prime \prime}\left(\epsilon_{i j}\right)\left(v^{\mathrm{T}} \delta_{i j}\right)^{2}-\frac{1}{6 n^{\frac{3}{2}}} \phi_{\tau_{2}}^{\prime \prime \prime}\left(\epsilon_{i j}^{*}\right)\left(v^{\mathrm{T}} \delta_{i j}\right)^{3}\right\} \\
& \triangleq I_{7}+I_{8}+I_{9},
\end{aligned}
$$

where $\epsilon_{i j}^{*}$ lies between $\epsilon_{i j}-v^{\mathrm{T}} \delta_{i j} / \sqrt{n}$ and $\epsilon_{i j}$. Recall that $\bar{X}_{i j}=\left(X_{i 1}, \cdots, X_{i j}\right)^{\mathrm{T}}$. It should be emphasized that $\delta_{i j}$ consists of both covariates and $\varepsilon_{i 1}, \cdots, \varepsilon_{i, j-1}$, and thus $\epsilon_{i j}$ is independent of $\delta_{i j}$ conditional on $\bar{X}_{i j}$. In addition, from (8), we have

$$
E\left(\delta_{i j} \delta_{i j}^{\mathrm{T}}\right)=\left(\begin{array}{cc}
E\left(X_{i j} X_{i j}^{\mathrm{T}}\right) & 0 \\
0 & E\left(\zeta_{i j} \zeta_{i j}^{\mathrm{T}}\right)
\end{array}\right)
$$

Then, for $I_{7}$, we have

$$
\begin{aligned}
E\left(I_{7}\right) & =-\frac{1}{n} \sum_{i=1}^{n} \sum_{j=2}^{m_{i}} E\left\{\phi_{\tau_{2}}^{\prime}\left(\epsilon_{i j}\right) v^{\mathrm{T}} \delta_{i j} / \sqrt{n}\right\} \\
& =-\frac{1}{n} \sum_{i=1}^{n} \sum_{j=2}^{m_{i}} E\left\{E\left(\phi_{\tau_{2}}^{\prime}\left(\epsilon_{i j}\right) \mid X_{i j}\right) E\left(v^{\mathrm{T}} \delta_{i j} \mid \bar{X}_{i j}\right) / \sqrt{n}\right\} \\
& =0
\end{aligned}
$$

and

$$
\begin{aligned}
\operatorname{var}\left(I_{7}\right) & =E\left\{\frac{1}{n} \sum_{i=1}^{n} \sum_{j=2}^{m_{i}} \phi_{\tau_{2}}^{\prime}\left(\epsilon_{i j}\right) v^{\mathrm{T}} \delta_{i j} / \sqrt{n}\right\}^{2} \\
& =\frac{1}{n^{3}} \sum_{i=1}^{n} \sum_{j=2}^{m_{i}} v^{\mathrm{T}} E\left\{\phi_{\tau_{2}}^{\prime}\left(\epsilon_{i j}\right)^{2} \delta_{i j} \delta_{i j}^{\mathrm{T}}\right\} v \\
& =\frac{1}{n^{3}} \sum_{i=1}^{n} \sum_{j=2}^{m_{i}} v^{\mathrm{T}} E\left\{G_{2}\left(X_{i j}, \tau_{2}\right) E\left(\delta_{i j} \delta_{i j}^{\mathrm{T}} \mid \bar{X}_{i j}\right)\right\} v \\
& =O\left(\|v\|^{2} n^{-2}\right)=O\left(C^{2} n^{-2}\right) .
\end{aligned}
$$

Therefore, we can deduce that $I_{7}=E\left(I_{7}\right)+O_{P}\left(\sqrt{\operatorname{var}\left(I_{7}\right)}\right)=O_{P}\left(\mathrm{Cn}^{-1}\right)$. Similarly, we get $I_{9}=O_{P}\left(n^{-3 / 2}\right)$. For $I_{8}$, we have

$$
\begin{aligned}
I_{8} & =\frac{1}{2 n^{2}} \sum_{i=1}^{n} \sum_{j=2}^{m_{i}} v^{\mathrm{T}} E\left\{F_{2}\left(X_{i j}, \tau_{2}\right) E\left(\delta_{i j} \delta_{i j}^{\mathrm{T}} \mid \bar{X}_{i j}\right)\right\} v\left(1+o_{P}(1)\right) \\
& =O\left(\|\nu\|^{2} n^{-1}\right)=O\left(C^{2} n^{-1}\right) .
\end{aligned}
$$

Note that $\|v\|=C$, we can choose a sufficiently large $C$ such that $I_{8}$ dominates both $I_{7}$ and $I_{9}$ with a probability of at least $1-\rho$. From condition (C8), we get $F_{2}\left(x, \tau_{2}\right)<0$, then (2) holds. The proof of Lemma 2 is finished.

Proof of Theorem 2. Let $\hat{\alpha}_{i j}=\delta_{i j}^{\mathrm{T}}\left(\hat{\theta}-\theta_{0}\right)$, then $\hat{\theta}$ satisfies the following equation:

$$
\begin{aligned}
0 & =\frac{1}{n} \sum_{i=1}^{n} \sum_{j=2}^{m_{i}} \delta_{i j} \phi_{\tau_{2}}^{\prime}\left(\epsilon_{i j}-\hat{\alpha}_{i j}\right) \\
& =\frac{1}{n} \sum_{i=1}^{n} \sum_{j=2}^{m_{i}} \delta_{i j}\left\{\phi_{\tau_{2}}^{\prime}\left(\epsilon_{i j}\right)-\phi_{\tau_{2}}^{\prime}\left(\epsilon_{i j}\right) \hat{\alpha}_{i j}+\frac{1}{2} \phi_{\tau_{2}}^{\prime \prime \prime}\left(\epsilon_{i j}^{*}\right) \hat{\alpha}_{i j}^{2}\right\} \\
& \triangleq I_{10}+I_{11}+I_{12}
\end{aligned}
$$

where $\epsilon_{i j}^{*}$ locates between $\epsilon_{i j}-\hat{\alpha}_{i j}$ and $\epsilon_{i j}$. It can be easily obtained that

$$
\begin{aligned}
I_{11} & =-\frac{1}{n} \sum_{i=1}^{n} \sum_{j=2}^{m_{i}} \phi_{\tau_{2}}^{\prime \prime}\left(\epsilon_{i j}\right) \delta_{i j} \delta_{i j}^{\mathrm{T}}\left(\hat{\theta}-\theta_{0}\right) \\
& =-\frac{1}{n} \sum_{i=1}^{n} \sum_{j=2}^{m_{i}} E\left\{F_{2}\left(X_{i j}, \tau_{2}\right) E\left(\delta_{i j} \delta_{i j}^{\mathrm{T}} \mid \bar{X}_{i j}\right)\right\}\left(\hat{\theta}-\theta_{0}\right)+o_{P}(1),
\end{aligned}
$$

and

$$
I_{12}=\frac{1}{2 n} \sum_{i=1}^{n} \sum_{j=2}^{m_{i}} \phi_{\tau_{2}}^{\prime \prime \prime}\left(\epsilon_{i j}^{*}\right) \delta_{i j}\left\{\delta_{i j}^{\mathrm{T}}\left(\hat{\theta}-\theta_{0}\right)\right\}^{2}=O_{P}\left(\left\|\hat{\theta}-\theta_{0}\right\|^{2}\right)=o_{P}\left(I_{11}\right) .
$$

Then
$\frac{1}{n} \sum_{i=1}^{n} \sum_{j=2}^{m_{i}} E\left\{F_{2}\left(X_{i j}, \tau_{2}\right) E\left(\delta_{i j} \delta_{i j}^{\mathrm{T}} \mid \bar{X}_{i j}\right)\right\} \sqrt{n}\left(\hat{\theta}-\theta_{0}\right)=\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=2}^{m_{i}} \delta_{i j} \phi_{\tau_{2}}^{\prime}\left(\epsilon_{i j}\right)+o_{P}(1)$.
Notice that $E\left(\delta_{i j} \phi_{t_{2}}^{\prime}\left(\epsilon_{i j}\right)\right)=0$. For any $(p+q) \times 1$ vector $\eta$, let
$\psi_{i}=\sum_{j=2}^{m_{i}}\left(\eta^{\mathrm{T}} \delta_{i j}\right) \phi_{\tau_{2}}^{\prime}\left(\epsilon_{i j}\right)$. By $C_{r}$ inequality and condition (C1), we have

$$
\begin{aligned}
E\left|\psi_{i}\right|^{3} & \leq M \sum_{j=2}^{m_{i}} E\left|\left(\eta^{\mathrm{T}} \delta_{i j}\right) \phi_{\tau_{2}}^{\prime}\left(\epsilon_{i j}\right)\right|^{3} \\
& =M \sum_{j=2}^{m_{i}} E\left\{E\left(\left|\phi_{\tau_{2}}^{\prime}\left(\epsilon_{i j}\right)\right|^{3} \mid X_{i j}\right) E\left(\left|\eta^{\mathrm{T}} \delta_{i j}\right|^{3} \mid \bar{X}_{i j}\right)\right\} \\
& \leq K_{2},
\end{aligned}
$$

where $K_{2}$ is a constant independent of $i$. Moreover,

$$
\begin{aligned}
\frac{1}{n} \sum_{i=1}^{n} \operatorname{var}\left(\psi_{i}\right) & =\frac{1}{n} \sum_{i=1}^{n} E\left(\psi_{i}^{2}\right)=\frac{1}{n} \sum_{i=1}^{n} E\left\{\sum_{j=2}^{m_{i}}\left(\eta^{\mathrm{T}} \delta_{i j}\right) \phi_{\tau_{2}}^{\prime}\left(\epsilon_{i j}\right)\right\}^{2} \\
& =\frac{1}{n} \sum_{i=1}^{n} \sum_{j=2}^{m_{i}} E\left\{\left(\eta^{\mathrm{T}} \delta_{i j}\right)^{2} \phi_{\tau_{2}}^{\prime}\left(\epsilon_{i j}\right)^{2}\right\} \\
& =\frac{1}{n} \sum_{i=1}^{n} \sum_{j=2}^{m_{i}} \eta^{\mathrm{T}} E\left\{G_{2}\left(X_{i j}, \tau_{2}\right) E\left(\delta_{i j} \delta_{i j}^{\mathrm{T}} \mid \bar{X}_{i j}\right)\right\} \eta \\
& \rightarrow K_{3}>0 .
\end{aligned}
$$

So the proof is completed by the Lyapunov central limit theorem and Slutsky's theorem.

## List of Main Symbols

| $\triangleq$ | "Define as" or "Denote as" |
| :---: | :---: |
| $I(A)$ | Indicative function of set $A$ |
| T | Vector or matrix transposition |
| $\xi=o(1)$ | $\xi$ is infinitesimal |
| $\xi=O(1)$ | $\xi$ is a bounded variable |
| $\xi_{n}=o_{P}\left(\eta_{n}\right)$ | For any $\varepsilon>0, P\left(\left\\|\xi_{n}\right\\| \geq \varepsilon\left\\|\eta_{n}\right\\|\right) \rightarrow 0$ |
| $\xi_{n}=O_{P}\left(\eta_{n}\right)$ | $\xi_{n} / \eta_{n}$ is bounded in probability |
| $N(\mu, \Sigma)$ | Normal distribution with mean $\mu$ and covariance matrix $\Sigma$ |
| $E(\cdot)$ | Expectation |
| $\operatorname{var}(\cdot)$ | Variance |
| $\\|\cdot\\|$ | $\ell_{2}$ norm |

