Journal of Data Science **3**(2005), 153-177

Influence Diagnostics for Linear Mixed Models

Temesgen Zewotir[1] and Jacky S. Galpin[2]
[1] *University of KwaZulu-Natal and*
[2] *University of the Witwatersrand*

Abstract: Linear mixed models are extremely sensitive to outlying responses and extreme points in the fixed and random effect design spaces. Few diagnostics are available in standard computing packages. We provide routine diagnostic tools, which are computationally inexpensive. The diagnostics are functions of basic building blocks: studentized residuals, error contrast matrix, and the inverse of the response variable covariance matrix. The basic building blocks are computed only once from the complete data analysis and provide information on the influence of the data on different aspects of the model fit. Numerical examples provide analysts with the complete pictures of the diagnostics.

Key words: Case deletion, influential observations, random effects, statistical diagnostics, variance components ratios.

1. Introduction

The linear mixed model provides flexibility in fitting models with various combinations of fixed and random effects, and is often used to analyze data in a broad spectrum of areas. It is well known that not all observations in a data set play an equal role in determining estimates, tests and other statistics. Sometimes the character of estimates in the model may be determined by only a few cases while most of the data are essentially ignored. It is important that the data analyst be aware of particular observations that have an unusually large influence on the results of the analysis. Such cases may be assessed as being appropriate and retained in the analysis, may represent inappropriate data and be eliminated from the analysis, may suggest that additional data need to be collected, may suggest the current modeling scheme is inadequate, or may indicate a data reading or data entry error. Regardless of the ultimate assessment of such cases, their identification is necessary before intelligent subject-matter-based decisions can be drawn.

In ordinary linear models such model diagnostics are generally available in statistical packages and standard textbooks on applied regression, see for example, Cook and Weisberg (1982), Chatterjee and Hadi (1986, 1988). Oman

(1995) notes that, for the mixed model, standard statistical packages generally concentrate on estimation of and testing hypotheses about various parameters and submodels without providing overall diagnostics of the model. There have been a limited number of studies in linear mixed model diagnostics. Fellner (1986) examined robust estimates of the variance components, focusing on finding outlier resistant estimates of the mixed model parameters by limiting the influence of the outliers on the estimates of the model parameters. Beckman, Nachtsheim and Cook (1987) applied the local influence method of Cook (1986) to the analysis of the linear mixed model. Although an assessment of the influence of a model perturbation is generally considered to be useful, as Lawrance (1990) remarked, a practical and well established approach to influence analysis in statistical modeling is based on case deletion. Christensen, Pearson and Johnson (1992) (hereafter CPJ) studied case deletion diagnostics, in particular the analog of Cook's distance, for diagnosing influential observations when estimating the fixed effect parameters and variance components.

Some work has been done for special cases of the mixed model. In particular Martin (1992), Haslett and Hayes (1998) and Haslett (1999) considered the fixed effect linear model with correlated covariance structure and Christensen, Johnson and Pearson (1992, 1993) considered spatial linear models. But the approaches of Martin (1992), Haslett and Hayes (1998) and Haslett (1999) cannot be directly applied to a mixed model unless the entire focus of analysis is on the fixed effects estimate. Similarly the spatial linear model where covariance matrix of the data are determined by the locations at which observations are taken cannot be directly applied to the linear mixed model. We are also mindful that for fixed effect linear models with correlated error structure Haslett (1999) showed that the effects on the fixed effects estimate of deleting each observation in turn could be cheaply computed from the fixed effects model predicted residuals. However, when the interest of analysis is on the linear mixed model parameters (fixed effects and variance components) and predictors and likelihood functions this is not sufficient.

The key to making deletion diagnostics useable is the development of efficient computational formulas, allowing one to obtain the case deletion diagnostics by making use of basic building blocks, computed only once for the full model. The goal of this paper is to supplement the work of CPJ with such information and extend the ordinary linear regression influence diagnostics approach to linear mixed models. A number of diagnostics are proposed, which are analogues of the Cook's distance (Cook, 1977), likelihood distance (Cook and Weisberg, 1982), the variance (information) ratio (Belsley, Kuh, and Welsch, 1980), the Cook-Weisberg statistic (Cook and Weisberg, 1980) and the Andrews-Pregibon statistic (Andrews and Pregibon, 1978). The diagnostics are applied to two examples.

2. Model Definition and Estimation

The general form of the linear mixed model is

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \epsilon$$

where β is a $p \times 1$ vector of unknown constants, the fixed effects of the model; $\mathbf{X}$ is an $n \times p$ known design matrix of fixed numbers associated with β; $\mathbf{Z} = [\mathbf{Z}_1|\mathbf{Z}_2|\ldots,|\mathbf{Z}_r]$, where $\mathbf{Z}_i$ is an $n \times q_i$ known design matrix of the random effect factor i; $\mathbf{u}' = [\mathbf{u}_1'|\mathbf{u}_2'|\ldots,|\mathbf{u}_r']]$, where $\mathbf{u}_i$ is a $q_i \times 1$ vector of random variables from $N(\mathbf{0}, \sigma_i^2\mathbf{I})$, $i = 1, 2, \ldots, r$; ϵ is an $n \times 1$ vector of error terms from $N(\mathbf{0}, \sigma_e^2\mathbf{I})$; and $\mathbf{u}_i$ and ϵ are mutually independent. One may also write $\mathbf{u} \sim N(\mathbf{0}, \sigma_e^2\mathbf{D})$, where $\mathbf{D}$ is block diagonal with the i-th block being $\gamma_i\mathbf{I}_{q_i}$, for $\gamma_i = \sigma_i^2/\sigma_e^2$, so that $\mathbf{y}$ has a multivariate normal distribution with $E(\mathbf{y}) = \mathbf{X}\beta$ and $var(\mathbf{y}) = \sigma_e^2\mathbf{H}$, for $\mathbf{H} = \mathbf{I} + \sum_{i=1}^r \gamma_i\mathbf{Z}_i\mathbf{Z}_i'$.

No consensus exists on the best way to estimate the parameters (see, e.g., Harville, 1977; Robinson, 1991 and Searle, Casella and McCulloch, 1992). Standard methods for the linear mixed models are maximum likelihood, restricted maximum likelihood (REML) and minimum norm quadratic unbiased estimation (MINQUE). The discussion in this article focuses on maximum likelihood, but the results can also be easily applied to REML and MINQUE.

If the γ_i's (and hence $\mathbf{H}$) are known, the maximum likelihood estimates of β, σ_e^2 and the realized value of $\mathbf{u}$ are given by $\hat{\beta} = W(\mathbf{X}, \mathbf{X})^{-1}W(\mathbf{X}, \mathbf{y})$, $\hat{\sigma}_e^2 = W(\mathbf{d}, \mathbf{d})/n$, $\tilde{\mathbf{u}} = \mathbf{D}\mathbf{Z}'\mathbf{H}^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta}) = \mathbf{D}\mathbf{W}(\mathbf{Z}, \mathbf{d})$ for $\mathbf{d} = \mathbf{y} - \mathbf{X}\hat{\beta}$, and where $W(\mathbf{A}, \mathbf{B}) = \mathbf{A}'\mathbf{H}^{-1}\mathbf{B}$, in the notation of the W-operator of Hemmerle and Hartley (1973). Here $\hat{\beta}$ is the best linear unbiased estimator (BLUE) of β and $\tilde{\mathbf{u}}$ is the best linear unbiased predictor (BLUP) of $\mathbf{u}$. If the γ_i's are unknown, the maximum likelihood estimates are substituted back into $\mathbf{D}$ (and/or $\mathbf{H}$) to obtain $\hat{\beta}, \hat{\sigma}_e^2$ and $\tilde{\mathbf{u}}$ (see, e.g., Harville, 1977; Searle *et al.*, 1992, SAS Institute, 1992 and McCulloch and Searle, 2001).

Conventional optimization methods, which require first and second order partial derivatives, may be applied to obtain the maximum likelihood estimate of $\gamma' = [\gamma_1, \ldots, \gamma_r]$. The theoretical advantage of the Newton Raphson (NR) procedure (quadratic convergence rate in the neighborhood of the solution) and the practical advantage of the Fisher scoring procedure (robustness toward poor starting values) lead to the development of a hybrid NR and Fisher scoring algorithm (Jennrich and Sampson, 1976 and SAS Institute, 1992), used in this study.

Using minimum variance quadratic unbiased estimators with zero variance components commonly called MIVQUE0 (Goodnight, 1978) as initial estimates of the variance components, the NR iterative procedure, $\gamma^{(t+1)} = \gamma^{(t)} - (\mathbf{F}^{(t)})^{-1}\mathbf{g}^{(t)}$, while for the Fisher scoring algorithm, $\gamma^{(t+1)} = \gamma^{(t)} - (E(\mathbf{F}^{(t)}))^{-1}\mathbf{g}^{(t)}$, where, $\mathbf{g}^{(t)}$

and $\mathbf{F}^{(t)}$ are the gradient (first derivative of the log-likelihood) and the Hessian (second derivative of the log -ikelihood) evaluated at $\gamma^{(t)}$, respectively. The t-th element of the gradient vector $\mathbf{g}$ is

$$-\frac{1}{2}\text{trace}(W(\mathbf{Z}_i,\mathbf{Z}_i)) + \frac{1}{2\sigma_e^2}W(\mathbf{d},\mathbf{Z}_i)W(\mathbf{Z}_i,\mathbf{d}).$$

The (i,j)-th, $i,j = 1,2,\ldots r$, element of $\mathbf{F}$ is

$$\frac{1}{2}\text{trace}(W(\mathbf{Z}_i,\mathbf{Z}_j))W(\mathbf{Z}_j,\mathbf{Z}_i)) - \frac{1}{\sigma_e^2}W(\mathbf{d},\mathbf{Z}_i)W(\mathbf{Z}_i,\mathbf{Z}_j)W(\mathbf{Z}_j,\mathbf{d}),$$

implying that the expected value of the (i,j)-th element of $\mathbf{F}$ is

$$-\frac{1}{2}\text{trace}(W(\mathbf{Z}_i,\mathbf{Z}_j)W(\mathbf{Z}_j,\mathbf{Z}_i)).$$

The fitted values of the response variable $\mathbf{y}$ are $\hat{\mathbf{y}} = \mathbf{X}\hat{\beta} + \mathbf{Z}\tilde{\mathbf{u}}$, giving the residuals $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$. The symmetric matrix $\mathbf{R} = \mathbf{H}^{-1} - \mathbf{H}^{-1}\mathbf{X}W(\mathbf{X},\mathbf{X})^{-1}\mathbf{X}'\mathbf{H}^{-1}$ transforms the observations into residuals. That is, $\mathbf{e} = \mathbf{R}\mathbf{y}$, and if the γ_i's (and hence $\mathbf{H}$) are known $\mathbf{e} \sim N(\mathbf{0}, \sigma_e^2\mathbf{R})$ and correlation between e_i and e_j is entirely determined by the elements of $\mathbf{R}$, $\text{corr}(e_i,e_j) = r_{ij}/\sqrt{r_{ii}r_{jj}}$. If the γ_i's are unknown and their estimates are used these results are useful only as an approximation. In fact such approximation is common in mixed model inference. For instance inference about β and $\mathbf{u}$ use similar approximation in the SAS procedure, PROC MIXED (Little, *et al.*, 1996 and Verbeke and Molenberghs, 1997). The i-th Studentized residual is defined as $t_i = e_i/(\hat{\sigma}_e\sqrt{r_{ii}})$.

3. Background, Notation and Update Formulae

Without loss of generality, we partition the matrices as if the i-th omitted observation is the first row; i.e., $i = 1$. Then

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_i' \\ \mathbf{X}_{(i)} \end{bmatrix}, \quad \mathbf{Z}_j = \begin{bmatrix} \mathbf{z}_{ji}' \\ \mathbf{Z}_{j(i)} \end{bmatrix}, \quad \mathbf{Z} = \begin{bmatrix} \mathbf{z}_i' \\ \mathbf{Z}_{(i)} \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_i \\ \mathbf{y}_{(i)} \end{bmatrix},$$
$$\text{and} \quad \mathbf{H} = \begin{bmatrix} h_{ii} & \mathbf{h}_i' \\ \mathbf{h_i} & \mathbf{H}_{[i]} \end{bmatrix}.$$

For notational simplicity, $\mathbf{A}_{(i)}$ denotes an $n \times m$ matrix $\mathbf{A}$ with the i-th row removed, $\mathbf{a}_i$ denotes the i-th row of $\mathbf{A}$, and a_{ij} denotes the (i,j)-th element of $\mathbf{A}$. Similarly, $\mathbf{a}_{(i)}$ denotes a vector $\mathbf{a}$ with the i-th element removed, and a_i denotes the i-th element of $\mathbf{a}$. Moreover, using a similar notation to CPJ, we define

$$\breve{\mathbf{a}}_i = \mathbf{a}_i - \mathbf{A}_{(i)}\mathbf{H}_{[i]}\mathbf{h}_i \qquad p_{ii} = \breve{\mathbf{x}}_i'W(\mathbf{X},\mathbf{X})^{-1}\breve{\mathbf{x}}_i$$

$$m_i = h_{ii} - \mathbf{h}'\mathbf{H}_{[i]}\mathbf{h}_i, \text{ and } \qquad W_{[i]}(\mathbf{A},\mathbf{B}) = \mathbf{A}'_{(i)}\mathbf{H}_{[i]}^{-1}\mathbf{B}_{(i)}.$$

Further, let $\hat{\beta}_{(i)}, \hat{\sigma}^2_{e(i)}$, $\tilde{\mathbf{u}}_{(i)}$ and $\hat{\gamma}_{(i)}$ denote the estimates of β, σ_e^2, $\mathbf{u}$ and γ when the i-th observation is deleted, respectively.

Theorem 1: For fixed (or known) variance components ratios,

$$W_{[i]}(\mathbf{A},\mathbf{B}) = W(\mathbf{A},\mathbf{B}) - \breve{\mathbf{a}}_i\breve{\mathbf{b}}'_i/m_i.$$

The proof is given in the Appendix. The updating formulae for $\hat{\beta}$, $\hat{\sigma}_e^2$ and $\tilde{\mathbf{u}}$ then become

$$\begin{aligned}
\hat{\beta}_{(i)} &= \hat{\beta} - \frac{1}{m_i - p_{ii}} W(\mathbf{X},\mathbf{X})^{-1}\breve{\mathbf{x}}'_i(\breve{y}_i - \breve{\mathbf{x}}'_i\hat{\beta}),\\
\hat{\sigma}_{e(i)} &= \frac{1}{n-1}\{n\hat{\sigma}_e^2 - \frac{1}{m_i - p_{ii}}(\breve{y}_i - \breve{\mathbf{x}}'_i\hat{\beta})^2\} \text{ and}\\
\tilde{\mathbf{u}}_{(i)} &= \tilde{\mathbf{u}} - \mathbf{D}\{\breve{\mathbf{z}}_i - W(\mathbf{Z},\mathbf{X})^{-1}\breve{\mathbf{x}}_i\}\frac{\breve{y}_i - \breve{\mathbf{x}}'_i\hat{\beta}}{(m_i - p_{ii})}
\end{aligned}$$

Because of the iterative nature of estimation a simple analytical expression for the change in the variance components ratios when a case is deleted cannot be obtained. However, we can derive analytic expressions for the change in γ if we use $\hat{\gamma}$ as a starting value and perform a one step iteration with the i-th observation omitted. In general, such one step diagnostics adequately estimate the fully iterated change in the parameter with the i-th observation deleted, for models with non linear parameters, see Pregibon (1981), who pioneered the idea.

Cox and Hinkley (1974, page 308) noted that, if started with a consistent estimate, the NR and Fisher scoring methods lead in one step to an asymptotically efficient estimate. Thus, since $\hat{\gamma}$, the MLE of γ (obtained as the final iterated estimate for the full model), is a consistent estimate (see Hartley and Rao, 1967) this leads in one step to an efficient estimate of $\gamma_{(i)}$.

Noting that at convergence $\gamma^{(t+1)} = \gamma^{(t)} - (\mathbf{F}^{(t)})^{-1}\mathbf{g}^{(t)}$ is equal to $\gamma^{(t+1)} = \gamma^{(t)} - (E(\mathbf{F}^{(t)}))^{-1}\mathbf{g}^{(t)}$ and making use of Theorem 1, the one step updating formula for variance components ratios becomes

$$\hat{\gamma}_{(i)} = \hat{\gamma} - (\mathbf{Q} - \mathbf{G})^{-1}\mathbf{g}_{(i)}$$

where

$$\begin{aligned}
\mathbf{Q}(j,k) &= E(\mathbf{F}(i,j)),\\
\mathbf{G}(j,k) &= \frac{1}{m_i}\{\frac{1}{2m_i}\mathrm{ssq}(\breve{\mathbf{z}}_{ji}\breve{\mathbf{z}}'_{ki}) - \breve{\mathbf{z}}'_{ji}W(\mathbf{Z}_j\mathbf{Z}_k)\breve{\mathbf{z}}_{ki}\}
\end{aligned}$$

for $j, k = 1, 2, \ldots, r$, and the j-th element of $\mathbf{g}_{(i)}$ is

$$\begin{aligned}\frac{\partial\lambda}{\partial\gamma_{j(i)}} &= \frac{1}{2}\mathrm{ssq}(W(\mathbf{Z}_j, \mathbf{d}))\left(\frac{1}{\hat{\sigma}^2_{e(i)}} - \frac{1}{\hat{\sigma}^2_e}\right) \\ &+W(\mathbf{d}, \mathbf{Z}_j)\left\{W(\mathbf{Z}_j, \mathbf{X})W(\mathbf{X}, \mathbf{X})^{-1}\breve{\mathbf{x}}_i - \breve{\mathbf{z}}_{ji}\right\}\frac{\breve{y} - \breve{\mathbf{x}}_i'\hat{\beta}}{\hat{\sigma}^2_{e(i)}(m_i - p_{ii})} \\ &+\frac{(\breve{y} - \breve{\mathbf{x}}_i'\hat{\beta})^2}{2\sigma^2_{e(i)}(m_i - p_{ii})}\mathrm{ssq}(W(\mathbf{Z}_j, \mathbf{X})W(\mathbf{X}, \mathbf{X})^{-1}\breve{\mathbf{x}}_i - \breve{\mathbf{z}}_{ji}) + \frac{1}{2m_i}\mathrm{ssq}(\breve{\mathbf{z}}_{ji}),\end{aligned}$$

where the notation $\mathrm{ssq}(\mathbf{A})$ is the sum of the squares of elements of $\mathbf{A}$. Clearly, the one step estimate can only be obtained by explicitly recomputing and reinverting $(\mathbf{Q} - \mathbf{G})$ for each i. This is an $r \times r$ matrix, where r is the number of random effects in the model. In most practical cases r will not be large enough to imply that an $r \times r$ matrix inversion is expensive.

CPJ used $\breve{\mathbf{x}}_i, \breve{\mathbf{z}}_i, \breve{\mathbf{z}}_{ji}, \breve{y}_i,$ p_{ii} and m_i as the basic building blocks of the case deletion statistics. All are a function of a row (or column) of $\mathbf{H}$ and $\mathbf{H}_{[i]}^{-1}$. But in the iterative estimation process for the full model, it is not necessary to explicitly compute and store $\mathbf{H}$, an $n \times n$ matrix. Moreover, $\mathbf{H}_{[i]}$ is an $(n-1)\times(n-1)$ matrix, and its inversion and storage for each $i, i = 1, 2, \ldots, n$, would be computationally expensive. Although CPJ, overcame this last problem to a certain extent by noting that one can write $\mathbf{H}_{[i]}^{-1} = \mathbf{C}_{[i]} - \frac{1}{c_{ii}}\mathbf{c}_i\mathbf{c}_i'$, for

$$\mathbf{C} = \mathbf{H}^{-1} = \begin{bmatrix} c_{ii} & \mathbf{c}_i' \\ \mathbf{c}_i & \mathbf{C}_{[i]} \end{bmatrix},$$

it is still essential that $\breve{\mathbf{x}}_i, \breve{\mathbf{z}}_i, \breve{\mathbf{z}}_{ji}, \breve{y}_i,$ p_{ii} and m_i be computed using $\mathbf{H}_{[i]}^{-1}$, for each of the n cases deleted in turn. This is extremely time consuming. It is, therefore, important to find a solution to the problem, which we solve using other basic building blocks.

Theorem 2: (Basic theorem on efficient updating): If $\mathbf{C}_i' = [c_{ii}\ \mathbf{c}_i']$, so that $\mathbf{C}_i$ is the i-th column of $\mathbf{H}^{-1}$ and c_{ii} is the i-th diagonal element of $\mathbf{H}^{-1}$ then

$$\begin{aligned}&(1)\ \ m_i = \frac{1}{c_{ii}}, \qquad (2)\ \ \breve{\mathbf{x}}_i = \frac{1}{c_{ii}}\mathbf{X}'\mathbf{C}_i \\ &(3)\ \ \breve{\mathbf{z}}_{ji} = \frac{1}{c_{ii}}\mathbf{Z}_j'\mathbf{C}_i \ \text{ and } \ (4)\ \ \breve{y}_i = \frac{1}{c_{ii}}\mathbf{y}'\mathbf{C}_i.\end{aligned}$$

The proof is given in the Appendix. In summary, once we have $\mathbf{H}^{-1}$ a very efficient updating formula can be obtained by applying Theorem 2. Obtaining

$\mathbf{H}^{-1}$ seems to require inversion of an $n \times n$ matrix, but the formula $\mathbf{H}^{-1} = \mathbf{I} - \mathbf{Z}(\mathbf{D}^{-1} + \mathbf{Z}\mathbf{Z})^{-1}\mathbf{Z}'$ reduces the work, and implies that there is no need to compute and store either $\mathbf{H}_{[i]}^{-1}$ or $\mathbf{H}$.

Making use of Theorem 2, $\mathbf{C}, \mathbf{R}$, and $\mathbf{e}$ the updating formulae become

$$\begin{aligned}
\hat{\beta} - \hat{\beta}_{(i)} &= W(\mathbf{X}, \mathbf{X})^{-1}\mathbf{X}'\mathbf{C}_i \frac{e_i}{r_{ii}}, \\
\hat{\sigma}^2_{e(i)} &= \frac{n}{n-1}\hat{\sigma}^2_e - \frac{e_i^2}{(n-1)r_{ii}} = \hat{\sigma}^2_e \left(\frac{n}{n-1} - \frac{t_i^2}{n-1} \right), \\
\tilde{\mathbf{u}} - \tilde{\mathbf{u}}_{(i)} &= \mathbf{D}\mathbf{Z}'\mathbf{R}_i \frac{e_i}{r_{ii}} \\
\frac{\partial \lambda}{\partial \gamma_{j(i)}} &= \frac{1}{2\hat{\sigma}^2_e} \left(\frac{t_i^2 - 1}{n - t_i^2} \right) \mathrm{ssq}(\mathbf{Z}_j'\mathbf{e}) - \frac{t_i(n-1)}{(n-t_i^2)\sqrt{r_{ii}\hat{\sigma}^2_e}} \mathbf{e}'\mathbf{Z}_j\mathbf{Z}_j'\mathbf{R}_i \\
&\quad + \frac{t_i^2(n-1)}{2r_{ii}(n-t_i^2)} \mathrm{ssq}(\mathbf{Z}_j'\mathbf{R}_i) + \frac{1}{2c_{ii}} \mathrm{ssq}(\mathbf{C}_i'\mathbf{Z}_j),
\end{aligned}$$

and

$$\mathbf{G}(j,k) = \frac{1}{c_{ii}} \left\{ \frac{1}{2c_{ii}} \mathrm{ssq}(\mathbf{Z}_j'\mathbf{C}_i\mathbf{C}_i'\mathbf{Z}_k) - \mathbf{C}_i'\mathbf{Z}_j W(\mathbf{Z}_j, \mathbf{Z}_k)\mathbf{Z}_k'\mathbf{C}_i \right\},$$

$j, k = 1, 2, \ldots, r$.

As it can be seen from Table 3.1, when the i-th case is deleted the number of arithmetic operations involved is substantially reduced by $(4n-1)(n-1) - 2(p+q+1)$, which is at least $(2n - 1/2)(2n-3) - 1/2$, since $p + q + 1 \leq n$. Thus, at least $n\{(2n-1/2)(2n-3) - 1/2\}$ operations and the computation and storage of $\mathbf{H}$ are reduced in the overall basic blocks construction for a case deletion updating formulae. For large n, which is frequently the case, the improvement is substantial. Moreover, the rearrangement of the elements of $\mathbf{H}$ and $\mathbf{H}^{-1}$ in order to compute $\mathbf{H}_{[i]}^{-1}$ which creates a computational difficulty and defeats the compactness of the updating formula is avoided. Note that $\mathbf{H}_{[i]}^{-1}\mathbf{h}_i$ is common for m_i, $\breve{\mathbf{x}}_i, \breve{\mathbf{z}}_i$ and $\breve{y}_i$, and the number of arithmetic operations required to compute $\mathbf{H}_{[i]}^{-1}\mathbf{h}_i$ is counted only once.

Besides the computational advantage in a case deletion diagnostics, as it can be seen from Section 4, a more general, compact and appealing diagnostic tools can be easily obtained from our updating formulae.

For spatial linear models, $\mathbf{y} = \mathbf{X}\beta + \epsilon$ with $var(\epsilon) = \sigma_0^2\mathbf{V}$, where the covariance matrix of the data is determined by the locations at which observations are taken, $m_i, \breve{\mathbf{x}}_i$ and $\breve{y}_i$ were used as basic building blocks for the covariance function diagnostics (Christensen *et al.*, 1993) and prediction (i.e., prediction of future observation) diagnostics (Christensen *et al.*, 1992). In such cases too the computational advantage of our method relative to Christensen *et al.* (1992, 1993) is large by using $\mathbf{V}$ in place of $\mathbf{H}$.

Table 3.1: The number of operations required to compute the basic building blocks for ith case deletion updating

	$\mathbf{H}_{[i]}^{-1}$	$\mathbf{H}_{[i]}^{-1}\mathbf{h}_i$	m_i	$\breve{\mathbf{x}}_i$
CPJ	$(n-1)(2n-1)$	$2(n-1)^2$	$(2n-1)$	$p(2n-1)$
Ours	$-$	$-$	1	$p(2n+1)$

	$\breve{\mathbf{z}}_i$	$\breve{y}_i$	Total
CPJ	$q(2n-1)$	$(2n-1)$	$(2n-1)(n+p+q+1)$ $+2(n-1)^2$
Ours	$q(2n+1)$	$(2n+1)$	$(2n+1)(p+q+1)+1$

In the fixed effects models $\mathbf{y} = \mathbf{X}\beta + \epsilon$ with $var(\epsilon) = \sigma_0^2\mathbf{V}$, the generalized least squares estimate of β is given by $\hat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$. Noting that instead of the commonly used residuals the predicted residuals could be easily obtained in such a model, Haslett (1999) simplified the aspects of work discussed by Martin (1992) and Hayes and Haslett (1998) and showed that predicted residuals (which they referred to as "conditional residuals") play a key role in the diagnostics. But by making use of the two theorems (and using $\mathbf{V}$ in place of $\mathbf{H}$), the simplicity and role of the natural residuals is pronounced and that gives more insight to ascribe the source of influence to easily interpretable components such as residuals and leverages. It must be also noted that conditional residual can be easily obtained as $e_{(i)} = y_i - \mathbf{x}_i'\hat{\beta}_{(i)} - \mathbf{z}_i'\tilde{\mathbf{u}}_{(i)} = e_i/r_{ii}$.

The updating formulae assume that the variance components ratios are correct. CPJ, Banerjee and Frees (1997) and Haslett (1999) also make this assumption. Thus it seems important we first examine the variance components ratios diagnostics and take any necessary measure in order to build confidence that the variance components ratios estimate are equally sensitive to each observation. However, since the maximum likelihood estimate of the variance components ratios and the maximum likelihood estimate of fixed effects are independent such assessment might not be that essential. This is clearly demonstrated from our examples (Section 5) that the inclusion or exclusion of variance components ratios influential observations has no remarkable effect on the influence of fixed/random effects and the likelihood function.

4. Measures of Influence

We propose and investigate a number of diagnostics for the variance components ratios, fixed effects parameters, prediction of the response variable and of random effects, and the likelihood function.

4.1 Influence on variance components ratios

The general diagnostic tools for variance components ratios are the analogues of the Cook's distance and the information ratio.

4.1.1 Analogue of Cook's distance

The analogue of Cook's distance measure for variance components γ, is

$$\begin{aligned} \mathrm{CD}_i(\gamma) &= (\hat{\gamma}_{(i)} - \hat{\gamma})'[var(\hat{\gamma})]^{-1}(\hat{\gamma}_{(i)} - \hat{\gamma}) \\ &= -\mathbf{g}'_{(i)}(\mathbf{Q} - \mathbf{G})^{-1}\mathbf{Q}(\mathbf{Q} - \mathbf{G})\mathbf{g}_{(i)} \\ &= \mathbf{g}'_{(i)}(\mathbf{I}_r + var(\hat{\gamma})\mathbf{G})^{-2} var(\hat{\gamma})\mathbf{g}_{(i)}. \end{aligned}$$

Large values of $\mathrm{CD}_i(\gamma)$ highlight points for special attention.

4.1.2 Analogue of the information ratio

The analogue of the information ratio measures the change in the determinant of the maximum likelihood estimate's information matrix, giving

$$\begin{aligned} IR(\gamma) &= \frac{\det(Inf(\hat{\gamma}_{(i)}))}{\det(Inf(\hat{\gamma}))} = \frac{\det(\mathbf{Q} - \mathbf{G})}{\det(\mathbf{Q})} \\ &= \frac{\det(\mathbf{Q})\det(\mathbf{I}_r - \mathbf{Q}^{-1}\mathbf{G})}{\det(\mathbf{Q})} = \det(\mathbf{I}_r + var(\hat{\gamma})\mathbf{G}), \end{aligned}$$

where $\det(A)$ denotes the determinant of the square matrix $\mathbf{A}$. Ideally when all observations have equal influence on the information matrix, $IR(\gamma)$ is approximately one. Deviation from unity indicates that the i-th observation is potentially influential. Since $var(\hat{\gamma})$ and $\mathbf{I}_r$ are fixed for all $i = 1, 2, \ldots, n$, we observe that $IR(\gamma)$ will be a function of $\mathbf{G}$ which is a function of $\mathbf{C}_i$ and $c_{ii} = 1 - \mathbf{z}'_i(\mathbf{D}^{-1} + \mathbf{Z}\mathbf{Z})^{-1}\mathbf{z}_i$. Since c_{ii} is a function of the number of observations falling in the random effect levels, the values of $IR(\gamma)$ tend to be identical for a fairly balanced data set.

4.2 Influence on fixed effects parameter estimates

4.2.1 Analogue of Cook's distance

The Cook's distance can be extended to measure influence on the fixed effects in the mixed models by defining

$$\begin{aligned}
\mathrm{CD}_i(\beta) &= (\hat{\beta}_{(i)} - \hat{\beta})' W(\mathbf{X}, \mathbf{X})(\hat{\beta}_{(i)} - \hat{\beta})/(p\hat{\sigma}_e^2) \\
&= \mathbf{C}_i' \mathbf{X} W(\mathbf{X}, \mathbf{X})^{-1} \mathbf{C}_i e_i^2/(p r_{ii}^2 \hat{\sigma}_e^2) \\
&= \frac{(c_{ii} - r_{ii})t_i^2}{r_{ii} p}.
\end{aligned}$$

Large values of $\mathrm{CD}_i(\beta)$ indicate points for further consideration.

4.2.2 Analogue of the variance ratio

The variance ratio measures the change of the determinant of the variance of $\hat{\beta}$ when the i-th case is deleted:

$$\begin{aligned}
VR_i(\beta) &= \frac{\det(var(\hat{\beta}_{(i)}))}{\det(var(\hat{\beta}))} = \frac{\det(\hat{\sigma}_{e(i)}^2 W_{[i]}(\mathbf{X}, \mathbf{X})^{-\mathbf{1}})}{\det(\hat{\sigma}_e^2 W(\mathbf{X}, \mathbf{X})^{-1})} \\
&= \left(\frac{n - t_i^2}{n-1}\right)^p \frac{c_{ii}}{c_{ii} - \mathbf{C}_i \mathbf{X} W(\mathbf{X}, \mathbf{X})^{-1} \mathbf{X}' \mathbf{C}_i} \\
&= \left(\frac{n - t_i^2}{n-1}\right)^p \frac{c_{ii}}{r_{ii}}.
\end{aligned}$$

As this is close to 1 if the point is not influential, it seems sensible to use the relative measure $|VR_i(\beta) - 1|$ as a criterion for assessing the influence of the i-th observation on the variance of $\hat{\beta}$. The larger the statistic $|VR_i(\beta) - 1|$, the higher the influence of the i-th observation.

If one uses the trace instead of the determinant, the VR becomes

$$\begin{aligned}
VR_i(\beta) &= \mathrm{trace}\{[var(\hat{\beta})]^{-1}(var(\hat{\beta}_{(i)}))\} \\
&= \frac{\hat{\sigma}_{e(i)}}{\hat{\sigma}_e^2} \mathrm{trace}\{W(\mathbf{X}, \mathbf{X}) W_{[i]}(\mathbf{X}, \mathbf{X})^{-1}\} \\
&= \frac{n - t_i^2}{n-1} \mathrm{trace}\left\{\mathbf{I} + \mathbf{X}' \mathbf{C}_i \mathbf{C}_i' \mathbf{X} W(\mathbf{X}, \mathbf{X})^{-1} \frac{1}{r_{ii}}\right\} \\
&= \frac{n - t_i^2}{n-1} \left(\frac{c_{ii}}{r_{ii}} + p - 1\right).
\end{aligned}$$

Since $VR_i(\beta)$ will be close to p if removing the i-th point does not change the trace, we could use the relative measure $|VR_i(\beta) - p|$ as a criterion for assessing the influence of the i-th observation on the variance of $\hat{\beta}$. The larger the statistic $|VR_i(\beta) - p|$, the higher the influence of the i-th observation.

4.2.3 Analogue of the Cook-Weisberg statistic

Another statistic used to measure the change of the confidence ellipsoid volume of β is the Cook-Weisberg statistic. Under the assumption $\mathbf{y} \sim N(\mathbf{X}\beta, \sigma_e^2\mathbf{H})$, $\hat{\beta} \sim N(\beta, \sigma_e^2 W(\mathbf{X}, \mathbf{X})^{-1})$. The $100(1-\alpha)\%$ confidence ellipsoid for β is then given by $E = \{\beta : (\beta - \hat{\beta})' W(\mathbf{X}, \mathbf{X})(\beta - \hat{\beta})/(p\hat{\sigma}_e^2) \leq F(\alpha, p, n-p)\}$. When the i-th observation is omitted $E_{(i)} = \{\beta : (\beta - \hat{\beta}_{(i)})' W_{[i]}(\mathbf{X}, \mathbf{X})(\beta - \hat{\beta}_{(i)})/(p\sigma_{e(i)}^2) \leq F(\alpha, p, n-p-1)\}$. Cook and Weisberg (1980) proposed the logarithm of the ratio $E_{(i)}$ to E as a measure of influence on the volume of confidence ellipsoid for in the multiple linear regression model, namely

$$CW_i = \frac{\text{volume}(E_{(i)})}{\text{volume}(E)}, \quad i = 1, 2, \ldots, n.$$

Noting that the volume of the ellipsoid is proportional to the inverse of the square root of the determinant of the associated matrix of the quadratic form, the analogous measure for fixed effects β becomes

$$\begin{aligned} CW_i &= \log\left\{\frac{\det(W(\mathbf{X}, \mathbf{X})^{1/2}\hat{\sigma}_{e(i)}^p (F(\alpha, p, n-p))^{p/2})}{\det(W_{[i]}(\mathbf{X}, \mathbf{X}))^{1/2}\hat{\sigma}_e^p (F(\alpha; p, n-p-1))^{p/2}}\right\} \\ &= \frac{1}{2}\log\left(\frac{c_{ii}}{r_{ii}}\right) + \frac{p}{2}\log\left(\frac{n-t_i}{n-1}\right) + \frac{p}{2}\log\left(\frac{F(\alpha; p, n-p)}{F(\alpha; p, n-p-1)}\right). \end{aligned}$$

If CW_i is negative (positive) then the volume of the confidence ellipsoid is decreasing (increasing) and increases (decreases) precision after deleting the ith observation. Regardless of CW_i's sign, a large $|CW_i|$ indicates a strong influence on $\hat{\beta}$ or/and $var(\hat{\beta})$. The constant $(p/2)\log(F(\alpha; p, n-p)/F(\alpha; p, n-p-1))$ does not affect the detection of an influential observation, but it plays important role in determining the sign of CW_i. Clearly the Cook-Weisberg statistic, CW_i, can be simplified as

$$CW_i = \frac{1}{2}\log(VR_i(\beta)) + \frac{p}{2}\log\left(\frac{F(\alpha; p, n-p)}{F(\alpha; p, n-p-1)}\right).$$

Thus apart from the F values, CW_i is identical to $VR_i(\beta)$. In fact, for large n, the ratio of the F values is close to 1, and hence $CW_i \approx \frac{1}{2}\log(VR_i(\beta))$.

4.2.4 Analogue of the Andrews Pregibon statistic

Yet another measure based on the volume of the confidence ellipsoid is the Andrews Pregibon statistic:

$$AP_i = -\frac{1}{2}\log\left\{\frac{\mathbf{e}'_{(i)}\mathbf{e}_{(i)}\det(W_{[i]}(\mathbf{X},\mathbf{X}))}{\mathbf{e}'\mathbf{e}\det(W(\mathbf{X},\mathbf{X}))}\right\}, \quad i = 1, 2, \ldots, n.$$

Note that

$$\begin{aligned}\mathbf{e}_{(i)} &= \mathbf{y}_{(i)} - \mathbf{X}_{(i)}\hat{\beta}_{(i)} - \mathbf{Z}_{(i)}\tilde{\mathbf{u}}_{(i)} \\ &= \mathbf{y}_{(i)} - \mathbf{X}_{(i)}\hat{\beta} - \mathbf{Z}_{(i)}\tilde{\mathbf{u}} \\ &\quad + (\mathbf{X}_{(i)}W(\mathbf{X},\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}_i + \mathbf{Z}_{(i)}\mathbf{D}\mathbf{Z}'\mathbf{R}_i)\frac{e_i}{r_{ii}}.\end{aligned}$$

For $\mathbf{R}_{i(i)}$ and $\mathbf{C}_{i(i)}$ the i-th column of $\mathbf{R}$ and $\mathbf{H}^{-1}$ without the i-th element, respectively,

$$\begin{aligned}\mathbf{Z}_{(i)}\mathbf{D}\mathbf{Z}'\mathbf{R}_i &= (\mathbf{H}_{(i)} - \mathbf{I}_{(i)})\mathbf{H}^{-1}(\mathbf{I}_i - \mathbf{X}W(\mathbf{X},\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}_i) \\ &= -\mathbf{R}_{i(i)} - \mathbf{X}_{(i)}W(\mathbf{X},\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}_i,\end{aligned}$$

which then yields

$$\begin{aligned}\mathbf{e}_{(i)} &= \mathbf{y}_{(i)} - \mathbf{X}_{(i)}\hat{\beta} - \mathbf{Z}_{(i)}\tilde{\mathbf{u}} - \mathbf{R}_{i(i)}\frac{e_i}{r_{ii}}, \quad \text{and} \\ \mathbf{e}'_{(i)}\mathbf{e}_{(i)} &= \mathrm{ssq}(\mathbf{y}_{(i)} - \mathbf{X}_{(i)}\hat{\beta} - \mathbf{Z}_{(i)}\tilde{\mathbf{u}} - \mathbf{R}_{i(i)}(e_i/r_{ii})) \\ &= \mathbf{e}'\mathbf{e} + \frac{e_i^2}{r_{ii}^2}\mathrm{ssq}(\mathbf{R}_i) - \frac{2e_i}{r_{ii}}\mathbf{e}'\mathbf{R}_i = \mathrm{ssq}(\mathbf{e} - \frac{e_i}{r_{ii}}\mathbf{R}_i).\end{aligned}$$

It immediately follows that

$$AP_i = \frac{1}{2}\log\left(\frac{c_{ii}}{r_{ii}}\right) + \frac{1}{2}\log(\mathbf{e}'\mathbf{e}) - \frac{1}{2}\log(\mathrm{ssq}(\mathbf{e} - \frac{e_i}{r_{ii}}\mathbf{R}_i)).$$

Obviously, the larger the statistic AP_i, the stronger the i-th observation's influence on the fit.

4.3 Influence on random effect prediction

The proposed diagnostic measure examines the squared distance from the complete data predictor of the random effects to the i-th case deleted predictor of the random effects, relative to the variance of the random effects, $var(\mathbf{u}) = \sigma_e^2\mathbf{D}$. This is the analogue to the Cook distance (Cook, 1977) and can be written as

$$\begin{aligned} CD_i(\mathbf{u}) &= (\tilde{\mathbf{u}} - \tilde{\mathbf{u}}_{(i)})'\mathbf{D}^{-1}(\tilde{\mathbf{u}} - \tilde{\mathbf{u}}_{(i)})/\hat{\sigma}_e^2 = \frac{t_i^2}{r_{ii}}\mathbf{R}_i'\mathbf{ZDDZ}'\mathbf{R}_i \\ &= \frac{t_i^2}{r_{ii}}\mathbf{R}_i'(\mathbf{H}-\mathbf{I})\mathbf{R}_i = t_i^2\left(1 - \frac{\mathrm{ssq}(\mathbf{R}_i)}{r_{ii}}\right). \end{aligned}$$

A large $CD_i(\mathbf{u})$ indicates that the i-th observation is influential in predicting the random effects.

4.4 Influence on the likelihood function

Because of the computational intractability in estimating the variance components ratios, we consider the estimate of $\theta' = [\beta', \sigma_e^2]$, assuming γ fixed. For $\hat{\theta}(\gamma)$ and $\hat{\theta}_{(i)}(\gamma)$, the maximum likelihood estimates of θ based on n and $n-1$ observations, respectively, let $\lambda(\hat{\theta}(\gamma))$ and $\lambda(\hat{\theta}_{(i)}(\gamma))$ be the log likelihood function evaluated at $\hat{\theta}(\gamma)$ and $\hat{\theta}_{(i)}(\gamma)$, respectively. The distance between the likelihood functions measures the influence of the i-th observation on the likelihood function, giving the likelihood distance LDi (cf Cook and Weisberg, 1982) as

$$\mathrm{LD}_i = 2\{\lambda(\hat{\theta}(\gamma)) - \lambda(\hat{\theta}_{(i)}(\gamma)\}$$

where

$$\begin{aligned} \lambda(\hat{\theta}(\gamma)) &= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\hat{\sigma}_e^2 - \frac{1}{2}\log(\det(\mathbf{H})) - \frac{n}{2} \quad \text{and} \\ \lambda(\hat{\theta}_{(i)}(\gamma)) &= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\hat{\sigma}_{e(i)}^2 - \frac{1}{2}\log(\det(\mathbf{H})) \\ &\quad -(\mathbf{y} - \mathbf{X}\hat{\beta}_{(i)})'\mathbf{H}^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta}_{(i)})/(2\hat{\sigma}_{e(i)}^2). \end{aligned}$$

Furthermore, since

$$(\mathbf{y} - \mathbf{X}\hat{\beta}_{(i)})'\mathbf{H}^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta}_{(i)}) = n\hat{\sigma}_e^2 + \frac{e_i^2}{r_{ii}^2}(c_{ii} - r_{ii}),$$

we obtain

$$\mathrm{LD}_i = n\log\left(\frac{n - t_i^2}{n-1}\right) + \frac{[n + t_i^2(c_{ii} - r_{ii})](n-1)}{r_{ii}(n - t_i^2)} - n.$$

If LD_i is large then the i-th observation is influential on the likelihood function.

4.5 Influence on Linear Functions of $\hat{\beta}$

In some cases we may be interested in assessing the influence of the i-th observation on a given subset of β, or more generally, on s linearly independent combinations of β namely $\psi = \mathbf{L}'\beta$ where $\mathbf{L}'$ is $s \times p$ matrix with $\mathrm{rank}(\mathbf{L}) = s$, using the maximum likelihood estimate of ψ, $\hat{\psi} = \mathbf{L}'\hat{\beta}$. Omitting the i-th observation, the MLE of ψ is $\hat{\psi}_{(i)} = \mathbf{L}'\hat{\beta}_{(i)}$. Using the generalized Cook's statistic as a measure of the influence of the i-th observation on $\hat{\psi}$ we obtain

$$\begin{aligned}
\mathrm{CD}_i(\psi) &= (\hat{\psi} - \hat{\psi}_{(i)})'(\mathbf{L}'W(\mathbf{X},\mathbf{X})\mathbf{L})^{-1}((\hat{\psi} - \hat{\psi}_{(i)})/(s\hat{\sigma}_e^2) \\
&= (\hat{\beta}_{(i)} - \hat{\beta})'\mathbf{M}(\hat{\beta}_{(i)} - \hat{\beta})/(s\hat{\sigma}_e^2) \\
&= \mathbf{C}_i'\mathbf{X}W(\mathbf{X},\mathbf{X})^{-1}\mathbf{M}W(\mathbf{X},\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}_i t_i^2/(sr_{ii}),
\end{aligned}$$

where $\mathbf{M} = \mathbf{L}(\mathbf{L}'W(\mathbf{X},\mathbf{X})^{-1}\mathbf{L})^{-1}\mathbf{L}'$. For positive definite $\mathbf{M}$, we have

$$\begin{aligned}
&\mathbf{C}_i{}'\mathbf{X}W(\mathbf{X},\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}_i - \mathbf{C}_i'\mathbf{X}W(\mathbf{X},\mathbf{X})^{-1}\mathbf{M}W(\mathbf{X},\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}_i \\
&= \mathbf{C}_i{}'\mathbf{X}W(\mathbf{X},\mathbf{X})^{-1/2}[\mathbf{I} - W(\mathbf{X},\mathbf{X})^{-1/2}\mathbf{M}W(\mathbf{X},\mathbf{X})^{-1/2}]W(\mathbf{X},\mathbf{X})^{-1/2}\mathbf{X}'\mathbf{C}_i \\
&\geq 0
\end{aligned}$$

and

$$\mathbf{C}_i'\mathbf{X}W(\mathbf{X},\mathbf{X})\mathbf{X}'\mathbf{C}_i t_i^2/(sr_{ii}) \geq \mathbf{C}_i'\mathbf{X}W(\mathbf{X},\mathbf{X})^{-1}\mathbf{M}W(\mathbf{X},\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}_i t_i^2/(sr_{ii}),$$

so that

$$\begin{aligned}
\mathrm{CD}_i(\beta)p/s &= (c_{ii} - r_{ii})t_i^2/(sr_{ii}) \\
&\geq \mathbf{C}_i'\mathbf{X}W(\mathbf{X},\mathbf{X})^{-1}\mathbf{M}W(\mathbf{X},\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}_i t_i^2/(sr_{ii}) \\
&= \mathrm{CD}_i(\psi)
\end{aligned}$$

where p/s is fixed. This means that $\mathrm{CD}_i(\psi)$ does not need to be calculated unless $\mathrm{CD}_i(\beta)$ is large.

One can further simplify $\mathrm{CD}_i(\psi)$ by imposing some constraints on $\mathbf{L}$. If our interest is centered on the influence of the i-th observation on t elements of β (which we assume, without loss of generality, to be the last t components of β) we have $\mathbf{L}' = [\mathbf{0} \,|\, \mathbf{I}_t]$ where $\mathbf{0}$ is a $t \times (p-t)$ matrix of zeros. Partitioning $\mathbf{X}$ into $\mathbf{X} = [\mathbf{X}_1 \,|\, \mathbf{X}_2]$ we obtain

$$W(\mathbf{X},\mathbf{X})^{-1} = \begin{bmatrix} W(\mathbf{X}_1,\mathbf{X}_1) & W(\mathbf{X}_1,\mathbf{X}_2) \\ W(\mathbf{X}_2,\mathbf{X}_1) & W(\mathbf{X}_2,\mathbf{X}_2) \end{bmatrix}^{-1},$$

and using the standard matrix identity for the inverse of a partitioned matrix we further simplify

$$\begin{aligned}
& W(\mathbf{X},\mathbf{X})^{-1}\mathbf{M}W(\mathbf{X},\mathbf{X})^{-1} \\
&= W(\mathbf{X},\mathbf{X})^{-1}\mathbf{L}(\mathbf{L}'W(\mathbf{X},\mathbf{X})^{-1}\mathbf{L})^{-1}\mathbf{L}'W(\mathbf{X},\mathbf{X})^{-1} \\
&= W(\mathbf{X},\mathbf{X})^{-1} - \begin{bmatrix} W(\mathbf{X}_1,\mathbf{X}_2)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}
\end{aligned}$$

from which it follows that

$$\begin{aligned}
\mathrm{CD}_i(\psi) &= \frac{t_i^2}{tr_{ii}}(c_{ii} - r_{ii} - \mathbf{C}_i'\mathbf{X}_1 W(\mathbf{X}_1,\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{C}_i) \\
&= \frac{p}{t}\mathrm{CD}_i(\beta)(1 - \frac{1}{c_{ii} - r_{ii}}\mathbf{C_i}'\mathbf{X}_1 W(\mathbf{X}_1,\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{C}_i),
\end{aligned}$$

which measures the influence of the i-th observation on the last t elements of β.

If our interest is centered on the influence of one particular effects' level, namely β_k, $t = 1, \mathbf{L}' = [\mathbf{0}\,|\,1]$, $\mathbf{X}_1 = \mathbf{X}_{[k]}$ and $\mathbf{X}_2 = X_k$ and

$$\mathrm{CD}_i(\psi) = p\mathrm{CD}_i(\beta)(1 - \frac{1}{c_{ii} - r_{ii}}\mathbf{C_i}'\mathbf{X}_{[k]} W(\mathbf{X}_{[k]},\mathbf{X}_{[k]})^{-1}\mathbf{X}_{[k]}'\mathbf{C}_i),$$

which measures the influence of the i-th observation on the k-th element of β.

5. Application

5.1 Aerosol data

Beckman *et al.* (1987) developed local influence diagnostics for the mixed model and illustrated their methodology by using aerosol penetration data. CPJ used the same data for comparison and illustration of their diagnostic methods. We use the same data.

The aerosol data involves high-efficiency particulate air filter cartridges used in commercial respirators to prevent or reduce the respiration of toxic fumes, radionuclides, dusts, mists, and other particulate matter. The aim is to determine the factors that contribute most to the variability in penetration of the filters and to determine whether the standard aerosol can be replaced by an alternative aerosol in quality assurance testing. Filters fail a quality test if the percentage of penetration of an aerosol is too large. Two aerosols were crossed with two filter manufacturers. Within each manufacturer, three filters were used to evaluate the penetration of the two aerosols. The aerosols and manufacturers were taken as fixed effects and the filters nested within manufacturers were taken as a random effect. The data are presented in Table 5.1.

Figures 5.1-5.4 give the index plots of the influence measures. Cases 33 and 31 stand out as the most influential on the filter variance ratio, $\mathrm{CD}(\gamma)$, with

Table 5.1: Percent penetration of two aerosols using three filters from manufacturers

	Manufacturer 1 ($j = 1$)			Manufacturer 2 ($j = 2$)		
Filter 1 ($k = 1$)	Filter 2 ($k = 2$)	Filter 3 ($k = 3$)	Filter 1 ($k = 1$)	Filter 2 ($k = 2$)	Filter 3 ($k = 3$)	
Aerosol 1 ($i = 1$)	0.750	0.082	0.082	0.600	4.100	1.000
	0.770	0.085	0.076	0.680	5.000	1.800
	0.840	0.096	0.077	0.870	1.800	2.700
Aerosol 2 ($i = 2$)	0.910	0.660	2.170	1.120	0.160	0.150
	0.830	0.830	1.520	1.100	0.110	0.120
	0.950	0.610	1.580	1.120	0.260	0.120

Figure 5.1 Index plot of CD(γ) and |IR-1|

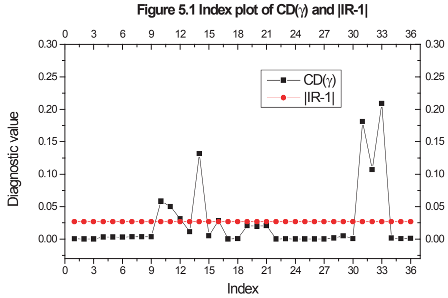


Figure 5.2 Index plot of CD(β), |VR-1|, |CW| and AP

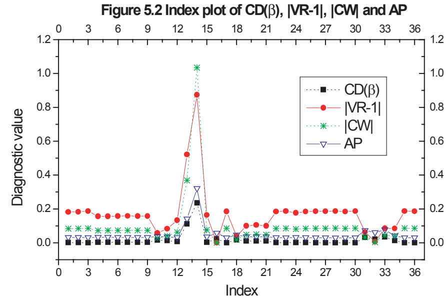


Figure 5.3 Index plot of CD(**u**)

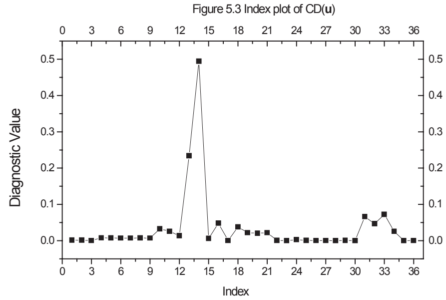
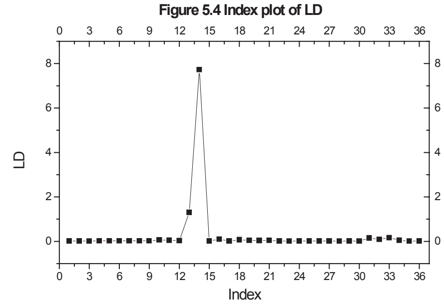


Figure 5.4 Index plot of LD



cases 14 also being indicated as influential. However, Beckman *et al.* and CPJ ranked case 14 as the most influential observation on the filter variance, with cases 13, 31, 32 and 33 regarded as somewhat less influential on the filter variance.

Table 5.2: Maximum likelihood filter variacne ratio estimate for aerosol data

	MLE	% of change
Full data	0.21516832	
Case 33 deleted	0.34529161	60.475%
Case 31 deleted	0.33581581	56.071%
Case14 deleted	0.31175359	44.888%

Figure 5.5 Index plot of CD(β), |VR-1|, |CW| and AP when case 33 is deleted

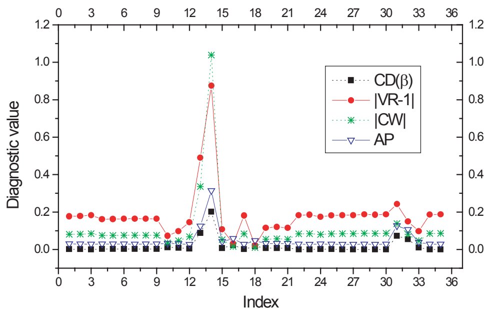


Figure 5.6 Index plot of CD(**u**) when case 33 is deleted

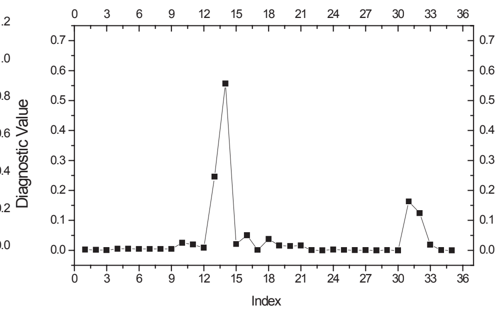


Figure 5.7 Index plot of LD when case 33 is deleted

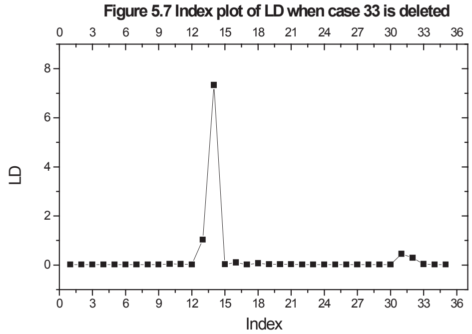


Table 5.2 gives the variance ratio estimate for cases 33, 31 and 14 deleted. The result indicates that case 33 is the most influential point while case 14 is the third most influential case.

Figures 5.2-5.4 show the index plots of the influence measures for fixed effects, prediction of random effects and the log-likelihood distance, respectively. In all of these, case 14 is by far the most influential point. Case 13 is clearly the second most influential case. In the overall analyses of aerosol data cases 13 and 14 have a large influence. However, before dealing with the influential cases on the

Table 5.3: Metallic oxide analysis data

Raw material	Lot	Sample 1				Sample 2			
		Chemist 1		Chemist 2		Chemist 1		Chemist 2	
Type 1	1	4.1	4.0	4.3	4.3	4.1	4.0	4.1	4.1
	2	4.1	4.0	4.0	3.9	4.2	4.2	3.7	4.6
	3	3.5	3.5	3.4	3.6	3.4	3.3	4.0	3.5
	4	4.2	4.2	4.2	4.3	4.1	3.7	4.1	4.6
	5	3.7	3.8	3.3	3.3	3.2	3.1	3.1	3.2
	6	4.0	4.2	3.8	4.2	4.1	4.3	4.2	4.1
	7	4.0	3.8	3.8	4.0	3.6	3.8	3.9	3.8
	8	3.8	3.9	4.0	3.9	4.0	4.0	4.2	4.0
	9	4.2	4.5	4.3	4.1	3.8	3.7	3.8	3.8
	10	3.6	4.0	4.0	3.7	3.9	4.1	4.2	3.7
	11	4.6	4.6	4.0	3.4	4.4	4.5	3.9	4.1
	12	3.3	2.9	3.2	3.9	2.9	3.7	3.3	3.4
	13	4.5	4.5	4.0	4.2	3.7	4.0	4.0	3.9
	14	3.8	3.8	3.5	3.6	4.3	4.1	3.8	3.8
	15	4.2	4.1	3.8	3.8	3.8	3.8	3.9	3.9
	16	4.2	3.4	3.7	4.1	4.4	4.5	4.0	4.0
	17	3.3	3.4	3.9	4.0	2.2	2.3	2.4	2.7^b
	18	3.6	3.7	3.6	3.5	4.1	4.0	4.4	4.2
Type 2	1	3.4	3.4	3.6	3.5	3.7	3.5	3.1	3.4
	2	4.2	4.1	4.3	4.2	4.2	4.2	4.3	4.2
	3	3.5	3.5	4.2	4.5	3.4	3.7	3.9	4.0
	4	3.4	3.3	3.5	3.1	4.2	4.2	3.3	3.1
	5	3.2	2.8	3.1	2.7	3.0	3.0	3.2	2.7
	6	0.2	0.7	0.8	0.7	0.3	0.4	0.2	−1.0
	7	0.9	0.6	0.3	0.6	1.0	1.1	0.7	1.0
	8	3.3	3.5	3.5	3.4	3.9	3.7	3.7	3.7
	9	2.9	2.6	2.8	2.9	3.1	3.1	2.9	2.7
	10	3.8	3.8	3.9	3.8	3.4	3.6	4.0	3.8
	11	3.8	3.4	3.6	3.8	3.8	3.6	3.9	4.0
	12	3.2	2.5	3.0	3.5	4.3	3.7	3.8	3.8
	13	3.4	3.4	3.3	3.3	3.5	3.5	3.2	3.3

Note: Table value is metal content minus 80 in percent by weight

fixed/random effects and the log likelihood function, scrutinizing influential points on the variance components ratio seems essential.

Accordingly, case 33 was deleted and Figures 5.5-5.7 present the diagnostic

Table 5.4: The Top 5 influential cases

ID	$CD_i(\gamma)$	ID	$CD_i(\beta)$
(2,6,2, 2,1)	0.468932	(2,6,2,2,2)	0.00118
(1,2,2,2,2)	0.129775	(2,12,1,1,2)	0.00041
(1,11,1,2,2)	0.125141	(2,12,2,1,1)	0.00041
(1,2,2,2,1)	0.117674	(1,11,1,2,2)	0.00029
(2,12,2,1,1)	0.101159	(1,2,2,2,2)	0.00026
ID	$\lvert CR_i - 1\rvert$	ID	$\lvert CW_i\rvert$
(2,6,2,2,2)	0.26717	(2,6,2,2,2)	0.155420
(2,12,1,1,2)	0.09283	(2,12,1,1,2)	0.048712
(2,12,2,1,1)	0.09134	(2,12,2,1,1)	0.047890
(1,11,1,2,2)	0.09103	(1,11,1,2,2)	0.047723
(1,2,2,2,2)	0.07938	(1,2,2,2,2)	0.041354
ID	AP_i	ID	$CD_i(u)$
(2,6,2,2,2)	0.06554	(2,6,2,2,2)	3.15971
(2,6,2,2,1)	0.03450	(2,12,1,1,2)	1.11102
(1,2,2,2,1)	0.02850	(2,12,2,1,1)	1.09466
(1,2,2,2,2)	0.02820	(1,11,1,2,2)	1.09466
(2,12,2,1,1)	0.02475	(1,2,2,2,2)	1.0912
ID	LD_i		
(2,6,2,2,2)	1.38359		
(2,12,1,1,2)	0.13386		
(2,12,2,1,1)	0.12939		
(1,11,1,2,2)	0.12803		
(1,2,2,2,2)	0.09615		

The order in the ID is (raw material type, lot, sample, chemist, duplicate analysis number).

measures for the fixed/random effects, the response variable and the log likelihood function without case 33. Many of the general features of Figure 5.5 are similar to the full data plot Figure 5.2. Likewise, Figure 5.6 and Figure 5.3, and Figure 5.7 and Figure 5.4 are similar. It seems that case 33 has little effect on the influence of each observation on the fixed/random effects and the log-likelihood function, while it has a noticeable effect on the variance components ratio.

5.2 Metallic oxide data

Fellner (1987) introduced the robust estimation of variance components and illustrated his methodology by using metallic oxide data. The metallic oxide analysis data given by Fellner (1987) were taken from a sampling study designed to explore the effects of process and measurement variation on the properties of lots of metallic oxide. The data are shown in Table 5.3. Two samples were drawn from each lot. Duplicate analyses were then performed by each of two chemists, with a pair of chemists being randomly selected for each sample.

A nested model was fitted, with raw material type considered fixed, and lot, sample and chemist considered random.

The top five influential points on variance components ratios, fixed effect parameters, prediction of random effects and likelihood function are given in Tables 5.4. The order in the identification is (raw material type, lot, sample, chemist, duplicate analysis number). Observation (2,6,2,2,1) is the most influential on the variance components ratios estimate. On the other hand, observation (2,6,2,2,2) is by far the most influential on the fixed and random effects estimate and likelihood function. There is no large second most influential observation. Since in a balanced design the value of $IR_i(\gamma)$ is identical for all i, $IR_i(\gamma)$ did not indicate any influential point on the variance of variance components ratios estimate.

Table 5.5 gives the maximum likelihood variance components ratios estimates with the full data and with case (2,6,2,2,1) removed. But case (2,6,2,2,2) is not influential on the variance components ratios, even if it is an outlier and influential on the fixed and random effects estimate and the likelihood function. Outliers need not necessarily be influential. In order to clear the doubt we also include the estimates when case (2,6,2,2,2) is removed.

Table 5.5: Metallic oxide data: Maximum likelihood estimate of variance components ratios from full-data and the data with cases (2,6,2,2,1) and (2,6,2,2,2) removed

	Full data	(a)	% change	(b)	% change
Lot	13.0791	14.9490	14.30%	14.2893	9.25%
Sample	1.0019	1.1538	15.17%	1.0774	7.54%
Chemist	0.7379	1.0174	37.88%	0.8031	8.84%

(a)= case (2,6,2,2,1) deleted; (b)= case (2,6,2,2,2) deleted.

The estimates of the variance components ratios are small with the full data. Deleting case (2,6,2,2,1) results in an increase in the variance components ratios of the lot effect, sample effect and chemist effect far larger than that due to deleting case (2,6,2,2,2). Undoubtedly the cumulative increase in the variance components ratios will have a substantial effect on the overall fit. Thus case (2,6,2,2,1) should be put under scrutiny.

We further assessed if the inclusion or exclusion of case (2,6,2,2,1), which is the most influential case on the variance components ratio, does have a remarkable effect on the influence measures of the fixed/random effects and the log-likelihood function. The result shows that the top five influential cases under CD$beta$), $|VR-1|$, $|CW|$, CD($\mathbf{u}$) and LD remained as the top five influential cases and case (2,6,2,2,2) is by far the most influential point on the fixed and random effects estimate and likelihood function. But under AP statistic cases (1,2,2,2,1), (1,2,2,2,2) and (2,12,2,1,1), become the top three influential points and case (2,6,2,2,2) becomes the seventh influential observation. This is, perhaps the removed case and case (2,6,2,2,2) belongs in the same cell and the removed case was also the second top influential case under AP statistic.

6. Concluding Remarks

We have presented various diagnostic measures that appear useful and can play an important role in linear mixed model data analysis. All the diagnostic measures are a function of the basic building blocks: Studentized residuals, error contrast matrix, and the inverse of the response variable covariance matrix. The basic building blocks are computed only once from the complete data analysis.

Although no formal cutoff points are presented for these measures, it appears that relative comparisons such as rankings or simple index plots are a promising and practical approach to pinpoint influential observations. The basic building blocks, particularly $\mathbf{C}$, $\mathbf{R}$ and $\mathbf{e}$ play an important role in the mixed model diagnostics. Characterization of the distributions and properties of $\mathbf{C}$ and $\mathbf{R}$ are not studied. This is a bottleneck for distributional properties and cutoff point for the diagnostic measures. This is an issue for future research.

Generally we fit a linear mixed model by specifying the covariance structure of $\mathbf{u}$ and ϵ. Here we have assumed that $\mathbf{D}$ is block diagonal with the i-th block being $\gamma_i \mathbf{I}_{q_i}$. In the case where $\mathbf{D}$ is a known unstructured covariance matrix, i.e., the covariance of $\mathbf{u}$ is arbitrary positive definite matrix, the diagnostic measures for the fixed effects, random effects and likelihood function derived in this study would be used as they are. However, if $\mathbf{D}$ is unknown, extending our diagnostics to assess influence on such unstructured covariance estimates is an area of future research.

Acknowledgements

The work of the first author is supported by German Academic Exchange Service (DAAD) and is grateful for the support.

Appendix: Proof of Theorems

Proof of theorem 1

$W(\mathbf{A},\mathbf{B}) = [\mathbf{a}_i, bfA'_{(i)}]\mathbf{H}^{-1}[\mathbf{b}'_i, \mathbf{B}_{(i)}]'$. Applying the standard matrix identity for the inverse of a partitioned matrix $\mathbf{H}$, we obtain

$$\begin{aligned}
W(\mathbf{A},\mathbf{B}) &= [\mathbf{a}_i, \mathbf{A}'_{(i)}] \begin{bmatrix} 1/m_i & -\mathbf{h}'_i\mathbf{H}^{-1}_{[i]}/m_i \\ -\mathbf{h}'_i\mathbf{H}^{-1}_{[i]}/m_i & \mathbf{H}^{-1}_{[i]} + \mathbf{H}^{-1}_{[i]}\mathbf{h}_i\mathbf{h}'_i\mathbf{H}^{-1}_{[i]}/m_i \end{bmatrix} \begin{bmatrix} \mathbf{b}'_i \\ \mathbf{B}_{(i)} \end{bmatrix} \\
&= \mathbf{a}_i\mathbf{b_i}'/m_i - \mathbf{A}'_{(i)}\mathbf{H}^{-1}_{[i]}\mathbf{h}_i\mathbf{b}'_i/m_i - \mathbf{a}_i\mathbf{h}'_i\mathbf{H}^{-1}_{[i]}\mathbf{B}_{(i)}/m_i \\
&\quad + \mathbf{A}'_{(i)}\mathbf{H}^{-1}_{[i]}\mathbf{B}_{(i)} + \mathbf{A}'_{(i)}\mathbf{H}^{-1}_{[i]}\mathbf{h}_i\mathbf{h}'_i\mathbf{H}^{-1}_{[i]}\mathbf{B}_{(i)}/m_i \\
&= \mathbf{A}'_{(i)}\mathbf{H}^{-1}_{[i]}\mathbf{B}_{(i)} + \mathbf{a}_i(\mathbf{b}'_i - \mathbf{h}'_i\mathbf{H}^{-1}_{[i]}\mathbf{B}_{(i)})/m_i \\
&\quad - \mathbf{A}'_{(i)}\mathbf{H}^{-1}_{[i]}\mathbf{h}_i(\mathbf{b}'_i - \mathbf{h}'_i\mathbf{H}_{[i]}\mathbf{B}_{(i)})/m_i \\
&= \mathbf{A}'_{(i)}\mathbf{H}^{-1}_{[i]}\mathbf{B}_{(i)} + (\mathbf{a}_i - \mathbf{A}'_{(i)}\mathbf{H}^{-1}_{[i]}\mathbf{h}_i)(\mathbf{b}'_i - \mathbf{h_i}'\mathbf{H}^{-1}_{[i]}\mathbf{B}_{(i)})/m_i \\
&= \mathbf{A}'_{(i)}\mathbf{H}^{-1}_{[i]}\mathbf{B}_{(i)} + \breve{\mathbf{a}}_i\breve{\mathbf{b}}'_i/m_i = W_{[i]}(\mathbf{A},\mathbf{B}) + \breve{\mathbf{a}}_i\breve{\mathbf{b}}'_i/m_i \qquad \square
\end{aligned}$$

Proof of theorem 2.

From the inverse of a partitioned matrix $\mathbf{H}$ and $\mathbf{H}^{-1}$, and the uniqueness property of the inverse of a positive definite matrix,

$$\begin{bmatrix} c_{ii} & \mathbf{c}'_i \\ \mathbf{c}_i & \mathbf{C}_{[i]} \end{bmatrix} = \begin{bmatrix} 1/m_i & -\mathbf{h}'_i\mathbf{H}^{-1}_{[i]}/m_i \\ -\mathbf{H}^{-1}_{[i]}\mathbf{h}_i/m_i & \mathbf{H}^{-1}_{[i]} + \mathbf{H}^{-1}_{[i]}\mathbf{h}_i\mathbf{h}'_i\mathbf{H}^{-1}_{[i]}/m_i \end{bmatrix}.$$

Then $m_i = 1/c_{ii}$ follows immediately. Further, $\mathbf{c}_i = -\mathbf{H}^{-1}_{[i]}\mathbf{h}_i/m_i$, $\mathbf{c}_i/c_{ii} = -\mathbf{H}^{-1}_{[i]}\mathbf{h}_i$,

$$\begin{aligned}
\breve{\mathbf{x}}_i &= \mathbf{x}_i - \mathbf{X}'_{(i)}\mathbf{H}^{-1}_{[i]}\mathbf{h}_i = \mathbf{x}_i + \frac{1}{c_{ii}}\mathbf{X}'_{[i]}\mathbf{c}_i \\
&= \frac{1}{c_{ii}}(c_{ii}\mathbf{x}_i + \mathbf{X}'_{[i]}\mathbf{c}_i) = \frac{1}{c_{ii}}[\mathbf{x}_i, \mathbf{X}'_{[i]}] \begin{bmatrix} c_{ii} \\ \mathbf{c}_i \end{bmatrix} = \frac{1}{c_{ii}}\mathbf{X}'\mathbf{C}_i.
\end{aligned}$$

Similarly,

$$\begin{aligned}\breve{\mathbf{z}}_{ji} &= \mathbf{z}_{ji} - \mathbf{Z}'_{j(i)}\mathbf{H}_{[i]}^{-1}\mathbf{h}_i = \frac{1}{c_{ii}}\mathbf{Z}'_j\mathbf{C}_{\mathbf{i}}, \quad \text{and} \\ \breve{y}_i &= y_i - \mathbf{y}'_{(i)}\mathbf{H}_{[i]}^{-1}\mathbf{h}_i = \frac{1}{c_{ii}}\mathbf{y}'\mathbf{C}_i \ . \qquad \square\end{aligned}$$

References

Andrews, D. F. and Pregibon, D. (1978). Finding outliers that matter. *Journal of the Royal Statistical Society, Series B* **40**, 85-93.

Banerjee, M. and Frees, E. W. (1997). Influence diagnostics for linear longitudinal models. *Journal of the American Statistical Association* **92**, 999-1005.

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

Belsley, D. A., Kuh, E. and Welsch, R. E. (1980). *Regression Diagnostics.* Wiley.

Chatterjee, S. and Hadi, A. S. (1986). Influential observation, high leverage points, and outliers in linear regression (with discussion). *Statistical Science***1**, 379-416.

Chatterjee, S. and Hadi, A. S., (1988). *Sensitivity Analysis in Linear Regression.* Wiley.

Christensen, R., Johnson, W. and Pearson, L. M. (1992). Prediction diagnostics for spatial linear models. *Biometrika***79**, 583-591.

Christensen, R., Johnson, W. and Pearson, L. M. (1993). Covariance function diagnostics for spatial linear models. *Mathematical Geology* **25**, 145-160.

Christensen, R., Pearson, L. M. and Johnson, W. (1992). Case deletion diagnostics for mixed models. *Technometrics* **34**, 38-45.

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). *Journal of the Royal Statistical Society, Series B* **48**, 133-169.

Cook, R. D. and Weisberg, S. (1980). Characterization of an empirical influence function for detecting influential cases in regression. *Technometrics* **22**, 495-508.

Cook, R. D. and Weisberg, S. (1982). *Residual and Influence in Regression.* Chapman and Hall, London.

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

Fellner, W. H. (1986). Robust estimation of variance components. *Technometrics* **28**, 51-60.

Goodnight, J. H. (1978). Computing MIVQUE0 estimates of variance components. Technical Report R-105, SAS Institute Inc.

Hartley, H. O. and Rao, J. N. K. (1967). Maximum likelihood estimation of the mixed analysis of variance model. *Biometrika* **54**, 93-108.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems (with discussion). *Journal of the American Statistical Association* **72**, 320-340.

Haslett, J. and Hayes, K. (1998). Residuals for the linear model with general covariance structure. *Journal of the Royal Statistical Society, Series B* **60**, 201-215.

Haslett, S. (1999). A simple derivation of deletion diagnostics for the general linear model with correlated errors. *Journal of the Royal Statistical Society, Series B* **61**, 603-609.

Hemmerle, W. J. and Hartley, H. O., 1973. Computing maximum likelihood estimates for the mixed AOV model using the W transformation. *Technometrics* **15**, 819-831.

Jennrich, R. I. and Sampson, P. F., 1976. Newton-Raphson and related algorithms for maximum likelihood estimation of variance components. *Technometrics* **18**, 11-18.

Lawrance, A. J. (1990). Local and deletion influence. In *Directions in Robust Statistics and Diagnostics, Part I* (Edited by W. A. Stahel and S. Weisberg), Springer Verlag. 141-157.

Littell, R. C., Milliken, G. A., Stroup, W. W., and Wolfinger, R. D. (1996). SAS system for mixed models. SAS Institute Inc.

Martin, R. J. (1992). Leverage, influence and residuals in regression models when observations are correlated. *Communications in Statistics, Theory and Methods* **21**, 1183-1212.

McCulloch, C. E. and Searle, S. R. (2001). *Generalized, Linear and Mixed Models.* Wiley.

Oman, S. D. (1995). Checking the assumptions in mixed model analysis of variance: A residual analysis approach. *Computational Statistics and Data Analysis* **20**, 309-330.

Pregibon, D. (1981). Logistic regression diagnostics. *Annals of Statistics* **9**, 705-724.

Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects (with discussion). *Statistical Science* **6**, 15-51.

SAS Institute Inc. (1992). *SAS Technical Report*, p. 229. SAS Institute, Cary, NC.

Searle, S. R., Casella, G. and McCulloch, C. E. (1992). *Variance Components.* Wiley.

Verbeke G. and G. Molenberghs (1997). *Linear Mixed Models in Practice: A SAS Oriented Approach.* Springer-Verlag.

Received February 4, 2004; accepted May 2, 2004.

Temesgen Zewotir (corresponding author)
School of Statistics and Actuarial Science
University of KwaZulu-Natal
Pietermaritzburg Campus
Private Bag X01, Scottsville 3209, SOUTH AFRICA.
zewotir@ukzn.ac.za

Jacky S Galpin
School of Statistics and Actuarial Science
University of the Witwatersrand
Private Bag 3, P. O. Wits 2050, SOUTH AFRICA.
jacky@galpin.co.za