Computational Statistics and Data Analysis 66 (2013) 202–216

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

New prediction method for the mixed logistic model applied in a marketing problem

Karin Ayumi Tamura *, Viviana Giampaoli

Institute of Mathematics and Statistics, University of São Paulo, Brazil

ARTICLE INFO

Article history:
Received 27 April 2012
Received in revised form 9 April 2013
Accepted 9 April 2013
Available online 16 April 2013

Keywords:
Mixed logistic regression
Prediction
Random effects
Marketing application

ABSTRACT

When units belong to a specific group, such as employees nested within companies, the data present a hierarchical structure that can be modeled by using mixed models. In addition to fixed effects, these models estimate the random effects for each group. The problem of assigning values to the random effects for new groups in order to predict the outcome in a future period, based on the parameters previously estimated, is the focus of this article. The empirical best predictor (EBP) has been applied to the logistic mixed model, but when there is more than one random effect, the processing time required to solve the multidimensional integrals is costly. A new methodology is proposed based on linear regression that considers the relationship among the random effects and the covariates aggregated at the group level. A comparison among the linear regression prediction method (LRPM), EBP, and the ordinary logistic model is provided through simulation studies and an application study with a mobile company. The results indicate that LRPM drastically reduced the computational effort, and at the same time, maintained a similar level of prediction in relation to EBP.

1. Introduction

The traditional (or ordinary) regression model considers that observations in the dataset are independent. The assumption of independence is violated when data are hierarchically structured, because there is correlation among units that belong to a higher level, indicating that they share the same environment or have similar characteristics. In practice, we can apply a mixed model in the social sciences to investigate how the workplace affects an employee's productivity or in marketing when we are interested in investigating how mobile phone usage by an employee affects the purchase of the additional service within a company. In both examples, the data come from a hierarchical structure, with employees grouped within companies.

The mixed models are adequate for this type of data structure and can be seen as a hierarchical system of the regression of two equations, allowing for the inclusion of variables at both levels (individual and group) with individual observations aggregated within groups. In this type of model, there are random effects that take into account the variability between groups by the random intercept and/or the random slopes. By ignoring this relationship, it omits the importance of the random effects and may invalidate many traditional statistical techniques.

In the context of hierarchical data, Raudenbush and Bryk (2002) presented the multilevel generalized linear models, in which these models with two levels are known as generalized linear mixed models (GLMM).

* Correspondence to: Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão, 1010, CEP 05508-090, São Paulo, Brazil. Tel.: +55 11 3091 6129; fax: +55 11 3091 6134.
E-mail addresses: karinat@ime.usp.br, karin.ayumi.tamura@gmail.com (K.A. Tamura).

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csda.2013.04.006

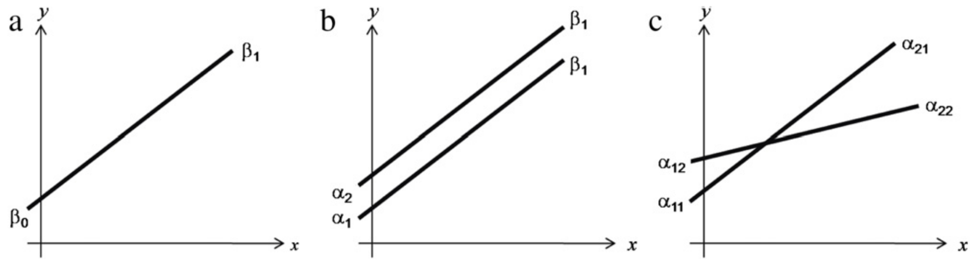

Fig. 1. A comparison among traditional and mixed models with one and two random effects.

The mixed logistic regression is a model in the class of the GLMM and is suitable for binary outcomes, such as the purchase of an additional product, the solicitation of an additional subscription, or the cancellation of a service. These events set up a binary response as follows: 1—the client presented the target event, or 0—the client did not present the target event.

Before proceeding, it is important to note the difference between the terms *prediction* and *estimation* of the random effects. The distinction of these two concepts becomes clearest when considering the marketing application: if there is a group of clients who purchased an additional product, then it seems reasonable to describe the ''purchase probability'' based on the information of usage as an estimation problem. On the other hand, if the interest is to find the probability of a new customer buying an additional product, then this is a prediction problem.

By considering the mixed models in the context of prediction, the challenge is to find the value of the outcome at the observation level, which is a function of the random effects estimated per group. For new groups, it is not possible to predict the response variable using the traditional approach, because we do not know the estimated values of their random effects.

To predict future observations in linear mixed models, Robbinson (1991) provides a broad view with examples and application of the empirical best linear unbiased prediction (EBLUP).

A general review of prediction methods in GLMM can be seen, for example, in Jiang (2003, 2007) or Jiang and Lahiri (2006). Research in prediction involving GLMM generally approaches the case in which the model considers one random effect. In particular, for the mixed logistic regression, Jiang and Lahiri (2001) and Skrondal and Rabe-Hesketh (2009) present methods for prediction of the random effect or the outcome for particular cases. In order to predict the outcome using a mixed logistic model at the observation level, an approach in the literature was presented in Tamura and Giampaoli (2010), in which the authors developed the expression of the EBP considering only the random intercept, which was based on Jiang and Lahiri (2001). An extension of this method considering k random effects was presented by Tamura and Giampaoli (2011).

In this work, we present a new approach to predict the random effects of future groups considering a linear regression model, named as the linear regression prediction method (LRPM). The LRPM takes into account the relationship between the covariates at the group level and the estimated random effects through a linear regression model.

This alternative method was studied in order to reduce the computational effort, given that the EBP presented in Tamura and Giampaoli (2011) involved substantial time processing to solve the multidimensional integrals.

We fitted the mixed model by considering two random effects, the intercept and one slope; however, the estimation and prediction methods can be implemented for $k > 2$ random effects.

Other contributions of this study included the analysis of the bias of the estimation methods and the performance of the prediction methods, which considered the mixed and the traditional models. The mixed model considered the Penalized Quasi-Likelihood (PQL) and Laplace estimation methods, and the traditional model considered Maximum Likelihood estimation. We illustrated the results through simulation studies and an application using a real dataset.

This paper is organized as follows. In Section 2, we described the mixed logistic regression and the EBP for k random effects presented by the authors Tamura and Giampaoli (2011). The LRPM for the mixed logistic regression considering k random effects was described in Section 3. In Section 4, we studied the behavior of the estimate parameters and the performance of the models and the prediction methods through simulations. Application in a real dataset involving a cross-sell model of a mobile operator is provided in Section 5. Finally, in Section 6, we discussed the results, the advantages of the prediction methods for the mixed logistic regression, and possibilities of future work related to the subject.

2. Related works

In the marketing problem, employees of the same company use the mobile service for business purposes. If some employees cancel the mobile service, then it is reasonable to consider that the cancellations are correlated among the employees in the same company. In this example exists dependence among the employees (observations) inside the company (group), and the random part of the mixed models can capture these correlations.

Consider a linear model with a continuous outcome that is a particular model of GLMM. Fig. 1 illustrates the difference between a traditional and mixed model, in which (a) is the traditional model $y = \beta_0 + \beta_1 x$; (b) is the mixed model with the random intercept $y = \beta_0 + \beta_1 x + \alpha_i$; and (c) is the mixed model with the random intercept and one random slope $y = \beta_0 + \alpha_{1i} + \alpha_{2i} x$, with $i = 1$ and 2, indexing the groups. Note that the model in Fig. 1(c) allows the flexibility for considering not only different intercepts to the groups, but also different slopes (non-parallel lines between groups).

Of course, the model in Fig. 1(b) may not capture all the correlations among the observations. Therefore, the inclusion of covariates in the random part of the model (Fig. 1(c)) may provide a richer class of models by considering further information regarding the correlation structure of the groups.

A particular model of GLMM is the mixed logistic regression that considers the binary outcome of the Bernoulli density, with logit link function. In this article, we considered this particular model for predicting the outcome; however the prediction methods presented here can be easily extended to other models belonging to GLMM.

The first approach to predict the outcome of a mixed logistic regression was proposed by Jiang and Lahiri (2001), in which the outcome was at the group level. Based on Jiang and Lahiri (2001), Tamura and Giampaoli (2010) proposed prediction at the observation level (observation nested into the group). In both articles, the authors only considered the intercept as the random effect. In this section, we briefly present the mixed logistic model with k random effects and describe the approach to predict the outcome probability, proposed by Tamura and Giampaoli (2011).

2.1. Mixed logistic regression model

Let y_{ij} be the binary response, with $j = 1, \ldots, n_i$ and $i = 1, \ldots, q$, according to a Bernoulli distribution with parameter μ_{ij}. The density function of this model, belonging to the exponential family, is given by

$$f(y_{ij}|\boldsymbol{\alpha}_i) = \exp\left[\left(\frac{a_{ij}}{\phi}\right)\left(y_{ij}\gamma_{ij} - b\left(\gamma_{ij}\right)\right) + c\left(y_{ij}, \left(\frac{\phi}{a_{ij}}\right)\right)\right],$$

in which $b(\gamma_i) = \log(1 + e^{\gamma_{ij}})$ and $c(\cdot, \cdot)$ are known functions associated with the exponential family; a_{ij} is a weight and ϕ is known, with γ_{ij} associated with $\mu_{ij} = E(y_{ij}|\boldsymbol{\alpha}_i)$. Conditional on $\boldsymbol{\alpha}_i$, the probability of the j-th observation in the i-th group is given by $P(y_{ij} = 1|\boldsymbol{\alpha}_i) = \mu_{ij}$, and the mixed logistic regression is defined by the linear predictor

$$\text{logit}(P(y_{ij} = 1|\boldsymbol{\alpha}_i)) = \log\left(\frac{\mu_{ij}}{1 - \mu_{ij}}\right) = \mathbf{x}_{ij}^t\boldsymbol{\beta} + \mathbf{z}_{ij}^t\boldsymbol{\alpha}_i, \quad (1)$$

in which $\mathbf{x}_{ij}^t$ is a vector $(1 \times p)$ of known covariates, $\boldsymbol{\beta}$ is a vector $(p \times 1)$ of fixed effects coefficients, $\mathbf{z}_{ij}^t$ is a vector $(1 \times k)$ of known associated covariates and $\boldsymbol{\alpha}_i$ is a vector $(k \times 1)$ of the random effects, in which α_{1i} is the random intercept and $(\alpha_{2i}, \ldots, \alpha_{ki})$ are the random slopes of the vector $\boldsymbol{\alpha}_i^t = (\alpha_{1i}, \alpha_{2i}, \ldots, \alpha_{ki})$. Furthermore, the assumption of this model is that $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_q$ are i.i.d. with $\boldsymbol{\alpha}_i \sim \mathcal{N}_k(\mathbf{0}, \boldsymbol{\Sigma})$, in which $\boldsymbol{\Sigma}$ is the unknown covariance matrix of the random effects.

The exact likelihood function has complicated contributions, and approximation techniques can be used to estimate the parameters of the model (1). In this article, we used two different techniques, that are available in the R software: PQL, proposed by Breslow and Clayton (1993), and the Laplace approximation of integrals. An overview of the GLMM and these two estimation methods can be found, for example, in Demidenko (2004) and McCulloch and Searle (2001).

2.2. EBP method

To develop the EBP method with k random effects, Tamura and Giampaoli (2011) considered the multivariate linear transformation $\boldsymbol{\alpha_i} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\xi_i}$, with $\boldsymbol{\alpha}_i \sim \mathcal{N}_k(\mathbf{0}, \boldsymbol{\Sigma})$ and $i \in Q = \{1, \ldots, q\}$. The authors showed, based on Rencher (1998), that $\boldsymbol{\xi}_i \sim \mathcal{N}_k(\mathbf{0}, \mathbf{I})$, with $\mathbf{I}$ representing the identity matrix. Inspired by this transformation, the authors proposed the EBP expression for predicting the outcome probability of the j-th observation within the l-th of a new group, with $l \notin Q$, given by

$$\hat{p}_{lj} = \frac{\exp(\mathbf{x}_{lj}^t\hat{\boldsymbol{\beta}})\int_{\xi_1}\cdots\int_{\xi_k}\frac{\exp((y_{l.}+1)\mathbf{z}_{lj}^t\boldsymbol{\alpha}^*)}{1+\exp(\mathbf{x}_{lj}^t\hat{\boldsymbol{\beta}}+\mathbf{z}_{lj}^t\boldsymbol{\alpha}^*)}\prod_{l=1}^{n_i}\frac{1}{1+\exp(\mathbf{x}_{lj}^t\hat{\boldsymbol{\beta}}+\mathbf{z}_{lj}^t\boldsymbol{\alpha}^*)}f(\xi_1, \ldots, \xi_k)d\xi_1\ldots d\xi_k}{\int_{\xi_1}\cdots\int_{\xi_k}\exp(y_{l.}\mathbf{z}_{lj}^t\boldsymbol{\alpha}^*)\prod_{l=1}^{n_i}\frac{1}{1+\exp(\mathbf{x}_{lj}^t\hat{\boldsymbol{\beta}}+\mathbf{z}_{lj}^t\boldsymbol{\alpha}^*)}f(\xi_1, \ldots, \xi_k)d\xi_1\ldots d\xi_k}, \quad (2)$$

in which $\hat{\boldsymbol{\beta}}$ is estimated by model (1), $f(\xi_1, \ldots, \xi_k) = f(\xi_1)\cdots f(\xi_k)$ and $f(\xi_m)$ is the univariate standard normal density, with $m = 1, \ldots, k$. For the l-th new group, since $\boldsymbol{\alpha_l}$ is unknown, the notation for prediction was considered as $\boldsymbol{\alpha}^*$, in which $\boldsymbol{\alpha}^* = \hat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{\xi}$, with $\boldsymbol{\alpha}^* = (\alpha_1^*, \alpha_2^*, \ldots, \alpha_k^*)^t$, i.e.,

$$\begin{pmatrix}\alpha_1^*\\ \alpha_2^*\\ \vdots\\ \alpha_k^*\end{pmatrix} = \begin{pmatrix}\xi_1\hat{\upsilon}_{11} + \xi_2\hat{\upsilon}_{12} + \cdots + \xi_k\hat{\upsilon}_{1k}\\ \xi_1\hat{\upsilon}_{21} + \xi_2\hat{\upsilon}_{22} + \cdots + \xi_k\hat{\upsilon}_{2k}\\ \vdots\\ \xi_1\hat{\upsilon}_{k1} + \xi_2\hat{\upsilon}_{k2} + \cdots + \xi_k\hat{\upsilon}_{kk}\end{pmatrix},$$

with $\hat{\upsilon}_{mm'}$, with $m = 1, \ldots, k$ $m' = 1, \ldots, k$, be the variance components of matrix $\hat{\boldsymbol{\Sigma}}^{1/2}$, estimated by model (1). In expression (2), since we do not know the outcome at the group level, we considered $y_{l.} = n_l/2$, which represents equal probabilities for success and failure of the observations inside the l-th group. In addition, the integrals have no closed-form expression and they should be evaluated numerically with an appropriate method for multidimensional integration.

3. Linear regression prediction method

In this section, we propose a method for predicting the random effect for new groups, since the EBP expression given in (2) requires a large computational effort for calculating the multidimensional integrals.

The hypothesis is to predict the random effect for the new groups, considering the relationship between the covariates and random effects estimates (response variable) through linear regression models.

Suppose that the parameters (fixed and random effects) of the mixed logistic model were estimated by the model (1). Then, the function for the prediction of the probability outcome can be written by

$$\hat{\mu}_{ij} = \frac{\exp\{\mathbf{x}_{ij}^t\hat{\boldsymbol{\beta}} + \mathbf{z}_{ij}^t\hat{\boldsymbol{\alpha}}_i\}}{1 + \exp\{\mathbf{x}_{ij}^t\hat{\boldsymbol{\beta}} + \mathbf{z}_{ij}^t\hat{\boldsymbol{\alpha}}_i\}} \tag{3}$$

with $j = 1, \ldots, n_i$, $i = 1, \ldots, q$, and $i \in Q = \{1, \ldots, q\}$.

If the interest was to predict the outcome based on the groups considered in model (3), then the prediction would be straightforward. However, since we do not know the individual value for the l-th new group, with $l \notin Q$, the strategy to explain the new random effects is based on their covariates. In this case, the methodology considers, once the estimates $\hat{\boldsymbol{\alpha}}_i$ from the interest parameter $\boldsymbol{\alpha}_i$ (with $i \in Q$) are obtained, the known values $\hat{\boldsymbol{\alpha}}_i$ as the outcome of additional regression models. Thus, the estimate parameters of these regression models are used to predict the random effects for the l-th new group, now with $l \notin Q$.

LRPM is described as follows. After obtaining the estimate parameters of model (3), all of the covariates available at the observation level should be aggregated at the group level, i.e.,

$$\mathbf{w}_i^t = (\mathbf{x}_{i.}^t, \mathbf{z}_{i.}^t), \tag{4}$$

with $i \in Q$. There are several ways of aggregating variables. For quantitative variables, one can use functions such as mean, median, minimum, maximum, and others. For qualitative variables, one approach is to use the mode. The choice of the aggregation function to be used depends on the problem, and moreover the same variable can be tested by using different approaches for summarizing the information at the group level.

By considering the covariates $\mathbf{w}_i^t$, we will fit one regression model to each m-th random effect estimate, with $m = 1, \ldots, k$. The dataset considered to fit models is called the training dataset. Thus, we looked for a model

$$\hat{\alpha}_{mi} = f(\mathbf{w}_{mi}^t\boldsymbol{\lambda}_{\boldsymbol{m}}) \tag{5}$$

that was able to explain the relationship between the covariates and the random effects, with $i = 1, \ldots, q$ indexing groups and $m = 1, \ldots, k$ indexing the random effects, where $\boldsymbol{\lambda}_{\boldsymbol{m}} = (\lambda_{m1}, \ldots, \lambda_{mp})^t$ is the vector of unknown regression coefficients and $\mathbf{w}_{mi}$ is the vector of known covariates ($p \times 1$) of the i-th group and the m-th random effect, aggregated at the group level as described in (4).

Based on the assumption of model (1), i.e., the random effects follow a normal distribution, it seemed reasonable to assume that the general model (5) may be particularized to a linear regression model, given by

$$\hat{\alpha}_{mi} = \mathbf{w}_{mi}^t\boldsymbol{\lambda}_{\boldsymbol{m}} + \varepsilon_{mi}, \tag{6}$$

with $\varepsilon_{mi} \sim \mathcal{N}(0, \sigma_m^2)$, independents. The parameter λ_{m1} is the fixed intercept and $(\lambda_{m2}, \ldots, \lambda_{mp})^t$ are the fixed slopes of the vector $\boldsymbol{\lambda}_{\boldsymbol{m}}$. The parameters can be performed by the usual estimation methods, such as Least Squares or Maximum Likelihood.

In the assumption of model (1), as $\boldsymbol{\alpha}_{\boldsymbol{i}}$ has a multivariate normal distribution, then each marginal m of random effects (α_{mi}) has univariate normal distribution. Thus, if the mixed model has a set of k random effects, we fit k-linear independent models assuming that the empirical random effects follow the normal distribution.

Notably, the proposed methodology using the model (6) implies that $\alpha_{mi} \sim \mathcal{N}(\mathbf{w}_{mi}^t\boldsymbol{\lambda}_{\boldsymbol{m}}, \sigma_m^2)$, i.e., we relaxed the assumption of the model (1) that the average α_{mi} is zero, and we considered $E(\alpha_{mi}) = \mathbf{w}_{mi}^t\boldsymbol{\lambda}_{\boldsymbol{m}}$ by postulating that $\boldsymbol{\Sigma} = \mathrm{diag}({\sigma_1}^2, \ldots, {\sigma_k}^2)$. An alternative that takes into account the correlation between the outcomes (random effects) could be the multivariate linear regression, but is not considered in this approach.

In order to identify the set of covariates that explain the outcome (the m-th empirical random effect), several methods have been described in the literature for selecting regression models. Some known procedures, such as backward, forward, stepwise R^2-adjusted, residual mean square s^2, Mallow's C_p, Akaike Information Criterion (AIC), and the Bayes Information Criterion (BIC), can be used. For more details regarding these procedures, see for example Hocking (1976), Draper and Smith (1998) and Faraway (2002).

These procedures are useful when there are several predictors in the database, and all covariates available should be tested, including those that were not part of the fitted mixed model, in order to find additional information to explain the random effects. Here, we use the procedure *stepAIC* available in the library *MASS* in the R software. This procedure uses the AIC criterion that is defined as

$$\mathrm{AIC} = -2\log(\text{likelihood}) + 2p,$$

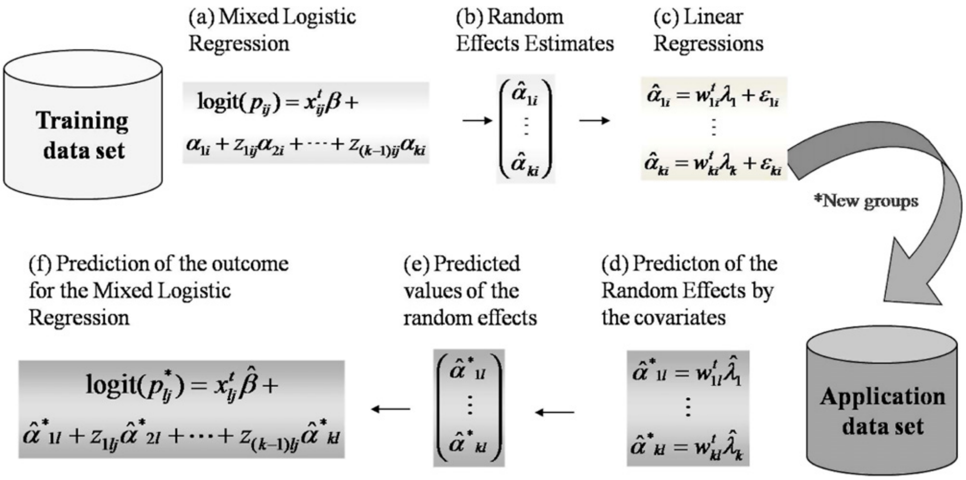

Fig. 2. A summary of the linear regression prediction method.

with p representing the number of predictors in the model in question. We wanted to minimize AIC in each step of removing predictors, because a model with all predictors fits better. The stop rule of removing the predictors is associated with the p-value criterion and stops when all p-values are less than the critical value. The multicollinearity problem in the selection of the covariates should be investigated, in order to be sure that the model is fitted correctly.

Based on (6), we selected the final model by using the AIC criterion for each m-th random effect, in which $\hat{\boldsymbol{\lambda}}_{\boldsymbol{m}} = (\hat{\lambda}_{m1}, \ldots, \hat{\lambda}_{mp'})^t$ is the vector of estimate regression coefficients and $\mathbf{w}_{mi}$ is the vector of known covariates ($p' \times 1$), with $p' \leq p$. However, the AIC criterion might be very time-consuming, if the number of covariates are large and equivalent to 2^p, in which p is the number of potential predictors (covariates).

Here, for predicting the random effects for the l-th new group (application dataset), with $l \notin Q$, we use the following regression equation:

$$\hat{\alpha}^*_{ml} = \mathbf{w}^t_{ml}\hat{\boldsymbol{\lambda}}_{\boldsymbol{m}}. \tag{7}$$

Finally, using (7) and the mixed logistic equation, it is possible to predict the outcome probability for the j-th observation within the m-th new group by

$$\hat{\mu}^*_{lj} = \frac{\exp\{\mathbf{x}^t_{lj}\hat{\boldsymbol{\beta}} + \mathbf{z}^t_{lj}\hat{\boldsymbol{\alpha}}^*_l\}}{1 + \exp\{\mathbf{x}^t_{lj}\hat{\boldsymbol{\beta}} + \mathbf{z}^t_{lj}\hat{\boldsymbol{\alpha}}^*_l\}}, \tag{8}$$

in which $\hat{\boldsymbol{\alpha}}^*_l = (\hat{\alpha}^*_{1l}, \ldots, \hat{\alpha}^*_{kl})$.

Notably, we were not interested in inferring or interpreting the parameters of the linear model (7). We were only seeking a "rule" to predict the random effects.

The summary of the LRPM is presented in Fig. 2, by steps:

- *step* (a): In the training dataset, the mixed logistic model was fitted;
- *step* (b): Extract the matrix random effects ($q \times k$) and separate each vector of random effect ($q \times 1$). In this step, aggregate all covariates at the group level by using a diversity of aggregation functions;
- *step* (c): For each m-th random effect estimate, fit a linear regression by using a stepwise procedure with the AIC criterion;
- *step* (d): In the application dataset (new groups), for each m-th random effect estimate, predict the random effect based on the covariates selected in step (c) and the estimate parameters provided by the linear model;
- *step* (e): Provide the predicted values for all random effects of the new groups;
- *step* (f): Insert the predicted random effects in the mixed logistic model function by considering the estimate values of the fixed effects obtained by step (a). Finally, obtain the outcome probability prediction for the mixed logistic regression.

4. Simulation study

We used two models with the following estimation methods: the traditional (or ordinary) logistic regression using Maximum Likelihood and the mixed logistic regression using PQL and the Laplace approximation.

The mixed model with the random intercept and one random slope is given by

$$\text{logit}[P(y_{ij} = 1|(\alpha_{1i}, \alpha_{2i}))] = x_{ij}\beta + \alpha_{1i} + z_{ij}\alpha_{2i}, \tag{9}$$

Table 1
Simulation scenarios.

Scenario	σ_2	n
1	2.5	15
2	5	15
3	10	15
4	10	6
5	10	3

and the traditional logistic regression by

$$\text{logit}[P(y_{ij} = 1)] = x_{ij}\beta. \quad (10)$$

In models (9) and (10), β is the slope associated with the fixed effect, and in model (9) α_{1i} is the random intercept and α_{2i} is the random slope associated with covariate z_{ij}, with $j = 1, \ldots, n$ and $i = 1, \ldots, q$. In addition, $\boldsymbol{\alpha_i} = (\alpha_{1i}, \alpha_{2i})^t \sim \mathcal{N}_2(\mathbf{0}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}.$$

We considered the following values for the simulation studies:

- total number of observations (N): 120;
- number of observations within the group (n): 3, 6, and 15;
- slope associated with the fixed effect (β): 1;
- standard deviation of the random intercept (σ_1): 2.5;
- standard deviation of the random slope (σ_2): 2.5, 5 and 10;
- covariance of α_1 and α_2 (σ_{12}): 1.

In simulation studies, some combinations of values n and σ_2 were used and five scenarios were analyzed by fixing the values $N = 120$, $\beta = 1$, $\sigma_1 = 2.5$ and $\sigma_{12} = 1$, as given in Table 1.

The models presented in simulation studies considered the binary outcome y and the covariate x associated with the fixed effect, which followed a standard normal distribution. We considered the covariate z, which was associated with the random slope, following a normal distribution with mean 3 and variance 2. Then, we generated the observations of two covariates, x_{ij} and z_{ij}, per group, where each group had n observations. The covariates were generated independently from the groups.

For each group, we generated the random effect $\boldsymbol{\alpha_i}$ with $\boldsymbol{\alpha_i} \sim \mathcal{N}_2(\mathbf{0}, \boldsymbol{\Sigma})$.

Without loss of generality and assuming q groups, each had n observations, resulting in a total of $(q \times n) = N$ observations. Note that the same $\boldsymbol{\alpha_i}$ was used for each n observation within the same group.

Probabilities were calculated using the expression given by (1), considering the true model with two random effects, i.e., the mixed logistic model with a random intercept and one random slope.

Using the probability generated in the previous step, the binary response was generated using the Bernoulli distribution. In this step, the database was formed with a binary response variable (y) and two covariates (x and z).

The database was separated in the training and the application datasets. Therefore, the groups were selected alternately: the groups indexed by even groups were allocated in the training dataset, while the odd groups were used in the application dataset, assuming that the groups were randomly indexed.

For each simulation scenario, we generated 500 datasets, each with sample size N with q groups, with the covariates x and z generated only once at the beginning of each simulation scenario.

At the end of each simulated scenario, we calculated predictive performance of the models for the training and application datasets, which is described in the results of the simulation studies.

The objective was to compare the estimate of the parameters and to evaluate the performance measurements of the models between the five simulated scenarios.

Note that, in order to evaluate the prediction of the outcome, considering the mixed models, the random effects were known in the training dataset, while for new groups (application dataset), we used EBP and LRPM to predict the outcome.

4.1. Results

Analysis of the estimate parameters

In order to compare the estimate parameters, we carried out models (10) with Maximum Likelihood and (9) with PQL and Laplace estimation methods by using the procedures *glm*, *glmmPQL*, and *glmer*, respectively, which are available in the R software, version 2.10.1, see R Development Core Team (2010).

The following notations were used to reference the models:

- MTrad: Traditional logistic regression model adjusted through Maximum Likelihood,

Table 2
Average of estimate parameters ($\hat{\beta}$, $\hat{\sigma}_1$, $\hat{\sigma}_2$ e $\hat{\sigma}_{12}$) in training dataset.

Estimation method		Scenario				
		1	2	3	4	5
	β	1	1	1	1	1
MTrad	Average of $\hat{\beta}$	0.296	0.248	0.204	0.138	0.073
MPQL		0.973	0.935	0.892	0.730	0.124
MLap		0.948	0.851	0.823	0.539	0.394
	σ_1	2.5	2.5	2.5	2.5	2.5
MPQL	Average of $\hat{\sigma}_1$	2.135	2.177	2.179	2.446	1.460
MLap		2.119	1.930	1.835	1.081	0.724
	σ_2	2.5	5	10	10	10
MPQL	Average of $\hat{\sigma}_2$	1.493	1.930	2.408	3.104	1.172
MLap		1.724	2.308	3.018	3.858	4.111
	σ_{12}	1	1	1	1	1
MPQL	Average of $\hat{\sigma}_{12}$	0.019	0.205	0.184	−0.196	−0.176
MLap		0.192	0.230	0.190	0.398	0.682

- MPQL: Mixed logistic regression with two random effects adjusted through PQL,
- MLap: Mixed logistic regression with two random effects adjusted through Laplace.

The estimate parameters of each model for each scenario are shown in Table 2. We observed that MTrad, in general, presented an average of $\hat{\beta}$ more biased than MPQL and MLap. By comparing scenarios 1, 2, and 3, we noticed that as the values of the standard deviation of the random slope increased, the average of $\hat{\beta}$ became more biased. In addition, analyzing scenarios 3, 4, and 5, we concluded that the average of $\hat{\beta}$ became more biased when the size of the group n decreased.

The mixed models underestimated the average values of $\hat{\sigma}_1$, $\hat{\sigma}_2$ and $\hat{\sigma}_{12}$ compared to their true values. The average of $\hat{\sigma}_1$ estimated by MLap was more biased than MPQL. We observed that the average $\hat{\sigma}_1$ was more biased for group size $n = 3$.

For the parameter σ_2, the average estimate values using the Laplace method were less biased than using the PQL method. We noticed that when the value of σ_2 increased, the average of $\hat{\sigma}_2$ was closer to the true value σ_2 for both estimation methods. In addition, for MLap, when the value of n decreased, the average of $\hat{\sigma}_2$ showed a tendency to go to the true value, although the values were in the same level. In all scenarios, the average of $\hat{\sigma}_2$ was underestimated, mainly when the value of σ_2 increased.

Finally, analyzing σ_{12}, the average of $\hat{\sigma}_{12}$ of MLap was less biased than MPQL in all cases. In addition, when the value of n decreased, the average $\hat{\sigma}_{12}$ was close to its true value σ_{12}, but only for the Laplace method. Remarkably, the estimate values were underestimated in all scenarios, presenting values close to zero. The bias of the estimate values of $\hat{\sigma}_{12}$ gives an advantage to the LRPM, given that the methodology assumes that there is no correlation between the random effects.

The biases observed in the components of variance were expected for the mixed model with binary response. Engel (1998) suggests that estimation procedures may yield seriously biased estimates for components of variance (underestimated by as much as 50%) and intra-class correlation, and Raudenbush et al. (2000) discuss that bias is most serious when the random effects have large variance.

In addition to the scenarios previously presented, we studied the cases when N assumed values different from 120 and when σ_{12} was zero. In the first case, we concluded that regardless of the size of the database, the relationship between group size (n) was relative to the total number of observations (N). In the second case, because the methods of estimation of the mixed model are biased and produced average estimate values of the correlation between the random effects close to zero, the results were similar when $\sigma_{12} = 1$.

Performance prediction

In order to evaluate the performance prediction of the outcome, we considered the following performance measurements: Md.SR, sensitivity, specificity, and KS statistics. A brief description of the performance measures is given as follows:

- The squared residual measures the distance of true probability versus the predicted probability. Empirically, we noticed that the distribution of the squared residuals was positively asymmetric, and therefore we used the median to represent the centrality of the residuals, which is calculated as Md.SR $=$ median$[(p_{11}-\hat{p}_{11})^2, (p_{12}-\hat{p}_{12})^2, \ldots, (p_{qn_q}-\hat{p}_{qn_q})^2]$, in which p_{ij} denotes the observed probability for the j-th observation in the i-th group with $i = 1, \ldots, q$ and $j = 1, \ldots, n_i$ and $\hat{p}_{ij}$ the predicted probability.
- Sensitivity measures the percentage of accuracy in event 1, which indicates how good the model is to predict the outcome 1 when the observed response is truly 1. Specificity measures the accuracy within the complementary event, which indicates how good the model is at predicting the outcome 0 when the observed response is truly 0. For more details about these measurements see, for example, Hosmer and Lemeshow (2000) or Collett (2003).

Table 3
Performance measures (Md.SR, sensitivity, specificity, and KS) in the training dataset.

Estimation method	Scenario				
	1	2	3	4	5
	Md.SR				
MTrad	0.18	0.20	0.22	0.23	0.24
MPQL	0.00	0.00	0.00	0.00	0.04
MLap	0.00	0.00	0.00	0.00	0.00
	Sensitivity				
MTrad	64.26	60.14	59.15	63.15	56.29
MPQL	91.62	93.11	93.89	95.54	97.11
MLap	91.67	92.70	93.76	95.11	96.47
	Specificity				
MTrad	54.07	53.63	53.52	55.43	52.31
MPQL	90.86	92.50	93.45	94.95	97.60
MLap	90.76	92.64	93.33	95.22	96.88
	KS				
MTrad	18.32	13.77	12.67	18.59	16.29
MPQL	82.46	85.62	87.33	90.49	64.96
MLap	82.43	85.34	87.09	90.33	93.37

- Based on sensitivity and specificity, we calculated nonparametric KS statistics, which test the hypothesis that two populations have the same distribution for a given random variable and is calculated by KS $=$ max |sensitivity−(1-specificity)| for different cut-offs. For more details, see, for example, Conover (1999).

In sensitivity, specificity, and KS measurements, the greater the value, the better the discriminative power of the model. On the other hand, the smaller the value of Md.SR, the better the discriminative power of the model.

In Table 3 we present the performance measures for the training dataset. Analyzing this table, we noticed that the mixed model, by using PQL and Laplace, were considered more suitable for the hierarchical data structure than the ordinary model, because the performance measures for the mixed models presented more accurate values in terms of prediction in all scenarios. When we compared the prediction measures between PQL and Laplace estimation methods, we noticed that in general PQL was slightly superior to Laplace.

We evaluated the performance values for the new groups in the application dataset presented in Table 4. For the mixed model, we applied the EBP and LRPM for each estimation method in order to predict the outcome for these new groups. In general, the prediction methods for the mixed model presented superior performance measures compared to the traditional model, except for sensibility in the LRPM case. However, the LRPM presented accurate values of performance compared to the traditional model, such as when considering the measurements as Md.SR and KS. A comparison of the prediction methods showed that both were similar, although LRPM was slightly superior to EBP for the measurements Md.SR and KS. LRPM presented superior performance measures compared to EBP for specificity, but was inferior with regard to sensitivity.

Of note, in the training dataset, the difference in the performance measures between the ordinary and mixed models was larger than the application dataset. This result was expected, since we did not know the values of the random effects for new groups in the application dataset.

In order to compare the simulated scenarios, we built graphics for Md.SR and KS measurements. Fig. 3 shows that when KS values are interpreted as the value σ_2 increased, the LRPM had a larger difference from the traditional model compared to EBP. In addition, MPQL and MLap presented values of KS that were very similar, indicating that the estimation method did not have an influence on the prediction of response.

In the same figure, we noted that there was no influence of the σ_2 values when we compare the traditional model to the mixed models because the lines were approximately parallel. This result could be a consequence of the bias caused by underestimation of σ_2, especially when the value of σ_2 increased. The same conclusions obtained by KS can be seen by analyzing Md.SR values.

In Fig. 4, we observed that when the group size increased, the prediction methods for mixed models were more effective than the traditional model. We observed a stronger discriminative power for LRPM compared to EBP, in particular when $n = 6$ and $n = 15$. The exception was MLap (EBP), in which the values of Md.SR and KS were similar for group sizes $n = 3$ and $n = 6$.

In summary, the following practical conclusions can be drawn for the EBP and LRPM in relation to the traditional method to predict the outcome for new groups:

- The EBP and LRPM performed better than the traditional model for all presented scenarios by considering Md.SR and KS.
- The performance of the EBP and LRPM methodology is weakly related to the bias of the estimate parameters.
- EBP and LRPM methods performed better when the group size increased and when the standard deviation of the random effects (σ_1, σ_2) increased.

Table 4
Performance measures (Md.SR, sensitivity, specificity and KS) in application dataset, using the EBP and LRPM.

Prediction method	Estimation method	Scenario				
		1	2	3	4	5
		Md.SR				
Traditional	MTrad	0.20	0.23	0.24	0.24	0.25
EBP	MPQL	0.17	0.19	0.21	0.20	0.24
	MLap	0.16	0.20	0.21	0.20	0.20
LRPM	MPQL	0.15	0.16	0.19	0.21	0.24
	MLap	0.15	0.15	0.19	0.19	0.20
		Sensitivity				
Traditional	MTrad	64.21	63.63	60.84	62.54	60.51
EBP	MPQL	66.22	62.97	64.55	65.46	62.40
	MLap	67.05	63.46	62.82	65.37	60.67
LRPM	MPQL	53.10	51.23	49.62	52.50	59.85
	MLap	54.28	52.92	50.93	56.63	56.08
		Specificity				
Traditional	MTrad	55.13	53.12	54.38	52.99	53.68
EBP	MPQL	55.71	56.64	52.50	53.56	52.50
	MLap	54.84	55.58	55.27	52.91	55.23
LRPM	MPQL	69.71	69.37	70.44	64.61	55.50
	MLap	68.85	68.47	70.13	60.88	59.96
		KS				
Traditional	MTrad	19.35	16.74	15.22	15.53	14.20
EBP	MPQL	21.93	19.62	17.05	19.02	14.90
	MLap	21.89	19.04	18.09	18.28	15.90
LRPM	MPQL	22.81	20.60	20.06	17.12	15.36
	MLap	23.13	21.39	21.05	17.51	16.03

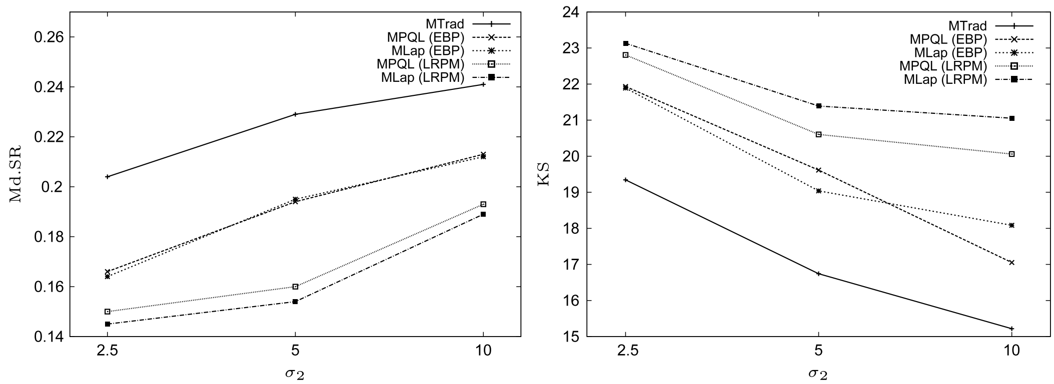

Fig. 3. Md.SR and KS values in scenarios 1, 2 and 3.

Computational issues.

To explore the computational complexity of EBP (e.g. scenario 3) considering the total of 60,000 observations ($N = 120$ multiplied to 500 replicas) with the group size $n = 15$, it took approximately one month to predict the outcome (in the application data with 30,000 observations) by using an ordinary computer, while using the LRPM, the time required was less than 10 min. In the EBP case, the processing time depended on the size of the dataset and the quantity of observations in the groups. The greater these two values, the slower the time of processing, given that the quantity of replicas remained constant.

5. Application: cross-sell model in telecommunication branch

In this section, we present an application of the mixed logistic regression model in direct marketing for the telecommunication field. We considered a dataset with hierarchical structure and applied the prediction methods in a future period.

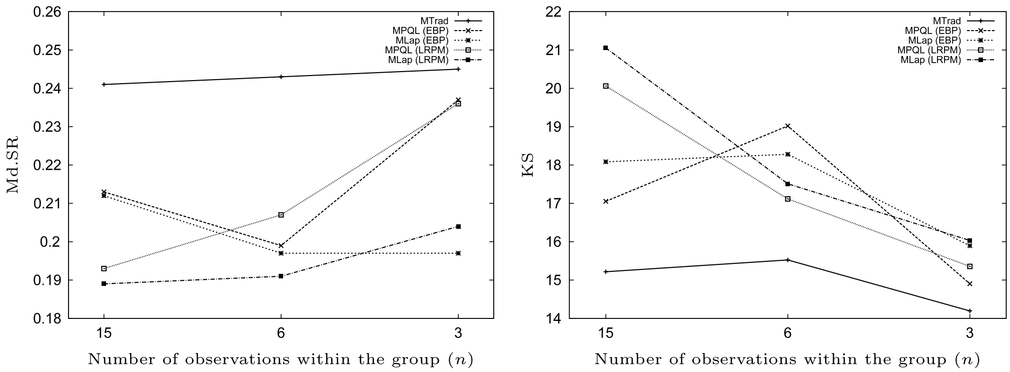

Fig. 4. Md.SR and KS values in scenarios 3, 4 and 5.

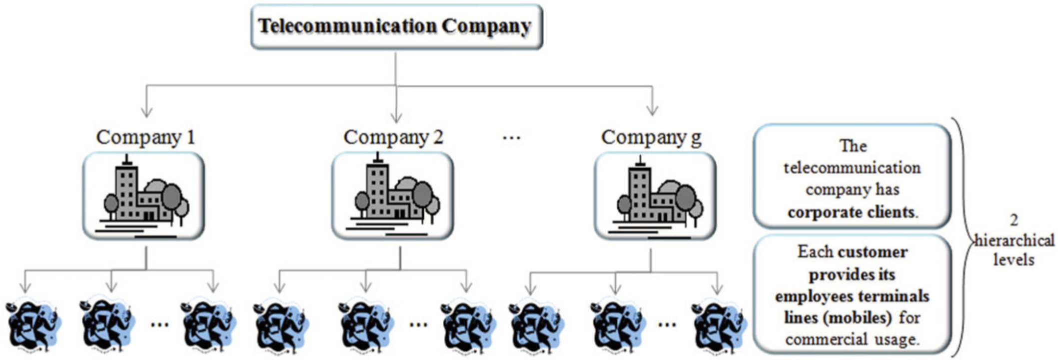

Fig. 5. Hierarchical dataset structure.

5.1. Problem

In direct marketing surveys, cross-selling is defined as "selling an additional product or service to an existing customer". The objectives of cross-selling can be either to increase the income or to protect the relationship with the client. Often, cross-selling involves offering the customer items that complement the original purchase. Butera (2000) discusses the importance of cross-selling and states that although a client does not need an additional product, she or he may need it in the future. Moreover, when the customers are satisfied with the customer service, they are more likely to return to the company than to go to the competitor.

Recently, companies have stored information of their customers in a database. The information stored can be useful for identifying client profiles in order to support the customer relationship management (CRM) strategy.

This problem involves a mobile operator that wants to approach its customers by using the telemarketing channel. This sales channel has high cost, and a statistical model could help by selecting a minimum number of customers with a maximum effectiveness to sell the service.

The aim is to obtain a score or probability in which the mobiles can be sorted and prioritized, which would identify the most likely customers to purchase an additional service. This type of model in a business market is known as a cross-sell model.

The logistic regression model has been used for cross-selling, and mixed logistic regression is appropriate when the database has a hierarchical structure. In this problem, the hierarchical structure can be represented by corporate customers (groups) with employees nested into the company (see Fig. 5).

The additional product that we treat in this problem is the short message service (SMS), which is a text messaging service of a mobile that allows the exchange of short text messages between cell phone devices. There are three ways to acquire the service: paying individually by usage, paying for a full day, or signing up for the monthly service. If the service is purchased per day or per month, then customers can send unlimited SMS within the period.

The monthly package is the most economical for the customer among the three service options provided, since the owner uses the service frequently.

5.2. Dataset

The telecommunication company stores the data monthly. In order to analyze the data, a specific month should be selected as the reference month. The behavior of the corporate customers or their terminal lines were observed from the

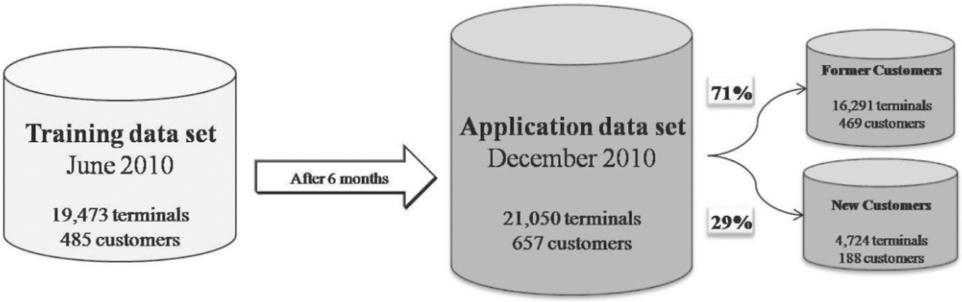

Fig. 6. Training and application datasets.

Table 5
Summary statistics—quantity of terminal lines per customer.

Dataset	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
June 2010	1	22	28	40	42	487
December 2010	1	19	23	32	35	442

reference month to prior months. The aim was to predict the outcome one month after the reference month. The binary outcome of the model was defined as:

- 1—terminal line acquired the SMS monthly package one month after the reference month;
- 0—terminal line did not acquire the SMS monthly package one month after the reference month.

We considered two period of analysis, June 2010 and December 2010. Data from June 2010 were considered as the training dataset and were used to the model the data. For prediction in a future period, we used the dataset from December 2010 (application dataset).

Fig. 6 shows the quantity of corporate customers and terminal lines (mobiles) in each period. In December 2010, 29% of customers were new clients, and the remaining clients were in the dataset from June 2010.

An interesting application of this problem is that the quantity of terminal lines varies according to the corporate customer. In Table 5 we can see the summary statistics regarding the quantity of terminals per customer for the periods of June 2010 and December 2010. On average, there are approximately thirty-five lines per customer with a median of twenty-five.

There was an abundance of information in the original dataset, which could be used as covariates. These covariates were collected from the corporate customer and the terminal levels as follows:

- Corporate customer: geographical location, age of founding of the company, activity branch, time in months of the company with the mobile operator, quantity of active terminals, value of the invoice, and others.
- Terminal: time in months of the terminal with the mobile operator, indicator of using the data package, indicator of using e-mail, indicator of paying SMS per usage, quantity of paying SMS service per day, indicator of paying SMS service monthly in the past, minutes of calls received, minutes of outgoing calls, minutes of usage roaming, percentage of minutes that the terminal represents inside the company, percentage of money of which the terminal represents inside the invoice of the company, and others.

5.3. Fitted models

We analyzed the data by considering the initial mixed model with all covariates described in Section 5.2. The selection models were led in such a way that the final model is well specified in both fixed and random parts. After the model selection procedure, the final mixed logistic regression with 2 random effects is given by

$$\mathrm{logit}[P(y_{ij} = 1)|(\alpha_{1i}, \alpha_{2i})] = \mathbf{x}_{ij}^t\boldsymbol{\beta} + \alpha_{1i} + z_{ij}\alpha_{2i}, \tag{11}$$

in which i indexes the groups (or corporate customers), with $i = 1, \ldots, 485$; and j indexes the observations (or terminal lines) inside the groups, with $j = 1, \ldots, n_i$. Some covariates (factors) were categorized and the first category of each variable was considered to be the reference cell, such as: quantity of paying for SMS service per day less than 2, for factor 1; and minutes of usage roaming equal to zero, for factor 3. The elements of $\boldsymbol{\beta}$ and the categories of the covariates in the fixed part of the model are described above:

- β_0: intercept;
- Factor 1: $x_{1_{ij}}$ representing the quantity of paying for SMS service per day equal to 2, associated with β_1; $x_{2_{ij}}$ representing the quantity of paying for SMS service per day greater than 2, associated with β_2;
- Factor 2: $x_{3_{ij}}$ representing the indicator of paying for SMS service monthly, associated with β_3;

Table 6
Estimate parameters, S.E., and p-values, for the traditional and mixed models.

Parameter	Traditional		Mixed	
	Estimate (s.e.)	p-value	Estimate (s.e.)	p-value
β_0	−4.90(0.09)	<0.01	−9.31 (0.47)	<0.01
β_1	1.26(0.17)	<0.01	0.59 (0.23)	0.01
β_2	1.68(0.29)	<0.01	1.07 (0.38)	<0.01
β_3	2.34(0.31)	<0.01	1.63 (0.40)	<0.01
β_4	0.70(0.18)	<0.01	0.56 (0.25)	0.02
β_5	0.30(0.26)	0.25	−0.04 (0.35)	0.91
β_6	1.54(0.35)	<0.01	−0.71 (0.44)	0.10
β_7	0.79(0.29)	0.01	1.97 (0.43)	<0.01
(σ_1, σ_2)			(5.66, 2.41)	
$\rho = \sigma_{12}/\sigma_1\sigma_2$			(−0.68)	

- Factor 3: $x_{4_{ij}}$ representing the minutes of usage roaming greater than 0 and less than 10, associated with β_4; $x_{5_{ij}}$ representing the minutes of usage roaming greater than 10, associated with β_5;
- Factor 4: $x_{6_{ij}}$ representing the indicator of using e-mail, associated with β_6;
- Factor 5: $x_{7_{ij}}$ representing the indicator of using a data package, associated with β_7.

The covariate z_{ij}, in the random part of the model, is the percentage of minutes that the j-th terminal represents inside the i-th corporate customer, associated with α_{2i}.

In addition, we investigated the random intercept mixed model, but in this case, one of the factors (factor 4) was not significant (p-value <0.05). In order to ensure no loss of information that could compromise the prediction and the comparison among the models, we considered the final mixed model given by (11).

We compare model (11) with a traditional model by considering the same set of covariates, and the fixed effect associated with z_{ij} was not significant (p-value >0.05). Therefore, the traditional model that could be comparable with (11) is given by

$$\text{logit}[P(y_{ij} = 1)] = \mathbf{x}_{ij}^t \boldsymbol{\beta}. \tag{12}$$

The methods considered for the estimation models (11) and (12) were Laplace and Maximum Likelihood. It is important to underscore that the PQL method did not achieve convergence in the estimation of parameters of model (11).

By considering the models (11) and (12), a significance test of the "fixed factors" was determined by ANOVA, considering Chi-squared statistics. The aim of the test was to identify the significance of a factor (in the fixed part of the model) given the others in the model. All of the factors were significant (p-value <0.05) in models (11) and (12).

In Table 6, we analyzed the estimate parameters, standard errors (s.e.), and p-values of models (11) and (12). The parameter estimates associated with fixed slopes presented positive signals, indicating that the usage of the additional services (SMS, roaming, e-mail, and data package) positively explains the target event, with the exception of β_6. Note that some categories of both models were not significant (p-value >0.05), but the decision of including the covariates was based on the ANOVA of nested models, rather than the test of individual categories.

The diagnosis of the mixed (11) and traditional (12) models was conducted by using normalized randomized quantile residuals: if the quantile residuals follow the normal distribution, by Dunn and Smyth (1996), then the model is correct. Rigby and Stasinopoulos (2005) used this type of residuals in order to compare different models of generalized additive models for location, scale and shape (GAMLSS), which include the mixed models.

Fig. 7 provides the diagnostic plots of a mixed model with two random effects (11). Fig. 7(c) and (d) provide QQ-plots of the empirical random intercept and slope, respectively, indicating some departure of each random effect from a normal distribution. This result is expected, as shown in Huang (2009), in which the authors suggest that the normality assumption may be unrealistic in some real applications. Neuhaus et al. (1992), McCulloch and Neuhaus (2011) and Neuhaus et al. (2012) use simulation studies to show that the misspecification of the random effects distribution for a mixed model is robust to the assumption of the lack of normality of the random effects. Thus, in this dataset, the residuals appear satisfactory, since the quantile residuals follow the normal distribution (e.g., Fig. 7(a) and (b)).

Fig. 8 provides the diagnostic plots of traditional model (12): (a) quantile residuals against index and (b) QQ-plot of the quantile residuals, which indicate that the quantile residuals follow the normal distribution.

5.4. Predictive performance

In this subsection, we evaluated the predictive performance of the traditional and mixed models performed in training and application datasets, which was based on the fitted models presented in Section 5.3.

The squared residuals (Md.SR) were not used because we did not know the observed probability, and in this case, we only used this measurement in the simulation studies.

In Table 7, we observed sensitivity, specificity, and KS values for the training and application datasets. In the training dataset, an analysis of all of the performance measures showed that the mixed model presented a higher index than the traditional model, indicating that the mixed model is more appropriate for the hierarchical structure.

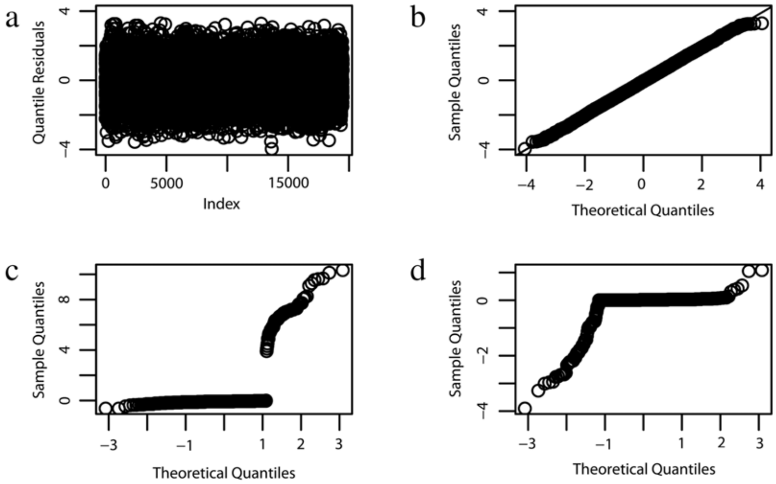

Fig. 7. Diagnostic plots for the mixed model with two random effects (11): quantile residuals against index, (b) QQ-plot of the quantile residuals, (c) QQ-plot of the random intercept, and (d) QQ-plot of the random slope.

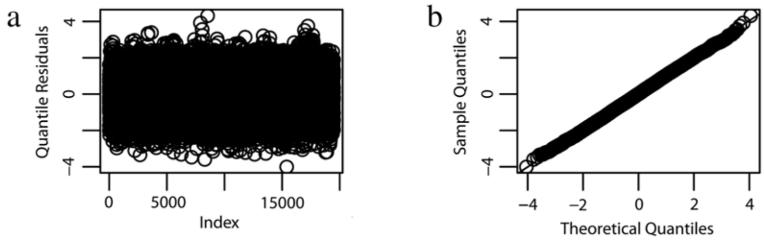

Fig. 8. Diagnostic plots for the traditional model: (a) quantile residuals against index, and (b) QQ-plot of the quantile residuals.

Table 7
Performance measures for the traditional and mixed models, in the training and the application datasets.

Performance measures	Traditional	Mixed	EBP	LRPM
Training dataset				
Sensitivity	56.09	99.57		
Specificity	82.95	87.77		
KS	39.04	87.34		
Application dataset				
Former customers				
Sensitivity	55.22	82.09		
Specificity	91.24	76.01		
KS	46.46	58.00		
New customers				
Sensitivity	42.86		62.86	65.70
Specificity	74.92		71.91	69.96
KS	17.77		34.77	37.90
All customers				
Sensitivity	49.58		78.39	69.92
Specificity	90.62		73.91	81.46
KS	40.19		52.30	51.37

Our challenge was to maintain the performance prediction in the future period, even if new clients appear. In the application dataset (December 2010), we did not know the individual random effects of the mixed model, and therefore EBP and LRPM were used to predict the outcome. In the EBP method, it was possible to predict the outcome by using Eq. (2), with the covariates and the estimate parameters presented in Table 6 for the mixed model.

On the other hand, in the LRPM, the covariates were aggregated at the customer level, since the rule depends on a set of covariates. For example, we may choose the mean as an aggregation function that was the sum of the service in the past of all terminals divided by the quantity of terminals of the company. The aggregation functions adopted in the final models used to summarize information from the terminal lines to the company level were mean, median, and maximum.

After the stepwise method, the set of covariates that explained the random effects aggregated at the customer level were as follows:

- Model for the random intercept: indicator of paying for SMS service monthly in the past, quantity of paying for SMS service per day, quantity of active terminals, minutes of calls received, value of the invoice, and indicator of paying for SMS per usage.
- Model for the random slope: indicator of paying for SMS service monthly in the past, quantity of paying for SMS service per day, and indicator of paying for SMS per usage.

In the application dataset (December 2010), we analyzed the performance measures using three different perspectives: only the former customers, only new customers, and all customers.

For the former customers, we used the estimate of random effects found in the training dataset. Note, the sensitivity, specificity, and KS values of the mixed models were higher than in the traditional model. We observed that the KS value was stable when we compared the former customer of the application dataset with the training dataset.

For new customers, we used the prediction methods EBP and LRPM. For the sensitivity and KS measurements, the prediction methods generated higher values than the traditional model.

Finally, by joining former and new customers in the application dataset, we concluded that the mixed model generated higher values of sensitivity, specificity, and KS than the traditional model. Compared to the mixed model, the traditional logistic regression presented unbalanced values of sensitivity and specificity. Thus, the mixed model combined with the prediction methods EBP and LRPM is adequate for a dataset with a hierarchical structure, regardless of the empirical random effects departing from a normal distribution.

6. Concluding remarks

The prediction of a future observation in a mixed regression is a problem that has been studied because of its utility in important applications, such as longitudinal studies, social science research, or direct marketing applications.

More recently, advances have been made in prediction methods for the models belonging to the exponential family, and in particular, the mixed logistic regression model. Although there is a prediction method of future observations in mixed logistic regression, when the model considers k random effects, the current methodology (EPB) can take extensive computational time.

In this article, we proposed the LRPM as a new method of predicting the outcome of the logistic mixed model. This methodology relaxed the assumption that the average of the random effect is attributed to the specific average for each group. In contrast to the LRPM, EBP considers the posterior expectations and uses the covariance matrix of the random effects with the vector of the means equal to zero.

Through simulation and application studies, EBP and LRPM can be used to predict the outcome for the mixed logistic model and provides better performance in prediction than traditional logistic regression. In addition, the LRPM provided slightly better results in terms of prediction than the EBP method for some measurements, such as KS and specificity.

In the direct marketing application, we considered the prediction in a future dataset by using the individual values of the random effects for the former customers and the prediction methods for new customers. We have concluded that EBP and LRPM are robust, even if the empirical random effects depart from a normal distribution.

Finally, by using the proposed methodology, we found that LRPM drastically reduced the computational effort and, at the same time, maintained the same level of prediction as the EBP method. Using the novel method (LRPM), one may predict the outcome with a similar processing time as the traditional logistic model.

For future work, the prediction methods used in this article might be extended for a different class of models. For instance, hierarchical generalized linear models (HGLM), proposed by Lee et al. (2006), allow for the assumption of different distributions of the random effects besides the normal distribution. Lombardía and Sperlich (2012) proposed a new class of semi-mixed effects model that considers that the random effects follow a non-parametric distribution, and determines the position of the model between the random model and fixed effects model. In addition, Esteban et al. (2012) presented the mixed model with time correlation.

In order to compare the prediction methods, the mean squared error prediction (MSEP) might be studied in the context of new groups, as introduced in Jiang (2007). In addition, it would be interesting to study the misspecification of the random effects distribution in the problem of prediction for new groups by considering EBP and LRPM. In the case of LRPM, an extension of the methodology might be multivariate linear regression that considers the correlation between the outcomes (random effects). An investigation involving non-parametrical models to predict the random effects, instead of a linear regression, will be the focus of future studies.

Acknowledgments

This work received partial financial support from FAPESP and CNPq, Brazil. We thank Rubens Stephan, Marcelo Sousa, Eduardo Taniguchi and Caio Soares for providing the dataset of mobile operators and helping us in the manipulation of the dataset used in the application study.

References

Breslow, N.E., Clayton, D.G., 1993. Approximate inference in generalized linear mixed models. J. Amer. Statist. Assoc. 88, 9–25.
Butera, A.M., 2000. Cross-selling: capitalizing on the opportunities. Hoosier Banker 84 (7), 14–16.
Collett, D., 2003. Modelling Binary Data, second ed. Chapman & Hall, London.
Conover, W.J., 1999. Practical Nonparametric Statistics. John Wiley and Sons, New York.
Demidenko, E., 2004. Mixed Models: Theory and Applications. In: Wiley Series in Probability and Statistics, New York.
Draper, N.R., Smith, H., 1998. Applied Regression Analysis, third ed. John Wiley and Sons, New York.
Dunn, P.K., Smyth, G.K., 1996. Randomised quantile residuals. J. Comput. Graph. Statist. 5, 236–244.
Engel, B., 1998. A simple illustration of the failure of PQL, IRREML and APHL as approximate ML methods for multilevel methods for binary data. Biom. J. 40, 141–151.
Esteban, M.D., Morales, D., Perez, A., Santamaria, L., 2012. Small area estimation of poverty proportions under area-level time models. Comput. Statist. Data Anal. 56 (10), 2840–2855.
Faraway, J., 2002. Practical regression and anova using R. http://www.stat.lsa.umich.edu/~faraway/book.
Hocking, R.R., 1976. The analysis and selection of variables in linear regression. Biometrics 32, 1–49.
Hosmer, D.W., Lemeshow, S., 2000. Applied Logistic Regression, second ed. John Wiley and Sons, New York.
Huang, X., 2009. Diagnosis of random-effect model misspecification in generalized linear mixed models for binary response. Biometrics 65, 361–368.
Jiang, J., 2003. Empirical best prediction for small-area inference based on generalized linear mixed models. J. Statist. Plann. Inference 111, 117–127.
Jiang, J., 2007. Linear and Generalized Linear Mixed Models and Their Applications. Springer, New York.
Jiang, J., Lahiri, P., 2001. Empirical best prediction for small area inference with binary data. Ann. Inst. Statist. Math. 53, 217–243.
Jiang, J., Lahiri, P., 2006. Mixed model prediction and small area estimation. Test 15, 1–96.
Lee, Y., Nelder, J.A., Pawitan, Y., 2006. Generalized Linear Models with Random Effects: Unified Analysis via H-Likelihood. Chapman & Hall, London.
Lombardía, J.M., Sperlich, S., 2012. A new class of semi-mixed effects models and its application in small area estimation. Comput. Statist. Data Anal. 56, 2903–2917.
McCulloch, C.E., Neuhaus, J.M., 2011. Misspecifying the shape of a random effects distribution: why getting it wrong may not matter. Statist. Sci. 26 (3), 388–402.
McCulloch, C.E., Searle, S.R., 2001. Generalized Linear, and Mixed Models. John Wiley and Sons, New York.
Neuhaus, J.M., Hauck, W.W., Kalbfleisch, J.D., 1992. The effects of mixture distribution misspecification when fitting mixed-effects logistic models. Biometrika 79 (4), 755–762.
Neuhaus, J.M., McCulloch, C.E., Boylan, R., 2012. Estimation of covariate effects in generalized linear mixed models with a misspecified distribution of random intercepts and slopes. Statist. Med. Early view.
Raudenbush, S.W., Bryk, A.S., 2002. Hierarchical Linear Models: Applications and Data Analysis Methods, second ed. Sage Publications, Newbury Park, CA.
Raudenbush, S.W., Yang, M., Yosef, M., 2000. Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate LA. J. Comput. Graph. Statist. 9 (1), 141–157.
R Development Core Team, 2010. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, ISBN: 3-900051-07-0.
Rencher, A., 1998. Multivariate Statistical Inference and Applications. John Wiley and Sons, New York.
Rigby, R., Stasinopoulos, D., 2005. Generalized additive models for location, scale and shape. Appl. Statist. 54 (3), 507–554.
Robbinson, G.K., 1991. That BLUP is a good thing: the estimation of random effects (with discussion). Statist. Sci. 6, 15–51.
Skrondal, A., Rabe-Hesketh, S., 2009. Prediction in multilevel generalized linear model. J. Roy. Statist. Soc. Ser. A 172, 659–687.
Tamura, K.A., Giampaoli, V., 2010. Prediction in mixed logistic regression. Comm. Statist. Simul. Comput. 39 (6), 1083–1096.
Tamura, K.A., Giampaoli, V., 2011. Prediction for an observation in a new cluster for mixed Logistic Regression considering k random coefficients. In: 26th International Workshop on Statistical Modelling, Vol. 1, pp. 593–596.