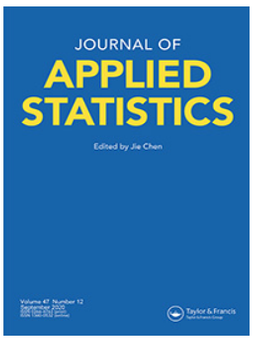

Journal of Applied Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cjas20

The multinomial logistic regression model for predicting the discharge status after liver transplantation: estimation and diagnostics analysis

E. M. Hashimoto , E. M. M. Ortega , G. M. Cordeiro , A. K. Suzuki & M. W. Kattan

To cite this article: E. M. Hashimoto , E. M. M. Ortega , G. M. Cordeiro , A. K. Suzuki & M. W. Kattan (2020) The multinomial logistic regression model for predicting the discharge status after liver transplantation: estimation and diagnostics analysis, Journal of Applied Statistics, 47:12, 2159-2177, DOI: 10.1080/02664763.2019.1706725

To link to this article: https://doi.org/10.1080/02664763.2019.1706725

Published online: 24 Dec 2019.

JOURNAL OF APPLIED STATISTICS
2020, VOL. 47, NO. 12, 2159–2177
https://doi.org/10.1080/02664763.2019.1706725

Taylor & Francis
Taylor & Francis Group

The multinomial logistic regression model for predicting the discharge status after liver transplantation: estimation and diagnostics analysis

E. M. Hashimoto[a], E. M. M. Ortega[b], G. M. Cordeiro[c], A. K. Suzuki[d] and M. W. Kattan[e]

[a]Departamento Acadêmico de Matemática, Universidade Tecnológica Federal do Paraná, Londrina, PR, Brazil; [b]Departamento de Ciências Exatas, Universidade de S ao Paulo, Piracicaba, SP, Brazil; [c]Departamento de Estatística, Universidade Federal de Pernambuco, Recife, PE, Brazil; [d]Departamento de Matemática Aplicada e Estatística, Universidade de S ao Paulo, S ao Carlos, SP, Brazil; [e]Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH, USA

ABSTRACT

The multinomial logistic regression model (MLRM) can be interpreted as a natural extension of the binomial model with logit link function to situations where the response variable can have three or more possible outcomes. In addition, when the categories of the response variable are nominal, the MLRM can be expressed in terms of two or more logistic models and analyzed in both frequentist and Bayesian approaches. However, few discussions about post modeling in categorical data models are found in the literature, and they mainly use Bayesian inference. The objective of this work is to present classic and Bayesian diagnostic measures for categorical data models. These measures are applied to a dataset (status) of patients undergoing kidney transplantation.

ARTICLE HISTORY
Received 13 June 2018
Accepted 13 December 2019

KEYWORDS
Categorical data; diagnostic analysis; multinomial distribution; nominal response; regression model

1. Introduction

Categorical data are very common in areas related to public health. For instance, when testing a new drug, the observed response can be if the patient improved (yes or no), or with respect to surgery, the patient can be classified into three categories (released from the hospital, remained hospitalized or died). In both cases, the attributes do not have a natural scale, so they are called nominal categorical variables [1]. However, since categorical data are classified as attributes or qualities, they cannot be analyzed using methods proposed for numerical data [31]. Hence, in this case, more specific methods for these types of data are required.

Logistic regression is one of the most important methods for statistical modeling of data. Even when the response of interest is not originally binary, some researchers have dichotomized the response such that the probability of success can be modeled via logistic regression. Then, the logistic regression is more often employed to measure the relation between a dichotomous response variable (binomial) and a set of explanatory variables, but with some modifications it can also be used when the response variable is polychotomous

CONTACT A. K. Suzuki suzuki@icmc.usp.br

(multinomial). The extension of a binomial to a multinomial model is easily illustrated when the response variable has three categories (trinomial).

Several papers have proposed models to analyze categorical data. Albert and Chib [2] presented an analysis of nominal and ordinal categorical data from a Bayesian standpoint. Hartzel *et al.* [14] introduced a model with random effects to analyze nominal and ordinal categorical data. Ren [25] worked with the logistic model to predict the toxic action of phenols. Fagerland *et al.* [13] presented a goodness-of-fit test for nominal categorical data. Lee *et al.* [20] utilized a logistic regression model to fit acute gastrointestinal hemorrhage data.

For the inference approach, the model parameters can be estimated using the maximum likelihood method. However, the hypotheses and confidence interval tests are based on the asymptotic distribution of the maximum likelihood estimators, since the response variable is not continuous. Alternatively, we can use the Bayesian inference and the bootstrap estimator.

On the other hand, after modeling, our purpose is to detect possible influential observations that may cause distortions in the results. Andersen [3] presented measures such as Cook's distance and leverage for categorical data models, while Steen *et al.* [28] presented local influence measures for ordinal data. However, in the context of the Bayesian inference, there is little discussion about the use of sensitivity analysis. For this reason, our aim is to present diagnostic measures in the Bayesian context for nominal categorical data.

Thus, our research focuses on the following contributions by consolidating the multinomial logistic regression model (MLRM):

- First, we consider three types of estimation: classic, Bayesian and bootstrap for the MLRM.
- Second, a diagnostic analysis is proposed for this regression model considering two approaches: frequentist and Bayesian.
- Third, a real liver transplant dataset is used to illustrate the application of the MLRM.

The article is organized as follows. In Section 2, we define the MLRM and discuss three methods (maximum likelihood, Bayesian analysis and bootstrap re-sampling) for estimating the model parameters. Diagnostic measures are presented in Section 3 by considering case-deletion and normal curvatures of local influence under some perturbation schemes with censored observations. In Section 4, we fit the MLRM to a real data set from patients who have undergone liver transplantation. Finally, Section 5 offers some conclusions.

2. Model definition

In developing models for polychotomous responses having nominal scale, Fagerland *et al.* [13] defined a random variable Y that takes values in the set $\{1, \ldots, c\}$, where c is the number of categories, where $Y = 1$ is the reference category, i.e. the remaining categories are compared with the benchmark category. The choice of the reference category is arbitrary and based on previous knowledge of the researcher.

We can transform the nominal category of the response variable into a numerical scale. Fagerland *et al.* [13] expressed the categorical response variable Y using a vector of binary responses. In other words, let $\widetilde{\mathbf{Y}} = (\widetilde{Y}_2, \ldots, \widetilde{Y}_c)^\top$ be a vector of dummy variables such

that $\widetilde{Y}_j = 1$ if $Y = j$ and $\widetilde{Y}_j = 0$ otherwise, for $j = 2, \ldots, c$, given that category 1 is the reference value. In this way, $\widetilde{\mathbf{Y}}$ is a random variable with a multinomial distribution, say $\widetilde{\mathbf{Y}} \sim MN(\mathbf{p}, 1)$, and probability mass function

$$f(\tilde{\mathbf{y}}) = \frac{1}{\prod_{j=1}^{c} \tilde{y}_j!} \prod_{j=1}^{c} p_j^{\tilde{y}_j}, \tag{1}$$

where $\mathbf{p} = (p_1, \ldots, p_c)^\top$ is a vector of probabilities such that $\sum_{j=1}^{c} p_j = 1$.

For a set of k explanatory variables, denoted by $\mathbf{x} = (x_1, \ldots, x_k)^\top$, we define

$$p_j(\mathbf{x}) = P(Y = j|\mathbf{x}), \quad j = 1, \ldots, c,$$

which is a multinomial probability such that $\sum_{j=1}^{c} p_j(\mathbf{x}) = 1$.

Hedeker [15] and Fagerland *et al.* [13] defined the odds ratio for the jth category compared to the first category by

$$\frac{p_j(\mathbf{x})}{p_1(\mathbf{x})}, \quad \text{for} \quad 2 \leq j \leq c.$$

Thus,

$$\log\left[\frac{p_j(\mathbf{x})}{p_1(\mathbf{x})}\right] = \alpha_j + \boldsymbol{\beta}_j^\top \mathbf{x}, \quad j = 2, \ldots, c,$$

where α_j is the intercept and $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{jk})^\top$ is the vector of covariates for the jth category of the response variable.

The multinomial probabilities for each category are defined (for $j = 2, \ldots, c$) by

$$p_j(\mathbf{x}) = \frac{\exp(\alpha_j + \boldsymbol{\beta}_j^\top \mathbf{x})}{1 + \sum_{k=2}^{c} \exp(\alpha_j + \boldsymbol{\beta}_j^\top \mathbf{x})} \tag{2}$$

and

$$p_1(\mathbf{x}) = \frac{1}{1 + \sum_{k=2}^{c} \exp(\alpha_j + \boldsymbol{\beta}_j^\top \mathbf{x})}. \tag{3}$$

For $c = 2$, $p_1(\mathbf{x}) = 1 - p_2(\mathbf{x})$, and Equations (2) and (3) reduce to the logit link function of a binary response variable.

Suppose now that the covariate X_1 takes only two levels $\{0, 1\}$. The odds ratio between $X = 0$ and $X = 1$ in relation to the answer $Y = 2$, assuming the other covariates are fixed, is given by

$$\frac{P(Y = 2|X_1 = 1, X_2 = x_2, \ldots, Xk = x_k)}{P(Y = 2|X_1 = 0, X_2 = x_2, \ldots, Xk = x_k)} = \exp(\beta_{11}).$$

Similarly, the odds ratio between X_0 and X_1 in relation to the answer $Y = c$, assuming the other covariates are fixed, has the form

$$\frac{P(Y = c|X_1 = 1, X_2 = x_2, \ldots, Xk = x_k)}{P(Y = c|X_1 = 0, X_2 = x_2, \ldots, Xk = x_k)} = \exp(\beta_{c1}).$$

2.1. Maximum likelihood estimation

Consider n observations $(\widetilde{Y}_1, \mathbf{x}_1) \ldots, (\widetilde{Y}_n, \mathbf{x}_n)$ of a sample, where $\mathbf{x}_i = (x_{i1}, \ldots, x_{ik})^\top$ is the vector of covariates associated with the ith individual $(i = 1, \ldots, n)$ and $\widetilde{Y}_i$ is an independent random variable such that $\widetilde{Y}_i \sim MN(\mathbf{p}(\mathbf{x}_i), 1)$. The probability function in Equation (1) can be expressed as

$$f(\tilde{\mathbf{y}}_i) = \frac{1}{\prod_{j=1}^{c} \tilde{y}_{ij}!} \prod_{j=1}^{c} p_j(\mathbf{x}_i)^{\tilde{y}_{ij}}, \quad \tilde{y}_{ij} = 0, 1, \tag{4}$$

where $p_j(\mathbf{x}_i)$ is defined in Equations (2) and (3), for $i = 1, \ldots, n$ and $j = 1, \ldots, c$.

The likelihood function for the vector $\boldsymbol{\theta} = (\alpha_2, \ldots, \alpha_c, \boldsymbol{\beta}_2^\top, \ldots, \boldsymbol{\beta}_c^\top)^\top$ of model parameters follows from (2) and (3) as

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \left\{ \frac{1}{\prod_{j=1}^{c} \tilde{y}_{ij}!} \prod_{j=1}^{c} [p_j(\mathbf{x}_i)]^{\tilde{y}_{ij}} \right\}.$$

By removing the quantity that does not depend on $\boldsymbol{\theta}$, the likelihood function is

$$L(\boldsymbol{\theta}) \simeq \prod_{i=1}^{n} \prod_{j=1}^{c} [p_j(\mathbf{x}_i)]^{\tilde{y}_{ij}}.$$

By replacing the probabilities by the linear predictors, we obtain

$$L(\boldsymbol{\theta}) \simeq \prod_{i=1}^{n} \prod_{j=2}^{c} \left[\exp(\alpha_j + \boldsymbol{\beta}_j^\top \mathbf{x}_i) \right]^{\tilde{y}_{ij}} \left[\frac{1}{1 + \sum_{j=2}^{c} \exp(\alpha_j + \boldsymbol{\beta}_j^\top \mathbf{x}_i)} \right]. \tag{5}$$

Finally, the log-likelihood function for $\boldsymbol{\theta}$ has the form

$$l(\boldsymbol{\theta}) \simeq \sum_{i=1}^{n} \left\{ \left[\sum_{j=2}^{c} \tilde{y}_{ij} (\alpha_j + \boldsymbol{\beta}_j^\top \mathbf{x}_i) \right] - \log \left[1 + \sum_{j=2}^{c} \exp(\alpha_j + \boldsymbol{\beta}_j^\top \mathbf{x}_i) \right] \right\}. \tag{6}$$

Note that the number of parameters increases according to the number of categories that comprise the response variable. Thus, it is necessary for the sample size of each category to be reasonably large to obtain a reliable inference [29].

In this way, the MLE $\widehat{\boldsymbol{\theta}}$ can be calculated by maximizing (6). There are several optimization procedures such as the MaxBFGS routine in the matrix programming language Ox (see, [10]), which require only the original estimator function rather than their derivatives. Hypothesis tests and standard errors of the estimators in $\widehat{\boldsymbol{\theta}}$ can be based on the asymptotic normal approximation

$$\widehat{\boldsymbol{\theta}}^\top \sim \mathrm{N}_{((c-1)+(c-1)k)} \left\{ \boldsymbol{\theta}^\top, -\ddot{\mathbf{L}}^{-1}(\boldsymbol{\theta}) \right\},$$

where $-\ddot{\mathbf{L}}(\boldsymbol{\theta}) = \{\partial^2 l(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top\}$ is the $((c-1) + (c-1)k) \times ((c-1) + (c-1)k)$ observed information matrix.

2.2. Inference: bootstrap estimator

The bootstrap re-sampling method was proposed by Efron [11] and treats the observed sample as if it represents the population. From the information obtained from such a sample, B bootstrap samples of similar size to that of the observed sample are generated, from which it is possible to estimate various characteristics of the population, such as mean, variance, percentiles and so on.

The re-sampling method may be non-parametric or parametric. In this study, the non-parametric bootstrap method is considered according to which the distribution function F can be estimated by the empirical distribution $\hat{F}$.

Let $\mathbf{T} = (T_1, \ldots, T_n)$ be an observed random sample and $\hat{F}$ be the empirical distribution of $\mathbf{T}$. Thus, a bootstrap sample $\mathbf{T}^*$ is constructed by re-sampling with replacement of n elements from the sample $\mathbf{T}$. For B bootstrap samples generated, $T_1^*, \ldots, T_B^*$, the bootstrap replication of the parameter of interest for the bth sample is given by [11]

$$\hat{\boldsymbol{\theta}}_b^* = s(T_b^*),$$

that is, is the value of $\hat{\boldsymbol{\theta}}$ for sample T_b^*, $b = 1, \ldots, B$.

The bootstrap estimator of the standard error [12], say $\widehat{EP}_B$, is the standard deviation of these bootstrap samples, namely

$$\widehat{EP}_B = \left[\frac{1}{(B-1)} \sum_{b=1}^{B} \left(\hat{\theta}_b^* - \bar{\theta}_B \right)^2 \right]^{1/2},$$

where $\bar{\theta}_B = 1/B \sum_{b=1}^{B} \hat{\theta}_b^*$.

It is generally sufficient to present good results to determine the bootstrap estimates if $B \geq 200$ [12]. However, to achieve greater accuracy, a reasonably high B value must be considered. In this study, we set $B = 1000$ bootstrap samples. We describe the bias corrected and accelerated (BCa) techniques for constructing approximated confidence intervals based on the bootstrap re-sampling method. For further details on bootstrap intervals, see, for example, [7,9,12].

The bootstrap interval based on the BCa method assumes that the percentiles used in delimiting the bootstrap confidence intervals depend on the corrections for tendency $\hat{a}$ and acceleration $\hat{z}_0$.

The bias correction value $\hat{z}_0$ is generated based on the proportion of estimations of bootstrap samples that are smaller than the original estimate $\hat{\boldsymbol{\theta}}$. It is given by

$$\hat{z}_0 = \Phi^{-1} \left(\frac{\sharp(\hat{\boldsymbol{\theta}}_b^* < \hat{\boldsymbol{\theta}})}{B} \right), \quad b = 1, \ldots, B,$$

where $\Phi^{-1}(\cdot)$ is the standard normal quantile function, B is the number of generated bootstrap samples, $\hat{\boldsymbol{\theta}}$ is the MLE of the observed sample and $\hat{\boldsymbol{\theta}}_b^*$ is the MLE of the bth bootstrap

sample. Let $\hat{\boldsymbol{\theta}}_{(i)}$ be the MLE of $\boldsymbol{\theta}$ without the ith observation. We have

$$\hat{a} = \frac{\sum_{i=1}^{n}\left[\hat{\boldsymbol{\theta}}_{(\cdot)} - \hat{\boldsymbol{\theta}}_{(i)}\right]^3}{6\left\{\sum_{i=1}^{n}\left[\hat{\boldsymbol{\theta}}_{(\cdot)} - \hat{\boldsymbol{\theta}}_{(i)}\right]^2\right\}^{3/2}}.$$

Note that $\hat{\boldsymbol{\theta}}_{(\cdot)} = \sum_{i=1}^{n}\hat{\boldsymbol{\theta}}_{(i)}/n$ and n is the sample size.

Hence, the BCa bootstrap interval of coverage $100(1-2\alpha)\%$ has the form

$$\left[\hat{\boldsymbol{\theta}}^*_{(B\alpha_1)}, \hat{\boldsymbol{\theta}}^*_{(B\alpha_2)}\right],$$

where

$$\alpha_1 = \Phi\left\{\hat{z}_0 + \frac{\hat{z}_0 + \Phi^{-1}(\alpha)}{1 - \hat{a}\left[\hat{z}_0 + \Phi^{-1}(\alpha)\right]}\right\} \quad \text{and}$$

$$\alpha_2 = \Phi\left\{\hat{z}_0 + \frac{\hat{z}_0 + \Phi^{-1}(1-\alpha)}{1 - \hat{a}\left[\hat{z}_0 + \Phi^{-1}(1-\alpha)\right]}\right\}.$$

Here, the quantities α_1 and α_2 are corrections to the bootstrap percentiles.

2.3. Inference: Bayesian analysis

The Bayesian method allows to incorporate previous knowledge of the parameters through an informative prior distribution. If there is no background knowledge for the parameters, we can consider a non-informative prior structure. Consider that $\boldsymbol{\theta}$ has a prior distribution, say $\pi(\boldsymbol{\theta})$. In addition, assuming independent distributions, we define

$$\pi(\boldsymbol{\theta}) = \prod_{j=2}^{c}\left[\pi(\alpha_j)\prod_{v=1}^{k}\pi(\beta_{vj})\right],$$

where $\alpha_j \sim N(0, 10^6)$ for $j = 2, \ldots, c$ and $\beta_{vj} \sim N(0, 10^6)$ for $v = 1, \ldots, k$. Thus, all hyper-parameters are specified to express non-informative priors, since there is no previous information about these parameters.

Let $\mathcal{D} = (\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_n)$ be a vector of n observations. The joint posterior of $\boldsymbol{\theta}$ has the form

$$\pi(\boldsymbol{\theta}|\mathcal{D}) \propto L(\boldsymbol{\theta})\pi(\boldsymbol{\theta}), \tag{7}$$

where $L(\boldsymbol{\theta})$ is the likelihood function (5).

By using $\pi(\boldsymbol{\theta}|\mathcal{D})$ in (7) and the MCMC method, we can estimate the parameters of the MLRM under the Bayesian approach. In this case, we use the `R` package `rjags` [23,24].

3. Checking model

3.1. Influence diagnostics: classic analysis

After fitting the MLRM, it is important to check the model assumptions as well as to conduct a robustness study in order to detect influential or extreme observations that can

cause distortions in the results of the analysis. Numerous approaches have been proposed in the literature to detect influential or outlying observations. For example, Russo *et al.* [26] studied influence diagnostics in autoregressive nonlinear elliptical models and Ibacache-Pulgar *et al.* [16] adopted local influence in elliptical multivariate regression models. Ortega *et al.* [21] developed influence diagnostics for the power series beta Weibull regression model for predicting breast carcinoma with censored data. Cruz *et al.* [6] derived curvature calculations under various perturbation schemes in the log-odd log-logistic Weibull regression model. Lachos *et al.* [19] performed influence diagnostics in spatial models with censored response and Schuma *et al.* [27] investigated local influence for censored regression models with autoregressive errors. We develop a similar methodology to detect influential subjects in the MLRM.

A first tool to perform sensitivity analysis, as stated before, is by means of global influence starting from case-deletion, which is a common approach to study the effect of dropping the *i*th observation from a data set. The case-deletion MLRM can be expressed as

$$\left[\exp(\alpha_j+\boldsymbol{\beta}_j^\top\mathbf{x}_{(i)})\right]^{\tilde{y}_{(i)j}}\left[\frac{1}{1+\sum_{j=2}^{c}\exp(\alpha_j+\boldsymbol{\beta}_j^\top\mathbf{x}_{(i)})}\right], \tag{8}$$

where $i=1,\ldots,n$, $j=2,\ldots,c$ and a quantity with subscript '(i)' means the original quantity without the ith observation. For model (8), the log-likelihood function of $\boldsymbol{\theta}$ is denoted by $l_{(i)}(\boldsymbol{\theta})$. Let $\hat{\boldsymbol{\theta}}_{(i)}=(\hat{\alpha}_{2(i)},\ldots,\hat{\alpha}_{c(i)},\widehat{\boldsymbol{\beta}}_{2(i)}^\top,\ldots,\widehat{\boldsymbol{\beta}}_{c(i)}^\top)^\top$ be the MLE of $\boldsymbol{\theta}$ from maximizing $l_{(i)}(\boldsymbol{\theta})$. To assess the influence of the ith observation on the MLE $\hat{\boldsymbol{\theta}}=(\hat{\alpha}_2,\ldots,\hat{\alpha}_c,\widehat{\boldsymbol{\beta}}_2^\top,\ldots,\widehat{\boldsymbol{\beta}}_c^\top)^\top$, the basic idea is to compare the difference between $\hat{\boldsymbol{\theta}}_{(i)}$ and $\hat{\boldsymbol{\theta}}$. If deletion of an observation seriously influences the estimates, more attention should be paid to that observation. Hence, if $\hat{\boldsymbol{\theta}}_{(i)}$ is far from $\hat{\boldsymbol{\theta}}$, then the ith observation is regarded as influential. A first measure of the global influence is defined as the standardized norm of $\hat{\boldsymbol{\theta}}-\hat{\boldsymbol{\theta}}_{(i)}$ (generalized Cook distance) giving by

$$GD_i(\boldsymbol{\theta})=(\hat{\boldsymbol{\theta}}-\hat{\boldsymbol{\theta}}_{(i)})^\top\big[-\ddot{\mathbf{L}}(\boldsymbol{\theta})\big](\hat{\boldsymbol{\theta}}-\hat{\boldsymbol{\theta}}_{(i)}).$$

A second common measure of the difference between $\hat{\boldsymbol{\theta}}_{(i)}$ and $\hat{\boldsymbol{\theta}}$ is the likelihood distance

$$LD_i(\boldsymbol{\theta})=2\Big\{l(\hat{\boldsymbol{\theta}})-l(\hat{\boldsymbol{\theta}}_{(i)})\Big\}.$$

A different approach for influence diagnostics was suggested by Cook [4] to give some weights to the observations instead of removing them. Local influence calculation can be carried out for model (2). If likelihood displacement $LD(\boldsymbol{\omega})=2\{l(\hat{\boldsymbol{\theta}})-l(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}})\}$ is utilized, where $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ denotes the MLE under the perturbed model, the normal curvature for $\boldsymbol{\theta}$ at the direction $\mathbf{d}$, $\|\,\mathbf{d}\,\|=1$, is given by $C_{\mathbf{d}}(\boldsymbol{\theta})=2|\mathbf{d}^\top\boldsymbol{\Delta}^\top[\ddot{\mathbf{L}}(\boldsymbol{\theta})]^{-1}\boldsymbol{\Delta}\mathbf{d}|$. Here, $\boldsymbol{\Delta}$ is a $((c-1)+(c-1)k)\times n$ matrix depending on the perturbation scheme, whose elements are $\Delta_{vi}=\partial^2 l(\boldsymbol{\theta}|\boldsymbol{\omega})/\partial\phi_v\partial\omega_i$, $i=1,\ldots,n$ and $v=1,2,\ldots,(c-1)+(c-1)k$, evaluated at $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega}_0$, where $\boldsymbol{\omega}_0$ is the no perturbation vector. For the MLRM, the elements of $\ddot{\mathbf{L}}(\boldsymbol{\theta})$ can be obtained numerically. We can also calculate normal curvatures $C_{\mathbf{d}}(\alpha_2),\ldots,C_{\mathbf{d}}(\alpha_c)$ and $C_{\mathbf{d}}(\boldsymbol{\beta}_2),\ldots,C_{\mathbf{d}}(\boldsymbol{\beta}_c)$ to perform various index plots. For instance, the index plot of $\mathbf{d}_{max}$, the eigenvector corresponding to $C_{\mathbf{d}_{max}}$, the largest eigenvalue of the matrix

$\mathbf{B} = -\boldsymbol{\Delta}^\top[\ddot{\mathbf{L}}(\boldsymbol{\theta})]^{-1}\boldsymbol{\Delta}$. The index plots of $C_{\mathbf{d}_i}(\alpha_2), \ldots, C_{\mathbf{d}_i}(\alpha_c)$ and $C_{\mathbf{d}_i}(\boldsymbol{\beta}_2), \ldots, C_{\mathbf{d}_i}(\boldsymbol{\beta}_c)$ are called the total local influence, where $\mathbf{d}_i$ denotes an $n \times 1$ vector of zeros with one at the ith position. Thus, the curvature at direction $\mathbf{d}_i$ becomes $C_i = 2|\boldsymbol{\Delta}_i^\top[\ddot{\mathbf{L}}(\boldsymbol{\theta})]^{-1}\boldsymbol{\Delta}_i|$, where $\boldsymbol{\Delta}_i^\top$ denotes the ith row of $\boldsymbol{\Delta}$. It is common to point out those observations such that $C_i \geq 2\bar{C}$, where $\bar{C} = \frac{1}{n}\sum_{i=1}^n C_i$.

Further, we calculate the matrix $\boldsymbol{\Delta}$ for two perturbation schemes from model (2) and its log-likelihood function (6), namely

$$\boldsymbol{\Delta} = (\boldsymbol{\Delta}_{vi})_{\left[((c-1)+(c-1)k)\times n\right]} = \left(\frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\theta}_v \boldsymbol{\omega}_i}\right)_{\left[((c-1)+(c-1)k)\times n\right]},$$

where $v = 1, 2, \ldots, (c-1)+(c-1)k$ and $i = 1, \ldots, n$.

First, we consider a case-weight perturbation which modifies the weight given to each subject in the log-likelihood. Let $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$ be a vector of weights. In this case, the log-likelihood function has the form

$$l(\boldsymbol{\theta}|\boldsymbol{\omega}) \simeq \sum_{i=1}^n \left\{ \omega_i \left[\sum_{j=2}^c \tilde{y}_{ij}(\alpha_j + \boldsymbol{\beta}_j^\top \mathbf{x}_i) \right] - \omega_i \log\left[1 + \sum_{j=2}^c \exp(\alpha_j + \boldsymbol{\beta}_j^\top \mathbf{x}_i) \right] \right\},$$

where $0 \leq \omega_i \leq 1$, $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$. The matrix $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_{\alpha_2}, \ldots, \boldsymbol{\Delta}_{\alpha_c}, \boldsymbol{\Delta}_{\boldsymbol{\beta}_2}^\top, \ldots, \boldsymbol{\Delta}_{\boldsymbol{\beta}_c}^\top)^\top$ is obtained numerically.

Second, we consider an additive perturbation on a particular continuous explanatory variable, namely x_t, by setting $x_{it\omega} = x_{it} + \omega_i S_x$, where S_x is a scaled factor, $\omega_i \in \mathbf{R}$. [4] described a general scheme for perturbing the whole design matrix X in the linear regression model. This perturbation has a more complicated impact on the estimate. For the MLRM, this approach leads to the perturbed log-likelihood function

$$l(\boldsymbol{\theta}|\boldsymbol{\omega}) \simeq \sum_{i=1}^n \left\{ \left[\sum_{j=2}^c \tilde{y}_{ij}(\alpha_j + \boldsymbol{\beta}_j^\top \mathbf{x}_i^*) \right] - \log\left[1 + \sum_{j=2}^c \exp(\alpha_j + \boldsymbol{\beta}_j^\top \mathbf{x}_i^*) \right] \right\},$$

where $(\boldsymbol{\beta}_j^\top \mathbf{x}_i^*) = \beta_{j1}x_{i1} + \ldots + \beta_{jt}(x_{it} + \omega_i S_x) + \ldots + \beta_{jk}x_{ik}$ and $\boldsymbol{\omega}_0 = (0, \ldots, 0)^\top$. The matrix $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_{\alpha_2}, \ldots, \boldsymbol{\Delta}_{\alpha_c}, \boldsymbol{\Delta}_{\boldsymbol{\beta}_2}^\top, \ldots, \boldsymbol{\Delta}_{\boldsymbol{\beta}_c}^\top)^\top$ can be obtained numerically.

3.2. Influence diagnostics: Bayesian analysis

From the Bayesian point of view, the case-deletion influence analysis is based on ψ-divergence [22,30] to determine possible influential observations. In this case, we consider $\mathcal{D}^{(-i)}$ as the data without the ith observation. Then, let $D_\psi(P; P_{(-i)})$ be the ψ-divergence between P and $P_{(-i)}$, where P denotes the posterior distribution of $\boldsymbol{\theta}$ for the complete data and $P_{(-i)}$ the posterior distribution without the ith observation. The ψ-divergence measure is defined by

$$D_\psi(P, P_{(-i)}) = \int_{\boldsymbol{\theta}\in\Theta} \psi\left(\frac{\pi(\boldsymbol{\theta}|\mathcal{D}^{(-i)})}{\pi(\boldsymbol{\theta}|\mathcal{D})}\right) \pi(\boldsymbol{\theta}|\mathcal{D})\, d\boldsymbol{\theta},$$

where $\psi(\cdot)$ is a convex function with $\psi(1) = 0$ and Θ is the parameter space. Several choices for ψ are given in [8]. For example,

- the Kullback–Leibler (K–L) divergence corresponds to $\psi(z) = -\log(z)$,
- the J-distance (or the symmetric version of the K–L divergence) is defined by $\psi(z) = (z-1)\log(z)$,
- the L_1 norm (or distance variational) follows when $\psi(z) = 0.5|z-1|$.

Further, we consider $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(Q)}$ as being a sample of size Q of $\pi(\boldsymbol{\theta}|D)$. Then, $D_\psi(P, P_{(-i)})$ can be calculated numerically by

$$\widehat{D}_\psi(P, P_{(-i)}) = \frac{1}{Q}\sum_{q=1}^{Q} \psi\left(\frac{\widehat{CPO}_i}{f(\tilde{\mathbf{y}}_i|\boldsymbol{\theta}^{(q)})}\right),$$

where

$$\widehat{CPO}_i = \left\{\frac{1}{Q}\sum_{q=1}^{Q}\frac{1}{f(\tilde{\mathbf{y}}_i|\boldsymbol{\theta}^{(q)})}\right\}^{-1}$$

is the numerical approximation of the conditional predictive ordinate statistic of the ith observation and $f(\tilde{\mathbf{y}}_i|\boldsymbol{\theta}^{(q)})$ is the probability function (4) of the ith observation [17].

The quantity $D_\psi(P, P_{(-i)})$ can be interpreted as the ψ-divergence of the effect of deleting the ith observation from the full data on the joint posterior distribution of $\boldsymbol{\theta}$. For example, the K–L divergence is given by

$$\begin{aligned} D_{\mathrm{K--L}}(P, P_{(-i)}) &= -E_{\boldsymbol{\theta}|\mathcal{D}}\left\{\log(CPO_i)\right\} + E_{\boldsymbol{\theta}|\mathcal{D}}\left\{\log\left[f(\tilde{\mathbf{y}}_i|\boldsymbol{\theta})\right]\right\} \\ &= -\log(CPO_i) + E_{\boldsymbol{\theta}|\mathcal{D}}\left\{\log\left[f(\tilde{\mathbf{y}}_i|\boldsymbol{\theta})\right]\right\}. \end{aligned} \tag{9}$$

Hence, we obtain the estimate of the divergence measure in (9) as follows

$$\widehat{D}_{\mathrm{K--L}}(P, P_{(-i)}) = -\log(\widehat{CPO}_i) + \frac{1}{Q}\sum_{q=1}^{Q}\log[f(\tilde{\mathbf{y}}_i|\boldsymbol{\theta}^{(q)})].$$

In addition, according to Peng and Dey [22] and Weiss [30], a cutoff for the value of the divergence measure can be determined as follows. Considering a biased coin with success probability p, the ψ-divergence between the biased and unbiased coins is

$$D_\psi(g_0, g_1) = \int \psi\left(\frac{g_0(x)}{g_1(x)}\right) g_1(x)\,\mathrm{d}x,$$

where $g_0(x) = p^x(1-p)^{1-x}$ and $g_1(x) = 0.5$, for $x = 0$ or 1. So, if $D_\psi(g_0, g_1) = d_\psi(p)$, the quantity d_ψ is given by

$$d_\psi(p) = \frac{\psi(2p) + \psi(2(1-p))}{2}. \tag{10}$$

In this way, we note that $d_\psi(p)$ in (10) increases when p departs from 0.50, and yet, $d_\psi(p)$ is symmetric around $p = 0.50$ and reaches its minimum at $p = 0.50$. At this point, $d_\psi(0.5) = 0$ and $g_0 = g_1$.

Hence, if we take $p > 0.90$ (or $p < 0.20$) to represent a strong bias in a coin, $d_{L_1}(0.90) = 0.40$, and the ith observation can be considered influential when $d_{L_1} > 0.40$. Analogously,

for the K–L divergence, we may consider an influential observation when $d_{K-L} > 0.51$. On the other hand, if the J-distance is used, an observation can be considered influential when $d_J > 0.88$.

4. Application

The liver is one of the largest organs in the human body and of vital importance for the digestive system. Moreover, the liver is responsible for a series of metabolic functions. Some diseases can cause hepatic insufficiency leading to death of the organ and consequently the need for a transplant. Terminal hepatic disease can be divided into two types: acute and chronic. The acute type is caused by medicines and the Hepatitis A virus. In turn, for the chronic type, the most common causes are: hepatic lesions caused by drinking alcohol, Hepatitis B or C virus, and non-alcoholic steatohepatitis (NASH), among others. Liver transplant is indicated for patients suffering from hepatic cirrhosis, having a life expectancy lower than 20% for the next 12 months, congenital diseases like biliary atresia, Wilson's disease, paramyloidosis, and cholangitis diseases such as primary sclerosing cholangitis. We illustrate the potentiality of the proposed model by using a part of the data set obtained from a study conducted by Kelly *et al.* [18]. These data come from a follow-up study of 772 patients who underwent orthotopic liver transplant between January 2004 and April 2010 at the Cleveland Clinic. After the surgery, 43 patients died, 213 remained hospitalized and 516 were released from the hospital. The authors indicated that the mortality rate was very low, and the occurrence of deaths was more closely related with the clinical conditions of the patient before surgery than the operation itself.

The response variable is then defined as $y = 1$ (Facility), $y = 2$ (Expired) and $y = 3$ (Home). We consider the following explanatory variables to find possible factors to predict the probabilities in these three categories for a given patient just after surgery:

- x_1 : Karnofsky score (%)
- x_2 : Age (years);
- x_3 : International normalized ratio (INR);
- x_4 : Creatinine (mg/dl);
- x_5 : Diabete mellitus (No = 0 and Yes = 1);
- x_6 : Bilirubin (mg/dl);
- x_7 : Albumin (g/dl);
- x_8 : Body mass index (kg/m^2);
- x_9 : Renal dialysis (No = 0 and Yes = 1);
- x_{10} : Gender (Female = 0 and Male = 1).

We use the MaxBFGS subroutine of the matrix programming language `Ox 7.10` to fit the MLRM to the current data. The initial values for all parameters are set to 0.5. We adopt the stepwise method to select among these ten explanatory variables, the most important ones, and obtain $x_1, x_2, x_3, x_5, x_7, x_9$ and x_{10}. On the other hand, we simulate two chains of size 300,000 for each parameter, disregarding the first 60,000 to eliminate the possible effects of the initial values and avoid autocorrelation problems, to obtain the Bayesian estimates. We consider a spacing of size 100, thus obtaining an effective sample size of 4, 800 on

which the posterior inference is based. In addition, the MCMC convergence is monitored according to the methods recommended by Cowles and Calin [5].

In Table 1, we list the MLEs and non-parametric bootstrap estimates of the parameters with their standard errors (SEs) and confidence intervals. Similarly, we give the posterior summaries (Mean, Standard Deviation (SD) and 95% Credible Interval (C.I.)) for the parameters of interest in Table 2. We note from the figures in Tables 1 and 2 that the estimates of the model parameters obtained by the three estimation methods are somewhat close, since their values belong to the confidence intervals. Further, the estimates from the non-parametric bootstrap method present larger SEs.

Based on the figures in Tables 1 and 2, we can note that the results of the point estimation of the three methodologies are similar, but the classical approach has lower standard errors,

Table 1. MLEs and non-parametric bootstrap estimates from the fitted MLRM to the liver transplantation data.

	MLE				Non-parametric bootstrap		
Parameter	Estimate	SE	*p*-Value	95% C.I.	Estimate	SE	95% Bca interval
α_2	0.1916	1.3034	0.8831	(−2.3630, 2.7462)	0.4871	1.8757	(−2.7975, 3.1362)
α_3	0.1070	0.8326	0.8977	(−1.5248,1.7389)	0.0734	0.8987	(−1.6479, 1.1503)
β_{12}	0.0017	0.0086	0.8437	(−0.0151, 0.0185)	0.00002	0.0105	(−0.0144, 0.0205)
β_{13}	0.0468	0.0058	< 0.0001	(0.0355, 0.0581)	0.0473	0.0061	(0.0368, 0.0568)
β_{22}	−0.0321	0.0174	0.0644	(−0.0662, 0.0019)	−0.0317	0.0192	(−0.0622, −0.0004)
β_{23}	−0.0423	0.0102	< 0.0001	(−0.0624, −0.0223)	−0.0436	0.0111	(−0.0600, −0.0243)
β_{32}	0.0760	0.1957	0.6976	(−0.3075, 0.4595)	−0.0548	0.4244	(−0.7305, 0.5034)
β_{33}	−0.3254	0.1622	0.0448	(−0.6432, −0.0075)	−0.3970	0.2198	(−0.6564, −0.0116)
β_{52}	−0.3896	0.3732	0.2965	(−1.1212, 0.3419)	−0.4149	0.0396	(−1.0552, 0.2219)
β_{53}	−0.5172	0.1969	0.0086	(−0.9032, −0.1312)	−0.5301	0.2103	(−0.8730, -0.1864)
β_{72}	−0.1499	0.2695	0.5781	(−0.6781, 0.3783)	−0.1855	0.3768	(−0.6970, 0.5243)
β_{73}	0.3091	0.1420	0.0295	(0.0308, 0.5875)	0.3034	0.1479	(0.0828, 0.5738)
β_{92}	−1.0684	0.5720	0.0618	(−2.1896, 0.0528)	−1.3646	2.3143	(−2.3749, −0.1401)
β_{93}	−0.9442	0.2861	0.0010	(−1.5048, −0.3835)	−0.9423	0.2904	(−1.4126, −0.4450)
β_{102}	0.6287	0.3899	0.1068	(−0.1354, 1.3928)	0.6458	0.4147	(−0.0226, 1.3154)
β_{103}	0.3460	0.1977	0.0800	(−0.0414, 0.7335)	0.3405	0.2066	(0.0376, 0.7226)

Table 2. Bayesian estimates from the fitted MLRM to liver transplantation data.

Parameter	Mean	SD	95% C.I.
α_2	0.6537	1.5503	(−2.4025, 3.6750)
α_3	0.4977	0.9546	(−1.3813, 2.3600)
β_{12}	0.0018	0.0086	(−0.0149, 0.0189)
β_{13}	0.0476	0.0058	(0.0365, 0.0591)
β_{22}	−0.0321	0.0176	(−0.0666, 0.0038)
β_{23}	−0.0432	0.0103	(−0.0633, −0.0225)
β_{32}	0.0487	0.2116	(−0.4027, 0.4325)
β_{33}	−0.3396	0.1688	(−0.6759, −0.0184)
β_{52}	−0.4238	0.3793	(−1.1990, 0.2995)
β_{53}	−0.5233	0.1954	(−0.9045, −0.1482)
β_{72}	−0.1611	0.2711	(−0.7031, 0.3547)
β_{73}	0.3156	0.1402	(0.0363, 0.5906)
β_{92}	−1.1590	0.6045	(−2.4562, −0.0859)
β_{93}	−0.9458	0.2908	(−1.5313, −0.3702)
β_{102}	0.6739	0.3923	(−0.0584, 1.4460)
β_{103}	0.3499	0.2011	(−0.0499, 0.7360)

i.e. it is the most efficient method among the three. So, we decide to continue the analysis using the classical approach.

Next, we analyze the sensitivity of the fitted regression model to extreme values or possible influential points as described in Section 3. The plots of the global influence measures ($GD_i(\boldsymbol{\theta})$ and $LD_i(\boldsymbol{\theta})$) are displayed in Figure 1. We can note that the case ♯688 is a possible influential observation.

By using the local influence measures described in Section 3 for case-weight perturbation, the value $C_{\mathbf{d}_{max}} = 2.1760$ is calculated as a maximum curvature. We provide the index plots of $\mathbf{d}_{max}(\boldsymbol{\theta})$ in Figure 2(a) and the index plots of $C_i(\boldsymbol{\theta})$ for all points in Figure 2(b). The most influential cases on $\widehat{\boldsymbol{\theta}}$ are the observations ♯243, ♯589 and ♯688.

Further, we examine the influence of perturbations on the KS, Age, INR and Albumin variables. The value of the maximum curvature is $C_{\mathbf{d}_{max}}(\boldsymbol{\theta}) = 1.4947$ for the KS variable, 1.2451 for the Age variable, 1.3406 for the INR variable and 1.2788 for the Albumin variable. We display the plots of $\mathbf{d}_{max}(\boldsymbol{\theta})$ and $C_i(\boldsymbol{\theta})$ in Figure 3(a ,b), respectively, for the KS variable. We display the plots os these two quantities for the Age, INR and Albumin

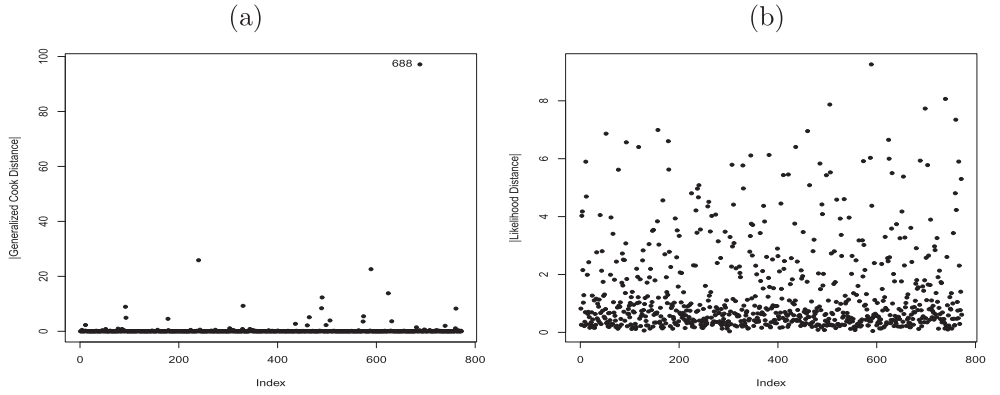

Figure 1. (a) Index plot of $GD_i(\boldsymbol{\theta})$ for $\boldsymbol{\theta}$ on the liver transplantation data. (b) Index plot of $LD_i(\boldsymbol{\theta})$ for $\boldsymbol{\theta}$ on the liver transplantation data.

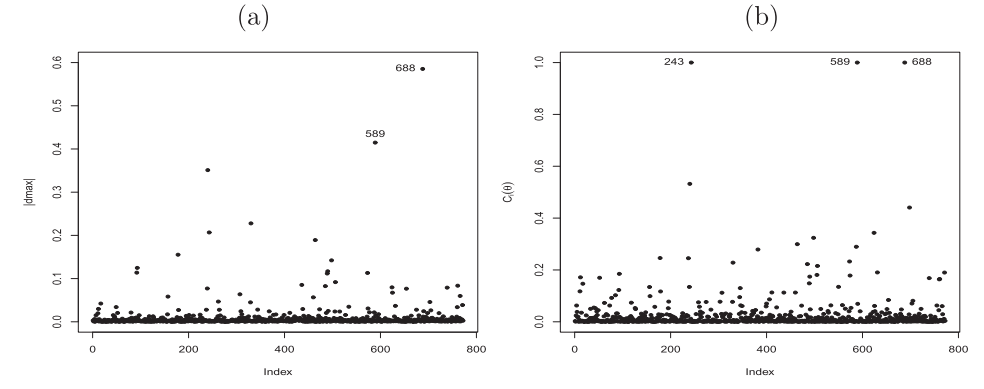

Figure 2. (a) Index plots of $\mathbf{d}_{max}$ under case-weight perturbation scheme. (b) Index plots of local total for $C_i(\boldsymbol{\theta})$ under case-weight perturbation scheme.

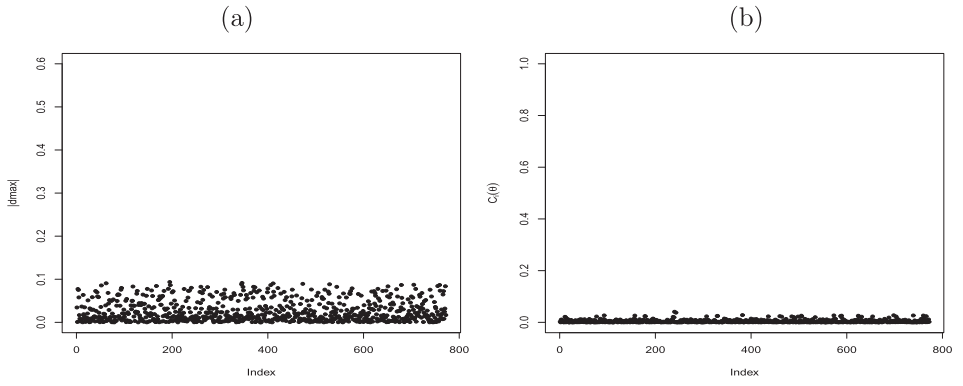

Figure 3. (a) Index plots of $\mathbf{d}_{max}$ under *Karnosfky score* variable perturbation scheme. (b) Index plots of local total for $C_i(\boldsymbol{\theta})$ under *Karnosfky score* variable perturbation scheme.

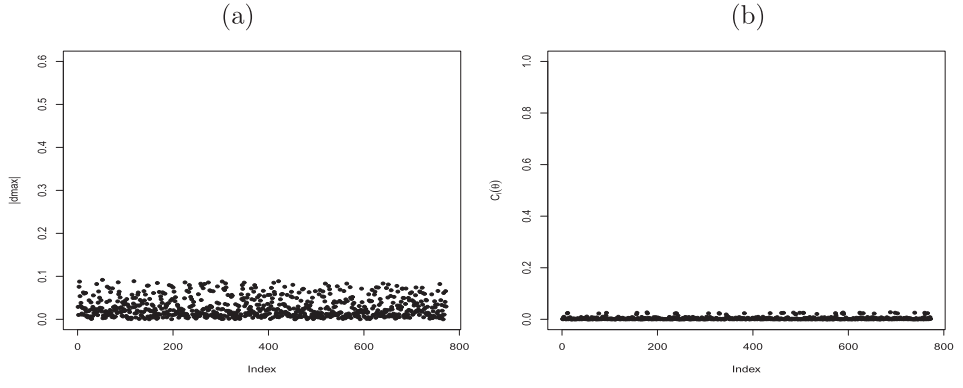

Figure 4. (a) Index plots of $\mathbf{d}_{max}$ under *Age* variable perturbation scheme. (b) Index plots of local total for $C_i(\boldsymbol{\theta})$ under *Age* variable perturbation scheme.

variables in Figures 4, 5, and 6, respectively. These plots suggest that there is no influential observation.

In order to detect possible influential observations in the posterior distribution of the parameters of the model, we present in Figure 7 the index plots of the three ψ-divergence measures obtained in the posterior sample of the parameters of the model. In Figure 7(a), we note that for the K–L measure, the cases ♯243, ♯589 and ♯688 are possible influential observations in the posterior distribution. For the estimates of the J-distance measure shown in Figure 7(b), the cases ♯243, ♯589 and ♯688 are detected as influential observations, whereas for the measure L_1 norm shown in Figure 7(c), the influential cases are ♯243, ♯589 and ♯688.

In Table 3, we present the estimates of the three divergences measures for the observations that are detected as influential (highlighted in bold) by at least one measure.

Hence, the diagnostic analysis indicates four observations (♯243, ♯589 and ♯688) as potentially influential. Based on the global influence, local influence and ψ-divergence measures, we note that the observations ♯243, ♯589 and ♯688 can be chosen as possible influential points. They have the following characteristics:

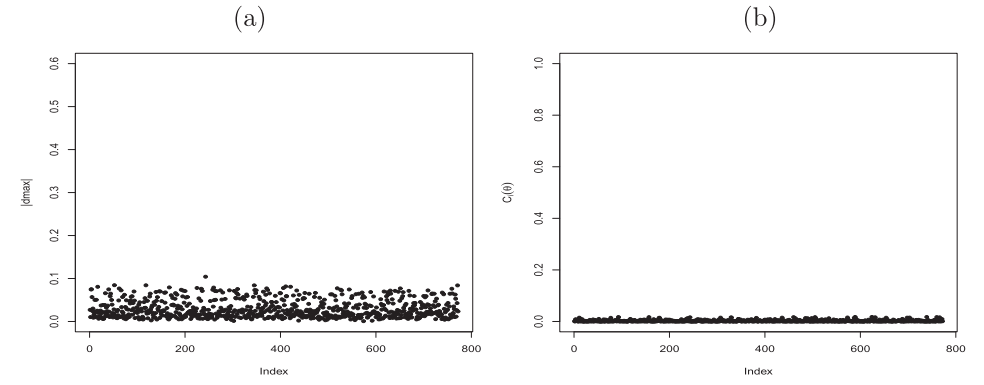

Figure 5. (a) Index plots of $\mathbf{d}_{max}$ under *INR* variable perturbation scheme. (b) Index plots of local total for $C_i(\boldsymbol{\theta})$ under *INR* variable perturbation scheme.

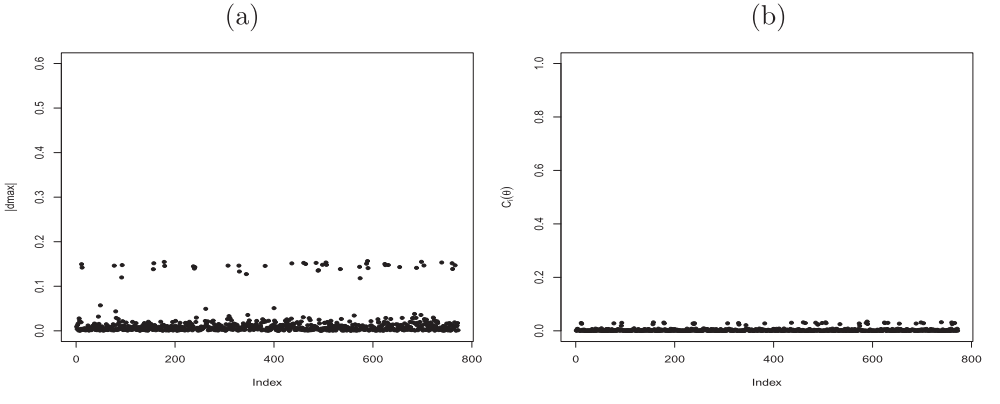

Figure 6. (a) Index plots of $\mathbf{d}_{max}$ under *Albumin* variable perturbation scheme. (b) Index plots of local total for $C_i(\boldsymbol{\theta})$ under *Albumin* variable perturbation scheme.

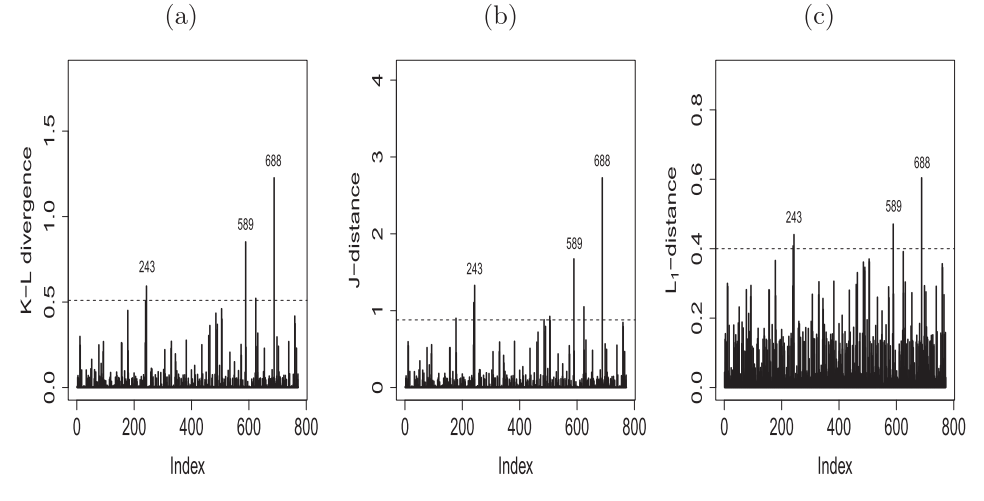

Figure 7. Divergence measure: (a) K–L divergence. (b) *J*-distance. (c) L_1 distance.

Table 3. Bayesian diagnostics.

Observation	d_{K-L}	d_J	d_{L_1}
243	**0.5940**	**1.3302**	**0.4409**
589	**0.8525**	**1.6749**	**0.4711**
688	**1.2272**	**2.7296**	**0.6044**

- Patient ♯243 (Home): female with age of 39 years, without history of diabetes mellitus and without use of renal dialysis, as well as Karnofsky score of 70%, albumin level equal to 2.8 g/dl and INR of 9.0;
- Patient ♯589 (Expired): male with age of 50 years, without history of diabetes mellitus and without use of renal dialysis, as well as Karnofsky score of 60%, albumin level equal to 6.6 g/dl and INR of 8.6;
- Patient ♯688 (Expired): male with age of 56 years, with history of diabetes mellitus and without use of renal dialysis, as well as Karnofsky score of 20%, albumin level equal to 2.9 g/dl and INR of 8.6.

In order to reveal the impact of these four observations on the parameter estimates, we refit the model by eliminating individually each of these four cases. In Tables 4 and 5, we show the relative changes (in percentages) of the parameter estimates defined by $\mathbf{RC}_{\theta_j} = [(\hat{\theta}_j - \hat{\theta}_{j(I)})/\hat{\theta}_j] \times 100$, parameter estimates and the corresponding p-values, where $\hat{\alpha}_{j(I)}$ denotes the MLE of θ_j after the set 'I' of observations being removed.

Tables 4 and 5 reveal that there are no significant changes. The greatest change occurs in the model intercept, but there is no significant change in the covariates of the regression model. This indicates that the fitted MLRM is robust to possible influential points. Hence, the estimated probabilities are obtained from the following equations:

- Expired:

$$p_2(\mathbf{x}_i) = \left(\frac{1}{fx}\right) \exp\{0.1916 + 0.0017x_{1i} - 0.0321x_{2i} + 0.0760x_{3i}\} \times \exp\{-0.3896x_{5i} - 0.1499x_{7i} - 1.0684x_{9i} + 0.6287x_{10i}\}.$$

Table 4. Relative changes [-**RC**-in %], estimates and the corresponding p-values in parentheses for the regression coefficients to explain the categorical data.

Set{I}	β_{01}	β_{02}	β_{11}	β_{12}	β_{21}	β_{22}	β_{31}	β_{32}
A	–	–	–	–	–	–	–	
	0.1916	−0.1070	0.0017	0.0468	−0.0321	−0.0423	0.0760	−0.3254
	(0.8831)	(0.8977)	(0.8437)	(< 0.0001)	(0.0644)	(< 0.0001)	(0.6976)	(0.0448)
A–{♯243}	[54]	[380]	[−35]	[3]	[0]	[1]	[−47]	[−52]
	0.0886	0.2994	0.0023	0.0453	−0.0321	-0.0420	0.1117	−0.4962
	(0.9466)	(0.7290)	(0.7980)	(< 0.0001)	(0.0633)	(< 0.0001)	(0.6172)	(0.0096)
A-{♯589}	[−287]	[−66]	[−29]	[0]	[3]	[0]	[8]	[2]
	0.7420	−0.1774	0.0022	0.0468	−0.0310	−0.0425	0.0699	−0.3174
	(0.5743)	(0.8317)	(0.7976)	(< 0.0001)	(0.0774)	(< 0.0001)	(0.7286)	(0.0489)
A–{♯688}	[−593]	[182]	[76]	[0]	[−14]	[−2]	[503]	[−18]
	1.3286	0.0881	0.0004	0.0466	−0.0365	−0.0432	−0.3065	−0.3852
	(0.3639)	(0.9162)	(0.9639)	(< 0.0001)	(0.0379)	(< 0.0001)	(0.3251)	(0.0176)

Table 5. Relative changes [-**RC**-in %], estimates and the corresponding *p*-values in parentheses for the regression coefficients to explain the the categorical data.

Set{I}	β_{51}	β_{52}	β_{71}	β_{72}	β_{91}	β_{92}	β_{101}	β_{102}
A	–	–	–	–	–	–	–	
	−0.3896	−0.5172	−0.1499	0.3091	−1.0684	−0.9442	0.6287	0.3460
	(0.2965)	(0.0086)	(0.5781)	(0.0295)	(0.0618)	(0.0010)	(0.1068)	(0.0800)
A–{♯243}	[0]	[−4]	[3]	[8]	[−1]	[1]	[1]	[−3]
	−0.3876	−0.5359	−0.1459	0.2853	−1.0747	−0.9357	0.6214	0.3567
	(0.2987)	(0.0067)	(0.5877)	(0.0450)	(0.0606)	(0.0011)	(0.1093)	(0.0716)
A–{♯589}	[9]	[0]	[−165]	[−8]	[4]	[−1]	[1]	[−1]
	−0.3538	−0.5184	−0.3969	0.3331	−1.0279	−0.9493	0.6224	0.3499
	(0.3459)	(0.0085)	(0.1587)	(0.0208)	(0.0730)	(0.0009)	(0.1123)	(0.0768)
A–{♯688}	[−38]	[−5]	[−50]	[4]	[8]	[1]	[−7]	[−2]
	−0.5360	−0.5408	−0.2247	0.2963	−0.9805	−0.9390	0.6696	0.3538
	(0.1624)	(0.0062)	(0.4133)	(0.0371)	(0.0865)	(0.0011)	(0.0927)	(0.0739)

- Home:

$$p_3(\mathbf{x}_i) = \left(\frac{1}{fx}\right) \exp\{-0.1070 + 0.0468x_{1i} - 0.0423x_{2i} - 0.3254x_{3i}\} \times \exp\{-0.5172x_{5i} + 0.3091x_{7i} - 0.9442x_{9i} + 0.3460x_{10i}\}.$$

- Facility:

$$p_1(\mathbf{x}_i) = \frac{1}{fx},$$

where

$$\begin{aligned} fx = 1 &+ \exp\{0.1916 + 0.0017x_{1i} - 0.0321x_{2i} + 0.0760x_{3i}\} \\ &\times \exp\{-0.3896x_{5i} - 0.1499x_{7i} - 1.0684x_{9i} + 0.6287x_{10i}\} \\ &+ \exp\{-0.1070 + 0.0468x_{1i} - 0.0423x_{2i} - 0.3254x_{3i}\} \\ &\times \exp\{-0.5172x_{5i} + 0.3091x_{7i} - 0.9442x_{9i} + 0.3460x_{10i}\}. \end{aligned}$$

Further, from the MLEs in Table 1, we can conclude:

- Expired (2)
  i. Each increase of 1% in the Karnofsky score leads to an expected increase of 0.17% in the chance of death of an individual who remains in the hospital, considering the other variables fixed;
  ii. Each additional year of age leads to an expected decrease of 3.16% in the chance of death of an individual who remains in the hospital, considering the other variables fixed;
  iii. Each unit increase in the INR leads to an expected increase of 7.90% in the chance of death of an individual who remains in the hospital, considering the other variables fixed;
  iv. In the comparison between patients without and with diabetes mellitus, a patient with the disease who remains in the hospital has a 32.27% lower chance of dying, considering the other variables fixed;

v. Each increase of 1 g/dl in the albumin level leads to an expected reduction of 13.92% in the chance of death of an individual who remains in the hospital, considering the other variables fixed;
vi. By comparing patients without and with renal dialysis, a patient undergoing dialysis who remains in the hospital has a 65.64% lower chance of dying, considering the other variables fixed;
vii. By comparing female and male patients, males who remain in the hospital have an 87.52% higher chance of dying, considering the other variables fixed;

- Home (3)
  i. Each 1% increase in the Karnofsky score leads to an expected increase of 4.79% in the chance of a patient who remains in the hospital of later being released, considering the other variables fixed;
  ii. Each additional year of age leads to an expected reduction of 4.14% in the chance of a patient who remains in the hospital of later being released, considering the other variables fixed;
  iii. Each unit increase in the INR leads to an expected reduction of 27.78% in the chance of a patient who remains in the hospital of later being released, considering the other variables fixed;
  iv. By comparing patients without and with diabetes mellitus, a patient with the disease who remains in the hospital has a 40.38% lower chance of later being released, considering the other variables fixed;
  v. Each increase of 1 g/dl in the albumin level leads to an expected increase of 36.22% in the chance of a patient who remains in the hospital of later being released, considering the other variables fixed;
  vi. By comparing patients without and with renal dialysis, a patient undergoing dialysis who remains in the hospital has a 61.10% lower chance of later being released, considering the other variables fixed; and
  vii. By comparing female and male patients, males who remain in the hospital have a 41.34% higher chance of later being released, considering the other variables fixed.

5. Concluding remarks

In this work, we study the multinomial logistic regression model (MLRM) which is an extension of the binomial logistic model. We consider classical inference, bootstrap and Bayesian procedures to estimate the parameters of the baseline-category for the current regression model. They all produced very similar estimates, but with a difference in the significance of the model. We also consider diagnostic measures based on global and local influences and ψ-divergence. The diagnostic measures in both approaches identify the same individuals, which show that the common diagnostic measures in classical models can also be used in categorical data models. We fit the proposed model to a real data set regarding liver transplantation at the Cleveland Clinic to illustrate the diagnostic measures reported. Based on the fitted model to this data set, we conclude that women undergoing renal dialysis and suffering from diabetes mellitus are more likely to remain in the hospital than be released, and several others facts. Further, we identify three patients as possible influential observations. Among these, the patient ♯688 attracts the most attention, because

his characteristics do not satisfy the conditions expected for a fatal outcome, thus indicating that more information should be obtained about this patient.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

[1] A. Agresti, *An Introduction to Categorical Data Analysis*, John Wiley & Sons, Hoboken, NJ, 2007.

[2] J.H. Albert and S. Chib, *Bayesian analysis of binary and polychotomous response data*, J. Am. Stat. Assoc. 49 (1993), pp. 327–335.

[3] E.B. Andersen, *Diagnostics in categorical data analysis*, J. Royal Stat. Soc. B 54 (1992), pp. 781–791.

[4] R. Cook, *Assessment of local influence*, J. Royal Stat. Soc. Ser. B 48 (1986), pp. 133–169.

[5] M.K. Cowles and B.P. Carlin, *Markov chain Monte Carlo convergence diagnostics: A comparative review*, J. Am. Stat. Assoc. 91 (1996), pp. 883–904.

[6] J.N. Da Cruz, E.M.M. Ortega and G.M. Cordeiro, *The log-odd log-logistic Weibull regression model: Modelling, estimation, influence diagnostics and residual analysis*, J. Stat. Comput. Simul. 86 (2016), pp. 1516–1538.

[7] A. Davison and D. Hinkley, *Bootstrap Methods and Their Application*, Cambrigde University Press, New York, 1997.

[8] D. Dey and L. Birmiwal, *Robust Bayesian analysis using divergence measures*, Stat. Probab. Lett. 20 (1994), pp. 287–294.

[9] T. DiCiccio and B. Efron, *Bootstrap confidence intervals*, Stat. Sci. 11 (1996), pp. 189–212.

[10] J.A. Doornik, *An Introduction to OxMetrics 5*, Timberlake Consultants Press, London, 2007.

[11] B. Efron, *Bootstrap methods: Another look at the jackknife*, Ann. Stat. 7 (1979), pp. 1–26.

[12] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hill, New York, NY, 1993.

[13] M.W. Fagerland, D.W. Hosmer and A.M. Bofin, *Multinomial goodness-of-fit tests for logistic regression models*, Stat. Med. 27 (2008), pp. 4238–4253.

[14] J. Hartzel, A. Agresti and B. Caffo, *Multinomial logit random effects models*, Stat. Modelling. 1 (2001), pp. 81–102.

[15] D. Hedeker, *A mixed-effects multinomial logistic regression model*, Stat. Med. 22 (2003), pp. 1433–1446.

[16] G. Ibacache-Pulgar, G.A. Paula and M. Galea, *Influence diagnostics for elliptical semiparametric mixed models*, Stat. Modelling. 12 (2012), pp. 165–193.

[17] J.G. Ibrahim, M. Chen and D. Sinha, *Bayesian Survival Analysis*, Springer-Verlag, New York, 2001.

[18] D.M. Kelly, R. Bennett, N. Brown, J. McCoy, D. Boerner, C. Yu, B. Eghtesad, W. Barsoum, J.J. Fung and M.W. Kattan, *Predicting the discharge status liver transplantation at a single center: A new approach for a new era*, Liver. Transpl. 18 (2012), pp. 796–802.

[19] V.H. Lachos, A.M. Larissa, T. Barbosa, A.W. Garay and D.K. Dey, *Influence diagnostics in spatial models with censored response*, Environmetrics 28 (2017), pp. 1–21.

[20] K. Lee, H. Ahn, H. Moon, R.L. Kodell and J.J. Chen, *Multinomial logistic regression ensembles*, J. Biopharm. Stat. 23 (2013), pp. 681–694.

[21] E.M.M. Ortega, G.M. Cordeiro, A.K. Campelo, M.W. Kattan and V.G. Cancho, *A power series beta Weibull regression model for predicting breast carcinoma*, Stat. Med.34 (2015), pp. 1366–1388.

[22] F. Peng and D. Dey, *Bayesian analysis of outlier problems using divergence measures*, Canadian J. Stat. – La Revue Canadienne De Statistique 23 (1995), pp. 199–213.

[23] M. Plummer, *rjags: Bayesian graphical models using MCMC*, R package version 3-10, 2013. Package available at http://CRAN.R-project.org/package = rjags

[24] R Development Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2011.
[25] S. Ren, *Prediction of the mechanisms of toxic action of phenols: Baseline-category logit models*, Aquat. Toxicol. 65 (1995), pp. 111–115.
[26] C. Russo, G.A. Paula, F.J.A. Cysneyros and R. Aoki, *Influence diagnostics and/or autoregressive nonlinear elliptical models for correlated data*, J. Appl. Stat. 39 (2012), pp. 1049–1067.
[27] F.L. Schumacher, V.H. Lachos and F.E. Vilca-Lavra, *Influence diagnostics for censored regression models with autoregressive errors*, Aust. New Zealand J. Stat. 60 (2018), pp. 209–229.
[28] K. van Steen, G. Molenberghs, G. Verbeke and H. Thijs, *A local influence approach to sensitivity analysis of incomplete longitudinal ordinal data*, Stat. Model. 1 (2001), pp. 125–142.
[29] W. Tang, H. He and X.M. Tu, *Applied Categorical and Count Data Analysis*, Taylor & Francis Group, Boca Raton, 2012.
[30] R. Weiss, *An approach to Bayesian sensitivity analysis*, J. Royal Stat. Soc. 58 (1996), pp. 739–750.
[31] B. Xu, X. Feng and R.D. Burdine, *Categorical data analysis in experimental biology*, Dev. Biol. 348 (2010), pp. 3–11.