Chilean Journal of Statistics
Vol. 3, No. 1, April 2012, 57–73

Statistical Modeling
Research Paper

On linear mixed models and their influence diagnostics applied to an actuarial problem

Luis Gustavo Bastos Pinho, Juvêncio Santos Nobre* and Sílvia Maria de Freitas

Department of Applied Mathematics and Statistics, Federal University of Ceará, Fortaleza, Brazil

(Received: 10 June 2011 · Accepted in final form: 23 September 2011)

Abstract

In this paper, we motivate the use of linear mixed models and diagnostic analysis in practical actuarial problems. Linear mixed models are an alternative to traditional credibility models. Frees et al. (1999) showed that some mixed models are equivalent to some widely used credibility models. The main advantage of linear mixed models is the use of diagnostic methods. These methods may help to improve the model choice and to identify outliers or influential subjects which deserve better attention by the insurer. As an application example, the data set in Hachemeister (1975) is modeled by a linear mixed model. We can conclude that this approach is superior to the traditional credibility one since the former is more flexible and allows the use of diagnostic methods.

Keywords: Credibility models · Hachemeister model · Linear mixed models · Diagnostics · Local influence · Residual analysis.

Mathematics Subject Classification: Primary 62J05 · Secondary 62J20.

1. Introduction

One of the main concerns in actuarial science is to predict the future behavior for the aggregate amount of claims of a certain contract based on its past experience. By accurately predicting the severity of the claims, the insurer is able to provide a fairer and thus more competitive premium.

Statistical analysis in actuarial science generally belongs to the class of repeated measures studies, where each subject may be observed more than once. By subject we mean each element of the observed set which we want to investigate. Workers of a company, class of employees, and different states are possible examples of subjects in actuarial science. To model actuarial data, a large variety of statistical models can be used, but it is usually difficult to choose a model due to the data structure, in which within-subject correlation is often seen. Correlation misspecification may lead to erroneous analysis. In some cases this error is very severe. A clear example may be seen in Demidenko (2004, pp. 2-3) and a similar artificial situation is reproduced in Figure 1, that shows the relation between the number of claims and the number of policy holders of an insurer for nine different regions within a country. In each region the two variables are measured once a year on the

*Corresponding author. Juvêncio Santos Nobre. Departamento de Estatística e Matemática Aplicada. Universidade Federal do Ceará. Fortaleza/CE-Brazil. CEP:60.440-900. Email: juvencio@ufc.br

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)

http://www.soche.cl/chjs

same day for three consecutive years. In Figure 1(a) we do not consider the within-region (within-subject) correlation. The dashed line is a simple linear regression and suggests that the more the policy holders, the less claims occur. In Figure 1(b) we joined the observations for each region by a solid line. It is clear now that the number of claims increases with the number of policy holders.

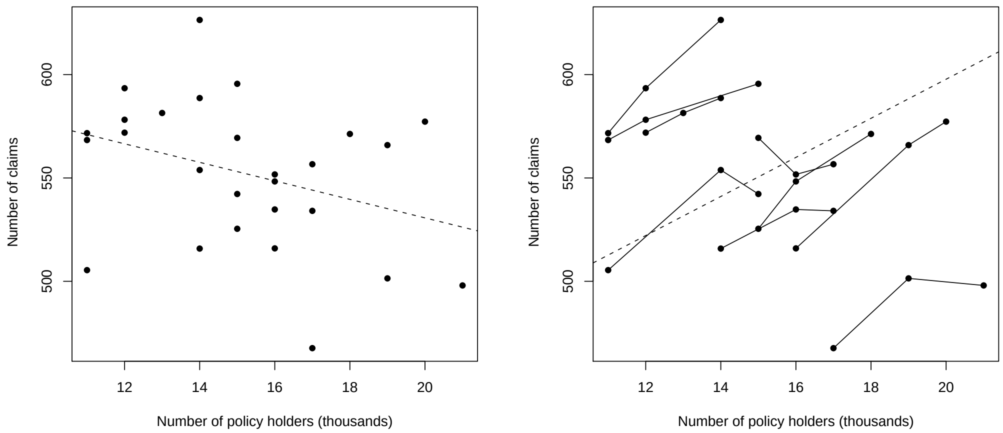

Figure 1. (a) Not considering the within-subject correlation, (b) considering the within-subject correlation.

It is necessary to take into consideration that each region may have a particular behavior which should be modeled, but only this is usually not enough. Techniques summarized under the name of diagnostic procedures may help to identify issues of concern, such as high influential observations, which may distort the analysis. For linear homoskedastic models, a well known diagnostic procedure is the residual plot. For linear mixed models better types of residuals are defined. Besides residual techniques, which are useful, there is a less used class of diagnostic procedures, which includes case deletion and measuring changes in the likelihood of the adjusted model under minor perturbations. Several important issues may not be noticed without the aid of these last diagnostics methods.

For introductory information regarding regression models and respective diagnostic analysis; see Cook and Weisberg (1982) or Drapper and Smith (1998). For a comprehensive introduction to linear mixed models, see Verbeke and Molenberghs (2000), McCulloch and Searle (2001) and Demidenko (2004). Diagnostic analysis of linear mixed models were presented and discussed in Beckman et al. (1987), Christensen and Pearson (1992), Hilden-Minton (1995), Lesaffre and Verbeke (1998), Banerjee and Frees (1997), Tan et al. (2001), Fung et al. (2002), Demidenko (2004), Demidenko and Stukel (2005), Zewotir and Galpin (2005), Gumedze et al. (2010) and Nobre and Singer (2007, 2011).

The seminal work of Frees et al. (1999) showed some similarities and equivalences between mixed models and some well known credibility models. Applications to data sets in actuarial context may be seen in Antonio and Beirlant (2006). Our contribution is to show how to use diagnostic methods for linear mixed models applied to actuarial science. We illustrate how to identify outliers and influential observations and subjects. We also show how to use diagnostics as a tool for model selection. These methods are very important and usually overlooked by most of the actuaries.

This paper is divided as follows. In Section 2 we present a motivational example using a well known data set. In Section 3 we briefly present the linear mixed models. Section 4 contains a short introduction to the diagnostic methods used in the example. In Section 5 we present an application based on the motivational example. Section 6 shows some conclusions. Finally, in an Appendix, we present mathematical details of some formulas and expressions used in the text.

2. Motivational Example

For a practical example, consider the Hachemeister (1975) data on private passenger bodily injury insurance. The data were collected from five states (subjects) in the US, through twelve trimesters between July 1970 and June 1973, and show the mean claim amount and the total number of claims in each trimester. The data may be found in the `actuar` package (see Dutang et al., 2008) from `R` (R Development Core Team, 2009) and are partially shown in Table 1.

Table 1. Hachemeister's data.

Trimester	State	Mean claim amount	Number of claims
1	1	1738	7861
1	2	1364	1622
1	3	1759	1147
1	4	1223	407
1	5	1456	2902
2	1	1642	9251
⋮	⋮	⋮	⋮
12	1	2517	9077
12	2	1471	1861
12	3	2059	1121
12	4	1306	342
12	5	1690	3425

In Figure 2 we plot the individual profiles for each state and the mean profile. It suggests that the claims have a different behavior along the trimesters for each state. One may notice that the claims from state 1 are greater than those from other states for almost every observation, and the claims from states 2 and 3 seem to grow more slowly than those from state 1. If the insurer wants to accurately predict the severity, the subjects' individual behavior must also be modeled. Traditionally this is possible with the aid of credibility models; see, e.g., Bühlmann (1967), Hachemeister (1975) and Dannenburg et al. (1996). These models assign weights, known as credibility factors, to a pair of different estimates of severity.

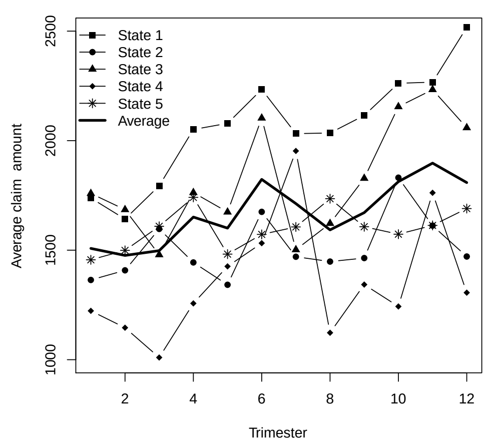

Figure 2. Individual profiles and mean profile for Hachemeister (1975) data.

Credibility models may be functionally defined as

$$ZB + (1 - Z)C,$$

where A represents the severity in a given state, Z is a credibility factor restricted to $[0, 1]$, B is a priori estimate of the expected severity for the same estate and C is a posteriori estimate also of the expected severity. Considering a particular state, B may be equal to the sample mean of the severity of its observations and C equal to the overall sample mean of the data in the same period.

Frees et al. (1999) showed that it is possible to find linear mixed models equivalent to some known credibility models, such as Bühlmann (1967) and Hachemeister (1975) models. Information about linear mixed models is provided in the next section.

3. Linear Mixed Models

Linear mixed models are a popular alternative to analyze repeated measures. Such models may be functionally expressed as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{e}_i, \quad i = 1, \ldots, k, \tag{1}$$

where $\mathbf{y}_i = (y_1, y_2, \ldots, y_{n_i})^\top$ is a $n_i \times 1$ vector of the observed values of the response variable for the ith subject, $\mathbf{X}_i$ is a $n_i \times p$ known full rank matrix, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, also known as fixed effects, which are used to model $\mathrm{E}[\mathbf{y}_i]$, $\mathbf{Z}_i$ is a $n_i \times q$ known full rank matrix, $\mathbf{b}_i$ is a $q \times 1$ vector of latent variables, also known as random effects, used to model the within-subject correlation structure, and $\boldsymbol{e}_i = (e_{i1}, e_{i2}, \ldots, e_{in_i})^\top$ is the $n_i \times 1$ random vector of (within-subject) measurement errors. It is usually also assumed that $\boldsymbol{e}_i \overset{ind}{\sim} \mathcal{N}_{n_i}(\mathbf{0}, \sigma^2\mathbf{I}_{n_i})$, where $\mathbf{I}_{n_i}$ denotes the identity matrix of order n_i for $i = 1, \ldots, k$, $\mathbf{b}_i \overset{iid}{\sim} \mathcal{N}_q(\mathbf{0}, \sigma^2\mathbf{G})$ for $i = 1, \ldots, k$ in which $\mathbf{G}$ is a $q \times q$ positive definite matrix, and $\boldsymbol{e}_i$ and $\mathbf{b}_j$ are independent $\forall\, i, j$. Under these assumption, this is called a homoskedastic conditional independence model. It is possible to rewrite model given in Equation (1) in a more concise way as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{e}, \tag{2}$$

where $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_k^\top)^\top$, $\mathbf{X} = (\mathbf{X}_1^\top, \ldots, \mathbf{X}_k^\top)^\top$, $\mathbf{Z} = \bigoplus_{i=1}^k \mathbf{Z}_i$, $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_k^\top)^\top$ and $\boldsymbol{e} = (\boldsymbol{e}_1^\top, \ldots, \boldsymbol{e}_k^\top)^\top$, with $\bigoplus$ representing the direct sum.

It can be shown that, conditional on known covariance parameters of the model, that is conditional to the elements of $\mathbf{G}$ and σ^2, the best linear unbiased estimator (BLUE) for $\boldsymbol{\beta}$ and the best linear unbiased predictor (BLUP) for $\mathbf{b}$ are given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{y}, \tag{3}$$

and

$$\hat{\mathbf{b}} = \mathbf{D}\mathbf{Z}^\top\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}),$$

respectively, where $\mathbf{D} = \sigma^2\mathbf{G}$, $\mathbf{V} = \sigma^2(\mathbf{I}_n + \mathbf{Z}\mathbf{G}\mathbf{Z}^\top)$, with $n = \sum_{i=i}^k n_i$; see Hachemeister (1975).

Maximum likelihood (ML) and restricted maximum likelihood (RML) methods can be used to estimate the variance components of the model. The latter, proposed in Patterson and Thompson (1971), is usually chosen since it often generates less biased estimators related to the variance structure. When estimates for $\mathbf{V}$ are used in Equation (3) to obtain $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{b}}$, they are called empirical BLUE (EBLUE) and empirical BLUP (EBLUP), respectively. Usually the estimation of the parameters involves the use of iterative methods for maximizing the likelihood function.

Linear mixed models are not the only way to deal with repeated measures studies. Other popular alternatives are the generalized estimation equations (see Liang and Zeger, 1986; Diggle et al., 2002) and multivariate models as seen in Johnson and Whichern (1982) and Vonesh and Chinchilli (1997). But usually these alternatives are more restrictive than linear mixed models, and they only model the marginal expected value of the response variable.

4. Diagnostic Methods

Diagnostic methods comprehend techniques whose purpose is to investigate the plausibility and robustness of the assumptions made when choosing a model. It is possible to divide the techniques shown here in two classes: residual analysis, which investigates the assumptions on the distribution of errors and presence of outliers; and sensitivity analysis, which analyzes the sensitivity of a statistical model when subject to minor perturbations. Usually, it would be far more difficult, or even impossible, to observe these aspects in a traditional credibility model.

In the context of traditional linear models (homoskedastic and independent), examples of diagnostic methods may be seen in Hoaglin and Welsch (1978), Belsley et al. (1980) andCook and Weisberg (1982). Linear mixed models, extensions and generalizations are briefly discussed here and may be seen in Beckman et al. (1987), Christensen and Pearson (1992), Hilden-Minton (1995), Lesaffre and Verbeke (1998), Banerjee and Frees (1997), Tan et al. (2001), Fung et al. (2002), Demidenko (2004), Demidenko and Stukel (2005), Zewotir and Galpin (2005), Nobre and Singer (2007, 2011) and Gumedze et al. (2010).

4.1 Residual analysis

In the linear mixed models class three different kinds of residuals may be considered. The conditional residuals: $\hat{\boldsymbol{e}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\mathbf{b}}$, the EBLUP: $\mathbf{Z}\hat{\mathbf{b}}$, and the marginal residuals: $\hat{\boldsymbol{\xi}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$. These predict respectively conditional error $\boldsymbol{e} = \mathbf{y} - \mathrm{E}[\mathbf{y}|\mathbf{b}] = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}$, random effects $\mathbf{Z}\mathbf{b} = \mathrm{E}[\mathbf{y}|\mathbf{b}] - \mathrm{E}[\mathbf{y}]$ and the marginal error $\boldsymbol{\xi} = \mathbf{y} - \mathrm{E}[\mathbf{y}] = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$. Each of the mentioned residuals is useful to verify some assumption of the model, as seen in Nobre and Singer (2007) and briefly presented next.

4.1.1 Conditional residuals

To identify cases with a possible high influence on $\hat{\sigma}^2$ in linear mixed models, Nobre and Singer (2007) suggested the standardization for the conditional residual given by

$$\hat{e}_i^* = \frac{\hat{e}_i}{\sigma\sqrt{q_{ii}}},$$

where q_{ii} represents the ith element in the main diagonal of $\mathbf{Q}$ defined as

$$\mathbf{Q} = \sigma^2(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{V}^{-1}).$$

Under normality assumptions on $\boldsymbol{e}$, this standardization identifies outlier observations and subjects; see Nobre and Singer (2007). To do so, the same authors consider the quadratic form $M_I = \mathbf{y}^\top\mathbf{Q}\mathbf{U}_I(\mathbf{U}_I^\top\mathbf{Q}\mathbf{U}_I)^{-1}\mathbf{U}_I^\top\mathbf{Q}\mathbf{y}$, where $\mathbf{U}_I = (u_{ij})_{(n\times k)} = (\mathbf{U}_{i_1}, \ldots, \mathbf{U}_{i_k})$, with $\mathbf{U}_i$ representing the ith column of the identity matrix of order n. To identify an outlier subject let I be the index set of the subject observations and evaluate M_I for this subset.

Table 2. Diagnostic techniques involving residuals.

Diagnostic	Graph
Linearity of fixed effects	$\hat{\boldsymbol{\xi}}$ vs. explanatory variables (fitted values)
Presence of outliers	$\hat{\boldsymbol{e}}$ vs. observation index
Homoskedasticity of the conditional errors	$\hat{\boldsymbol{e}}$ vs. fitted values
Normality of the conditional errors	QQ plot for the least confounded residuals
Presence of outlier subjects	Mahalanobis distance vs. observation index
Normality of the fixed effects	weighted QQ plot for $\hat{\mathbf{b}}_i$

4.1.2 Confounded residuals

It can be shown that, under the assumptions made by model given in Equation (1), we have

$$\hat{\boldsymbol{e}} = \mathbf{RQ}\boldsymbol{e} + \mathbf{RQZb} \quad \text{and} \quad \mathbf{Z}\hat{\mathbf{b}} = \mathbf{ZGZ}^{\top}\mathbf{QZb} + \mathbf{ZGZ}^{\top}\mathbf{Q}\boldsymbol{e},$$

where $\mathbf{R} = \sigma^2\mathbf{I}_n$. These identities tell us that $\hat{\boldsymbol{e}}$ and $\mathbf{Z}\hat{\mathbf{b}}$ depend on $\mathbf{b}$ and $\boldsymbol{e}$ and thus are called confounded residuals; see Hilden-Minton (1995). To verify the normality of the conditional errors using only $\hat{\boldsymbol{e}}$ may be misleading because of the presence of $\mathbf{b}$ in the above formulas. Hilden-Minton (1995) defined the confounding fraction as the proportion of variability in $\widehat{\boldsymbol{e}}$ due to the presence of $\mathbf{b}$. The same work suggested the use of a linear transformation $\mathbf{L}$ such that $\mathbf{L}^{\top}\widehat{\boldsymbol{e}}$ has the least confounding fraction possible. The suggested transformation also generates uncorrelated homoskedastic residuals. It is more appropriated to analyze the assumption of normality in the conditional errors using $\mathbf{L}^{\top}\widehat{\boldsymbol{e}}$ instead of $\widehat{\boldsymbol{e}}$ as suggested by Hilden-Minton (1995) and verified by simulation in Nobre and Singer (2007).

4.1.3 EBLUP

The EBLUP is useful to identify outlier subjects given that it represents the distance between the population mean value and the value predicted for the ith subject. A way of using the EBLUP to search for outliers subjects is to use the Mahalanobis distance (see Waternaux et al., 1989), $\zeta_i = \hat{\mathbf{b}}_i^{\top}(\widehat{\mathrm{Var}}[\hat{\mathbf{b}}_i - \mathbf{b}_i])^{-1}\hat{\mathbf{b}}_i$. It is also possible to use the EBLUP to verify the random effects normality assumption. For more information; see Nobre and Singer (2007). In Table 2 we summarize diagnostic techniques involving residuals discussed in Nobre and Singer (2007).

4.2 Sensitivity analysis

Influence diagnostic techniques are used to detect observations that may produce excessive influence in the parameters estimates. There are two main approaches of such techniques: global influence, which is usually based on case deletion; and local influence, which introduces small perturbations in different components of the model.

In normal homoskedastic linear regression, examples of sensitivity measures are the Cook distance, DFFITS and the COVRATIO; see Cook (1977), Belsley et al. (1980) and Chatterjee and Hadi (1986, 1988).

4.2.1 Global influence

A simple way to verify the influence of a group of observations in the parameters estimates is to remove the group and observe the changes in the estimation. The group of observations are influential if the changes are considerably large. However, in LMM, it may not be practical to reestimate the parameters every time a set of observations is removed. To avoid doing so, Hilden-Minton (1995) presented an update formula for the

BLUE and BLUP. Let $I = \{i_1, \ldots, i_k\}$ be the index set of the removed observations and $\boldsymbol{U}_I = (\boldsymbol{U}_{i_1}, \ldots, \boldsymbol{U}_{i_k})$. Hilden-Minton (1995) showed that

$$\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(I)} = (\boldsymbol{X}^\top \boldsymbol{M} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{M} \boldsymbol{U}_I \hat{\boldsymbol{\phi}}_{(I)} \quad \text{and} \quad \hat{\boldsymbol{b}} - \hat{\boldsymbol{b}}_{(I)} = \boldsymbol{D} \boldsymbol{Z}^\top \boldsymbol{Q} \boldsymbol{U}_I \hat{\boldsymbol{\phi}}_{(I)},$$

where the subscript (I) indicates that the estimates were obtained without the observations indexed by I and $\hat{\boldsymbol{\phi}}_{(I)} = (\boldsymbol{U}_I^\top \boldsymbol{Q} \boldsymbol{U}_I)^{-1} \boldsymbol{U}_I^\top \boldsymbol{Q} \boldsymbol{y}$.

A suggestion to measure the influence on the parameters estimates in linear mixed models is to use the Cook distance (see Cook, 1977) given by

$$D_I = \frac{(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_{(I)})^\top (\boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{X})^{-1} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_{(I)})}{c} = \frac{(\boldsymbol{y} - \hat{\boldsymbol{y}}_{(I)})^\top \boldsymbol{V}^{-1} (\boldsymbol{y} - \hat{\boldsymbol{y}}_{(I)})}{c},$$

such as seen in Christensen and Pearson (1992) and Banerjee and Frees (1997), where c is a scale factor. However, it was pointed out by Tan et al. (2001) that D_I is not always able to measure the influence on the estimation properly in the mixed models class. The same authors suggest the use of a measure similar to the Cook distance, but conditional to BLUP ($\hat{\mathbf{b}}$). The conditional Cook distance is defined for the ith observation as

$$D_i^{\text{cond}} = \sum_{j=1}^{k} \frac{\mathbf{P}_{j(i)}^\top \text{Var}[\mathbf{y}|\mathbf{b}]^{-1} \mathbf{P}_{j(i)}}{(n-1)k + p}, \quad i = 1, \ldots, k,$$

where $\mathbf{P}_{j(i)} = \hat{\mathbf{y}}_j - \hat{\mathbf{y}}_{j(i)} = (\mathbf{X}_j \hat{\boldsymbol{\beta}} + \mathbf{Z}_j \hat{\mathbf{b}}_j) - (\mathbf{X}_j \hat{\boldsymbol{\beta}}_{(i)} + \mathbf{Z}_j \hat{\mathbf{b}}_{j(i)})$. The same authors decomposed $D_i^{\text{cond}} = D_{i1}^{\text{cond}} + D_{i2}^{\text{cond}} + D_{i3}^{\text{cond}}$ and commented the interpretation of each part of the decomposition. D_{i1}^{cond} is related to the influence in the fixed effects, D_{i2}^{cond} is related to the influence on the predicted values and D_{i3}^{cond} to the covariance of the BLUE and the BLUP, which should be close to zero if the model is valid.

When all the observations from a subject are deleted, it is not possible to obtain the BLUP for the random effects of that subject, making it impossible to obtain D_I^{cond} as stated above. For this purpose, Nobre (2004) suggested using $D_I^{\text{cond}} = (n_i)^{-1} \sum_{j \in I} D_j^{\text{cond}}$, where I indexes the observation from a subject, as a way to measure the influence of a subject on the parameters estimates when its observations are deleted.

There are natural extensions of leverage measures for linear mixed models. These can be seen in Banerjee and Frees (1997), Fung et al. (2002), Demidenko (2004) and Nobre (2004). However, they only provide information about leverage regarding fitted marginal values. This has two main limitations as commented in Nobre and Singer (2011). First we may be interested in detecting high-leverage within-subject observations. Second, in some cases the presence of high-leverage within-subject observations does not imply that the subject itself is detected as a high-leverage subject. Suggestions of how to evaluate the within-subject leverage may be seen in Demidenko and Stukel (2005) and Nobre and Singer (2011).

4.2.2 Local influence

The concept of local influence was proposed by Cook (1986) and consists in analyzing the sensitivity of a statistical model when subjected to small perturbations. It is suggested to use an influence measure called the likelihood displacement. Considering the model described in Equation (2), up to a constant, the log-likelihood function may be written as

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{k} L_i(\boldsymbol{\theta}) = -\frac{1}{2} \sum_{i=1}^{k} \left\{ \ln |\mathbf{V}_i| + (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})^\top \mathbf{V}^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right\}.$$

The likelihood displacement is defined as $LD(\boldsymbol{\omega}) = 2\{L(\hat{\boldsymbol{\theta}}) - L(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})\}$, where $\boldsymbol{\omega}$ is a $l \times 1$ perturbations vector in an open set $\boldsymbol{\Omega} \in \mathbb{R}^l$; $\boldsymbol{\theta}$ is the parameters vector of the model, including covariance parameters; $\hat{\boldsymbol{\theta}}$ is the ML estimate of $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ is the ML estimate when the model is perturbed. It is necessary to assume that $\boldsymbol{\omega}_0$ exists such that $L(\hat{\boldsymbol{\theta}}) = L(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}_0})$ and such that LD has its first and second derivatives in a neighborhood of $(\hat{\boldsymbol{\theta}}^\top, \boldsymbol{\omega}_0^\top)^\top$. Cook (1986) considered a $\mathbb{R}^{l+1}$ surface formed by the influence function $\boldsymbol{\alpha}(\boldsymbol{\omega}) = (\boldsymbol{\omega}^\top, LD(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}))^\top$ and the normal curvature in the vicinity of $\boldsymbol{\omega}_0$ in the direction of a vector $\mathbf{d}$, denoted by $C_{\mathbf{d}}$. In this case, the normal curvature is given by

$$C_{\mathbf{d}} = 2|\mathbf{d}^\top \mathbf{H}^\top \ddot{\mathbf{L}}^{-1} \mathbf{H} \mathbf{d}|,$$

where $\ddot{\mathbf{L}} = \partial^2 L(\boldsymbol{\theta})/\partial \boldsymbol{\theta}^\top \partial \boldsymbol{\theta}$ and $\mathbf{H} = \partial^2 L(\boldsymbol{\theta}_{\boldsymbol{\omega}})/\partial \boldsymbol{\theta}^\top \partial \boldsymbol{\omega}$ both evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$; see Cook (1986). It can be shown that $C_{\mathbf{d}}$ always lies between the minimum and maximum eigenvalue of the matrix $\ddot{\mathbf{F}} = -\mathbf{H}^\top \ddot{\mathbf{L}}^{-1} \mathbf{H}$, so $\mathbf{d}_{\max}$, the eigenvector associated to the highest eigenvalue, gives information about the direction that exhibits more sensitivity of $LD(\boldsymbol{\theta})$ in a $\boldsymbol{\omega}_0$ neighborhood. Beckman et al. (1987) made some comments on the effectiveness of the local influence approach. Lesaffre and Verbeke (1998) and Nobre (2004) showed some examples of perturbation schemes in the linear mixed models context.

Perturbation scheme for the covariance matrix of the conditional errors. To verify the sensitivity of the model to the conditional homoskedasticity assumption, perturbations are inserted in the covariance matrix of the conditional errors. This can be done by considering $\mathrm{Var}[\boldsymbol{\varepsilon}] = \sigma^2 \boldsymbol{\Lambda}(\boldsymbol{\omega})$, where $\boldsymbol{\Lambda}(\boldsymbol{\omega}) = \mathrm{diag}(\boldsymbol{\omega})$, with $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_N)^\top$, the perturbation vector. For this case we have $\boldsymbol{\omega}_0 = \mathbf{1}_N$. The log-likelihood function in this case is given by

$$L = L(\boldsymbol{\theta}_{\boldsymbol{\omega}}) = -\frac{1}{2}\left\{\ln|\boldsymbol{V}(\boldsymbol{\omega})| + (\boldsymbol{y} - \boldsymbol{X\beta})^\top \boldsymbol{V}(\boldsymbol{\omega})^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})\right\},$$

where $\boldsymbol{V}_{\boldsymbol{\omega}} = \boldsymbol{ZDZ}^\top + \sigma^2 \boldsymbol{\Lambda}(\boldsymbol{\omega})$.

Perturbation scheme for the response. For the local influence approach, Beckman et al. (1987) proposed the perturbation scheme

$$\boldsymbol{y}(\boldsymbol{\omega}) = \boldsymbol{y} + s\boldsymbol{\omega},$$

where s represents a scale factor and $\boldsymbol{\omega}$ is a $n \times 1$ perturbation vector. For this scheme we have $\boldsymbol{\omega}_0 = \mathbf{0}$, with $\mathbf{0}$ representing the $n \times 1$ null vector. In this case, the perturbed log-likelihood function is proportional to

$$L(\boldsymbol{\theta}_{\boldsymbol{\omega}}) = -\frac{1}{2}(\boldsymbol{y} + s\boldsymbol{\omega} - \boldsymbol{X\theta})^\top \boldsymbol{V}^{-1}(\boldsymbol{y} + s\boldsymbol{\omega} - \boldsymbol{X\beta}).$$

Perturbation scheme for the random effects covariance matrix. It is possible to assess the sensitivity of the model in relation to the random effects homoskedasticity assumption by perturbing the matrix $\mathbf{G}$. Nobre (2004) suggested the use of $\mathrm{Var}[\boldsymbol{b}_i] = \omega_i \boldsymbol{G}$ as a perturbation scheme. In this case $\boldsymbol{\omega}$ is a $q \times 1$ vector and $\boldsymbol{\omega}_0 = \mathbf{1}_q$. The perturbed log-likelihood function is proportional to

$$L(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{i=1}^{k}\left\{\ln|\boldsymbol{V}_i(\boldsymbol{\omega})| + (\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^{-1}\boldsymbol{V}(\boldsymbol{\omega})^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})\right\}.$$

Perturbation scheme for the weighted case. Verbeke (1995) and Lesaffre and Verbeke (1998) suggested perturbing the log-likelihood function as

$$L(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^{k} \omega_i L_i(\boldsymbol{\theta}).$$

Such a perturbation scheme is appropriate for measuring the influence of the ith subject using the normal curvature in its direction and is given by

$$C_i = 2|\mathbf{d}_i^\top \mathbf{H}^\top \ddot{\mathbf{L}}^{-1} \mathbf{H} \mathbf{d}_i|,$$

where $\mathbf{d}_i$ is a vector whose entries are 1 in the ith coordinate and zero everywhere else. Verbeke (1995) showed that if C_i has a high value, then the ith subject has great influence in the value of $\widehat{\boldsymbol{\theta}}$. A threshold of twice the mean value of all C_j's helps to decide whether or not the observation is influential.

Lesaffre and Verbeke (1998) extracted from C_i some interpretable measures. They especially propose using $\|\mathcal{X}_i\mathcal{X}_i^\top\|^2$, $\|\mathcal{R}_i\|^2$ $\|\mathcal{Z}_i\mathcal{Z}_i^\top\|^2$, $\|\mathbf{I}_{n_i} - \mathcal{R}_i\mathcal{R}_i^\top\|^2$ and $\|\widehat{\mathbf{V}}_i^{-1}\|^2$, where $\mathcal{X}_i = \hat{\mathbf{V}}^{-1/2}\mathbf{X}_i$, $\mathcal{Z}_i = \hat{\mathbf{V}}_i^{-1/2}\mathbf{Z}_i$, $\mathcal{R}_i = \hat{\mathbf{V}}_i^{-1/2}\hat{\boldsymbol{e}}_i$, to evaluate the influence of the ith subject in the model parameter estimates. The actual interpretation of each of these terms can be seen in the original paper.

4.2.3 Conform local influence

The C_d measure proposed by Cook (1986) is not invariant to scale re-parametrization. To obtain a similar standardized measure and make it more comparable, Poon and Poon (1999) used the conform normal curvature instead of the normal curvature given by

$$B_d(\boldsymbol{\theta}) = \frac{2|\mathbf{d}^\top \mathbf{H}^\top \ddot{L}^{-1} \mathbf{H}\mathbf{d}|}{\|2\mathbf{H}^\top \ddot{L}^{-1}\mathbf{H}\|}.$$

It can be shown that $0 \leq B_d(\boldsymbol{\theta}) \leq 1$ to $\mathbf{d}$ direction and that B_d is invariant to conform scale re-parametrization. A re-parametrization is said to be conform if its jacobian $\mathbf{J}$ is such that $\mathbf{J}^\top\mathbf{J} = t\mathbf{I}_s$, to some real t and integer s. They showed that if $\lambda_1, \ldots, \lambda_l$ are the eigenvalues of $\ddot{\mathbf{F}}$ matrix with $\mathbf{v}_1, \ldots, \mathbf{v}_l$ representing the respective normalized eigenvectors, then the value of the conform normal curvature in $\mathbf{v}_i$ direction is equal to $\lambda_i/\sqrt{\sum_{i=1}^{l} \lambda_i^2}$ and $\sum_{i=1}^{l} B_{v_i}^2(\boldsymbol{\theta}) = 1$. If every eigenvector has the same conform normal curvature, its value is equal to $1/\sqrt{l}$. Poon and Poon (1999) proposed to use this measure as a referential to measure the intensity of the local influence of an eigenvector. It can also be shown that when $\mathbf{d}$ has the direction of $\mathbf{d}_{\max}$ the conform normal curvature also attains its maximum. In this way, the normal curvature and the conform normal curvature are equivalent methods.

5. Application

According to Frees et al. (1999), the random coefficient models are equivalent to the Hachemeister linear regression model which is used for the example data in Hachemeister (1975). The random coefficient model to the data in Table 1 may be described as

$$y_{ij} = \alpha_i + j\beta_i + e_{ij}, \quad i = 1, \ldots, 5, \quad j = 1, \ldots, 12,$$

where y_{ij} represents the average claim amount for state i in the jth trimester, $\alpha_i = \alpha + a_i$ and $\beta_i = \beta + b_i$, with fixed α and β, and $(a_i, b_i)^\top \sim \mathcal{N}_2(\mathbf{0}, \mathbf{D})$, in which $\mathbf{D}$ is a 2×2 covariance matrix. In order to find a possible simpler model, we used `R` to apply the asymptotic likelihood ratio test described in Giampaoli and Singer (2009) to compare the suggested random coefficients model and a random intercept model. The p-value obtained from the test was 0.0514. It indicates that it may be enough to consider the random effect for the intercept only. This decision is also supported by the Bayesian information criterion (BIC), which is equal to 808.3 for the single random effect model and 811.6 for the model with two random effects. We may also use another set of tests, involving bootstrap, monte-carlo and permutational methods, to investigate whether or not should we prefer the random intercept model. These tests may be seen in Crainiceanu and Ruppert (2004), Greven et al. (2008) and Fitzmaurice et al. (2007). However, this is very distant from our goals and is not discussed here. For the sake of simplicity and based on the presented reasons we shall use the random intercept model, which differs a little from the model proposed by Frees et al. (1999). Thus, the model to be adjusted for the data in this example is

$$y_{ij} = \alpha_i + j\beta + \epsilon_{ij}, \quad i = 1, \ldots, 5, \quad j = 1, \ldots, 12, \tag{4}$$

where $\alpha_i = \alpha + a_i$, β and ϵ are the same as defined before. Assume also that $\mathrm{Var}[\epsilon_{ij}] = \sigma^2_\epsilon$ and $\mathrm{Var}[a_i] = \sigma^2_a$.

The model parameter estimates were obtained by the RML method using the `lmer()` function from the `lme4` package in `R`. The standard errors were obtained from `SAS` © (SAS Institute Inc., 2004) using the proc `MIXED`. The estimates are shown in Table 3.

Table 3. Model parameter estimates.

Parameter	α	β	σ^2_ϵ	σ^2_a
Estimate	1460.32	32.41	32981.53	73398.25
SE	131.07	6.79	6347.17	24088.00

Figure 3 shows the five conditional regression lines obtained from the linear mixed model given in Equation (4). The adjusted model clearly suggests that the claim amount is higher in state 1. Also it suggests a similarity in the claim amounts from states 2 and 4. Besides that, we can expect a smaller risk from policies in state 5, since they are much closer to the respective adjusted conditional line. Further information is explored by the diagnostic analysis commented next.

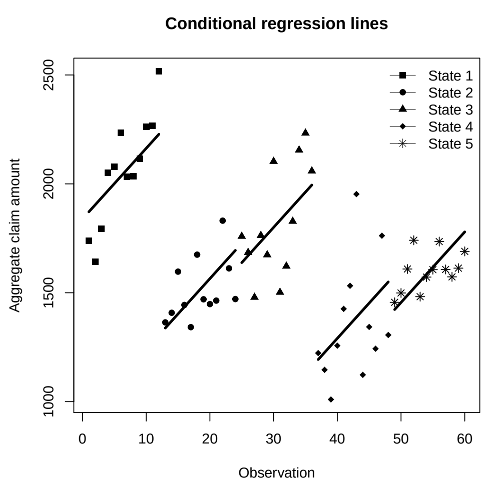

Figure 3. Conditional regression lines.

5.1 Diagnostic analysis

The standardized residuals proposed by Nobre and Singer (2007) suggest that observation 4.7 (obtained from state 4 in the seventh trimester) may be considered an outlier as shown in Figure 4(a). According to the QQ plot in Figure 4(b) it is reasonable to assume that the conditional errors are normally distributed. The Mahalanobis distance in Figure 4(c) was normalized to fit the interval $[0, 1]$ and suggests that the first state may be an outlier. The measure M_I proposed by Nobre and Singer (2007) in Figure 4(d), also normalized, suggests that none of the states have outlier observations. The Mahalonobis distance should not be confounded with M_I. The first is based on the EBLUP and the last is based on the conditional errors, and thus they have different meanings. For both analyses, an observation is highlighted if it is greater than twice the mean of the measures.

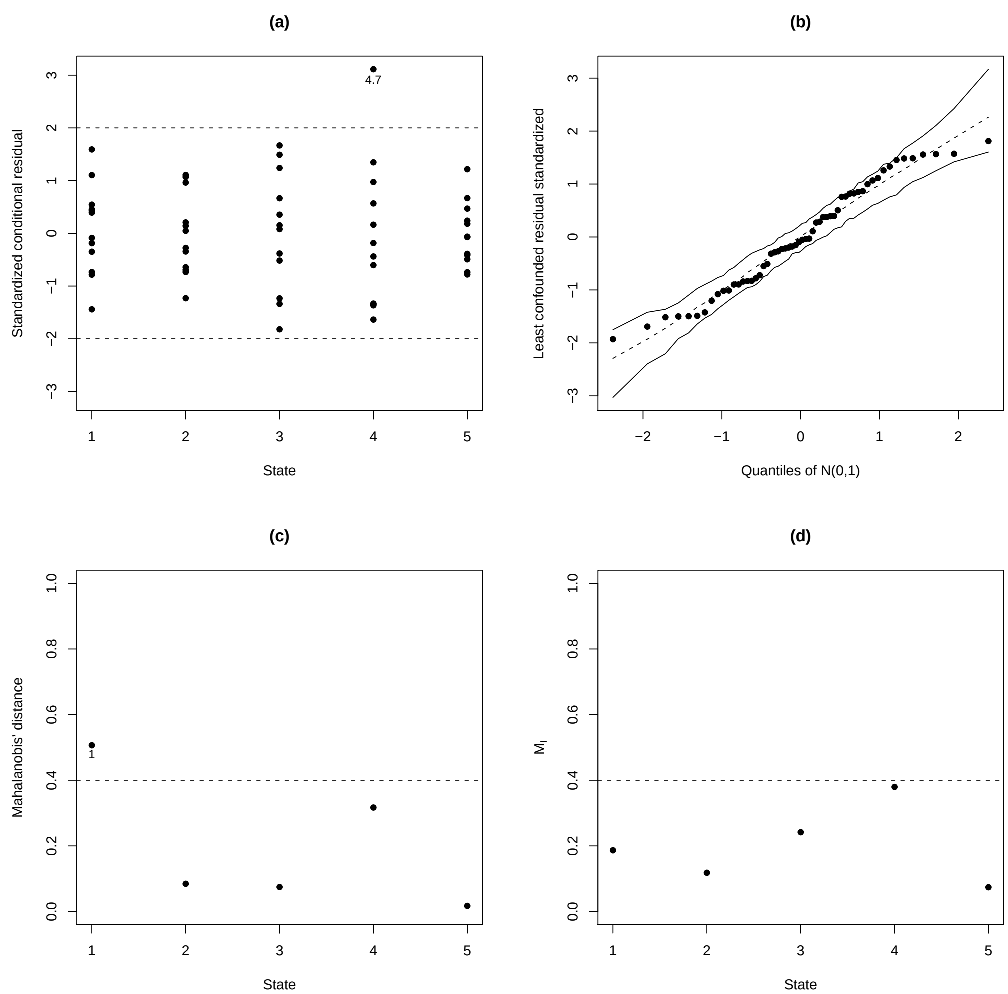

Figure 4. Residual analysis: (a) standardized residuals, (b) least confounding residuals, (c) EBLUP, (d) values for M_I.

The conditional Cook distance is shown in Figure 5. The distances were normalized for comparison. Figure 5(a) suggests that observation 4.7 is influential in the model estimates. The first term of the distance decomposition suggests that no observations were influential in the estimate of $\boldsymbol{\beta}$ as shown in Figure 5(b). The second term of the decomposition suggests that observation 4.7 is potentially influential in the prediction of $\mathbf{b}$ as seen on Figure 5(c). The last term, D_{i3}, is as close to zero as expected and is omitted.

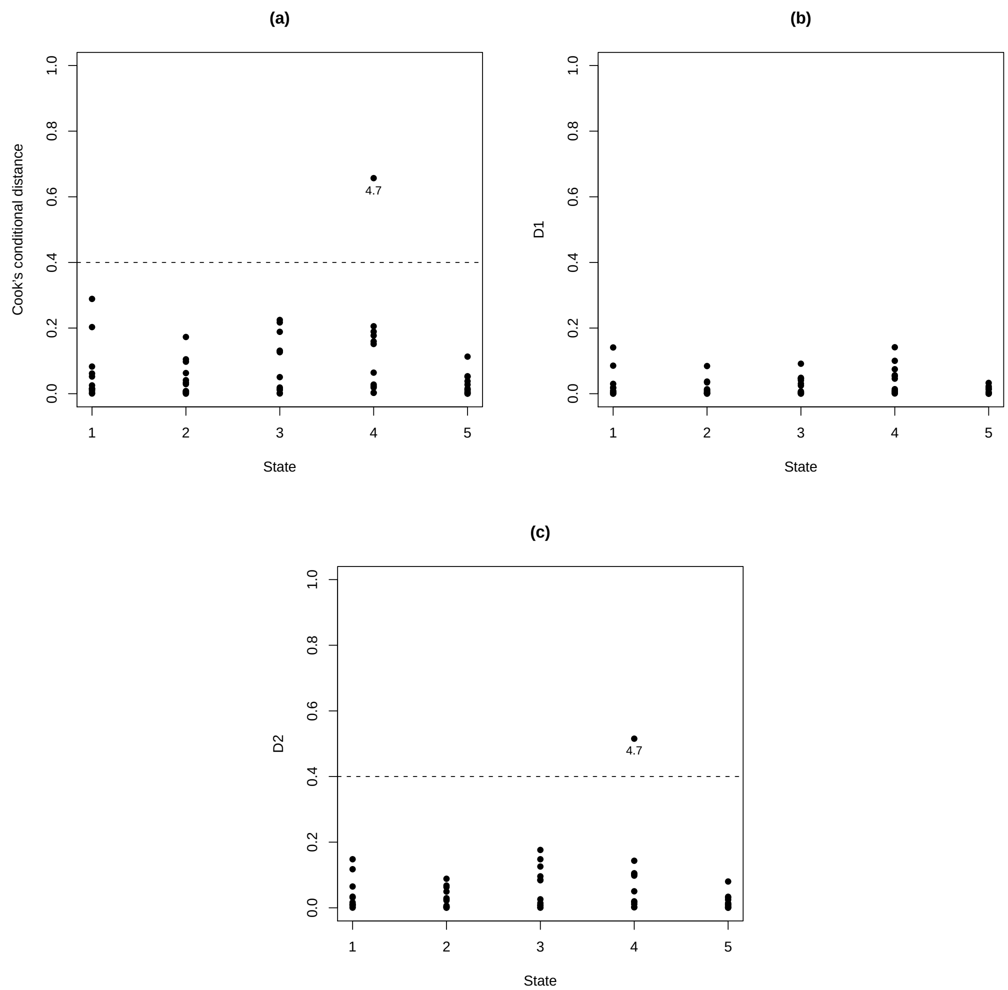

Figure 5. (a) Conditional Cook distance, (b) D_{i1}, (c) D_{i2}.

Figure 6 shows the local influence analysis using three different perturbation schemes. The first, in Figure 6(a), is related to the conditional errors covariance matrix, as suggested in Beckman et al. (1987), and indicates that the observations from the fourth state, especially 4.7, are possibly influential in the homoskedasticity and independence assumption for the conditional errors. Notice that it is possible to explain the influence of observation 4.7 analyzing Figure 2. This observation has a value considerably higher than the others from the same state. Figure 6(b) demonstrates the perturbation scheme for the covariance matrix associated to the random effects as shown in Nobre (2004). Alternative perturbation schemes for this case can be seen at Beckman et al. (1987). These schemes suggest that all states are equally influential in the random effects covariance matrix estimate. Finally, there are evidences that the observations in the fourth state may not be well predicted by the model; see Figure 6(c).

After the diagnostic we proceed to a confirmatory analysis by removing the observations from states 1 and 4, one at a time and then both at the same time. The new estimates are shown in Table 4. For each parameter, we calculate the relative change in the estimated values, defined for parameter θ, as

$$RC(\theta) = \left| \frac{\hat{\theta}_{(i)} - \hat{\theta}}{\hat{\theta}} \right| \times 100\%.$$

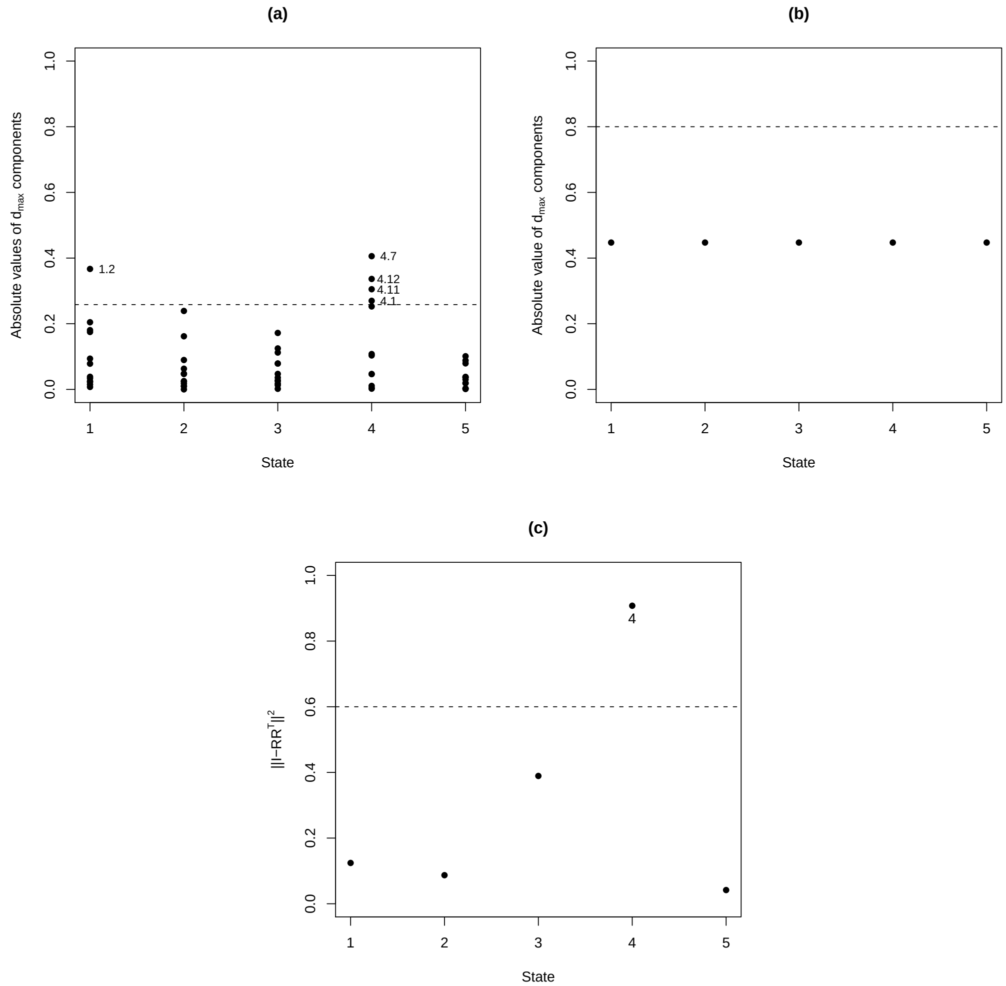

Figure 6. Perturbation schemes: (a) conditional covariance matrix, (b) random effects' covariance matrix, (c) values for $\|\mathbf{I}_{n_i} - \mathcal{R}_i\mathcal{R}_i^\top\|^2$.

Table 4. Estimates and relative changes for the model given in Equation (4) parameter estimates with and without states 1 and 4.

Situation	α	β	σ_ϵ^2	σ_a^2
Complete data	1460.32	32.41	32981.53	73398.25
Without State 1	1408.63 (3.67%)	25.26 (22.06%)	34666.31 (5.11%)	34335.64 (53.22%)
Without State 4	1530.94 (4.61%)	33.50 (3.36%)	24940.12 (24.38%)	59214.50 (19.32%)
Without States 1 and 4	1485.56 (1.70%)	24.32 (24.96%)	24497.48 (25.72%)	23707.07 (67.70%)

If all five states were equally influential, we would expect the value for RC to lie around $1/5 = 20\%$ after removing a state. If $RC(\theta)$ exceeds two times this value, that is 40%, for some parameter θ we consider the state was potentially influential. It is possible to conclude that three observations from state 1 were influential in the within-subject variance estimate. From Figure 2, one can explain this influence noticing that all the observations from state 1 had higher values compared to the other states. Notice that such influence was not detected in Figure 5(b), but was pointed out by the Mahalanobis distance in Figure 4(c). Removing state 1 from our analysis and running every diagnostic procedure again we detect no excessive influence and the only issue is the observation 4.7, which is still an outlier. From this result the model is validated and it is assumed to be robust and ready for use.

6. Conclusions

The use of linear mixed models in actuarial science should be encouraged given their capability to model the within-subject correlation, their flexibility and the presence of diagnostic tools. Insurers should not use a model without validating it first. For the specific example seen here, the decision makers may consider a different approach for state 1. After removing observations from state 1 there was a relative change of more than 50% in the random effect variance estimate, which reflects significantly in the premium estimate. Such analysis would not be possible in the traditional credibility models approach. This illustrates how the model can be used to identify different sources of risk and can be used in portfolio management. Linear mixed models are also usually easier to understand and to present, when compared to standard actuarial methods, such as the credibility models and Bayesian approach for determining the fair premium. The natural extension of this work is to repeat the estimation and diagnostic procedures, adapting what is necessary, to the generalized linear mixed models, which are also useful to actuarial science. Some works have already been made in this area; see, e.g., Antonio and Beirlant (2006). It is also interesting to continue a further analysis of the example in Hachemeister (1975), using the diagnostic procedures again when weights are introduced to the covariance matrix of the conditional residuals in the random coefficient models, and to evaluate the robustness of the linear mixed models equivalent to the other classic credibility models. Again, this care is justified because the fairest premium is more competitive in the market.

Appendix

We present here expressions for matrix $\boldsymbol{H}$ and the derivatives seen in the different perturbation schemes presented in Section 4.2.2. These calculations are taken from Nobre (2004) and are presented here to make this text more self-content.

Appendix A. Perturbation Scheme for the Covariance Matrix of the Conditional Errors

Let $\boldsymbol{H}^{(k)}$ be the kth column of $\boldsymbol{H}$ and f be the number of distinct components of matrix $\boldsymbol{D}$, then

$$\boldsymbol{H}^{(k)} = \left\{ \left(\frac{\partial^2 L(\boldsymbol{\omega})}{\partial \omega_k \partial \boldsymbol{\beta}} \right)^\top, \frac{\partial^2 L(\boldsymbol{\omega})}{\partial \omega_k \partial \sigma^2}, \frac{\partial^2 L(\boldsymbol{\omega})}{\partial \omega_k \partial \theta_1}, \ldots, \frac{\partial^2 L(\boldsymbol{\omega})}{\partial \omega_k \partial \theta_f} \right\}^\top,$$

where

$$\frac{\partial^2 L(\boldsymbol{\omega})}{\partial \omega_k \partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}};\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \boldsymbol{X}^\top \boldsymbol{D}_k \hat{\boldsymbol{e}},$$

$$\frac{\partial^2 L(\boldsymbol{\omega})}{\partial \omega_k \partial \sigma^2} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}};\boldsymbol{\omega}=\boldsymbol{\omega}_0} = -\frac{1}{2} \left\{ \hat{\sigma}^{-2} \operatorname{tr} \left[\boldsymbol{D}_k \boldsymbol{Z} \widehat{\boldsymbol{D}} \boldsymbol{Z}^\top \right] - 2\hat{\boldsymbol{e}}^\top \boldsymbol{D}_k \widehat{\boldsymbol{V}}^{-1} \hat{\boldsymbol{e}} + \hat{\sigma}^{-2} \hat{\boldsymbol{e}}^\top \boldsymbol{D}_k \hat{\boldsymbol{e}} \right\},$$

$$\frac{\partial^2 L(\boldsymbol{\omega})}{\partial \omega_k \partial \theta_i} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}};\boldsymbol{\omega}=\boldsymbol{\omega}_0} = -\frac{1}{2} \left\{ \operatorname{tr} \left[\boldsymbol{D}_k \boldsymbol{Z} \dot{\boldsymbol{D}}_i \boldsymbol{Z}^\top \right] - 2\hat{\boldsymbol{e}}^\top \boldsymbol{D}_k \boldsymbol{Z} \dot{\boldsymbol{D}}_i \boldsymbol{Z}^\top \hat{\boldsymbol{e}} \right\},$$

with

$$\boldsymbol{D}_k = \frac{\partial \boldsymbol{V}^{-1}(\boldsymbol{\omega})}{\partial \omega_k}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}};\boldsymbol{\omega}=\boldsymbol{\omega}_0} = -\sigma^2 \widehat{\boldsymbol{V}}^{(k)}(\widehat{\boldsymbol{V}}^{(k)})^\top, \quad \text{e} \quad \dot{\boldsymbol{D}}_i = \frac{\partial \boldsymbol{D}}{\partial \theta_i}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}};\boldsymbol{\omega}=\boldsymbol{\omega}_0},$$

and $\boldsymbol{V}^{(k)}$ representing the kth column of $\boldsymbol{V}^{-1}$.

Appendix B. Perturbation Scheme for the Response

It can be shown that

$$\frac{\partial^2 L(\boldsymbol{\omega})}{\partial \boldsymbol{\omega} \partial \boldsymbol{\beta}}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}};\boldsymbol{\omega}=\boldsymbol{\omega}_0} = s\widehat{\boldsymbol{V}}^{-1}\boldsymbol{X},$$

$$\frac{\partial^2 L(\boldsymbol{\omega})}{\partial \boldsymbol{\omega} \partial \sigma^2}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}};\boldsymbol{\omega}=\boldsymbol{\omega}_0} = s\widehat{\boldsymbol{V}}^{-1}\widehat{\boldsymbol{V}}^{-1}\hat{\boldsymbol{e}},$$

$$\frac{\partial^2 L(\boldsymbol{\omega})}{\partial \boldsymbol{\omega} \partial \theta_i}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}};\boldsymbol{\omega}=\boldsymbol{\omega}_0} = s\boldsymbol{V}^{-1}\boldsymbol{Z}\dot{\boldsymbol{D}}_i\boldsymbol{Z}^\top\widehat{\boldsymbol{V}}^{-1}\hat{\boldsymbol{e}},$$

implying

$$\boldsymbol{H}^\top = s\boldsymbol{V}^{-1}\left[\boldsymbol{X}, \widehat{\boldsymbol{V}}^{-1}\hat{\boldsymbol{e}}, \boldsymbol{Z}\dot{\boldsymbol{D}}_1\boldsymbol{Z}^\top\widehat{\boldsymbol{V}}^{-1}\hat{\boldsymbol{e}}, \ldots, \boldsymbol{Z}\dot{\boldsymbol{D}}_f\boldsymbol{Z}^\top\widehat{\boldsymbol{V}}^{-1}\hat{\boldsymbol{e}}\right].$$

Appendix C. Perturbation Scheme for the Random Effects Covariance Matrix

For this scheme we have

$$\boldsymbol{H}^{(k)} = \left\{\left(\frac{\partial^2 L(\boldsymbol{\omega})}{\partial \omega_k \partial \boldsymbol{\beta}}\right)^\top, \frac{\partial^2 L(\boldsymbol{\omega})}{\partial \omega_k \partial \sigma^2}, \frac{\partial^2 L(\boldsymbol{\omega})}{\partial \omega_k \partial \theta_1}, \ldots, \frac{\partial^2 L(\boldsymbol{\omega})}{\partial \omega_k \partial \theta_f}\right\}^\top.$$

It can be shown that

$$\frac{\partial^2 L(\boldsymbol{\omega}_k)}{\partial \boldsymbol{\omega} \partial \boldsymbol{\beta}}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}};\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \boldsymbol{X}_k^\top\widehat{\boldsymbol{V}}_k^{-1}\boldsymbol{Z}_k\widehat{\boldsymbol{G}}\boldsymbol{Z}_k^\top\widehat{\boldsymbol{V}}_k^{-1}\hat{\boldsymbol{e}}_k,$$

$$\frac{\partial^2 L(\boldsymbol{\omega}_k)}{\partial \boldsymbol{\omega} \partial \sigma^2}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}};\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \operatorname{tr}\left[\widehat{\boldsymbol{V}}_k^{-1}\boldsymbol{Z}_k\widehat{\boldsymbol{G}}_k\boldsymbol{Z}_k^\top\right]^\top - 2\hat{\mathrm{e}}_k^\top\widehat{\boldsymbol{V}}_k^{-1}\boldsymbol{Z}_k\widehat{\boldsymbol{G}}\boldsymbol{Z}_k^\top\widehat{\boldsymbol{V}}_k^{-1}\widehat{\boldsymbol{V}}_k^{-1}\hat{\mathrm{e}}_k,$$

$$\frac{\partial^2 L(\boldsymbol{\omega}_k)}{\partial \boldsymbol{\omega} \partial \theta_i}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}};\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \operatorname{tr}\left[\widehat{\boldsymbol{V}}_k^{-1}\boldsymbol{Z}_k\widehat{\boldsymbol{G}}_k\boldsymbol{Z}_k^\top\widehat{\boldsymbol{V}}_k^{-1}\boldsymbol{Z}_k\dot{\boldsymbol{G}}_i\boldsymbol{Z}_k^\top\right] - \hat{\boldsymbol{e}}_k^\top\widehat{\boldsymbol{V}}_k^{-1}\boldsymbol{Z}_k\widehat{\boldsymbol{G}}\boldsymbol{Z}_k^\top\widehat{\boldsymbol{V}}_k^{-1}\boldsymbol{Z}_k\dot{\boldsymbol{G}}_i\boldsymbol{Z}_k^\top\widehat{\boldsymbol{V}}_k^{-1}\hat{\boldsymbol{e}}_k.$$

Acknowledgements

We are grateful to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq project # 564334/2008-1) and Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP), Brazil, for partial financial support. We also thank an anonym referee and the executive editor for their careful and constructive review.

References

Antonio, K., Beirlant, J., 2006. Actuarial statistics with generalized linear mixed models. Insurance: Mathematics and Economics, 75, 643–676.

Banerjee, M., 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., Cook, R.D., 1987. Diagnostics for mixed-model analysis of variance. Technometrics, 29, 413–426.

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

Bühlmann, H., 1967. Experience rating and credibility. ASTIN Bulletin, 4, 199–207.

Chatterjee, S., Hadi, A.S., 1986. Influential observations, high leverage points, and outliers in linear regression (with discussion). Statistical Science, 1, 379–393.

Chatterjee, S., Hadi, A.S., 1988. Sensitivity Analysis in Linear Regression. John Wiley & Sons, New York.

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

Cook, R.D., 1977. Detection of influential observation in linear regression. Technometrics, 19, 15–28.

Cook, R.D., 1986. Assessment of local influence (with discussion). Journal of The Royal Statistical Society Series B - Statistical Methodology, 48, 117–131.

Cook, R.D., Weisberg, S., 1982. Residuals and Influence in Regression. Chapman and Hall, London.

Crainiceanu, C.M., Ruppert, D., 2004. Likelihood ratio tests in linear mixed models with one variance component. Journal of The Royal Statistical Society Series B - Statistical Methodology, 66, 165–185.

Dannenburg, D.R., Kaas, R., Goovaerts, M.J., 1996. Practical actuarial credibility models. Institute of Actuarial Science and Economics, University of Amsterdam, Amsterdam.

Demidenko, E., 2004. Mixed Models - Theory and Applications. Wiley, New York.

Demidenko, E., Stukel, T.A., 2005. Influence analysis for linear mixed-effects models, Statistics in Medicine, 24, 893–909.

Diggle, P.J., Heagerty, P., Liang, K.Y., Zeger, S.L., 2002. Analysis of Longitudinal Data. Oxford Statistical Science Series.

Drapper, N.R., Smith, N., 1998. Applied Regression Analysis. Wiley, New York.

Dutang, C., Goulet, V., Pigeon, M., 2008. Actuar: an R package for actuarial science. Journal of Statistical Software, 25, 1–37.

Fitzmaurice, G.M., Lipsitz, S.R., Ibrahim, J.G., 2007. A note on permutation tests for variance components in multilevel generalized linear mixed models. Biometrics, 63, 942–946.

Frees, E.W., Young, V.R., Luo, Y., 1999. A longitudinal data analysis interpretation of credibility models. Insurance: Mathematics and Economics, 24, 229–247.

Fung, W.K., Zhu, Z.Y., Wei, B.C., He, X., 2002. Influence diagnostics and outliers tests for semiparametric mixed models. Journal of The Royal Statistical Society Series B - Statistical Methodology, 64, 565–579.

Giampaoli, V., Singer, J., 2009. Restricted likelihood ratio testing for zero variance components in linear mixed models. Journal of Statistical Planning and Inference, 139, 1435–1448.

Greven, S., Crainiceanu, C.M., Kuchenhoff, H., Peters, A., 2008. Likelihood ratio tests for variance components in linear mixed models. Journal of Computational and Graphical Statistics, 17, 870–891.

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.

Hachemeister, C.A., 1975. Credibility for regression models with application to trend. Proceedings of the Berkeley Actuarial Research Conference on Credibility, pp. 129–163.

Hilden-Minton, J.A., 1995. Multilevel diagnostics for mixed and hierarchical linear models. Ph.D. Thesis. University of California, Los Angeles.

Hoaglin, D.C., Welsch, R.E., 1978. The hat matrix in regression and ANOVA. American Statistical Association, 32, 17–22.

Johnson, R.A., Whichern, D.W., 1982. Applied Multivariate Stastical Analysis. Sixth edition. Prentice Hall. pp. 273–332.

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

Liang, K.Y., Zeger, S.L., 1986. Longitudinal analysis using generalized linear models. Biometrika, 73, 13–22.

McCulloch, C.E, Searle, S.R., 2001. Generalized, Linear, and Mixed Models. Wiley, New York.

Nobre, J.S., 2004. Mtodos de diagnstico para modelos lineares mistos. Unpublished Master Thesis (in portuguese). IME/USP, São Paulo.

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

Nobre, J.S., Singer, J.M., 2011. Leverage analysis for linear mixed models. Journal of Applied Statistics, 38, 1063–1072.

Patterson, H.D., Thompson, R., 1971. Recovery of interblock information when block sizes are unequal. Biometrika, 58, 545–554.

Poon, W.Y., Poon, Y.S., 1999. Conformal normal curvature and assessment of local influence. Journal of The Royal Statistical Society Series B - Statistical Methodology, 61, 51–61.

R Development Core Team, 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

SAS Institute Inc., 2004. SAS 9.1.3 Help and Documentation. SAS Institute Inc., Cary, North Carolina.

Tan, F.E.S., Ouwens, M.J.N., Berger, M.P.F., 2001. Detection of influential observation in longitudinal mixed effects regression models. The Statistician, 50, 271–284.

Verbeke, G., 1995. The linear mixed model. A critical investigation in the context of longitudinal data analysis. Ph.D. Thesis. Catholic University of Leuven, Faculty of Science, Department of Mathematics, Leuven, Belgium.

Verbeke, G., Molenberghs, G., 2000. Linear Mixed Models for Longitudinal Data. Springer.

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

Waternaux, C., Laird, N.M., Ware, J.H., 1989. Methods for analysis of longitudinal data: blood-lead concentrations and cognitive development. Journal of the American Statistical Association, 84, 33–41.

Zewotir, T., Galpin, J.S., 2005. Influence diagnostics for linear mixed models. Journal of Data Science, 3, 53–177.