TEST
https://doi.org/10.1007/s11749-018-0613-3

ORIGINAL PAPER

Influence diagnostics in mixed effects logistic regression models

Alejandra Tapia[1] · **Victor Leiva**[2] · **Maria del Pilar Diaz**[3] · **Viviana Giampaoli**[4]

Received: 14 March 2018 / Accepted: 9 September 2018

Abstract
Correlated binary responses are commonly described by mixed effects logistic regression models. This article derives a diagnostic methodology based on the Q-displacement function to investigate local influence of the responses in the maximum likelihood estimates of the parameters and in the predictive performance of the mixed effects logistic regression model. An appropriate perturbation strategy of the probability of success is established, as a form of assessing the perturbation in the response. The diagnostic methodology is evaluated with Monte Carlo simulations. Illustrations with two real-world data sets (balanced and unbalanced) are conducted to show the potential of the proposed methodology.

Keywords Approximation of integrals · Correlated binary responses · Metropolis–Hastings and Monte Carlo methods · Probability of success · R software

Mathematics Subject Classification 62J20 · 62J12

1 Introduction

Correlated binary response variables (responses hereafter) frequently occur in several areas such as agriculture, economics, medicine, psychology and sociology. The correlation of binary responses may be attributed to grouped data, longitudinal data

✉ Victor Leiva
victorleivasanchez@gmail.com
http://www.victorleiva.cl

1 Institute of Statistics, Faculty of Economic and Administration Sciences, Universidad Austral de Chile, Valdivia, Chile

2 School of Industrial Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

3 School of Nutrition, Faculty of Medical Sciences and INICSA-CONICET, Universidad Nacional de Córdoba, Córdoba, Argentina

4 Institute of Mathematics and Statistics, Universidade de São Paulo, São Paulo, Brazil

Published online: 21 September 2018

or repeated measurements; see, for example, Stiratelli et al. (1984), Diggle et al. (1996) and Hosmer et al. (2013). A commonly used model for analyzing this kind of data is the mixed effects logistic regression model (MELRM hereafter). This model belongs to the class of generalized linear mixed models (McCulloch and Searle 2001; Jiang 2007), which accommodate the correlation structure through unobservable random variables considered as random effects. Thus, conditional on random effects, the responses have a distribution belonging to the exponential family, which in the case of the MELRM corresponds to the Bernoulli distribution with parameter given by the probability of success. There, the random effects are assumed to follow a multivariate normal distribution with zero mean vector and unknown variance–covariance matrix.

The estimation of parameters in generalized linear mixed models is often conducted with the maximum likelihood (ML) method. However, it is not an easy computational task, because the incorporation of random effects leads to a likelihood function of observed data that includes an integral, which generally does not have an analytical solution and is besides of high dimension. To solve this integral, numerical methods such as approximations of Laplace–AL—(Raudenbush et al. 2000) and adaptive Gauss–Hermite quadrature—AGHQ—(Pinheiro and Chao 2006) are often considered. Other alternative methods of estimation were also proposed for this type of models, such as restricted pseudo-likelihood (Wolfinger and O'Connell 1993) and penalized quasi-likelihood (Breslow and Clayton 1993). In particular, for the MELRM, the integral has no analytical solution and the above-mentioned procedures have been considered with different variants in the model and complexity of implementation; see, for example, Lesaffre and Spiessens (2001), Capanu et al. (2013) and Demidenko (2013).

The estimation of MELRM parameters is often carried out to predict the binary response with future observations (or measurements). In order to evaluate the predictive performance of this model, different indicators are used, such as accuracy (Acc), sensitivity (Sen) and specificity (Spe); see, for example, Hosmer et al. (2013).

A relevant issue which should be considered in all statistical modeling, once the estimation procedure is performed, is the influence diagnostic. Particularly, assessment of the stability of the ML estimates with respect to different schemes of uncertainty in the model or in the data is a widely studied topic. A technique to conduct this assessment is the deletion case, which analyzes the stability of the parameter estimates after removing an observation; see, for example, the classic book by Cook and Weisberg (1982). For other type of stability studies and biological applications of statistical models, see Stehlík et al. (2008). However, nowadays, the most studied diagnostic technique among researchers is local influence, which evaluates the stability of estimates under small perturbations in the model or data; see Cook (1986). It identifies the presence of cases that, under small modifications in the model or data, provoke large changes in the parameter estimates. The local influence technique has been applied to other statistical models than the original normal linear regression model. Some recent works on local influence include, for example, elliptical linear regression models (Liu 2000), log-linear negative binomial models (Svetliza and Paula 2001), missing data models (Zhu and Lee 2001), multivariate regression models (Díaz-García et al. 2003; Marchant et al. 2016), linear logistic regression models (Hossain and Islam 2003), time series models under elliptical distributions (Liu 2004), multinomial models (Nyangoma et al. 2006),

beta regression models (Rocha and Simas 2011), symmetric semiparametric models (Ibacache-Pulgar et al. 2013), spatial models (Assumpção et al. 2014; Bastiani et al. 2015; Garcia-Papani et al. 2018), generalized linear type models (Leiva et al. 2014), vector autoregressive models (Liu et al. 2015), varying precision models (Santos-Neto et al. 2016) and survival analysis models (Leão et al. 2017).

The earliest work using the local influence approach in linear mixed models is attributed to Lesaffre and Verbeke (1998). As local influence is a likelihood-based technique, its usage in generalized linear mixed models has the same problem of integrals above mentioned, which was solved by Ouwens et al. (2001). As mentioned, Zhu and Lee (2001) derived local influence for missing data models and there defined a type of likelihood displacement which is known as Q-displacement function. Based on this work, Zhu and Lee (2003) proposed to treat the random effects of generalized linear mixed models as missing data and used the conditional expectation of the complete-data log-likelihood function to estimate the model parameters and to detect local influence. An important issue in the local influence analysis corresponds to the selection of an appropriate perturbation. This is because, in general, it is known that to arbitrarily perturb the model or data can lead to unreliable results in relation to observations detected as influential. For this purpose, Zhu et al. (2007) proposed a methodology for selecting an appropriate perturbation, which is based on the observed-data log-likelihood function. Chen et al. (2010) used this perturbation selection for local influence analysis in generalized linear mixed models. The most recent work on local influence for generalized linear mixed models is attributed to Rakhmawati et al. (2017). However, note that within the family of generalized linear mixed models we have members with a discrete response, as particular cases. In these cases, it is not possible to perturb the response as usual in local influence techniques. To our best knowledge, local influence studies to detect how a binary response affects the estimates and predictive performance of the MELRM have not been addressed to date.

The MELRM is widely used in different areas to analyze data with binary response, covariates and random effects. Due to the discrete nature of this response, taking only zero and one values, standard local influence diagnostic techniques do not apply for perturbing the response of this model. Then, one can perturb the probability of success associated with the binary response of the MELRM. Therefore, the novelty of the present work is in deriving local influence for the MELRM, perturbing the associated probability of success. We use the local influence technique based on the Q-displacement function and an appropriate perturbation strategy for the probability of success in order to evaluate local influence of the measurements (observations of the binary response), but not of the subjects. With this perturbation strategy, we are able to detect influential observations which can cause disproportionate effects in the estimates and/or in the predictive performance of the MELRM considering the value of its binary response. This allows us to avoid misleading ML estimates and to improve the predictive performance of the model. For more details about the perturbation of the probability of success, see Sect. 2.5.

The main objective of this paper is to derive a methodology of local influence in the MELRM to detect how the response affects the estimates and predictive performance of the model by using an appropriate perturbation of the associated probability of success.

This objective is conducted to investigate the local influence of binary responses on the ML estimates of the MELRM parameters and in the predictive performance of this model by using the Q-displacement function. The diagnostic methodology is evaluated by Monte Carlo (MC) methods. Furthermore, as illustration, we use two real data sets related to seeds (unbalanced) and salamanders (balanced). The numerical results of this study are obtained with the aid of routines implemented by the authors in the R software, which are available under request; see www.R-project.org and R Core Team (2016).

The remainder of this article is organized as follows. In Sect. 2, we present the MELRM and derive a local influence analysis for such a model selecting an appropriate perturbation for the probability of success associated with the corresponding binary response. In Sect. 3, the results of the MC simulation study are presented to evaluate the performance of the methodology derived in Sect. 2. Illustrations with two real-world biological data sets of the derived methodology are analyzed in Sect. 4. Finally, conclusions and proposals for future research are discussed in Sect. 5.

2 Local influence in the mixed effects logistic regression model

2.1 The mixed effects logistic regression model

Consider the binary responses Y_{ij} with Bernoulli distribution of parameter p_{ij}, for $j = 1, \ldots, n_i$ and $i = 1, \ldots, I$. Consider also that $\boldsymbol{b}_i$ is a random vector of dimension p_2 with normal distribution of mean $\mathbf{0}_{p_2\times 1}$ and variance–covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\gamma})$, where $\mathbf{0}_{p_2\times 1}$ is a vector of zeros of dimension p_2 and $\boldsymbol{\gamma}$ is a vector of unknown variance and covariance components of dimension p_3, with $p_3 \leq p_2(p_2+1)/2$. It is assumed that the conditional distribution of Y_{ij} given $\boldsymbol{b}_i$ belongs to the exponential family with probability function given by

$$p_{Y_{ij}|\boldsymbol{b}_i}(y_{ij}) = \exp\left(y_{ij}\log\left(\frac{p_{ij}}{1-p_{ij}}\right) - \log\left(\frac{1}{1-p_{ij}}\right)\right), \tag{1}$$

mean expressed as $\mathrm{E}(Y_{ij}|\boldsymbol{b}_i) = p_{ij}$ and variance defined as $\mathrm{V}(Y_{ij}|\boldsymbol{b}_i) = p_{ij}(1-p_{ij})$. Thus, the MELRM is defined by (1) and by the systematic component

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\top}\boldsymbol{b}_i,$$

where $\log(p_{ij}/(1-p_{ij}))$ is the logit link function and $\boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\top}\boldsymbol{b}_i$ is the linear predictor, with $\boldsymbol{x}_{ij} = (x_{ij1}, \ldots, x_{ijp_1})^{\top}$ and $\boldsymbol{z}_{ij} = (z_{ij1}, \ldots, z_{ijp_2})^{\top}$ being vectors of dimension p_1 and p_2, respectively, containing values of the corresponding covariates. Here, $\boldsymbol{\beta}$ is the vector of unknown regression coefficients to be estimated of dimension p_1. Let $\boldsymbol{\psi} = (\boldsymbol{\beta}^{\top}, \boldsymbol{\gamma}^{\top})^{\top}$ be a unknown parameter vector of dimension $(p_1 + p_3)$ and $\boldsymbol{y}_{\mathrm{o}} = \{y_{ij}\colon j = 1, \ldots, n_i, i = 1, \ldots, I\}$ be the observed data set. Then, the observed-data log-likelihood function for $\boldsymbol{\psi}$ is defined as

$$\ell(\boldsymbol{\psi};\boldsymbol{y}_{\mathrm{o}}) = \sum_{i=1}^{I} \log\left(\int_{\mathbb{R}^{p_2}} \prod_{j=1}^{n_i} \exp\left(y_{ij}\log\left(\frac{p_{ij}}{1-p_{ij}}\right) - \log\left(\frac{1}{1-p_{ij}}\right)\right) \times \frac{1}{(2\pi)^{p_2/2}} \det(\boldsymbol{\Sigma})^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{b}_i^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{b}_i\right) \mathrm{d}\boldsymbol{b}_i\right). \tag{2}$$

For more details about the MELRM, see the books by Agresti (2003), Jiang (2007) and Hosmer et al. (2013).

Note that the function defined in (2) contains a multiple integral which has no analytical solution. Hence, a procedure that approximates this integral is required to solve it. In this work, the ML estimation of $\boldsymbol{\psi}$ is conducted using the function `glmer` of an `R` package named `lme4`, which uses AGHQ, but particularly we use 25 quadrature points. This procedure is only available when the model includes one random intercept. For two or more random effects, we use AL.

Deriving the local influence technique from (2) is not an easy task, because the integral has a high dimension and no analytical solution. Zhu and Lee (2003) treated the random effects as a missing (unobserved) data set, $\boldsymbol{y}_{\mathrm{u}} = \{\boldsymbol{b}_i : i = 1,\ldots,I\}$, and defined $\boldsymbol{y}_{\mathrm{c}} = (\boldsymbol{y}_{\mathrm{o}}, \boldsymbol{y}_{\mathrm{u}})$ as the complete data set, where $\boldsymbol{y}_{\mathrm{o}}$ is the observed data set, as mentioned. In general, if the size of the involved integral is large, the standard numerical integration methods may be intractable. In addition, in some cases, a large number of quadrature points can be required and then the AGHQ may slowly converge; see Lesaffre and Spiessens (2001). Thus, the ML estimates might be calculated alternatively by MC methods, such as the expectation–maximization (EM) algorithm; see Dempster et al. (1977) and McCulloch (1997). However, the MC methods are computationally intensive as well and may require many iterations with slow convergence; see Molenberghs and Verbeke (2005). Because of the ease of implementation and since we do not detect convergence problems, quadrature methods are used in this work. Then, the complete-data log-likelihood function for $\boldsymbol{\psi}$ of the MELRM is expressed as

$$\ell(\boldsymbol{\psi};\boldsymbol{y}_{\mathrm{c}}) = \sum_{i=1}^{I}\left\{\sum_{j=1}^{n_i}\left(y_{ij}\log\left(\frac{p_{ij}}{1-p_{ij}}\right) - \log\left(\frac{1}{1-p_{ij}}\right)\right) - \frac{1}{2}\boldsymbol{b}_i^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{b}_i - \frac{1}{2}\log(\det(\boldsymbol{\Sigma}))\right\}, \tag{3}$$

which is a relatively simple expression for local influence analysis; see Zhu and Lee (2003).

2.2 The local influence technique

Let $\boldsymbol{Y}_{\mathrm{c}} = (\boldsymbol{Y}_{\mathrm{o}}, \boldsymbol{Y}_{\mathrm{u}})$ be the random vector associated with the complete data set $\boldsymbol{y}_{\mathrm{c}} = (\boldsymbol{y}_{\mathrm{o}}, \boldsymbol{y}_{\mathrm{u}})$. Consider a perturbation vector $\boldsymbol{\omega} \in \Omega \subset \mathbb{R}^q$, with $q = n = \sum_{i=1}^{I} n_i$ in

our case, such that $\ell(\boldsymbol{\psi}, \boldsymbol{\omega}; \boldsymbol{y}_{\mathrm{c}})$ is the complete-data log-likelihood function of the perturbed model. It is assumed that there is a non-perturbation vector $\boldsymbol{\omega}_0 \in \Omega \subset \mathbb{R}^n$, such that $\ell(\boldsymbol{\psi}, \boldsymbol{\omega}_0; \boldsymbol{y}_{\mathrm{c}}) = \ell(\boldsymbol{\psi}; \boldsymbol{y}_{\mathrm{c}})$, for all $\boldsymbol{\psi}$. Let $\widehat{\boldsymbol{\psi}}(\boldsymbol{\omega})$ be the ML estimate of $\boldsymbol{\psi}$ for the perturbed model that maximizes $Q(\boldsymbol{\psi}, \boldsymbol{\omega})|_{\boldsymbol{\psi}=\widehat{\boldsymbol{\psi}}} = \mathrm{E}[\ell(\boldsymbol{\psi}, \boldsymbol{\omega}; \boldsymbol{Y}_{\mathrm{c}}|\boldsymbol{Y}_{\mathrm{o}} = \boldsymbol{y}_{\mathrm{o}})]|_{\boldsymbol{\psi}=\widehat{\boldsymbol{\psi}}}$, where $\widehat{\boldsymbol{\psi}}$ is the ML estimate of $\boldsymbol{\psi}$ and the expectation is calculated with respect to the conditional distribution of $\boldsymbol{Y}_{\mathrm{u}} = \boldsymbol{b}_i$ given $\boldsymbol{Y}_{\mathrm{o}} = \boldsymbol{y}_{\mathrm{o}}$ (note that, for example, the notation "$|_{\boldsymbol{A}=\boldsymbol{a}}$" means that the corresponding function is evaluated at $\boldsymbol{A} = \boldsymbol{a}$). To assess the influence of $\boldsymbol{\omega} \in \Omega \subset \mathbb{R}^n$, Zhu and Lee (2003) used the Q-displacement function given by

$$f_Q(\boldsymbol{\omega}) = 2\left(Q\left(\widehat{\boldsymbol{\psi}}\right) - Q\left(\widehat{\boldsymbol{\psi}}(\boldsymbol{\omega})\right)\right), \tag{4}$$

where $Q(\widehat{\boldsymbol{\psi}}) = Q(\boldsymbol{\psi}, \boldsymbol{\omega})|_{\boldsymbol{\psi}=\widehat{\boldsymbol{\psi}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0}$ and $Q(\widehat{\boldsymbol{\psi}}(\boldsymbol{\omega})) = Q(\boldsymbol{\psi}, \boldsymbol{\omega})|_{\boldsymbol{\psi}=\widehat{\boldsymbol{\psi}}(\boldsymbol{\omega}_0)}$. Function (4) is considered a metric of the difference between $\widehat{\boldsymbol{\psi}}$ and $\widehat{\boldsymbol{\psi}}(\boldsymbol{\omega})$ with respect to the objective function $Q(\boldsymbol{\psi})|_{\boldsymbol{\psi}=\widehat{\boldsymbol{\psi}}}$, which is greater than or equal to zero and achieves a global minimum at $\boldsymbol{\omega}_0$. Then, similarly to Zhu and Lee (2001), the influence graph of $f_Q(\boldsymbol{\omega})$ is defined as $\alpha(\boldsymbol{\omega}) = (\boldsymbol{\omega}^\top, f_Q(\boldsymbol{\omega}))^\top$. Note that the normal curvature $C_{f_Q,\boldsymbol{h}}$ of $\alpha(\boldsymbol{\omega})$ in $\boldsymbol{\omega}_0$, in the direction of a unit vector $\boldsymbol{h} \in \mathbb{R}^n$, is used to summarize the local behavior of $f_Q(\boldsymbol{\omega})$ and expressed as

$$C_{f_Q,\boldsymbol{h}} = -2\boldsymbol{h}^\top \ddot{\boldsymbol{Q}}_{\boldsymbol{\omega}_0}\boldsymbol{h} = 2\boldsymbol{h}^\top \boldsymbol{\Delta}_{\boldsymbol{\omega}_0}^\top \left(-\ddot{\boldsymbol{Q}}_{\boldsymbol{\psi}}\left(\widehat{\boldsymbol{\psi}}\right)\right)^{-1} \boldsymbol{\Delta}_{\boldsymbol{\omega}_0}\boldsymbol{h}, \tag{5}$$

where

$$\ddot{\boldsymbol{Q}}_{\boldsymbol{\omega}_0} = \frac{\partial^2 Q\left(\widehat{\boldsymbol{\psi}}(\boldsymbol{\omega})\right)}{\partial\boldsymbol{\omega}\partial\boldsymbol{\omega}^\top}$$

is a semi-positive definite matrix of dimension $n \times n$,

$$-\ddot{\boldsymbol{Q}}_{\boldsymbol{\psi}}\left(\widehat{\boldsymbol{\psi}}\right) = -\frac{\partial^2 Q\left(\widehat{\boldsymbol{\psi}}\right)}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^\top} \tag{6}$$

is a semi-positive definite matrix of dimension $(p_1 + p_3) \times (p_1 + p_3)$ and

$$\boldsymbol{\Delta}_{\boldsymbol{\omega}_0} = \left.\frac{\partial^2 Q(\boldsymbol{\psi}, \boldsymbol{\omega})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\omega}^\top}\right|_{\boldsymbol{\psi}=\widehat{\boldsymbol{\psi}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} \tag{7}$$

is the perturbation matrix of dimension $(p_1 + p_3) \times n$.

Following Zhu and Lee (2001), the normal curvature given in (5) is invariant under reparametrization of $\boldsymbol{\psi}$ and it can assume any value. Thus, based on Poon and Poon (1999) and the expression given in (5), Zhu and Lee (2001) proposed the conformal normal curvature of $\alpha(\boldsymbol{\omega})$ in $\boldsymbol{\omega}_0$, in the direction of a unit vector $\boldsymbol{h} \in \mathbb{R}^n$, by means of

$$B_{f_Q,\boldsymbol{h}} = \frac{C_{f_Q,\boldsymbol{h}}}{\mathrm{tr}\left(-2\ddot{\boldsymbol{Q}}_{\boldsymbol{\omega}_0}\right)}, \tag{8}$$

which, in addition to be invariant under reparametrization of $\boldsymbol{\psi}$, is invariant under conformal reparametrization of $\boldsymbol{\omega}$, and it can assume any value in the closed interval [0, 1].

Note that $-2\ddot{\boldsymbol{Q}}_{\boldsymbol{\omega}_0} = 2\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}^{\top}(-\ddot{\boldsymbol{Q}}_{\boldsymbol{\psi}}(\widehat{\boldsymbol{\psi}}))^{-1}\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$ may be expressed in terms of its spectral decomposition, that is, by

$$-2\ddot{\boldsymbol{Q}}_{\boldsymbol{\omega}_0} = \sum_{m=1}^{M} \lambda_m \boldsymbol{e}_m^{\top}\boldsymbol{e}_m, \tag{9}$$

where $(\lambda_1, \boldsymbol{e}_1), \ldots, (\lambda_M, \boldsymbol{e}_M)$ are pairs of eigenvalues and eigenvectors such that $\lambda_1 \geq \cdots \geq \lambda_M > \lambda_{M+1} = \cdots = \lambda_n = 0$ and $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_M)$ is an orthonormal basis of $\mathbb{R}^M$, with $M = p_1 + p_3$. Then, the normal curvature is defined in the direction of the observation i by

$$C_{f_Q,\boldsymbol{h}_i} = \sum_{m=1}^{M} \lambda_m \mathrm{e}_{m_i}^2, \tag{10}$$

with e_{m_i} being the ith component of $\boldsymbol{e}_m$ and $\boldsymbol{h}_i$ being a vector of dimension n with the ith component equal to one and the remaining values equal to zero. According to Zhu and Lee (2001), the observation i is influential if

$$B_{f_Q,\boldsymbol{h}_i} > \overline{B} + 2\mathrm{SD}(B),$$

where $\overline{B} = \sum_{i=1}^{n} B_{f_Q,\boldsymbol{h}_i}/n$ and $\mathrm{SD}(B)$ is the standard deviation of $B_{f_Q,\boldsymbol{h}_1}, \ldots, B_{f_Q,\boldsymbol{h}_n}$, with $B_{f_Q,\boldsymbol{h}_i}$ denoting the conformal normal curvature in the direction of the observation i, which is given from (8) and $B_{f_Q,\boldsymbol{h}_i} = C_{f_Q,\boldsymbol{h}_i}/\mathrm{tr}(-2\ddot{\boldsymbol{Q}}_{\boldsymbol{\omega}_0})$, for $i = 1, \ldots, n$, with $-2\ddot{\boldsymbol{Q}}$ and $C_{f_Q,\boldsymbol{h}_i}$ defined in (9) and (10), respectively.

2.3 Approximations of $-\ddot{\boldsymbol{Q}}_{\boldsymbol{\psi}}(\widehat{\boldsymbol{\psi}})$ and $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$

Since the conditions of regularity allow the exchange between integration and differentiation, the matrices given by (6) and (7) can be expressed as

$$-\ddot{\boldsymbol{Q}}_{\boldsymbol{\psi}}(\widehat{\boldsymbol{\psi}}) = \mathrm{E}\left(-\frac{\partial^2 \ell(\boldsymbol{\psi}; \boldsymbol{Y}_{\mathrm{c}}|\boldsymbol{Y}_{\mathrm{o}} = \boldsymbol{y}_{\mathrm{o}})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}^{\top}}\right)\Big|_{\boldsymbol{\psi}=\widehat{\boldsymbol{\psi}}} \tag{11}$$

and

$$\boldsymbol{\Delta}_{\boldsymbol{\omega}_0} = \mathrm{E}\left(\frac{\partial^2 \ell(\boldsymbol{\psi}, \boldsymbol{\omega}; \boldsymbol{Y}_{\mathrm{c}}|\boldsymbol{Y}_{\mathrm{o}} = \boldsymbol{y}_{\mathrm{o}})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\omega}^{\top}}\right)\Big|_{\boldsymbol{\psi}=\widehat{\boldsymbol{\psi}},\boldsymbol{\omega}=\boldsymbol{\omega}_0}, \tag{12}$$

respectively. However, the conditional expectation presented in the blocks of (11) and (12) cannot be calculated in closed form. Consequently, Zhu and Lee (2003) used the classic MC integration method to solve it of the following way. Let $\{\boldsymbol{y}_{\mathrm{u}}^{(s)}: s = 1, \ldots, S\}$

be data generated from the conditional distribution of $\boldsymbol{Y}_{\mathrm{u}} = \boldsymbol{b}_i$ given $\boldsymbol{Y}_{\mathrm{o}} = \boldsymbol{y}_{\mathrm{o}}$. Then,

$$-\ddot{\boldsymbol{Q}}_{\boldsymbol{\psi}}\left(\widehat{\boldsymbol{\psi}}\right) \approx \frac{1}{S - M_0} \sum_{s=M_0+1}^{S} -\frac{\partial^2 \ell\left(\boldsymbol{\psi}; \boldsymbol{y}_{\mathrm{o}}, \boldsymbol{y}_{\mathrm{u}}^{(s)}\right)}{\partial \boldsymbol{\psi}\, \partial \boldsymbol{\psi}^{\top}}\Bigg|_{\boldsymbol{\psi}=\widehat{\boldsymbol{\psi}}}$$

and

$$\boldsymbol{\Delta}_{\boldsymbol{\omega}_0} \approx \frac{1}{S - M_0} \sum_{s=M_0+1}^{S} \frac{\partial^2 \ell\left(\boldsymbol{\psi}, \boldsymbol{\omega}; \boldsymbol{y}_{\mathrm{o}}, \boldsymbol{y}_{\mathrm{u}}^{(s)}\right)}{\partial \boldsymbol{\psi}\, \partial \boldsymbol{\omega}^{\top}}\Bigg|_{\boldsymbol{\psi}=\widehat{\boldsymbol{\psi}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0},$$

where M_0 corresponds to the number of observations discarded to avoid an effect of dependence produced by the Metropolis–Hastings (MH) algorithm in the first iterations. Note that a usual value considered in practice is $M_0 = 1000$ observations. The conditional distribution of $\boldsymbol{Y}_{\mathrm{u}} = \boldsymbol{b}_i$ given $\boldsymbol{Y}_{\mathrm{o}} = \boldsymbol{y}_{\mathrm{o}}$ has a probability function which is proportional to

$$\exp\left(-\frac{1}{2}\boldsymbol{b}_i^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{b}_i + \sum_{j=1}^{n_i} y_{ij}\left(\boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\top}\boldsymbol{b}_i\right) - \log\left(1 + \exp\left(\boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\top}\boldsymbol{b}_i\right)\right)\right),$$

so that generating the observations $\{\boldsymbol{y}_{\mathrm{u}}^{(s)} : s = 1, \ldots, S\}$ is not trivial. Then, Zhu and Lee (2003) used the MH algorithm to generate observations as described next. The MH algorithm is initialized from an arbitrary value $\boldsymbol{b}_i^{(0)}$, which does not affect mostly the results due to that this algorithm works with a large number of observations. In addition, since M_0 observations are discarded, this also does the initial value $\boldsymbol{b}_i^{(0)}$ not be relevant; see details in Robert and Casella (1999). In the rth iteration of the algorithm, the following steps must be considered:

1. Given the current value of $\boldsymbol{b}_i^{(r-1)}$, generate a new candidate as $\boldsymbol{b}_i \sim \mathrm{N}_{p_2}(\boldsymbol{b}_i^{(r-1)}, \Gamma_i(\boldsymbol{0}_{p_2\times 1}))$, where, following the same notation as in Zhu and Lee (2003) and Xu et al. (2006),

$$\Gamma_i\left(\boldsymbol{0}_{p_2\times 1}\right) = \Gamma\left(\boldsymbol{b}_i\right)|_{\boldsymbol{b}_i=\boldsymbol{0}_{p_2\times 1}} = \left(\boldsymbol{\Sigma}^{-1} + \sum_{j=1}^{n_i} \frac{\exp\left(\boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\top}\boldsymbol{b}_i\right)}{\left(1 + \exp\left(\boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\top}\boldsymbol{b}_i\right)\right)^2} z_{ij}\boldsymbol{z}_{ij}^{\top}\right)^{-1}\Bigg|_{\boldsymbol{b}_i=\boldsymbol{0}_{p_2\times 1}}.$$

2. Obtain u from $U \sim \mathrm{U}(0, 1)$, that is, from a uniform distribution in $[0, 1]$. If $u \leqslant \alpha(\boldsymbol{b}_i^{(r-1)}, \boldsymbol{b}_i)$, then $\boldsymbol{b}_i^{(r)} = \boldsymbol{b}_i$, otherwise, consider $\boldsymbol{b}_i^{(r)} = \boldsymbol{b}_i^{(r-1)}$, where

$$\alpha\left(\boldsymbol{b}_i^{(r-1)}, \boldsymbol{b}_i\right) = \min\left\{\frac{p_{\boldsymbol{b}_i|\boldsymbol{Y}_{\mathrm{o}}=\boldsymbol{y}_{\mathrm{o}}}\left(\boldsymbol{b}_i, \boldsymbol{\psi}\right)}{p_{\boldsymbol{b}_i|\boldsymbol{Y}_{\mathrm{o}}=\boldsymbol{y}_{\mathrm{o}}}\left(\boldsymbol{b}_i^{(r-1)}, \boldsymbol{\psi}\right)}, 1\right\}$$

   is the probability of accepting a new candidate.
3. Repeat steps (1) and (2) for $r + 1$.

2.4 Second-order derivatives of $-\ddot{Q}_{\psi}(\widehat{\psi})$

To calculate the conformal normal curvature, it is necessary to obtain the derivatives included in $-\ddot{Q}_{\psi}(\widehat{\psi})$ and $\boldsymbol{\Delta}_{\omega_0}$. However, we present only the derivatives related to $-\ddot{Q}_{\psi}(\widehat{\psi})$, since they do not depend on the proposed perturbation strategy. Thus, from (3), the derivatives involved in $-\ddot{Q}_{\psi}(\widehat{\psi})$ are given by

$$\frac{\partial^2 \ell\left(\boldsymbol{\psi}; \boldsymbol{y}_{\mathrm{c}}\right)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^{\top}} = -\sum_{i=1}^{I} \sum_{j=1}^{n_i} \frac{\exp\left(\boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\top}\boldsymbol{b}_i\right)\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\top}}{\left(1 + \exp\left(\boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\top}\boldsymbol{b}_i\right)\right)^2},$$

$$\frac{\partial^2 \ell\left(\boldsymbol{\psi}; \boldsymbol{y}_{\mathrm{c}}\right)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}^{\top}} = \mathbf{0}_{p_1 \times p_3}, \quad \frac{\partial^2 \ell\left(\boldsymbol{\psi}; \boldsymbol{y}_{\mathrm{c}}\right)}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}^{\top}} = \mathbf{0}_{p_3 \times p_1},$$

$$\frac{\partial^2 \ell\left(\boldsymbol{\psi}; \boldsymbol{y}_{\mathrm{c}}\right)}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^{\top}} = \frac{I}{2}\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) - \left(\boldsymbol{\Sigma}^{-1} \sum_{i=1}^{I} \boldsymbol{b}_i \boldsymbol{b}_i^{\top} \boldsymbol{\Sigma}^{-1}\right) \otimes \boldsymbol{\Sigma}^{-1},$$

where $\otimes$ denotes the Kronecker product of matrices; see Caro-Lopera et al. (2012).

2.5 Appropriate perturbation for the probability of success

As mentioned, due to the binary nature in the response of the MELRM, standard local influence methods do not apply for perturbing this response. Then, to evaluate local influence of the measurements in the ML estimates and/or in the predictive performance of the MELRM, we perturb the probability of success associated with this binary response. We use the local influence technique based on the Q-displacement function defined in Sect. 2.2 and an appropriate perturbation strategy as detailed below. The methodology for local influence analysis derived here is based on the strategy of an appropriate multiplicative perturbation of the probability of success (AMPPS). This methodology allows us to detect influential observations evaluating how the binary response can cause disproportionate effects in the estimates and/or in the predictive performance of the MELRM, avoiding misleading about them.

Note that the AMPPS strategy allows us to detect the influence of each measurement for each subject, that is, the influence of measurements is evaluated, but not of the subjects. As we are using a model for repeated measurements, a measurement done to a subject may be detected as influential, but another measurement of the same subject could be not influential. Thus, with the AMPPS strategy, we are able to judge whether an observation is influential in the results or not. This is particularly of interest in the MELRM to study the effect of influential observations in the ML estimates of the model parameters and in its predictive performance by means of the Acc, Sen and Spe indicators. However, if we evaluate the influence of the subjects, we are unable to know whether a measurement (binary response) can cause or not disproportionate effects in the ML estimates and/or in the predictive performance of the MELRM. We could evaluate whether a subject is influential or not deleting this case from the analysis; see Xu et al. (2006).

Similarly to Nyangoma et al. (2006), a form to evaluate the perturbation of a binary response in the MELRM is through a strategy of multiplicative perturbation of the probability of success (MPPS) given by

$$p_{ij}(\omega_{ij}) = p_{ij}\omega_{ij}; \quad \omega_{ij} \in (0,\ 1], \quad j = 1, \ldots, n_i, \quad i = 1, \ldots, I. \tag{13}$$

Thus, according to Chen et al. (2010), the joint probability function of $\boldsymbol{Y}_{\mathrm{c}}$ under (13) is given by

$$p_{\boldsymbol{Y}_{\mathrm{c}}}\left(\boldsymbol{y}_{\mathrm{c}}; \boldsymbol{\psi}, \boldsymbol{\omega}\right) = \prod_{i=1}^{I}\prod_{j=1}^{n_i} \exp\left(\left(y_{ij} \log\left(\frac{p_{ij}\left(\omega_{ij}\right)}{1 - p_{ij}\left(\omega_{ij}\right)}\right) - \log\left(\frac{1}{1 - p_{ij}\left(\omega_{ij}\right)}\right)\right) - \frac{1}{2}\boldsymbol{b}_i^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{b}_i - \frac{1}{2}\log\left(\det\left(\boldsymbol{\Sigma}\right)\right)\right),$$

whose non-perturbation vector is $\boldsymbol{\omega}_0 = \mathbf{1}_{n\times 1}$, where $\mathbf{1}_{n\times 1}$ is a vector of ones of dimension n, with $n = \sum_{i=1}^{I} n_i$, such that $p_{\boldsymbol{Y}_{\mathrm{c}}}(\boldsymbol{y}_{\mathrm{c}}; \boldsymbol{\psi}, \boldsymbol{\omega}_0) = p_{\boldsymbol{Y}_{\mathrm{c}}}(\boldsymbol{y}_{\mathrm{c}}; \boldsymbol{\psi})$, for all $\boldsymbol{\psi}$. Consider the perturbed model given by $P \equiv \{p_{\boldsymbol{Y}_{\mathrm{c}}}(\boldsymbol{y}_{\mathrm{c}}; \boldsymbol{\psi}, \boldsymbol{\omega})\colon \boldsymbol{\omega} \in \Omega \subset \mathbb{R}^n\}$. Then, the Fisher expected information matrix of dimension $n \times n$ with respect to the perturbation vector $\boldsymbol{\omega}$ under P is expressed as

$$\boldsymbol{G}(\boldsymbol{\omega}) = (g_{ij}(\boldsymbol{\omega})), \tag{14}$$

where

$$g_{ij}\left(\boldsymbol{\omega}\right) = \mathrm{E}\left(\frac{\partial \log\left(p_{\boldsymbol{Y}_{\mathrm{c}}}\left(\boldsymbol{y}_{\mathrm{c}}; \boldsymbol{\psi}, \boldsymbol{\omega}\right)\right)}{\partial \omega_i} \frac{\partial \log\left(p_{\boldsymbol{Y}_{\mathrm{c}}}\left(\boldsymbol{y}_{\mathrm{c}}; \boldsymbol{\psi}, \boldsymbol{\omega}\right)\right)}{\partial \omega_j}\right)$$

and the expectation is calculated with respect to $p_{\boldsymbol{Y}_{\mathrm{c}}}(\boldsymbol{y}_{\mathrm{c}}; \boldsymbol{\psi}, \boldsymbol{\omega})$. The diagonal elements of the matrix $\boldsymbol{G}(\boldsymbol{\omega})$ defined in (14) are the variances of the score vector with respect to the components of $\boldsymbol{\omega}$, which indicate the amount of perturbation introduced by the corresponding components of $\boldsymbol{\omega}$. The off-diagonal elements of $\boldsymbol{G}(\boldsymbol{\omega})$ are the covariances of the score vector with respect to the components of $\boldsymbol{\omega}$, which represent the association between the different components of $\boldsymbol{\omega}$. Note that a perturbation is appropriate if it satisfies the following conditions: (i) $\boldsymbol{G}(\boldsymbol{\omega})$ is full rank in a small neighborhood of $\boldsymbol{\omega}_0$, to avoid redundant components of $\boldsymbol{\omega}$; (ii) the off-diagonal components are as small as possible, to avoid a strong association between the components of $\boldsymbol{\omega}$, and consequently, perturbations with strong ambiguous effects; and (iii) the differences between the components of the diagonal are as small as possible, so that the perturbations introduced by the components of $\boldsymbol{\omega}$ are uniform. Based on (i)–(iii), Chen et al. (2010) defined an appropriate perturbation satisfying that $\boldsymbol{G}(\boldsymbol{\omega}_0) = c\boldsymbol{I}_n$, where $c > 0$. In applications, although $\boldsymbol{G}(\boldsymbol{\omega}_0) \neq c\boldsymbol{I}_n$, we can always choose a new perturbation vector $\widetilde{\boldsymbol{\omega}}$ defined by

$$\widetilde{\boldsymbol{\omega}} = \boldsymbol{\omega}_0 + \boldsymbol{G}(\boldsymbol{\omega}_0)^{1/2}(\boldsymbol{\omega} - \boldsymbol{\omega}_0), \tag{15}$$

that is, we consider the perturbed model $\widetilde{P} = \{p_{\boldsymbol{Y}_c}(\boldsymbol{y}_c; \boldsymbol{\psi}, \boldsymbol{\omega}(\widetilde{\boldsymbol{\omega}}))\colon \widetilde{\boldsymbol{\omega}} \in \widetilde{\Omega} \subset \mathbb{R}^n\}$, where $\boldsymbol{\omega}(\widetilde{\boldsymbol{\omega}}) = \boldsymbol{\omega}_0 + \boldsymbol{G}(\boldsymbol{\omega}_0)^{-1/2}(\widetilde{\boldsymbol{\omega}} - \boldsymbol{\omega}_0)$ and $\widetilde{\Omega} = \{\boldsymbol{\omega}_0 + \boldsymbol{G}(\boldsymbol{\omega}_0)^{1/2}(\boldsymbol{\omega} - \boldsymbol{\omega}_0)\colon \boldsymbol{\omega} \in$

$\Omega \subset \mathbb{R}^n\}$. Under $\widetilde{P}$, we have $\boldsymbol{G}(\boldsymbol{\omega}_0) = c\boldsymbol{I}_n$. In our case, for $i = j$, the derivatives related to $\boldsymbol{G}(\boldsymbol{\omega}_0)$ are given by

$$\frac{\partial \log\left(p_{\boldsymbol{Y}_c}\left(\boldsymbol{y}_c;\boldsymbol{\psi},\boldsymbol{\omega}\right)\right)}{\partial \omega_i}\frac{\partial \log\left(p_{\boldsymbol{Y}_c}\left(\boldsymbol{y}_c;\boldsymbol{\psi},\boldsymbol{\omega}\right)\right)}{\partial \omega_j}\Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = y_{ij}\left(1-\exp\left(2\left(\boldsymbol{x}_{ij}^\top\boldsymbol{\beta}+\boldsymbol{z}_{ij}^\top\boldsymbol{b}_i\right)\right)\right)+\exp\left(2\left(\boldsymbol{x}_{ij}^\top\boldsymbol{\beta}+\boldsymbol{z}_{ij}^\top\boldsymbol{b}_i\right)\right),$$

whereas for $i \neq j$,

$$\frac{\partial \log\left(p_{\boldsymbol{Y}_c}\left(\boldsymbol{y}_c;\boldsymbol{\psi},\boldsymbol{\omega}\right)\right)}{\partial \omega_i}\frac{\partial \log\left(p_{\boldsymbol{Y}_c}\left(\boldsymbol{y}_c;\boldsymbol{\psi},\boldsymbol{\omega}\right)\right)}{\partial \omega_j}\Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = 0.$$

Then, based on (15), the AMPPS strategy is defined as

$$p_{ij}(\omega_{ij}(\widetilde{\omega}_{ij})) = p_{ij}\left(\omega_{ij_0} + g_{ij}(\boldsymbol{\omega}_0)^{-1/2}(\widetilde{\omega}_{ij} - \omega_{ij_0})\right). \tag{16}$$

The expectation involved in the blocks of $\boldsymbol{G}(\boldsymbol{\omega}_0)$ cannot be calculated in closed form. Consequently, Chen et al. (2010) used the classic MC integration method as follows. Generate a sample $\{\boldsymbol{b}_i^{(t)}; t = 1,\ldots,T\}$ from a normal distribution of mean vector $\boldsymbol{0}_{p_2\times 1}$ and covariance–variance matrix $\boldsymbol{\Sigma}$, approximating the elements of $\boldsymbol{G}(\boldsymbol{\omega}_0)$ by

$$g_{ij}\left(\boldsymbol{\omega}_0\right) \approx \frac{1}{T}\sum_{t=1}^{\mathrm{T}}\frac{\partial \log\left(p_{\boldsymbol{Y}_c}\left(\boldsymbol{y}_o,\boldsymbol{b}_i^{(t)};\boldsymbol{\psi},\boldsymbol{\omega}\right)\right)}{\partial \omega_i}\frac{\partial \log\left(p_{\boldsymbol{Y}_c}\left(\boldsymbol{y}_o,\boldsymbol{b}_i^{(t)};\boldsymbol{\psi},\boldsymbol{\omega}\right)\right)}{\partial \omega_j}\Big|_{\boldsymbol{\psi}=\widehat{\boldsymbol{\psi}},\boldsymbol{\omega}=\boldsymbol{\omega}_0}.$$

Note that, in the MPPS strategy presented in (13), we do not impose restrictions on the perturbation vector, whereas in the AMPPS strategy given in (16), the perturbation vector depends on $\boldsymbol{G}$, guaranteeing an appropriate perturbation. As mentioned, identification of influential cases is a very important step in data analysis. However, arbitrarily perturbing the model or data can lead to unreliable conclusions with respect to influence diagnostics. For example, considering unbalanced cluster (subjects) sizes in the perturbation scheme related to case-weights among clusters may lead to the inaccuracy identification of influential groups among all groups; see Chen et al. (2010).

2.6 Second-order derivatives of $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$

Under (13), the complete-data log-likelihood function of the perturbed model is given by

$$\ell\left(\boldsymbol{\psi},\boldsymbol{\omega};\boldsymbol{y}_c\right) = \sum_{i=1}^{I}\left\{\sum_{j=1}^{n_i}\left\{y_{ij}\log\left(\frac{p_{ij}\left(\omega_{ij}\right)}{1-p_{ij}\left(\omega_{ij}\right)}\right)-\log\left(\frac{1}{1-p_{ij}\left(\omega_{ij}\right)}\right)\right\}\right.$$
$$\left.-\frac{1}{2}\boldsymbol{b}_i^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{b}_i-\frac{1}{2}\log\left(\det\left(\boldsymbol{\Sigma}\right)\right)\right\}.$$

The non-perturbation vector is $\boldsymbol{\omega}_0 = \mathbf{1}_{n\times 1}$. Then, the derivatives involved in $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$ are

$$\left.\frac{\partial^2 \ell\left(\boldsymbol{\psi}, \boldsymbol{\omega}; \boldsymbol{y}_{\mathrm{c}}\right)}{\partial \boldsymbol{\beta} \partial \omega_{ij}}\right|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \left(y_{ij}-1\right) \exp\left(\boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\top}\boldsymbol{b}_i\right)\boldsymbol{x}_{ij}, \left.\frac{\partial^2 \ell\left(\boldsymbol{\psi}, \boldsymbol{\omega}; \boldsymbol{y}_{\mathrm{c}}\right)}{\partial \boldsymbol{\gamma} \partial \omega_{ij}}\right|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \mathbf{0}_{p_3\times 1}. \tag{17}$$

Under (16), the complete-data log-likelihood function of the perturbed model is given by

$$\ell\left(\boldsymbol{\psi}, \widetilde{\boldsymbol{\omega}}; \boldsymbol{y}_{\mathrm{c}}\right) = \sum_{i=1}^{I}\left\{\sum_{j=1}^{n_i}\left\{y \log\left(\frac{p_{ij}\left(\omega_{ij}\left(\widetilde{\omega}_{ij}\right)\right)}{1-p_{ij}\left(\omega_{ij}\left(\widetilde{\omega}_{ij}\right)\right)}\right) - \log\left(\frac{1}{1-p_{ij}\left(\omega_{ij}\left(\widetilde{\omega}_{ij}\right)\right)}\right)\right\} - \frac{1}{2}\boldsymbol{b}_i^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{b}_i - \frac{1}{2}\log\left(\det\left(\boldsymbol{\Sigma}\right)\right)\right\}.$$

The non-perturbation vector is $\widetilde{\boldsymbol{\omega}}_0 = \boldsymbol{\omega}_0 = \mathbf{1}_{n\times 1}$. Then, the derivatives involved in $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$ are

$$\begin{aligned}\left.\frac{\partial^2 \ell\left(\boldsymbol{\psi}, \widetilde{\boldsymbol{\omega}}; \boldsymbol{y}_{\mathrm{c}}\right)}{\partial \boldsymbol{\beta} \partial \widetilde{\omega}_{ij}}\right|_{\widetilde{\boldsymbol{\omega}}=\boldsymbol{\omega}_0} &= g_{ij}\left(\boldsymbol{\omega}_0\right)^{-1/2}\left(y_{ij}-1\right)\exp\left(\boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\top}\boldsymbol{b}_i\right)\boldsymbol{x}_{ij},\\ \left.\frac{\partial^2 \ell\left(\boldsymbol{\psi}, \widetilde{\boldsymbol{\omega}}; \boldsymbol{y}_{\mathrm{c}}\right)}{\partial \boldsymbol{\gamma} \partial \widetilde{\omega}_{ij}}\right|_{\widetilde{\boldsymbol{\omega}}=\boldsymbol{\omega}_0} &= \mathbf{0}_{p_3\times 1}.\end{aligned} \tag{18}$$

Note that, when $y_{ij} = 1$, expressions given in (17) and (18) are equal to zero. Thus, initially we carry out an analysis for observations with $y_{ij} = 0$. Subsequently, we alternate the values of y_{ij} and analyze the observations with $y_{ij} = 1$.

3 Monte Carlo simulation studies

3.1 Simulation model and notations

To illustrate the performance of the proposed methodology, we conduct MC simulation studies with $R = 100$ replications each. These studies are based in the MELRM given by $Y_{ij}|b_i \sim \text{Bernoulli}(p_{ij})$ and $b_i \sim \text{N}(0, \sigma^2)$, with systematic component expressed as

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_0 + \beta_1 x_{1ij} + b_i, \quad j = 1, \ldots, n_i, \quad i = 1, \ldots, I, \quad n = \sum_{i=1}^{I} n_i, \tag{19}$$

where $x_{1ij} = u_{ij} - 0.5$, with u_{ij} being a value obtained from $U \sim \text{U}(0, 1)$. In addition, β_0 and β_1 are the regression coefficients.

3.2 Scenario of the simulation

Data sets are generated from (19), fixing the true values of the parameters at $\beta_0 = 1$, $\beta_1 = 1$ and $\sigma^2 = 0.5$, with sample sizes $n = 90$ ($I = 30, n_i = 3$), $n = 360$ ($I = 60, n_i = 6$) and $n = 1080$ ($I = 120, n_i = 9$). We consider values of the perturbation given by $\widetilde{\omega}_{ij} = 0.75, 0.85, 0.95$. Perturbed replications of Y_{ij} under the AMPPS strategy for each data set are generated as follows. For i and $\widetilde{\omega}_{ij}$, we generate n_i observations of Y_{ij} from a Bernoulli distribution with parameter $p_{ij}(\omega_{ij}(\widetilde{\omega}_{ij})) = p_{ij}\omega_{ij}(\widetilde{\omega}_{ij})$, where $\omega_{ij}(\widetilde{\omega}_{ij}) = 1 + g_{ij}(\boldsymbol{\omega}_0)^{-1/2}(\widetilde{\omega}_{ij} - 1)$, with $g_{ij}(\boldsymbol{\omega}_0)$ being approximated with $T = 2000$ additional observations of b_i generated from the $\mathrm{N}(0, \sigma^2)$ distribution. For each replication, the ML estimate of $\boldsymbol{\psi} = (\beta_0, \beta_1, \sigma^2)^\top$ is obtained by AGHQ with 25 quadrature points. Initial values for estimating the parameters by AGHQ were considered as follows. Recall the function `glmer` of the `lme4` package is used, which employs a two-stage optimization process. During the first stage, the optimization is carried out over the $\boldsymbol{\gamma}$ parameter, using starting values by default indicated as "1" for diagonal elements and "0" for off-diagonal elements of the lower triangular matrix in the Cholesky decomposition, plus a vector of zeros for the fixed-effect coefficients. Then, both the estimated $\boldsymbol{\gamma}$ and the starting values for fixed-effect $\boldsymbol{\beta}$ coefficients from the first stage are used as initial values for the second stage of the optimization. In this second stage, the optimization is conducted over the $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$. For more details, see documentation of the `lme4` package in https://cran.r-project.org/web/packages/lme4/lme4.pdf.

The local influence analysis is performed with $S - M_0 = 400$ ($S = 500$, $M_0 = 100$) observations of b_i generated through the MH algorithm for approximating $-\ddot{\boldsymbol{Q}}_{\boldsymbol{\psi}}(\widehat{\boldsymbol{\psi}})$ and $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$. In addition, as mentioned in Sect. 2.3, the initial value $\boldsymbol{b}_i^{(0)}$ of $\boldsymbol{b}_i$ is not relevant due to the reasons there indicated. Note that as $T = 2000$ observations of b_i are generated from a normal distribution with zero mean and variance $\widehat{\sigma}^2$ for approximating the elements $g_{ij}(\boldsymbol{\omega}_0)$, where $\widehat{\sigma}^2$ is the ML estimate of σ^2. Here, we use the following terminology for the different types of influential cases, in relation to the observed value of the response or in the estimation procedure. If the observation is detected as influential and the value of the perturbed response is different from the value of the original response, the observation is identified as truly influential (TI hereafter). If the observation is detected as influential and the value of the perturbed response is equal to the value of the original response, the observation is identified as false influential (FI hereafter). If the observation is detected as influential independent of the value of the response, the observation is identified as potential influential (PI hereafter) in the ML estimate.

3.3 Results of the simulation

Table 1 reports the percentages of detection for each type of influential points in the simulation studies. For the different $\widetilde{\omega}_{ij}$, the percentages of detection of TI, FI and PI improve considerably as n increases. Specifically, for $\widetilde{\omega}_{ij} = 0.75$, the percentages of detection of TI, FI and PI are up to 54%, 65% and 99%, respectively. For $\widetilde{\omega}_{ij} = 0.85$, the percentages of detection of TI, FI and PI are up to 48%, 70% and 96%, respectively.

Table 1 Percentages (%) of detection of TI, FI and PI for indicated values of the simulation

n	I	n_i	$\widetilde{\omega}_{ij} = 0.75$			$\widetilde{\omega}_{ij} = 0.85$			$\widetilde{\omega}_{ij} = 0.95$		
			TI	FI	PI	TI	FI	PI	TI	FI	PI
90	30	3	37	35	72	34	38	72	31	40	71
			34	26	60	32	28	60	33	28	61
			39	30	69	38	33	71	35	37	72
360	90	6	43	52	95	34	59	93	28	62	90
			29	28	57	34	27	61	25	25	50
			28	24	52	27	17	44	17	18	35
			49	43	92	48	41	89	46	41	87
			45	30	75	40	34	74	33	31	64
			43	42	85	35	40	75	25	43	68
1080	120	9	54	38	92	37	52	89	36	45	81
			39	34	73	41	30	71	34	35	69
			40	33	73	37	39	76	32	32	64
			33	37	70	31	32	63	23	27	50
			19	21	40	19	13	32	18	10	28
			36	43	79	36	43	79	26	38	64
			23	14	37	26	16	42	22	11	33
			35	38	73	28	38	66	24	31	55
			34	65	99	26	70	96	17	73	90

For $\widetilde{\omega}_{ij} = 0.95$, the percentages of detection of TI, FI and PI are up to 46%, 73% and 90%, respectively. Consequently, the percentages of detection of TI and FI are satisfactory, indicating that the value of the response and its associated probability of success can determine its influence. In addition, the percentages of detection for PI are very satisfactory, showing that the proposed methodology is able to detect the perturbed observations as influential when they really are.

As an alternative way to detect influence in the MELRM, simulation studies can be performed for global influence analysis by deleting measurements and subjects, following the work of Xu et al. (2006), which also is based on the Q-displacement function. Some comments about future research for global influence in the MELRM are provided in the final section.

4 Numerical illustrations with biological data

4.1 Seeds data

Orobanche is a kind of parasitic plants without chlorophyll that grow on the roots of many *dicotyledonous* crop plants. To determine the factors that affect the germination of the seed for the species *Orobanche Aegyptiaca*, the following experiment was performed. A batch of seeds for the varieties *Orobanche Aegyptiaca* 75 (OA75) and *Orobanche Aegyptiaca* 73 (OA73) was spread on plates containing a dilution 1/125 of root extract for bean and cucumber plant. The number of seeds in the batch for the

combination of seed varieties and root extracts was different, constituting an unbalanced data set. The results of the experiment were originally reported by Crowder (1978). The binary responses corresponding to the germination or non-germination of the seeds in each plate are considered in this work. See more details on analyses of these data in Crowder (1978), Breslow and Clayton (1993), Zhu and Lee (2003) and Chen et al. (2010).

For the binary responses corresponding to the germination or non-germination of the seeds in each plate, we consider an MELRM given by $Y_{ij}|b_i \sim \text{Bernoulli}(p_{ij})$, with $b_i \sim \text{N}(0, \sigma^2)$ and

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{1ij} x_{2ij} + b_i,$$
$$j = 1, \ldots, n_i, \quad i = 1, \ldots, 21, n = 831,$$

where x_{1ij} is the observed value of the covariate "variety of the seed" (1 for OA73 and 0 for OA75), x_{2ij} is the observed value of the covariate "type of root extract" (1 for beans and 0 for cucumber), $x_{1ij}x_{2ij}$ is the interaction term, and β_0, β_1, β_2, β_3 are the regression coefficients.

The ML estimate of $\boldsymbol{\psi} = (\beta_0, \beta_1, \beta_2, \beta_3, \sigma^2)^\top$ is obtained by AGHQ with 25 quadrature points, and their values are presented in Table 2. The local influence analysis is carried out with $S - M_0 = 9000$ ($S = 10{,}000$, $M_0 = 1000$) observations of b_i generated through the MH algorithm for approximating $-\ddot{\boldsymbol{Q}}_{\boldsymbol{\psi}}(\widehat{\boldsymbol{\psi}})$ and $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$. Also, $T = 2000$ observations of b_i are generated from a normal distribution with zero mean and variance $\widehat{\sigma}^2$ for approximating the elements $g_{ij}(\boldsymbol{\omega}_0)$, where $\widehat{\sigma}^2$ is the ML estimate of σ^2 as defined in Sect. 3.2.

Figure 1a, b shows index plots of the conformal normal curvature for local influence with (a) $y_{ij} = 0$ and (b) $y_{ij} = 1$, respectively, under the AMPPS strategy. For $y_{ij} = 0$, we have the observations: #456 to #476 of plate (subject) #12; #532 to #548 of plate #13; and #646 to #678 of plate #15 are detected as influential. For $y_{ij} = 1$, the observations: #40 to #62 of plate #2; and #102 to #124 of plate #3 are detected as influential. For comparison, under the MPPS strategy, Fig. 1c, d shows index plots of the conformal normal curvature for local influence with (a) $y_{ij} = 0$ and (b) $y_{ij} = 1$, respectively. From this figure, note that the same results are obtained for both strategies. In order to evaluate the magnitude of the impact exerted in the ML estimates by individual influential observations or a set of them, we compare the ML estimates with those obtained by dropping such observations using the percentage error (PE) given by

$$\text{PE} = \left|\left(\widehat{\psi}_k - \widehat{\psi}_k^*\right)/\widehat{\psi}_k\right| \times 100\%,$$

where $\widehat{\psi}_k$ is the ML estimate of ψ_k obtained from the fit of the model with all observations and $\widehat{\psi}_k^*$ is the ML estimates of ψ_k obtained from the fit of the model excluding individual influential observations or a set of them, if these observations belong to the same subject, for $k = 1, \ldots, 5$.

Table 2 provides the results for two of the combinations that show the largest values of the PE, for each strategy. For the combinations that present the first largest values of

Table 2 Estimates, p-values and PE for seeds data

Dropped observations	Plates	Parameter	Estimate	p-value	PE
None	–	β_0	−0.548	0.001	–
		β_1	0.081	0.776	–
		β_2	1.340	<0.001	–
		β_3	−0.821	0.037	–
		σ	0.249	–	–
AMPPS strategy					
40–62, 102–124, 456–476, 532–548, 646–678	2, 3, 12, 13, 15	β_0	−1.998	0.009	264.356
		β_1	1.215	0.253	1409.102
		β_2	5.149	<0.001	284.421
		β_3	−4.570	0.003	456.339
		σ	1.523	–	512.803
40–62, 102–124, 456–476, 646–678	2, 3, 12, 15	β_0	−1.908	0.006	248.051
		β_1	1.165	0.220	1346.567
		β_2	4.198	<0.001	213.411
		β_3	−3.644	0.007	343.691
		σ	1.353	–	442.598
MPPS strategy					
40–62, 102–124, 456–476, 532–548, 646–678	2, 3, 12, 13, 15	β_0	−1.998	0.009	264.356
		β_1	1.215	0.253	1409.102
		β_2	5.149	<0.001	284.421
		β_3	−4.570	0.003	456.339
		σ	1.523	–	512.803
40–62, 102–124, 456–476, 646–678	2, 3, 12, 15	β_0	−1.908	0.006	248.051
		β_1	1.165	0.220	1346.567
		β_2	4.198	<0.001	213.411
		β_3	−3.644	0.007	343.691
		σ	1.353	–	442.598

the PE, note that by dropping the combinations #40 to #62, #102 to #124, #456 to #476, #532 to #548, and #646 to #678, the ML estimates provide very large variations, being them affected by the AMPPS strategy. In addition, the p-value associated with β_3 is less than 0.01, becoming the interaction significant at 1% instead of 5% before the removal. However, for the other estimates, despite the large variations, no changes are obtained in relation to the significance of the covariates. For the MPPS strategy, we observe the same results. For the combinations that present the second largest values of the PE, note that for both strategies when dropping the combinations #40 to #62, #102 to #124, #456 to #476, and #646 to #678, the ML estimates present very large variations, being them affected by the respective strategies. Furthermore, once again the p-value associated with β_3 is less than 0.01, becoming the interaction significant at 1% instead of 5% before the removal. Nevertheless, for the other estimates, despite the large variations, no changes are detected in relation to the significance of the covariates. Note

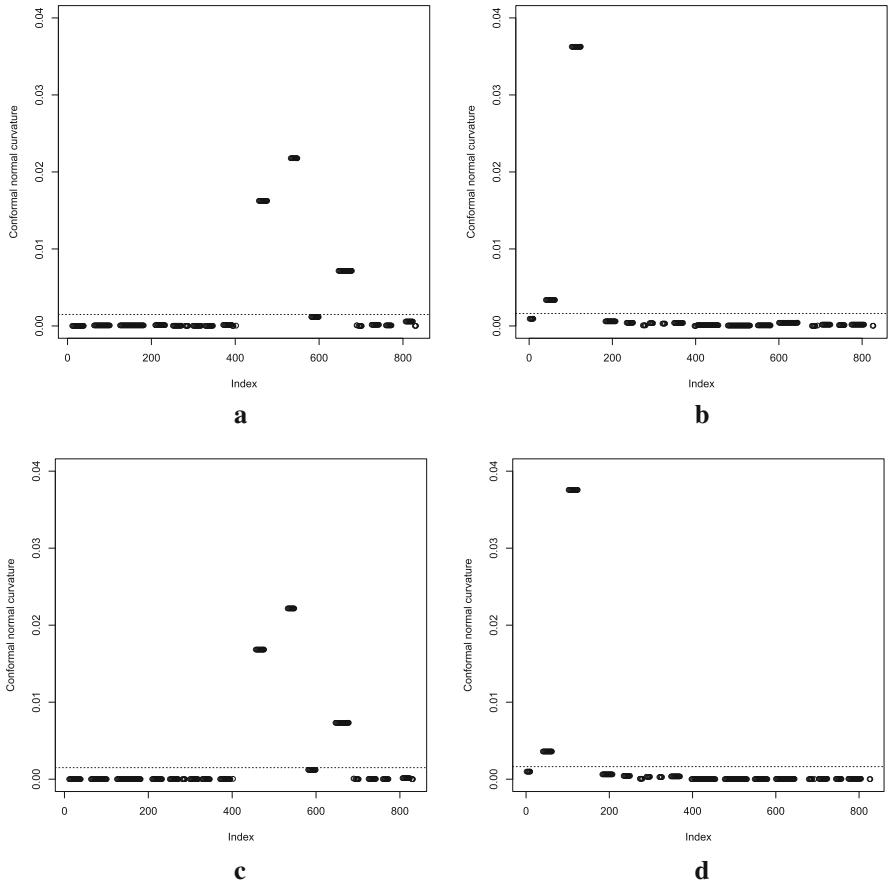

Fig. 1 Index plots of conformal normal curvature for seeds data. **a** $y_{ij} = 0$ and **b** $y_{ij} = 1$, under the AMPPS strategy and **c** $y_{ij} = 0$ and **d** $y_{ij} = 1$, under the MPPS strategy

that, in this application, both MPPS and AMPPS strategies did not show differences in relation to the points detected as influential. Then, the combinations were the same and, therefore, the p-values and PE also. To determine how the removal of combinations of influential cases affects the predictive performance of the model, we calculate the Acc, Sen and Spe indicators. Table 3 reports these results. For the data set with all observations, values for the Acc, Sen and Spe indicators are 0.637, 0.644 and 0.629, respectively. Hence, as noted, removal of combinations of influential cases leads to substantial increase in the predictive indicators. In summary, the results of the proposed methodology lead to very large variations in the ML estimates of the parameters, and to changes related to significance of a covariate of the model. Moreover, these results imply a substantial increase in the Acc, Sen and Spe indicators, improving considerably the predictive performance of the model, when the observations detected as influential are removed.

4.2 Salamander data

A mating experiment of salamanders conducted in the summer season was presented by McCullagh and Nelder (1983). This experiment involves two populations of salamanders: *Rough Butt* (RB) and *Whiteside* (WS). In their natural habitat, these two populations are geographically isolated from each other. The main question is whether barriers to interbreeding have evolved so that matings within the population are more successful than those between populations. A total of 40 salamanders were used from the two populations, with ten males and ten females. Each female was paired six times with three males from her own population and three males from the other, constituting a balanced data set. Then, binary responses corresponding to the success or failure of the mating were obtained. These data have been studied also by other authors; see, for example, Breslow and Clayton (1993), Larsen et al. (2000) and Jiang (2007). We consider an MELRM to describe these data as follows.

Let Y_{ij} be the response for mating female i and male j. Then, $Y_{ij}|b_i^{\mathrm{f}}, b_j^{\mathrm{m}} \sim$ Bernoulli(p_{ij}), with $b_i^{\mathrm{f}} \sim \mathrm{N}(0, \sigma_{\mathrm{f}}^2)$ and $b_j^{\mathrm{m}} \sim \mathrm{N}(0, \sigma_{\mathrm{m}}^2)$ being independent, and

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{1ij} x_{2ij} + b_i^{\mathrm{f}} + b_j^{\mathrm{m}},$$
$$j = 1, \ldots, 6, \quad i = 1, \ldots, 20, n = 120,$$

where x_{1ij} is the observed indicator covariate of females in the population WS (1 for WS and 0 for RB), x_{2ij} is the observed covariate of males in the population WS (1 for WS and 0 for RB) and $x_{1ij}x_{2ij}$ is the interaction term. Here, β_0, β_1, β_2, β_3 are the regression coefficients and b_i^{f}, b_j^{m} are the random effects associated with females or males in the pair, respectively. The ML estimate of $\boldsymbol{\psi} = (\beta_0, \beta_1, \beta_2, \beta_3, \sigma_{\mathrm{f}}^2, \sigma_{\mathrm{m}}^2)^\top$ is obtained by AL, and their values are presented in Table 4.

The local influence analysis is carried out again with $S - M_0 = 9000$ ($S = 10{,}000$, $M_0 = 1000$) observations of b_i^{f} and b_j^{m} generated through the MH algorithm for approximating $-\ddot{\boldsymbol{Q}}_{\boldsymbol{\psi}}(\widehat{\boldsymbol{\psi}})$ and $\boldsymbol{\Delta}_{\boldsymbol{\omega}_0}$. Furthermore, once again, we consider $T = 2000$ observations of b_i^{f} generated from a normal distribution with zero mean and variance $\widehat{\sigma}_{\mathrm{f}}^2$, as well as $T = 2000$ observations of b_j^{m} generated from a normal

Table 3 Acc, Sen and Spe indicators for seeds data

Dropped observations	Plates	Sen	Spe	Acc
None	–	0.644	0.629	0.637
AMPPS strategy				
40–62, 102–124, 456–476, 532–548, 646–678	2, 3, 12, 13, 15	0.737	0.754	0.745
40–62, 102–124, 456–476, 646–678	2, 3, 12, 15	0.737	0.718	0.728
MPPS strategy				
40–62, 102–124, 456–476, 532–548, 646–678	2, 3, 12, 13, 15	0.737	0.754	0.745
40–62, 102–124, 456–476, 646–678	2, 3, 12, 15	0.737	0.718	0.728

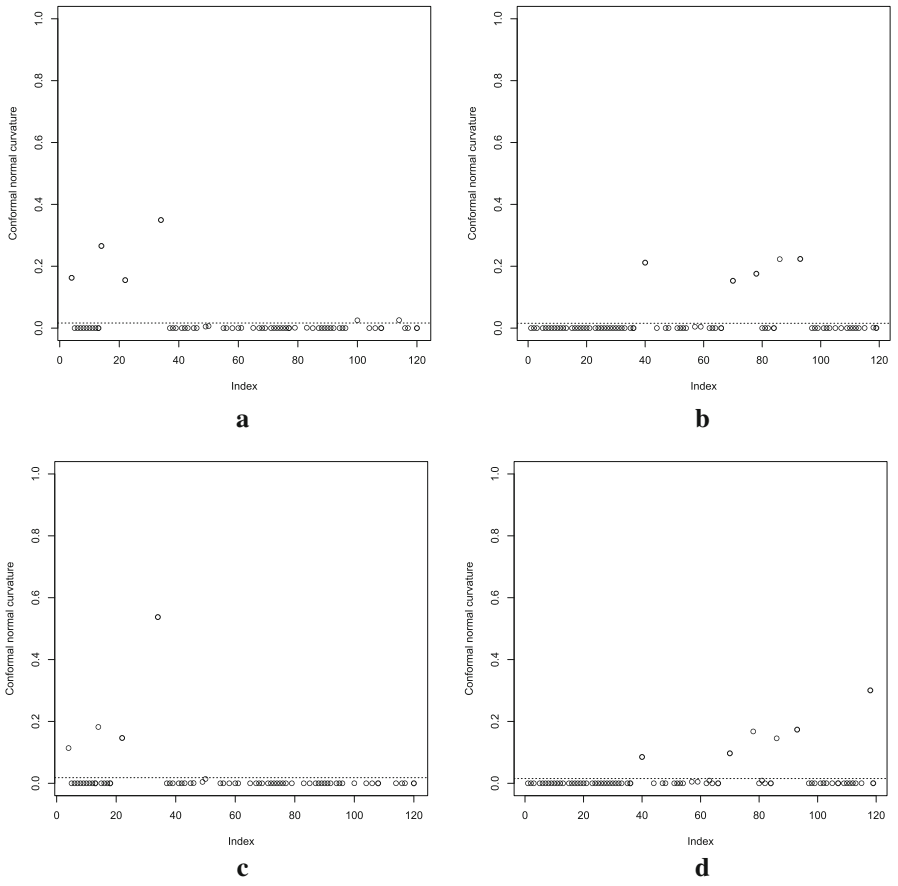

Fig. 2 Index plots of conformal normal curvature for the salamander data: **a** $y_{ij} = 0$ and **b** $y_{ij} = 1$, under the AMPPS strategy, and **c** $y_{ij} = 0$ and **d** $y_{ij} = 1$, under the MPPS strategy

distribution with mean zero and variance $\widehat{\sigma}^2_{\mathrm{m}}$, for approximating the elements $g_{ij}(\boldsymbol{\omega}_0)$, where $\widehat{\sigma}^2_{\mathrm{f}}$ and $\widehat{\sigma}^2_{\mathrm{m}}$ are as defined in Sect. 3.2.

For the AMPPS strategy, Fig. 2a, b shows index plots of the conformal normal curvature for local influence with $y_{ij} = 0$ and $y_{ij} = 1$, respectively. For $y_{ij} = 0$, the pairs (1,11), (3,11), (4,15), (6,7), (17,2) and (19,4) are detected as influential, but if $y_{ij} = 1$, the pairs (7,9), (12,17), (13,16), (15,18) and (16,13) are identified as influential. For comparison, under the MPPS strategy, Fig. 2c, d displays index plots of the conformal normal curvature for local influence with $y_{ij} = 0$ and $y_{ij} = 1$, respectively. For $y_{ij} = 0$, the pairs (17,2) and (19,4) are no longer detected as influential, whereas for $y_{ij} = 1$, the pair (20,3) is considered as influential, in addition to the cases already considered as influential by the AMPPS strategy.

Table 4 reports the results for the combinations of pairs that present the largest values of PE for each strategy. We observe that by dropping the combination of

Table 4 Estimates, p-values and PE for salamander data

Dropped (female, male) observations	Parameter	Estimate	p-value	PE
None	β_0	1.335	0.042	–
	β_1	− 2.940	0.003	–
	β_2	− 0.422	0.525	–
	β_3	3.181	0.003	–
	σ^{f}	1.255	–	–
	σ^{m}	0.269	–	–
AMPPS strategy				
(1,11), (3,11), (4,15), (6,7), (7,9), (12,17)	β_0	2.211	0.066	65.549
	β_1	− 4.382	0.011	49.015
	β_2	0.255	0.781	160.337
	β_3	3.210	0.017	0.904
	σ^{f}	2.383	–	89.860
	σ^{m}	0.609	–	126.874
MPPS strategy				
(1,11), (3,11), (4,15), (7,9), (12,17), (16,3)	β_0	1.787	0.106	33.845
	β_1	− 4.046	0.014	37.611
	β_2	0.526	0.557	224.644
	β_3	2.809	0.033	11.714
	σ^{f}	2.350	–	87.292
	σ^{m}	0.627	–	133.613

pairs (1,11), (3,11), (4,15), (6,7), (7,9), (12,17), the estimates of β_0, β_1, β_2, σ_{f}^2 and σ_{m}^2 provide variations which are up to approximately 160%, for the AMPPS strategy. For comparison, under the MPPS strategy, we note that by removing the combination of pairs (1,11), (3,11), (4,15), (7,9), (12,17), (16,3), the estimates of β_0, β_1, β_2, σ_{f}^2 and σ_{m}^2 present variations which are up to approximately 224%. Notice that, in this application, differences between the points detected by both MPPS and AMPPS strategies were obtained. Therefore, the removal of combinations produced different effects with both strategies, noting that the points detected by the AMPPS strategy are more influential in terms of p-values and PE values. With respect to the predictive performance of the model, Table 5 presents the results of the Acc, Sen and Spe indicators. Observe that by dropping the combination of pairs (1,11), (3,11), (4,15), (6,7), (7,9), (12,17), the Acc, Sen and Spe indicators increase approximately 10%, for the AMPPS strategy. In addition, for the MPPS strategy, note that by removing the combination of pairs (1,11), (3,11), (4,15), (7,9), (12,17), (16,3), the Acc, Sen and Spe indicators increase approximately 8%. In summary, the results of the proposed methodology lead to large variations in the ML estimates, but not to inferential changes. Furthermore, the detection and removal of influential observations allowed a considerable improvement in the predictive performance of the model.

Table 5 Acc, Sen and Spe indicators for salamander data

Dropped (female, male) observations	Sen	Spe	Acc
None	0.814	0.820	0.816
AMPPS strategy			
(1,11), (3,11), (4,15), (6,7), (7,9), (12,17)	0.912	0.913	0.912
MPPS strategy			
(1,11), (3,11), (4,15), (7,9), (12,17), (16,13)	0.896	0.894	0.895

5 Conclusions and future research

In this article, we derived a perturbation strategy for the binary responses of the mixed effects logistic regression model. This strategy allowed us to investigate the influence of these responses in the maximum likelihood estimates and in the predictive performance of the model. The derived strategy was studied using the local influence technique based on the Q-displacement function presented by Zhu and Lee (2003). Due to the discrete nature of binary response, taking only zero and one values, standard local influence diagnostic techniques do not apply for perturbing the response of this model. The proposed strategy corresponds to an appropriate multiplicative perturbation of the probability of success associated with this response, as a form of evaluating the perturbation in the response. It should be noted that it is possible to perturb the probability of success in different ways, but an immediate form is the multiplicative perturbation. However, since to arbitrarily perturb the model or the data may lead to unreliable results about local influence, the appropriate perturbation of the probability of success was considered. A very important point of the proposed perturbation strategy is that it allowed us to investigate the influence of binary responses in the predictive performance of the mixed effects logistic regression model, converting it into a powerful diagnostic tool when the purpose of the statistical analysis is the prediction. The proposed perturbation strategy permitted us to investigate the influence of individual responses or of sets of them, but not of the subjects.

From the results of the illustrations with two biological data sets, we showed that it is possible to detect influential observations of the binary responses that allow us to avoid misleading maximum likelihood estimates and to improve the predictive performance of the model, obtaining better results with the appropriate multiplicative perturbation of the probability of success, in both balanced and unbalanced data sets. It should be noted that these conclusions are based on the percentage error analysis, considering the individual influential observations or sets of them, as a way of taking into account their joint effect.

From the results of the Monte Carlo simulation, we showed that it is possible to detect the perturbed observations as influential, providing additional information with respect to the nature of the influence. Thus, local influence diagnostics can be established by both the value of binary response and the probability of success. Simulations and illustrations were carried out in two stages: first, when the values of the binary responses assume the value equal to zero and second, when the values of the binary responses assume the value equal to one. However, this procedure can be automated.

As a recommendation, if one want to select one diagnostic tool for the mixed effects logistic regression model, our research showed that the proposed methodology is useful in this selection. Thus, it allows the practitioners to obtain valuable information about maximum likelihood estimates, predictive performance and binary responses that need additional scrutiny in this type of models, enabling us to get a better perspective on the data analytic consequences when influential observations are detected.

As an alternative way to detect influence, global influence analysis for the mixed effects logistic regression model can be performed by deleting measurements and subjects, following the work of Xu et al. (2006), which also is based on the Q-displacement function. This will allow us to contrast the results of local influence proposed in the present work with a global influence study in a future work.

Acknowledgements The authors thank the Editors and two referees for their constructive comments on an earlier version of this manuscript which resulted in this improved version. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, by HPC resources provided by the Information Technology Superintendence of the University of São Paulo, and also by CNPq from Brazil; as well as by the Chilean Council for Scientific and Technology Research (CONICYT) through fellowship "Becas-Chile" (A. Tapia) and FONDECYT 1160868 Grant (V. Leiva) from the Chilean government.

References

Agresti A (2003) Categorical data analysis, vol 482. Wiley, New York

Assumpção RAB, Uribe-Opazo MA, Galea M (2014) Analysis of local influence in geostatistics using student-t distribution. J Appl Stat 41:2323–2341

Breslow NE, Clayton DG (1993) Approximate inference in generalized linear mixed models. J Am Stat Assoc 88(421):9–25

Capanu M, Gönen M, Begg CB (2013) An assessment of estimation methods for generalized linear mixed models with binary outcomes. Stat Med 32:4550–4566

Caro-Lopera F, Leiva V, Balakrishnan N (2012) Connection between the Hadamard and matrix products with an application to matrix-variate Birnbaum–Saunders distributions. J Multivar Anal 104:126–139

Chen F, Zhu H-T, Song X-Y, Lee S-Y (2010) Perturbation selection and local influence analysis for generalized linear mixed models. J Comput Graph Stat 19:826–842

Cook RD (1986) Assessment of local influence. J R Stat Soc B 48:133–169

Cook RD, Weisberg S (1982) Residuals and influence in regression. Chapman and Hall, London

Crowder MJ (1978) Beta-binomial ANOVA for proportions. J R Stat Soc C 27:34–37

De Bastiani F, Cysneiros AHMA, Uribe-Opazo MA, Galea M (2015) Influence diagnostics in elliptical spatial linear models. TEST 24:322–340

Demidenko E (2013) Mixed models: theory and applications with R. Wiley, Hoboken

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc B 39:1–38

Díaz-García J, Galea M, Leiva V (2003) Influence diagnostics for elliptical multivariate linear regression models. Commun Stat Theory Methods 32:625–641

Diggle PJ, Liang K-Y, Zeger SL (1996) Analysis of longitudinal data. Oxford University Press, London

Garcia-Papani F, Leiva V, Uribe-Opazo MA, Aykroyd RG (2018) Birnbaum–Saunders spatial regression models: diagnostics and application to chemical data. Chemom Intell Lab Syst 177:114–128

Hosmer DW, Lemeshow S, Sturdivant RX (2013) Applied logistic regression. Wiley, Hoboken

Hossain M, Islam MA (2003) Application of local influence diagnostics to the linear logistic regression models. Dhaka Univ J Sci 51:269–278

Ibacache-Pulgar G, Paula GA, Cysneiros FJA (2013) Semiparametric additive models under symmetric distributions. TEST 22:103–121

Jiang J (2007) Linear and generalized linear mixed models and their applications. Springer, New York

Larsen K, Petersen JH, Budtz-Jørgensen E, Endahl L (2000) Interpreting parameters in the logistic regression model with random effects. Biometrics 56:909–914
Leão J, Leiva V, Saulo H, Tomazella V (2017) Birnbaum–Saunders frailty regression models: diagnostics and application to medical data. Biomet J 59:291–314
Leiva V, Santos-Neto M, Cysneiros FJA, Barros M (2014) Birnbaum–Saunders statistical modelling: a new approach. Stat Model 14:21–48
Lesaffre E, Spiessens B (2001) On the effect of the number of quadrature points in a logistic random-effects model: an example. J R Stat Soc C 50:325–335
Lesaffre E, Verbeke G (1998) Local influence in linear mixed models. Biometrics 54:570–582
Liu S (2000) On local influence in elliptical linear regression models. Stat Pap 41:211–224
Liu S (2004) On diagnostics in conditionally heteroskedastic time series models under elliptical distributions. J Appl Probab 41:393–406
Liu Y, Ji G, Liu S (2015) Influence diagnostics in a vector autoregressive model. J Stat Comput Simul 85:2632–2655
Marchant C, Leiva V, Cysneiros FJA, Vivanco JF (2016) Diagnostics in multivariate generalized Birnbaum–Saunders regression models. J Appl Stat 43:2829–2849
McCullagh P, Nelder JA (1983) Generalized linear models. Chapman and Hall, London
McCulloch CE (1997) Maximum likelihood algorithms for generalized linear mixed models. J Am Stat Assoc 92:162–170
McCulloch S, Searle S (2001) Generalized, linear and mixed models. Wiley, New York
Molenberghs G, Verbeke G (2005) Models for discrete longitudinal data. Springer, New York
Nyangoma SO, Fung WK, Jansen RC (2006) Identifying influential multinomial observations by perturbation. Comput Stat Data Anal 50:2799–2821
Ouwens MJNM, Tan FES, Berger MPF (2001) Local influence to detect influential data structures for generalized linear mixed models. Biometrics 57:1166–1172
Pinheiro JC, Chao EC (2006) Efficient Laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear mixed models. J Comput Graph Stat 15:58–81
Poon WY, Poon YS (1999) Conformal normal curvature and assessment of local influence. J R Stat Soc B 61:51–61
R Core Team (2016) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
Rakhmawati TW, Molenberghs G, Verbeke G, Faes C (2017) Local influence diagnostics for generalized linear mixed models with overdispersion. J Appl Stat 44:620–641
Raudenbush SW, Yang M, Yosef M (2000) Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation. J Comput Graph Stat 9:141–157
Robert CP, Casella G (1999) Monte Carlo statistical methods. Springer, New York
Rocha AV, Simas AB (2011) Influence diagnostic in a general class of beta regression models. TEST 20:95–119
Santos-Neto M, Cysneiros FJA, Leiva V, Barros M (2016) Reparameterized Birnbaum–Saunders regression models with varying precision. Electron J Stat 10:2825–2855
Stehlík M, Rodríguez-Díaz JM, Müller WG, López-Fidalgo J (2008) Optimal allocation of bioassays in the case of parametrized covariance functions: an application to lung's retention of radioactive particles. TEST 17:56–68
Stiratelli R, Laird N, Ware JH (1984) Random effects models for serial observations with binary responses. Biometrics 40:961–971
Svetliza CF, Paula GA (2001) On diagnostics in log-linear negative binomial models. J Stat Comput Simul 71:231–244
Wolfinger R, O'Connell M (1993) Generalized linear mixed models: a pseudo-likelihood approach. J Stat Comput Simul 48(3–4):233–243
Xu L, Lee SY, Poon WY (2006) Deletion measures for generalized linear mixed effects models. Comput Stat Data Anal 51:1131–1146
Zhu H-T, Lee S-Y (2001) Local influence for incomplete-data models. J R Stat Soc B 63:111–126
Zhu H-T, Lee S-Y (2003) Local influence for generalized linear mixed models. Can J Stat 31:293–309
Zhu H, Ibrahim JG, Lee S, Zhang H (2007) Perturbation selection and influence measures in local influence analysis. Ann Stat 35:2565–2588