NIASRA
NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

National Institute for Applied Statistics Research Australia

University of Wollongong, Australia

Working Paper

18-21

An Efficient Resampling Scheme for Outlier Detection in Linear Mixed Models

Alison B. Smith and Brian R. Cullis

National Institute for Applied Statistics Research Australia, University of Wollongong, Wollongong NSW 2522, Australia Phone +61 2 4221 5076, Fax +61 2 4221 4998. Email: karink@uow.edu.au

An efficient resampling scheme for outlier detection in linear mixed models

Alison B. Smith and Brian R. Cullis

[a] *Centre for Bioinformatics and Biometrics, National Institute for Applied Statistics Research Australia, University of Wollongong, Wollongong, NSW, Australia*

Abstract

The alternative outlier model, which was developed in the context of ordinary linear models, assumes that outliers arise from an error term with inflated variance. We generalise this for the linear mixed model framework and derive score tests for inflated variance based on the null (no outlier) model. The tests are applicable for residuals and also for other random effects in the model. Correlated effects can be accommodated. In order to address multiple testing issues we propose a parametric bootstrap approach to obtain thresholds for the score tests. Since full parametric bootstrap methods are computationally demanding, we propose an efficient resampling scheme that does not involve re-fitting the model in each iteration. A simulation study shows that our approach provides accurate Type I error rates for the residual outlier tests.

Keywords:
Alternative outlier model, Residual Maximum Likelihood, Score tests, Multiple testing

1. Introduction

The linear mixed model has become a widely used tool for data analysis in a range of applications including the agricultural, biological, medical and environmental sciences. Model checking using data driven diagnostics is as important for the linear mixed model as for the ordinary linear model. A frequently occurring problem is the existence of outliers. The term outlier has been used in various and sometimes ambiguous ways in the literature. In this paper the definition given in Langford and Lewis (1998) is adopted, namely that "an outlier or outlying observation in the data set is an observation which appears to be inconsistent with the rest of the data, relative to the assumed model." There are two factors that complicate outlier detection in linear mixed models, namely the existence of multiple random terms rather than a single random (error) term and the possibility of correlated rather than independent effects.

There is extensive literature on the detection of outliers for ordinary linear models but only a few papers, including Christensen et al. (1992b,a); Langford and Lewis (1998); Haslett and Hayes (1998); Haslett and Haslett (2007); Gumedze et al. (2010) discuss the problem in the context of linear mixed models. Christensen et al. (1992b) proposed case-deletion diagnostics for detecting influential observations, both in terms of the estimation of fixed effects and variance components in a linear mixed model. Their approach was developed with respect to a model with multiple random terms, each of which (including the error term) comprised a set of independently and identically distributed effects. The work was extended in Christensen et al. (1992a) for detecting influential observations when interpolating from a spatial linear model with a general covariance structure. Both Christensen et al. (1992b) and Christensen et al. (1992a) provided case-deletion diagnostics that could be used to rank observations in terms of their influence. No attempt was made to determine a threshold value.

Langford and Lewis (1998) considered outlier detection for multilevel data that may be analysed using a linear mixed model in which distinct sets of random effects are used to represent the error terms for each level. They pointed out the complexity of outlier detection for such data since "... one has to consider, for example, at what level or levels a particular response is outlying ...". Langford and Lewis (1998) used diagnostics derived for ordinary linear models and applied them to estimates of the errors for any particular level.

The most general framework for outliers in a linear mixed model is presented in Haslett and Hayes (1998); Haslett and Haslett (2007). These papers focussed on the definition and role of residuals (that is, estimates of error) in linear mixed models with a general covariance structure. In terms of their role Haslett and Haslett (2007) stated that "... residuals are widely recommended in the context of model criticism. This includes the identification of outlying observations ...". In terms of the definition they noted that there are ambiguities in what is meant by the term residual when the data are correlated. Haslett and Haslett (2007) suggested that for the linear mixed model there are three fundamental types of scalar residuals and they call these the marginal, model-specified and full-conditional residuals. In the remainder of this paper, and as done in Haslett and Haslett (2007), the latter will be abbreviated to conditional residuals. Haslett and Hayes (1998) and Haslett and Haslett (2007) showed the importance of the role of conditional residuals in outlier diagnostics, in particular deletion diagnostics.

Gumedze et al. (2010) specifically considered variance component models, that is, linear mixed models in which each set of random effects, and the residuals, are assumed independent. They used a variance shift outlier model (also known as an alternative outlier model) and proposed likelihood ratio and score test statistics to determine whether individual observations have inflated variance. A (full) parametric bootstrap procedure was used to obtain thresholds and to account for multiple testing. In the variance component model setting this approach appears to provide reasonable Type I error rates but as Gumedze et al. (2010) point out, the bootstrap procedure is computationally demanding.

In this paper we address the issue of outlier detection using a similar framework to Gumedze et al. (2010), namely an extension of the alternative outlier model discussed by Cook et al. (1982) for ordinary linear models. In this approach outliers are assumed to arise from an error term with inflated variance. Cook et al. (1982) used maximum likelihood (ML) estimation for variance parameters (including the extra variance associated with potential outliers). Thompson (1985) considered the same model but used Residual Maximum Likelihood (REML, Patterson and Thompson, 1971) estimation for variance parameters. In this paper we extend the alternative outlier model for use in a linear mixed model setting and propose that it be used as a diagnostic tool to detect outliers, either at the error level and/or with respect to other sets of random effects in the model. We do not restrict attention to variance component models but also allow for correlated effects. In order to diagnose an inflated variance for an effect we use a score test based on the fit of the null (no outlier) model. The REML score for an extra variance is shown to be a function of the conditional residuals as defined by Haslett and Haslett (2007). The diagnosis of outliers is likely to involve the sequential examination of all effects in a set (for example all error effects). In our context this implies the conduct of multiple hypothesis tests so that a pointwise significance level for the score test based on a chi-square approximation is inappropriate. We determine a threshold using an efficient resampling scheme that is computationally less demanding than a full parametric bootstrap approach and is similar to that used by Zou et al. (2004).

The paper is arranged as follows. In Section 2 we commence with a brief overview of linear mixed models theory necessary for the development of our outlier detection procedure. In Section 2.2 we describe the alternative outlier model of Cook et al. (1982), that is, in the context of ordinary linear models. Then in Section 2.3 we propose an extension for linear mixed models. The score test procedure is developed in Sections 2.3.1 and 2.3.2 and graphical tools are discussed in Section 2.3.3. The methodology is applied to two examples in Section 3. In Section 4 we give the results of a simulation study to investigate the performance of the score test in terms of Type I error rates. Some concluding remarks are given in Section 5.

2. The linear mixed model

The linear mixed model for an $n \times 1$ data vector $\boldsymbol{y}$ can be written as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\tau} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e} \tag{1}$$

where $\boldsymbol{\tau}$ is the $p \times 1$ vector of fixed effects with associated $n \times p$ design matrix $\boldsymbol{X}$ (assumed to have rank $r_x \leq p$), $\boldsymbol{u}$ is the $b \times 1$ vector of random effects with associated $n \times b$ design matrix $\boldsymbol{Z}$ and $\boldsymbol{e}$ is the $n \times 1$ vector of errors. We assume that

$$\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{e} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \ \theta \begin{bmatrix} \boldsymbol{G} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R} \end{bmatrix} \right) \tag{2}$$

The matrices $\boldsymbol{G}$ and $\boldsymbol{R}$ are functions of unknown parameters. We write $\boldsymbol{G} = \boldsymbol{G}(\boldsymbol{\gamma})$ and $\boldsymbol{R} = \boldsymbol{R}(\boldsymbol{\phi})$ where $\boldsymbol{\gamma}$ and $\boldsymbol{\phi}$ are the parameter vectors associated with the random effects and errors respectively. The vector of random effects is often composed of q sub-vectors $\boldsymbol{u}_i$ so that $\boldsymbol{u} = (\boldsymbol{u}_1^\top \ldots \boldsymbol{u}_q^\top)^\top$. We assume that $\boldsymbol{u}_i$ comprises b_i effects so that $\boldsymbol{u}$ comprises $b = \sum_{i=1}^q b_i$ effects. The sub-vectors are assumed to be independent with $\text{var}\,(\boldsymbol{u}_i) = \theta \boldsymbol{G}_i$ so that $\boldsymbol{G}$ is a block diagonal matrix given by $\oplus_{i=1}^q \boldsymbol{G}_i$.

The variance model in equation (2) is specified with an overall scale parameter, θ. In most applications the variance matrix for the errors can be expressed as a scaled matrix, in which case θ represents the associated unknown scale parameter (sometimes denoted σ^2). All other variance component parameters are then expressed as ratios with respect to θ. In some applications θ is fixed at a value of 1. In the following we allow for complete generality but note that θ has a key role in the procedure for determining thresholds for the outlier score tests.

Under the assumptions in equation (2) we have

$$\boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{X\tau},\ \boldsymbol{H}\right)$$

where $\boldsymbol{H} = \theta\boldsymbol{V}$ and $\boldsymbol{V} = \boldsymbol{ZGZ}^\top + \boldsymbol{R}$.

2.1. Model estimation

The first step in fitting the model in equation (1) is the estimation of the variance parameters which requires calculation of the REML scores for θ and the elements of $\boldsymbol{\kappa} = (\boldsymbol{\gamma}^\top, \boldsymbol{\phi}^\top)^\top$. These are given by

$$\begin{aligned} U(\theta) &= -\tfrac{1}{2}\left[\nu/\theta - \boldsymbol{y}^\top\boldsymbol{Py}/\theta^2\right] && (3) \\ U(\kappa_i) &= -\tfrac{1}{2}\left[\operatorname{tr}\left(\boldsymbol{P}\dot{\boldsymbol{V}}_i\right) - \boldsymbol{y}^\top\boldsymbol{P}\dot{\boldsymbol{V}}_i\boldsymbol{Py}/\theta\right] && (4) \end{aligned}$$

where $\nu = n - r_x$ and $\boldsymbol{P} = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}\left(\boldsymbol{X}^\top\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-}\boldsymbol{X}^\top\boldsymbol{V}^{-1}$ with $\left(\boldsymbol{X}^\top\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-}$ being any generalised inverse of $\left(\boldsymbol{X}^\top\boldsymbol{V}^{-1}\boldsymbol{X}\right)$. The "dot" notation indicates a derivative so that $\dot{\boldsymbol{V}}_i = \partial\boldsymbol{V}/\partial\kappa_i,\ \ i = 1 \ldots n_k$ where n_k is the number of variance parameters in $\boldsymbol{\kappa}$.

The REML estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\kappa}$ are obtained by equating the scores to zero. Note in particular that, given $\boldsymbol{\kappa}$, an estimate of θ can be obtained from equation (3) as

$$\hat{\theta} = \boldsymbol{y}^\top\boldsymbol{Py}/\nu \qquad (5)$$

Best linear unbiased predictions (BLUPs) of the random effects and errors in equation (1) can be obtained as

$$\tilde{\boldsymbol{u}} = \boldsymbol{GZ}^\top\boldsymbol{Py} \quad \text{and} \quad \tilde{\boldsymbol{e}} = \boldsymbol{RPy}$$

It is also informative to consider the (scalar) conditional residuals of Haslett and Hayes (1998) and Haslett and Haslett (2007). If we let $\boldsymbol{d}_i$ be the $n \times 1$ vector with a value of one in position i and zeros elsewhere, the conditional residual for observation $i (= 1 \ldots n)$ is given by

$$\boldsymbol{d}_i^\top\boldsymbol{Py}/p_{ii}$$

where $p_{ii} = \boldsymbol{d}_i^\top\boldsymbol{Pd}_i$ is the i^{th} diagonal element of $\boldsymbol{P}$. Note that the variance for the i^{th} conditional residual is given by θ/p_{ii} so that, as in Haslett and Hayes (1998), we can define a Studentised conditional residual as

$$t_i = \boldsymbol{d}_i^\top\boldsymbol{Py}/\sqrt{\theta p_{ii}}$$

It will be shown in later sections that there is a fundamental link between Studentised conditional residuals and outlier detection in linear mixed models using the alternative outlier model.

2.2. Alternative outlier model

Cook et al. (1982) examined the modelling of a single outlier in an ordinary linear model as arising from an error term with inflated variance. In this model the variance of all observations apart from the i^{th} is assumed to be σ^2 and the i^{th} has variance $(1+\omega_i)\sigma^2$, where $\omega_i > 0$. Using the notation of equation (1) this so-called alternative outlier model (AOM) can be written as

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X\tau} + \boldsymbol{e}_i \\ \text{where } \boldsymbol{e}_i &= \boldsymbol{e} + \boldsymbol{d}_i\delta_i \end{aligned} \tag{6}$$

and $\boldsymbol{d}_i$ is as previously defined. It is assumed that $\text{var}(\boldsymbol{e}) = \theta\boldsymbol{I}_n$ (so that $\theta = \sigma^2$) and δ_i is a random effect with zero mean and variance $\omega_i\theta$. Thus the variance of the i^{th} observation is $(1+\omega_i)\theta$.

In the AOM of equation (6) it is required to estimate $\boldsymbol{\tau}, \theta, \omega_i$ and the position i of the outlier. In practice, Cook et al. (1982) proposed to fix the value of i and use Maximum Likelihood (ML) to obtain estimates of the parameters. Repeating for each of the n possible values of i gives n values of the likelihood function. The position of the outlier is that value of i associated with the largest of these likelihood values and the associated values of $\boldsymbol{\tau}, \theta$ and ω_i are their ML estimates. Thompson (1985) proposed an analogous approach but using REML estimation. As he pointed out, an advantage of using REML is that the observation with the largest Studentised residual will be selected as the outlier whereas under ML this is not necessarily the case.

2.3. Alternative outlier mixed model

We now generalise the approach in Section 2.2 for a linear mixed model and will call this an alternative outlier mixed model (AOMM). The AOM as considered by Cook et al. (1982) and Thompson (1985) is used not only to identify the outlier but also to specifically accommodate it in the analysis as a down-weighted observation. We propose the use of the model purely as a diagnostic tool. The practitioner may then choose an appropriate course of action. The first response would be to ascertain whether the outlier was due to erroneous data and if so, replace with the correct value (if available) or a missing value indicator.

As discussed in Langford and Lewis (1998) outliers in the context of a linear mixed model may arise at any level. There is a natural distinction between outliers associated with the errors (where the variance inflation is linked to $\theta\boldsymbol{R}$) and outliers associated with the random effects (variance inflation linked to $\theta\boldsymbol{G}$). Accordingly we develop a separate AOMM for these two types of effects and label them as AOMM-R and AOMM-G.

2.3.1. AOMM-R

As in Cook et al. (1982) we assume that an outlier arises from an error with inflated variance. Under the null hypothesis of no outliers, the linear mixed model is as in equation (1). Under the alternative hypothesis of the i^{th} error being an outlier ($i = 1 \dots n$), the linear mixed model can be written as

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X\tau} + \boldsymbol{Zu} + \boldsymbol{e}_i \\ \text{where} \;\; \boldsymbol{e}_i &= \boldsymbol{e} + \boldsymbol{d}_i\delta_i \\ \Rightarrow \boldsymbol{y} &= \boldsymbol{X\tau} + \boldsymbol{Zu} + \boldsymbol{d}_i\delta_i + \boldsymbol{e} \end{aligned} \tag{7}$$

where $\boldsymbol{d}_i$ is as previously defined and δ_i is a random effect with zero mean and variance $\theta\omega_i$ where $\omega_i > 0$. Under the AOMM-R of equation (7) the variance matrix for the data vector is given by $\boldsymbol{H} = \boldsymbol{\theta V}$ where

$$\boldsymbol{V} = \boldsymbol{ZGZ}^\top + \omega_i \boldsymbol{d}_i\boldsymbol{d}_i^\top + \boldsymbol{R} \tag{8}$$

To examine whether the i^{th} error effect is an outlier we develop an approximate score test for the null hypothesis of $H_0 : \omega_i = 0$ against $H_a : \omega_i > 0$. As in Verbyla (1993) we base the approach on the REML score for ω_i evaluated at the null value of $\omega_i = 0$. Using equations (4) and (8) this is given by

$$\begin{aligned} U_0(\omega_i) &= -\tfrac{1}{2}\left[\mathrm{tr}\,(\boldsymbol{P}_0\boldsymbol{d}_i\boldsymbol{d}_i^\top) - \boldsymbol{y}^\top\boldsymbol{P}_0\boldsymbol{d}_i\boldsymbol{d}_i^\top\boldsymbol{P}_0\boldsymbol{y}/\theta_0\right] \\ &= -\tfrac{1}{2}\left[p_{0_{ii}} - (\boldsymbol{d}_i^\top\boldsymbol{P}_0\boldsymbol{y})^2/\theta_0\right] \end{aligned} \tag{9}$$

where $p_{0_{ii}}$ is the i^{th} diagonal element of $\boldsymbol{P}_0$ and the subscript of "0" indicates that all quantities are obtained from the null (no outlier) model.

In the ordinary linear mixed model setting it is common practice to assess every observation to determine whether it is an outlier. Thus, for example, Cook et al. (1982) and Thompson (1985) applied a separate AOM for each of the n observations. In our setting the analogous procedure is the application of a series of AOMM-R of the form in equation (7) and associated pointwise score tests for the null hypotheses $H_0 : \omega_i = 0, \;\; i = 1 \dots n$. The score test statistic for $H_0 : \omega_i = 0$ (evaluated at $\omega_i = 0$) has the form

$$W_0(\omega_i) = [U_0(\omega_i)]^2\, \mathcal{I}_0^{(\omega_i,\omega_i)} \tag{10}$$

where $\mathcal{I}_0^{(\omega_i,\omega_i)}$ is the element of the inverse of the expected information matrix for $(\omega_i, \theta, \boldsymbol{\kappa}^\top)^\top$ corresponding to ω_i and evaluated under the null. The complete expected information matrix can be written as

$$\left[\begin{array}{cc|c} \mathcal{I}_0(\omega_i,\omega_i) & \mathcal{I}_0(\omega_i,\theta) & \mathcal{I}_0(\omega_i,\boldsymbol{\kappa}^\top) \\ \mathcal{I}_0(\theta,\omega_i) & \mathcal{I}_0(\theta,\theta) & \mathcal{I}_0(\theta,\boldsymbol{\kappa}^\top) \\ \hline \mathcal{I}_0(\boldsymbol{\kappa},\omega_i) & \mathcal{I}_0(\boldsymbol{\kappa},\theta) & \mathcal{I}_0(\boldsymbol{\kappa},\boldsymbol{\kappa}^\top) \end{array}\right]$$

where the partitioning separates ω_i and θ from $\boldsymbol{\kappa}$. Expressions for individual elements of this matrix are given in Appendix A. We proceed as in Gumedze et al. (2010) and compute the test whilst holding the parameters in $\boldsymbol{\kappa}$ fixed so that $\mathcal{I}_0^{(\omega_i,\omega_i)}$ is obtained using only the partition corresponding to ω_i and θ. Using the results in Appendix A we then have

$$\begin{aligned} \mathcal{I}_0^{(\omega_i,\omega_i)} &= \left[\mathcal{I}_0(\omega_i,\omega_i) - \mathcal{I}_0(\omega_i,\theta)^2/\mathcal{I}_0(\theta,\theta)\right]^{-1} \\ &= \frac{2\nu}{(\nu-1)p_{0_{ii}}^2} \end{aligned}$$

so that the test statistic in equation (10) takes the form

$$\begin{aligned} W_0(\omega_i) &= \frac{\nu}{2(\nu-1)}\left(\frac{(\boldsymbol{d}_i^\top \boldsymbol{P}_0 \boldsymbol{y})^2}{\theta_0 p_{0_{ii}}} - 1\right)^2 && (11) \\ &= \frac{\nu}{2(\nu-1)}\left(t_{0_i}^2 - 1\right)^2 && (12) \end{aligned}$$

where t_{0_i} is the Studentised conditional residual for observation i evaluated under the null model. Equation (11) or equivalently (12) holds for $t_{0_i}^2 > 1$, otherwise $W_0(\omega_i) = 0$. Note that in practice, the score test statistic is calculated using the REML estimates of θ and $\boldsymbol{\kappa}$ from the fit of the null (no outlier) model.

In order to assess the significance of individual tests given the multiple testing scenario we need to evaluate the distribution of $\max_i W_0(\omega_i)$. This can be obtained using a resampling scheme. We propose a computationally efficient method that is similar to the one used in Zou et al. (2004). The steps in the procedure are as follows:

1. Fit the linear mixed model to obtain REML estimates of θ and $\boldsymbol{\kappa}$. Denote these by θ_0 and $\boldsymbol{\kappa}_0$ and thence form $\boldsymbol{V}_0$ and $\boldsymbol{P}_0$. Calculate the score test statistic $W_0(\omega_i), \quad i = 1 \ldots n$ using equation (11).

2. Given the estimates of the variance parameters from the model fit we propose to simulate score test statistics for each outlier variance parameter $\omega_i, \quad i = 1 \ldots n$. The number of simulations will be denoted N_s. Equation (11) shows that the simulations can be based on the distribution of $\boldsymbol{P}_0\boldsymbol{y}$ which is given by $\mathcal{N}(\boldsymbol{0}, \theta_0\boldsymbol{P}_0)$. Thus if $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$ then $\sqrt{\theta_0}\boldsymbol{P}_0\boldsymbol{Q}_0^\top\boldsymbol{z}$ has the same distribution as $\boldsymbol{P}_0\boldsymbol{y}$, where $\boldsymbol{Q}_0$ denotes the Choleski decomposition of $\boldsymbol{V}_0$ (that is, such that $\boldsymbol{Q}_0^\top\boldsymbol{Q}_0 = \boldsymbol{V}_0$). Thus in practice, for the j^{th} simulation ($j = 1 \ldots N_s$) we obtain $\boldsymbol{z}_j$ by sampling from a $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$ distribution and compute the simulated value of $\boldsymbol{P}_0\boldsymbol{y}$, which we will label as $(\boldsymbol{P}\boldsymbol{y})_j$, as

$$(\boldsymbol{P}\boldsymbol{y})_j = \sqrt{\theta_0}\boldsymbol{P}_0\boldsymbol{Q}_0^\top\boldsymbol{z}_j$$

We could then compute the score test statistic using equation (11) but with $\boldsymbol{P}_0\boldsymbol{y}$ replaced by $(\boldsymbol{P}\boldsymbol{y})_j$. However, substantial improvements in the Type I error rates can be achieved by noting that the simulation of $\boldsymbol{P}_0\boldsymbol{y}$ allows for an updated estimate of θ. Thus, using equation (5) and the result that $\boldsymbol{P}\boldsymbol{V}\boldsymbol{P} = \boldsymbol{P}$ we can compute

$$\theta_j = (\boldsymbol{P}\boldsymbol{y})_j^\top\boldsymbol{V}_0(\boldsymbol{P}\boldsymbol{y})_j/\nu$$

and thence the score test statistic as

$$W_j(\omega_i) = \begin{cases} \frac{\nu}{2(\nu-1)}\left(\frac{(\boldsymbol{d}_i^\top(\boldsymbol{P}\boldsymbol{y})_j)^2}{\theta_j p_{0_{ii}}} - 1\right)^2 & \frac{(\boldsymbol{d}_i^\top(\boldsymbol{P}\boldsymbol{y})_j)^2}{\theta_j p_{0_{ii}}} > 1 \\ 0 & \text{otherwise} \end{cases}$$

3. We repeat step 2 N_s times to obtain an $n \times N_s$ matrix of simulated score test statistics. We then compute $M_j = \max_i W_j(\omega_i), \quad j = 1 \ldots N_s$.
4. In order to obtain a global Type I error rate α for the n hypothesis tests we calculate the $100(1-\alpha)$ percentile of the N_s values M_j. If the observed value of the score test statistic $W_0(\omega_i)$ exceeds this threshold then the null hypothesis $H_0 : \omega_i = 0$ is rejected.

The calculations involved in the simulations are based on the matrix $\boldsymbol{P}_0$ from the original analysis. This is evaluated once and used repeatedly in steps 2 and 3. This is in contrast to the standard full parametric bootstrap approach that would require the re-fitting of the model in each simulation iteration. This is important for large data-sets and/or complex mixed models where model estimation may be very time consuming.

2.3.2. AOMM-G

We now consider an outlier model with respect to the random effects $\boldsymbol{u}$. Recall that $\boldsymbol{u}$ comprises q sub-vectors so here we partition as $(\boldsymbol{u_A}^\top, \boldsymbol{u_{\bar{A}}}^\top)^\top$ where $\boldsymbol{u_A}$ is the $b_A \times 1$ sub-vector of random effects we wish to investigate for outliers and $\boldsymbol{u_{\bar{A}}}$ comprises the remaining $q-1$ sub-vectors. We partition $\boldsymbol{Z}$ conformably as $\boldsymbol{Z} = [\boldsymbol{Z_A} \;\; \boldsymbol{Z_{\bar{A}}}]$ then write the linear mixed model in equation (1) as

$$\boldsymbol{y} = \boldsymbol{X\tau} + \boldsymbol{Z_A u_A} + \boldsymbol{Z_{\bar{A}} u_{\bar{A}}} + \boldsymbol{e}$$

and we let $\text{var}\,(\boldsymbol{u_A}) = \theta \boldsymbol{G_A}$ and $\text{var}\,(\boldsymbol{u_{\bar{A}}}) = \theta \boldsymbol{G_{\bar{A}}}$ so that $\boldsymbol{G}$ is a block diagonal matrix given by $\boldsymbol{G_A} \oplus \boldsymbol{G_{\bar{A}}}$.

An outlier is assumed to arise from a random effect with inflated variance. The associated linear mixed model for the case of the k^{th} random effect in $\boldsymbol{u_A}$ being an outlier $(k = 1 \ldots b_A)$ can be written as

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X\tau} + \boldsymbol{Z_A u_{A}}_k + \boldsymbol{Z_{\bar{A}} u_{\bar{A}}} + \boldsymbol{e} \\ \text{where} \quad \boldsymbol{u_A}_k &= \boldsymbol{u_A} + \boldsymbol{f}_k \eta_k \end{aligned}$$

and $\boldsymbol{f}_k$ is a $b_A \times 1$ vector with a value of one in position k and zeros elsewhere. The random effect η_k is assumed to have zero mean and variance $\theta\lambda_k$ where $\lambda_k > 0$. The model and associated scaled variance matrix for the data vector can then be written as

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X\tau} + \boldsymbol{Zu} + \boldsymbol{Z_A f}_k \eta_k + \boldsymbol{e} \\ \boldsymbol{V} &= \boldsymbol{ZGZ}^\top + \lambda_k (\boldsymbol{Z_A f}_k)(\boldsymbol{Z_A f}_k)^\top + \boldsymbol{R} \end{aligned} \tag{13}$$

Note that this model and variance matrix can also be written as

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X\tau} + \boldsymbol{Zu} + \dot{\boldsymbol{d}}_k \eta_k + \boldsymbol{e} \\ \boldsymbol{V} &= \boldsymbol{ZGZ}^\top + \lambda_k \dot{\boldsymbol{d}}_k \dot{\boldsymbol{d}}_k^\top + \boldsymbol{R} \end{aligned} \tag{14}$$

where $\dot{\boldsymbol{d}}_k = \boldsymbol{Z_A f}_k$. In the case where $\boldsymbol{Z_A}$ is a standard design matrix that corresponds to a factor with b_A levels, this means that $\dot{\boldsymbol{d}}_k$ is an $n \times 1$ vector with values of one in positions corresponding to the k^{th} level of this factor and zeros elsewhere. The form of the model in equation (14) appears similar to that in equation (8), suggesting that there may be a relationship between the AOMM-G for a random effect and an AOMM-R for a specific set of errors, namely for all observations corresponding to the particular random effect. A key difference, however, is that the variance matrix $\lambda_k \dot{\boldsymbol{d}}_k \dot{\boldsymbol{d}}_k^\top$ in the AOMM-G involves covariances between those observations.

We proceed as in section 2.3.1, that is, we conduct pointwise score tests for the null hypotheses $H_0 : \lambda_k = 0, \quad k = 1 \ldots b_A$. Using the form of $\boldsymbol{V}$ in equation (13) leads to the REML score for λ_k evaluated at the null value of $\lambda_k = 0$ being given by

$$U_0(\lambda_k) = -\tfrac{1}{2}\left[a_{0_{kk}} - (\boldsymbol{f}_k^\top \boldsymbol{Z_A}^\top \boldsymbol{P}_0 \boldsymbol{y})^2/\theta_0\right]$$

where $a_{0_{kk}}$ is the k^{th} diagonal element of $\boldsymbol{Z_A}^\top \boldsymbol{P}_0 \boldsymbol{Z_A}$. The score test has the same form as in equation (10), and as in section 2.3.1 the parameters in $\boldsymbol{\kappa}$ are held fixed so that the test statistic is given by

$$W_0(\lambda_k) = \frac{\nu}{2(\nu-1)}\left(\frac{(\boldsymbol{f}_k^\top \boldsymbol{Z_A}^\top \boldsymbol{P}_0 \boldsymbol{y})^2}{\theta_0 a_{0_{kk}}} - 1\right)^2 \qquad (15)$$

$$= \frac{\nu}{2(\nu-1)}\left(s_{0_k}^2 - 1\right)^2 \qquad (16)$$

where s_{0_k} is evaluated at the null value and is a function of the BLUP of $\boldsymbol{u_A}$. Equation (15) or equivalently (16) holds for $s_{0_k}^2 > 1$, otherwise $W_0(\lambda_k) = 0$. Note that in practice, the score test statistic is calculated using the REML estimates of θ and $\boldsymbol{\kappa}$ from the fit of the null (no outlier) model.

The resampling scheme for obtaining thresholds is analogous to that described in Section 2.3.1.

2.3.3. Graphical tools

A key part of model checking, in particular the examination of data for outliers is the use of graphical displays of residuals. The score test statistic in equation (12) shows that Studentised conditional residuals, t_i, are fundamentally important for outliers associated with the errors. The observation with the largest absolute Studentised conditional residual is the most likely candidate for rejection of the hypothesis of $H_0 : \omega_i = 0$. The scenario for random effect outliers is analogous, with equation (16) showing that the quantities s_k are the key. Clearly, visual displays of t_i and s_k will be illuminating in terms of identifying potential outliers, at the error and random effect levels, respectively. We find normal probability plots particularly useful for this purpose.

We note that the quantities t_i and s_k are easily obtained in the statistical software package ASReml-R (Butler et al., 2017), using the "aom" option in the function call. This returns two-column matrices with information corresponding to AOMM-R and AOMM-G, respectively. Using the notation in this paper, individual elements in the second column of the matrix for AOMM-R are defined in Butler et al. (2017) to be

$$\boldsymbol{d}_i^\top \left(\theta \boldsymbol{R}\right)^{-1} \tilde{\boldsymbol{e}} / \sqrt{p_{ii}/\theta} = \boldsymbol{d}_i^\top \boldsymbol{P}\boldsymbol{y} / \sqrt{\theta p_{ii}} = t_i$$

Similarly, individual elements in the second column of the matrix for AOMM-G are the quantities s_k. It is therefore straightforward to construct diagnostic plots based on these quantities, irrespective of whether formal tests are undertaken. In fact, we recommend use of normal probability plots of Studentised conditional residuals as a first course of action in order to identify obviously erroneous data, for example when decimal points have been misplaced. These should be dealt with prior to undertaking formal tests otherwise more subtle outliers may be missed.

3. Examples

3.1. Nicotine data

We consider the study first presented by Wagner and Thaggard (1979) in which 10 samples were analysed for nicotine content (using gas-liquid chromatography) by each of 14 laboratories. There was a total of 138 observations since 2 were missing. This study has been analysed by several authors, including Christensen et al. (1992b) and Rocke (1983), the latter of whom presented the full data-set. The data have been analysed using a mixed model with fixed effects for samples and random effects for laboratories. In terms of equation (1) we have $n = 138$ observations, $p = 10$ fixed effects and $b = 14$ random effects (with $q = 1$). It is assumed that $\text{var}\left(\boldsymbol{u}\right) = \theta\gamma \boldsymbol{I}_{14}$ and $\text{var}\left(\boldsymbol{e}\right) = \theta \boldsymbol{I}_{138}$, where θ is the error variance and γ is the laboratory variance component ratio. Fitting this model to the data yielded the REML estimates $\hat{\theta} = 7.70 \times 10^{-4}$ and $\hat{\gamma} = 2.19$. A normal probability plot of the Studentised conditional residuals from this model is shown in Figure 1(a). Observations with "large" absolute Studentised conditional residuals have been identified on this graph. These data points were also identified by Christensen et al. (1992b) as having the highest influence on the estimation of fixed effects and variance components.

We then applied the AOMM-R for the test of $H_0 : \omega_i = 0,\ i = 1 \dots 138$ using the approach described in Section 2.3.1 and with $N_s = 50,000$ simulations. The observed score test statistics $W_0(\omega_i)$ are shown in Figure 1(b), together with the 95% threshold value which was 63.4. Figure 1(b) shows

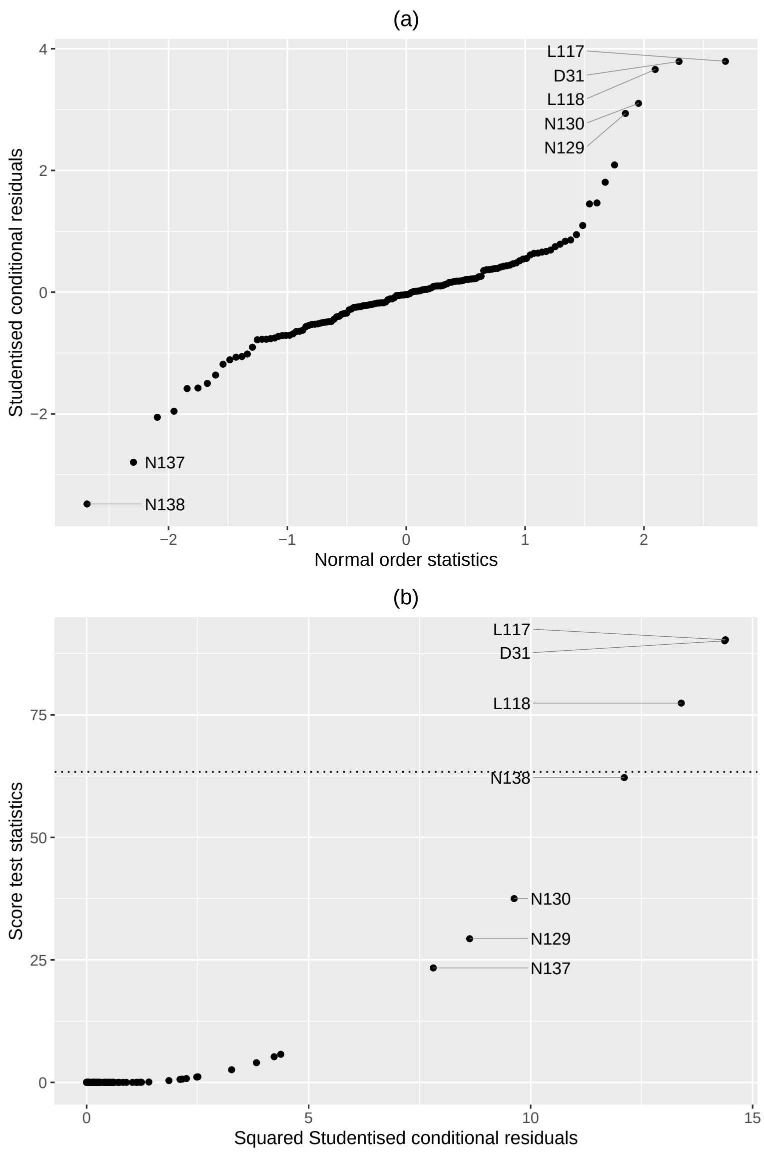

Figure 1: Nicotine data (a) normal probability plot for Studentised conditional residuals and (b) score test statistics for AOMM-R plotted against squared Studentised conditional residuals. The 7 cases with the largest (absolute) Studentised conditional residuals have been labelled with their laboratory and case number. The dashed line in (b) indicates the 95% threshold value for the test statistics.

that the null hypothesis of no additional error variance was rejected in 3 instances, corresponding to cases 117, 31 and 118. The test statistics for cases 129, 130, 137 and 138 did not exceed the threshold (although case 138 was close with a test statistic of 62.2). Note that Christensen et al. (1992b) included these 4 cases when they referred to a total of 7 cases "... that appear to have substantial influence ...". Figure 1(b) also shows the relationship between the score test statistics and the squared Studentised conditional residuals as established in equation (12).

In their (original) analysis of these data, Wagner and Thaggard (1979) rejected laboratory N as being inconsistent with the rest. We have formally assessed the outlier status of individual laboratories using the AOMM-G. First note that the value of s_k for laboratory N (-3.29) was substantially lower than for the other laboratories (values ranged from -0.58 to 0.92). In terms of the hypothesis tests we used the resampling approach of Section 2.3.2 and with $N_s = 50,000$ simulations. The observed score test statistic for laboratory N was 48.46 which exceeded the 95% threshold value of 31.65. Laboratory N was therefore identified as being an outlier in the sense that the associated effect had an inflated variance compared with the effects for other laboratories.

3.2. Variety trial data

We consider data on grain yield from a variety trial in which 90 durum wheat varieties were grown. There were 2 replicate plots of each variety, making a total of 180 plots arranged in a rectangular array of 6 columns by 30 rows. The design was resolvable with replicate blocks being aligned with columns (block 1 = columns 1-3; block 2 = columns 4-6). The data were analysed using a linear mixed model with a fixed effect for the overall mean, random effects for the replicate blocks, random effects for varieties and a spatial model for the errors of the form advocted by Gilmour et al. (1997). In terms of equation (1) we have $n = 180$ observations, $p = 1$ fixed effect and $q = 2$ sets of random effects, denoted $\boldsymbol{u}_1$ for the replicate block effects and $\boldsymbol{u}_2$ for the variety effects. It was assumed that $\text{var}(\boldsymbol{u}_1) = \theta\gamma_1\boldsymbol{I}_2$, $\text{var}(\boldsymbol{u}_2) = \theta\gamma_2\boldsymbol{I}_{90}$ and $\text{var}(\boldsymbol{e}) = \theta\boldsymbol{\Sigma}_{\boldsymbol{c}}(\phi_1) \otimes \boldsymbol{\Sigma}_{\boldsymbol{r}}(\phi_2)$, where γ_1 and γ_2 are the replicate block and variety variance component ratios and $\boldsymbol{\Sigma}_{\boldsymbol{c}}(\phi_1)$ and $\boldsymbol{\Sigma}_{\boldsymbol{r}}(\phi_2)$ are correlation matrices corresponding to autoregressive processes of order 1 for the column and row dimensions respectively. The parameters ϕ_1 and ϕ_2 are the autocorrelation parameters and θ is the variance of the spatial process. Fitting this model to the data yielded the REML estimates

$\hat{\theta} = 0.1285$; $\hat{\gamma}_2 = 0.18$; $\hat{\phi}_1 = 0.49$ and $\hat{\phi}_2 = 0.75$. The variance parameter γ_1 for replicate blocks was estimated at the boundary value of 0.

Figure 2(a) shows the residual diagnostic plot used by various authors after the conduct of a spatial analysis (see Gilmour et al., 1997, for example) in which the BLUPs of the errors, that is, $\tilde{\boldsymbol{e}}$, are plotted against row number for each column. The error BLUPs are used as residuals in this type of plot as they reveal the spatial patterns which are strong in these data. It is possible to identify potential outliers on these plots as points that depart from the neighbouring spatial pattern. Thus the observation in Column 3, Row 3 is a potential candidate. However, outliers are more clearly seen by examining a normal probability plot of the Studentised conditional residuals as given in Figure 2(b). We also applied the AOMM-R test using the simulation approach described in section 2.3.1 using $N_s = 50,000$ simulations. The null hypothesis of no outlier was rejected for a single observation, namely that associated with the field plot in Column 3, Row 3.

4. Simulation study

A simulation study was undertaken in order to assess the performance of the score test approach in terms of Type I error rates. The model considered was a one-way classification with b groups and r replicates per group, thence $n = rb$ observations. In terms of equation (1) we had $\boldsymbol{\tau}$ comprising a single effect (the overall mean, μ) and $\boldsymbol{u}$ comprising a single set of random effects associated with the groups. The variance models were $\text{var}(\boldsymbol{u}) = \theta\gamma\boldsymbol{I}_b$ and $\text{var}(\boldsymbol{e}) = \theta\boldsymbol{I}_n$. Without loss of generality θ was set to 1 in all cases. The AOMM-R score tests were conducted for the factorial combinations of 3 values of b (10, 25 and 50) by 3 values of γ (0.1, 1.0 and 2.0) by 2 values of r (2 and 4). Additionally a scenario similar to the variety trial example was considered with $b = 90$ and $r = 2$ and using 3 values of γ (0.1, 1.0 and 2.0). We also considered spatial models for the errors, assuming 6 columns and 30 rows and with $\gamma = 0.1$ and 3 pairs of spatial correlations (ϕ_1 for columns and ϕ_2 for rows), namely (0.4, 0.7), (0.3, 0.6) and (0.2, 0.5). Thus there was a total of 24 scenario. For each of these we performed 2000 simulations and conducted the AOMM-R score tests for outliers as described in section 2.3.1 (using $N_s = 50,000$ "inner" simulations to determine 95% thresholds). The empirical Type I error rate was calculated as the proportion of the 2000 simulations in which one or more score test statistics exceeded the threshold. The results presented in Table 1 clearly show the accuracy of the Type I error

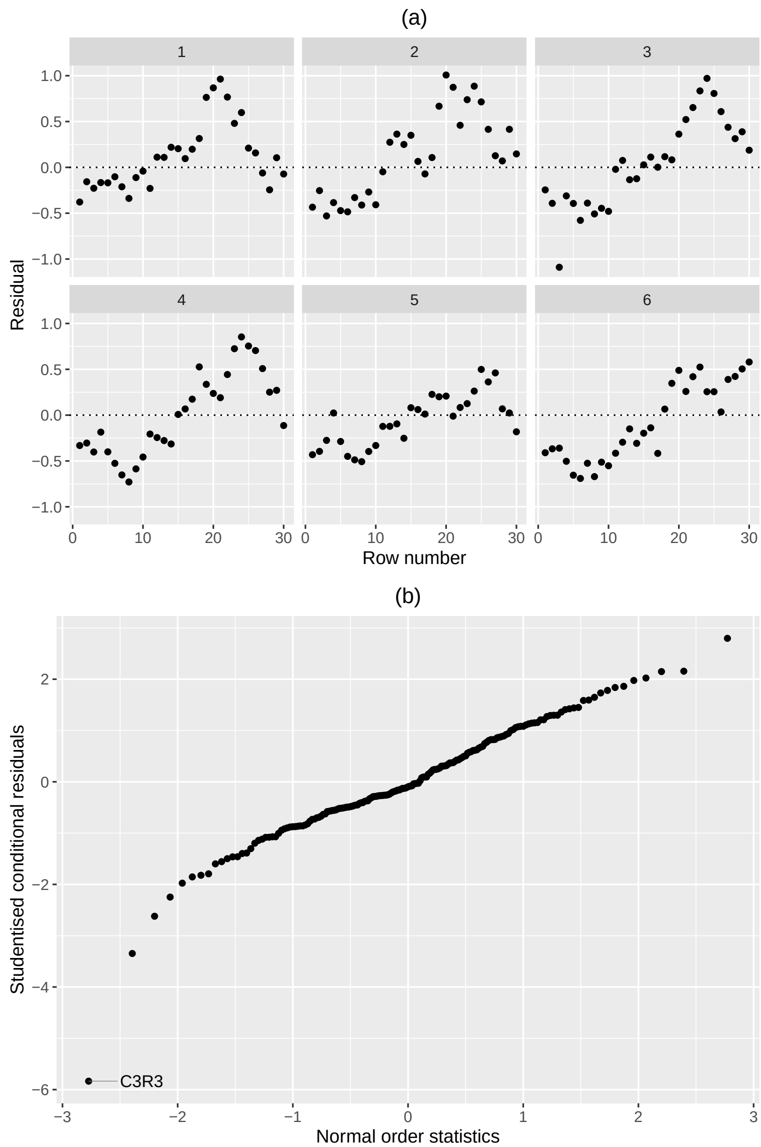

Figure 2: Variety trial data (a) residuals ($\tilde{\boldsymbol{e}}$) plotted against Row number for each Column and (b) normal probability plot of Studentised conditional residuals (labelled point corresponds to Column 3, Row 3).

Table 1: Simulation study: empirical Type I error rates for AOMM-R score tests in a one-way random effects model with b groups and r replicates per group. Overall scale parameter $\theta = 1$ and groups variance is $\theta\gamma$. For the case of $b = 90$, $r = 2$ and $\gamma = 0.1$ a two-dimensional (row by column) separable autoregressive spatial model was fitted for 3 different combinations of autocorrelation parameters (ϕ_1 and ϕ_2). The nominal Type I rate is 0.05.

		Replicates per group		Spatial correlations ϕ_1, ϕ_2		
Groups	γ	2	4	0.4, 0.7	0.3, 0.6	0.2, 0.5
10	0.1	0.0495	0.0545			
10	1.0	0.0450	0.0425			
10	2.0	0.0430	0.0505			
25	0.1	0.0500	0.0520			
25	1.0	0.0440	0.0415			
25	2.0	0.0470	0.0430			
50	0.1	0.0510	0.0515			
50	1.0	0.0465	0.0490			
50	2.0	0.0445	0.0515			
90	0.1	0.0525	–	0.0475	0.0485	0.0500
90	1.0	0.0490	–			
90	2.0	0.0505	–			

rates for all scenario, including those that involved correlated error (spatial) models.

We also conducted AOMM-G score tests for the factorial combinations of 3 values of b (10, 25 and 50) by 2 values of γ (1.0 and 2.0) by 2 values of r (2 and 4). Additionally the $b = 90$ and $r = 2$ scenario was considered. Thus there was a total of 14 scenario. As with the AOMM-R tests, the empirical Type I error rate was calculated as the proportion of 2000 simulations in which one or more score test statistics exceeded the threshold. The results presented in Table 2 reveal the AOMM-G score tests to be conservative, particularly for small numbers of groups.

5. Concluding remarks

In this paper we have addressed the issue of outlier identification in linear mixed models using an extension of the alternative outlier model proposed for ordinary linear models. The distinction has been made between outlier models associated with the errors and with the random effects, and these have

Table 2: Simulation study: empirical Type I error rates for AOMM-G score tests in a one-way random effects model with b groups and r replicates per group. Overall scale parameter $\theta = 1$ and groups variance is $\theta\gamma$. The nominal Type I rate is 0.05.

		Replicates per group	
Groups	γ	2	4
10	1.0	0.0100	0.0035
10	2.0	0.0120	0.0030
25	1.0	0.0235	0.0250
25	2.0	0.0360	0.0195
50	1.0	0.0315	0.0390
50	2.0	0.0395	0.0320
90	1.0	0.0400	–
90	2.0	0.0380	–

been denoted AOMM-R and AOMM-G, respectively. In order to diagnose an inflated variance for an effect we have developed score tests based on the fit of the null (no outlier) model. Importantly, the AOMM-R score test statistics have a strong link with Studentised conditional residuals, so we would recommend use of the latter in informal diagnostic plots as part of the routine examination of model adequacy (see Lee et al., 2006, for example). Note that we are referring here to the conditional residuals of Haslett and Hayes (1998) and Haslett and Haslett (2007) which are provided in ASReml-R (Butler et al., 2017). These residuals are not to be confused with those of the same name in, for example, Nobre and Singer (2007), and as provided in other linear mixed models software packages including SAS (SAS Institute, Inc, 2013).

Formal score tests were developed for both the AOMM-R and AOMM-G, with thresholds determined using an efficient resampling scheme. This offers a pragmatic solution as a full parametric bootstrap method is prohibitively time-consuming for larger data-sets (also see Gumedze et al., 2010). Application of the method for AOMM-R score tests in the nicotine and variety trial examples required only 1.95 and 2.72 seconds respectively for 50,000 simulations. As a comparison, the model fitting took 0.08 and 0.10 seconds so that 50,000 repeats of this process would require over an hour for each of the examples. Importantly the scheme allowed updating of the overall scale parameter, thereby introducing some level of uncertainty associated with variance parameter estimation. This appears to be adequate for AOMM-R scores tests as supported by the results of the simulation study. The empirical Type I error rates were extremely close to the nominal rate of 0.05 for all scenario considered, including correlated error models. This is a key result since the identification of "data" outliers, that is, associated with the errors, is a recommended step in all linear mixed model analyses. The simulation study for AOMM-G score tests showed that the resampling scheme resulted in conservative tests, particularly for small numbers of random effects. In contrast to the identification of data outliers, the investigation of random effect outliers occurs much less frequently and could therefore be deemed less critical. Never-the-less the resampling scheme presented here may still be of use. Modifications aimed at improving the Type I error rates for AOMM-G score tests are the subject of further research.

Having identified outliers, the issue remains as to a course of action. One option is to re-fit the linear mixed model with a separate (extra) variance component for each outlier so they are down-weighted in the analysis (Gumedze et al., 2010). Another is to delete the observations altogether. Either way, we believe it is important to confer with the scientist and to err on the side of parsimony if adding extra variance parameters, or caution if deleting data.

Finally, as demonstrated in the variety trial example, the method is applicable in the case of correlated effects so is not restricted to variance component models. We note however, that further extensions are required for multivariate linear mixed models since outliers may occur in the full multivariate sense or in a subspace.

Funding

AS was supported by the Grains Research and Development Corporation (GRDC) through the EssCargoT project (UW00010).

Appendix A. Expected Information matrix

Elements of the expected information matrix for θ and elements of $\boldsymbol{\kappa}$ are given by

$$\begin{aligned}
\mathcal{I}(\theta,\theta) &= \tfrac{1}{2}\nu/\theta^2 \\
\mathcal{I}(\theta,\kappa_r) &= \tfrac{1}{2}\mathrm{tr}\left(\boldsymbol{P}\dot{\boldsymbol{V}}_r\right)/\theta \\
\mathcal{I}(\kappa_r,\kappa_s) &= \tfrac{1}{2}\mathrm{tr}\left(\boldsymbol{P}\dot{\boldsymbol{V}}_r\boldsymbol{P}\dot{\boldsymbol{V}}_s\right)
\end{aligned}$$

Elements corresponding to the AOMM-R outlier variance parameter ω_i are

$$\begin{aligned} \mathcal{I}(\omega_i, \omega_i) &= \tfrac{1}{2} p_{ii}^2 \\ \mathcal{I}(\omega_i, \theta) &= \tfrac{1}{2} p_{ii} / \theta \\ \mathcal{I}(\omega_i, \kappa_r) &= \tfrac{1}{2} \mathrm{tr} \left(\boldsymbol{d}_i^\top \boldsymbol{P} \dot{\boldsymbol{V}}_r \boldsymbol{P} \boldsymbol{d}_i \right) \end{aligned}$$

Elements corresponding to the AOMM-G outlier variance parameter λ_k are

$$\begin{aligned} \mathcal{I}(\lambda_k, \lambda_k) &= \tfrac{1}{2} a_{kk}^2 \\ \mathcal{I}(\lambda_k, \theta) &= \tfrac{1}{2} a_{kk} / \theta \\ \mathcal{I}(\lambda_k, \kappa_r) &= \tfrac{1}{2} \mathrm{tr} \left(\boldsymbol{f}_k^\top \boldsymbol{Z_A}^\top \boldsymbol{P} \dot{\boldsymbol{V}}_r \boldsymbol{P} \boldsymbol{f}_k \boldsymbol{Z_A} \right) \end{aligned}$$

References

Butler, D.G., Cullis, B.R., Gilmour, A.R., Gogel, B.J., Thompson, R., 2017. ASReml-R Reference Manual Version 4. Technical Report. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Christensen, R., Johnson, W., Pearson, L.M., 1992a. Prediction diagnostics for spatial linear models. Biometrika 79, 583–591.

Christensen, R., Pearson, L.M., Johnson, W., 1992b. Case-deletion diagnostics for mixed models. Technometrics 34, 38–45.

Cook, R., Holschuh, N., Weisberg, S., 1982. A note on an alternative outlier model. Journal of the Royal Statistical Society, Series B 44, 370–376.

Gilmour, A.R., Cullis, B.R., Verbyla, A.P., 1997. Accounting for natural and extraneous variation in the analysis of field experiments. Journal of Agricultural, Biological, and Environmental Statistics 2, 269–273.

Gumedze, F.N., Welham, S.J., Gogel, B.J., Thompson, R., 2010. A variance shift model for detection of outliers in the linear mixed model. Computational Statistics and Data Analysis 54, 2128–2144. URL: `http://dx.doi.org/10.1016/j.csda.2010.03.019`, doi:10.1016/j.csda.2010.03.019.

Haslett, J., Haslett, S., 2007. The three basic types of residuals for a linear model. International Statistical Review 75, 1–24.

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

Langford, I., Lewis, T., 1998. Outliers in multilevel data. Journal of the Royal Statistical Society, Series A 161, 121–160.

Lee, Y., Nelder, J., Pawitan, Y., 2006. Generalized Linear Models with Random Effects: Unified Analysis via H-likelihood. Boca Raton: Chapman and Hall/CRC.

Nobre, J., Singer, J., 2007. Residual analysis for linear mixed models. Biometrical Journal 49, 863–875.

Patterson, H.D., Thompson, R., 1971. Recovery of interblock information when block sizes are unequal. Biometrika 31, 100–109.

Rocke, D.M., 1983. Robust statistical analysis of interlaboratory studies. Biometrika 70, 421–431.

SAS Institute, Inc, 2013. SAS/STAT 13.1 Users Guide. SAS Institute. Cary, NC.

Thompson, R., 1985. A note on restricted maximum likelihood estimation with an alternative outlier model. Journal of the Royal Statistical Society Series B 47, 53–55.

Verbyla, A.P., 1993. Modelling variance heterogeneity: residual maximum likelihood and diagnostics. Journal of the Royal Statistical Society, Series B 55, 493–508.

Wagner, J.R., Thaggard, N.A., 1979. Gas-liquid chromatographic determination of nicotine contained on cambridge filter pads. Journal of the Association of Official Analytical Chemists 62, 229–236.

Zou, F., Fine, J., Hu, J., Lin, D., 2004. An efficient resampling method for assessing genome-wide statistical significance in mapping quantitative trait loci. Genetics 168, 2307–2316.