Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 99 (2008) 1860–1877
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

Case deletion diagnostics in multilevel models

Lei Shi [a], Gemai Chen [b,c,*]

[a] College of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming, Yunnan 650221, PR China
[b] Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, T2N 1N4, Canada
[c] Department of Applied Mathematics, Southwest Jiaotong University, Chengdu, Sichuan, PR China

Received 24 January 2007
Available online 17 February 2008

Abstract

This paper studies case deletion diagnostics for multilevel models. Using subset deletion, diagnostic measures for identifying influential units at any level are developed for both fixed and random parameters. Two approximate update formulae are derived. The first formula uses one-step approximation, while the second formula also includes the impact of estimating the random parameter. Two examples are used to illustrate the methodology developed.

AMS 2000 subject classifications: primary 62J20; secondary 62J05; 62H99

Keywords: Multilevel models; Case deletion; Influential observations; Cook's distance

1. Introduction

Influence diagnostics in ordinary least squares regression is well studied in the literature, see [4,7,5,8,1]. Recently, identification of influential observations has been studied for more complex models with correlated errors and/or unknown covariance matrices. For example, Martin [15] suggested several influence measures in general linear models with correlated errors. Christensen et al. [6] developed an update formula to study the influence of observations in mixed models. Banerjee and Frees [3] studied influence diagnostics based on subject deletion in linear longitudinal models. Haslett [12] suggested a simpler case-deletion measure using marginal

* Corresponding author at: Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, T2N 1N4, Canada.
E-mail address: gchen@math.ucalgary.ca (G. Chen).

0047-259X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2008.01.023

and conditional residuals. Hodges [14] studied case influence in hierarchical models based on a reformulation of the model and an approximate case deletion formula, and Haslett and Dillane [13] developed a case deletion diagnostic for estimates of the variance components in linear mixed models. We note however that no influence diagnostic measures have been studied for multilevel models.

For models with unknown covariance matrices, it is usually necessary to estimate both regression coefficients and parameters in covariance matrices in an iterative manner. In all of the above mentioned works, the authors conduct a separate influence analysis on estimates of regression coefficients and estimates of parameters in covariance matrices. In doing so, authors often ignore the fact that they are using a case deletion approach and yet plug in estimates based on the entire data set for convenience. As pointed out by Hodges [14, p. 507], Atkinson [2, p. 521], and Haslett and Dillane [13, p. 142], this practice is flawed because deletion of data points affects estimation of both regression coefficients and parameters in covariance matrices. So far, little has been done to look after this issue.

In this paper we study subset deletion diagnostic measures for multilevel models, where regression coefficients are called the fixed parameter and the index of covariance matrix is called the random parameter. Case deletion measures based on Cook's distance are developed to identify influential units at any level for both fixed parameter estimation and random parameter estimation. A special effort is made to remedy the common flaw discussed above.

The rest of the paper is organized as below. In Section 2 we introduce multilevel models and their parameter estimation. In Section 3 we develop our diagnostic measures. Two approximate update formulae are obtained to make it easy to use our diagnostic measures. In Section 4 we specialize the general results of Section 3 to the case of two-level models. In Section 5 we use two real data sets to illustrate the use of our diagnostic measures, and in Section 6 we provide some discussions and draw conclusions. Technical details are provided in the Appendix.

2. Multilevel model

A general multilevel model can be written as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \quad \mathbf{e} = \mathbf{Z}^{(s)}\mathbf{e}^{(s)} + \cdots + \mathbf{Z}^{(1)}\mathbf{e}^{(1)}, \tag{2.1}$$

where $\mathbf{Y}$ is an $N \times 1$ response vector, $\mathbf{X}$ is a known $N \times p$ design matrix of explanatory variables related to the unknown $p \times 1$ fixed (effects) parameter β, $\mathbf{e}^{(k)}$ $(k = 1, 2, \ldots, s)$ is the random error at level k and $\mathbf{Z}^{(k)}$ is the associated design matrix.

Let $\mathbf{V} = \mathrm{cov}(\mathbf{e})$ denote the unknown $N \times N$ covariance matrix of $\mathbf{e}$. In applications involving multilevel models, $\mathbf{V}$ can often be expressed as $\mathbf{V} = \mathbf{V}(\theta)$, where θ is an $R \times 1$ vector called the random (effects) parameter and satisfies $\mathrm{vec}(\mathbf{V}(\theta)) = \mathbf{Z}^{*}\theta$, where $\mathbf{Z}^{*}$ is a known matrix and vec is the vector operator stacking the columns of a matrix. If θ is known, the fixed parameter β is estimated as

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}. \tag{2.2}$$

Let $\hat{\mathbf{e}} = \mathbf{Y} - \mathbf{X}\hat{\beta}$ and $\mathbf{Y}^{**} = \mathrm{vec}(\hat{\mathbf{e}}\hat{\mathbf{e}}')$. Thinking from the point of $E(\mathbf{Y}^{**}) \approx \mathbf{Z}^{*}\theta$, the random parameter θ is estimated by

$$\hat{\theta} = (\mathbf{Z}^{*'}\mathbf{V}^{*-1}\mathbf{Z}^{*})^{-1}\mathbf{Z}^{*'}\mathbf{V}^{*-1}\mathbf{Y}^{**}, \tag{2.3}$$

where $\mathbf{V}^{*} = \mathbf{V} \otimes \mathbf{V}$, and $\otimes$ is the Kronecker product. Iterative generalized least square (IGLS) estimation consists of an iterative procedure which alternates between estimation of the fixed parameter β and the random parameter θ until convergence. Covariance matrices of $\hat{\beta}$ and $\hat{\theta}$ at convergence are given respectively by [11, p. 40]

$$\operatorname{cov}(\hat{\beta}) = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}, \qquad \operatorname{cov}(\hat{\theta}) = 2(\mathbf{Z}^{*'}\hat{\mathbf{V}}^{*-1}\mathbf{Z}^{*})^{-1}, \tag{2.4}$$

where $\hat{\mathbf{V}} = \mathbf{V}(\hat{\theta})$. Under the normality assumption for $\mathbf{e}^{(k)}$ $(k = 1, \dots s)$, $\hat{\beta}$ and $\hat{\theta}$ are the maximum likelihood estimators of β and θ, respectively [9].

3. Case deletion measures

We consider the general subset deletion approach. Let a denote the index of a subset of observations. Let $\hat{\beta}_{[a]}$ and $\hat{\theta}_{[a]}$ denote the IGLS estimators of β and θ respectively when the observations indexed by a are removed from the corresponding vectors and matrices, then we have

$$\begin{aligned} \hat{\beta}_{[a]} &= (\mathbf{X}'_{[a]}\mathbf{V}^{-1}_{[a]}(\hat{\theta}_{[a]})\mathbf{X}_{[a]})^{-1}\mathbf{X}'_{[a]}\mathbf{V}^{-1}_{[a]}(\hat{\theta}_{[a]})\mathbf{Y}_{[a]}, \\ \hat{\theta}_{[a]} &= (\mathbf{Z}^{*'}_{[a]}\mathbf{V}^{*-1}_{[a]}(\hat{\theta}_{[a]})\mathbf{Z}^{*}_{[a]})^{-1}\mathbf{Z}^{*'}_{[a]}\mathbf{V}^{*-1}_{[a]}(\hat{\theta}_{[a]})\mathbf{Y}^{**}_{[a]}, \end{aligned} \tag{3.1}$$

where $Y^{**}_{[a]} = \operatorname{vec}(\hat{\mathbf{e}}_{[a]}\hat{\mathbf{e}}'_{[a]})$, $\hat{\mathbf{e}}_{[a]} = \mathbf{Y}_{[a]} - \mathbf{X}_{[a]}\hat{\beta}_{[a]}$, $\mathbf{V}^{*}_{[a]}(\hat{\theta}_{[a]}) = \mathbf{V}_{[a]}(\hat{\theta}_{[a]}) \otimes \mathbf{V}_{[a]}(\hat{\theta}_{[a]})$, and similarly for the other vectors and matrices. Cook's distance can then be used to measure the magnitude of influence through

$$\begin{aligned} C_a(\hat{\beta}) &= (\hat{\beta}_{[a]} - \hat{\beta})'\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X}(\hat{\beta}_{[a]} - \hat{\beta}), \\ C_a(\hat{\theta}) &= (\hat{\theta}_{[a]} - \hat{\theta})'\mathbf{Z}^{*'}\hat{\mathbf{V}}^{*-1}\mathbf{Z}^{*}(\hat{\theta}_{[a]} - \hat{\theta})/2. \end{aligned} \tag{3.2}$$

Because $\hat{\beta}_{[a]}$ and $\hat{\theta}_{[a]}$ are obtained through iterations, it is intractable to obtain exact update formulae from the full data based $\hat{\beta}$ and $\hat{\theta}$. One common approach is to replace $\hat{\beta}_{[a]}$ and $\hat{\theta}_{[a]}$ on the right side of (3.1) by $\hat{\beta}$ and $\hat{\theta}$ to get the so called one-step approximation

$$\begin{aligned} \tilde{\beta}_{[a]} &= (\mathbf{X}'_{[a]}\hat{\mathbf{V}}^{-1}_{[a]}\mathbf{X}_{[a]})^{-1}\mathbf{X}'_{[a]}\hat{\mathbf{V}}^{-1}_{[a]}\mathbf{Y}_{[a]}, \\ \tilde{\theta}_{[a]} &= (\mathbf{Z}^{*'}_{[a]}\hat{\mathbf{V}}^{*-1}_{[a]}\mathbf{Z}^{*}_{[a]})^{-1}\mathbf{Z}^{*'}_{[a]}\hat{\mathbf{V}}^{*-1}_{[a]}\mathbf{Y}^{**}_{0[a]}, \end{aligned} \tag{3.3}$$

where $Y^{**}_{0[a]} = \operatorname{vec}(\hat{\mathbf{e}}_{0[a]}\hat{\mathbf{e}}'_{0[a]})$, $\hat{\mathbf{e}}_{0[a]} = \mathbf{Y}_{[a]} - \mathbf{X}_{[a]}\hat{\beta}$, $\hat{\mathbf{V}}^{*}_{[a]} = \hat{\mathbf{V}}_{[a]} \otimes \hat{\mathbf{V}}_{[a]}$, and $\hat{\mathbf{V}}_{[a]} = \mathbf{V}_{[a]}(\hat{\theta})$. We call this approach type I approximation.

To help state our update formulae, we let $a = \{i_1, \dots, i_m\}$ and $\mathbf{D}_a = (d_{i_1}, \dots, d_{i_m})$, where d_{i_k} is an $N \times 1$ vector with the i_kth element equal to 1 and the rest equal to zero, $k = 1, \dots, m$. Let $\mathbf{I}_{[a]}$ denote the result of removing the rows in the $N \times N$ identity matrix that are indexed by a. Some properties of $\mathbf{I}_{[a]}$ and $\mathbf{D}_a$ are listed bellow:

$$\begin{aligned} &(1)\ \mathbf{X}_{[a]} = \mathbf{I}_{[a]}\mathbf{X},\ \mathbf{V}_{[a]} = \mathbf{I}_{[a]}\mathbf{V}\mathbf{I}'_{[a]}, \\ &(2)\ \mathbf{I}'_{[a]}\mathbf{I}_{[a]} = \mathbf{I}_N - \mathbf{D}_a\mathbf{D}'_a, \quad \mathbf{I}_{[a]}\mathbf{I}'_{[a]} = \mathbf{I}_{N-m}, \quad \mathbf{I}_{[a]}\mathbf{D}_a = 0, \\ &(3)\ \mathbf{V}^{-1}_{[a]} = \mathbf{I}_{[a]}(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{D}_a(\mathbf{D}'_a\mathbf{V}^{-1}\mathbf{D}_a)^{-1}\mathbf{D}'_a\mathbf{V}^{-1})\mathbf{I}'_{[a]}. \end{aligned} \tag{3.4}$$

Properties (1) and (2) are obvious, and Property (3) can be checked directly.

The following theorem provides details for update formulae under type I approximation.

Theorem 3.1. *Let* $\hat{\beta}$ *and* $\hat{\theta}$ *denote IGLS estimates of* β *and* θ *respectively,* $\hat{\mathbf{V}} = \mathbf{V}(\hat{\theta})$*, and* $\hat{\mathbf{r}} = \hat{\mathbf{V}}^{-1}\hat{\mathbf{e}}$*. Then we have*

$$\begin{aligned}\tilde{\beta}_{[a]} &= \hat{\beta} - (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{D}_a(\mathbf{D}_a'\hat{\mathbf{Q}}\mathbf{D}_a)^{-1}\mathbf{D}_a'\hat{\mathbf{r}},\\ \tilde{\theta}_{[a]} &= \hat{\theta} - (\mathbf{Z}^{*'}\hat{\mathbf{V}}^{*-1}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*'}\hat{\mathbf{V}}^{*-1}(\mathbf{I}-\mathbf{B}_a)\mathbf{Y}^{**},\end{aligned} \tag{3.5}$$

where $\hat{\mathbf{Q}} = \hat{\mathbf{V}}^{-1} - \hat{\mathbf{V}}^{-1}\mathbf{X}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}$*,* $\mathbf{B}_a = \mathbf{Z}^*(\mathbf{Z}^{*'}(\mathbf{N}_a\otimes\mathbf{N}_a)\mathbf{Z}^*)^{-1}\mathbf{Z}^{*'}(\mathbf{N}_a\otimes\mathbf{N}_a)$*, and* $\mathbf{N}_a = \hat{\mathbf{V}}^{-1} - \hat{\mathbf{V}}^{-1}\mathbf{D}_a(\mathbf{D}_a'\hat{\mathbf{V}}^{-1}\mathbf{D}_a)^{-1}\mathbf{D}_a'\hat{\mathbf{V}}^{-1}$.

Let $\hat{\mathbf{Q}}_{aa} = \mathbf{D}_a'\hat{\mathbf{Q}}\mathbf{D}_a$ and $\hat{\mathbf{r}}_a = \mathbf{D}_a'\hat{\mathbf{r}}$, then our case deletion measures based on type I approximation are

$$\begin{aligned}C_a^I(\hat{\beta}) &= \hat{\mathbf{r}}_a'\hat{\mathbf{Q}}_{aa}^{-1}\hat{\mathbf{P}}_{aa}\hat{\mathbf{Q}}_{aa}^{-1}\hat{\mathbf{r}}_a,\\ C_a^I(\hat{\theta}) &= \mathbf{Y}^{**'}(\mathbf{I}-\mathbf{B}_a)'\hat{\mathbf{P}}^*(\mathbf{I}-\mathbf{B}_a)\mathbf{Y}^{**}/2,\end{aligned} \tag{3.6}$$

where $\hat{\mathbf{P}} = \hat{\mathbf{V}}^{-1}\mathbf{X}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}$ and $\hat{\mathbf{P}}^* = \hat{\mathbf{V}}^{*-1}\mathbf{Z}^*(\mathbf{Z}^{*'}\hat{\mathbf{V}}^{*-1}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*'}\hat{\mathbf{V}}^{*-1}$. These measures allow us to conduct diagnostics for general multilevel models at any level.

Type I approximation is easy to implement but has the flaw discussed in Section 1. For example, if observations indexed by a are highly influential for the estimation of β and θ, then substituting $\hat{\beta}$ and $\hat{\theta}$ into (3.1) to get (3.3) admits the subset deletion framework but essentially ignores this influence or mixes this influence with other matters. To address this issue, we study a new approximation based on a Taylor expansion of $\mathbf{V}_{[a]}(\hat{\theta}_{[a]})$ around $\hat{\theta}$, namely,

$$\mathbf{V}_{[a]}(\hat{\theta}_{[a]}) = \mathbf{V}_{[a]}(\hat{\theta}) + \dot{\mathbf{V}}_{[a]}(\hat{\theta})(\hat{\theta}_{[a]} - \hat{\theta}) + o(\|\hat{\theta}_{[a]} - \hat{\theta}\|^2), \tag{3.7}$$

where $\dot{\mathbf{V}}_{[a]}(\hat{\theta}) = \partial\mathbf{V}_{[a]}(\theta)/\partial\theta|_{\hat{\theta}}$ is an $(N-m)\times(N-m)\times R$ array if a has m indices, and $\|\cdot\|$ denotes the Euclidean norm. By substituting (3.7) into (3.1) and ignoring the second order term $o(\|\hat{\theta}_{[a]} - \hat{\theta}\|^2)$, we derive two new update formulae which we call type II approximation.

Theorem 3.2. *Assume* $o(\|\hat{\theta}_{[a]} - \hat{\theta}\|^2)$ *is negligible, then our update formulae under type II approximation are*

$$\begin{aligned}\bar{\beta}_{[a]} &= \hat{\beta} - (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}[\mathbf{D}_a(\mathbf{D}_a'\hat{\mathbf{Q}}\mathbf{D}_a)^{-1}\mathbf{D}_a'\hat{\mathbf{r}} + (\mathbf{M}_a\hat{\mathbf{r}}\otimes\mathbf{M}_a)'\mathbf{Z}^*(\bar{\theta}_{[a]} - \hat{\theta})],\\ \bar{\theta}_{[a]} &= \hat{\theta} + (\mathbf{Z}^{*'}\mathbf{W}_a\mathbf{Z}^*)^{-1}\mathbf{Z}^{*'}\mathrm{vec}(\mathbf{M}_a\hat{\mathbf{r}}\hat{\mathbf{r}}'\mathbf{M}_a' - \mathbf{N}_a),\end{aligned} \tag{3.8}$$

where $\mathbf{M}_a = \mathbf{I}_N - \hat{\mathbf{Q}}\mathbf{D}_a(\mathbf{D}_a'\hat{\mathbf{Q}}\mathbf{D}_a)^{-1}\mathbf{D}_a'$*,* $\mathbf{W}_a = \mathbf{N}_a\otimes\mathbf{N}_a - \mathbf{N}_a\otimes(\mathbf{N}_a\mathbf{V}(\hat{\theta}_{[a]}^*)\mathbf{N}_a) - (\mathbf{N}_a\mathbf{V}(\hat{\theta}_{[a]}^*)\mathbf{N}_a)\otimes\mathbf{N}_a + \mathbf{M}_a\hat{\mathbf{Q}}\otimes(\mathbf{M}_a\hat{\mathbf{r}}\hat{\mathbf{r}}'\mathbf{M}_a') + \mathbf{M}_a\hat{\mathbf{r}}\hat{\mathbf{r}}'\mathbf{M}_a'\otimes\mathbf{M}_a\hat{\mathbf{Q}}$*,* $\hat{\theta}_{[a]}^* = (\mathbf{Z}^{*'}(\mathbf{N}_a\otimes\mathbf{N}_a)\mathbf{Z}^*)^{-1}\mathbf{Z}^{*'}\mathrm{vec}(\mathbf{M}_a\hat{\mathbf{r}}\hat{\mathbf{r}}'\mathbf{M}_a')$*, and* $\hat{\mathbf{Q}}$ *and* $\mathbf{N}_a$ *are given in Theorem 3.1.*

We see from Theorem 3.2 that the first equation in (3.8) has an extra term compared with $\tilde{\beta}_{[a]}$ in Theorem 3.1, which is due to the estimation of θ when cases indexed by a are removed. Our case deletion measures under type II approximation are

$$\begin{aligned}C_a^{\mathrm{II}}(\hat{\beta}) &= C_a^I(\hat{\beta}) + \Delta C_a(\hat{\beta}),\\ C_a^{\mathrm{II}}(\hat{\theta}) &= \mathrm{vec}'(\mathbf{M}_a\hat{\mathbf{r}}\hat{\mathbf{r}}'\mathbf{M}_a' - \mathbf{N}_a)\mathbf{G}_a\hat{\mathbf{V}}^{*-1}\mathbf{G}_a\mathrm{vec}(\mathbf{M}_a\hat{\mathbf{r}}\hat{\mathbf{r}}'\mathbf{M}_a' - \mathbf{N}_a)/2,\end{aligned} \tag{3.9}$$

where $\mathbf{G}_a = \mathbf{Z}^*(\mathbf{Z}^{*'}\mathbf{W}_a\mathbf{Z}^*)^{-1}\mathbf{Z}^{*'}$, $C_a^I(\hat{\beta})$ is given in (3.6), and

$$\Delta C_a(\hat{\beta}) = (\bar{\theta}_{[a]} - \hat{\theta})'\mathbf{Z}^*(\mathbf{M}_a\hat{\mathbf{r}} \otimes \mathbf{M}_a)'\hat{\mathbf{P}} \times [(\mathbf{M}_a\hat{\mathbf{r}} \otimes \mathbf{M}_a)\mathbf{Z}^{*'}(\bar{\theta}_{[a]} - \hat{\theta}) + 2\mathbf{D}_a(\mathbf{D}_a'\hat{\mathbf{Q}}\mathbf{D}_a)^{-1}\mathbf{D}_a'\hat{\mathbf{r}}].$$

We note that $C_a^{\mathrm{II}}(\hat{\beta})$ has two parts. The first part is the deletion measure of $\hat{\beta}$ when the random parameter θ is fixed at $\hat{\theta}$. The second part $\Delta C_a(\hat{\beta})$ measures the change of the estimator of β due to the estimation of θ when the observations indexed by a are deleted. If $\hat{\theta}_{[a]} \approx \hat{\theta}$, $C_a^{\mathrm{II}}(\hat{\beta})$ reduces to $C_a^I(\hat{\beta})$.

To compute our type I and type II subset deletion measures, one needs to deal with some complicated matrices, such as $\mathbf{Z}^*$, $\mathbf{V}^{*-1}$, $(\mathbf{M}_a\hat{\mathbf{r}} \otimes \mathbf{M}_a)'\mathbf{Z}^*$, and $\mathbf{Z}^{*'}\mathbf{W}_a\mathbf{Z}^*$. However computation can be simplified because in common applications of multilevel models $\mathbf{V}$ can be written as $\mathbf{V}(\theta) = \sum_{i=1}^R \mathbf{A}_i\theta_i$, where $\mathbf{A}_i$ has the same dimensions as $\mathbf{V}$ and is symmetric and R is the dimension of θ. As a result, $\mathbf{Z}^* = (\mathrm{vec}(\mathbf{A}_1), \ldots, \mathrm{vec}(\mathbf{A}_R))$ and some of the matrices in Eq. (3.8) can be decomposed into several simpler matrices. For example, $(\mathbf{M}_a\hat{\mathbf{r}} \otimes \mathbf{M}_a)'\mathbf{Z}^*$ in the first equation of (3.8) is an $N \times R$ matrix and its jth column is given by

$$(\mathbf{M}_a\hat{\mathbf{r}} \otimes \mathbf{M}_a)'\mathrm{vec}(\mathbf{A}_j) = \mathbf{M}_a'\mathbf{A}_j\mathbf{M}_a\hat{\mathbf{r}}.$$

$\mathbf{Z}^{*'}\mathrm{vec}(\mathbf{M}_a\hat{\mathbf{r}}\hat{\mathbf{r}}'\mathbf{M}_a' - \mathbf{N}_a)$ in the second equation of (3.8) is an $R \times 1$ vector and its jth element is given by

$$\mathrm{vec}'(\mathbf{A}_j)\mathrm{vec}(\mathbf{M}_a\hat{\mathbf{r}}\hat{\mathbf{r}}'\mathbf{M}_a' - \mathbf{N}_a) = \mathrm{tr}(\mathbf{M}_a\hat{\mathbf{r}}\hat{\mathbf{r}}'\mathbf{M}_a'\mathbf{A}_j - \mathbf{N}_a\mathbf{A}_j).$$

Similarly, $\mathbf{Z}^{*'}\mathbf{W}_a\mathbf{Z}^*$ is an $R \times R$ matrix, and its (j, k)th element is

$$\mathrm{vec}'(\mathbf{A}_j)\mathbf{W}_a\mathrm{vec}(\mathbf{A}_k) = \mathrm{tr}(\mathbf{A}_j\mathbf{N}_a\mathbf{A}_k\mathbf{N}_a - 2\mathbf{A}_j\mathbf{N}_a\mathbf{V}(\tilde{\theta}_{[a]})\mathbf{N}_a\mathbf{A}_k\mathbf{N}_a + 2\mathbf{A}_j\mathbf{M}_a\hat{\mathbf{Q}}\mathbf{A}_k\mathbf{M}_a\hat{\mathbf{r}}\hat{\mathbf{r}}'\mathbf{M}_a').$$

4. Results for two-level models

We use two-level models to illustrate the use of subset deletion measures developed in Section 3. We first consider the case of deletion measures from a type I approximation. Let a_i denote the indices for all observations in unit i at level 2. Our case deletion measures for the ith unit at level 2 become

$$C_{a_i}^I(\hat{\beta}) = \hat{\mathbf{r}}_i'\hat{\mathbf{Q}}_{ii}^{-1}\hat{\mathbf{P}}_{ii}\hat{\mathbf{Q}}_{ii}^{-1}\hat{\mathbf{r}}_i, \qquad C_{a_i}^I(\hat{\theta}) = \hat{\mathbf{r}}_i^{*'}\hat{\mathbf{Q}}_{ii}^{*-1}\hat{\mathbf{P}}_{ii}^*\hat{\mathbf{Q}}_{ii}^{*-1}\hat{\mathbf{r}}_i^*/2, \tag{4.1}$$

where $\hat{\mathbf{r}}_i = \hat{\mathbf{V}}_i^{-1}\hat{\mathbf{e}}_i$, $\hat{\mathbf{e}}_i = \mathbf{y}_i - \mathbf{X}_i\hat{\beta}$, $\hat{\mathbf{P}}_{ii} = \hat{\mathbf{V}}_i^{-1}\mathbf{X}_i(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}_i'\hat{\mathbf{V}}_i^{-1}$, $\hat{\mathbf{Q}}_{ii} = \hat{\mathbf{V}}_i^{-1} - \hat{\mathbf{P}}_{ii}$, $\hat{\mathbf{r}}_i^* = \hat{\mathbf{V}}_i^{*-1}\hat{\mathbf{e}}_i^*$, $\hat{\mathbf{e}}_i^* = \mathbf{Y}_i^{**} - \mathbf{Z}_i^*\hat{\theta}$, $\hat{\mathbf{Q}}_{ii}^* = \hat{\mathbf{V}}_i^{*-1} - \hat{\mathbf{P}}_{ii}^*$, and $\hat{\mathbf{P}}_{ii}^* = \hat{\mathbf{V}}_i^{*-1}\mathbf{Z}_i^*(\mathbf{Z}^{*'}\hat{\mathbf{V}}^{*-1}\mathbf{Z}^*)^{-1}\mathbf{Z}_i^{*'}\hat{\mathbf{V}}_i^{*-1}$, where $\mathbf{V}_i^* = \mathbf{V}_i \otimes \mathbf{V}_i$, and $\mathbf{Z}_i^*$ satisfies $\mathrm{vec}(\mathbf{V}_i) = \mathbf{Z}_i^*\theta$.

The first equation of (4.1) corresponds to the case deletion measure for the estimate of regression coefficients in linear models with correlated error [12] or linear longitudinal models [3], while the second equation is a new case deletion measure for random parameter estimation in multilevel models. Following the standard approach in influence analysis, a suitable measure of the leverage of unit i for the fixed parameter estimate can be defined as

$$h_i^{(2)} = \sum_j \hat{p}_{i,jj}/\hat{v}_i^{jj}, \tag{4.2}$$

where $\hat{p}_{i,jj}$ and $\hat{v}_i^{jj}$ denote the jth diagonal elements of $\hat{\mathbf{P}}_{ii}$ and $\hat{\mathbf{V}}_i^{-1}$, respectively. The motivation of this definition is that $\hat{p}_{i,jj}/\hat{v}_i^{jj}$ is the generalized leverage value of the jth observation in the ith unit for the fixed parameter (see (4.4)).

To study the influence of individual observations at level 1, let $a = \{i_j\}$ denote the index of the jth observation in the ith unit, and let $D_a = \mathbf{d}_{i_j}$ be the $N \times 1$ vector with the i_jth element equal to 1 and the rest equal to zero. Then $\mathbf{D}_a'\hat{\mathbf{Q}}\mathbf{D}_a = \hat{v}_i^{jj} - \hat{p}_{i,jj}$, the jth diagonal element of the matrix $\hat{\mathbf{Q}}_{ii}$, and $\mathbf{D}_a\hat{\mathbf{r}} = \hat{r}_{ij}$, the jth element of the vector $\hat{\mathbf{r}}_i$. Our case deletion measure for $\hat{\beta}$ now becomes

$$C_{ij}^I(\hat{\beta}) = \left(\frac{\hat{r}_{ij}}{\sqrt{\hat{v}_i^{jj} - \hat{p}_{i,jj}}}\right)^2 \frac{\hat{p}_{i,jj}}{\hat{v}_i^{jj} - \hat{p}_{i,jj}} = t_{ij}^2 \frac{h_{ij}^{(1)}}{1 - h_{ij}^{(1)}}, \tag{4.3}$$

where $t_{ij} = \hat{r}_{ij}/\sqrt{\hat{v}_i^{jj} - \hat{p}_{i,jj}}$ is the standardized residual, and

$$h_{ij}^{(1)} = \hat{p}_{i,jj}/\hat{v}_i^{jj} \tag{4.4}$$

is defined here as the generalized leverage value for the fixed parameter at level 1. It is noted that $h_{ij}^{(1)}$ turns out to be the same as the generalized leverage defined by Christensen, et al. [6] for linear mixed models. A simplified formula for the case deletion measure for $\hat{\theta}$ does not seem achievable.

For the case deletion measures from type II approximation, there are no simpler formulae available. However the computation is not difficult using techniques suggested at the end of Section 3. For example, when the influence of units at level 2 is of interest, the matrices involving a_i are $\mathbf{N}_{a_i}$, $\mathbf{M}_{a_i}\hat{\mathbf{Q}}$ and $\mathbf{M}_{a_i}\hat{\mathbf{r}}$. In this case, we see that $\mathbf{N}_{a_i}$ is block diagonal with the ith block a zero matrix and the jth block $\hat{\mathbf{V}}_j^{-1}$ for $j \neq i$. The (k, l)th block of matrix $\mathbf{M}_{a_i}\hat{\mathbf{Q}}$ is given by $\hat{\mathbf{Q}}_{kl} - \hat{\mathbf{Q}}_{ki}\hat{\mathbf{Q}}_{ii}^{-1}\hat{\mathbf{Q}}_{il}$ $(k, l = 1, \ldots, n)$ and $\mathbf{M}_{a_i}\hat{\mathbf{r}} = (\tilde{\mathbf{r}}_{1i}', \ldots, \tilde{\mathbf{r}}_{ni}')'$ with $\tilde{\mathbf{r}}_{ji} = \hat{\mathbf{r}}_j - \hat{\mathbf{Q}}_{ji}\hat{\mathbf{Q}}_{ii}^{-1}\hat{\mathbf{r}}_i$, where $\hat{\mathbf{Q}}_{ji} = -\hat{\mathbf{V}}_j^{-1}\mathbf{X}_j(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}_i'\hat{\mathbf{V}}_i^{-1}$ for $j \neq i$. The computation of case deletion measures for studying the influence of observations at level 1 can be treated similarly.

5. Illustrations

Two examples are used here to illustrate the use of our case deletion measures.

5.1. Serum bilirubin data

The data from [17, p. 280–292] has been studied by Shi and Ojeda [16] using the following two-level model

$$\begin{aligned} y_{ij} &= z_i\alpha + \gamma_{00} + \gamma_{10}t_{ij} + \gamma_{20}t_{ij}^2 + \gamma_{01}w_i + \gamma_{11}(t_{ij} \times w_i) \\ &\quad + \gamma_{21}(t_{ij}^2 \times w_i) + e_{1i}^{(2)}t_{ij} + e_{2i}^{(2)}t_{ij}^2 + e_{ij}^{(1)}, \end{aligned} \tag{5.1}$$

where y_{ij} is the serum bilirubin measurement on the ith patient in the jth week and t_{ij} is the corresponding week $(0, 1, 2, 3, 4)$, $i = 1, \ldots, 66$, $j = 1, \ldots, m_i$, m_i is the number of observations on the ith patient (several patients died during the study), w_i is the baseline serum

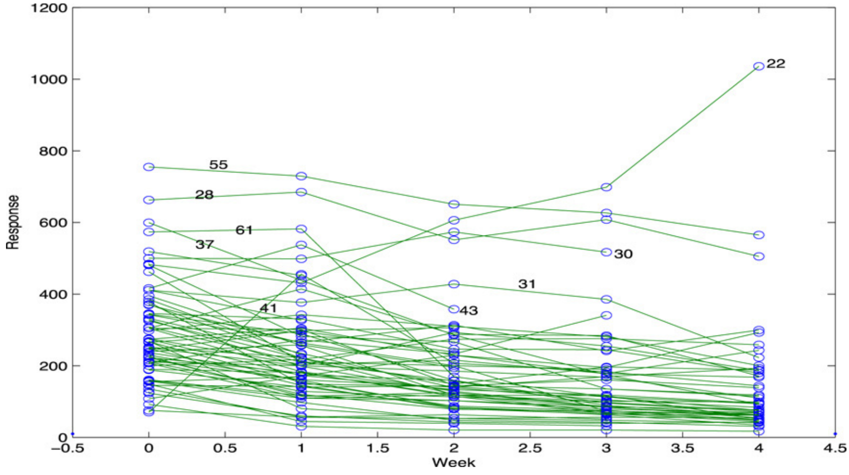

Fig. 1. Plot of y_{ij} versus t_{ij}.

bilirubin value, i.e., $y_{i0} = w_i$, and z_i is the indicator variable ($z_i = 1$ if the ith patient was treated; $z_i = 0$ otherwise). This model assumes that $\mathbf{V} = \text{diag}(\mathbf{V}_1, \ldots, \mathbf{V}_n)$, where $\mathbf{V}_i = \sigma^2 \mathbf{I}_{m_i} + \mathbf{T}_i \Omega \mathbf{T}_i'$, $\mathbf{T}_i = (\mathbf{t}_i^{(1)}, \mathbf{t}_i^{(2)})$, $\mathbf{t}_i^{(1)} = (t_{i1}, \ldots, t_{im_i})'$, $\mathbf{t}_i^{(2)} = (t_{i1}^2, \ldots, t_{im_i}^2)'$, $\Omega = \text{cov}(\mathbf{e}_i^{(2)}) = (\sigma_{kl})_{2\times 2}$, and $\text{var}(e_{ij}^{(1)}) = \sigma^2$. The fixed parameter is $\beta = (\alpha, \gamma_{00}, \gamma_{10}, \gamma_{20}, \gamma_{01}, \gamma_{11}, \gamma_{21})'$, and the random parameter is $\theta = (\sigma^2, \sigma_{11}, \sigma_{12}, \sigma_{22})'$. The IGLS estimates of β and θ are given in [16].

To judge the performance of our case deletion measures, we calculate the actual change of the parameter estimates due to case deletion for fixed and random parameters. These changes are measured by Cook's distances

$$\begin{aligned} AC_a(\hat{\beta}) &= (\hat{\beta} - \hat{\beta}_{[a]})' \mathbf{X}' \hat{\mathbf{V}}^{-1} \mathbf{X} (\hat{\beta} - \hat{\beta}_{[a]}), \\ AC_a(\hat{\theta}) &= (\hat{\theta} - \hat{\theta}_{[a]})' \mathbf{Z}^{*'} \hat{\mathbf{V}}^{*-1} \mathbf{Z}^* (\hat{\theta} - \hat{\theta}_{[a]})/2, \end{aligned} \tag{5.2}$$

where $\hat{\beta}_{[a]}$ and $\hat{\theta}_{[a]}$ are exactly calculated using the IGLS iterative algorithm, respectively, when observations indexed by a are deleted from the data. We compare diagnostic results obtained from Type I and Type II approximations with the results from actual changes to see the effectiveness of the two approximations.

We first focus on the influence of units (i.e. patients) at level 2. Shio and Ojeda [16] concluded, using local influence, that patients 22, 61, 37, 55, 28 and 30 in order are influential on $\hat{\beta}$, and patients 41, 22, and 61 in order are influential on $\hat{\theta}$. Fig. 1 plots y_{ij} versus t_{ij} for each patient, in which we see that these patients seem unusual compared to the other patients.

Fig. 2(a) is the index plot of the generalized leverage values for $\hat{\beta}$ at level 2 (defined as in (4.2)). We see that patients 55, 28, 37 and 61 have unusual leverage values. From the raw data plot in Fig. 1 we see that patients 55 and 28 have higher responses in all weeks, while patients 37 and 61 have larger responses in the baseline serum bilirubin value (w_i). Fig. 2(b) gives a scatter plot of level 2 predicted residuals of $\mathrm{e}_{1i}^{(2)}$ and $\mathrm{e}_{2i}^{(2)}$ in model (5.1). The plot reflects a strong negative correlation over patients ($\hat{\text{cor}}(\mathrm{e}_{1i}^{(2)}, \mathrm{e}_{2i}^{(2)}) = -0.8638$), and patients 22 and 41 are unusual. Fig. 2(c) and Fig. 2(d) give the index plots of case deletion measures for $\hat{\beta}$ and $\hat{\theta}$, respectively.

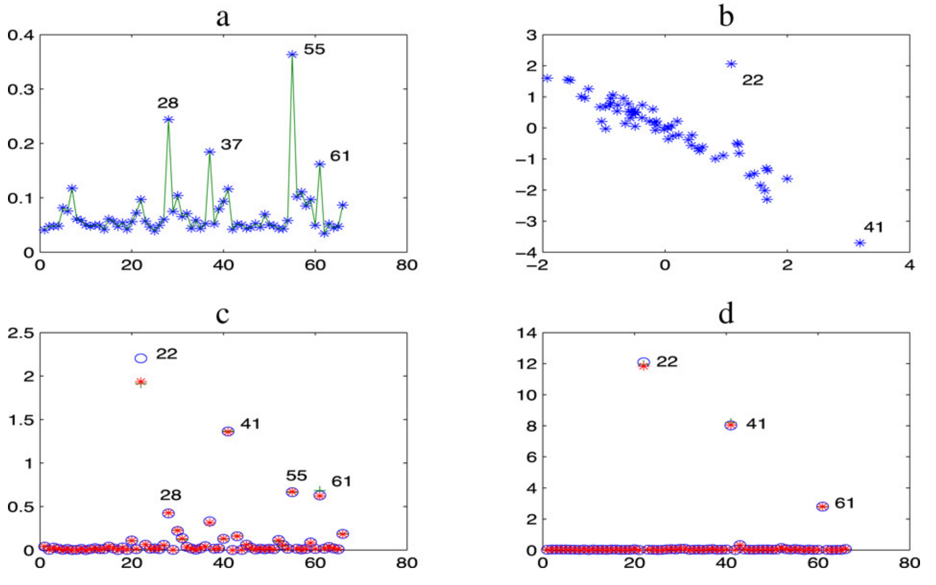

Fig. 2. (a) Index plot of generalized leverage values for $\hat{\beta}$ at level 2. (b) Scatter plot of predicted residuals of $e_{1i}^{(2)}$ and $e_{2i}^{(2)}$. (c) Index plot of the case deletion measure for $\hat{\beta}$ at level 2. (d) Index plot of the case deletion measure for $\hat{\theta}$ at level 2. In (c) and (d), o is for the actual change, + for Type I approximation and * for type II approximation.

In Fig. 2(c), we find that type I and type II approximations give almost the same results, and both approximate the actual changes very well. Patients 22 and 41 are highly influential, and patients 55, 61 and 28 are influential. These results are consistent with those of Shi and Ojeda [16]. Influential patterns for $\hat{\theta}$ as shown in Fig. 2(d) are very clear and the two approximations give the identical results as those from the actual changes.

Second, we examine which observations at level 1 are influential. Let (i, j) denote the jth observation in the ith unit. The index plots of generalized leverage values and standardized residuals are given in Fig. 3(a) and Fig. 3(b), respectively. From Fig. 3(a) we find that observations (55, 1), (28, 1), (55, 5) and (37, 1) have large leverage values. Fig. 3(b) indicates that observations (41, 2), (61, 2), (22, 5), (43, 2) and (61, 3) have large absolute residuals.

Case deletion measures for $\hat{\beta}$ and $\hat{\theta}$ obtained from three ways, namely actual change, type I approximation and type II approximation, are given in Fig. 3(c) and Fig. 3(d), respectively. Except for some observations labeled in the plots, type I and type II approximations are close to each other and together close to the actual change. However, for observation (22, 5), type II approximation is closer to the actual change than type I approximation. Specifically, the influential observations on $\hat{\beta}$ detected by type I approximation are (41, 2), (61, 2) and (22, 5), and the most influential observation is (41, 2). However, type II approximation finds the highly influential observations as (22, 5), (41, 2) and (61, 2), which are completely consistent with results based on actual changes. Similar conclusions can be drawn from Fig. 3(d) for influential patterns on $\hat{\theta}$. Therefore this example shows that type II approximation correctly identifies true influential patterns for serum bilirubin data, while type I approximation under-estimates the influence from observation (22, 5) on both fixed and random parameter estimations.

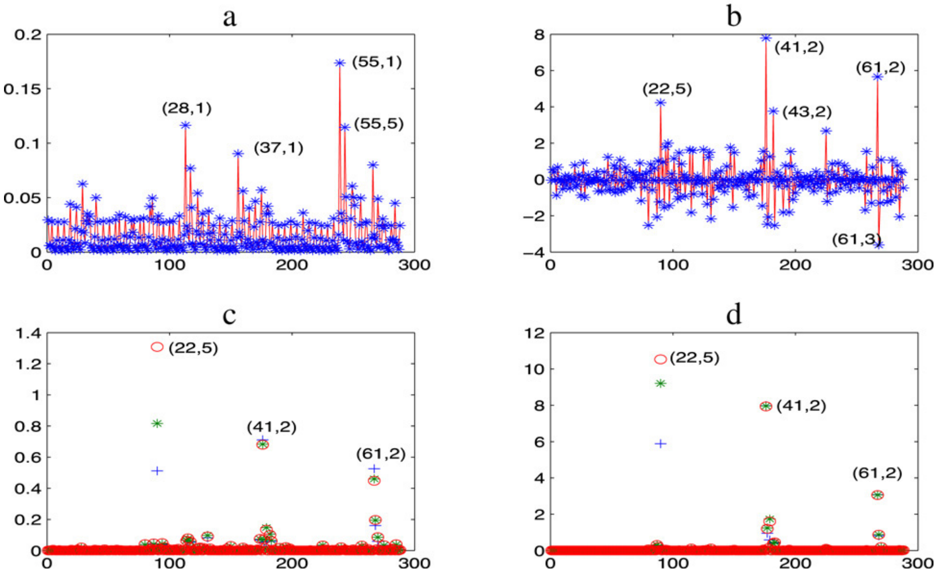

Fig. 3. (a) Index plot of generalized leverage values at level 1. (b) Index plot of standardized residuals at level 1. (c) Index plot of the case deletion measure for $\hat{\beta}$ at level 1. (d) Index plot of the case deletion measure for $\hat{\theta}$ at level 1. In (c) and (d), o is for the actual change, + for Type I approximation and * for type II approximation.

5.2. JSP data

The second example uses a data set that has been extensively analyzed by Goldstein [11]. The data consists of 728 pupils in 48 primary schools in Inner London as a part of the 'Junior School Project' (JSP). Two measurement occasions are considered: the first one was when the pupils were in their fourth year of schooling (8 years old), and the second one was three years later in their final year of primary school (11 years old). We have the scores from mathematics tests administered on these two occasions together with information collected on the social background of the pupils and their gender. Goldstein [11] suggested a two-level model given below to fit this data,

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + \mathrm{e}_{0i}^{(2)} + \mathrm{e}_{1i}^{(2)} x_{1ij} + \mathrm{e}_{ij}^{(1)}, \tag{5.3}$$

where y_{ij} = 11-year-old score for the jth pupil in the ith school, x_{1ij} = 8-year-old score centered by the sample mean, x_{2ij} = gender (1 for boy, 0 for girl), x_{3ij} = social class (1 for non-manual, 0 for manual). School ID counts from 1 to 50, but number 10 and 43 are missing, so we have 48 schools in the data. The IGLS estimates of the parameters using a software called MLn were given in [11, p. 49].

We first look at the residuals. Fig. 4(a) and (b) are index plots of level 2 predicted residuals for intercept and 8-year-old score, respectively. In Fig. 4(a), school 31 has the largest residual, while in Fig. 4(b), schools 38 and 33 have large absolute residuals. However outlying patterns are not clear. Fig. 4(c) is a scatter plot of the level 2 predicted residuals, indicating a negative correlation ($\hat{\mathrm{cor}}(\mathrm{e}_{0i}^{(2)}, \mathrm{e}_{1i}^{(2)}) = -0.706$). Fig. 4(d) shows the index plot of generalized leverage values for a fixed parameter estimation. We see that schools 48, 31 and 33 have large leverage values.

Now we use the diagnostic measures developed in this paper to study the influence of schools at level 2 on parameter estimates. Fig. 5(a) gives the index plot of case deletion measures for $\hat{\beta}$ at

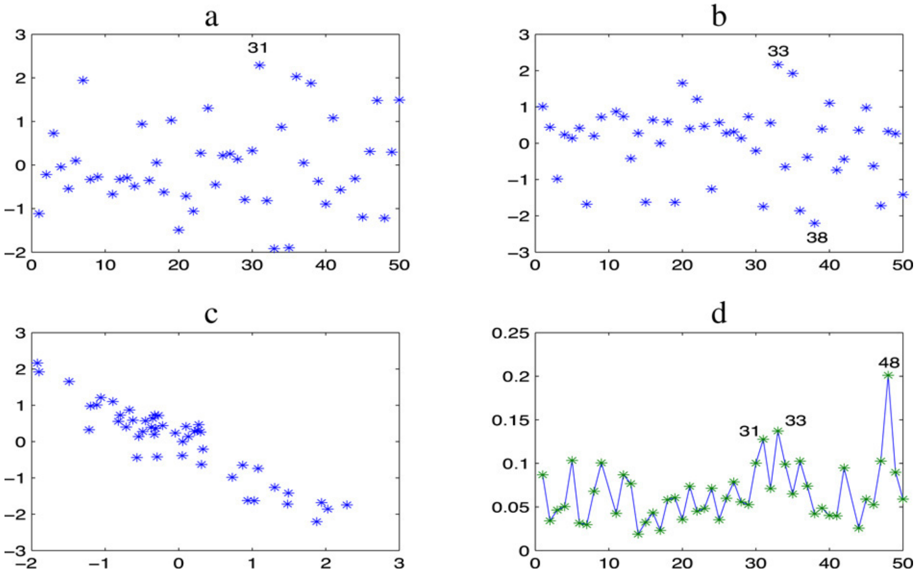

Fig. 4. (a) Index plot of level 2 predicted residuals for intercept. (b) Index plot of level 2 predicted residuals for 8-year-old score. (c) Scatter plot of predicted level 2 residuals. (d) Index plot of generalized leverage values for $\hat{\beta}$ at level 2.

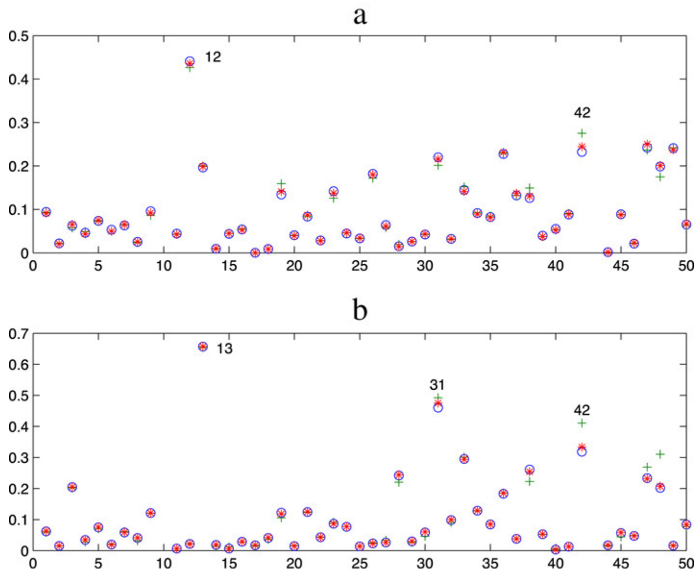

Fig. 5. (a) Index plot of case deletion measures for $\hat{\beta}$ at level 2 for JSP data. (b) Index plot of case deletion measures for $\hat{\theta}$ at level 2 for JSP data. Symbol o is for the actual change, + for Type I approximation, and * for type II approximation.

level 2 calculated from a type I approximation, a type II approximation and the actual change. We find that school 12 is highly influential, based on these two approximations, and the performance of type I and II approximations is very close to that of the actual change. However, a type I approximation over-estimates the actual influence for school 42.

Fig. 5(b) is the index plot of case deletion measures for $\hat{\theta}$ at level 2. It is seen that schools 13 and 31 are highly influential on $\hat{\theta}$. We see again that a Type II approximation follows the actual change very well, while a type I approximation tends to overestimate or underestimate the actual influence.

6. Discussions and conclusions

Subset deletion measures are very useful in multilevel models because they can be easily adopted to analyze the influence of units at any level. We have derived two approximate update formulae in this paper. One-step or type I approximation is simple and easy to interpret based on residuals and leverages. However this approximation conducts separate diagnostics for fixed parameters and random parameters, and has a weakness of ignoring or mixing actual influence. For this reason, we have derived a type II approximation which provides more effective diagnostics as shown in the examples.

The formulae in Theorems 3.1 and 3.2 are derived under the general structure for $\mathbf{V}(\theta)$ such that $\mathrm{vec}(\mathbf{V}(\theta)) = \mathbf{Z}^*\theta$ for some $\mathbf{Z}^*$, thus the results can be applied to linear mixed models with variance components, namely,

$$\mathbf{Y} = \mathbf{X}\beta + \sum_{j=1}^{r} \mathbf{Z}_j\gamma_j + \mathbf{e}, \tag{6.1}$$

where β is the fixed parameter, γ_j $(q_j \times 1)$, $j = 1, \ldots, r$, are random, independent and zero mean vectors with $\mathrm{cov}(\gamma_j) = \sigma_j^2\mathbf{I}_{q_j}$, $\mathbf{Z}_j$ are $N \times q_j$ design matrices for the random effects, and $\mathbf{e}$ is a random error vector (independent of γ_j) with $\mathrm{cov}(\mathbf{e}) = \sigma_0^2\mathbf{I}_N$. Christensen et al. [6] and Haslett and Dillane [13] studied the deletion diagnostic for restricted maximum likelihood estimation (REMLE) of variance components in model (6.1). We can apply the results of Section 3 to study the case deletion diagnostic for the MLE of the variance components because MLE is equivalent to IGLS under the normality assumption. From $\mathrm{cov}(\mathbf{Y}) = \mathbf{V} = \sum_{j=0}^{r}\sigma_j^2\mathbf{Z}_j\mathbf{Z}_j'$, where $\mathbf{Z}_0 = \mathbf{I}_N$, we have $\theta = (\sigma_0^2, \sigma_1^2, \ldots, \sigma_r^2)'$ and

$$\mathbf{Z}^* = (\mathrm{vec}(\mathbf{Z}_0\mathbf{Z}_0'), \mathrm{vec}(\mathbf{Z}_1\mathbf{Z}_1'), \ldots, \mathrm{vec}(\mathbf{Z}_r\mathbf{Z}_r')). \tag{6.2}$$

The MLE of θ at convergence is $\hat{\theta} = \mathbf{A}^{-1}\mathbf{b}$, where $\mathbf{A}$ is an $(r+1) \times (r+1)$ matrix with the (k,l)th element given by $a_{kl} = \mathrm{tr}(\mathbf{Z}_k'\hat{\mathbf{V}}^{-1}\mathbf{Z}_l\mathbf{Z}_l'\hat{\mathbf{V}}^{-1}\mathbf{Z}_k)$, and $\mathbf{b}$ is an $(r+1) \times 1$ vector with the lth element $\hat{\mathbf{r}}'\mathbf{Z}_l\mathbf{Z}_l'\hat{\mathbf{r}}$ $(k, l = 0, 1, \ldots, r)$. If type I approximation is used, then the case deletion measure for $\hat{\theta}$ in (3.6) reduces to

$$C_a^I(\hat{\theta}) = \mathbf{g}_a'\mathbf{A}^{-1}\mathbf{g}_a/2,$$

where $\mathbf{g}_a = (g_{a,0}, \ldots, g_{a,r})'$, $g_{a,i} = \hat{\mathbf{r}}'\mathbf{Z}_i\mathbf{Z}_i'\hat{\mathbf{r}} - \mathbf{a}_i'\mathbf{A}_a^{-1}\mathbf{r}_{za}$, $\mathbf{a}_i = (a_{i0}, \ldots, a_{ir})'$, the (k,l)th element of $\mathbf{A}_a$ and the lth element of $\mathbf{r}_{za}$ are given by

$$(\mathbf{A}_a)_{kl} = \mathrm{tr}(\mathbf{K}_{a,kl}\mathbf{K}_{a,kl}'), \qquad (\mathbf{r}_{za})_l = \mathbf{Z}_l'\hat{\mathbf{r}} - \tilde{\mathbf{Z}}_{l,a}'\mathbf{V}^{aa}\hat{\mathbf{r}}_a,$$

respectively, $\mathbf{K}_{a,kl} = \mathbf{Z}_k'\mathbf{N}_a\mathbf{Z}_l = \tilde{\mathbf{Z}}_k'\hat{\mathbf{V}}\tilde{\mathbf{Z}}_k - \tilde{\mathbf{Z}}_{k,a}\hat{\mathbf{V}}^{aa}\tilde{\mathbf{Z}}_{l,a}$, $\tilde{\mathbf{Z}}_k = \hat{\mathbf{V}}^{-1}\mathbf{Z}_k$, $\tilde{\mathbf{Z}}_{k,a}$ is the matrix containing rows of $\tilde{\mathbf{Z}}_k$ indexed by a, and $\hat{\mathbf{V}}^{aa}$ is the sub-matrix of $\hat{\mathbf{V}}^{-1}$ indexed by a. If type II approximation is employed, the calculation can be done by substituting (6.2) into (3.9) and using the techniques suggested at the end of Section 3. The influence diagnostic for $\hat{\beta}$ as shown in the first equation of (3.8) gives an insight on how $\hat{\theta}_{[a]}$ affects the influence measure for $\hat{\beta}$.

The case deletion diagnostics in this paper are derived under the framework of IGLS, which is equivalent to MLE under the normality assumption. However IGLS estimators are biased [11]. Unbiased estimators can be obtained by the restricted unbiased iterative generalized least squares (REIGLS) estimation in multilevel models, which is equivalent to the restricted maximum likelihood estimation (REMLE) under the normality assumption [10]. It would be interesting to extend the results of this paper to the REIGLS estimation in multilevel models.

Acknowledgments

We wish to thank the Editor, the Associate Editor and a referee for their helpful comments and suggestions which have helped clarify the presentation of the paper. The first author's research was supported by National Natural Science Foundation of China (10761010) and the second authors's research was supported by Natural Sciences and Engineering Research Council of Canada.

Appendix A. Proof of Theorem 3.1

Let $\hat{\mathbf{V}} = \mathbf{V}(\hat{\theta})$, $\hat{\mathbf{V}}_{[a]} = \mathbf{V}_{[a]}(\hat{\theta})$. From (3.4) and using the fact that $\mathbf{N}_a\mathbf{D}_a = 0$, we have

$$\mathbf{X}'_{[a]}\hat{\mathbf{V}}^{-1}_{[a]}\mathbf{X}_{[a]} = \mathbf{X}'\mathbf{N}_a\mathbf{X}, \qquad \mathbf{X}'_{[a]}\hat{\mathbf{V}}^{-1}_{[a]}\mathbf{Y}_{[a]} = \mathbf{X}'\mathbf{N}_a\mathbf{Y}, \tag{A.1}$$

where $\mathbf{N}_a = \hat{\mathbf{V}}^{-1} - \hat{\mathbf{V}}^{-1}\mathbf{D}_a(\mathbf{D}'_a\hat{\mathbf{V}}^{-1}\mathbf{D}_a)^{-1}\mathbf{D}'_a\hat{\mathbf{V}}^{-1}$. Therefore

$$\tilde{\beta}_{[a]} = (\mathbf{X}'\mathbf{N}_a\mathbf{X})^{-1}\mathbf{X}'\mathbf{N}_a\mathbf{Y}. \tag{A.2}$$

Note that

$$\begin{aligned}(\mathbf{X}'\mathbf{N}_a\mathbf{X})^{-1} &= (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1} + (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{D}_a[\mathbf{D}'_a\hat{\mathbf{V}}^{-1}\mathbf{D}_a \\ &\quad - \mathbf{D}'_a\hat{\mathbf{V}}^{-1}\mathbf{X}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{D}_a]^{-1}\mathbf{D}'_a\hat{\mathbf{V}}^{-1}\mathbf{X}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1} \\ &= (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1} + (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{D}_a \\ &\quad \times (\mathbf{D}'_a\hat{\mathbf{Q}}\mathbf{D}_a)^{-1}\mathbf{D}'_a\hat{\mathbf{V}}^{-1}\mathbf{X}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1},\end{aligned} \tag{A.3}$$

where $\hat{\mathbf{Q}} = \hat{\mathbf{V}}^{-1} - \hat{\mathbf{V}}^{-1}\mathbf{X}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}$. Thus

$$\begin{aligned}\tilde{\beta}_{[a]} &= \hat{\beta} - (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{D}_a(\mathbf{D}'_a\hat{\mathbf{V}}^{-1}\mathbf{D}_a)^{-1}\mathbf{D}'_a\hat{\mathbf{V}}^{-1}\mathbf{Y} \\ &\quad + (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{D}_a(\mathbf{D}'_a\hat{\mathbf{Q}}\mathbf{D}_a)^{-1}\mathbf{D}'_a\hat{\mathbf{V}}^{-1}\mathbf{X}\hat{\beta} \\ &\quad - (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{D}_a(\mathbf{D}'_a\hat{\mathbf{Q}}\mathbf{D}_a)^{-1}\mathbf{D}'_a\hat{\mathbf{P}}\mathbf{D}_a(\mathbf{D}'_a\hat{\mathbf{V}}^{-1}\mathbf{D}_a)^{-1}\mathbf{D}'_a\hat{\mathbf{V}}^{-1}\mathbf{Y} \\ &= \hat{\beta} - (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{D}_a(\mathbf{D}'_a\hat{\mathbf{Q}}\mathbf{D}_a)^{-1}\mathbf{D}'_a\hat{\mathbf{V}}^{-1}\hat{\mathbf{e}},\end{aligned}$$

because $\hat{\mathbf{Q}} = \hat{\mathbf{V}}^{-1} - \hat{\mathbf{P}}$, $\hat{\mathbf{e}} = \mathbf{Y} - \mathbf{X}\hat{\beta}$ and $\hat{\beta} = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{Y}$.

Below we prove the second equation in Theorem 3.1. Let $\mathbf{Z}^*_{[a]}$ denote the design matrix of θ when observations indexed by a are removed. Then

$$\mathrm{vec}[\mathbf{V}_{[a]}(\theta)] = \mathbf{Z}^*_{[a]}\theta.$$

Since

$$\mathrm{vec}[\mathbf{V}_{[a]}(\theta)] = \mathrm{vec}(\mathbf{I}_{[a]}\mathbf{V}(\theta)\mathbf{I}'_{[a]}) = (\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})\mathrm{vec}(\mathbf{V}(\theta)) = (\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})\mathbf{Z}^*\theta,$$

we have

$$\mathbf{Z}^*_{[a]} = (\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})\mathbf{Z}^*. \tag{A.4}$$

Using the properties in (3.4), there is

$$\hat{\mathbf{V}}^{*-1}_{[a]} = \hat{\mathbf{V}}^{-1}_{[a]} \otimes \hat{\mathbf{V}}^{-1}_{[a]} = (\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})(\mathbf{N}_a \otimes \mathbf{N}_a)(\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})'. \tag{A.5}$$

From (3.4) and $\mathbf{N}_a\mathbf{D}_a = 0$, we have

$$\begin{aligned}
\mathbf{Z}^{*'}_{[a]}\mathbf{V}^{*-1}_{[a]}(\hat{\theta})\mathbf{Z}^*_{[a]} &= \mathbf{Z}^{*'}(\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})'(\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})(\mathbf{N}_a \otimes \mathbf{N}_a)(\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})'(\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})\mathbf{Z}^* \\
&= \mathbf{Z}^{*'}(\mathbf{I}'_{[a]}\mathbf{I}_{[a]} \otimes \mathbf{I}'_{[a]}\mathbf{I}_{[a]})(\mathbf{N}_a \otimes \mathbf{N}_a)(\mathbf{I}'_{[a]}\mathbf{I}_{[a]} \otimes \mathbf{I}'_{[a]}\mathbf{I}_{[a]})\mathbf{Z}^* \\
&= \mathbf{Z}^{*'}((\mathbf{I}_N - \mathbf{D}_a\mathbf{D}'_a) \otimes (\mathbf{I}_N - \mathbf{D}_a\mathbf{D}'_a))(\mathbf{N}_a \otimes \mathbf{N}_a) \\
&\quad \cdot((\mathbf{I}_N - \mathbf{D}_a\mathbf{D}'_a) \otimes (\mathbf{I}_N - \mathbf{D}_a\mathbf{D}'_a))\mathbf{Z}^* \\
&= \mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^*,
\end{aligned} \tag{A.6}$$

and

$$\begin{aligned}
&\mathbf{Z}^{*'}_{[a]}\mathbf{V}^{*-1}_{[a]}(\hat{\theta})\mathbf{Y}^{**}_{0[a]} \\
&\quad = \mathbf{Z}^{*'}(\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})'(\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})(\mathbf{N}_a \otimes \mathbf{N}_a)(\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})' \\
&\qquad \times \mathrm{vec}[(\mathbf{Y}_{[a]} - \mathbf{X}_{[a]}\hat{\beta})(\mathbf{Y}_{[a]} - \mathbf{X}_{[a]}\hat{\beta})'] \\
&\quad = \mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)(\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})'(\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})\mathrm{vec}[(\mathbf{Y} - \mathbf{X}\hat{\beta})(\mathbf{Y} - \mathbf{X}\hat{\beta})'] \\
&\quad = \mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathrm{vec}[\hat{\mathbf{e}}\hat{\mathbf{e}}'].
\end{aligned}$$

Thus

$$\begin{aligned}
\tilde{\theta}_{[a]} &= (\mathbf{Z}^{*'}_{[a]}\mathbf{V}^{*-1}_{[a]}(\hat{\theta})\mathbf{Z}^*_{[a]})^{-1}\mathbf{Z}^{*'}_{[a]}\mathbf{V}^{*-1}_{[a]}(\hat{\theta})\mathbf{Y}^{**}_{0[a]} \\
&= (\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^*)^{-1}\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Y}^{**},
\end{aligned} \tag{A.7}$$

which gives

$$\tilde{\theta}_{[a]} = \hat{\theta} - (\mathbf{Z}^{*'}\hat{\mathbf{V}}^{*-1}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*'}\hat{\mathbf{V}}^{*-1}(\mathbf{I} - \mathbf{B}_a)\mathbf{Y}^{**}, \tag{A.8}$$

as claimed in Theorem 3.1.

Appendix B. Proof of (4.1)

The first equation of (4.1) is obvious. For the second equation, note that

$$\mathbf{D}_{a_i} \hat{=} \mathbf{D}_i = (\mathbf{0}_{m_i \times n_1}, \ldots, \mathbf{I}_{m_i}, \ldots, \mathbf{0}_{m_i \times n_n})'$$

is an $N \times m_i$ matrix, and

$$\mathbf{V} = \mathrm{diag}(\mathbf{V}_1, \ldots, \mathbf{V}_n) = \sum_j \mathbf{D}_j\mathbf{V}_j\mathbf{D}'_j.$$

Suppose $\mathbf{Z}_j^*$ satisfies $\mathrm{vec}(\mathbf{V}_j) = \mathbf{Z}_j^*\theta$, then we have

$$\mathbf{Z}^*\theta = \mathrm{vec}(\mathbf{V}) = \sum_j (\mathbf{D}_j \otimes \mathbf{D}_j)\mathrm{vec}(\mathbf{V}_j) = \sum_j (\mathbf{D}_j \otimes \mathbf{D}_j)\mathbf{Z}_j^*\theta,$$

which gives

$$\mathbf{Z}^* = \sum_j (\mathbf{D}_j \otimes \mathbf{D}_j)\mathbf{Z}_j^*.$$

Note that $\mathbf{N}_{a_i} = \hat{\mathbf{V}}^{-1} - \hat{\mathbf{V}}^{-1}\mathbf{D}_i\hat{\mathbf{V}}_i\mathbf{D}_i'\hat{\mathbf{V}}^{-1}$ and $\mathbf{D}_j'\mathbf{N}_{a_i}\mathbf{D}_k = \hat{\mathbf{V}}_j^{-1}$ for $j = k \neq i$ and $= 0$ otherwise, we have

$$\mathbf{Z}^{*'}(\mathbf{N}_{a_i} \otimes \mathbf{N}_{a_i})\mathbf{Z}^* = \sum_{j\neq i} \mathbf{Z}_j^{*'}\hat{\mathbf{V}}_j^{*-1}\mathbf{Z}_j^* = \mathbf{Z}^{*'}\hat{\mathbf{V}}^{*-1}\mathbf{Z}^* - \mathbf{Z}_i^{*'}\hat{\mathbf{V}}_i^{*-1}\mathbf{Z}_i^*,$$

$$\mathbf{Z}^{*'}(\mathbf{N}_{a_i} \otimes \mathbf{N}_{a_i})\mathbf{Y}^{**} = \sum_{j\neq i} \mathbf{Z}_j^{*'}\hat{\mathbf{V}}_j^{*-1}\mathbf{Y}_j^{**} = \mathbf{Z}^{*'}\hat{\mathbf{V}}^{*-1}\mathbf{Y}^{**} - \mathbf{Z}_i^{*'}\hat{\mathbf{V}}_i^{*-1}\mathbf{Y}_i^{**}.$$

Thus

$$\begin{aligned}
\tilde{\theta}_{[a_i]} &= (\mathbf{Z}^{*'}\hat{\mathbf{V}}^{*-1}\mathbf{Z}^* - \mathbf{Z}_i^{*'}\hat{\mathbf{V}}_i^{*-1}\mathbf{Z}_i^*)^{-1}(\mathbf{Z}^{*'}\hat{\mathbf{V}}^{*-1}\mathbf{Y}^{**} - \mathbf{Z}_i^{*'}\hat{\mathbf{V}}_i^{*-1}\mathbf{Y}_i^{**}) \\
&= [(\mathbf{Z}^{*'}\hat{\mathbf{V}}^{*-1}\mathbf{Z}^*)^{-1} + (\mathbf{Z}^{*'}\hat{\mathbf{V}}^{*-1}\mathbf{Z}^*)^{-1}\mathbf{Z}_i^{*'}\hat{\mathbf{V}}_i^{*-1}\hat{\mathbf{Q}}_{ii}^{*-1}\hat{\mathbf{V}}_i^{*-1}\mathbf{Z}_i^*(\mathbf{Z}^{*'}\hat{\mathbf{V}}^{*-1}\mathbf{Z}^*)^{-1}] \\
&\quad \cdot (\mathbf{Z}^{*'}\hat{\mathbf{V}}^{*-1}\mathbf{Y}^{**} - \mathbf{Z}_i^{*'}\hat{\mathbf{V}}_i^{*-1}\mathbf{Y}_i^{**}) \\
&= \hat{\theta} - (\mathbf{Z}^{*'}\hat{\mathbf{V}}^{*-1}\mathbf{Z}^*)^{-1}\mathbf{Z}_i^{*'}\hat{\mathbf{V}}_i^{*-1}\hat{\mathbf{Q}}_{ii}^{*-1}\mathbf{V}_i^{*-1}\hat{\mathbf{e}}_i^*, \qquad \text{(B.1)}
\end{aligned}$$

where $\hat{\mathbf{e}}_i^* = \mathbf{Y}_i^{**} - \mathbf{Z}_i^*\hat{\theta}$, $\hat{\mathbf{Q}}_{ii}^* = \hat{\mathbf{V}}_i^{*-1} - \hat{\mathbf{P}}_{ii}^*$ and $\hat{\mathbf{P}}_{ii}^* = \hat{\mathbf{V}}_i^{*-1}\mathbf{Z}_i^*(\mathbf{Z}^{*'}\hat{\mathbf{V}}^{*-1}\mathbf{Z}^*)^{-1}\mathbf{Z}_i^{*'}\hat{\mathbf{V}}_i^{*-1}$. Therefore,

$$C_{a_i}^I(\hat{\theta}) = \hat{\mathbf{r}}_i^{*'}\hat{\mathbf{Q}}_{ii}^{*-1}\hat{\mathbf{P}}_{ii}^*\hat{\mathbf{Q}}_{ii}^{*-1}\hat{\mathbf{r}}_i^*/2, \qquad \text{(B.2)}$$

where $\hat{\mathbf{r}}_i^* = \mathbf{V}_i^{*-1}\hat{\mathbf{e}}_i^*$.

Appendix C. Proof of Theorem 3.2

The IGLS of β and θ when the observations indexed by a are deleted are given in (3.1). Suppose we use the first order Taylor expansion to approximate $\mathbf{V}_{[a]}(\hat{\theta}_{[a]})$,

$$\mathbf{V}_{[a]}(\hat{\theta}_{[a]}) = \mathbf{V}_{[a]}(\hat{\theta}) + \mathbf{V}_{[a]}^{(1)} + o(\eta_a^2), \qquad \text{(C.1)}$$

where $\eta_a = \|\hat{\theta}_{[a]} - \hat{\theta}\|$, and $\mathbf{V}_{[a]}^{(1)} = \dot{\mathbf{V}}_{[a]}(\hat{\theta})(\hat{\theta}_{[a]} - \hat{\theta}) = o(\eta_a)$. It is easy to see that

$$\mathbf{V}_{[a]}^{(1)} = \mathbf{V}_{[a]}(\hat{\theta}_{[a]}) - \mathbf{V}_{[a]}(\hat{\theta}) + o(\eta_a^2) = \mathbf{I}_{[a]}(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{I}_{[a]}' + o(\eta_a^2). \qquad \text{(C.2)}$$

Therefore $\mathbf{V}_{[a]}^{(1)}$ satisfies

$$\begin{aligned}
\mathbf{I}_{[a]}'\mathbf{V}_{[a]}^{(1)}\mathbf{I}_{[a]} &= \mathbf{I}_{[a]}'\mathbf{I}_{[a]}(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{I}_{[a]}'\mathbf{I}_{[a]} + o(\eta_a^2) \\
&= (\mathbf{I}_N - \mathbf{D}_a\mathbf{D}_a')(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))(\mathbf{I}_N - \mathbf{D}_a\mathbf{D}_a') + o(\eta_a^2). \qquad \text{(C.3)}
\end{aligned}$$

Using the result that for any positive definite matrix $\mathbf{A}$ and small ϵ

$$[\mathbf{A} + \mathbf{B}\epsilon + o(\epsilon^2)]^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\epsilon + o(\epsilon^2),$$

and noting that $\hat{\mathbf{V}}^{(1)}_{[a]} = o(\eta_a)$, we have

$$\begin{aligned}\mathbf{V}^{-1}_{[a]}(\hat{\theta}_{[a]}) &= (\mathbf{V}_{[a]}(\hat{\theta}) + \hat{\mathbf{V}}^{(1)}_{[a]} + o(\eta_a^2))^{-1} \\ &= \hat{\mathbf{V}}^{-1}_{[a]} - \hat{\mathbf{V}}^{-1}_{[a]}\mathbf{V}^{(1)}_{[a]}\hat{\mathbf{V}}^{-1}_{[a]} + o(\eta_a^2), \end{aligned} \tag{C.4}$$

where $\hat{\mathbf{V}}^{-1}_{[a]} = \mathbf{V}^{-1}_{[a]}(\hat{\theta})$. From (A.1), (C.3) and (C.4), we find

$$\begin{aligned}&(\mathbf{X}'_{[a]}\mathbf{V}^{-1}_{[a]}(\hat{\theta}_{[a]})\mathbf{X}_{[a]})^{-1} \\ &\quad = (\mathbf{X}'_{[a]}\hat{\mathbf{V}}^{-1}_{[a]}\mathbf{X}_{[a]} - \mathbf{X}'_{[a]}\hat{\mathbf{V}}^{-1}_{[a]}\mathbf{V}^{(1)}_{[a]}\hat{\mathbf{V}}^{-1}_{[a]}\mathbf{X}_{[a]} + o(\eta_a^2))^{-1} \\ &\quad = (\mathbf{X}'_{[a]}\hat{\mathbf{V}}^{-1}_{[a]}\mathbf{X}_{[a]})^{-1} + (\mathbf{X}'_{[a]}\hat{\mathbf{V}}^{-1}_{[a]}\mathbf{X}_{[a]})^{-1}\mathbf{X}'_{[a]}\hat{\mathbf{V}}^{-1}_{[a]}\mathbf{V}^{(1)}_{[a]}\hat{\mathbf{V}}^{-1}_{[a]}\mathbf{X}_{[a]}(\mathbf{X}'_{[a]}\hat{\mathbf{V}}^{-1}_{[a]}\mathbf{X}_{[a]})^{-1} + o(\eta_a^2) \\ &\quad = (\mathbf{X}'\mathbf{N}_a\mathbf{X})^{-1} + (\mathbf{X}'\mathbf{N}_a\mathbf{X})^{-1}\mathbf{X}'\mathbf{I}'_{[a]}\mathbf{I}_{[a]}\mathbf{N}_a\mathbf{I}_{[a]}\mathbf{V}^{(1)}_{[a]}\mathbf{I}_{[a]}\mathbf{N}_a\mathbf{I}'_{[a]}\mathbf{I}_{[a]}\mathbf{X}(\mathbf{X}'\mathbf{N}_a\mathbf{X})^{-1} + o(\eta_a^2) \\ &\quad = (\mathbf{X}'\mathbf{N}_a\mathbf{X})^{-1} + (\mathbf{X}'\mathbf{N}_a\mathbf{X})^{-1}\mathbf{X}'\mathbf{N}_a(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{N}_a\mathbf{X}(\mathbf{X}'\mathbf{N}_a\mathbf{X})^{-1} + o(\eta_a^2).\end{aligned}$$

Similarly

$$\mathbf{X}'_{[a]}\mathbf{V}^{-1}_{[a]}(\hat{\theta}_{[a]})\mathbf{Y}_{[a]} = \mathbf{X}'\mathbf{N}_a\mathbf{Y} - \mathbf{X}'\mathbf{N}_a(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{N}_a\mathbf{Y} + o(\eta_a^2).$$

Substituting the above two equations into the first equation of (3.1), we have

$$\hat{\beta}_{[a]} = \tilde{\beta}_{[a]} - (\mathbf{X}'\mathbf{N}_a\mathbf{X})^{-1}\mathbf{X}'\mathbf{N}_a(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{N}_a(\mathbf{Y} - \mathbf{X}\tilde{\beta}_{[a]}) + o(\eta_a^2). \tag{C.5}$$

Using (A.3), we have

$$(\mathbf{X}'\mathbf{N}_a\mathbf{X})^{-1}\mathbf{X}'\mathbf{N}_a = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{M}'_a,$$

and

$$\begin{aligned}\mathbf{N}_a(\mathbf{Y} - \mathbf{X}\tilde{\beta}_{[a]}) &= \mathbf{N}_a[\mathbf{I}_N + \mathbf{X}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{D}_a(\mathbf{D}'_a\hat{\mathbf{Q}}\mathbf{D}_a)^{-1}\mathbf{D}'_a\hat{\mathbf{V}}^{-1}]\hat{\mathbf{e}} \\ &= [\hat{\mathbf{V}}^{-1} - \hat{\mathbf{V}}^{-1}\mathbf{D}_a(\mathbf{D}'_a\hat{\mathbf{V}}^{-1}\mathbf{D}_a)^{-1}\mathbf{D}'_a\hat{\mathbf{V}}^{-1}] \\ &\quad \times [\mathbf{I}_N + \mathbf{X}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{D}_a(\mathbf{D}'_a\hat{\mathbf{Q}}\mathbf{D}_a)^{-1}\mathbf{D}'_a\hat{\mathbf{V}}^{-1}]\hat{\mathbf{e}} \\ &= [\mathbf{I}_N + \hat{\mathbf{P}}\mathbf{D}_a(\mathbf{D}'_a\hat{\mathbf{Q}}\mathbf{D}_a)^{-1}\mathbf{D}'_a - \hat{\mathbf{V}}^{-1}\mathbf{D}_a(\mathbf{D}'_a\hat{\mathbf{V}}^{-1}\mathbf{D}_a)^{-1}\mathbf{D}'_a \\ &\quad - \hat{\mathbf{V}}^{-1}\mathbf{D}_a(\mathbf{D}'_a\hat{\mathbf{V}}^{-1}\mathbf{D}_a)^{-1}\mathbf{D}'_a\hat{\mathbf{P}}\mathbf{D}_a(\mathbf{D}'_a\hat{\mathbf{Q}}\mathbf{D}_a)^{-1}\mathbf{D}'_a]\hat{\mathbf{r}} \\ &= [\mathbf{I}_N + \hat{\mathbf{P}}\mathbf{D}_a(\mathbf{D}'_a\hat{\mathbf{Q}}\mathbf{D}_a)^{-1}\mathbf{D}'_a - \hat{\mathbf{V}}^{-1}\mathbf{D}_a(\mathbf{D}'_a\hat{\mathbf{Q}}\mathbf{D}_a)^{-1}\mathbf{D}'_a]\hat{\mathbf{r}} \\ &= [\mathbf{I}_N - \hat{\mathbf{Q}}\mathbf{D}_a(\mathbf{D}'_a\hat{\mathbf{Q}}\mathbf{D}_a)^{-1}\mathbf{D}'_a]\hat{\mathbf{r}} = \mathbf{M}_a\hat{\mathbf{r}},\end{aligned} \tag{C.6}$$

where $\hat{\mathbf{r}} = \hat{\mathbf{V}}^{-1}\hat{\mathbf{e}}$ and $\mathbf{M}_a = \mathbf{I}_N - \hat{\mathbf{Q}}\mathbf{D}_a(\mathbf{D}'_a\hat{\mathbf{Q}}\mathbf{D}_a)^{-1}\mathbf{D}'_a$. It follows that

$$\begin{aligned}\hat{\beta}_{[a]} &= \tilde{\beta}_{[a]} - (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{M}'_a(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{M}_a\hat{\mathbf{r}} + o(\eta_a^2) \\ &= \tilde{\beta}_{[a]} - (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}(\mathbf{M}_a\hat{\mathbf{r}} \otimes \mathbf{M}_a)'\mathrm{vec}(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta})) + o(\eta_a^2)\end{aligned}$$

$$= \hat{\beta} - (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}[\mathbf{D}_a(\mathbf{D}_a'\hat{\mathbf{Q}}\mathbf{D}_a)^{-1}\mathbf{D}_a'\hat{\mathbf{r}} + (\mathbf{M}_a\hat{\mathbf{r}} \otimes \mathbf{M}_a)'\mathbf{Z}^*(\hat{\theta}_{[a]} - \hat{\theta})] + o(\eta_a^2). \tag{C.7}$$

Next we consider the update formula for $\hat{\theta}$. We observe from (C.4) that

$$\begin{aligned}\mathbf{V}_{[a]}^{*-1}(\hat{\theta}_{[a]}) &= \mathbf{V}_{[a]}^{-1}(\hat{\theta}_{[a]}) \otimes \mathbf{V}_{[a]}^{-1}(\hat{\theta}_{[a]})\\ &= \hat{\mathbf{V}}_{[a]}^{-1} \otimes \hat{\mathbf{V}}_{[a]}^{-1} - (\hat{\mathbf{V}}_{[a]}^{-1} \otimes \hat{\mathbf{V}}_{[a]}^{-1}\mathbf{V}_{[a]}^{(1)}\hat{\mathbf{V}}_{[a]}^{-1} + \hat{\mathbf{V}}_{[a]}^{-1}\mathbf{V}_{[a]}^{(1)}\hat{\mathbf{V}}_{[a]}^{-1} \otimes \hat{\mathbf{V}}_{[a]}^{-1}) + o(\eta_a^2)\\ &= (\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})[\mathbf{N}_a \otimes \mathbf{N}_a - \mathbf{N}_a \otimes \mathbf{N}_a\mathbf{I}_{[a]}'\mathbf{V}_{[a]}^{(1)}\mathbf{I}_{[a]}\mathbf{N}_a\\ &\quad + \mathbf{N}_a\mathbf{I}_{[a]}'\mathbf{V}_{[a]}^{(1)}\mathbf{I}_{[a]}\mathbf{N}_a \otimes \mathbf{N}_a](\mathbf{I}_{[a]}' \otimes \mathbf{I}_{[a]}') + o(\eta_a^2).\end{aligned} \tag{C.8}$$

From (A.3), (A.4) and (C.8), we have

$$\begin{aligned}&(\mathbf{Z}_{[a]}^{*'}\mathbf{V}_{[a]}^{*-1}(\hat{\theta}_{[a]})\mathbf{Z}_{[a]}^{*})^{-1}\\ &= [\mathbf{Z}_{[a]}^{*'}\hat{\mathbf{V}}_{[a]}^{*-1}\mathbf{Z}_{[a]}^{*} - \mathbf{Z}_{[a]}^{*'}(\hat{\mathbf{V}}_{[a]}^{-1} \otimes \hat{\mathbf{V}}_{[a]}^{-1}\mathbf{V}_{[a]}^{(1)}\hat{\mathbf{V}}_{[a]}^{-1} + \hat{\mathbf{V}}_{[a]}^{-1}\mathbf{V}_{[a]}^{(1)}\hat{\mathbf{V}}_{[a]}^{-1} \otimes \hat{\mathbf{V}}_{[a]}^{-1})\mathbf{Z}_{[a]}^{*} + o(\eta_a^2)]^{-1}\\ &= [\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^* - \mathbf{Z}^{*'}(\mathbf{I}_{[a]}' \otimes \mathbf{I}_{[a]}')(\mathbf{I}_{[a]}\mathbf{N}_a\mathbf{I}_{[a]}' \otimes \mathbf{I}_{[a]}\mathbf{N}_a\mathbf{I}_{[a]}'\mathbf{V}_{[a]}^{(1)}\mathbf{I}_{[a]}\mathbf{N}_a\mathbf{I}_{[a]}'\\ &\quad + \mathbf{I}_{[a]}\mathbf{N}_a\mathbf{I}_{[a]}'\mathbf{V}_{[a]}^{(1)}\mathbf{I}_{[a]}\mathbf{N}_a\mathbf{I}_{[a]}' \otimes \mathbf{I}_{[a]}\mathbf{N}_a\mathbf{I}_{[a]}')(\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})\mathbf{Z}^* + o(\eta_a^2)]^{-1}\\ &= [\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^* - \mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a\mathbf{I}_{[a]}'\mathbf{V}_{[a]}^{(1)}\mathbf{I}_{[a]}\mathbf{N}_a\\ &\quad + \mathbf{N}_a\mathbf{I}_{[a]}'\mathbf{V}_{[a]}^{(1)}\mathbf{I}_{[a]}\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^* + o(\eta_a^2)]^{-1}\\ &= (\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^*)^{-1} + (\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^*)^{-1}\{\mathbf{Z}^{*'}[\mathbf{N}_a \otimes \mathbf{N}_a\mathbf{I}_{[a]}'\mathbf{V}_{[a]}^{(1)}\mathbf{I}_{[a]}\mathbf{N}_a +\\ &\quad + \mathbf{N}_a\mathbf{I}_{[a]}'\mathbf{V}_{[a]}^{(1)}\mathbf{I}_{[a]}\mathbf{N}_a \otimes \mathbf{N}_a]\mathbf{Z}^*\}(\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^*)^{-1} + o(\eta_a^2)\\ &= (\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^*)^{-1} + (\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^*)^{-1}\mathbf{Z}^*[\mathbf{N}_a \otimes \mathbf{N}_a(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{N}_a\\ &\quad + \mathbf{N}_a(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{N}_a \otimes \mathbf{N}_a]\mathbf{Z}^{*'}(\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^*)^{-1} + o(\eta_a^2).\end{aligned} \tag{C.9}$$

Since

$$\mathbf{Y}_{[a]}^{**} = (\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})\mathrm{vec}((\mathbf{Y} - \mathbf{X}\hat{\beta}_{[a]})(\mathbf{Y} - \mathbf{X}\hat{\beta}_{[a]})'),$$

we have

$$\begin{aligned}&\mathbf{Z}_{[a]}^{*'}\mathbf{V}_{[a]}^{*-1}(\hat{\theta}_{[a]})\mathbf{Y}_{[a]}^{**}\\ &= \mathbf{Z}^{*'}(\mathbf{I}_{[a]}' \otimes \mathbf{I}_{[a]}')[(\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})(\mathbf{N}_a \otimes \mathbf{N}_a - \mathbf{N}_a \otimes \mathbf{N}_a\mathbf{I}_{[a]}'\mathbf{V}_{[a]}^{(1)}\mathbf{I}_{[a]}\mathbf{N}_a\\ &\quad - \mathbf{N}_a\mathbf{I}_{[a]}'\mathbf{V}_{[a]}^{(1)}\mathbf{I}_{[a]}\mathbf{N}_a \otimes \mathbf{N}_a)(\mathbf{I}_{[a]}' \otimes \mathbf{I}_{[a]}')](\mathbf{I}_{[a]} \otimes \mathbf{I}_{[a]})\\ &\quad \times \mathrm{vec}((\mathbf{Y} - \mathbf{X}\hat{\beta}_{[a]})(\mathbf{Y} - \mathbf{X}\hat{\beta}_{[a]})') + o(\eta_a^2)\\ &= \mathbf{Z}^{*'}[\mathbf{N}_a \otimes \mathbf{N}_a - \mathbf{N}_a \otimes \mathbf{N}_a(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{N}_a - \mathbf{N}_a(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{N}_a \otimes \mathbf{N}_a]\\ &\quad \times \mathrm{vec}((\mathbf{Y} - \mathbf{X}\hat{\beta}_{[a]})(\mathbf{Y} - \mathbf{X}\hat{\beta}_{[a]})') + o(\eta_a^2).\end{aligned} \tag{C.10}$$

Note that

$$\begin{aligned}\hat{\beta}_{[a]} &= \tilde{\beta}_{[a]} - (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{M}_a'(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{M}_a\hat{\mathbf{r}} + o(\eta_a^2)\\ &= \tilde{\beta}_{[a]} - \hat{\beta}_{[a]}^{(1)} + o(\eta_a^2),\end{aligned} \tag{C.11}$$

where $\hat{\beta}^{(1)}_{[a]} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{M}'_a(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{M}_a\hat{\mathbf{r}}$. Thus

$$\mathbf{Y} - \mathbf{X}\hat{\beta}_{[a]} = \hat{\mathbf{e}}_{[a]0} + \mathbf{X}\hat{\beta}^{(1)}_{[a]}, \tag{C.12}$$

where $\hat{\mathbf{e}}_{[a]0} = \mathbf{Y} - \mathbf{X}\tilde{\beta}_{[a]}$. Substituting (C.9) and (C.10) into the second equation of (3.1) and using (C.12), we have that

$$\begin{aligned}
\hat{\theta}_{[a]} &= (\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^*)^{-1}\mathbf{Z}^{*'}[\mathbf{N}_a \otimes \mathbf{N}_a - \mathbf{N}_a \otimes \mathbf{N}_a(\hat{\mathbf{V}}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{N}_a \\
&\quad - \mathbf{N}_a(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{N}_a \otimes \mathbf{N}_a]\mathrm{vec}((\mathbf{Y} - \mathbf{X}\hat{\beta}_{[a]})(\mathbf{Y} - \mathbf{X}\hat{\beta}_{[a]})') \\
&\quad + (\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^*)^{-1}\mathbf{Z}^{*'}[\mathbf{N}_a \otimes \mathbf{N}_a(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{N}_a \\
&\quad + \mathbf{N}_a(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{N}_a \otimes \mathbf{N}_a]\mathbf{Z}^*(\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^*)^{-1}\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a) \\
&\quad \times \mathrm{vec}((\mathbf{Y} - \mathbf{X}\hat{\beta}_{[a]})(\mathbf{Y} - \mathbf{X}\hat{\beta}_{[a]})') + o(\eta_a^2) \\
&= (\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^*)^{-1}\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathrm{vec}((\mathbf{Y} - \mathbf{X}\hat{\beta}_{[a]})(\mathbf{Y} - \mathbf{X}\hat{\beta}_{[a]})') \\
&\quad - (\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^*)^{-1}\mathbf{Z}^{*'}[\mathbf{N}_a \otimes \mathbf{N}_a(\hat{\mathbf{V}}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{N}_a \\
&\quad + \mathbf{N}_a(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{N}_a \otimes \mathbf{N}_a][\mathbf{I} - \mathbf{Z}^*(\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^*)^{-1}\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)] \\
&\quad \times \mathrm{vec}((\mathbf{Y} - \mathbf{X}\tilde{\beta}_{[a]})(\mathbf{Y} - \mathbf{X}\tilde{\beta}_{[a]})') + o(\eta_a^2) \\
&= (\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^*)^{-1}\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathrm{vec}((\hat{\mathbf{e}}_{[a]0} + \mathbf{X}\hat{\beta}^{(1)}_{[a]})(\hat{\mathbf{e}}_{[a]0} + \mathbf{X}\hat{\beta}^{(1)}_{[a]})') \\
&\quad - (\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^*)^{-1}\mathbf{Z}^{*'}[\mathbf{N}_a \otimes \mathbf{N}_a(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{N}_a \\
&\quad + \mathbf{N}_a(\hat{\mathbf{V}}_{[a]} - \hat{\mathbf{V}})\mathbf{N}_a \otimes \mathbf{N}_a][\mathrm{vec}(\hat{\mathbf{e}}_{[a]0}\hat{\mathbf{e}}'_{[a]0}) - \mathbf{Z}^*\hat{\theta}^*_{[a]}] + o(\eta_a^2) \\
&= \hat{\theta}^*_{[a]} + (\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^*)^{-1}\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathrm{vec}(\hat{\mathbf{e}}_{[a]0}\hat{\beta}^{(1)'}_{[a]}\mathbf{X}' + \mathbf{X}\hat{\beta}^{(1)}_{[a]}\hat{\mathbf{e}}_{[a]0}) \\
&\quad - (\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^*)^{-1}\mathbf{Z}^{*'}[\mathbf{N}_a \otimes \mathbf{N}_a(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{N}_a \\
&\quad + \mathbf{N}_a(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{N}_a \otimes \mathbf{N}_a][\mathrm{vec}(\hat{\mathbf{e}}_{[a]0}\hat{\mathbf{e}}'_{[a]0}) - \mathbf{Z}^*\hat{\theta}^*_{[a]}] + o(\eta_a^2),
\end{aligned}$$

where $\hat{\theta}^*_{[a]} = (\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^*)^{-1}\mathbf{Z}^{*'}\mathrm{vec}(\mathbf{M}_a\hat{\mathbf{r}}\hat{\mathbf{r}}'\mathbf{M}'_a)$. The above used the fact that $\hat{\beta}^{(1)}_{[a]} = o(\eta_a)$ and $\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}) = o(\eta_a)$. Now note that $\mathbf{N}_a\hat{\mathbf{e}}_{[a]0} = \mathbf{N}_a(\mathbf{Y} - \mathbf{X}\tilde{\beta}_{[a]}) = \mathbf{M}_a\hat{\mathbf{r}}$ (see (C.6)) and

$$\begin{aligned}
\mathbf{N}_a\mathbf{X}\hat{\beta}^{(1)}_{[a]} &= (\mathbf{I}_N - \hat{\mathbf{V}}^{-1}\mathbf{D}_a(\mathbf{D}'_a\hat{\mathbf{V}}^{-1}\mathbf{D}_a)^{-1}\mathbf{D}'_a)\hat{\mathbf{P}}\mathbf{M}'_a(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{M}_a\hat{\mathbf{r}} \\
&= (\mathbf{N}_a - \mathbf{M}_a\hat{\mathbf{Q}})(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{M}_a\hat{\mathbf{r}},
\end{aligned} \tag{C.13}$$

we immediately have

$$\begin{aligned}
\hat{\theta}_{[a]} &= \hat{\theta}^*_{[a]} + (\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^*)^{-1}\mathbf{Z}^{*'}\mathrm{vec}[\mathbf{M}_a\hat{\mathbf{r}}\hat{\mathbf{r}}'\mathbf{M}'_a(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))(\mathbf{N}_a - \mathbf{M}_a\hat{\mathbf{Q}})' \\
&\quad + (\mathbf{N}_a - \mathbf{M}_a\hat{\mathbf{Q}})(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{M}_a\hat{\mathbf{r}}\hat{\mathbf{r}}'\mathbf{M}'_a - \mathbf{N}_a(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{M}_a\hat{\mathbf{r}}\hat{\mathbf{r}}'\mathbf{M}'_a \\
&\quad - \mathbf{M}_a\hat{\mathbf{r}}\hat{\mathbf{r}}'\mathbf{M}'_a(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{N}_a + \mathbf{N}_a(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{N}_a\mathbf{V}(\hat{\theta}^*_{[a]})\mathbf{N}_a \\
&\quad + \mathbf{N}_a\mathbf{V}(\hat{\theta}^*_{[a]})\mathbf{N}_a(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta}))\mathbf{N}_a] + o(\eta_a^2) \\
&= \hat{\theta}^*_{[a]} + (\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^*)^{-1}\mathbf{Z}^{*'}[\mathbf{N}_a \otimes \mathbf{N}_a\mathbf{V}(\hat{\theta}^*_{[a]})\mathbf{N}_a + \mathbf{N}_a\mathbf{V}(\hat{\theta}^*_{[a]})\mathbf{N}_a \otimes \mathbf{N}_a \\
&\quad - \mathbf{M}_a\hat{\mathbf{Q}} \otimes \mathbf{M}_a\hat{\mathbf{r}}\hat{\mathbf{r}}'\mathbf{M}'_a - \mathbf{M}_a\hat{\mathbf{r}}\hat{\mathbf{r}}'\mathbf{M}'_a \otimes \mathbf{M}_a\hat{\mathbf{Q}}]\mathrm{vec}(\mathbf{V}(\hat{\theta}_{[a]}) - \mathbf{V}(\hat{\theta})) + o(\eta_a^2) \\
&= \hat{\theta} + \hat{\theta}^*_{[a]} - \hat{\theta} + (\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^*)^{-1}\mathbf{Z}^{*'}[\mathbf{N}_a \otimes \mathbf{N}_a\mathbf{V}(\hat{\theta}^*_{[a]})\mathbf{N}_a + \mathbf{N}_a\mathbf{V}(\hat{\theta}^*_{[a]})\mathbf{N}_a \otimes \mathbf{N}_a \\
&\quad - \mathbf{M}_a\hat{\mathbf{Q}} \otimes \mathbf{M}_a\hat{\mathbf{r}}\hat{\mathbf{r}}'\mathbf{M}'_a - \mathbf{M}_a\hat{\mathbf{r}}\hat{\mathbf{r}}'\mathbf{M}'_a \otimes \mathbf{M}_a\hat{\mathbf{Q}}]\mathbf{Z}^*(\hat{\theta}_{[a]} - \hat{\theta}) + o(\eta_a^2).
\end{aligned} \tag{C.14}$$

Multiplying both sides of (C.14) by $\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^*$, we have

$$(\mathbf{Z}^{*'}\mathbf{W}_a\mathbf{Z}^*)(\hat{\theta}_{[a]} - \hat{\theta}) = \mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^*(\hat{\theta}^*_{[a]} - \hat{\theta}) + o(\eta_a^2), \tag{C.15}$$

where

$$\begin{aligned}\mathbf{W}_a &= \mathbf{N}_a \otimes \mathbf{N}_a - \mathbf{N}_a \otimes \mathbf{N}_a\mathbf{V}(\hat{\theta}^*_{[a]})\mathbf{N}_a - \mathbf{N}_a\mathbf{V}(\hat{\theta}^*_{[a]})\mathbf{N}_a \otimes \mathbf{N}_a \\ &\quad + \mathbf{M}_a\hat{\mathbf{Q}} \otimes \mathbf{M}_a\hat{\mathbf{r}}\hat{\mathbf{r}}'\mathbf{M}_a' + \mathbf{M}_a\hat{\mathbf{r}}\hat{\mathbf{r}}'\mathbf{M}_a' \otimes \mathbf{M}_a\hat{\mathbf{Q}}.\end{aligned}$$

Since

$$\mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^*\hat{\theta}^*_{[a]} = \mathbf{Z}^{*'}\mathrm{vec}(\mathbf{M}_a\hat{\mathbf{r}}\hat{\mathbf{r}}'\mathbf{M}_a'),\ \ \mathbf{N}_a\hat{\mathbf{V}}\mathbf{N}_a = \mathbf{N}_a,$$

from (C.15), we have

$$\begin{aligned}(\mathbf{Z}^{*'}\mathbf{W}_a\mathbf{Z}^*)(\hat{\theta}_{[a]} - \hat{\theta}) &= \mathbf{Z}^{*'}\mathrm{vec}(\mathbf{M}_a\hat{\mathbf{r}}\hat{\mathbf{r}}'\mathbf{M}_a') - \mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathbf{Z}^*\hat{\theta} + o(\eta_a^2) \\ &= \mathbf{Z}^{*'}\mathrm{vec}(\mathbf{M}_a\hat{\mathbf{r}}\hat{\mathbf{r}}'\mathbf{M}_a') - \mathbf{Z}^{*'}(\mathbf{N}_a \otimes \mathbf{N}_a)\mathrm{vec}(\mathbf{V}(\hat{\theta})) + o(\eta_a^2) \\ &= \mathbf{Z}^{*'}[\mathrm{vec}(\mathbf{M}_a\hat{\mathbf{r}}\hat{\mathbf{r}}'\mathbf{M}_a') - \mathrm{vec}(\mathbf{N}_a)) + o(\eta_a^2) \\ &= \mathbf{Z}^{*'}\mathrm{vec}(\mathbf{M}_a\hat{\mathbf{r}}\hat{\mathbf{r}}'\mathbf{M}_a' - \mathbf{N}_a) + o(\eta_a^2).\end{aligned} \tag{C.16}$$

Discarding $o(\eta_a^2)$ in (C.7) and (C.16), and using $\bar{\beta}_{[a]}$ and $\bar{\theta}_{[a]}$ to denote the corresponding approximations, we have the result in Theorem 3.2.

References

[1] A. Atkinson, M. Riani, Robust Diagnostic Regression Analysis, Springer, New York, 2000.

[2] A.C. Atkinson, Discussion on Some algebra and geometry for hierarchical models, applied to diagnostics (by J. S. Hodges), Journal of the Royal Statistical Society, B 60 (1998) 521–523.

[3] M. Banerjee, E.W. Frees, Influence diagnostics for linear longitudinal models, Journal of the American Statistical Association 92 (1997) 999–1005.

[4] D.A. Belsley, E. Kuh, R.E. Welsch, Regression Diagnostics: Identifying Influential Data and Sources of Collinearity, John Wiley, New York, 1980.

[5] S. Chatterjee, A.S. Hadi, Sensitivity Analysis in Linear Regression, John Wiley, New York, 1988.

[6] R. Christensen, L.M. Pearson, W.O. Johnson, Case-deletion diagnostics for mixed models, Technometrics 34 (1992) 38–45.

[7] R.D. Cook, S. Weisberg, Residuals and Influence in Regression, Chapman and Hall, New York, 1982.

[8] R.D. Cook, S. Weisberg, Applied Regression Including Computing and Graphics, Wiley, New York, 1999.

[9] H. Goldstein, Multilevel mixed linear model analysis using iterative generalized least squares, Biometrika 73 (1986) 43–56.

[10] H. Goldstein, Restricted unbiased iterative generalized least squares estimation, Biometrika 76 (1989) 622–623.

[11] H. Goldstein, Multilevel Statistical Models, second edition, Halsted Press, New York, 1995.

[12] J. Haslett, A simple derivation of deletion diagnostic results for the general linear model with correlated errors, Journal of the Royal Statistical Society, B 61 (1999) 603–609.

[13] J. Haslett, D. Dillane, Application of 'delete=replace' to deletion diagnostics for variance component estimation in the linear mixed model, Journal of the Royal Statistical Society, B 66 (2004) 131–141.

[14] J.S. Hodges, Some algebra and geometry for hierarchical models, applied to diagnostics, Journal of the Royal Statistical Society, B 60 (1998) 497–536.

[15] R.J. Martin, Leverage, influence and residuals in regression models when observations are correlated, Communication in Statistics — Theory and Method 21 (1992) 1183–1212.

[16] L. Shi, M.M. Ojeda, Local influence in multilevel regression for growth curve, Journal of Multivariate Analysis 91 (2004) 282–304.

[17] E.F. Vonesh, V.M. Chinchilli, Linear and Nonlinear Models for the Analysis of Repeated Measurements, Marcel Dekker, New York, 1997.