J. R. Statist. Soc. B (2001)
63, *Part* 1, *pp.* 111–126

Local influence for incomplete-data models

Hong-Tu Zhu

University of Victoria, Canada

and Sik-Yum Lee

Chinese University of Hong Kong, People's Republic of China

[Received June 1999. Final revision July 2000]

Summary. This paper proposes a method to assess the local influence in a minor perturbation of a statistical model with incomplete data. The idea is to utilize Cook's approach to the conditional expectation of the complete-data log-likelihood function in the EM algorithm. It is shown that the method proposed produces analytic results that are very similar to those obtained from a classical local influence approach based on the observed data likelihood function and has the potential to assess a variety of complicated models that cannot be handled by existing methods. An application to the generalized linear mixed model is investigated. Some illustrative artificial and real examples are presented.

Keywords: EM algorithm; Generalized linear mixed model; Local influence; Model perturbation; Normal curvature

1. Introduction

An assessment of local influence in minor perturbations of a statistical model is important in data analysis. Cook (1986) proposed a simple and unified approach for such an assessment. This important work used the geometric normal curvature (Bates and Watts, 1980) to characterize the behaviour of an influence graph based on a well-behaved likelihood function. In recent years, it has been applied to various statistical models; see for example Beckman *et al*. (1987) for the mixed model analysis of variance, Escobar and Meeker (1992) for regression models with censored data and Tsai and Wu (1992) for transformation models, among others. However, it is very difficult to obtain local influence measures for some complex models because the building-blocks in the associated diagnostic measures involve intractable integrals. In addition, the underlying computational burden is heavy for problems with missing data.

The EM algorithm (Dempster *et al*., 1977) is a powerful method for computing maximum likelihood estimates for incomplete-data problems. Many complicated statistical models can be analysed by treating latent variables in the model as hypothetical missing data and applying the EM algorithm. See Lee and Poon (1998), McCulloch (1997), Meng and van Dyk (1997) and Wei and Tanner (1990), among many others.

Inspired by the power and wide applicability of the EM algorithm, we propose in this paper a method to assess local influence for incomplete data. The idea is to apply a procedure similar

Address for correspondence: Sik-Yum Lee, Department of Statistics, Chinese University of Hong Kong, Shatin, Hong Kong, People's Republic of China.
E-mail: sylee@sparc2.sta.cuhk.edu.hk

 1369–7412/01/63111

to that in Cook (1986) to the conditional expectation of the complete-data log-likelihood function instead of the observed data log-likelihood function. This procedure not only produces analytic results that are very similar to those obtained from a classical local influence approach but also makes possible an assessment of local influence for complicated models.

The paper is organized as follows. In Section 2, we introduce an objective function and propose a procedure to assess local influence in a model perturbation for incomplete data. Some nice properties and motivation for the procedure are discussed. To illustrate the methodology proposed two data examples are presented in Section 3. In Section 4, we apply the procedure to generalized linear mixed models and illustrate this with a real example. Technical proofs are given in Appendix A.

2. Local influence under incomplete data

To introduce the new methodology and some notation, we briefly review the EM algorithm. Let $Y_c = (Y_o, Y_m)$ be the complete-data set with a density $p(Y_c|\theta)$ parameterized by an r-dimensional parameter vector $\theta \in \Theta \subseteq R^r$, where Y_o and Y_m are the observed data and the missing data respectively. The complete-data log-likelihood

$$L_c(\theta|Y_c) = \log\{p(Y_c|\theta)\}$$

is simple in most statistical applications, whereas the observed data log-likelihood

$$L_o(\theta|Y_o) = \log\{p(Y_o|\theta)\}$$

is complicated. A standard EM algorithm consists of two steps: the expectation (E) step and the maximization (M) step. The E-step evaluates

$$Q(\theta|\theta^{(r)}) = E\{L_c(\theta|Y_c)|Y_o, \theta^{(r)}\},$$

where the expectation is taken with respect to the conditional distribution $p(Y_m|Y_o, \theta^{(r)})$. The M-step determines a $\theta^{(r+1)}$ that maximizes $Q(\theta|\theta^{(r)})$. Under mild conditions, the sequence $\{\theta^{(r)}\}$ obtained from the EM algorithm iterations converges to the maximum likelihood estimate $\hat{\theta}$ (Wu, 1983).

2.1. Motivation

Consider a perturbation vector $\omega = (\omega_1, \ldots, \omega_p)^T$ varying in an open region $\Omega \subset \mathbb{R}^r$. Let $L_o(\theta, \omega|Y_o)$ and $L_c(\theta, \omega|Y_c)$ be the observed data and complete-data log-likelihoods for the perturbed model. We assume that there is an ω^0 such that $L_o(\theta, \omega^0|Y_o) = L_o(\theta|Y_o)$ and $L_c(\theta, \omega^0|Y_c) = L_c(\theta|Y_c)$ for all θ. Let $\hat{\theta}_o(\omega)$ be the maximum likelihood estimator of θ from $L_o(\theta, \omega|Y_o)$. For simple statistical models, Cook (1986) considered the likelihood displacement function

$$\mathrm{LD}(\omega) = 2[L_o(\hat{\theta}|Y_o) - L_o\{\hat{\theta}_o(\omega)|Y_o\}]$$

and used it to assess the local influence of a minor perturbation. Although this approach is very useful, we encounter serious difficulties when applying it to complicated models, because of the intractable likelihood function. Hence, it is natural to consider alternatives to replace $\mathrm{LD}(\omega)$.

Motivated by recent advances relating to the EM algorithm, we propose the following Q-displacement function as an alternative to $\mathrm{LD}(\omega)$:

$$f_Q(\omega) = 2[Q(\hat{\theta}|\hat{\theta}) - Q\{\hat{\theta}(\omega)|\hat{\theta}\}] \tag{1}$$

where $\hat{\theta}(\omega)$ is the estimate of θ which maximizes

$$Q(\theta, \omega|\hat{\theta}) = E\{L_c(\theta, \omega|Y_c)|Y_o, \hat{\theta}\}.$$

This function can be regarded as a measure of the difference between $\hat{\theta}$ and $\hat{\theta}(\omega)$; it is greater than or equal to 0 and achieves its global minimum at ω^0. When no perturbation is introduced, $\hat{\theta}(\omega^0)$ equals the maximum likelihood estimate $\hat{\theta}$. Attention should be paid to situations where key results of the analysis are seriously influenced by a minor perturbation of $Q(\theta, \omega|\hat{\theta})$. It will be shown empirically in Section 3 that the approach proposed gives good results in some well-understood situations, so we can expect it to work well in other settings. Moreover, when there are no missing data, $f_Q(\omega)$ reduces to $\mathrm{LD}(\omega)$. Thus, normal curvature based on $f_Q(\omega)$ can be regarded as a generalization of the normal curvature based on $\mathrm{LD}(\omega)$.

A reason for using $\mathrm{LD}(\omega)$ is that it may be interpreted in terms of the asymptotic confidence region $\{\theta\colon 2\{L_o(\hat{\theta}) - L_o(\theta)\} \leqslant \chi^2_r(\alpha)\}$, where $\chi^2_r(\alpha)$ is the upper α-point of a χ^2-distribution with r degrees of freedom (Cook, 1986). Another motivation for our procedure is an analogous result for $f_Q(\omega)$. Let I_{obs} and I_{mis} be the observed information matrix and the missing information matrix (Louis, 1982) respectively. Under mild conditions, I_{obs} is positive definite with spectral decomposition $I_{\text{obs}} = B\Gamma B^{\mathrm{T}}$, where $\Gamma = \mathrm{diag}(\gamma_1, \ldots, \gamma_r)$. Let $I_{\text{obs}}^{1/2} = B\Gamma^{1/2}B^{\mathrm{T}}$ and $I_{\text{obs}}^{-1/2} = B\Gamma^{-1/2}B^{\mathrm{T}}$, where $\Gamma^{1/2} = \mathrm{diag}(\gamma_1^{1/2}, \ldots, \gamma_r^{1/2})$ and $\Gamma^{-1/2} = \mathrm{diag}(\gamma_1^{-1/2}, \ldots, \gamma_r^{-1/2})$.

Theorem 1.

(a) Let θ_0 be the true parameter vector. Then, if the maximum eigenvalue of the semi-positive definite matrix $I_{\text{obs}}^{-1/2} I_{\text{mis}} I_{\text{obs}}^{-1/2}$ converges to 0 in probability,

$$2\{Q(\hat{\theta}|\hat{\theta}) - Q(\theta_0|\hat{\theta})\} \xrightarrow{L} \chi^2_r,$$

where $\rightarrow^L$ denotes convergence in distribution.

(b) If D is a non-negative matrix with eigenvalues $\lambda'_1 \geqslant \ldots \geqslant \lambda'_r \geqslant 0$, and, if $I_{\text{obs}}^{-1/2} I_{\text{mis}} I_{\text{obs}}^{-1/2}$ converges to D in probability, then

$$2\{Q(\hat{\theta}|\hat{\theta}) - Q(\theta_0|\hat{\theta})\} \xrightarrow{L} X,$$

where X has the same distribution as $\Sigma_{i=1}^r (1 + \lambda'_i) Z_i^2$, with $Z_1, \ldots, Z_r$ independently and identically distributed as $N[0, 1]$.

Proofs for theorem 1 and the other theorems are sketched in Appendix A. It can be seen from equation (8) in Appendix A that the Q-distance $2\{Q(\hat{\theta}|\hat{\theta}) - Q(\theta_0|\hat{\theta})\}$ is different from the likelihood distance

$$2\{L_o(\hat{\theta}) - L_o(\theta_0)\} = (\hat{\theta} - \theta_0)^{\mathrm{T}} I_{\text{obs}} (\hat{\theta} - \theta_0) + o_p(\|\hat{\theta} - \theta_0\|^2).$$

If the amount of missing data is much less than observed data, condition (a) of theorem 1 is satisfied. The Q-distance and the likelihood distance are then closely related; in particular, their asymptotic distributions are both χ^2_r. Theorem 1, part (b), gives the asymptotic distribution of the Q-distance when the amount of missing data may be large. It indicates that the difference between the asymptotic distributions of these distances is then more substantial. On the basis of theorem 1, an asymptotic confidence region for θ is $\{\theta\colon 2\{Q(\hat{\theta}|\hat{\theta}) - Q(\theta|\hat{\theta})\} \leqslant X(\alpha)\}$, where $X(\alpha)$ is the upper α-point of the distribution of X.

Moreover, the theorem can be applied to test hypothesis H_0: $\theta = \theta_0$ against H_A: $\theta \neq \theta_0$. It can also be applied to other inference, especially to statistical models with missing data.

2.2. Q-displacement function and normal curvature

Following the arguments in Cook (1986), the influence graph of $f_Q(\omega)$ is defined as

$$\alpha(\omega) = (\omega^T, f_Q(\omega))^T. \tag{2}$$

The normal curvature $C_{f_Q,\mathbf{h}}$ of $\alpha(\omega)$ at ω^0 in the direction of a unit vector $\mathbf{h}$ from $\mathbb{R}^p$ can be used to summarize the local behaviour of $f_Q(\omega)$. Define

$$\ddot{Q}_{\omega^0} = \partial^2 Q\{\hat{\theta}(\omega)|\hat{\theta}\}/\partial\omega\partial\omega^T|_{\omega=\omega^0},$$

$$\ddot{Q}_\theta(\hat{\theta}) = \partial^2 Q(\theta|\hat{\theta})/\partial\theta\partial\theta^T|_{\theta=\hat{\theta}}$$

and

$$\Delta_\omega = \partial^2 Q(\theta, \omega|\hat{\theta})/\partial\theta\partial\omega^T|_{\theta=\hat{\theta}(\omega)}.$$

Using the rationale given in Cook (1986), it can be shown that the normal curvature $C_{f_Q,\mathbf{h}}$ of $\alpha(\omega)$ at ω^0 is

$$C_{f_Q,\mathbf{h}} = -2\mathbf{h}^T\ddot{Q}_{\omega^0}\mathbf{h} = 2\mathbf{h}^T\Delta_{\omega^0}^T\{-\ddot{Q}_\theta(\hat{\theta})\}^{-1}\Delta_{\omega^0}\mathbf{h}. \tag{3}$$

Although Cook's (1986) normal curvature is very useful, it may take any value and is not invariant under a uniform change of scale. Hence, there is no objective criterion to judge its size and the relative size of directions corresponding to large normal curvatures. Recently, on the basis of LD(ω), Poon and Poon (1999) introduced a conformal normal curvature which is a one-to-one function of the normal curvature and takes values in [0, 1]. Also, this curvature is invariant under conformal reparameterization of ω. A *conformal reparameterization* is a smooth map ϕ: $\Omega \to \Phi$ from the domain Ω to a new domain Φ of the same dimension such that the Jacobian matrix of ϕ, M, is non-singular throughout Ω and there is a positive number τ satisfying $MM^T = \tau \mathrm{ID}_r$, where ID_r is an $r \times r$ identity matrix. An example of a conformal reparameterization is $\phi(\omega) = M\omega + c$. See Poon and Poon (1999) for more discussion and examples. On the basis of these nice properties, objective bench-marks to judge size were constructed. Inspired by Poon and Poon (1999), a conformal normal curvature is introduced for our procedure. On the basis of equation (3), the conformal normal curvature $B_{f_Q,\mathbf{h}}$ at ω^0 in a unit direction $\mathbf{h}$ is defined as

$$B_{f_Q,\mathbf{h}} = \frac{-2\mathbf{h}^T\ddot{Q}_{\omega^0}\mathbf{h}}{\mathrm{tr}(-2\ddot{Q}_{\omega^0})}. \tag{4}$$

The norm of $-2\ddot{Q}_{\omega_0}$ is taken as the trace of the matrix, differently from Poon and Poon (1999). Two important properties of $B_{f_Q,\mathbf{h}}$ are given by the following theorem.

Theorem 2.

(a) *Invariance under reparameterization of* θ—if ψ: $\Theta \to \Psi$ is differentiable with a non-singular Jacobian, then $C_{f_Q,\mathbf{h}}$ and $B_{f_Q,\mathbf{h}}$ are invariant with respect to any reparameterization $\psi = \psi(\theta)$.
(b) *Invariance under conformal reparameterization of* ω—let $\phi = \phi(\omega)$ be a conformal reparameterization of ω. Then $B_{f_Q,\mathbf{h}}$ in any unit direction at ω^0 is invariant with respect to the conformal reparameterization $\phi = \phi(\omega)$.

Theorem 2 points out the important property that $C_{f_Q,\mathbf{h}}$ and $B_{f_Q,\mathbf{h}}$ are invariant with respect to any reparameterization of θ. Hence, results obtained via $C_{f_Q,\mathbf{h}}$ or $B_{f_Q,\mathbf{h}}$ are parameterization independent. Moreover, for a conformal reparameterization of ω, $B_{f_Q,\mathbf{h}}$ is still invariant.

2.3. Assessment of local influence

The previous arguments suggest that we should assess local influence by using $-\ddot{Q}_{\omega^0}$, or equivalently Δ_{ω^0} and $-\ddot{Q}_\theta(\hat{\theta})$. Under regularity conditions, $-\ddot{Q}_{\omega^0}$ is semipositive definite. To give a clear picture of $-2\ddot{Q}_{\omega^0}$, we consider its spectral decomposition

$$-2\ddot{Q}_{\omega^0} = \sum_{i=1}^{p} \lambda_i \mathbf{e}_i \mathbf{e}_i^{\mathrm{T}}, \tag{5}$$

where $\{(\lambda_i, \mathbf{e}_i)\}_{i=1}^p$ are eigenvalue–eigenvector pairs of $-2\ddot{Q}_{\omega^0}$ with $\lambda_1 \geqslant \ldots \geqslant \lambda_r > \lambda_{r+1} = \ldots = \lambda_p = 0$ and eigenvectors $\{\mathbf{e}_i\}_{i=1}^p$. Since $\Sigma_{j=1}^p e_{ij}^2 = 1$, if all elements e_{ij} in $\mathbf{e}_i$ are identical, then they all equal $p^{-1/2}$, which can be used as a bench-mark to assess the size of each case. Also, since the $\tilde{\lambda}_i = \lambda_i / \Sigma_{k=1}^r \lambda_k$ sum to 1, if all $\tilde{\lambda}_i$ are identical, then they all equal $1/r$, which can be used as a bench-mark to judge the size of an eigenvalue.

To detect influential observations, Cook (1986) pointed out that $\mathbf{h}_{\max} = \mathbf{e}_1$ provides important information for assessing the influence of a minor perturbation. Some previous work (Lesaffre and Verbeke, 1998; Poon and Poon, 1999) proposed to inspect all $C_{f_Q,\mathbf{u}_j}$ for further information, where $\mathbf{u}_j$ is a basic perturbation vector with jth entry 1 and other entries 0. Since $\mathrm{tr}(-2\ddot{Q}_{\omega^0}) = \Sigma_{i=1}^r \lambda_i$, it can be seen that

$$\begin{aligned} C_{f_Q,\mathbf{u}_j} &= \sum_{i=1}^{r} \lambda_i e_{ij}^2, \\ B_{f_Q,\mathbf{u}_j} &= \sum_{i=1}^{r} \tilde{\lambda}_i e_{ij}^2. \end{aligned} \tag{6}$$

Theorem 3. For any unit direction $\mathbf{h}$, $B_{f_Q,\mathbf{h}}$ satisfies the inequality $0 \leqslant B_{f_Q,\mathbf{h}} \leqslant 1$. In particular, $B_{f_Q,\mathbf{e}_k} = \tilde{\lambda}_k$.

Theorem 3 indicates that $B_{f_Q,\mathbf{h}}$ is a normalized measure, and its magnitude can be easily interpreted. Hence, objective bench-marks to judge size can be constructed as follows. Consider all eigenvalue–eigenvector pairs of $-2\ddot{Q}_{\omega^0}$. An eigenvector $\mathbf{e}_i$ of $-2\ddot{Q}_{\omega^0}$ is called m_0 influential if $B_{f_Q,\mathbf{e}_i} \geqslant m_0/r$. Since each component of $\mathbf{e}_i$ can be greater or smaller than 0, it is difficult to judge size in some cases. Moreover, it seems to be unreasonable to treat all the $\{e_i\}_{i=1}^p$ equally. On the basis of these considerations, we propose to inspect $\mathbf{e}_i^2 = (e_{i1}^2, \ldots, e_{ip}^2)^{\mathrm{T}}$ with the standardized eigenvalue $\tilde{\lambda}_i$ of $\mathbf{e}_i$ for $i = 1, \ldots, p$. We call the weighted sum of all m_0-influential eigenvectors,

$$M(m_0) = \sum_{i:\, \tilde{\lambda}_i \geqslant m_0/r} \tilde{\lambda}_i \mathbf{e}_i^2,$$

the *aggregate contribution vector of all m_0-influential eigenvectors*. The jth component of $M(m_0)$ is $M(m_0)_j = \Sigma_{i=1}^K \tilde{\lambda}_i e_{ij}^2$, where K is the largest i such that $\tilde{\lambda}_i \geqslant m_0/r$. On the basis of $\{M(m_0)_j\colon j = 1, \ldots, p\}$, we can assess the size of each case. In particular, when $m_0 = 0$, $M(0) = \Sigma_{i=1}^r \tilde{\lambda}_i \mathbf{e}_i^2$.

Theorem 4. The mean of $\{M(m_0)_j\}$ is $\overline{M}(m_0) = \Sigma_{i=1}^K \tilde{\lambda}_i / p$. In particular, $M(0)_j = B_{f_Q,\mathbf{u}_j}$ for any j, and their mean is $\overline{M}(0) = 1/p$.

Theorem 4 indicates that $M(0)_j$ and $B_{f_Q,\mathbf{u}_j}$ of the basic perturbation vector $\mathbf{u}_j$ are closely related. On the basis of theorems 3 and 4, we can use $\overline{M}(m_0)$ as a bench-mark to determine the significance of contributions from an individual case. For instance, Poon and Poon (1999) proposed to use $2\,\overline{M}(m_0)$, which is similar to what is often done for leverage in linear regression analysis. An alternative that takes into account the variation of $M(m_0)_j$ is to take

$$\overline{M}(m_0) + 2\,S_M(m_0)$$

as a bench-mark, where $S_M(m_0)$ is the corresponding sample standard error. This bench-mark works quite well in our empirical studies.

3. Examples

In this section we show that the method proposed gives analytic results that are similar to those obtainable from the local influence approach (Cook, 1986) based on the classical observed data likelihood. The idealized example and the aerosol data from Beckman *et al.* (1987) will be used.

3.1. An idealized example

Beckman *et al.* (1987), page 417, analysed an idealized data set with $\mathrm{LD}(\omega)$ via Cook's (1986) approach and three perturbation schemes. We use the same perturbation schemes and the same data, and then follow them in adding two standard errors to a single data point and to a single random effect. In our approach, $\mathbf{h}_{\max}$ and $M(0)_j$ are derived from $f_Q(\omega)$. To save space, we only report local influence measures based on the whole parameter vector. The data set was constructed from the model

$$y_{ij} = \mu + A_i + \epsilon_{ij},$$

where $i = 1, \ldots, 10$, $j = 1, 2, 3$ and $\mu = 0$, so $n = 30$. It is assumed that $A_i \sim N(0, \sigma_A^2)$, $\epsilon_{ij} \sim N(0, \sigma^2)$ and A_i and ϵ_{ij} are independent with $\sigma_A = 3$ and $\sigma = 1$. Here, $A_1, \ldots, A_{10}$ are treated as missing data. The complete-data log-likelihood function is

$$L_c(\theta|Y_c) = -15\,\log(\sigma^2) - 5\,\log(\sigma_A^2) - \sum_{i=1}^{10}\sum_{j=1}^{3}\frac{(y_{ij}-\mu-A_i)^2}{2\sigma^2} - \sum_{i=1}^{10}\frac{A_i^2}{2\sigma_A^2}.$$

On differentiating $L_c(\theta|Y_c)$ with respect to $\theta = (\mu, \sigma^2, \sigma_A^2)^{\mathrm{T}}$ twice, we have

$$-\frac{\partial^2 L_c(\theta|Y_c)}{\partial\theta\,\partial\theta^{\mathrm{T}}} = \begin{pmatrix} \dfrac{30}{\sigma^2} & \displaystyle\sum_{i=1}^{10}\sum_{j=1}^{3}\frac{\hat{y}_{ij}}{\sigma^4} & 0 \\ \displaystyle\sum_{i=1}^{10}\sum_{j=1}^{3}\frac{\hat{y}_{ij}}{\sigma^4} & \displaystyle\sum_{i=1}^{10}\sum_{j=1}^{3}\frac{\hat{y}_{ij}^2}{\sigma^6} - \frac{15}{\sigma^4} & 0 \\ 0 & 0 & \displaystyle\sum_{i=1}^{10}\frac{A_i^2}{\sigma_A^6} - \frac{5}{\sigma_A^4} \end{pmatrix}$$

where $\hat{y}_{ij} = y_{ij} - \mu - A_i$. Since

$$A_i|Y_o \sim N\left\{(3+\sigma^2\sigma_A^{-2})^{-1}\sum_{j=1}^{3}(y_{ij}-\mu),\ (3\sigma^{-2}+\sigma_A^{-2})^{-1}\right\},$$

an evaluation of the conditional expectation is straightforward.

First, we consider a perturbation scheme via an $n \times 1$ vector ω such that $\mathrm{var}(\epsilon_{ij}) = \sigma^2 \omega_{ij}^{-1}$. In this case, $\omega^0 = \mathbf{1}$ and

$$\frac{\partial^2 L_c(\theta, \omega | Y_c)}{\partial \omega_{ij} \partial \theta^T} = \left(\frac{\hat{y}_{ij}}{\sigma^2}, \frac{\hat{y}_{ij}^2}{2\sigma^4}, 0 \right),$$

so perturbing σ^2 does not involve σ_A^2. Figs 1(a) and 1(b) display index plots of $\mathbf{h}_{\max}$ and $M(0)_j$, in which the indices are the corresponding data point numbers. We observe that all the $M(0)_j$ are smaller than the bench-mark $0.083 = 1/30 + 2 \times 0.025$, so there is no influential data point. Figs 1(c) and 1(d) are reconstructions of Figs 1(a) and 1(b) after adding $2\sigma = 2$ to $y_{10,3}$. Only $M(0)_{30}$ is greater than the bench-mark $0.201 = 1/30 + 2 \times 0.084$. Thus, a minor change in the error variance of $y_{10,3}$ will have a strong influence on the analysis. Increasing the 10th group $\{y_{10,1}, y_{10,2}, y_{10,3}\}$ by $2\sigma_A = 6$ has little influence on local influence measures, so the resulting plots are not shown.

Secondly, we consider a perturbation scheme via a 10×1 vector ω such that $\mathrm{var}(A_i) = \omega_i^{-1} \sigma_A^2$; here

$$\frac{\partial^2 L_c(\theta, \omega | Y_c)}{\partial \omega_i \partial \theta^T} = \left(0, 0, \frac{A_i^2}{2\sigma_A^4} \right)$$

and perturbation of σ_A^2 does not influence μ and σ^2. Plots of $\mathbf{h}_{\max}$ and $M(0)_j$ are presented in Fig. 2, where the indices refer to group $i = 1, \ldots, 10$. Fig. 2(a) shows that the smallest and the largest $M(0)_j$ are smaller than the bench-mark $0.43 = 1/10 + 2 \times 0.165$. From Figs 2(c) and 2(d), the effect of a minor modification $A_{10} + 6$ is remarkable; $M(0)_{10}$ is greater than the bench-mark $0.670 = 1/10 + 2 \times 0.285$. Since $y_{10,3} + 2\sigma$ has little influence, the corresponding results are not presented.

Finally, we consider a perturbation of the response vector such that $y_{ij}(\omega) = y_{ij} + \omega_{ij}$. In this case, $\omega^0 = \mathbf{0}$ represents a non-perturbation situation and

$$\frac{\partial^2 L_c(\theta, \omega | Y_c)}{\partial \omega_{ij} \partial \theta^T} = \left(\frac{1}{\sigma^2}, \frac{\hat{y}_{ij}}{\sigma^4}, 0 \right).$$

Index plots for all parameters are presented in Figs 3(a) and 3(b). Those for $\mathbf{h}_{\max}$ and $M(0)_j$ are very similar to Figs 1(a) and 1(b). After $y_{10,3}$ has been increased by 2σ, it is identified to be influential in Fig. 3(c) and its corresponding $M(0)_{30}$ is greater than the bench-mark $0.092 = 1/30 + 2 \times 0.029$, but not as dramatically as in Fig. 1(c). When the 10th group is perturbed, plots of $\mathbf{h}_{\max}$ and $M(0)_j$ are very similar to Figs 1(a) and 1(b). For brevity, the details are not displayed. These results suggest that diagnostics obtained via a perturbation of the responses may be less sensitive than diagnostics derived from a perturbation to the model.

Results obtained via our approach are very similar to those obtained in Beckman *et al.* (1987) via the classical local influence approach based on the observed data likelihood. Index plots of $\mathbf{h}_{\max}$ presented in Figs 1–3 are virtually the same as the corresponding plots given in Fig. 2 of Beckman *et al.* (1987). In addition, index plots $M(0)_j$ provide more information to assess the local influence, and it seems that the bench-mark proposed works well in this analysis.

3.2. *Analysis of the aerosol data*

Beckman *et al.* (1987) also analysed an aerosol data set with the mixed model

$$y_{ijkl} = \mu + \alpha_i + \beta_j + z_{jk} + \epsilon_{ijkl}, \qquad \text{for } i, j = 1, 2 \text{ and } k, l = 1, 2, 3,$$

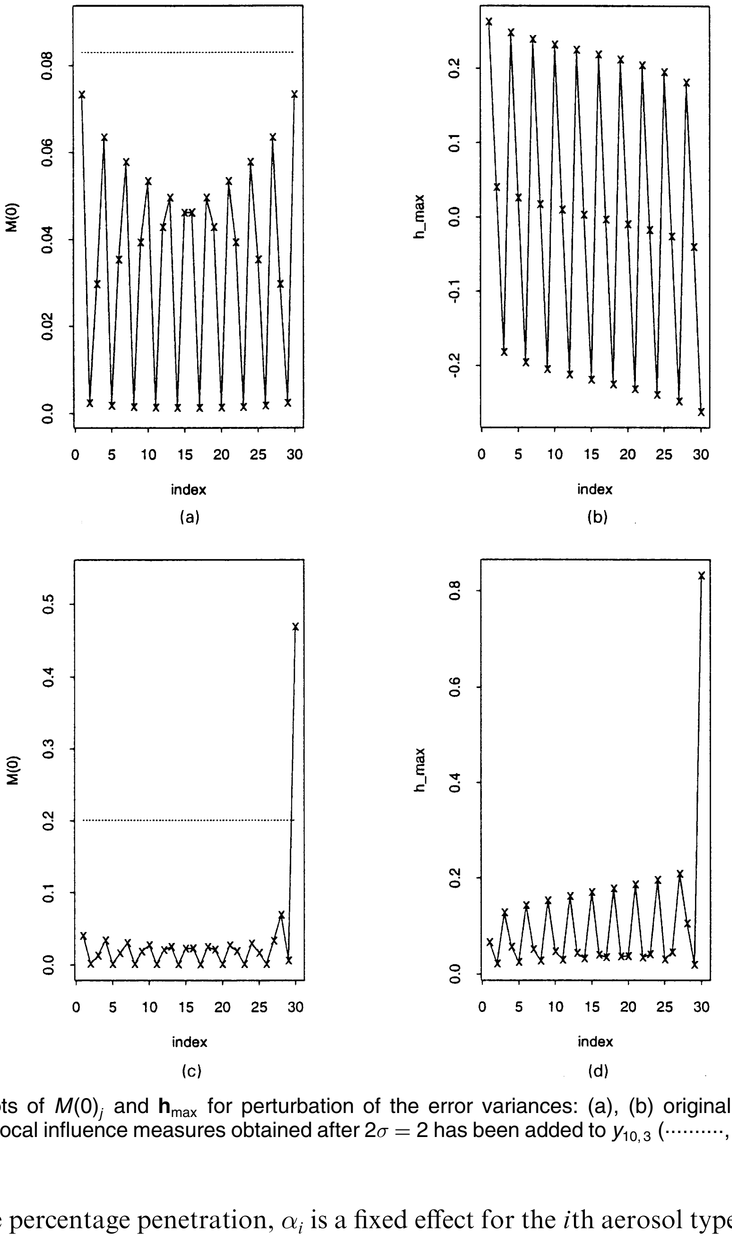

Fig. 1. Index plots of $M(0)_j$ and $\mathbf{h}_{max}$ for perturbation of the error variances: (a), (b) original local influence measures; (c), (d) local influence measures obtained after $2\sigma = 2$ has been added to $y_{10,3}$ (··········, bench-mark for $M(0)$)

where y_{ijkl} is the percentage penetration, α_i is a fixed effect for the ith aerosol type, β_j is a fixed effect for the jth filter manufacturer, z_{jk} is a random effect for the kth filter nested within the jth manufacturer and ϵ_{ijkl} is the error associated with the lth replication in the (ijk)th cell. It is assumed that $z_{jk} \sim N[0, \sigma_F^2]$, $\epsilon_{ijkl} \sim N[0, \sigma^2]$, and z_{jk} and ϵ_{ijkl} are independent. The maximum likelihood estimate of $\theta = (\mu, \alpha_1, \beta_1, \sigma^2, \sigma_F^2)^T$ was found to be $(0.992, 0.197, -0.597, 0.136, 0.633)^T$. Beckman *et al.* (1987) found that cases y_{1221} and y_{1222} are influential, with y_{1221} more influential than y_{1222}.

We consider a perturbation scheme via a 36×1 perturbation vector ω such that $\text{var}(\epsilon_{ijkl}) = \sigma^2/\omega_{ijkl}$. Here, z_{jk} ($j = 1, 2$; $k = 1, 2, 3$) are treated as Y_m. The complete-data log-likelihood function is

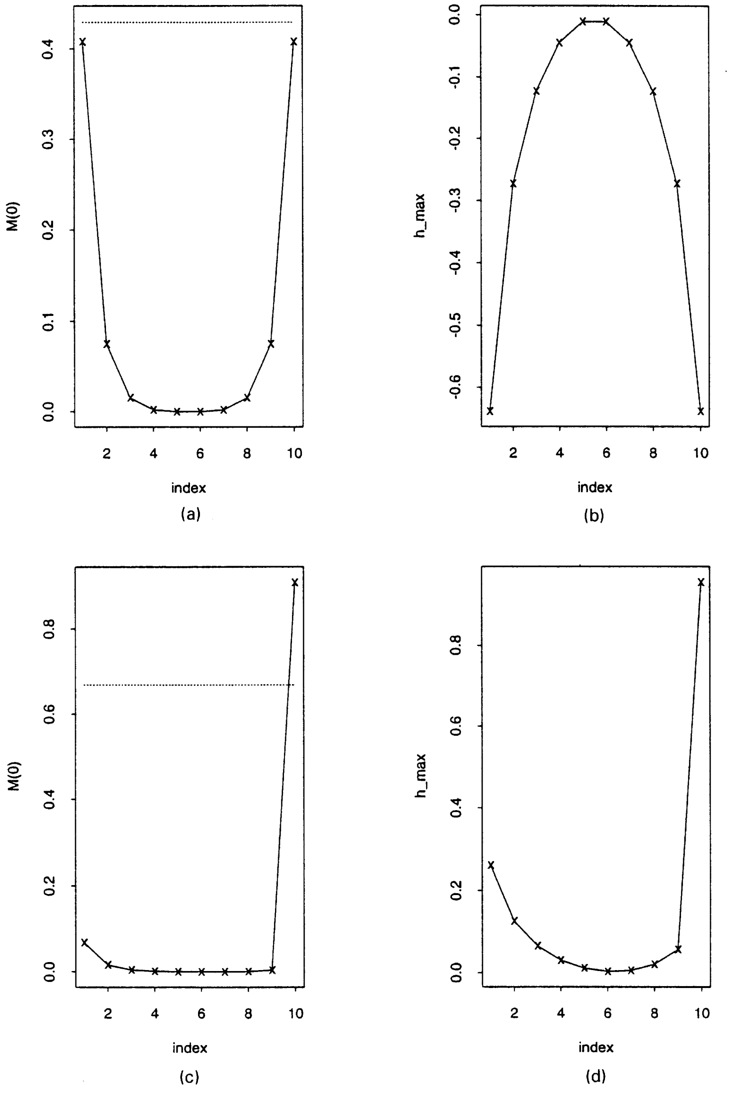

Fig. 2. Index plots of $M(0)_j$ and $\mathbf{h}_{\max}$ for perturbation of the random-effects variances: (a), (b) original local influence measures; (c), (d) local influence measures obtained after $2\sigma_A = 6$ has been added to A_{10} (··········, bench-mark for $M(0)$)

$$L_c(\theta, \omega | Y_c) = -3 \log(\sigma_F^2) - 18 \log(\sigma^2) + \frac{1}{2} \sum_{i,j,k,l} \log(\omega_{ijkl}) - \frac{1}{2} \sum_{i,j,k,l} \frac{\omega_{ijkl} \mathbf{e}_{ijkl}^2}{\sigma^2} - \frac{1}{2} \sum_{j,k} \frac{z_{jk}^2}{\sigma_F^2},$$

where $\mathbf{e}_{ijkl} = y_{ijkl} - \mu - \alpha_i - \beta_j - z_{jk}$. By differentiation,

$$\frac{\partial^2 L_c(\theta, \omega | Y_c)}{\partial \omega_{ijkl} \partial \theta} = \left(1, (-1)^{i-1}, (-1)^{j-1}, 0, \frac{1}{2} \frac{\mathbf{e}_{ijkl}}{\sigma^2}\right) \frac{\mathbf{e}_{ijkl}}{\sigma^2},$$

and non-zero elements in $-\partial^2 L_c(\theta | Y_c)/\partial\theta\partial\theta^{\mathrm{T}}$ are given by

$$\frac{\partial^2 L_c(\theta|Y_c)}{\partial\mu\partial\mu} = \frac{\partial^2 L_c(\theta|Y_c)}{\partial\alpha_1\partial\alpha_1} = \frac{\partial^2 L_c(\theta|Y_c)}{\partial\beta_1\partial\beta_1} = -\frac{36}{\sigma^2},$$

$$-\frac{\partial^2 L_c(\theta|Y_c)}{\partial\mu\partial\sigma^2} = \sum_{i,j,k,l} \frac{\mathbf{e}_{ijkl}}{\sigma^4},$$

$$-\frac{\partial^2 L_c(\theta|Y_c)}{\partial\sigma_F^2\partial\sigma_F^2} = \sum_{j,k} \frac{z_{jk}^2}{\sigma_F^6} - \frac{3}{\sigma_F^4},$$

$$-\frac{\partial^2 L_c(\theta|Y_c)}{\partial\sigma^2\partial\sigma^2} = \sum_{i,j,k,l} \frac{\mathbf{e}_{ijkl}^2}{\sigma^6} - \frac{18}{\sigma^4},$$

$$-\frac{\partial^2 L_c(\theta|Y_c)}{\partial\alpha_1\partial\sigma^2} = \sum_{i,j,k,l} (-1)^{i+1}\frac{\mathbf{e}_{ijkl}}{\sigma^4},$$

$$-\frac{\partial^2 L_c(\theta|Y_c)}{\partial\beta_1\partial\sigma^2} = \sum_{i,j,k,l} (-1)^{j+1}\frac{\mathbf{e}_{ijkl}}{\sigma^4}.$$

Since

$$z_{jk}|\tilde{Y}_{jk}, \hat{\theta} \sim N\left\{(6+\sigma^2\sigma_F^{-2})^{-1}\sum_{i,l}(y_{ijkl}-\mu-\alpha_i-\beta_j), (6\sigma^{-2}+\sigma_F^{-2})^{-1}\right\},$$

where $\tilde{Y}_{jk} = (y_{1jk1}, y_{1jk2}, y_{1jk3}, y_{2jk1}, y_{2jk2}, y_{2jk3})^{\mathrm{T}}$, we can replace z_{jk} and z_{jk}^2 by their expectations $E(z_{jk}|\hat{\theta}, Y)$ and $E(z_{jk}^2|\hat{\theta}, Y)$. Thus, Δ_{ω^0} and $\ddot{Q}_\theta(\hat{\theta})$ can be obtained. These basic building-blocks for obtaining the diagnostic measures are easier to obtain than those presented in Beckman *et al.* (1987) with the observed data likelihood function. Fig. 4 shows the index plots of $M(0)_j$ and $\mathbf{h}_{\max}$. The bench-mark for $M(0)_j$ is $0.199 = 1/36 + 2 \times 0.086$. From Fig. 4(a) and this bench-mark, y_{1221} is identified as influential, with its $M(0)_j$ about 0.187. Again, y_{1221} is more influential than y_{1222}. Fig. 4(b) is very similar to Fig. 3(a) of Beckman *et al.* (1987). If we only consider $\mathbf{h}_{\max}$, then y_{1221} and y_{1222} are identified as influential. Hence our procedure gives essentially the same analytic results as would be obtained from a local influence approach with the classical observed data likelihood.

4. Extension to the generalized linear mixed model

In this section we illustrate the potential of our approach in analysing models in which it is very difficult or even impossible to achieve local influence analysis via the observed data likelihood. We sketch the methodology for generalized linear mixed models.

4.1. Generalized linear mixed models

The analysis of generalized linear mixed models has received much attention in the recent literature (see for example Breslow and Clayton (1993)). However, even estimation is non-trivial because it usually involves the computation of high dimensional integrals. Several approaches have been proposed, notably Breslow and Clayton (1993) using Laplace's method, Zeger and Karim (1991) using the Gibbs sampler, McCulloch (1997) using a Monte Carlo Newton–Raphson algorithm and Aitkin (1999) using the EM algorithm with numerical integration implemented as finite mixture maximum likelihood.

Data typically consist of a response y_{ij}, covariate vectors $\mathbf{x}_{ij}$ $(s_1 \times 1)$ and $\mathbf{z}_{ij}$ $(s_2 \times 1)$ for

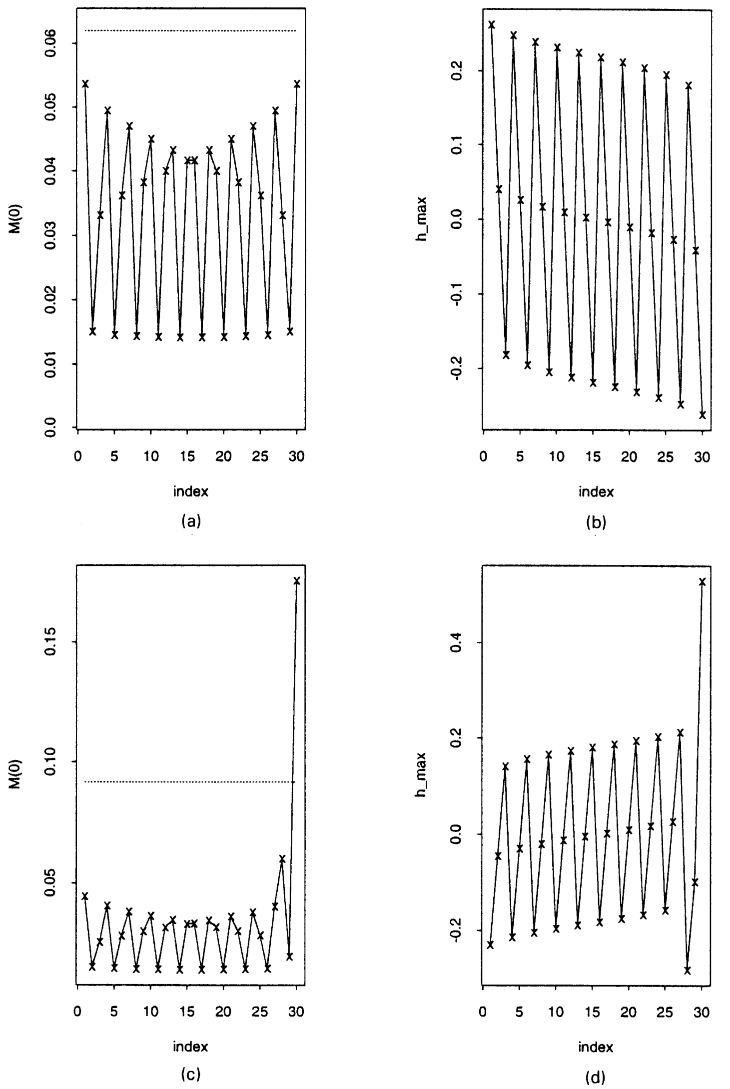

Fig. 3. Index plots of $M(0)_j$ and $\mathbf{h}_{\max}$ for perturbation of the response vector: (a), (b) original local influence measures; (c), (d) local influence measures obtained after $2\sigma = 2$ has been added to $y_{10,3}$ (··········, bench-mark for $M(0)$)

$i = 1, \ldots, I$, $j = 1, \ldots, n_i$ and $n = \Sigma_{i=1}^{I} n_i$. It is assumed that, conditional on $\mathbf{b}_i$, y_{ij} follows a density of the form

$$p(y_{ij}|\mathbf{b}_i) = \exp\left\{ \frac{y_{ij}\theta_{ij} - a(\theta_{ij})}{\phi} + c(y_{ij}, \phi) \right\}.$$

Conditional means $\mu_{ij} = E(y_{ij}|\mathbf{b}_i) = \dot{a}(\theta_{ij})$ are related to regression coefficients β and $\mathbf{b}_i$ via the link relationship

$$\mu_{ij} = \dot{a}\{k(\mathbf{x}_{ij}^{\mathrm{T}}\beta + \mathbf{z}_{ij}^{\mathrm{T}}\mathbf{b}_i)\},$$

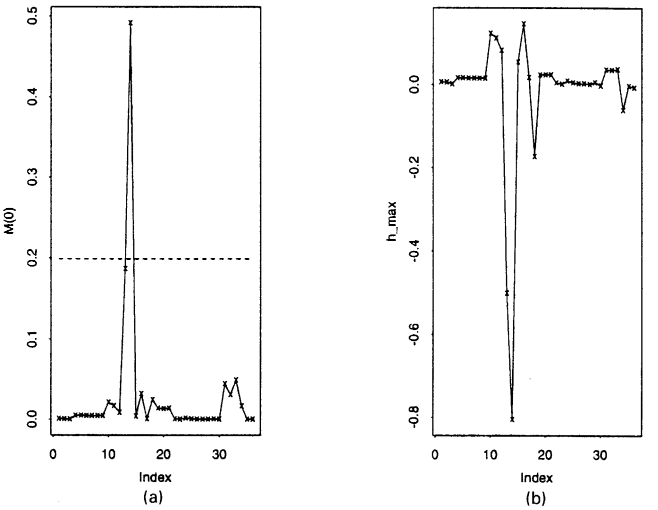

Fig. 4. Index plots of $M(0)_j$ and $\mathbf{h}_{max}$ for the aerosol data (- - - - -, bench-mark for $M(0)$)

where $\dot{a}(u) = \mathrm{d}a(u)/\mathrm{d}u$ and $k(\cdot)$ is a known link function. In the following, we assume that $\mathbf{b}_i \sim N(\mathbf{0}, \Sigma)$. The observed data log-likelihood involves intractable integrals of high dimension, so it is very difficult to assess the local influence on the basis of this function.

To apply our proposed method, $\{\mathbf{b}_i;\ i = 1, \ldots, I\}$ are treated as missing data. The complete-data log-likelihood is

$$L_c(\theta | Y_c) = \sum_{i=1}^{I} \sum_{j=1}^{n_i} \frac{y_{ij}\, k(\mathbf{x}_{ij}^T \beta + \mathbf{z}_{ij}^T \mathbf{b}_i) + a\{k(\mathbf{x}_{ij}^T \beta + \mathbf{z}_{ij}^T \mathbf{b}_i)\}}{\phi} - \sum_{i=1}^{I} \frac{1}{2} \mathbf{b}_i^T \Sigma^{-1} \mathbf{b}_i - \frac{I}{2} \log |\Sigma|.$$

For a given perturbation scheme, it is straightforward to derive $\partial^2 L_c(\theta | Y_c)/\partial\theta\partial\theta$ and $\partial^2 L_c(\theta, \omega | Y_c)/\partial\theta\partial\omega$. However, the conditional expectations involved in building-blocks Δ_{ω^0} and $\ddot{Q}_\theta(\hat{\theta})$ of the local influence measure cannot be evaluated in closed form. We handle this difficulty via Monte Carlo integration. Let $\{Y_m^{(s)}, s = 1, \ldots, S\}$ be a sample randomly drawn from the conditional distribution $p(Y_m | Y_o, \hat{\theta})$. Then the building-blocks are computed as

$$\begin{aligned} \Delta_{\omega^0} &\approx \frac{1}{S} \sum_{s=1}^{S} \left. \frac{\partial^2 L_c(\theta, \omega | Y_o, Y_m^{(s)})}{\partial\theta\partial\omega^T} \right|_{(\omega^0, \hat{\theta})}, \\ \ddot{Q}_\theta(\hat{\theta}) &\approx \frac{1}{S} \sum_{s=1}^{S} \left. \frac{\partial^2 L_c(\theta | Y_o, Y_m^{(s)})}{\partial\theta\partial\theta^T} \right|_{\hat{\theta}}. \end{aligned} \tag{7}$$

Techniques for sampling from a general density include rejection sampling, importance resampling (Rubin, 1987), the Gibbs sampler (Geman and Geman, 1984) and the Metropolis–Hasting algorithm (Metropolis *et al.*, 1953; Hasting, 1970) among others.

4.2. An illustrative example with longitudinal data

Thall and Vail (1990), Table 2, analysed a data set from a clinical trial of 59 epileptics who were randomly assigned to treatment ($T = 1$) and placebo ($T = 0$) groups as an adjuvant to the standard chemotherapy. Each patient reported the number of seizures in each of four 2-week

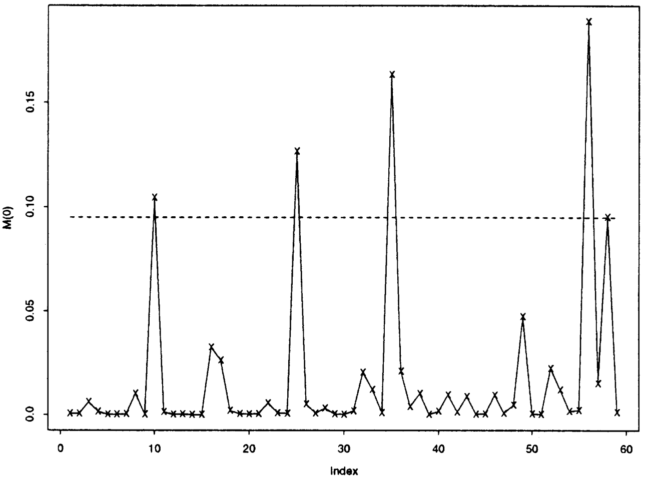

Fig. 5. Index plots of $M(0)_j$ for the longitudinal data: patients are identified according to the identification number in Table 2 of Thall and Vail (1990) (- - - - -, bench-mark for $M(0)$)

observation periods. The response variable y_{ij}, the seizure count for patient i on the jth visit, is assumed to be conditionally Poisson distributed with mean μ_{ij} such that

$$\log\{\mu_{ij}(\mathbf{b}_i)\} = \mathbf{x}_{ij}^{\mathrm{T}}\beta + b_{i1} + b_{i2}\,\mathrm{Visit}_j/10,$$

where $\mathbf{b}_i = (b_{i1}, b_{i2})^{\mathrm{T}}$. The covariates $\mathbf{x}_{ij}$ are the intercept term, the logarithm of a pre-experiment base-line count B of seizures, treatment T, their interaction $B \times T$, the logarithm of the patient's age and a variable Visit_j for each of four clinic visits (−3, −1, 1, 3). We assume that random effects b_{i1} and b_{i2} are independent, $b_{i1} \sim N(0, \sigma_{11}^2)$ and $b_{i2} \sim N(0, \sigma_{22}^2)$. This gives a random-effect model which is not invariant under the location change in the Visit_j covariate. The maximum likelihood estimate of θ is $(\hat{\beta}^{\mathrm{T}}, \hat{\sigma}_{11}, \hat{\sigma}_{22}) = (-1.36, 0.88, -0.93, 0.34, 0.48, -0.27, 0.50, 0.72)$.

On the basis of a plot of random effects (b_{i1}, b_{i2}), Breslow and Clayton (1993) pointed out that patient 135 has a marked improvement over time after an initially high seizure rate, patients 227, 225 and 112 have the highest overall count level relative to the expectation based on the covariables and patient 232 has especially low or zero counts. These patients were regarded informally as 'outliers' (see also Thall and Vail (1990)). As an illustration of our procedure, we consider a perturbation scheme such that

$$\mathrm{var}(b_{i1}) = \omega_i^{-1}\sigma_{11}^2$$

for $i = 1, \ldots, 59$. Expressions for matrices $\partial^2 L_c(\theta, \omega|Y_c)/\partial\theta\partial\omega^{\mathrm{T}}$ and $\partial^2 L_c(\theta|Y_c)/\partial\theta\partial\theta^{\mathrm{T}}$ can be derived via standard matrix calculus. Building-blocks Δ_{ω^0} and $-\ddot{Q}_\theta(\hat{\theta})$ cannot be evaluated in closed form, so they are approximated as in expression (7) by random observations $\{\mathbf{b}_i^{(s)} = (b_{i1}^{(s)}, b_{i2}^{(s)});\ s = 1, \ldots, S;\ i = 1, \ldots, 59\}$ sampled from the conditional distribution $p(\mathbf{b}_i|Y_o, \hat{\theta})$ via the Metropolis–Hasting algorithm (Gamerman, 1997; Zeger and Karim, 1991). The algorithm is implemented as follows: at the rth iteration with a current value $\mathbf{b}_i^{(r)}$, a new candidate $\mathbf{b}_i$ is generated from $N\{\mathbf{b}_i^{(r)}, \mathbf{C}(\mathbf{0})\}$, where

$$C(\mathbf{b}_i) = \begin{pmatrix} \sum_{j=1}^{4} \mu_{ij}(\mathbf{b}_i) + \sigma_{11}^{-2} & \sum_{j=1}^{4} \mu_{ij}(\mathbf{b}_i)\,\mathrm{Visit}_j \\ \sum_{j=1}^{4} \mu_{ij}(\mathbf{b}_i)\,\mathrm{Visit}_j & \sum_{j=1}^{4} \mu_{ij}(\mathbf{b}_i)\,\mathrm{Visit}_j^2 + \sigma_{22}^{-2} \end{pmatrix}^{-1}.$$

The probability of accepting this new candidate is

$$\min\{1, p(\mathbf{b}_i|Y_o, \hat{\theta})/p(\mathbf{b}_i^{(r)}|Y_o, \hat{\theta})\}.$$

We used a burn-in phase of 2000 cycles and then further collected $S = 10000$ random observations with a spacing of 2. Thus, 22000 iterations were run to calculate Δ_{ω^0} and $-\ddot{Q}_\theta(\hat{\theta})$. Results obtained from $M(0)_j$ are presented in Fig. 5. From Fig. 5 and using $0.095 = 1/59 + 2 \times 0.029$ as a bench-mark for $M(0)_j$, the 112th, 135th, 225th, 227th and 232nd patients are identified as influential. This confirms the finding of Breslow and Clayton (1993).

Acknowledgements

The authors would like to thank the Editor, the Associate Editor and two referees for their very helpful comments that led to a much improved presentation. The assistance of Michael Leung and Esther Tam in preparing the manuscript is gratefully acknowledged.

The work described in this paper was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region (project CUHK4356/00H).

Appendix A

A.1. Proof of theorem 1

Under mild conditions, the asymptotic distribution of $I_{\mathrm{obs}}^{1/2}(\hat{\theta} - \theta_0)$ is $N[0, \mathrm{ID}_r]$ (Cox and Hinkley, 1974). Since the complete-information matrix $-\ddot{Q}(\hat{\theta}|\hat{\theta})$ is the sum of I_{obs} and I_{mis}, it follows that

$$\begin{aligned} 2\{Q(\hat{\theta}|\hat{\theta}) - Q(\theta_0|\hat{\theta})\} &= (\hat{\theta} - \theta_0)^{\mathrm{T}}\{-\ddot{Q}(\hat{\theta}|\hat{\theta})\}(\hat{\theta} - \theta_0) + o_p(\|\hat{\theta} - \theta_0\|^2) \\ &= (\hat{\theta} - \theta_0)^{\mathrm{T}} I_{\mathrm{obs}}^{1/2}(\mathrm{ID}_r + I_{\mathrm{obs}}^{-1/2} I_{\mathrm{mis}} I_{\mathrm{obs}}^{-1/2}) I_{\mathrm{obs}}^{1/2}(\hat{\theta} - \theta_0) + o_p(\|\hat{\theta} - \theta_0\|^2). \end{aligned} \tag{8}$$

Assertions in theorem 1 can be proved by this result and the multivariate Slutzky theorem (Lehmann (1999), page 283).

A.2. Proof of theorem 2

(a) Let $\psi = \psi(\theta)$ and $\theta = \theta(\psi)$ be the transformation of θ and its inverse respectively. If $\hat{\theta}$ is the maximizer of θ under $L_o(\theta|Y_o)$, then $\hat{\psi} = \psi(\hat{\theta})$ is the maximizer of ψ under $L_o(\psi|Y_o) = L_o\{\theta(\psi)|Y_o\}$. Let $Q(\psi, \omega|\hat{\psi}) = Q\{\theta(\psi), \omega|\hat{\theta}\}$; then

$$\frac{\partial^2 Q(\psi, \omega|\hat{\psi})}{\partial\psi\,\partial\omega^{\mathrm{T}}} = \left(\frac{\partial\theta}{\partial\psi^{\mathrm{T}}}\right)^{\mathrm{T}} \frac{\partial Q(\theta, \omega|\hat{\theta})}{\partial\theta\,\partial\omega^{\mathrm{T}}}$$

and

$$\frac{\partial^2 Q(\psi, \omega|\hat{\psi})}{\partial\psi\,\partial\psi^{\mathrm{T}}} = \left(\frac{\partial\theta}{\partial\psi^{\mathrm{T}}}\right)^{\mathrm{T}} \frac{\partial^2 Q(\theta, \omega|\hat{\theta})}{\partial\theta\,\partial\theta^{\mathrm{T}}} \frac{\partial\theta}{\partial\psi^{\mathrm{T}}} + \left[\left(\frac{\partial Q(\theta, \omega|\hat{\theta})}{\partial\theta}\right)^{\mathrm{T}}\right]\left[\frac{\partial^2\theta}{\partial\psi\,\partial\psi^{\mathrm{T}}}\right],$$

where $[\cdot][\cdot]$ denotes array multiplication; see Wei (1998) for details. It follows from $\dot{Q}_\theta(\hat{\theta}) = \mathbf{0}$ that $C_{f_Q,\mathbf{h}}$ and $B_{f_Q,\mathbf{h}}$ are invariant with respect to reparameterization of θ.

(b) The second part can be proved via the same reasoning as in Poon and Poon (1999).

A.3. Proof of theorem 3

Since $\{\mathbf{e}_i\}_{i=1}^p$ is a standardized orthonormal basis in $\mathbb{R}^p$, it follows from equation (6) that

$$B_{f_Q,\mathbf{e}_k} = \mathbf{e}_k^{\mathrm{T}}(-2\ddot{Q}_{\omega^0})\mathbf{e}_k \bigg/ \sum_{i=1}^r \lambda_i = \lambda_k \bigg/ \sum_{i=1}^r \lambda_i = \tilde{\lambda}_k.$$

For any unit direction $\mathbf{h}$, $\mathbf{h} = \Sigma_{i=1}^p h_i\mathbf{e}_i$ and $\Sigma_{i=1}^p h_i^2 = 1$. Therefore,

$$B_{f_Q,\mathbf{h}} = \sum_{i=1}^p h_i^2 B_{f_Q,\mathbf{e}_i} = \sum_{i=1}^p \tilde{\lambda}_i h_i^2.$$

Thus, by $0 \leqslant \tilde{\lambda}_i \leqslant 1$, we have $0 \leqslant B_{f_Q,\mathbf{h}} \leqslant \Sigma_{i=1}^p h_i^2 = 1$.

A.4. Proof of theorem 4

Since $\Sigma_{j=1} e_{ij}^2 = 1$, we obtain

$$\overline{M}(m_0) = \sum_{j=1}^p M(m_0)_j/p = \sum_{i=1}^K \tilde{\lambda}_i \sum_{j=1}^p e_{ij}^2/p = \sum_{i=1}^K \tilde{\lambda}_i/p.$$

If $m_0 = 0$, then $M(0)_j = \Sigma_{i=1}^r \tilde{\lambda}_i e_{ij}^2$ and $\overline{M}(0) = 1/p$. It follows from equations (6) that $M(0)_j = B_{f_Q,\mathbf{u}_j}$.

References

Aitkin, M. (1999) A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, **55**, 117–128.

Bates, D. M. and Watts, D. G. (1980) Relative curvature measures of nonlinearity (with discussion). *J. R. Statist. Soc.* B, **42**, 1–25.

Beckman, R. J., Nachtsheim, C. J. and Cook, R. D. (1987) Diagnostics for mixed model analysis of variance. *Technometrics*, **29**, 413–426.

Breslow, N. E. and Clayton, D. G. (1993) Approximation inference in generalized linear mixed models. *J. Am. Statist. Ass.*, **88**, 9–25.

Cook, R. D. (1986) Assessment of local influence (with discussion). *J. R. Statist. Soc.* B, **48**, 133–169.

Cox, D. R. and Hinkley, D. V. (1974) *Theoretical Statistics*. London: Chapman and Hall.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. R. Statist. Soc.* B, **39**, 1–38.

Escobar, L. A. and Meeker, W. Q. (1992) Assessing influence in regression analysis with censored data. *Biometrics*, **48**, 507–528.

Gamerman, D. (1997) Sampling from the posterior distribution in generalized linear mixed models. *Statist. Comput.*, **7**, 57–68.

Geman, S. and Geman, D. (1984) Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Trans. Pattn Anal. Mach. Intell.*, **6**, 721–741.

Hastings, W. K. (1970) Monte Carlo sampling methods using Markov chains and their application. *Biometrika*, **57**, 97–109.

Lee, S. Y. and Poon, W. Y. (1998) Analysis of two-level structural equation models via EM type algorithms. *Statist. Sin.*, **8**, 749–766.

Lehmann, E. L. (1999) *Elements of Large-sample Theory*. New York: Springer.

Lesaffre, E. and Verbeke, G. (1998) Local influence in linear mixed models. *Biometrics*, **54**, 570–582.

Louis, T. A. (1982) Finding the observed information matrix when using EM algorithm. *J. R. Statist. Soc.* B, **44**, 226–233.

McCulloch, C. E. (1997) Maximum likelihood algorithms for generalized linear mixed models. *J. Am. Statist. Ass.*, **92**, 162–170.

Meng, X.-L. and van Dyk, D. (1997) The EM algorithm—an old folk song sung to a fast new tune (with discussion). *J. R. Statist. Soc.* B, **59**, 511–567.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953) Equations of state calculations by fast computing machine. *J. Chem. Phys.*, **21**, 1087–1091.

Poon, W.-Y. and Poon, Y. S. (1999) Conformal normal curvature and assessment of local influence. *J. R. Statist. Soc.* B, **61**, 51–61.

Rubin, D. B. (1987) *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

Thall, P. F. and Vail, S. C. (1990) Some covariance models for longitudinal count data with overdispersion. *Biometrics*, **46**, 657–671.

Tsai, C. L. and Wu, X. Z. (1992) Transformation model diagnostics. *Technometrics*, **34**, 197–202.
Wei, B. C. (1998) *Exponential Family Nonlinear Models*. Singapore: Springer.
Wei, G. C. G. and Tanner, M. A. (1990) Posterior computations for censored regression data. *J. Am. Statist. Ass.*, **85**, 829–839.
Wu, C. F. J. (1983) On the convergence properties of the EM algorithm. *Ann. Statist.*, **11**, 95–103.
Zeger, S. L. and Karim, M. R. (1991) Generalized linear models with random effects: a Gibbs sampling approach. *J. Am. Statist. Ass.*, **86**, 79–86.