

Robust estimation in joint mean–covariance regression model for longitudinal data

Xueying Zheng · Wing Kam Fung · Zhongyi Zhu

Received: 14 December 2011 / Revised: 17 July 2012 / Published online: 27 November 2012
© The Institute of Statistical Mathematics, Tokyo 2012

Abstract In this paper, we develop robust estimation for the mean and covariance jointly for the regression model of longitudinal data within the framework of generalized estimating equations (GEE). The proposed approach integrates the robust method and joint mean–covariance regression modeling. Robust generalized estimating equations using bounded scores and leverage-based weights are employed for the mean and covariance to achieve robustness against outliers. The resulting estimators are shown to be consistent and asymptotically normally distributed. Simulation studies are conducted to investigate the effectiveness of the proposed method. As expected, the robust method outperforms its non-robust version under contaminations. Finally, we illustrate by analyzing a hormone data set. By downweighing the potential outliers, the proposed method not only shifts the estimation in the mean model, but also shrinks the range of the innovation variance, leading to a more reliable estimation in the covariance matrix.

Keywords Covariance matrix · Generalized estimating equation · Longitudinal data · Modified Cholesky decomposition · Robustness

X. Zheng · W. K. Fung (✉)
Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, China
e-mail: wingfung@hku.hk

X. Zheng
e-mail: xyzheng@hku.hk

Z. Zhu
Department of Statistics, School of Management, Fudan University, Shanghai 200433, China
e-mail: zhuzy@fudan.edu.cn

1 Introduction

Longitudinal data are often characterized by the dependence of repeated observations over time within the same subject. Observations within the same subject are prone to be correlated. In a marginal model, the typical correlation among a subject's repeated measurements is not of primary interest, but it should be taken into account to make proper inference. In fact, ignoring the within-subject correlation can result in an inefficient estimator of a regression model, see [Qu et al. \(2000\)](#) and [Wang \(2003\)](#). A well-modeled covariance can decrease the bias of mean estimation for longitudinal data including missing values [Daniels and Zhao \(2003\)](#) and maintain a reliable estimation for the covariance matrix even when the mean model is not correctly specified [Pan and Mackenzie \(2003\)](#). In some cases, the correlation structure plays the role as important as the mean structure, which suggests that the estimation of the covariance matrix is crucial in longitudinal study.

To overcome these challenges, there has been substantial recent literature considering the mean and covariance matrix simultaneously. [Pourahmadi \(1999, 2000\)](#) developed a parametric joint mean and covariance model in the framework of GEE by the modified Cholesky decomposition, but their method does not deal with irregular observed measurements. [Pan and Mackenzie \(2003\)](#) exploited a reparameterisation of the marginal covariance matrix to extend Pourahmadi's work to irregular cases. [Wu and Pourahmadi \(2003\)](#) proposed non-parametric smoothing to regularize the estimation of covariance matrix, using the two-step estimation of [Fan and Zhang \(2000\)](#). To relax the parametric assumption in mean and covariance structure, [Fan et al. \(2007\)](#) and [Fan and Wu \(2008\)](#) proposed semiparametric model for the mean and covariance structure. However, they only considered the normal data or nearly normal data. For longitudinal data analysis, [Liang and Zeger \(1986\)](#) introduced the technique of generalized estimating equations and generalized estimating equations were developed for both mean and covariance parameters in [Ye and Pan \(2006\)](#). To relax the parametric assumption posed in [Ye and Pan \(2006\)](#), [Leng et al. \(2010\)](#) proposed a semiparametric model for mean and covariance structure.

GEE approach takes advantage of the built-in robustness since no specification of the full likelihood is required. However, in a longitudinal data set, one outlier in the subject level may generate a set of outliers in the sample due to repeated measurements. Therefore, robustness study is required for the reason that estimating equations are highly sensitive to outliers in the sample. Robust regression methods have been developed for estimation on mean parameters and covariance matrix estimation separately. An incomplete list of recent works on the robust GEE methods includes [Cantoni \(2004\)](#), [He et al. \(2005\)](#), [Wang et al. \(2005a\)](#), [Qin and Zhu \(2007\)](#), [Qin et al. \(2009\)](#) and [Croux et al. \(2012\)](#).

However, as far as we know, there is little discussion on robust estimation on joint mean and covariance model. [Croux et al. \(2012\)](#) considered robustification on the mean and covariance where they set up estimating equations for both the mean and the dispersion parameter. The constraint of their approach is that they assumed an inflexible covariance structure determined by two parameters. In this article, following [Ye and Pan \(2006\)](#) and [He et al. \(2005\)](#), we establish a set of robust generalized estimating equations for analyzing a parametric joint mean and covariance regression

model for longitudinal data. Robust generalized estimating equations using bounded scores and leverage-based weights are developed for the mean and covariance to achieve robustness against outliers. The advantage of the proposed joint model relies on modeling the covariance matrix with moderate number of parameters, rather than assuming a fixed structure or introducing unstructured covariance matrix causing the curse of dimensionality.

Similar to He et al. (2005), the Mallows-type weights are used to downweigh the effect of leverage points when a bounded score function on the Pearson residuals is employed to reduce the effect of outliers in the response. The Mallows-type weights have also been used by Qin and Zhu (2007) in generalized semiparametric mixed model and Qin et al. (2009) in generalized partial linear mixed model for longitudinal data analysis. The resulting estimators for the regression coefficients in both the mean and covariance are shown to be consistent and asymptotically normally distributed. In simulation studies, for one thing, we apply the robust method in the joint model to obtain better estimators for both mean and covariance parameters under contaminations. On the other side, we find robustifications in mean and in covariance matrix estimation are both necessary, owing to the fact that the triple-robust estimating method performs far better than the mean-robust estimating method. In the analysis of the hormone data analysis, the main advantage of the robust method lies in successfully detecting both the subject-level and observation-level potential influential outliers, which results in a more reliable estimation in both the mean and covariance.

The rest of the article is organized as follows: In Sect. 2, we formulate the robust joint mean and covariance model and introduce the estimation methods. Theoretical properties are established in this section as well. Simulation studies are presented in Sect. 3. Finally, we carry out a hormone data set analysis to illustrate the proposed method in Sect. 4.

2 Methodology

2.1 Joint mean–covariance model

Suppose that we have a sample of m subjects. Let $y_i = (y_{i1}, \dots, y_{in_i})'$ be the n_i repeated measurements at time point $t_i = (t_{i1}, \dots, t_{in_i})'$ of the i th subject. Let $E(y_i) = \mu_i = (\mu_{i1}, \dots, \mu_{in_i})'$ and $Cov(y_i) = \Sigma_i$ be the $n_i \times 1$ mean vector and $n_i \times n_i$ covariance matrix of y_i , respectively.

According to the positive definite property of the covariance matrix Σ_i , there exists a unique lower triangular matrix Φ_i with 1's being the diagonal entries and a unique diagonal matrix D_i with positive diagonals such that

$$\Phi_i \Sigma_i \Phi_i' = D_i. \quad (1)$$

As indicated by Pourahmadi (1999), Φ_i and D_i have clear statistical interpretation. The lower-diagonal entries of Φ_i are the negatives of the autoregressive coefficients ϕ_{ijk} defined in

$$\hat{y}_{ij} = \mu_{ij} + \sum_{k=1}^{j-1} \phi_{ijk}(y_{ik} - \mu_{ik}), \quad (2)$$

which is the linear least squares predictor of y_{ij} based on its predecessors $y_{i(j-1)}, \dots, y_{i1}$. The diagonal entries σ_{ij}^2 of D_i can be seen as the innovation variance $\sigma_{ij}^2 = \text{var}(\epsilon_{ij})$, where $\epsilon_{ij} = y_{ij} - \hat{y}_{ij}$.

The modified Cholesky decomposition can guarantee Σ_i to be positive definite thus takes the advantage that ϕ_{ijk} and σ_{ij}^2 are unconstrained. Similar to [Ye and Pan \(2006\)](#), we adopt three generalized linear models for the mean, generalized autoregressive parameters and innovation variances:

$$g(\mu_{ij}) = x'_{ij}\beta, \quad \phi_{ijk} = z'_{ijk}\gamma, \quad \log \sigma_{ij}^2 = z'_{ij}\lambda, \quad (3)$$

where $i = 1, \dots, m$, $j = 1, \dots, n_i$, $k = 1, \dots, j - 1$, x_{ij} , z_{ijk} and z_{ij} are $p \times 1$, $q \times 1$ and $d \times 1$ vectors of covariates, and β , γ and λ are the associated parameters. The known link function $g(\cdot)$ is assumed to be monotone and differentiable. The covariates z_{ijk} and z_{ij} may contain the baseline covariates, the time and the associated interactions. Orthogonal form for the polynomials of the time are recommended as the covariates for the autoregressive components by [Ye and Pan \(2006\)](#):

$$z_{ijk} = (1, (t_{ij} - t_{ik}), (t_{ij} - t_{ik})^2, \dots, (t_{ij} - t_{ik})^{q-1}).$$

In the above models, estimation for the autoregressive coefficients and innovation variances are treated as important as the estimation for the mean.

2.2 Robust generalized estimating equations for the mean

To estimate the regression parameter β in Sect. 2.1, generalized estimating equations ([Liang and Zeger 1986](#)) can be represented as

$$\sum_{i=1}^m \dot{\mu}'_i V_i^{-1} (y_i - \mu_i) = 0, \quad (4)$$

where V_i is the covariance matrix of y_i assumed as $V_i = A_i^{1/2} R(\alpha) A_i^{1/2}$ and $\dot{\mu}_i$ is the first derivative of $\mu(\cdot)$ evaluated at $(x_{i1}, \dots, x_{in_i})' \beta$. Here, A_i is a diagonal matrix with the marginal variance of y_i as the diagonal component, and $R(\alpha)$ is a working correlation matrix.

[He et al. \(2005\)](#) mentioned that the GEE approach has some built-in robustness since it requires no specification of the full likelihood. However, estimating equations are highly sensitive to outliers in the sample. In longitudinal studies, an outlier in a subject-level measurement can generate multiple outliers in the sample. [Sinha \(2004\)](#) also emphasized that a small proportion of the data may come from an arbitrary distribution rather than the distribution in the assumption, i.e. the deviations from underlying distributions, which can result in outliers or influential observations in the

data. Therefore, robust method is desirable to mitigate the effect of outliers and to obtain bounded influence functions.

To improve the efficiency of the GEE estimator under contamination, He et al. (2005) defined the following robust generalized estimating equations (RGEE) for the mean parameter:

$$\sum_{i=1}^m X_i' \Delta_i(\mu_i(\beta))(V_i^\beta)^{-1} h_i^\beta(\mu_i(\beta)) = 0,$$

where $X_i = (x_{i1}, \dots, x_{in_i})'$, $\Delta_i(\mu_i(\theta)) = \text{diag}\{\dot{\mu}_{i1}(\beta), \dots, \dot{\mu}_{in_i}(\beta)\}$, with $\dot{\mu}_{ij}(\cdot)$ denoting the first derivative of $\mu(\cdot)$ evaluated at $x'_{ij}\beta$; and $h_i^\beta(\mu_i(\beta)) = W_i^\beta [\psi^\beta(\mu_i(\beta)) - C_i^\beta(\mu_i(\beta))]$. In $h_i^\beta(\mu_i(\beta))$, they employed a Huber function ψ on the Pearson residual and a weighting matrix W_i^β to control the influence from outliers. $C_i^\beta(\mu_i(\beta))$ is developed to ensure the consistency of the estimating equation. Detailed explanation of the notations are in the following subsection.

2.3 Robust estimating equations for joint mean and covariance model

In the previous subsections, we have constructed the joint mean–covariance model. However, we have introduced robust generalized estimating equation only for the mean. Now we propose the following robustified generalized estimating equations for $\theta = (\beta', \gamma', \lambda')'$

$$U(\theta) = (U_1(\beta)', U_2(\gamma)', U_3(\lambda'))'.$$

To be specific, the detailed estimating equations for the mean, generalized autoregressive parameters and innovation variances are

$$U_1(\beta) = \sum_{i=1}^m X_i' \Delta_i(\mu_i(\beta))(V_i^\beta)^{-1} h_i^\beta(\mu_i(\beta)) = 0, \tag{5}$$

$$U_2(\gamma) = \sum_{i=1}^m T_i'(V_i^\gamma)^{-1} h_i^\gamma(\hat{r}_i(\gamma)) = 0, \tag{6}$$

$$U_3(\lambda) = \sum_{i=1}^m Z_i' D_i(V_i^\lambda)^{-1} h_i^\lambda(\sigma_i^2(\lambda)) = 0, \tag{7}$$

where $h_i^\beta(\mu_i(\beta)) = W_i^\beta [\psi^\beta(\mu_i(\beta)) - C_i^\beta(\mu_i(\beta))]$, $h_i^\gamma(\hat{r}_i(\gamma)) = W_i^\gamma [\psi^\gamma(\hat{r}_i(\gamma)) - C_i^\gamma(\hat{r}_i(\gamma))]$ and $h_i^\lambda(\sigma_i^2(\lambda)) = W_i^\lambda [\psi^\lambda(\sigma_i^2(\lambda)) - C_i^\lambda(\sigma_i^2(\lambda))]$ act as the core of the estimating equations with ψ_i^β , ψ_i^γ , ψ_i^λ , C_i^β , C_i^γ , C_i^λ , W_i^β , W_i^γ and W_i^λ to be specified later; $X_i = (x_{i1}, \dots, x_{in_i})'$ and $Z_i = (z_{i1}, \dots, z_{in_i})'$, r_i and \hat{r}_i are $n_i \times 1$ vectors with j th components $r_{ij} = y_{ij} - \mu_{ij}$ and $\hat{r}_{ij} = E(r_{ij}|r_{i1}, \dots, r_{i(j-1)}) = \sum_{k=1}^{j-1} \phi_{ijk} r_{ik}$. We denote $\sum_{k=1}^0$ as zero when $j = 1$. In $U_3(\lambda)$, ϵ_i^2 and σ_i^2 are $n_i \times 1$ vectors

with j th components ϵ_{ij}^2 and σ_{ij}^2 , respectively, where $\epsilon_{ij} = y_{ij} - \hat{y}_{ij}$. We have $E(\epsilon_i^2) = \sigma_i^2$. $\Delta_i(\mu_i(\theta)) = \text{diag}\{\dot{\mu}_{i1}(\beta), \dots, \dot{\mu}_{in_i}(\beta)\}$, with $\dot{\mu}_{ij}(\cdot)$ denoting the first derivative of $\mu(\cdot)$ evaluated at $x'_{ij}\beta$; $T'_i = \partial\hat{r}'_i/\partial\gamma$ is the $q \times n_i$ matrix with j th column $\partial\hat{r}_{ij}/\partial\gamma = \sum_{k=1}^{j-1} r_{ik}\phi_{ijk}$; $D_i = \text{diag}\{\sigma_{i1}^2, \dots, \sigma_{in_i}^2\}$.

Furthermore, $V_i^\beta = A_i^{-1/2}\Sigma_i$, A_i is the diagonal elements of Σ_i ; $V_i^\gamma = D_i^{1/2}$; $V_i^\lambda = \tilde{A}_i^{-1/2}\tilde{\Sigma}_i$, \tilde{A}_i is the diagonal elements of $\tilde{\Sigma}_i$. Similar to [Ye and Pan \(2006\)](#), the sandwich working covariance structure $\tilde{\Sigma}_i = B_i^{1/2}R_i(\delta)B_i^{1/2}$ can be used to model the true $\tilde{\Sigma}_i = \text{Var}(\epsilon_i^2)$ with $B_i = 2\text{diag}\{\sigma_{i1}^4, \dots, \sigma_{in_i}^4\}$ and $R_i(\delta)$ mimics the correlation between ϵ_{ij}^2 and ϵ_{ik}^2 by introducing a new parameter δ . Typical structures for $R_i(\delta)$ include compound symmetry and AR(1). Although no particular suggestion on how to choose the structure and the value of δ was provided [Ye and Pan \(2006\)](#), the parameter δ has little effect on the estimation in practice, which is also confirmed in the simulation study reported in the later section. Moreover, we will show that the working correlation structure also has little effect on the estimates. In fact, r_i in $U_2(\gamma)$ and ϵ_i^2 in $U_3(\lambda)$ play a role similar to that of y_i in $U_1(\beta)$ and they can be viewed as working responses. Hence the ideas behind Eqs. (6) and (7) are in agreement with that in Eq. (5), which enhance the importance of estimation for the covariance matrix.

2.4 Huber’s score function ψ and weights w_{ij}

In the core of the estimating equations, $\psi^\beta(\mu_i) = \psi(A_i^{-1/2}(y_i - \mu_i))$, $\psi^\gamma(\hat{r}_i) = \psi(D_i^{-1/2}(r_i - \hat{r}_i))$ and $\psi^\lambda(\sigma_i^2) = \psi(\tilde{A}_i^{-1/2}(\epsilon_i^2 - \sigma_i^2))$. The function $\psi(\cdot)$ is chosen to limit the influence of outliers in the response variable, and a common choice is Huber’s score function $\psi_c(x) = \min\{c, \max\{-c, x\}\}$ for some constant c , normally chosen to be between 1 and 2.

Huber’s score function is the most widely used robustness technique as a bounded function, truncating large Pearson residuals symmetrically, which ensures the asymptotic normality of the estimator. The tuning constant c controls the robustness and the level of asymptotic efficiency. In practice, $c = 1.345$, $c = 1.5$ or $c = 2$ can be used depending on the seriousness of the contamination in a data set. We do simulations on different c and find the choice of c is not critical to gaining a good robust estimate. In this article, we use $c = 2$ in our implementation, which is sufficient to prove the improvement of efficiency by adopting robust estimating equations.

To ensure Fisher consistency, we use $C_i^\beta(\mu_i) = E[\psi(A_i^{-1/2}(y_i - \mu_i))]$, $C_i^\gamma(\hat{r}_i) = E[\psi(D_i^{-1/2}(r_i - \hat{r}_i))]$ and $C_i^\lambda(\sigma_i^2) = E[\psi(\tilde{A}_i^{-1/2}(\epsilon_i^2 - \sigma_i^2))]$. Given the assumption that y_i are under normal distribution, the three expectations depend only on the choice of constant c in Huber’s score function. In the following parts, we use $C_i^\beta = 0$, $C_i^\gamma = 0$ and $C_i^\lambda = -0.05$ that calculated under normality assumption.

In general cases, unless the true distribution is correctly specified, the expectation C_i in estimating equations that used to ensure the Fisher consistency are not available. [Qin and Zhu \(2007\)](#) discussed the difficulty of calculating C_i . They mentioned that C_i can be calculated easily for binary data as y_{ij} only take values 0 and 1 while hard to

obtain for the data following other distributions as the calculation of expectation C_i involves intractable integrals. As an alternative, some numerical integration methods or approximations are required to achieve the expectation C_i in this situation. Wang et al. (2005b) provided a bias correction method for robust estimation functions.

The weighting matrix $W_i^\beta = \text{diag}(w_{i1}^\beta, \dots, w_{ini}^\beta)$, $W_i^\gamma = \text{diag}(w_{i1}^\gamma, \dots, w_{ini}^\gamma)$ and $W_i^\lambda = \text{diag}(w_{i1}^\lambda, \dots, w_{ini}^\lambda)$ are diagonal weighting matrices assigning weights to each observation. Here, diagonal entries w_{ij} can assign different weights on each observation, instead of assigning unique weight on observations from a single subject.

Following Qin et al. (2009), we choose the weight function w_{ij} as a function of the Mahalanobis distance in the form

$$w_{ij} = w(p_{ij}) = \min \left\{ 1, \left[\frac{b_0}{(p_{ij} - m_p)^T S_p^{-1} (p_{ij} - m_p)} \right]^{\rho/2} \right\},$$

with $\rho \geq 1$, m_p and S_p are some robust estimates of the location and scale of p_{ij} such as the minimum covariance determinant estimators, where p_{ij} is corresponding to the design space. As a result, w_{ij} is a weight function that can downweigh any leverage point in the design space. In the following simulation study, b_0 is chosen as the 95th percentile of the chi-squared distribution with degrees of freedom equal to the dimension of p_{ij} and ρ is fixed as 1. Moreover, since $z_{ij} = x_{ij}$ (the design spaces for β and λ are the same), we choose $p_{ij} = x_{ij}$ for all three weighting matrices and denote them as $W_i = \text{diag}(w_{i1}, \dots, w_{ini})$ for simplicity.

The three proposed robust estimation equations within the framework of generalized estimating equations that do not require the normal distribution assumption are extensions for the method illustrated in Ye and Pan (2006) since our equations can resist the contamination and downweigh the potential influential points. By introducing the modified Cholesky decomposition, the positive definiteness of the covariance matrix can be guaranteed. A different variance correlation decomposition on certain type of matrix is implemented in Fan et al. (2007). Furthermore, the dimension of the parameter space of the covariance matrix has been substantially reduced that allows us to consider the regression model for the generalized autoregressive parameters and innovation variances simultaneously with the mean. Most importantly, we apply Mallows-type robust estimations for the mean and covariance jointly for the regression model which enjoy thorough robustness comparing to the single robust estimating equation established in He et al. (2005).

2.5 Estimators of parameters

Quasi-Fisher scoring algorithm is applied in solving β , γ and λ iteratively. First we choose a starting value for β , γ and λ , respectively. If we choose the special case of working independence $R_i = I$, which implies a convenient starting value of γ and λ to be $\gamma^{(0)} = 0$ and $\lambda^{(0)} = 0$, then (5) no longer depends on γ and λ . Hence, an initial estimate $\beta^{(0)}$ of β is set to be the solution to (5) in this special case as the robust GEE estimator under working independence covariance structure.

Given Σ_i , we solve (5) to find the estimate of β using the iterative procedure

$$\beta^{(k+1)} = \beta^{(k)} + \left\{ \left[\sum_{i=1}^m X_i' \Delta_i (V_i^\beta)^{-1} \Gamma_i^\beta \Delta_i X_i \right]^{-1} \sum_{i=1}^m X_i' \Delta_i (V_i^\beta)^{-1} h_i^\beta(\mu_i(\beta)) \right\} \Bigg|_{\beta=\beta^{(k)}}, \tag{8}$$

where $\Gamma_i^\beta = E h_i^\beta(\mu_i(\beta)) = E \partial h_i^\beta(\mu_i) / \partial \mu_i |_{\mu_i=\mu_i(\beta)}$, for $i = 1, \dots, m$. In practice, if we know the distribution of the data set, such as normal or t distribution, we can calculate the expectation analytically. Otherwise, we can use the sample mean or sample median to approximate the expectation when we do not know the true underlying distribution.

Given β and λ , γ can be updated approximately through

$$\gamma^{(k+1)} = \left\{ \left[E \sum_{i=1}^m T_i'(V_i^\gamma)^{-1} \Gamma_i^\gamma T_i \right]^{-1} \sum_{i=1}^m T_i'(V_i^\gamma)^{-1} h_i^\gamma(\hat{r}_i(\gamma)) \right\} \Bigg|_{\gamma=\gamma^{(k)}}, \tag{9}$$

with $\Gamma_i^\gamma = E \dot{h}_i^\gamma(\hat{r}_i(\gamma)) = E \partial h_i^\gamma(\hat{r}_i) / \partial \hat{r}_i |_{\hat{r}_i=\hat{r}_i(\gamma)}$, for $i = 1, \dots, m$.

Finally, given β and γ , the innovation variance parameters λ can be updated using

$$\lambda^{(k+1)} = \lambda^{(k)} + \left\{ \left[\sum_{i=1}^m Z_i' D_i (V_i^\lambda)^{-1} \Gamma_i^\lambda D_i Z_i \right]^{-1} \sum_{i=1}^m Z_i D_i (V_i^\lambda)^{-1} h_i^\lambda(\sigma_i^2(\lambda)) \right\} \Bigg|_{\lambda=\lambda^{(k)}}, \tag{10}$$

where $\Gamma_i^\lambda = E \dot{h}_i^\lambda(\sigma_i^2(\lambda)) = E \partial h_i^\lambda(\sigma_i^2) / \partial \sigma_i^2 |_{\sigma_i^2=\sigma_i^2(\lambda)}$, for $i = 1, \dots, m$.

In summary, these sets of parameters can be estimated using weighted least squares. The main algorithm processes iteratively as follows:

- Step 1: Select an initial value $(\beta^{(0)'}, \gamma^{(0)'}, \lambda^{(0)'})'$ and use model (3) to form $\Phi_i^{(0)}$ and $D_i^{(0)}$. Then $\Sigma_i^{(0)}$, the starting value of Σ_i , is obtained.
- Step 2: Using the weighted least squares estimators (8)–(10) to calculate the estimators $\beta^{(1)}$, $\gamma^{(1)}$ and $\lambda^{(1)}$ of β , γ and λ , respectively.
- Step 3: Replace $\beta^{(0)}$, $\gamma^{(0)}$ and $\lambda^{(0)}$ with the estimators $\beta^{(1)}$, $\gamma^{(1)}$ and $\lambda^{(1)}$.

Repeat Steps 2–3 until convergence of the parameter estimators.

In simulation, the proposed robust method works well under different contaminations. When the sample size is moderate, the difficulty of the convergence in the non-robust method of the algorithm lies in the non-convergence of $\hat{\lambda}_m$ in most of the cases, especially in those having serious contaminations.

The robust method is supposed to outperform the non-robust method substantially in serious contaminations. However, the non-robust method has difficulties in obtaining a reliable result under heavy contaminations. Therefore, we can only compare them under mild contaminations in simulation studies.

2.6 Asymptotic properties and hypotheses testing

Following [Ye and Pan \(2006\)](#), we can obtain the following theorems.

Theorem 1 *Suppose there is only one root $\hat{\theta}_m = (\hat{\beta}'_m, \hat{\gamma}'_m, \hat{\lambda}'_m)'$ for the generalized estimating equations. Under some mild regularity conditions stated in Appendix, the generalized estimating equation estimator $\hat{\theta}_m = (\hat{\beta}'_m, \hat{\gamma}'_m, \hat{\lambda}'_m)'$ is strongly consistent for the true value $\theta_0 = (\beta'_0, \gamma'_0, \lambda'_0)'$; that is, $\hat{\theta}_m = (\hat{\beta}'_m, \hat{\gamma}'_m, \hat{\lambda}'_m)' \rightarrow \theta_0 = (\beta'_0, \gamma'_0, \lambda'_0)'$ almost surely as $m \rightarrow \infty$.*

We denote $V_m = (v_m^{kl})_{k,l=1,2,3}$ as the covariance matrix of the function $U(\theta)/\sqrt{m} = (U'_1(\beta), U'_2(\gamma), U'_3(\lambda))/\sqrt{m}$, where $v_m^{kl} = m^{-1}cov(U_k, U_l)$ for $k \neq l$ and $v_m^{kk} = m^{-1}var(U_k)$ ($k, l = 1, 2, 3$). For the following theorem, the covariance matrix V_m evaluated at the true value θ_0 is assumed to be positive definite. Furthermore, at θ_0 we assume that

$$V_m = \begin{pmatrix} v_m^{11} & v_m^{12} & v_m^{13} \\ v_m^{21} & v_m^{22} & v_m^{23} \\ v_m^{31} & v_m^{32} & v_m^{33} \end{pmatrix} \rightarrow V = \begin{pmatrix} v^{11} & v^{12} & v^{13} \\ v^{21} & v^{22} & v^{23} \\ v^{31} & v^{32} & v^{33} \end{pmatrix} \text{ as } m \rightarrow \infty.$$

Theorem 2 *Under some necessary regularity conditions stated in Appendix, the generalized estimating equation estimator $\hat{\theta}_m = (\hat{\beta}'_m, \hat{\gamma}'_m, \hat{\lambda}'_m)'$ is asymptotically normally distributed with*

$$\sqrt{m} \begin{pmatrix} \hat{\beta}_m - \beta_0 \\ \hat{\gamma}_m - \gamma_0 \\ \hat{\lambda}_m - \lambda_0 \end{pmatrix} \rightarrow N \left\{ 0, \begin{pmatrix} v^{11} & 0 & 0 \\ 0 & v^{22} & 0 \\ 0 & 0 & v^{33} \end{pmatrix}^{-1} \begin{pmatrix} v^{11} & v^{12} & v^{13} \\ v^{21} & v^{22} & v^{23} \\ v^{31} & v^{32} & v^{33} \end{pmatrix} \begin{pmatrix} v^{11} & 0 & 0 \\ 0 & v^{22} & 0 \\ 0 & 0 & v^{33} \end{pmatrix}^{-1} \right\}$$

in distribution as $m \rightarrow \infty$, where the matrices v^{kl} ($k, l = 1, 2, 3$) are evaluated at the true value $\theta = \theta_0$.

The proofs are given in Appendix.

Note that when the responses y_i are normally distributed, we have $v^{kl} = 0$ ($k \neq l$) and the asymptotic covariance matrix in Theorem 2 reduces to $\{\text{diag}(v^{11}, v^{22}, v^{33})\}^{-1}$.

For inference, we use a robust estimator for the covariance matrix of $\hat{\beta}$:

$$Cov(\hat{\beta}) = (\hat{H}_m)^{-1} \hat{K}_m (\hat{H}_m)^{-1},$$

where \hat{H}_m and \hat{K}_m are defined by

$$\hat{H}_m = \sum_{i=1}^m X'_i \Delta_i (V_i^\beta(\hat{\beta}_m))^{-1} \Gamma_i^\beta \Delta_i X_i,$$

$$\hat{K}_m = \sum_{i=1}^m X'_i \Delta_i (V_i^\beta(\hat{\beta}_m))^{-1} h_i^\beta(\mu_i(\hat{\beta}_m)) h_i^\beta(\mu_i(\hat{\beta}_m))' (V_i^\beta(\hat{\beta}_m))^{-1} \Delta_i X_i.$$

In the sandwich type estimator of covariance matrix, we adopt similar Mallows-type weights and Huber’s function to control the influence of outliers. The covariance

matrices of $\hat{\gamma}$ and $\hat{\lambda}$ can be estimated in an analogous way. In the same manner of He et al. (2005), we compare the average estimated standard errors and the Monte Carlo standard errors in simulations. Overall, we note the standard error estimation works well for different AR(1) correlation structure no matter the contamination exists or not. Similar findings have also been obtained in Leng et al. (2010). As a result, we consider the asymptotic covariance formula quite acceptable as the large-sample estimation.

For hypothesis testing, within the framework of generalized estimating equations, the quasi-score test based on the derivative of the generalized estimating equations may be constructed. See Ye and Pan (2006) for details.

3 Simulation study

In this section, simulations including contaminated cases are conducted to assess the performance of the proposed robust method. Four estimation methods are considered: NR refers to the non-robust method, which is given in Ye and Pan (2006). HR_m means the half-robust method on the mean. In other words, we only adopt the robust estimating equation (5), which is the estimating equation for the mean. In contrast, we have HR_c that stands for the other half-robust method on the covariance matrix only. The R (robust) method is our proposed method which includes all three robust estimating equations. Note that the non-robust estimators of β , γ and λ are defined through the same equations except that $\psi(x) = x$ and $W_i = I_i$, where I_i are $n_i \times n_i$ identity matrices.

Study 1. The following Gaussian linear model is used:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + e_{ij}, \quad i = 1, \dots, m; \quad j = 1, \dots, n_i,$$

where $m = 100$, $x_{ij} \sim N(0, 2)$, $\beta_1 = 1$, $\beta_0 = 0.5$ and $e_{ij} \sim N(0, \Sigma_i)$.

The error term $(e_{i1}, \dots, e_{in_i})$ is generated from a multivariate normal distribution with mean 0 and covariance Σ_i satisfying $T_i \Sigma_i T_i' = D_i$, where T_i and D_i are described in Sect. 2.1 with $z_{ijk} = (1, (t_{ij} - t_{ik}))'$ and $z_{ij} = x_{ij}$. Two specifications are considered: Case (1) $\gamma = (0.2, 0.3)'$, $\lambda = (-0.5, 0.2)'$ and Case (2) $\gamma = (0.2, 0)'$, $\lambda = (-0.5, 0.2)'$. The difference between these two cases lies in the choice of γ_2 .

Similar to the sampling scheme in Fan et al. (2007), the observation times are regularly scheduled but may be missing in practice. Missing at random is considered. More precisely, each subject has a set of scheduled time point $\{0, 1, \dots, 12\}$, in which each element (except time 0) has a 20% probability of being missing. A uniform $[0, 1]$ random variable is added to a non-missing scheduled time. This results in irregular (not on a grid) observed time points t_{ij} per individual and then t_{ij} is transformed onto interval $[0, 1]$.

To study robustness, we denote NC as no contamination situation and consider the following three contaminations:

C1: randomly choose 2% of x_{ij} to be $x_{ij} - 3$;

C2: randomly choose 2% of y_{ij} to be $y_{ij} + 6$;

C3: randomly choose 2% of x_{ij} to be $x_{ij} - 3$ and 2% of y_{ij} to be $y_{ij} + 6$;

Table 1 Simulation results of bias and MSE for β , γ and λ in Study 1 (Case 1)

	NC		C1		C2		C3	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\beta_0 = 0.5$								
NR	0.001	0.0004	0.059	0.0044	0.117	0.0159	0.174	0.0332
HR _m	0.000	0.0004	0.038	0.0022	0.045	0.0035	0.096	0.0115
HR _c	0.000	0.0004	0.056	0.0041	0.117	0.0159	0.170	0.0318
R	0.000	0.0004	0.029	0.0017	0.029	0.0020	0.067	0.0066
$\beta_1 = 1$								
NR	0.000	0.0001	−0.041	0.0019	0.001	0.0003	−0.042	0.0021
HR _m	0.000	0.0001	−0.030	0.0011	0.001	0.0002	−0.036	0.0016
HR _c	0.000	0.0001	−0.045	0.0022	0.001	0.0004	−0.043	0.0023
R	0.000	0.0001	−0.032	0.0012	0.001	0.0002	−0.035	0.0015
$\gamma_1 = 0.2$								
NR	0.001	0.0003	0.026	0.0014	0.057	0.0049	0.059	0.0050
HR _m	0.001	0.0003	0.026	0.0014	0.058	0.0051	0.061	0.0052
HR _c	0.001	0.0003	0.025	0.0013	0.053	0.0046	0.057	0.0048
R	0.001	0.0003	0.025	0.0013	0.054	0.0047	0.059	0.0051
$\gamma_2 = 0.3$								
NR	−0.033	0.0033	−0.102	0.0177	−0.270	0.0881	−0.295	0.1010
HR _m	−0.036	0.0033	−0.104	0.0180	−0.278	0.0926	−0.303	0.1056
HR _c	−0.037	0.0033	−0.097	0.0163	−0.250	0.0796	−0.284	0.0954
R	−0.041	0.0032	−0.100	0.0169	−0.257	0.0840	−0.298	0.1038
$\lambda_1 = −0.5$								
NR	0.000	0.0016	0.388	0.1532	0.958	0.9198	1.101	1.213
HR _m	0.000	0.0016	0.390	0.1541	0.961	0.9256	1.103	1.219
HR _c	−0.002	0.0025	0.186	0.0371	0.364	0.1369	0.546	0.302
R	−0.002	0.0025	0.181	0.0352	0.363	0.1360	0.548	0.305
$\lambda_2 = 0.2$								
NR	0.003	0.0000	−0.184	0.0346	−0.116	0.0166	−0.193	0.0389
HR _m	0.002	0.0004	−0.186	0.0352	−0.169	0.0167	−0.193	0.0391
HR _c	0.004	0.0005	−0.084	0.0078	−0.039	0.0024	−0.126	0.0172
R	0.004	0.0005	−0.081	0.0072	−0.036	0.0023	−0.125	0.0169

NR refers to the non-robust method; HR_m means the half-robust method on the mean; HR_c stands for the other half-robust method on the covariance matrix; R is the proposed robust method which adopts all three robust estimating equations

We consider 200 replications for the simulation. Table 1 shows the performance of the NR, HR_m, HR_c and R estimators for both the mean and the covariance in Case 1. It is found that the non-robust, half-robust and robust estimation perform equally well in the case of uncontaminated data (NC), although there is some loss of efficiency in the robust method with slightly larger MSE of λ_1 . However, for the contaminated data, the robust method generally achieves smaller biases resulting in smaller MSEs.

First we look into performance of estimating in β . For β_0 , we notice that in C1, C2 and C3, the half-robust method on mean outperforms the non-robust method and the half-robust method on covariance. Meanwhile, the robust method performs better than the half-robust method and therefore becomes the best performer for β_0 among the three. This does not happen in the estimation of β_1 . For β_1 , the robust method R and the half-robust method HR_m have similar performance. Both of them have much smaller MSEs than that of the non-robust method and the other half-robust method HR_c . However, no great difference can be detected between the estimation of the former two, i.e. R and HR_m .

Next we pay attention to the estimation for parameters in the covariance matrix. All four methods show little difference in estimating γ , because the corresponding covariates for γ only contain t , which has no contamination at all. It supports in one way that the proposed robust method performs equally well when there is no contamination in γ . On the other hand, the robust method has no advantage under no contamination. As for λ , while the half-robust method for mean performs as poor as the non-robust method (in both biases and MSEs), the robust estimators take great advantage uniquely (even slightly better than the half-robust method for covariance). For both λ_1 and λ_2 , the robust estimators have about half of the biases and one quarter of the MSEs as those of the non-robust and mean half-robust methods. Overall, the robust method performs favorably in comparison with the non-robust and the half-robust methods in estimation of the covariance matrix. This point supports a better estimate for the covariance matrix in the robust method resulting a better estimate for the mean parameter β_0 . In summary, the proposed robust method generally outperforms both the half-robust method and the non-robust method under different contaminations in the simulation study.

In previous discussion, we did not provide a method to choose the structure $R_i(\delta)$. Instead, we considered the typical AR(1) structure with δ equal to 0, 0.2, 0.5 and 0.8 for a test of sensitivity on the choice of δ . Table 2 summarizes the MSEs for β , γ and λ due to different values of δ in case 1 without contamination and under contamination C3. From the table, we observe very similar performance when we select $\delta = 0, 0.2$ or 0.5 . When $\delta = 0.8$, MSEs for the mean parameter β still stay close to those for other δ s. However, the increase of MSEs for λ cannot be ignored in the case without contamination. We suppose this choice of δ is a bit apart from the truth that leads to an inefficient estimation comparing with other choices of value for δ . In sum, we may conclude that the choice of δ or the structure has no significant influence in our simulation in the mean model. Similar conclusions can be obtained from the performances of estimation under C1 and C2 and thus we omit the results. In the rest of the article, we select an acceptable δ equal to 0 for convenience. We have considered simulations for Case 2 as well. The results are similar to those of Case 1 and so they are omitted for brevity.

In addition to the bias and MSE criteria that we considered, two loss functions are introduced to see how the four MSE methods work in estimating the covariance matrix. They are the entropy loss

$$L_1(\Sigma, \hat{\Sigma}) = m^{-1} \sum_{i=1}^m \{\text{trace}(\Sigma_i \hat{\Sigma}_i^{-1}) - \log |\Sigma_i \hat{\Sigma}_i^{-1}| - n_i\},$$

Table 2 Mean squared errors for estimates using different δ 's in Study 1 ($\times 100$)

δ		0	0.2	0.5	0.8
NC					
β_0	NR	0.034	0.034	0.034	0.034
	R	0.036	0.037	0.037	0.037
β_1	NR	0.009	0.009	0.009	0.009
	R	0.012	0.012	0.012	0.012
γ_1	NR	0.036	0.036	0.036	0.036
	R	0.037	0.037	0.037	0.037
γ_2	NR	0.333	0.333	0.334	0.335
	R	0.338	0.338	0.339	0.340
λ_1	NR	0.169	0.168	0.183	0.306
	R	0.270	0.270	0.293	0.456
λ_2	NR	0.044	0.050	0.064	0.076
	R	0.052	0.057	0.075	0.089
C3					
β_0	NR	3.14	3.14	3.14	3.16
	R	0.57	0.57	0.57	0.58
β_1	NR	0.21	0.21	0.21	0.21
	R	0.15	0.15	0.15	0.15
γ_1	NR	0.46	0.46	0.47	0.47
	R	0.46	0.46	0.46	0.46
γ_2	NR	9.71	9.71	9.79	9.87
	R	9.60	9.61	9.63	9.70
λ_1	NR	122	122	124	128
	R	29.7	29.9	30.9	33.1
λ_2	NR	3.68	3.64	3.62	3.61
	R	1.66	1.61	1.58	1.59

and the quadratic loss

$$L_2(\Sigma, \hat{\Sigma}) = m^{-1} \sum_{i=1}^m \text{trace}(\Sigma_i^{-1} \hat{\Sigma}_i - I_i)^2,$$

where Σ_i is the true covariance matrix and $\hat{\Sigma}_i$ is its estimator. Each of these losses is 0 when $\hat{\Sigma}_i = \Sigma_i$ and positive otherwise. The entropy loss is the same as the Kullback–Leibler loss after switching the roles of the covariance matrix and its inverse. As indicated by [Levina et al. \(2008\)](#), the entropy loss is a more appropriate measure if the covariance matrix itself is the primary object of interest.

Now we focus on the evaluation on the overall performance of the covariance matrix estimation. [Table 3](#) demonstrates that the robust method reduces both the entropy loss and the quadratic loss substantially especially when the contamination is heavier,

Table 3 Entropy loss and quadratic loss in estimating Σ in Study 1

	NC	C1	C2	C3
Entropy loss				
NR	0.05	0.81	2.51	3.40
HR _m	0.05	1.09	2.54	3.76
HR _c	0.06	0.27	0.77	1.38
R	0.06	0.29	0.76	1.41
Quadratic loss				
NR	0.26	5.32	25.1	29.9
HR _m	0.26	6.16	25.1	31.1
HR _c	0.42	1.52	5.52	10.2
R	0.42	1.55	5.50	10.2

although it has larger losses in the case of no contamination. Here we see clearly that the mean half-robust method has little improvement in resisting contamination in estimating the covariance matrix, since the losses are nearly the same as those of the non-robust method. In contrast, the half-robust method on covariance works as good as the robust method in covariance matrix estimation.

It is to be noted that the results in Table 3 are obtained from 200 replications with successful convergence in estimation. The robust method successfully converged in all situations. However, the non-robust and the half-robust methods did not converge in a few percents of the simulations. Thus, the robust method is recommended since the non-convergence problem should not be neglected. Moreover, it is also found that the larger the contamination, the poorer the convergence performance by the non-robust and half-robust methods. And this is one of the reasons why we only compare the performance of the four methods under relatively mild contaminations.

Study 2. This study is designed to compare the performance of the proposed robust method and the non-robust method when the data sets are from non-normal distributions. The half-robust methods are not chosen because they are outperformed by the robust method. The setting is similar to those in study 1 except that we consider the error terms $(e_{i1}, \dots, e_{in_i})$ which are drawn from (a) a multivariate t -distribution with 3 degrees of freedom and covariance matrix Σ_i and (b) a mixed multivariate normal distribution with 30 % coming from a normal distribution $N(-0.7 \times \mu_{mn}, \Sigma_i)$ and the other 70 % from $N(0.3 \times \mu_{mn}, \Sigma_i)$, where μ_{mn} will be specified later. Note that the error terms drawn from (b) are asymmetric in distribution. Only the case of no contamination and the case of 3rd contamination C3 are considered.

We report MSE of the mean parameter β_0 and β_1 in Table 4, together with the entropy loss and quadratic loss in Table 5. Results for error terms from normal distribution, t distribution and mixed normal distributions with $\mu_{mn} = 0.5, 1$ and 2 are listed (They are termed MM0.5, MM1 and MM2, respectively). It is predictable that MSE of β increases when the error terms are more asymmetric, for the reason that the bias corrected term C_i^β , C_i^γ and C_i^λ in the estimating equation can be misspecified. Meanwhile, the robust method always has smaller MSE for β and losses of the covariance matrix than the non-robust method under contamination.

Table 4 Mean squares errors for β_0 and β_1 in Study 2 ($\times 100$)

Distribution		Normal	t(3)	MN0.5	MN1	MN2
NC (no contamination)						
β_0	NR	0.035	0.036	0.087	0.277	1.008
	R	0.036	0.040	0.094	0.300	1.181
β_1	NR	0.009	0.010	0.010	0.013	0.015
	R	0.012	0.012	0.013	0.014	0.016
C3 (contamination 3)						
β_0	NR	3.33	3.33	3.43	3.68	4.50
	R	0.66	0.66	0.76	1.09	2.21
β_1	NR	0.20	0.20	0.20	0.20	0.20
	R	0.15	0.15	0.13	0.14	0.15

Table 5 Entropy loss and quadratic loss in estimating Σ in Study 2

Distribution		Normal	t(3)	MN0.5	MN1	MN2
NC (no contamination)						
Entropy loss	NR	0.035	0.039	0.265	1.351	3.476
	R	0.044	0.049	0.251	1.316	3.413
Quadratic loss	NR	0.216	0.216	1.002	3.452	7.976
	R	0.359	0.365	0.660	2.374	5.730
C3 (contamination 3)						
Entropy loss	NR	5.229	5.220	5.490	6.053	7.320
	R	2.178	2.164	2.523	3.357	5.067
Quadratic loss	NR	43.43	42.44	42.51	41.57	40.01
	R	13.70	13.61	14.42	15.31	17.51

Furthermore, we find it interesting that, under cases of asymmetric distribution errors, the robust covariance matrix estimator possesses even smaller entropy (and quadratic) losses than the non-robust estimator under no contamination (Table 5). It supports the view that the robust method cultivates a better estimation for the covariance matrix, which can be seriously affected by outliers, non-normal errors or misspecifications of the underlying distributions. In all, study 2 demonstrates that the proposed robust method is able to accommodate the effect of outliers and improve the efficiency of parameter estimation under non-normal or asymmetric distributions.

4 Real data analysis

We apply the proposed method to analyze the longitudinal data of a hormone study on progesterone (Zhang et al. 1998). This data set involves a total of 492 observations among the 34 subjects. The log-transformed progesterone level is taken to be the response (y_{ij}). Other than time (t), two covariates age (AGE) and body mass

Table 6 Regression coefficient estimates and standard deviations (in parentheses) for AGE and BMI of the hormone data

	GEE	NR	R
Intercept	0.95(0.11)	0.87(0.13)	0.80(0.14)
AGE	2.05(1.96)	2.48(2.16)	3.95(2.05)
BMI	-1.72(2.26)	-1.71(2.92)	-0.42(3.49)

index (BMI) are considered. Following Zhang et al. (1998) and other researchers who have imposed a non-linear relationship between y_{ij} and t_{ij} , we consider the following model:

$$\begin{aligned}
 y_{ij} &= \beta_0 + \beta_1 \text{AGE}_i + \beta_2 \text{BMI}_i + \beta_3 t_{ij} + \beta_4 t_{ij}^2 + \beta_5 t_{ij}^3 + e_{ij}, \\
 \phi_{ijk} &= (1, (t_{ij} - t_{ik}), (t_{ij} - t_{ik})^2, (t_{ij} - t_{ik})^3)' \gamma, \\
 \log \sigma_{ij}^2 &= (1, \text{AGE}_i, \text{BMI}_i, t_{ij}, t_{ij}^2, t_{ij}^3)' \lambda.
 \end{aligned}$$

Table 6 lists the intercepts and the regression coefficients for AGE and BMI obtained from both the robust and the non-robust methods as well as those of the GEE method for comparison. The effects of AGE and BMI are found to be insignificant for both the robust and non-robust methods due to the large standard errors. In Table 6, the response looks less negatively affected by the body mass index in the robust model than that in the non-robust model, while more positively affected by the age in the robust estimation. The obvious numerical differences of the estimates between the robust and non-robust methods implies that the data may be contaminated.

The weight functions w_{ij} in our robust method are calculated from $p_{ij} = (\text{AGE}_i, \text{BMI}_i)$. The heavily downweighted points are from subject 18 (a cluster of points from case 244 to case 263), with $w_{ij} = 0.459$. A closer inspection of the data set shows that subject 18 has an extremely high BMI of 38. To further look into robustness, we consider the standardized residual s_{ij} which is the j th component of $\hat{\Sigma}_i^{-1/2}(y_i - \hat{\mu}_i)$. Case 10 appears to be the most extreme point with $s_{ij} = -4.58$. The progesterone level of the 10th observation for subject 1 (case 10) is 2.46, which is very different from its neighborhood observations 9 and 11 measured one day before and one day after, with the progesterone level being 12.8 and 13.4 respectively. In fact, other 13 observations on the subject 1 range from 8.5 to 13.4 except this case. In particular, this observation is the lowest progesterone level in the whole data set. Therefore, we conclude that case 10 is a clear outlier from subject 1, which is consistent with Fung et al. (2002). When the sample size is moderate, a subject-level potential outlier can have significant influence on estimation and inference. Subject 24 is a potential outlier as the mean of its standardized residuals is 2.66, with s_{ij} of case 337–346 ranging from 2.09 to 3.80. Subject 24 turns out to be a young women with very low BMI and the highest average progesterone level. Consequently, we find the robust method downweights substantially the effect of both the subject-level and the observation-level outliers. This is the main reason that the robust method leads to a shift of estimation for the coefficients in the mean model compared with its non-robust

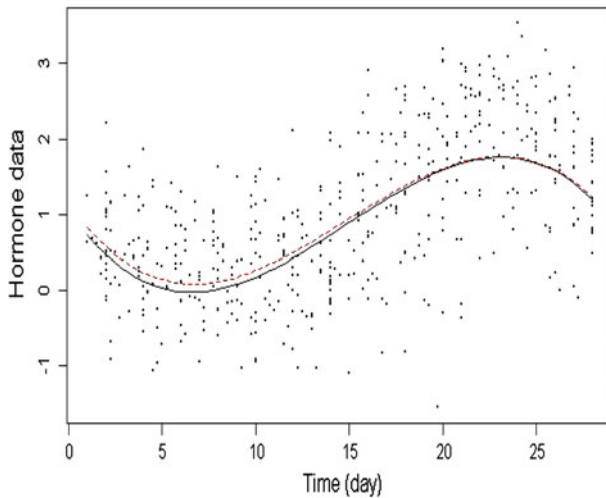


Fig. 1 Analysis of the hormone data. The fitted curve of the cubic polynomial of time, against time. The *solid line* is the robust estimate and the *dotted line* is the non-robust estimate

version. We believe that the robust estimation is more reliable. Nevertheless, the large standard errors of the estimates suggest that a much larger sample is needed to have any concrete finding.

Figure 1 displays the fitted curves for the cubic polynomial of time. From Fig. 1, we can see that the cubic polynomial of time decreases in the first 7 days and increases steadily later on. They reach a peak around the 23rd day in the cycle and then decrease again. The trajectory of the mean curve is very similar to that in Mao et al. (2011). Furthermore, although there are outliers in the data set, their effects are not too large on the estimation of the mean curve as we can see from Fig. 1 that the robust and non-robust estimates are rather close to each other.

Figure 2a plots the estimated generalized autoregressive parameters ϕ against the time lag between measurements in the same subject, which is also modeled as a cubic polynomial. The graph shows that the generalized autoregressive parameter decreases sharply from 0.6 to 0 if the time lag is less than 8 days and then drops slowly when the lag becomes larger. It is noted that the robust and non-robust estimates of the parameter, which is essentially a mean parameter as seen from (2), are also quite close to each other. From Fig. 2b, it is observed that the non-robust estimate fluctuates more intensely than the robust estimate for the innovation variance. The figure provides us some idea of how the robust method works. The shrinkage of estimation for the innovation variance (from above 0.6 to less than 0.5) suggests that our method can downweigh the effect of outliers to achieve robustness. Unlike the estimation on the mean parameter (Fig. 1), the non-robust estimate of the variance parameter (innovation variance) is strongly affected by outliers, and is rather different from the robust estimate (Fig. 2b). Although little difference can be detected in the mean model from Fig. 1, the robust method improves estimation for the covariance matrix, especially for the innovation variance.

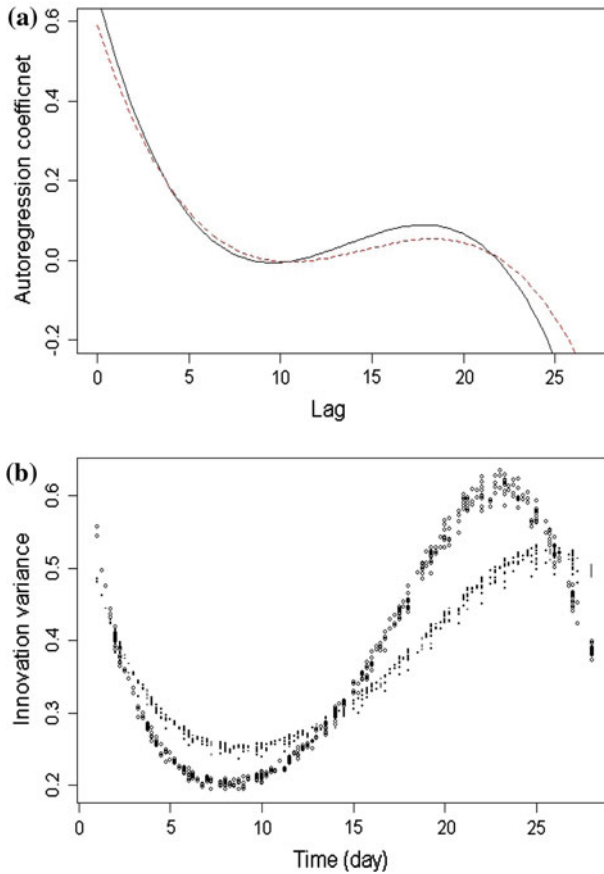


Fig. 2 Analysis of the hormone data. **a** Estimate for autoregressive coefficient against time lag. **b** Estimate for innovation variance against time. In **a**, the *solid line* is the robust estimate and the *dotted line* is the non-robust estimate. In **b**, the *solid circle* is the robust estimate and the *hollow circle* is the non-robust estimate

5 Discussion

In this paper, we propose simultaneous robust model for the mean and covariance matrix of longitudinal data. The proposed method has the following advantages and properties: (i) the robust covariance model guarantees the positive definiteness based on the covariance decomposition with a proper statistical interpretation (ii) it is able to control the influence of outliers in the mean and covariance model simultaneously that cultivates a more reliable estimation for the joint mean and covariance model (iii) the robust algorithm has a much greater chance to obtain a convergence solution than the non-robust algorithm. The robust estimating equations we proposed here should enhance the development of joint mean and covariance model for longitudinal data.

A limitation of the proposed method is that the model may include redundant covariates. If we have no prior knowledge of the covariance structure, then we are

prone to include all the time and mean associated variables. Redundant covariates may bring in outliers and also increase the computational burden. Robust estimating equations that can serve the goal of both estimating and penalizing the models with too many covariates are under development.

Appendix

Proofs

Regularity conditions:

- A1. We assume that the dimensions p , q and d of the covariates x_{ij} , z_{ij} and z_{ijk} are fixed and that $\{n_i\}$ is a bounded sequence of positive integers. The first four moments of y_{ij} exist.
- A2. The parameter space of $(\beta', \gamma', \lambda')'$, Θ , is a compact subset of R^{p+q+d} , and the true parameter value $(\beta'_0, \gamma'_0, \lambda'_0)'$ is in the interior of the parameter space Θ .
- A3. The covariates z_{ijk} and z_{ij} , the matrices W_i^{-1} are all bounded, meaning that all the elements of the vectors are bounded. The function $\dot{g}^{-1}(\cdot)$ has bounded second derivatives.

Proof of Theorem 1 For illustration we only give the proof that $\hat{\beta}_m \rightarrow \beta_0$ almost surely. The proofs for $\hat{\gamma}_m$ and $\hat{\lambda}_m$ are similar. According to McCullagh (1983), we have

$$\begin{aligned} \hat{\beta}_m - \beta_0 &= \left\{ \frac{1}{m} \sum_{i=1}^m X_i' \Delta_i (V_i^\beta)^{-1} \Gamma_i^\beta \Delta_i' X_i \right\}_{\beta=\beta_0}^{-1} \\ &\quad \times \left\{ \sum_{i=1}^m X_i' \Delta_i (V_i^\beta)^{-1} h_i^\beta(\mu_i(\beta)) \right\}_{\beta=\beta_0} + o_p(m^{-1/2}). \end{aligned}$$

On the other hand, the expectation and variance matrix of $U_{1i} = X_i' \Delta_i (V_i^\beta)^{-1} h_i^\beta(\mu_i(\beta))$ at $\beta = \beta_0$ are given by $E_0(U_{1i}) = 0$ and

$$\text{var}_0(U_{1i}) = \left\{ X_i' \Delta_i (V_i^\beta)^{-1} \Gamma_i^\beta \Delta_i' X_i \right\}_{\beta_0} = (G_i^0 X_i' X_i)' (V_i^{\beta_0})^{-1} \Gamma_i^{\beta_0} (G_i^0 X_i' X_i),$$

where $G_i^0 = \text{diag}\{\dot{g}^{-1}(x'_{i1}\beta_0), \dots, \dot{g}^{-1}(x'_{in_i}\beta_0)\}$ is an $n_i \times n_i$ diagonal matrix.

Since $V_i^\beta = A_i^{-1/2} \Sigma_i$ and $\Sigma_i^{-1} = \Phi_i' D_i^{-1} \Phi_i$, the variance can be further written as $\text{var}_0(U_{1i}) = (G_i^0 X_i' X_i)' \Phi_i (D_i^{-1} A_i^{-1/2}) \Phi_i' \Gamma_i^{\beta_0} (G_i^0 X_i' X_i)$. Condition A3 above implies that there exists a constant κ_0 such that $\text{var}_0(U_{1i}) \leq \kappa_0 1_{p \times p}$ for all i and all $\theta \in \Theta$, where $1_{p \times p}$ is the $p \times p$ matrix with all elements being 1's, meaning that all elements of $\text{var}_0(U_{1i})$ are bounded by κ_0 . Thus $\sum_{i=1}^\infty \text{var}_0(U_{1i})/i^2 < \infty$. By Kolmogorov's strong law of large numbers we know that

$$\left\{ \frac{1}{m} \sum_{i=1}^m X_i' \Delta_i (V_i^\beta)^{-1} h_i^\beta(\mu_i(\beta)) \right\}_{\beta=\beta_0} \rightarrow 0$$

almost surely as $m \rightarrow \infty$. In the same manner it can be shown that

$$\left\{ \frac{1}{m} \sum_{i=1}^m X_i' \Delta_i (V_i^\beta)^{-1} \Gamma_i^\beta \Delta_i' X_i \right\}_{\beta=\beta_0}$$

is a bounded matrix. This leads to $\hat{\beta}_m - \beta_0 \rightarrow 0$ almost surely as $m \rightarrow \infty$. The proof is complete. □

Proof of Theorem 2 First we give some notations. Define

$$H_m = \sum_{i=1}^m X_i' \Delta_i (V_i^\beta)^{-1} \Gamma_i^\beta \Delta_i X_i,$$

$$B_m = \sum_{i=1}^m T_i' (V_i^\gamma)^{-1} \Gamma_i^\gamma T_i,$$

$$C_m = \sum_{i=1}^m Z_i' D_i (V_i^\lambda)^{-1} \Gamma_i^\lambda D_i Z_i.$$

$$\tilde{U}_1(\beta) = \sum_{i=1}^m X_i' \Delta_{0i} (V_{0i}^\beta)^{-1} h_{0i}^\beta(\mu_{0i}(\beta)), \tag{11}$$

$$\tilde{U}_2(\gamma) = \sum_{i=1}^m T_i' (V_{0i}^\gamma)^{-1} h_{0i}^\gamma(\hat{r}_{0i}(\gamma)), \tag{12}$$

$$\tilde{U}_3(\lambda) = \sum_{i=1}^m Z_i' D_{0i} (V_{0i}^\lambda)^{-1} h_{0i}^\lambda(\sigma_{0i}^2(\lambda)). \tag{13}$$

$$\xi = H_m^{1/2}(\beta - \beta_0), \hat{\xi} = \xi(\hat{\beta}_m) = H_m^{1/2}(\hat{\beta}_m - \beta_0), \tilde{\xi} = H_m^{1/2} \tilde{U}_1;$$

$$\eta = B_m^{1/2}(\gamma - \gamma_0), \hat{\eta} = \eta(\hat{\gamma}_m) = B_m^{1/2}(\hat{\gamma}_m - \gamma_0), \tilde{\eta} = B_m^{1/2} \tilde{U}_2;$$

$$\zeta = C_m^{1/2}(\lambda - \lambda_0), \hat{\zeta} = \zeta(\hat{\lambda}_m) = C_m^{1/2}(\hat{\lambda}_m - \lambda_0), \tilde{\zeta} = C_m^{1/2} \tilde{U}_3.$$

Next we prove the following Lemma:

Lemma Under condition (A1)–(A3),

$$\|\hat{\xi} - \tilde{\xi}\| = o_p(1), \tag{14}$$

$$\|\hat{\eta} - \tilde{\eta}\| = o_p(1), \tag{15}$$

$$\|\hat{\zeta} - \tilde{\zeta}\| = o_p(1). \tag{16}$$

Define

$$\Psi(\xi) = H_m^{1/2} U_1(\beta) = H_m^{1/2} U_1(\xi), \tag{17}$$

$$\Phi(\xi) = H_m^{1/2} \tilde{U}_1(\xi) - \xi. \tag{18}$$

By condition (A1)–(A3), $\Psi(\xi)$ and U_1 give the same root for ξ . The solution of Φ is $\tilde{\xi}$. Following the proof of Theorem 1 in He et al. (2005), we immediately obtain that

$$\sup_{\|\xi\| \leq L} \|\Psi(\xi) - \Phi(\xi)\| = o_p(1), \quad \|\tilde{\xi}\| = O_p(1),$$

where L is a sufficiently large number. By Brouwer’s fixed-point theorem, (11) is verified. We can prove (12) and (13) similarly. \square

By Lemma, we only need to show the asymptotic normality of $(\tilde{\xi}', \tilde{\eta}', \tilde{\zeta}')'/\sqrt{m}$. This is equivalent to the asymptotic normality of $(\tilde{U}'_1, \tilde{U}'_2, \tilde{U}'_3)/\sqrt{m}$. Note that Conditions (A1)–(A3) imply that

$$E_0[\zeta' \{X'_i \Delta_{0i} (V_i^\beta)^{-1} h_i^\beta\} + \omega' \{T'_i (V_i^\gamma)^{-1} h_i^\gamma\} + \phi' \{Z'_i D_{0i} (V_i^\rho)^{-1} h_i^\lambda\}]^3 < \kappa,$$

for any $\zeta \in R^{p+K}$, $\omega \in R^q$ and $\phi \in R^{d+K'}$, where κ is a constant independent of i . Furthermore, we have

$$\begin{aligned} & \frac{1}{m} \sum_{i=1}^m V[\zeta' \{X'_i \Delta_{0i} (V_i^\theta)^{-1} h_i^\theta\} + \omega' \{T'_i (V_i^\gamma)^{-1} h_i^\gamma\} + \phi' \{Z'_i D_{0i} (V_i^\rho)^{-1} h_i^\rho\}] \\ &= (\zeta', \omega', \phi') \frac{1}{m} V_m(\zeta', \omega', \phi')' \rightarrow (\zeta', \omega', \phi')' V(\zeta', \omega', \phi')' > 0. \end{aligned}$$

Therefore, the asymptotic normality of $(\tilde{U}'_1, \tilde{U}'_2, \tilde{U}'_3)/\sqrt{m}$ is easily proved by multivariate Liapounov central limit theorem. Therefore,

$$\sqrt{m} \begin{pmatrix} \hat{\beta}_m - \beta_0 \\ \hat{\gamma}_m - \gamma_0 \\ \hat{\lambda}_m - \lambda_0 \end{pmatrix} = \begin{pmatrix} (H_m/m)^{-1} & 0 & 0 \\ 0 & (B_m/m)^{-1} & 0 \\ 0 & 0 & (C_m/m)^{-1} \end{pmatrix} \begin{pmatrix} \tilde{U}_1/\sqrt{m} \\ \tilde{U}_2/\sqrt{m} \\ \tilde{U}_3/\sqrt{m} \end{pmatrix} \tag{19}$$

$$\rightarrow N \left\{ 0, \begin{pmatrix} v^{11} & 0 & 0 \\ 0 & v^{22} & 0 \\ 0 & 0 & v^{33} \end{pmatrix}^{-1} \begin{pmatrix} v^{11} & v^{12} & v^{13} \\ v^{21} & v^{22} & v^{23} \\ v^{31} & v^{32} & v^{33} \end{pmatrix} \begin{pmatrix} v^{11} & 0 & 0 \\ 0 & v^{22} & 0 \\ 0 & 0 & v^{33} \end{pmatrix}^{-1} \right\} \tag{20}$$

The proof of Theorem 2 is completed. \square

Acknowledgments The authors are grateful to the reviewers, the Associate Editor, and the Co-Editor for their insightful comments and suggestions which have improved the manuscript significantly.

References

- Cantoni, E. (2004). A robust approach to longitudinal data analysis. *The Canadian Journal of Statistics*, 32, 169–180.
- Croux, C., Gijbels, I., Prosdocomo, I. (2012). Robust estimation of mean and dispersion functions in extended generalized additive models. *Biometrics*, 68, 31–44.
- Daniels, M., Zhao, Y. (2003). Modelling the random effects covariance matrix on longitudinal data. *Statistics in Medicine*, 22, 1631–1647.
- Fan, J., Huang, T., Li, R. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function. *Journal of the American Statistical Association*, 35, 632–641.
- Fan, J., Wu, Y. (2008). Semiparametric estimation of covariance matrices for longitudinal data. *Journal of the American Statistical Association*, 103, 1520–1533.
- Fan, J., Zhang, J. T. (2000). Two-step estimation of functional linear models with application to longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62, 303–322.
- Fung, W. K., Zhu, Z. Y., He, X. (2002). Influence diagnostics and outlier tests for semiparametric mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 565–579.
- He, X., Fung, W. K., Zhu, Z. Y. (2005). Robust estimation in generalized partial linear models for clustered data. *Journal of the American Statistical Association*, 472, 1176–1184.
- Leng, C., Zhang, W., Pan, J. (2010). Semiparametric mean-covariance regression analysis for longitudinal data. *Journal of the American Statistical Association*, 105, 181–193.
- Levina, E., Rothman, A. J., Zhu, J. (2008). Sparse estimation of large covariance matrices via a nested Lasso penalty. *Annals of Applied Statistics*, 2, 245–263.
- Liang, K. Y., Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13–22.
- Mao, J., Zhu, Z. Y., Fung, W. K. (2011). Joint estimation of mean-covariance model for longitudinal data with basis function approximations. *Computational Statistics and Data Analysis*, 55, 983–992.
- McCullagh, P. (1983). Quasi-likelihood functions. *Annals of Statistics*, 11, 59–67.
- Pan, J., Mackenzie, G. (2003). On modelling mean-covariance structures in longitudinal studies. *Biometrika*, 90, 239–244.
- Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. *Biometrika*, 86, 677–690.
- Pourahmadi, M. (2000). Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika*, 87, 425–435.
- Qin, G. Y., Zhu, Z. Y. (2007). Robust estimation in generalized semiparametric mixed models for longitudinal data. *Journal of Multivariate Analysis*, 98, 1658–1683.
- Qin, G. Y., Zhu, Z. Y., Fung, W. K. (2009). Robust estimation of covariance parameters in partial linear model for longitudinal data. *Journal of Statistical Planning and Inference*, 139, 558–570.
- Qu, A., Lindsay, B., Li, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika*, 87, 823–836.
- Sinha, S. (2004). Robust analysis of generalized linear mixed models. *Journal of the American Statistical Association*, 99, 451–460.
- Wang, N. (2003). Marginal nonparametric kernel regression accounting within-subject correlation. *Biometrika*, 90, 29–42.
- Wang, N., Carroll, R. J., Lin, X. (2005a). Efficient semiparametric marginal estimation for longitudinal/clustered data. *Journal of the American Statistical Association*, 100, 147–157.
- Wang, Y. G., Lin, X., Zhu, M. (2005b). Robust estimation functions and bias correction for longitudinal data analysis. *Biometrics*, 61, 684–691.
- Wu, W., Pourahmadi, M. (2003). Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika*, 90, 831–844.
- Ye, H., Pan, J. (2006). Modelling covariance structures in generalized estimating equations for longitudinal data. *Biometrika*, 93, 927–941.
- Zhang, D. W., Lin, X. H., Raz, J., Sowers, M. F. (1998). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association*, 93, 710–719.