



Deletion measures for generalized linear mixed effects models

Liang Xu^a, Sik-Yum Lee^{b,*}, Wai-Yin Poon^b

^a*Department of Mathematics, Southeast University, Nanjing, China*

^b*Department of Statistics, The Chinese University of Hong Kong, Hong Kong*

Received 9 May 2005; received in revised form 16 November 2005; accepted 16 November 2005

Available online 13 December 2005

Abstract

Generalized linear mixed models (GLMMs) have wide applications in practice. Similar to other data analyses, the identification of influential observations that may be potential outliers is an important step beyond estimation in GLMMs. Since the pioneering work of Cook in 1977, deletion measures have been applied to many statistical models for identifying influential observations. However, as this well-known approach is based on the observed-data likelihood, it is very difficult to apply it to developing diagnostic measures for GLMMs due to the complexity of the observed-data likelihood that involves multidimensional integrals. The objective of this article is to develop diagnostic measures for identifying influential observations. Deletion measures are developed on the basis of the conditional expectation of the complete-data log-likelihood at the E-step of a stochastic approximation Markov chain Monte Carlo algorithm. Making use of by-products of the estimation to compute building blocks of the proposed diagnostic measures and activating appropriate approximations, the proposed methods require little additional computation. The performance of the methods is illustrated by an artificial example, a real example, and some simulation studies.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Generalized linear mixed effects model; Deletion measures; Stochastic approximation; Metropolis–Hastings algorithm

1. Introduction

Generalized linear mixed models (GLMMs) are useful extensions of linear mixed models and generalized linear models for assessing additional components of variability due to latent random effects. GLMMs have wide applications in practice, such as in accommodating the overdispersion that is observed among outcomes that have nominally binomial or Poisson distributions (Williams, 1982; Breslow, 1984), in producing shrinkage estimates in the construction of maps of small area disease rates (Clayton and Kaldor, 1987; Manton et al., 1989), and in modeling the dependence outcome variables in longitudinal, repeated measures, or cluster designs (Stiratelli et al., 1984; Zeger et al., 1988). Due to the complexity of the models and the involvement of intractable high-dimensional integrals in the observed-data likelihood function, the maximum likelihood (ML) estimation of the parameters in GLMMs is difficult in the field. Recently, several procedures have been proposed to implement the ML estimation, such as those that are based on a Monte Carlo EM (MCEM) and Newton Raphson algorithm (McCulloch, 1997; Kuk and Cheng, 1999), an automated MCEM algorithm (Booth and Hobert, 1999), and a stochastic approximation algorithm with Markov chain Monte Carlo (SA–MCMC) method (Zhu and Lee, 2002).

* Corresponding author. Tel.: +86 2552388486; fax: +852 26037251.

E-mail addresses: xtree28@hotmail.com (L. Xu), sylee@sparc2.sta.cuhk.edu.hk (S.-Y. Lee).

On the other hand, influence analysis has been regarded as a crucial component in a thorough statistical analysis. The results of influence analysis can be used to identify implicit problems in a data set, and can be used to judge whether a conclusion is misleading or not. The identification of influential observations has received much attention: see [Cook and Weisberg \(1982\)](#), and [Chatterjee and Hadi \(1988\)](#), among others. However, although the analysis of many other statistical models indicates that the results obtained from ML estimation can be very sensitive to several extreme observations, not many results have been published for detecting influential observations in statistical analysis for the GLMMs.

There are two major approaches for detecting influential observations. The local influence approach ([Cook, 1986](#)) develops diagnostic measures by using the curvature of the influence graph of an appropriate function. For instance, for detecting influential observations for the GLMMs, [Ouwens et al. \(2001\)](#) worked with the log-likelihood function, and [Zhu and Lee \(2003\)](#) worked with a Q-displacement function. Applications of the Q-displacement function to other models include [Lee and Tang \(2004\)](#), [Lu and Song \(2005\)](#), [Lee et al. \(2006\)](#), and [Song and Lee \(2004\)](#). On the other hand, the deletion approach ([Cook, 1977](#)) is an intuitively appealing method. It develops a diagnostic measure by assessing a chosen quantity's change that is induced by the exclusion of individual data points from an analysis. Due to the pioneering work of [Cook \(1977\)](#), see also [Cook and Weisberg \(1982\)](#), deletion diagnostics such as Cook's distance or the likelihood distance have been applied to many statistical models: see for example [Banerjee and Frees \(1997\)](#), [Davison and Tsai \(1992\)](#), [Tan et al. \(2001\)](#), [Fung et al. \(2002\)](#), and [Haslett and Dillane \(2004\)](#), among others. However, the development of deletion diagnostic measures for GLMMs on the basis of Cook's approach is rather difficult (see, [Davidian and Giltinan, 1995](#); [Xiang et al., 2002](#)). The reason lies in that a GLMM may consist of latent variables with an observed-data likelihood function that involves intractable integrals, and hence the objective function and second derivatives that are usually involved in the basic building blocks of the deletion diagnostic measures are difficult to evaluate. Moreover, a major consideration in the development of diagnostic measures is the compatibility of the influence analysis procedure and the estimation method. In view of this, the objective of this article is to develop diagnostic measures to accompany the recently developed EM ([Dempster et al., 1977](#)) type algorithms, for example the SA-MCMC algorithm ([Zhu and Lee, 2002](#)).

In the EM approach, the ML estimation is formulated as a missing-data problem by treating the latent random effects as hypothetical missing data. For general statistical models with missing data, [Zhu et al. \(2001\)](#) developed a unified approach to obtaining deletion diagnostic measures. The approach focuses on a Q-displacement function that is based on the conditional expectation of the complete-data log-likelihood at the E-step of the EM algorithm rather than on the observed-data likelihood function as used by [Cook \(1977\)](#). It has been applied to construct deletion measures for various statistical models, see, for example [Lee and Lu \(2003\)](#), [Lee and Xu \(2003, 2004\)](#). In the current study, we apply this approach to develop deletion diagnostic measures for GLMMs. Again, the deletion measures are derived on the basis of the conditional expectations of the complete-data log-likelihood. These conditional expectations are not in closed form, and are approximated by the means of a sufficiently large number of observations that are simulated by the Metropolis–Hastings (MH) algorithm ([Metropolis et al., 1953](#); [Hastings, 1970](#)). As a set of observations has been generated during the process of ML estimation, the computation of the building blocks for the diagnostic measures is not difficult. In fact, the proposed procedure does not involve the computation of any high-dimensional complicated integrals, and requires only limited additional computation.

This article is organized as follows. Section 2 defines the GLMMs and briefly describes the ML estimation method that is based on the SA-MCMC algorithm proposed by [Zhu and Lee \(2002\)](#). The deletion measures are developed in Section 3. Section 4 presents results that are obtained from an artificial example, a real example, and simulation studies to illustrate performance of the deletion measures. A discussion is given in Section 5, and technical details are presented in the Appendices.

2. Model specification

Consider a data set with observations that are composed of a response y_{ij} and covariate vectors \mathbf{x}_{ij} ($p_1 \times 1$) and \mathbf{z}_{ij} ($p_2 \times 1$), where $j = 1, \dots, n_i$ represents an observation within the cluster $i = 1, \dots, N$. For example, a subject can be considered as a cluster, and repeated measurements for this subject i can be obtained at a total of n_i different time points that are represented by $j = 1, \dots, n_i$. It is assumed that conditional on a $p_2 \times 1$ vector \mathbf{b}_i of unobservable random variables, y_{ij} follows an exponential family distribution of the form ([McCullagh and Nelder, 1989](#))

$$p(y_{ij} | \mathbf{b}_i) = \exp[\phi \{y_{ij}\theta_{ij} - a(\theta_{ij})\} + c(y_{ij}, \phi)]. \quad (1)$$

The conditional mean and variance of y_{ij} given \mathbf{b}_i are $E(y_{ij} | \mathbf{b}_i) = \mu_{ij} = \dot{a}(\theta_{ij})$ and $\text{var}(y_{ij} | \mathbf{b}_i) = \ddot{a}(\theta_{ij})/\phi$, respectively, where $\dot{a}(u) = da/du$ and $\ddot{a}(u) = d^2a/du^2$. The GLMM is defined by (1) and the following systematic component:

$$g(\mu_{ij}) = \eta_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i \quad \text{or} \quad \theta_{ij} = k(\mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i), \quad (2)$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{p_1})^T$ is the vector that consists of the regression coefficients, $k(\cdot) = \dot{a}^{-1}(g^{-1}(\cdot))$ is a composite function of \dot{a}^{-1} and g^{-1} , $g(\cdot)$ is a known continuously differentiable function, and $\dot{a}^{-1}(\cdot)$ and $g^{-1}(\cdot)$ are inverse functions of $\dot{a}(\cdot)$ and $g(\cdot)$. The distribution of \mathbf{b}_i is assumed to be the normal $N[0, \boldsymbol{\Sigma}]$, where $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\gamma})$ depends on a $p_3 \times 1$ vector $\boldsymbol{\gamma}$ of unknown parameters that determine the variance components.

Let $\boldsymbol{\psi} = (\phi, \boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$ be the $p \times 1$ ($p = 1 + p_1 + p_3$) vector of unknown parameters, and let \mathbf{Y}_o represent all the observations on y_{ij} . The observed-data likelihood function for $\boldsymbol{\psi}$ has the form

$$p_o(\mathbf{Y}_o | \boldsymbol{\psi}) \propto \prod_{i=1}^N \int \prod_{j=1}^{n_i} p(y_{ij} | \mathbf{b}_i) |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{b}_i^T \boldsymbol{\Sigma}^{-1} \mathbf{b}_i\right) d\mathbf{b}_i. \quad (3)$$

The integral that is involved in $p_o(\mathbf{Y}_o | \boldsymbol{\psi})$ usually does not have an analytic solution, and its dimension is p_2 ; that is, the dimension of the random effect \mathbf{b}_i . Consequently, this function is usually intractable. We address this difficulty by treating $\{\mathbf{b}_i; i = 1, \dots, N\}$ as the missing-data set \mathbf{Y}_m , and formulating the problem as a missing value problem. Let $\mathbf{Y}_c = (\mathbf{Y}_o, \mathbf{Y}_m)$ denote the complete-data set, the complete-data log-likelihood can be written as

$$L_c(\boldsymbol{\psi} | \mathbf{Y}_c) = \sum_{i=1}^N \left(\sum_{j=1}^{n_i} [\phi \{y_{ij} \theta_{ij} - a(\theta_{ij})\} + c(y_{ij}, \phi)] - \frac{1}{2} \mathbf{b}_i^T \boldsymbol{\Sigma}^{-1} \mathbf{b}_i - \frac{1}{2} \log |\boldsymbol{\Sigma}| \right). \quad (4)$$

This complete-data log-likelihood function is relatively simple when compared with the observed-data likelihood function.

Based on this complete-data log-likelihood, a stochastic approximation (SA) and the Markov chain Monte Carlo (MCMC) method (Gu and Kong, 1998) is applied to find the ML estimates of the model parameters. This SA–MCMC algorithm requires the simulation of a sample of observations from the distribution $[\mathbf{Y}_m | \mathbf{Y}_o, \boldsymbol{\psi}]$, which is the conditional distribution given the current parameter value $\boldsymbol{\psi}$, and the observed data \mathbf{Y}_o . Hence, it is similar to the E-step of the MCEM type algorithm (Wei and Tanner, 1990). In contrast to an MCEM algorithm, the SA–MCMC algorithm applies the idea of stochastic approximation to update $\boldsymbol{\psi}$ by utilizing some matrices that involve the first and the second partial derivatives of $L_c(\boldsymbol{\psi} | \mathbf{Y}_c)$. Please see Gu and Kong (1998) and Zhu and Lee (2002) for details.

It should be noted that at the final iteration of the SA–MCMC or the MCEM algorithm, a sufficiently large sample $\{\mathbf{Y}_m^t : t = 1, \dots, T\}$ from $[\mathbf{Y}_m | \mathbf{Y}_o, \hat{\boldsymbol{\psi}}]$ is generated, where the ML estimate $\hat{\boldsymbol{\psi}}$ is the maximizer of the following Q-function:

$$Q(\boldsymbol{\psi} | \hat{\boldsymbol{\psi}}) = E\{L_c(\boldsymbol{\psi} | \mathbf{Y}_c) | \mathbf{Y}_o, \hat{\boldsymbol{\psi}}\}, \quad (5)$$

and the expectation is taken with respect to the conditional distribution of \mathbf{Y}_m given \mathbf{Y}_o and $\hat{\boldsymbol{\psi}}$. This generated sample can be used to create the building blocks of the diagnostic measures. In Section 3, we will provide the details of developing deletion diagnostic measures on the basis of the Q-function in (5).

3. Deletion measures

Deletion is a common approach to studying the effect of dropping some observations in a data set from an analysis. Due to the data structure of the GLMMs, two kinds of deletions are of interest. One is the deletion of individual observations within a cluster to judge how particular observations are likely to influence the analysis. The other is to delete all observations in cluster i , namely the vector $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})^T$, to study the overall impact of the cluster. To develop the measures, we will use specific notations and definitions. Let M_i be a set that collects selected observations which are to be examined for their influence. A quantity with a subscript “ $[M_i]$ ” is used to indicate that the quantity

is obtained from an analysis with all observations in the set M_i excluded from analysis. For example, if M_i contains all of the observations in the i th cluster, then $\mathbf{Y}_{o[M_i]}$ is the observed data with all of the observations for the i th cluster deleted from an analysis. Diagnostic measures for the two kinds of deletions can both be developed in a way to be described as follows with an appropriately chosen M_i .

To assess the influence of the observations in the set M_i on the ML estimate $\hat{\psi}$, we compare the difference between $\hat{\psi}_{[M_i]}$ and $\hat{\psi}$. If $\hat{\psi}_{[M_i]}$ is far from $\hat{\psi}$, then the observations in M_i are regarded as being influential because their influence on the ML estimate is substantial. A metric for measuring the distance between $\hat{\psi}_{[M_i]}$ and $\hat{\psi}$ is given by

$$GD_{[M_i]} = (\hat{\psi} - \hat{\psi}_{[M_i]})^T \{-\ddot{Q}(\hat{\psi})\} (\hat{\psi} - \hat{\psi}_{[M_i]}), \quad (6)$$

where $\ddot{Q}(\hat{\psi}) = \partial^2 Q(\hat{\psi} | \hat{\psi}) / \partial \hat{\psi} \partial \hat{\psi}^T |_{\hat{\psi}=\hat{\psi}}$. Another measure for the influence of the subset M_i is the following Q-distance function that is similar to the likelihood distance $LD_{[M_i]}$ (Cook and Weisberg, 1982),

$$QD_{[M_i]} = 2 \{Q(\hat{\psi} | \hat{\psi}) - Q(\hat{\psi}_{[M_i]} | \hat{\psi})\}. \quad (7)$$

In effect, a classical metric that is defined by the distance of the observed-data likelihood $p_o(\mathbf{Y}_o | \hat{\psi})$ and $p_o(\mathbf{Y}_o | \hat{\psi}_{[M_i]})$ is given by

$$\begin{aligned} LD_{[M_i]} &= 2 [\log p_o(\mathbf{Y}_o | \hat{\psi}) - \log p_o(\mathbf{Y}_o | \hat{\psi}_{[M_i]})] \\ &= -2 \log \frac{p_o(\mathbf{Y}_o | \hat{\psi}_{[M_i]})}{p_o(\mathbf{Y}_o | \hat{\psi})}. \end{aligned} \quad (8)$$

This $LD_{[M_i]}$ provides an indirect measure of the impact of M_i to the observed-data log-likelihood. Based on the following lemma, $LD_{[M_i]}$ in (8) can be expressed as the corresponding expectation of the complete-data likelihood, that is

$$LD_{[M_i]} = -2 \log \frac{p_o(\mathbf{Y}_o | \hat{\psi}_{[M_i]})}{p_o(\mathbf{Y}_o | \hat{\psi})} = -2 \log E \left\{ \frac{p_c(\mathbf{Y}_c | \hat{\psi}_{[M_i]})}{p_c(\mathbf{Y}_c | \hat{\psi})} | \mathbf{Y}_o, \hat{\psi} \right\}. \quad (9)$$

This relationship can be used to compute diagnostic measures.

Lemma. Let $p_o(\mathbf{Y}_o | \hat{\psi})$ and $p_c(\mathbf{Y}_c | \hat{\psi})$, respectively, denote the observed-data likelihood and the complete-data likelihood evaluated at $\hat{\psi}$. Then

$$\frac{p_o(\mathbf{Y}_o | \hat{\psi}_{[M_i]})}{p_o(\mathbf{Y}_o | \hat{\psi})} = E \left\{ \frac{p_c(\mathbf{Y}_c | \hat{\psi}_{[M_i]})}{p_c(\mathbf{Y}_c | \hat{\psi})} | \mathbf{Y}_o, \hat{\psi} \right\}. \quad (10)$$

The proof of this lemma is given in Appendix A.

To compute the diagnostic measures $GD_{[M_i]}$, $QD_{[M_i]}$, and $LD_{[M_i]}$, it is necessary to compute $\hat{\psi}_{[M_i]}$. We can use the SA-MCMC algorithm with $\hat{\psi}$ as the starting value to obtain $\hat{\psi}_{[M_i]}$. However, to conduct a thorough influence analysis, it may be necessary to compute $\hat{\psi}_{[M_i]}$ for various sets of M_i , and the computation can be quite heavy when N is large. To reduce the burden, the following approximation $\hat{\psi}_{[M_i]}^1$ (Zhu et al., 2001), which is similar to one-step approximation (Cook and Weisberg, 1982), is used

$$\hat{\psi}_{[M_i]}^1 = \hat{\psi} + \{-\ddot{Q}(\hat{\psi} | \hat{\psi})\}^{-1} \dot{Q}_{[M_i]}(\hat{\psi} | \hat{\psi}), \quad (11)$$

where $\dot{Q}_{[M_i]}(\hat{\psi} | \hat{\psi}) = \partial Q_{[M_i]}(\hat{\psi} | \hat{\psi}) / \partial \hat{\psi} |_{\hat{\psi}=\hat{\psi}}$. Hence, $\hat{\psi}_{[M_i]}^1$ depends on the available ML estimate $\hat{\psi}$, and it is not necessary to conduct the estimation again. Substituting (11) into (6), (7), and (9), we obtain the following approximations $LD_{[M_i]}^1$, $QD_{[M_i]}^1$, and $GD_{[M_i]}^1$ of $LD_{[M_i]}$, $QD_{[M_i]}$, and $GD_{[M_i]}$, respectively:

$$LD_{[M_i]}^1 = -2 \log E \left\{ \frac{p_c(\mathbf{Y}_c | \hat{\psi}_{[M_i]}^1)}{p_c(\mathbf{Y}_c | \hat{\psi})} | \mathbf{Y}_o, \hat{\psi} \right\}, \quad (12)$$

$$QD_{[M_i]}^1 = 2 \left\{ Q(\psi|\psi) - Q(\psi_{[M_i]}^1|\psi) \right\}, \quad (13)$$

$$GD_{[M_i]}^1 = \dot{Q}_{[M_i]}(\psi|\psi)^T \{-\ddot{Q}(\psi|\psi)\}^{-1} \dot{Q}_{[M_i]}(\psi|\psi). \quad (14)$$

As Cook and Weisberg (1982, p. 182) point out in the context of a regression model, it should be noted that: (i) for cases that are influential, the accuracy of one-step approximation is likely to be lower; and (ii) an accurate approximation to $\psi_{[M_i]}$ will not be needed as long as $\psi - \psi_{[M_i]}^1$ is sufficiently large to draw attention for further consideration.

In general, it has been shown by Zhu et al. (2001) that under mild conditions, deletion measures $GD_{[M_i]}$ and $QD_{[M_i]}$, as well as their approximations $GD_{[M_i]}^1$ and $QD_{[M_i]}^1$, have close asymptotic relationships and similar nice features. For example, the confidence regions corresponding to $GD_{[M_i]}$ and $QD_{[M_i]}$ are defined by the same distribution. In this paper, we present $GD_{[M_i]}^1$ and $QD_{[M_i]}^1$ deletion measures together with $LD_{[M_i]}^1$ for completeness and for cross-validation.

To compute the deletion measures, we need to calculate $Q(\psi|\psi)$, $Q(\psi_{[M_i]}^1|\psi)$, $\dot{Q}_{[M_i]}(\psi|\psi)$, $\ddot{Q}(\psi|\psi)$, or $LD_{[M_i]}^1$, as given in (12)–(14). All of these quantities have no explicit forms in the context of the GLMM model. We solve this difficulty via Monte Carlo integration. Let $\{\mathbf{Y}_m^t, t = 1, \dots, T\}$ be a sufficiently large sample of observations that are drawn randomly from the conditional distribution $[\mathbf{Y}_m | \mathbf{Y}_o, \psi]$ by the MH algorithm as in the ML estimation. The following quantities can be approximated as

$$Q(\psi|\psi) \approx \frac{1}{T} \sum_{t=1}^T L_c(\psi | \mathbf{Y}_o, \mathbf{Y}_m^t), \quad (15)$$

$$Q(\psi_{[M_i]}^1|\psi) \approx \frac{1}{T} \sum_{t=1}^T L_c(\psi_{[M_i]}^1 | \mathbf{Y}_o, \mathbf{Y}_m^t), \quad (16)$$

$$\ddot{Q}(\psi|\psi) \approx \frac{1}{T} \sum_{t=1}^T \frac{\partial^2 L_c(\psi | \mathbf{Y}_o, \mathbf{Y}_m^t)}{\partial \psi \partial \psi^T} \bigg|_{\psi=\psi}, \quad (17)$$

$$LD_{[M_i]}^1 \approx -2 \log \left\{ \frac{1}{T} \sum_{t=1}^T \frac{p_c(\mathbf{Y}_o, \mathbf{Y}_m^t | \psi_{[M_i]}^1)}{p_c(\mathbf{Y}_o, \mathbf{Y}_m^t | \psi)} \right\}. \quad (18)$$

The quantity $\dot{Q}_{[M_i]}(\hat{\psi}|\hat{\psi})$ is approximated as follows:

$$\dot{Q}_{[M_i]}(\psi|\psi) \approx \frac{1}{T} \sum_{t=1}^T \frac{\partial L_c(\psi | \mathbf{Y}_{o[M_i]}, \mathbf{Y}_m^{t*})}{\partial \psi} \bigg|_{\psi=\psi}, \quad (19)$$

where $\{\mathbf{Y}_m^{t*}, t = 1, \dots, T\}$ are simulated from the conditional distribution $[\mathbf{Y}_m | \mathbf{Y}_{o[M_i]}, \hat{\psi}]$. The derivatives that are involved in the above formulas are given in Appendix B.

It is worthy to note that as the density $p(\mathbf{b}_i | \mathbf{Y}_o, \psi)$ is nonstandard and complex, simulating the observations $\{\mathbf{Y}_m^t, t = 1, \dots, T\}$ from this conditional distribution is nontrivial. This difficulty is solved by using the MH algorithm (Metropolis et al., 1953; Hastings, 1970; Geman and Geman, 1984; Tanner, 1996). To generate random observations from the target density $p(\mathbf{b}_i | \mathbf{Y}_o, \psi)$, the MH algorithm is implemented as follows: at the r th iteration with a current value $\mathbf{b}_i^{(r)}$, a new candidate \mathbf{b}_i is generated from $N[\mathbf{b}_i^{(r)}, \Omega(0)]$, and the new candidate is accepted with the following probability:

$$\min \left\{ 1, \frac{p(\mathbf{b}_i | \mathbf{Y}_o, \psi)}{p(\mathbf{b}_i^{(r)} | \mathbf{Y}_o, \psi)} \right\}, \quad (20)$$

where $\Omega(0) = \Omega(\mathbf{b}_i) |_{\mathbf{b}_i=0} = \left\{ \Sigma^{-1} + \phi \sum_{j=1}^{n_i} \ddot{a}(k(\eta_{ij})) \dot{k}^2(\eta_{ij}) \mathbf{z}_{ij} \mathbf{z}_{ij}^T \right\}^{-1} \bigg|_{\mathbf{b}_i=0}$.

4. Examples and simulation studies

Example 1. A simulated example with artificially created outliers.

The main purpose of this artificial example is to reveal the performance of the proposed deletion measures when the data set does contain outliers. We consider a GLMM model in which the responses y_{ij} is distributed according to a binomial distribution, $B(n_{ij}, p_{ij})$, with $i = 1, \dots, 100$; $j = 1, 2$, and

$$\eta_{ij} = \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \beta_1 + \beta_2 x_{ij} + u_i + v_{ij}, \quad (21)$$

where $x_{i1} = 1$, $x_{i2} = 0$, for all i ; u_i and v_{ij} are respectively distributed as $N[0, \sigma_u^2]$ and $N[0, \sigma_v^2]$. The true population values of the parameters are: $\beta_1 = -2.0$, $\beta_2 = 1.5$, $\sigma_u^2 = 0.2$, and $\sigma_v^2 = 0.25$. The sample size n_{i1} are randomly selected from the range 100–2500, and n_{i2} are randomly selected from the range 10–700. We perturb the y_{ij} in the $i = 27, 47, 67$, and 87 studies in order to make these four studies influential. The (p_{i1}, p_{i2}) corresponding to the perturbed 27th, 47th, 67th, and 87th studies are (0.8, 0.05), (0.05, 0.8), (0.8, 0.8), and (0.05, 0.05). However, from (21) and the known population values of the parameters, the values of all p_{i1} should be relatively large and values of all p_{i2} should be relatively small. As a result, we expect the 47th study to be the most influential one. Based on the perturbed data set, ML estimates $\hat{\psi} = (\hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}_u^2, \hat{\sigma}_v^2)$ obtained from the SA–MCMC algorithm are equal to $(-1.990, 1.445, 0.168, 0.471)$.

Let M_i be the set that collects the two observations ($j = 1, 2$) in the study i . To study the effectiveness of the proposed deletion measures with the approximations that are described in Section 3, we obtained $\hat{\psi}_{[M_1]}^1, \dots, \hat{\psi}_{[M_{100}]}^1$ where $\hat{\psi}_{[M_i]}^1$ is the one-step approximation obtained via (11) with the i th study deleted, and then the deletion measures. The index plots of the deletion measures $GD_{[M_i]}^1$, $QD_{[M_i]}^1$, and $LD_{[M_i]}^1$ are presented in Fig. 1. We obtained the expected result that only the 27th, 47th, 67th, and 87th studies are identified as influential. Clearly, the 47th study is identified as the most influential one by all deletion measures. Deletion measures that are obtained by deleting an observation within each study have also been computed. Now M_k denotes the k th observation, for $k = 1, \dots, 200$. The index plots of $GD_{[M_k]}^1$, $QD_{[M_k]}^1$, and $LD_{[M_k]}^1$ are displayed in Fig. 2. These index plots indicate the expected result that only the first observation of the 27th study, both observations of the 47th study, the second observation of the 67th study, and the first observation of the 87th study are influential. In summary, the above empirical findings indicate that the deletion measures have detected what they are supposed to detect, and have given no false influential cases.

Example 2. Headache data.

The data set is taken from McKnight and Van Den Eeden (1993). It is obtained from a two-treatment, multiple-period crossover trial in which the number of headaches per week was repeatedly measured for 27 patients. In the first period, usually called the run-in period, each patient received the placebo. In the other four periods, the patients received either a placebo (P) or aspartame (A) in random order, using the double-blind crossover treatment design. To wash out the effects of the treatment of the foregoing periods, the periods were separated by one day. The four possible orderings of treatments were PAPAP, PAPPAP, PPAPA, and PPAAP. Almost all periods spanned seven days, but some periods had shorter durations. The data for the 27 subjects are shown in Table 1. This data set was analyzed by Ouwers et al. (2001). They analyzed the data set with the Poisson mixed model with known dispersion parameters, and used it to illustrate their local influence analysis method of GLMM. Considering each subject as a cluster, the Poisson mixed model is given by

$$E(y_{ij} | \boldsymbol{\beta}, \mathbf{b}_i) = t_{ij} \exp(\beta_1 + \text{DrugAsp}_{ij} \beta_2 + b_{0i}), \quad (22)$$

where $i = 1, \dots, 27$, $j = 1, \dots, n_i$ with $n_i \leq 5$; t_{ij} represents the number of days in the period, $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$ is the vector of fixed regression parameters, DrugAsp_{ij} is an observable covariate that takes the value 0 or 1 to indicate whether a placebo or aspartame was given to patient i at the time point j , and b_{0i} is the random intercept for patient i . Let δ be the standard deviation of the random intercept b_{0i} , the ML estimate obtained by the SA–MCMC algorithm is

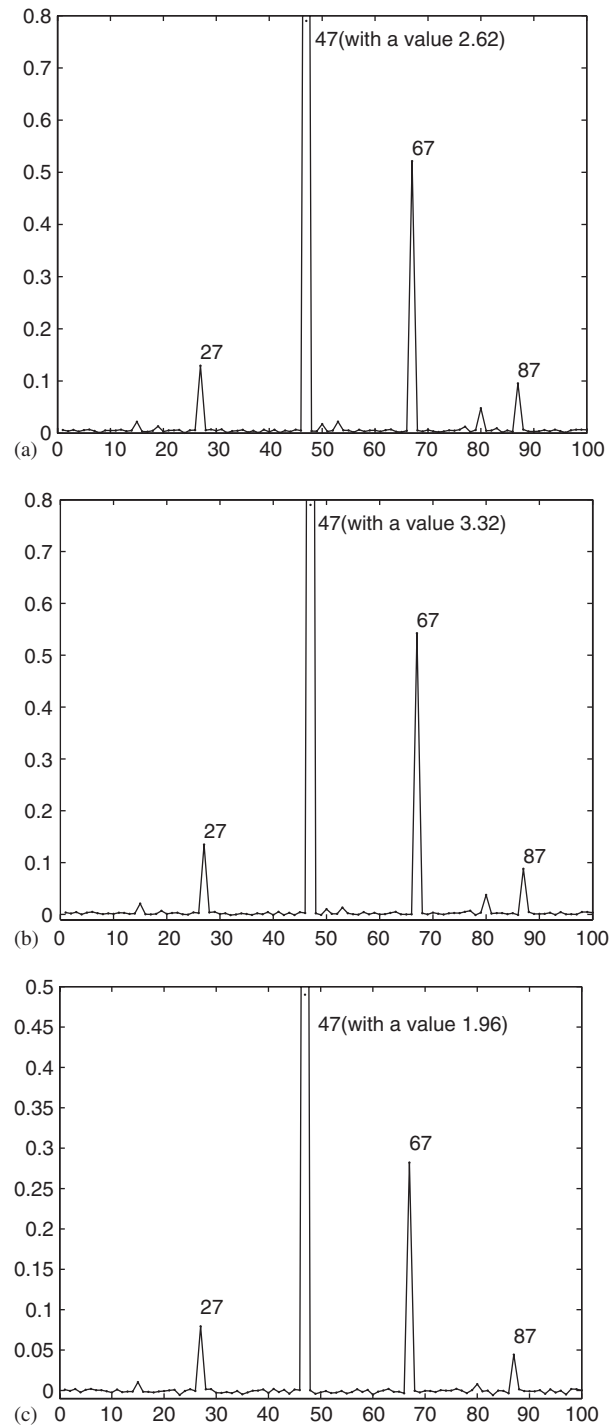


Fig. 1. Deletion measures (deletion of a study) for the simulated example. (a) $GD_{[M_i]}^1$ vs i , (b) $QD_{[M_i]}^1$ vs i , (c) $LD_{[M_i]}^1$ vs i .

$\psi^T = (\hat{\beta}_1, \hat{\beta}_2, \hat{\delta}) = (-1.717, 0.282, 0.695)$. These estimates are very close to those given in [Ouwens et al. \(2001\)](#). We also compute the proposed deletion measures to assess the influence of individual subject, as well as the observation at a specific time period.

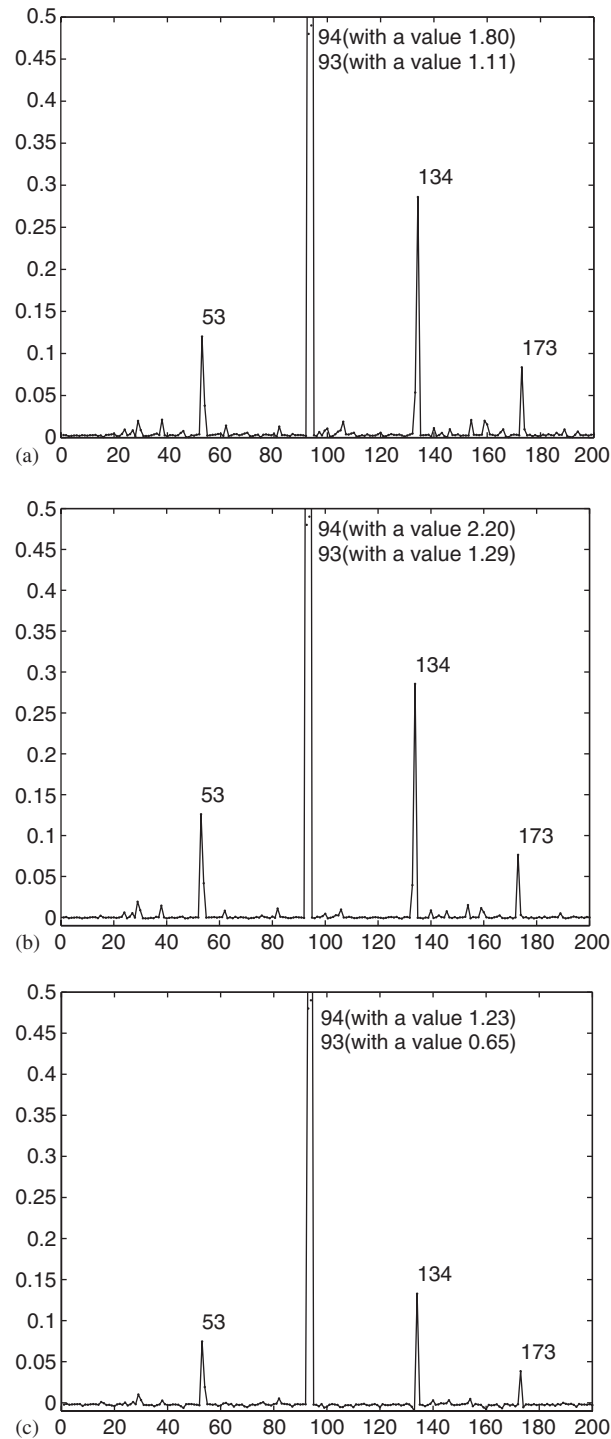


Fig. 2. Deletion measures (deletion of an observation) for the simulated example. (a) $GD^1_{[M_k]}$ vs k , (b) $QD^1_{[M_k]}$ vs k , (c) $LD^1_{[M_k]}$ vs k .

Firstly, we assess the influence of individual subject by considering M_i as the set that collects the observations in all periods for subject i . The deletion measures $GD^1_{[M_i]}$, $QD^1_{[M_i]}$, and $LD^1_{[M_i]}$ are calculated. The index plots of these measures with respect to i are presented in Fig. 3. The results for the three measures are consistent, all suggesting

Table 1

Headache data: number of days with headache (y), number of days (d) in each period, and treatment order for the 27 subjects

Treatment order	Subject number	Period									
		1		2		3		4		5	
		y	d	y	d	y	d	y	d	y	d
PAPAP	2	0	7	5	7	2	7	–	–	–	–
PAPAP	5	3	7	0	7	2	7	0	7	0	7
PAPAP	13	7	7	7	7	7	7	6	7	7	7
PAPAP	16	1	7	3	7	1	7	–	–	–	–
PAPAP	19	0	7	1	7	1	7	0	7	1	7
PAPAP	23	7	7	2	7	3	7	3	7	2	7
PAPAP	25	1	7	6	7	1	7	7	7	0	7
PAPPA	1	3	7	0	7	3	7	1	7	0	7
PAPPA	3	2	7	2	7	3	7	2	7	2	7
PAPPA	6	1	7	1	7	0	7	3	7	1	7
PAPPA	9	2	7	2	7	0	7	1	5	–	–
PAPPA	17	4	7	1	1	–	–	–	–	–	–
PAPPA	18	0	7	1	7	1	7	1	7	0	7
PAPPA	21	1	7	2	7	3	7	3	7	6	7
PAPPA	22	2	7	1	7	0	7	0	7	1	7
PPAPA	7	1	7	1	7	4	7	2	7	3	7
PPAPA	10	0	7	0	7	0	7	0	7	0	7
PPAPA	11	0	7	3	7	1	7	0	7	1	3
PPAPA	14	2	7	2	7	1	7	0	7	2	7
PPAPA	24	1	7	0	7	1	7	0	7	2	7
PPAPA	27	3	7	3	7	3	7	0	7	2	4
PPAAP	4	0	7	0	7	0	7	0	7	0	7
PPAAP	8	1	7	1	7	0	7	1	7	1	2
PPAAP	12	0	7	0	7	5	7	0	7	0	7
PPAAP	15	0	7	3	7	2	7	1	7	1	7
PPAAP	20	1	7	6	7	1	2	–	–	–	–
PPAAP	26	0	7	1	7	–	–	–	–	–	–

Source: McKnight and Van Den Eeden (1993).

that the 25th subject is the most influential. This conclusion coincides with the result of [Ouwens et al. \(2001\)](#). By examining the original data, it is found that this patient had a much longer headache time when he/she took Aspartame. These three measures also suggest that the influence of the 13th subject is relatively large. This patient, regardless of whether treated by a placebo or Aspartame, had a headache almost all of the time during the study period.

Secondly, we assess the influence of observation at specific time periods for each subject. The three proposed measures are computed by excluding the observation of subject i at each time period j in turn. The results are presented in [Fig. 4](#), which gives index plots of the measures with respect to k , where $k = 1, \dots, K$, and $K = \sum_{i=1}^{27} n_i$ with $n_i \leq 5$. In other words, each k represents a specific observation at time j for a subject i . The plots in [Fig. 4](#) show that the three measures are quite consistent, and they are relatively large when $k = 55$ ($i = 12, j = 3$), $k = 101$ ($i = 23, j = 1$), and $k = 114$ ($i = 25, j = 4$). The corresponding observations are identified as influential. These results suggest that the 25th patient is influential both on the subject and observation levels. For patients 12 and 23, observations in some periods are influential, but they are not influential as subjects. It is also found that the observations with $k = 7$ ($i = 2, j = 2$), $k = 89$ ($i = 20, j = 2$), $k = 95$ ($i = 21, j = 5$), and $k = 112$ ($i = 25, j = 2$) have relatively large influences, although the effect is less prominent. In contrast, there is no influential observation in a specific time period for patient 13, although the patient is influential at the subject level.

To illustrate the actual effect of the identified influential observations on the ML estimates, we compare the original ML estimates with those that are obtained by excluding some influential observations in the estimation. The results are presented in [Table 2](#). Subject 25 has substantial influence on the estimate of β_2 . The exclusion of this subject from the

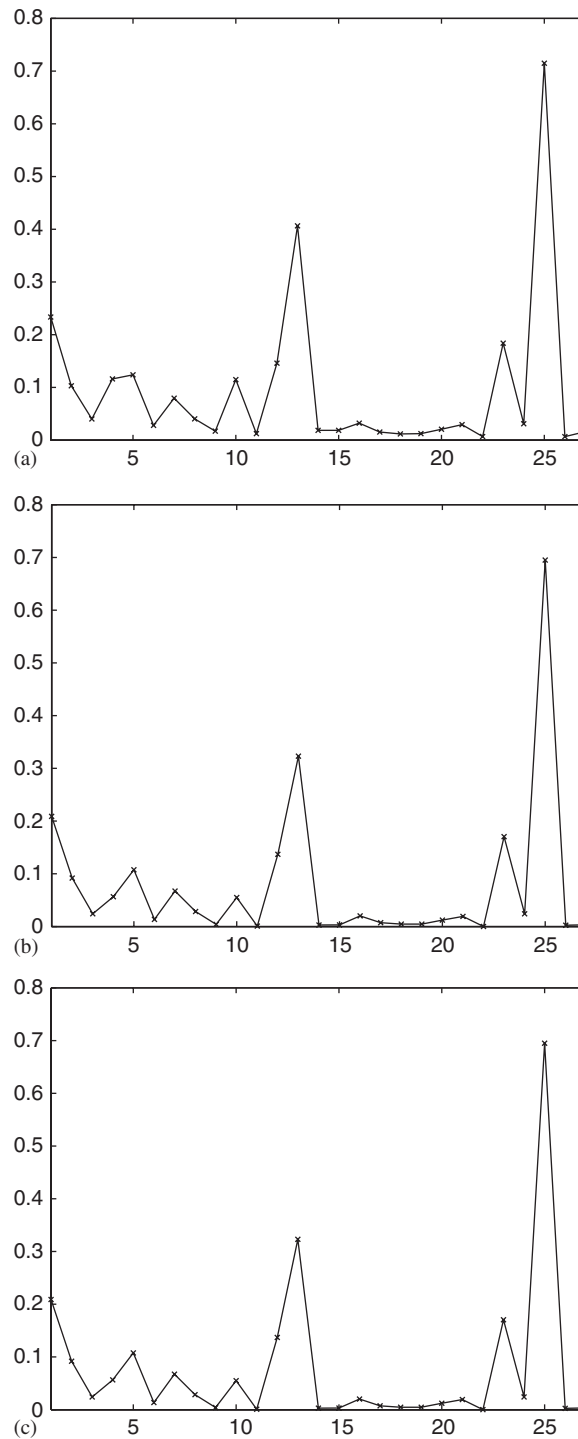


Fig. 3. Deletion measures (deletion of a subject) for the headache data. (a) $GD_{[M_i]}^1$ vs i , (b) $QD_{[M_i]}^1$ vs i , (c) $LD_{[M_i]}^1$ vs i .

estimation causes an approximately 40% change in the estimate of β_2 . Subject 13, in contrast, substantially influences the estimate of δ . The exclusion of this subject from the estimation leads to a change of 18.6% in the estimate of δ and 31.2% in the estimate of β_2 . Moreover, the exclusion of individual observations that have been identified as influential

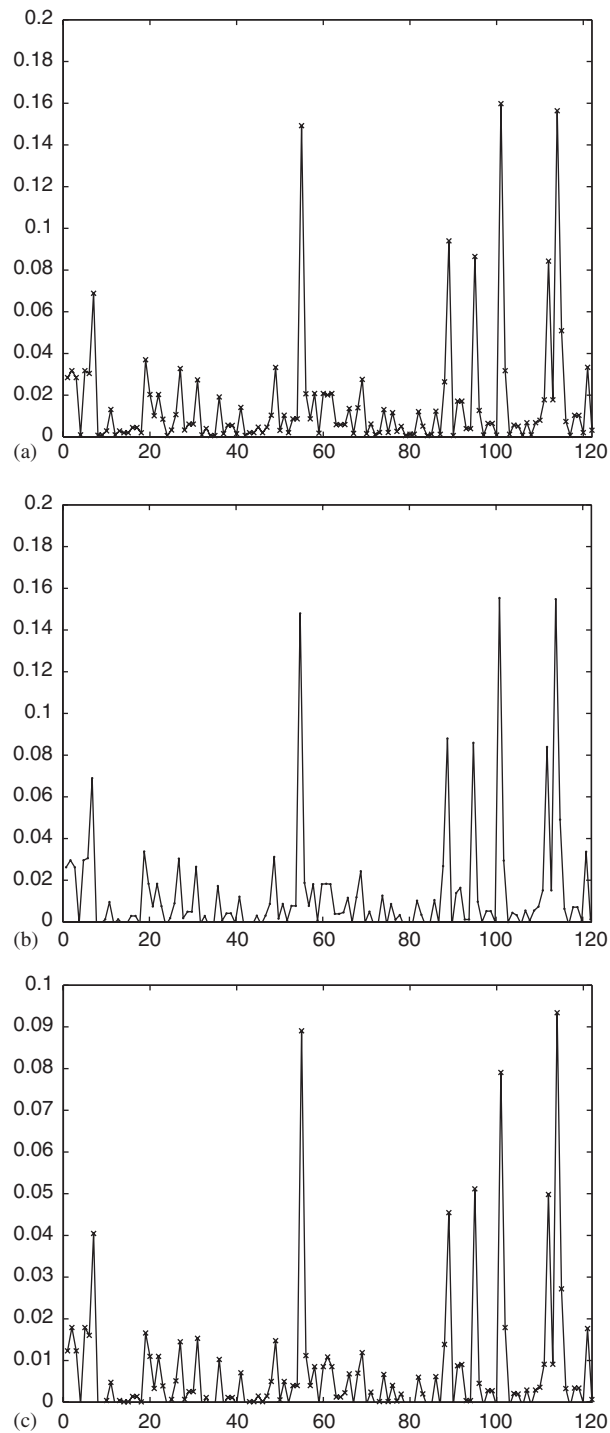


Fig. 4. Deletion measures (deletion of an observation) for the headache data. (a) $GD_{[M_k]}^1$ vs k , (b) $QD_{[M_k]}^1$ vs k , (c) $LD_{[M_k]}^1$ vs k .

from the estimation all lead to approximately 20% change in the ML estimate of β_2 . Only the observation with $k = 55$ substantially influences the estimate of δ . The estimate of β_1 , in contrast, is quite stable and is not highly influenced by individual observations or subjects.

Table 2

Headache data: the influence of influential observations

Observations deleted	$\hat{\beta}_{1[M_i]}$	$\frac{\hat{\beta}_{1[M_i]} - \hat{\beta}_1}{\hat{\beta}_1}$ (%)	$\hat{\beta}_{2[M_i]}$	$\frac{\hat{\beta}_{2[M_i]} - \hat{\beta}_2}{\hat{\beta}_2}$ (%)	$\hat{\delta}_{[M_i]}$	$\frac{\hat{\delta}_{[M_i]} - \hat{\delta}}{\hat{\delta}}$ (%)
No	−1.717	—	0.282	—	0.695	—
Subject 25	−1.667	−2.91	0.168	−40.4	0.709	2.0
Subject 13	−1.728	0.64	0.37	31.2	0.566	−18.6
Subjects 13 and 25	−1.766	2.9	0.190	−32.6	0.565	−18.7
Observation 55 ($i = 12, j = 3$)	−1.759	2.5	0.234	−17.0	0.769	10.7
Observation 101 ($i = 23, j = 1$)	−1.741	1.4	0.332	17.7	0.68	−2.2
Observation 114 ($i = 25, j = 4$)	−1.719	0.1	0.228	−19.2	0.678	−2.5
Observations 55, 101 and 114	−1.777	3.5	0.224	−20.6	0.734	5.6

4.1. Simulation studies

Two simulation studies are conducted to illustrate the effectiveness of the proposed deletion measures, and to compare with the results that are given by Xiang et al. (2002). Our one-step approximation deletion measures are based on an EM type estimation procedure which is different from the measures of Xiang et al. (2002), which are based on BLUP formulation (McGilchrist, 1994). The following Poisson linear mixed model is considered:

$$y_{ij}|u_i \sim \text{Poisson}(\mu_{ij}), \quad u_i \sim N(0, \sigma^2) \quad \text{and} \quad \log(\mu_{ij}) = \eta_{ij} = x_{ij}\beta + u_i, \quad (23)$$

where $i = 1, \dots, 10$, $j = 1, \dots, 30$. According to Xiang et al. (2002), x_{ij} is chosen as $j/30$ for all clusters. The true values of the parameters are: $\beta = 1.0$, $\sigma^2 = 0.1$. For illustrative purposes, only the performance of the deletion measures for clusters is considered. In Example 1, we have shown that the proposed deletion measures can successfully detect outliers in the data set. In the following simulation studies, there are no artificial outliers in the data sets, but different clusters have different influences. As the likelihood distance ($LD_{[M_i]}$, see (8)), which is based on the real estimates $\hat{\theta}$ and $\hat{\theta}_{[M_i]}$, is a classical metric to measure the influence of an observation, it will be taken as an objective value to judge the influence of a cluster. The cluster with the largest likelihood distance will be regarded as the most influential one, and the cluster with the second largest likelihood distance will be regarded as the second most influential one, and so on. The idea is the same as that in Xiang et al. (2002) who used an exact Cook's distance as an objective value to judge the influence of a cluster or a set of clusters. Here “the most influential one” does not necessarily mean that the cluster exerts substantial influence on the analysis. We will investigate whether or not the deletion measures based on one-step approximation can successfully identify the influential clusters in concordance with the exact $LD_{[M_i]}$ values. The proportions of correct identification of the most influential cluster and the first two most influential clusters among simulations are of interest.

Firstly, the proposed deletion procedure is applied to 101 randomly simulated data sets. The proportions of correct identification of the most influential cluster are 81.2% for $LD_{[M_i]}^1$, 86.1% for $QD_{[M_i]}^1$, and 85.1% $GD_{[M_i]}^1$, respectively. The proportions of correctly detecting the first two most influential clusters by $LD_{[M_i]}^1$, $QD_{[M_i]}^1$, and $GD_{[M_i]}^1$ are 45.5%, 63.4%, and 58.4%, respectively. For the same simulation design, the proportions of correct identification of the most influential cluster and the first two most influential clusters reported by Xiang et al. (2002) are 75% and 52%, respectively. It can be seen that our method is as efficient as the method given by Xiang et al. (2002). Furthermore, the relative frequencies of detecting the most influential cluster as the most influential one or the second most influential one by $LD_{[M_i]}^1$, $QD_{[M_i]}^1$, and $GD_{[M_i]}^1$ are 90.1%, 91.1%, and 93.1%, respectively. The diagnostic results for the data set with a median likelihood value among the 101 data sets are presented in Fig. 5. It can be seen from Fig. 5 that all the deletion measures successfully identify the first two most influential clusters.

Secondly, we further investigate the performance of our method on the “typical” data set in the sense that it has median value of the likelihoods among simulations. Ten thousand and one hundred data sets are randomly simulated and the value of likelihood of each data set is calculated. The “typical” data sets for every 101 data sets are selected. As a result, we have 100 data sets of which each has a median likelihood value among 101 randomly simulated data sets. The proportions of correct identification of the most influential cluster are 87.0% for $LD_{[M_i]}^1$, 86.0% for $QD_{[M_i]}^1$, and

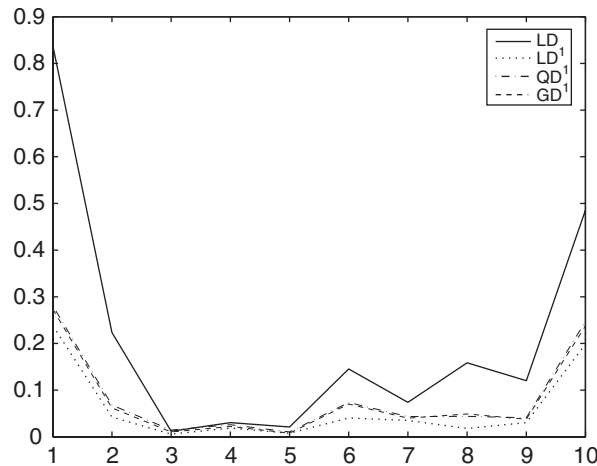


Fig. 5. Plots of $LD_{[M_i]}$, $LD^1_{[M_i]}$, $QD^1_{[M_i]}$ and $GD^1_{[M_i]}$ vs index of cluster for simulated data with the median likelihood.

85.0% for $GD^1_{[M_i]}$, respectively. The proportions of identification of the most influential cluster as the most influential one or the second most influential one are 95.0% for $LD^1_{[M_i]}$, 97.0% for $QD^1_{[M_i]}$, and 98.0% for $GD^1_{[M_i]}$, respectively.

Based on the results of our simulation studies, we may conclude that the performance of the proposed deletion measures for the GLMMs is acceptable, and more attention should be paid to the clusters with the first two largest deletion measures.

5. Discussion

The well-known approach of Cook (1977) has been extensively applied to identifying influential observations in analyzing statistical models. However, it is difficult to directly apply the results in Cook (1977) to GLMMs, because the observed-data likelihood of GLMMs involves intractable multiple integrals. Hence we have proposed a method to derive deletion measures for identifying influential observations by utilizing the nice features of the EM algorithm, the key idea of data augmentation, and the flexibility of MCMC methods. We have treated the latent random effects as hypothetical missing data, and have developed diagnostic measures on the basis of the Q-function instead of the complex observed-data likelihood function. We have demonstrated that the proposed influence measures, which are computed based on some by-products of the ML estimation, are rather easy to apply, require minimum additional effort to compute, and can efficiently identify influential observations.

It is well-known that the performance of deletion type measures may be discounted due to masking or swamping effects. Such effects are usually caused by the existence of a group rather than a single influential observation. One method to tackle the problem is to choose M_i appropriately so that it includes a set of observations rather than a single observation. However, although such multiple deletion method is intuitively convincing, it may not be feasible in practice because the number of combinations is usually insurmountable. Stepping searching procedure (Hadi, 1992; Atkinson and Mulira, 1993) is a popular approach in diagnostic literature to address the issue of masking effect. The development of a stepping searching procedure based on the proposed measures represents an interesting problem for further research. Robust procedures are another way to address the effect of outliers and influential observations in a statistical model. For example, He et al. (2002) developed M-estimators for the semiparametric models with longitudinal data. Cantoni (2004) proposed robust techniques for estimation, inference, and variable selection in the analysis of longitudinal data following a marginal mean modeling approach. The method of He et al. is insensitive to small proportions of contaminated data, and Canton's method can shed light on the outlying observations by checking the weights on the observations. Based on robust procedures such as those given by Cantoni (2004) and He et al. (2002), the diagnostic procedures developed in Rousseeuw and Leroy (1987) may be extended to statistical models with longitudinal data. For mixed effects models, the comparison of the efficiency of diagnostics that are based on robust and nonrobust estimation procedures may need further research.

Observations/subjects with maximum values of the deletion measures may not necessarily be influential. However, when such potential influential observations/subjects have been identified, further analysis that produces results similar to those given in Table 2 can be conducted to confirm whether they are actually “influential” or not. In effect, the problem of determining how large is large for a diagnostic measure has received considerable attention in the literature. While external scaling methods which determine cut-off values with reference to statistical distribution theories have been found useful for some simple measures (Belsley et al., 1980, p. 27), the natural gap approach (Lawrence, 1991) has also been widely applied to assess the magnitudes of measures for which statistical distributions are difficult to establish. In most practical applications, the natural gap approach, together with index plots, can always help for identifying observations that may require further attention.

Acknowledgements

The work described in this paper was partially supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region (Project Nos. CUHK4243/02H and CUHK 4242/03H). The authors are thankful for the constructive comments given by the reviewers.

Appendix A. The proof of Lemma 1

Proof. The right-hand side of (10) is equal to

$$\begin{aligned} & \int \frac{p_c(\mathbf{Y}_c | \hat{\psi}_{[M_i]})}{p_c(\mathbf{Y}_c | \hat{\psi})} p(\mathbf{Y}_m | \mathbf{Y}_o, \hat{\psi}) d\mathbf{Y}_m \\ &= \int \frac{p_o(\mathbf{Y}_o | \hat{\psi}_{[M_i]})}{p_o(\mathbf{Y}_o | \hat{\psi})} \frac{p(\mathbf{Y}_m | \mathbf{Y}_o, \hat{\psi}_{[M_i]})}{p(\mathbf{Y}_m | \mathbf{Y}_o, \hat{\psi})} p(\mathbf{Y}_m | \mathbf{Y}_o, \hat{\psi}) d\mathbf{Y}_m \\ &= \frac{p_o(\mathbf{Y}_o | \hat{\psi}_{[M_i]})}{p_o(\mathbf{Y}_o | \hat{\psi})} \int p(\mathbf{Y}_m | \mathbf{Y}_o, \hat{\psi}_{[M_i]}) d\mathbf{Y}_m. \end{aligned}$$

The proof is completed by the fact that the last integral is equal to 1. \square

Appendix B

$$\frac{\partial L_c(\psi | \mathbf{Y}_c)}{\partial \boldsymbol{\beta}} = \sum_{i=1}^N \left(\phi \sum_{j=1}^{n_i} \{y_{ij} - \mu_{ij}\} \dot{k}(\eta_{ij}) \mathbf{x}_{ij} \right),$$

$$\frac{\partial L_c(\psi | \mathbf{Y}_c)}{\partial \phi} = \sum_{i=1}^N \sum_{j=1}^{n_i} \left(y_{ij} \theta_{ij} - a(\theta_{ij}) + \frac{\partial c(y_{ij}, \phi)}{\partial \phi} \right),$$

$$\frac{\partial L_c(\psi | \mathbf{Y}_c)}{\partial \boldsymbol{\Sigma}} = -\frac{N}{2} \boldsymbol{\Sigma}^{-1} + \frac{1}{2} \boldsymbol{\Sigma}^{-1} \sum_{i=1}^N \mathbf{b}_i \mathbf{b}_i^T \boldsymbol{\Sigma}^{-1},$$

$$\frac{\partial^2 L_c(\psi | \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = \phi \sum_{i=1}^N \sum_{j=1}^{n_i} \left[-\ddot{a}(\theta_{ij}) \dot{k}^2(\eta_{ij}) + (y_{ij} - \mu_{ij}) \ddot{k}(\eta_{ij}) \right] \mathbf{x}_{ij} \mathbf{x}_{ij}^T,$$

$$\frac{\partial^2 L_c(\psi | \mathbf{Y}_c)}{\partial \phi \partial \boldsymbol{\beta}^T} = \sum_{i=1}^N \sum_{j=1}^{n_i} (y_{ij} - \mu_{ij}) \dot{k}(\eta_{ij}) \mathbf{x}_{ij},$$

$$\frac{\partial^2 L_c(\psi | \mathbf{Y}_c)}{\partial \phi^2} = \sum_{i=1}^N \sum_{j=1}^{n_i} \frac{\partial^2 c(y_{ij}, \phi)}{\partial \phi^2},$$

$$\frac{\partial^2 L_c(\psi | \mathbf{Y}_c)}{\partial \Sigma \partial \Sigma^T} = \frac{N}{2} \Sigma^{-1} \otimes \Sigma^{-1} - \left(\Sigma^{-1} \sum_{i=1}^N \mathbf{b}_i \mathbf{b}_i^T \Sigma^{-1} \right) \otimes \Sigma^{-1}.$$

References

- Atkinson, A.C., Mulira, H.M., 1993. The stalactite plot for the detection of multivariate outliers. *Statist. Comput.* 3, 27–45.
- Banerjee, M., Frees, E.W., 1997. Influence diagnostics for linear longitudinal models. *J. Amer. Statist. Assoc.* 92, 999–1005.
- Belsley, D.A., Kuh, E., Welsch, R.E., 1980. *Regression Diagnostics*. Wiley, New York.
- Booth, J.G., Hobert, J.P., 1999. Maximum generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *J. Roy. Statist. Soc. Ser. B* 61, 265–285.
- Breslow, N.E., 1984. Extra-Poisson variation in log-linear models. *Appl. Statist.* 33, 38–44.
- Cantoni, E., 2004. A robust approach to longitudinal data analysis. *Canad. J. Statist.* 32, 169–180.
- Chatterjee, S., Hadi, A.S., 1988. *Sensitivity Analysis of Linear Regression*. Wiley, New York.
- Clayton, D., Kaldor, J., 1987. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* 43, 671–681.
- Cook, R.D., 1977. Detection of influential observations in linear regression. *Technometrics* 19, 15–18.
- Cook, R.D., 1986. Assessment of local influence (with discussion). *J. Roy. Statist. Soc. Ser. B* 48, 133–169.
- Cook, R.D., Weisberg, S., 1982. *Residuals and Influence in Regression*. Chapman & Hall, New York.
- Davidian, M., Giltinan, M., 1995. *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall, London.
- Davison, A.C., Tsai, C.L., 1992. Regression model diagnostics. *Internat. Statist. Rev.* 60, 337–355.
- Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B* 39, 1–38.
- Fung, W.K., Zhu, Z.Y., Wei, B.C., He, X., 2002. Influence diagnostics and outlier test for semiparametric mixed models. *J. Roy. Statist. Soc. Ser. B* 64, 565–579.
- Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* 6, 721–741.
- Gu, M.G., Kong, F.H., 1998. A stochastic approximation algorithm with Markov chain Monte-Carlo method for incomplete data estimation problems. *Proc. Nat. Acad. Sci. USA* 95, 7270–7274.
- Hadi, A.S., 1992. Identifying multiple outliers in multivariate data. *J. Roy. Statist. Soc. Ser. B* 54, 761–771.
- Haslett, J., Dillane, D., 2004. Application of ‘delete=replace’ to deletion diagnostics for variance component estimation in the linear mixed model. *J. Roy. Statist. Soc. Ser. B* 66, 131–143.
- Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 97–109.
- He, X., Zhu, Z.Y., Fung, W.K., 2002. Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika* 89, 579–590.
- Kuk, A.Y.C., Cheng, Y.W., 1999. Pointwise and functional approximations in Monte Carlo maximum likelihood estimation. *Statist. Comput.* 9, 91–99.
- Lawrence, A.J., 1991. Local and deletion influence. In: Stahel, W., Weisberg, S. (Eds.), *Directions in Robust Statistics and Diagnostics, Part I*. Springer, Berlin, pp. 141–157.
- Lee, S.Y., Lu, B., 2003. Case-deletion diagnostics for nonlinear structural equation models. *Multivar. Behav. Res.* 38, 375–400.
- Lee, S.Y., Lu, B., Song, X.Y., 2006. Assessing local influence for nonlinear structural equation models with ignorable missing data. *Comput. Stat. Data An.* 50, 1356–1377.
- Lee, S.Y., Tang, N.S., 2004. Local influence analysis of nonlinear structural equation models. *Psychometrika* 69, 573–592.
- Lee, S.Y., Xu, L., 2003. Case-deletion diagnostics for factor analysis models with continuous and ordinal categorical data. *Sociol. Methods Res.* 31, 389–419.
- Lee, S.Y., Xu, L., 2004. Influence analyses of nonlinear mixed-effects models. *Comput. Statist. Data Anal.* 45, 321–341.
- Lu, S.Y., Song, X.Y., 2005. Local influence analysis for multivariate probit latent variable models. *J. Multivar. Anal.* in press.
- Manton, K.G., Woodbury, M.A., Stallard, E., Riggan, W.B., Creason, J.P., Pellom, A.C., 1989. Empirical Bayes procedures for stabilizing maps of U.S. cancer mortality rates. *J. Amer. Statist. Assoc.* 84, 637–650.
- McCullagh, P., Nelder, J.A., 1989. *Generalized Linear Models*. second ed. Chapman & Hall, London.
- McCulloch, C.E., 1997. Maximum likelihood algorithm for generalized linear mixed models. *J. Amer. Statist. Assoc.* 92, 162–170.
- McGilchrist, C.A., 1994. Estimation in generalized mixed models. *J. Roy. Statist. Soc. Ser. B* 56, 61–69.
- McKnight, B., Van Den Eeden, S.K., 1993. A conditional analysis for two-treatment multiple-period crossover designs with binomial or Poisson outcomes and subjects who drop out. *Statist. Med.* 12, 825–834.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equations of state calculations by fast computing machine. *J. Chem. Phys.* 21, 1087–1091.
- Ouwens, M.J.N., Tan, F.E.S., Berger, M.P.F., 2001. Local influence to detect influential data structures for generalized linear mixed models. *Biometrics* 57, 1166–1172.
- Rousseeuw, P.J., Leroy, A.M., 1987. *Robust Regression and Outlier Detection*. Wiley, New York.

- Song, X.Y., Lee, S.Y., 2004. Local influence analysis of two-level latent variable models with continuous and polytomous data. *Stat. Sinica* 14, 317–332.
- Stiratelli, R., Laird, N., Ware, J., 1984. Random effects models for serial observations with binary responses. *Biometrics* 40, 961–971.
- Tan, F.E.S., Ouwens, M.J.N., Berger, M.P.F., 2001. Detection of influential observations in longitudinal mixed effects regression models. *Statistician* 50, 271–284.
- Tanner, M.A., 1996. *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. second ed. Springer, New York.
- Wei, G.C.G., Tanner, M.A., 1990. A Monte Carlo implementation of the EM algorithm and the Poor man's data augmentation algorithm. *J. Amer. Statist. Assoc.* 85, 699–704.
- Williams, D.A., 1982. Extra-binomial variation in logistic linear models. *Appl. Statist.* 31, 144–148.
- Xiang, L., Tse, S.K., Lee, A.H., 2002. Influence diagnostics for generalized linear mixed models: applications to clustered data. *Comput. Statist. Data Anal.* 40, 759–774.
- Zeger, S.L., Liang, K.Y., Albert, P.S., 1988. Models for longitudinal data: a generalized estimation equation approach. *Biometrics* 44, 1049–1060.
- Zhu, H.T., Lee, S.Y., 2002. Analysis of generalized linear mixed models via a stochastic approximation algorithm with Markov chain Monte Carlo method. *Statist. Comput.* 12, 175–183.
- Zhu, H.T., Lee, S.Y., 2003. Local influence for generalized linear mixed models. *Canad. J. Statist.* 31, 293–309.
- Zhu, H.T., Lee, S.Y., Wei, B.C., Zhu, J., 2001. Case-deletion measures for models with incomplete data. *Biometrika* 88, 727–737.