Quality & Quantity manuscript No.
(will be inserted by the editor)

Logistic regression analysis with multidimensional random effects: A comparison of three approaches

Olga Lukočienė · Jeroen K. Vermunt

Received: date / Accepted: date

Abstract This paper investigates the performance of three types of random coefficients logistic regression models; that is, models using parametric, semi-parametric, and nonparametric specifications of the distribution of the random effects. Whereas earlier studies focussed on models with a single random effect, here we look at models with multidimensional random effects (intercepts and slopes). Moreover, also the performance of a semi-parametric approach – using mixture regression models with number of latent classes is selected using the BIC – is investigated.

One of the main conclusions of our study is that the good results obtained with the nonparametric approach in the unidimensional case do not generalize to the multidimensional case. Parametric and semi-parametric approaches are much better in terms of bias and relative efficiency than the nonparametric approach. For the fixed-effects estimation, a parametric approach is the preferred method when the underlying assumption of the parametric model holds. In other situations, the semi-parametric approach is the best choice.

Keywords Mixed models · Hierarchical models · Finite mixture models · Multilevel logistic regression analysis · Random coefficients · EM algorithm · Nonparametric maximum likelihood

O. Lukočienė
Department of Methodology and Statistics Tilburg University P.O. Box 90153 5000 LE Tilburg The Netherlands.
Tel.: +31134662544
Fax: +31134663002
E-mail: o.lukociene@uvt.nl

J.K. Vermunt
Department of Methodology and Statistics Tilburg University P.O. Box 90153 5000 LE Tilburg The Netherlands.
Tel.: +31134662544
Fax: +31134663002
E-mail: j.k.vermunt@uvt.nl

1 Introduction

During the last decades, multilevel regression analysis has become part of the standard statistical toolbox of researchers in the social and behavioral sciences as well as in the biomedical field. This statistical method is used for the analysis of data sets in which lower-level units are nested within higher-level units (Hox, 2002; Skrondal and Rabe-Hesketh, 2004; Snijders and Bosker, 1999). Examples include data sets with a nesting of persons within families, survey respondents within geographical units, patients within therapists, pupils within schools, employees within firms, and repeated measurements within subjects. The lower level of the hierarchical structure is often referred to as level-1 and the higher level as level-2.

Typical for multilevel data is that level-1 observations belonging to the same level-2 unit are more alike than level-1 units from different level-2 units, for example, because they share common environments, experiences, and interactions. The implications of this is that the responses of level-1 units within the same level-2 units are correlated and can thus not be treated as independent observations in the statistical analysis. Whereas in some applications this is perceived as a problem that should be dealt with when modeling multilevel data, in other applications the multilevel data structure is seen as containing valuable information on how groups (higher-level units) differ from each other, for example, in terms of the effects of explanatory variables on the outcome variable of interest (Bryk and Raudenbush, 1992; Hox, 1994; Snijders and Bosker, 1999).

The most popular approach for the analysis of such data sets is by means of multilevel models, which are also referred to as hierarchical, mixed, random-effects, or random-coefficients models (Bryk and Raudenbush, 1992; Hox, 1994; Longford, 1995; Snijders and Bosker, 1999). Whereas the terms "multilevel" and "hierarchical" refer to the data structure, the terms "mixed", "random-effects" and "random-coefficients" indicate what these models are from a more technical point of view. More specifically, these models capture differences between level-2 units – and thus also correlations between level-1 observations within level-2 units – by allowing one or more of the model parameters to vary randomly across level-2 units. Whereas the earliest developments and applications of multilevel regression models concerned linear models for continuous responses, these are nowadays also applied with discrete response variables. The most popular model for binary responses is the random-effects logistic regression model (Breslow and Clayton, 1993; Wolfinger and O'Connell, 1993).

A key issue in the specification of a multilevel regression model is that not only assumptions have to be made about the distribution of the residuals, but also about the distribution of the random effects, also referred to as the mixing distribution. The most common approach is to assume that it has a convenient parametric form, in most cases a normal distribution. However, as stressed by Aitkin (1999), parametric distributional assumptions about the random effects will usually not hold in practice, which may have serious implications for the parameter estimates. For example, various studies found that misspecification of the distribution of random effects results in a loss of efficiency of the fixed coefficient estimates (Agresti *et al.*, 2004; Heagerty and Kurland, 2001; Maas and Hox, 2004; Neuhaus *et al.*, 1992). Lukočienė (2008) not only confirmed this result for the random-intercept logistic regression model, but also showed that the estimate for the random-intercept variance may be severely biased when its distribution is misspecified.

Rather than using a parametric random-effects approach, it is also possible to use either a nonparametric or a semi-parametric approach. These two alternatives have in common that they are both latent class models; that is, a discrete mixing distribution with K nodes (latent classes) is used to approximate the underlying distribution with an unknown shape. The locations and weights corresponding to the nodes are quantities to be estimated. Al-

though the nonparametric and semi-parametric approach are similar, they are fundamentally different in how they determine the number of latent classes. In the former, the number of latent classes is increased till the likelihood function is maximized, which yields what is called the nonparametric maximum likelihood (NPML) estimator of the random-effects distribution (Heckman and Singer, 1984; Laird, 1978; Lindsay, 1995). As indicated by Leroux (1992) and Leroux and Puterman (1992), the NPML estimate may yield unnecessarily large numbers of latent classes and well fitting models with fewer latent classes may be preferred. Rather than increasing the number of latent classes till a saturation point is reached, it is also possible to decide about the number of classes using information criteria such as AIC and BIC. Note that this is what is usually done in mixture regression analysis (Vermunt and van Dijk, 2001; Wedel and DeSarbo, 1994), as well as in other types of latent class analyses. To distinguish this approach from NPML, we call it a semi-parametric approach.

This paper provides a comparison of the three random-effect approaches within the context of multilevel logistic regression analysis. It extends the work by (Lukočienė, 2008) on the comparison of parametric and NPML approaches for random-effects logistic regression analysis to the situation in which not only the intercept but also slopes are random coefficients, as is usual in social and behavioral science application of multilevel regression analysis (Kreft and de Leeuw, 1998; Singer, 1998; Snijders and Bosker, 1999). As far as we know, there are no studies investigating the performance of the NMPL approach when applied with multidimensional random effects. Moreover, we include the semi-parametric approach in the comparison. This is the commonly used latent-class based regression modeling approach for situations in which not only the intercept but also the slopes vary randomly across level-2 units. We focus on binary logistic regression models because these are more sensitive to specification issues in multilevel analysis than models for continuous response variables or counts (Agresti *et al.*, 2000, 2004).

Using a simulation study we wish to find out which of the three approaches – parametric, semi-parametric or nonparametric – should be used under different types of true random-effects distributions and specific features of the sample. More specifically, we are interested in whether it makes sense to use a nonparametric or semi-parametric model as an alternative when the underlying assumptions of the parametric model do not hold? Moreover, we wish to know whether it harms to use a nonparametric or semi-parametric model – say for practical reasons – when the assumptions of the parametric model hold?

The next section describes the multilevel logistic regression model of interest. Section 3 discusses the set up of the simulation study. Results of the simulation study are presented in Section 4. The last section summarizes the main conclusions and provides some practical recommendations.

2 The two-level logistic regression model

This section describes the two-level logistic regression model using the single equation mixed model formulation (Skrondal and Rabe-Hesketh, 2004). An alternative would be to use the hierarchical model formulation, which contains separate regression equations for the various hierarchical levels (Bryk and Raudenbush, 1992; Hox, 2002; Snijders and Bosker, 1999).

Let y_{ij} denote the binary response ($y_{ij} = 0, 1$) of the level-1 unit i, $i = 1, \ldots, n_j$, belonging to the level-2 unit j, $j = 1, \ldots, n$. Explanatory variables are referred to by $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_{ij}$, where the former concern the fixed and the later the random effects. The vector with fixed effects is denoted by $\boldsymbol{\beta}$ and the vector with the unobservable common random coefficients shared

by all level-1 units belonging to the $j^{\underline{th}}$ level-2 unit by $\boldsymbol{u}_j$. Let $\pi_{ij} = E(y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}, \boldsymbol{u}_j)$ be the conditional expectation of y_{ij}. The multilevel logistic regression model for y_{ij} takes on the following form:

$$\log \frac{\pi_{ij}}{1-\pi_{ij}} = \boldsymbol{\beta}'\boldsymbol{x}_{ij} + \boldsymbol{u}_j'\boldsymbol{z}_{ij}. \tag{1}$$

The typical assumption for the random coefficients $\boldsymbol{u}_j$ is that these are independently and identically distributed multivariate normal random variables with zero means and covariance matrix Σ_u. Consistent with this distributional assumption, parameters of the two-level logistic regression model may be estimated by maximum likelihood (ML), where construction of the likelihood function is simplified by the fact that the y_{ij} can be assumed to be independent within level-2 units conditionally on the observed predictors and the unobserved random effects. ML estimation involves maximizing the following marginal likelihood function:

$$L(\boldsymbol{\beta}, \Sigma_u) = \prod_{j=1}^{n} \int_{\boldsymbol{u}_j} \left[\prod_{i=1}^{n_j} \pi_{ij}^{y_{ij}} (1-\pi_{ij})^{1-y_{ij}} \right] f(\boldsymbol{u}_j; \Sigma_u) d\boldsymbol{u}_j, \tag{2}$$

where $\pi_{ij}^{y_{ij}}(1-\pi_{ij})^{1-y_{ij}}$ represents the Bernoulli distribution for the level-1 errors. Note that the fixed effects $\boldsymbol{\beta}$ and covariance matrix Σ_u are the unknown parameters to be estimated. The integral should be solve numerically, for example, using Gauss-Hermite quadrature, which is basically a discrete approximation of the multivariate normal integral. Algorithms for maximizing the resulting numerically integrated marginal likelihood are the EM algorithm (Agresti *et al.*, 2000; Bock and Aitkin, 1981; Dempster *et al.*, 1977) and gradient methods, such as the Fisher scoring (Longford, 1987) and Newton-Raphson algorithm (Pan and Thompson, 2003; Rabe-Hesketh *et al.*, 2004). In our study, we used numerical integration with 50 nodes per dimension. For maximization a combination of EM and Newton-Raphson was used, where the estimation process starts with EM iterations and switches to Newton-Raphson when the relative change in parameters is very small (Vermunt and Magidson, 2005).

As was indicated in the introduction, usually nothing or very little it is known about the underlying distribution of the random effects (Aitkin, 1999). To prevent possible misspecification, it may therefore be attractive to assume the random effects $\boldsymbol{u}_j$ come from an unspecified mixing distribution concentrated on a finite number of latent classes or mass points (Aitkin, 1999; Heckman and Singer, 1984; Laird, 1978; Vermunt, 1997). Let K denote the number of latent classes, k a particular latent class, and $\boldsymbol{u}_k^*$ the unknown values of the random effects $\boldsymbol{u}_j$ when level-2 unit j belongs to latent class k, and let $\pi_k = P(\boldsymbol{u}_j = \boldsymbol{u}_k^*)$ represent the probability that a randomly selected level-2 unit belongs to latent class k or in other words that the random effects correspond to the location of class k. Using such a K-class discrete characterization of the random effects distribution yields the following marginal likelihood function:

$$L(\boldsymbol{\beta}, \boldsymbol{u}^*, \boldsymbol{\pi}) = \prod_{j=1}^{n} \sum_{k=1}^{K} \left[\prod_{i=1}^{n_j} \pi_{ij|k}^{y_{ij}} (1-\pi_{ij|k})^{1-y_{ij}} \right] \pi_k, \tag{3}$$

where $\pi_{ij|k}$ is the conditional density function of y_{ij} given that level-2 unit j belongs to latent class k. The two-level logistic regression model can now be written as a model for $\pi_{ij|k}$; that is,

$$\log \frac{\pi_{ij|k}}{1-\pi_{ij|k}} = \boldsymbol{\beta}'\boldsymbol{x}_{ij} + \boldsymbol{u}_k^{*\prime}\boldsymbol{z}_{ij}. \tag{4}$$

The weights are restricted such that $\pi_k > 0$ and $\sum_{k=1}^{K} \pi_k = 1$. In addition one identifying location constraint should be imposed on each of the $M+1$ random coefficients, e.g. $\sum_{k=1}^{K} u^*_{mk}\pi_k = 0$, which implies that the $\boldsymbol{u}^*_k = (u^*_{0k}, \ldots, u^*_{mk}, \ldots, u^*_{Mk})$ are centered. The unknown parameters to be estimated are the fixed effects $\boldsymbol{\beta}$, $K-1$ free mass point locations per dimension $(u^*_{0k}, \ldots, u^*_{Mk})$ and $K-1$ free mass point weights π_k. Note that although the variances and covariance of the random effects are not model parameters, they can easily be estimated as follows (Vermunt and van Dijk, 2001):

$$\hat{\sigma}^2_m = \sum_{k=1}^{K} (u^*_{mk})^2 \pi_k \quad \text{and} \quad \hat{\sigma}_{mm'} = \sum_{k=1}^{K} u^*_{mk} u^*_{m'k} \pi_k.$$

Maximization of the marginal likelihood function in equation (3) for a specific K, as in the parametric case, can be achieved by means of the EM and/or Newton-Raphson algorithm. It is usually advised to use of multiple sets of starting values to reduce the likelihood of ending up in a local maximum.

To obtain the solution corresponding to the NPML estimate of the random effects distribution, we not only have to maximize (3) for specific values of K; but we simultaneously have to find the value of K – say K_{NPML} – that yields the largest marginal likelihood value. In other words, we have to find the saturation point at which increasing K no longer results in an increase of the likelihood function. A method to find K_{NPML} proposed by various authors involves introducing latent classes one by one using directional (Gateaux) derivatives (Böhning, 2000; Lindsay, 1983, 1995; Rabe-Hesketh *et al.*, 2003). A much simpler alternative approach is to estimate the model with a large number of latent classes, K_{MAX}. When $K_{MAX} > K_{NPML}$, the ML estimates for $\boldsymbol{u}^*_k$ will be equal for some latent classes and/or the estimate for π_k will be equal to zero for some latent classes (Böhning, 2000). In other words, classes may be merged (equal $\boldsymbol{u}^*_k$) and/or removed (π_k equal to zero). To prevent local maxima this procedure should be repeated with several sets of starting values. Moreover, to guarantee that also the more difficult to find mass points located at $-\infty$ and $+\infty$ are encountered when needed in the NPML solution, it is advisable to include latent classes located at these values in each starting set (Hartzel *et al.*, 2001; Wood and Hinde, 1987).

As already mentioned, the NPML estimates may yield unnecessarily large K (Leroux, 1992; Leroux and Puterman, 1992) and estimates with a smaller number of latent classes that describe the data sufficiently may be preferred. Moreover, the latent classes may have substantive interpretations which are useful for the study concerned. This yields an approach in which the value of K should be estimated, yielding what we called the semi-parametric random-effects modeling approach. In this approach the value of K is increased till the criterion used for model selection no longer improves. In our study, we will use the BIC (Schwarz, 1978) for deciding about the number of classes, as was for example done by Vermunt and van Dijk (2001); Wedel and DeSarbo (1994).

3 Design of the simulation study

This section describes the design of the simulation study. First, we discuss the design factors that were kept constant, and subsequently the ones that were varied. The key factors that were kept contant are the overall structure of the population model, the values of the fixed-effect parameters, and the values of the intraclass correlations for the random effects.

The population model we used is a two-level random coefficients logistic regression model with one level-1 and one level-2 explanatory variable. This model can be formulated as follows:

$$\log \frac{\pi_{ij}}{1-\pi_{ij}} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2j} + u_{0j} + u_{1j} z_{1ij}. \tag{5}$$

Here, both x_{1ij} and z_{1ij} represent the level-1 predictor (in fact, $x_{1ij} = z_{1ij}$), where x_{1ij} is used to define its fixed part and z_{1ij} its random part. The other fixed effects correspond to the intercept and the level-2 predictor x_{2j}. The two explanatory variables are assumed to be binary predictors taking on the values 0 and 1 with probability 0.5 independently of one another. For the fixed intercept β_0 and slopes β_1 and β_2, we used the same values across simulation replications. More specifically, we set their values to: $\beta_0 = -2$, $\beta_1 = \beta_2 = 2$. This yields large enough but not too extreme differences between the response probabilities for $u_{0j} = 0$ and $u_{1j} = 0$. More specifically, the corresponding response probabilities for the four possible combinations of explanatory variables are

$$P(y = 1|x_1 = 1, x_2 = 1, u_0 = u_1 = 0) = e^2/(1+e^2) = 0.88,$$
$$P(y = 1|x_1 = 1, x_2 = 0, u_0 = u_1 = 0) = e^0/(1+e^0) = 0.5,$$
$$P(y = 1|x_1 = 0, x_2 = 1, u_0 = u_1 = 0) = e^0/(1+e^0) = 0.5$$

and

$$P(y = 1|x_1 = 0, x_2 = 0, u_0 = u_1 = 0) = e^{-2}/(1+e^{-2}) = 0.12.$$

A second element that was kept constant in the simulation study is the overall importance of the random part, which can be expressed by means of the intraclass correlation (*ICC*). Although Hox and Maas (2001) found that the value of the *ICC* may affect the impact of a misspecification of the random effects distribution, the study by Lukočienė (2008) on the random-intercept logistic regression model found that parametric and nonparametric approaches are almost indistinguible when the *ICC* value is small (e.g. 0.1). We will therefore not investigate this situation again, but instead focus on the condition with a moderate *ICC* value of 0.3. For this *ICC* value, Lukočienė (2008) found important differences in the performance of the parametric and nonparametric approaches.

The *ICC* values can be set by using the fact that level-1 errors coming from a logistic distribution have a variance equal to $\pi^2/3$. Since $ICC = \sigma^2/(\sigma^2 + \pi^2/3)$, the variance of the random intercept σ_0^2 can be obtained by $\sigma_0^2 = ICC/(1-ICC)\,\pi^2/3$, which for $ICC = 0.3$ yields $\sigma_0^2 = 1.41$. Similar to Busing (1993) and Maas and Hox (2004), we used the same variance for the random slope as for the random intercept ($\sigma_1^2 = 1.41$ as well).

So far, we discussed only the elements that were not varied in the simulations study. The three design factors that were varied are the random effects distribution, the level-1 sample size, and the level-2 sample size. We wish to assess how the parametric, nonparametric, and semi-parametric models perform under different true random-effect distributions and whether the performance depends on the level-1 and level-2 sample sizes. The study by Lukočienė (2008) on the random-intercept logistic regression model showed that these are the main factors affecting the performance of the parametric and nonparametric approaches.

Data sets were generated using four distributional forms for the random effects, two continuous distributions (exponential and normal) and two discrete mixing distributions – one five-class distribution with class membership probabilities of 0.1, 0.2, 0.4, 0.2, and 0.1, respectively, and another four-class distribution with equal membership probabilities of 0.25. As demonstrated in Figure 1 and Figure 2, the locations of the classes of these two discrete mixing distribution were chosen in such a way that the random intercept and random slope would be uncorrelated, but strongly associated. With these four choices we have apart from the normal distribution, distributions that considerably deviate from normal in terms of skewness, kurtosis, discontinuity, and association between dimensions.

The other two factors that were varied are the level-1 and level-2 sample sizes. More specifically, for the number of level-2 units we used $n = 30$, 100, and 1000 and for the

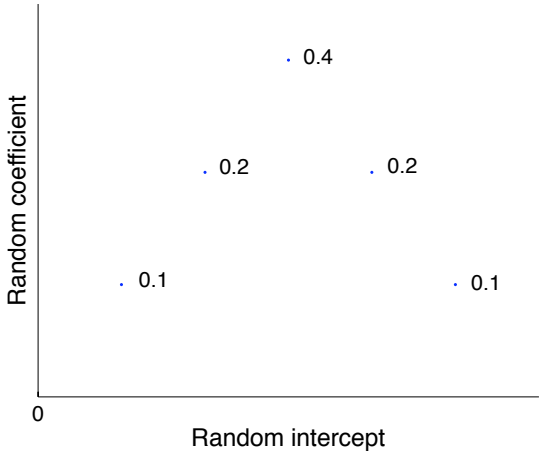

Fig. 1 Discrete mixing distribution with five classes

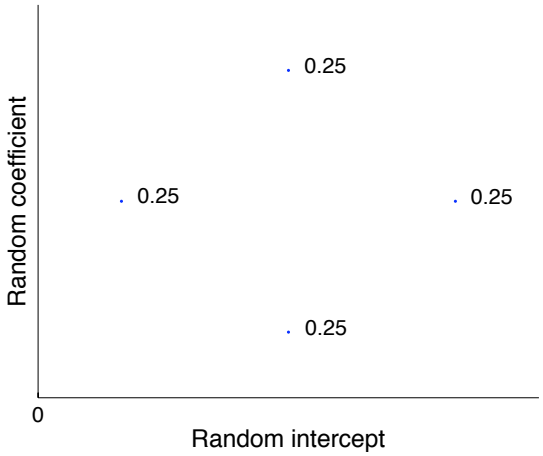

Fig. 2 Discrete mixing distribution with four classes

number of level-1 units $n_j = 10$, and 50. These sample sizes reflect the typical sample sizes in multilevel analysis (see also Kreft and de Leeuw (1998); Maas and Hox (2004); Lukočienė (2008)).

Combining the 3 design factors – distributional form, level-2 sample size, and level-1 sample size – yielded a total of $4 \times 2 \times 3 = 24$ conditions. We generated 1000 simulated data sets for each of these conditions. For each simulated data set, the unknown model parameters were estimated using the parametric approach assuming that random effects come from a normal distribution, the NPML approach, and the semi-parametric approach using BIC as the model selection criterion.

4 Results of the simulation study

The aim of the simulation study was to determine the bias and relative efficiency of the parametric, nonparametric, and semi-parametric random effects approaches under the different true random effects distributions and sample sizes. Let θ be one of the parameters of interest, which in our case are the fixed effects β_0, β_1, and β_2, and the standard deviations of the random effects distribution σ_0 and σ_1 which in the nonparametric and semi-parametric cases are computed from the nodes' locations and weights. The ML estimate of θ obtained in replication s, $s = 1, \ldots, 1000$, is denoted by $\widehat{\theta}_s$. Rather than using the more standard definitions of bias and relative efficiency – $E(\widehat{\theta}_s - \theta)$ and $E\left[(\widehat{\theta}_s - \theta)^2\right]$ – we used a more robust definition to prevent that the results are affected by a small number of replications with boundary estimates. More specifically, when using the NPML estimator, especially in the conditions with large number of level-2 units and small number of level-1 units, there is a positive probability that one of the latent classes is located at infinity. In our case, latent classes can have 4 such possible locations: $(-\infty, -\infty)$, $(-\infty, \infty)$, (∞, ∞), and/or $(\infty, -\infty)$. When such boundary estimates may occur $E(\widehat{\theta}_s - \theta)$ and $E\left[(\widehat{\theta}_s - \theta)^2\right]$ do not exist. This not only applies to σ_0 and σ_1, but also to β_0, β_1, and β_2. To prevent this problem from occurring we define bias as the median of $(\widehat{\theta}_s - \theta)$ and relative efficiency as the median of $|\widehat{\theta}_s - \theta|$. For similar approaches, see Agresti *et al.* (2004) and Galindo-Garre *et al.* (2004).

Below we first discuss the results for the fixed effects and then for the random effects.

4.1 Fixed effects

The first evaluation criterion of interest is the bias in the parameter estimates. Table 1 provides the estimated biases of the fixed effects for the level-2 sample sizes of 1000 and 30. The first three columns of Table 1 indicate the values for the design factors: level-2 sample size, level-1 sample size, and the random-effects distribution used to generate the data sets. The fourth column indicates which of the three approaches – parametric, nonparametric or semi-parametric – was used for the estimation of the parameters. The last three columns present the biases in the estimates of the intercept and the two slopes. Reported biases are marked by a "*" when they are larger than 5% of the true parameter value, and smaller values are considered negligible.

Table 1 shows that the bias of the fixed effects in the models estimated under the semi-parametric approach is negligible for all cases with true discrete and exponential underlying distributions, except for the fixed effect β_1 when $n = 30$ and true underlying discrete distribution has 5 latent classes. When the semi-parametric approach is applied with the true bivariate normal distribution, we see biases only for β_2. The NPML approach yields biased estimates for almost every parameter, where it makes no difference whether the true distribution is discrete or continuous. The parametric approach performs very well when the true underlying distribution is bivariate normal, in which case the bias in the fixed effects is always negligible. However, for other true underlying distributions, the parametric approach gives biased estimates for at least one of the fixed effects. The results for the medium level-2 unit sample size $n = 100$ (which are not shown) are very similar to the results obtained with $n = 1000$.

As was mentioned above, the second evaluation criterion of interest is the efficiency of the parameter estimates. Table 2 reports results on relative efficiency of the fixed effects obtained with the largest and smallest level-2 unit sample sizes $n = 1000$ and $n = 30$ (results

Table 1 Bias of the fixed effects for the conditions $n = 1000$ and 30

n	n_j	True distribution	Assumed	$\hat{\beta}_{0s} - \beta_0$	$\hat{\beta}_{1s} - \beta_1$	$\hat{\beta}_{2s} - \beta_2$
1000	10	Exponential	Normal	-0.12*	-0.03	0.04
			Nonparametric	-0.09	0.90*	0.04
			Semi-parametric	0.04	-0.09	-0.08
		Normal	Normal	0.00	0.01	0.03
			Nonparametric	-0.17*	0.28*	0.05
			Semi-parametric	0.06	-0.07	-0.06
		Discrete (4 classes)	Normal	-0.02	0.00	0.04
			Nonparametric	-0.11*	0.24*	0.05
			Semi-parametric	0.00	-0.01	-0.01
		Discrete (5 classes)	Normal	0.07	-0.21*	0.02
			Nonparametric	-0.02	0.13*	0.01
			Semi-parametric	0.00	-0.01	0.00
1000	50	Exponential	Normal	-0.04	-0.05	-0.03
			Nonparametric	-0.01	0.40*	0.01
			Semi-parametric	0.04	-0.04	-0.05
		Normal	Normal	0.00	0.00	0.02
			Nonparametric	-0.01	0.15*	0.03
			Semi-parametric	0.08	-0.03	-0.11*
		Discrete (4 classes)	Normal	-0.06	0.04	-0.24*
			Nonparametric	-0.01	0.13*	0.02
			Semi-parametric	0.00	0.00	0.01
		Discrete (5 classes)	Normal	0.05	-0.21*	0.05
			Nonparametric	-0.02	0.16*	0.01
			Semi-parametric	-0.01	0.00	0.00
30	10	Exponential	Normal	-0.10*	-0.02	0.01
			Nonparametric	-0.58*	2.35*	0.29*
			Semi-parametric	0.06	0.03	-0.06
		Normal	Normal	-0.02	0.07	0.02
			Nonparametric	-0.94*	2.46*	0.32*
			Semi-parametric	0.096	0.01	-0.18*
		Discrete (4 classes)	Normal	0.02	0.05	-0.03
			Nonparametric	-0.88*	2.19*	0.28*
			Semi-parametric	0.08	-0.02	-0.04
		Discrete (5 classes)	Normal	0.105*	-0.16*	-0.08
			Nonparametric	-0.16*	2.09*	0.02
			Semi-parametric	0.01	0.11*	0.00
30	50	Exponential	Normal	-0.06	-0.04	-0.08
			Nonparametric	-0.24*	1.59*	-0.14*
			Semi-parametric	0.05	-0.03	-0.01
		Normal	Normal	0.00	0.04	-0.09
			Nonparametric	-0.26*	0.92*	-0.34*
			Semi-parametric	0.07	0.01	-0.16*
		Discrete (4 classes)	Normal	-0.07	0.07	-0.15*
			Nonparametric	-0.15*	0.22*	0.04
			Semi-parametric	0.00	0.01	0.01
		Discrete (5 classes)	Normal	0.10*	-0.16*	-0.08
			Nonparametric	-0.17*	2.20*	0.01
			Semi-parametric	0.01	0.14*	0.00

* Cases with medians absolute value over 5%.

for $n = 100$ are again similar to the ones for $n = 1000$). Table 2 can be read similarly to Table 1.

The semi-parametric approach clearly outperforms the parametric and nonparametric approaches in cases when the true underlying distribution of random effects is discrete. However, when the true underlying random effects distribution is bivariate normal the para-

Table 2 Efficiency of the fixed effects for the conditions $n = 1000$ and 30

n	n_j	True distribution	Assumed	$\lvert\hat{\beta}_{0s} - \beta_0\rvert$	$\lvert\hat{\beta}_{1s} - \beta_1\rvert$	$\lvert\hat{\beta}_{2s} - \beta_2\rvert$
1000	10	Exponential	Normal	0.12	0.05	0.067
			Nonparametric	0.13	0.93	0.08
			Semi-parametric	0.05	0.11	0.070
		Normal	Normal	0.06	0.05	0.06
			Nonparametric	0.18	0.87	0.070
			Semi-parametric	0.07	0.07	0.071
		Discrete (4 classes)	Normal	0.06	0.07	0.09
			Nonparametric	0.11	0.30	0.08
			Semi-parametric	0.04	0.05	0.07
		Discrete (5 classes)	Normal	0.08	0.21	0.10
			Nonparametric	0.04	0.15	0.023
			Semi-parametric	0.03	0.04	0.021
1000	50	Exponential	Normal	0.06	0.05	0.07
			Nonparametric	0.04	0.40	0.04
			Semi-parametric	0.05	0.06	0.06
		Normal	Normal	0.04	0.04	0.07
			Nonparametric	0.05	0.23	0.09
			Semi-parametric	0.09	0.05	0.13
		Discrete (4 classes)	Normal	0.06	0.05	0.24
			Nonparametric	0.03	0.04	0.03
			Semi-parametric	0.02	0.02	0.02
		Discrete (5 classes)	Normal	0.07	0.21	0.10
			Nonparametric	0.04	0.16	0.021
			Semi-parametric	0.03	0.03	0.020
30	10	Exponential	Normal	0.29	0.27	0.388
			Nonparametric	0.80	3.07	0.50
			Semi-parametric	0.32	0.39	0.391
		Normal	Normal	0.29	0.29	0.38
			Nonparametric	1.08	3.94	0.59
			Semi-parametric	0.36	0.34	0.47
		Discrete (4 classes)	Normal	0.32	0.322	0.44
			Nonparametric	0.96	3.51	0.51
			Semi-parametric	0.31	0.320	0.36
		Discrete (5 classes)	Normal	0.27	0.23	0.35
			Nonparametric	0.28	2.13	0.16
			Semi-parametric	0.21	0.44	0.14
30	50	Exponential	Normal	0.27	0.20	0.36
			Nonparametric	0.37	1.85	0.38
			Semi-parametric	0.35	0.38	0.35
		Normal	Normal	0.27	0.21	0.41
			Nonparametric	0.45	1.64	0.58
			Semi-parametric	0.30	0.23	0.48
		Discrete (4 classes)	Normal	0.29	0.23	0.37
			Nonparametric	0.27	0.44	0.15
			Semi-parametric	0.18	0.18	0.14
		Discrete (5 classes)	Normal	0.28	0.23	0.34
			Nonparametric	0.29	2.22	0.16
			Semi-parametric	0.21	0.46	0.14

metric approach is most efficient. For the true underlying exponential distribution, the parametric and semi-parametric approaches perform equally well in terms of efficiency. We find a considerable lower efficiency under the nonparametric approach for almost every condition.

Table 3 Bias and efficiency of the random effects for the conditions $n = 1000$ and 30

n	n_j	True distribution	Assumed	$\hat{\sigma}_{0s} - \sigma_0$	$\hat{\sigma}_{1s} - \sigma_1$	$\lvert\hat{\sigma}_{0s} - \sigma_0\rvert$	$\lvert\hat{\sigma}_{1s} - \sigma_1\rvert$
1000	10	Exponential	Normal	0.35*	-0.17*	0.35	0.17
			Nonparametric	0.57*	6.15*	0.57	6.15
			Semi-parametric	-0.12*	-0.26*	0.13	0.32
		Normal	Normal	0.26*	0.28*	0.26	0.28
			Nonparametric	1.20*	6.59*	1.20	6.59
			Semi-parametric	-0.09*	-0.17*	0.10	0.18
		Discrete (4 classes)	Normal	0.29*	0.74*	0.29	0.74
			Nonparametric	0.16*	2.26*	0.16	2.26
			Semi-parametric	0.02	0.00	0.07	0.07
		Discrete (5 classes)	Normal	-0.02	-0.39*	0.07	0.39
			Nonparametric	0.03	2.25*	0.03	2.25
			Semi-parametric	-0.01	0.00	0.03	0.04
1000	50	Exponential	Normal	0.13*	-0.26*	0.13	0.26
			Nonparametric	-0.02	3.30*	0.07	3.30
			Semi-parametric	-0.05	-0.13*	0.06	0.16
		Normal	Normal	0.11*	0.18*	0.11	0.18
			Nonparametric	0.02	1.90*	0.05	1.90
			Semi-parametric	-0.05	-0.08*	0.04	0.08
		Discrete (4 classes)	Normal	0.24*	0.89*	0.24	0.89
			Nonparametric	0.02	0.09*	0.03	0.09
			Semi-parametric	0.00	0.01	0.02	0.02
		Discrete (5 classes)	Normal	-0.04	-0.39*	0.08	0.39
			Nonparametric	0.01	2.81*	0.04	2.81
			Semi-parametric	-0.01	0.00	0.03	0.04
30	10	Exponential	Normal	0.15*	-0.23*	0.61	0.81
			Nonparametric	2.45*	9.10*	2.45	9.10
			Semi-parametric	-0.18*	-0.47*	0.35	0.88
		Normal	Normal	0.05	0.29*	0.54	0.83
			Nonparametric	3.40*	10.69*	3.40	10.69
			Semi-parametric	-0.22*	-0.42*	0.39	0.75
		Discrete (4 classes)	Normal	0.14*	0.90*	0.57	1.06
			Nonparametric	2.78*	9.50*	2.78	9.50
			Semi-parametric	-0.27*	-0.33*	0.37	0.48
		Discrete (5 classes)	Normal	0.23*	-0.40*	0.42	0.52
			Nonparametric	0.18*	7.75*	0.24	7.75
			Semi-parametric	-0.05	0.12*	0.16	0.51
30	50	Exponential	Normal	0.25*	-0.24*	0.49	0.50
			Nonparametric	0.39*	6.87*	0.51	6.69
			Semi-parametric	-0.11*	-0.13*	0.26	0.47
		Normal	Normal	0.23*	0.39*	0.39	0.55
			Nonparametric	0.95*	6.35*	0.95	6.35
			Semi-parametric	-0.08*	-0.10*	0.18	0.28
		Discrete (4 classes)	Normal	0.37*	0.90*	0.45	0.91
			Nonparametric	0.14*	0.72*	0.21	0.72
			Semi-parametric	-0.05	0.00	0.15	0.14
		Discrete (5 classes)	Normal	0.22*	-0.41*	0.41	0.65
			Nonparametric	0.21*	7.93*	0.26	7.93
			Semi-parametric	-0.04	0.14*	0.16	0.51

* Cases with medians absolute value over 5%.

If we have a closer look at results presented in Table 1 and Table 2 from the perspective of the effects of the level-1 and level-2 sample sizes, it can be observed is that the nonparametric approach perform very bad with the smaller level-2 sample size, and this is enforced when also the level-1 sample size is small. The quality of the other two approaches

is less strongly affected by the sample sizes. However, when misspecified, the normal model performs worse when the level-1 sample size increases.

4.2 Random effects

Table 3 shows the results on bias and relative efficiency for the random effects obtained with sample sizes $n = 1000$ and $n = 30$. As in Table 1, biases larger than 5% of the true parameter value are marked by a "*". The semi-parametric approach yields negligible biases for both random effects when the true underlying distribution is discrete and the sample size is 1000. The parametric approach yields moderate biases for almost every condition. However, the obtained biases of the parametric estimates with true continuous underlying distributions in the smallest samples ($n = 30$ and $n_j = 10$) are smaller than for the semi-parametric and nonparametric estimates. In most other cases, the semi-parametric approach performs best showing the smallest bias for all true distributions.

The last three columns of Table 3 report the information on the efficiency of the random effects estimates (of the standard deviations of the random effects). For the random intercept, the semi-parametric approach outperforms the parametric and nonparametric approaches in all investigated conditions. The same applies to the random slope, except for one situation; that is, when $n_j = 10$ and the true underlying distribution is exponential, the parametric approach is the most efficient method.

Results on bias and relative efficiency of random effects for the medium level-2 sample size ($n = 100$) are again not presented because they are rather similar to the results obtained with $n = 1000$.

4.3 Remarks on semi-parametric and nonparametric approaches

The results reported in Tables 1, 2, and 3 show that the nonparametric approach performs worse than the semi-parametric approach in almost all investigated conditions. As explained earlier, the difference between these two approaches is that they use different methods for determining the number of latent classes. To see how the use of BIC worked out in our simulation study, let us take a look at the number of latent classes selected according to this criterion when the true distribution is discrete. More specifically, Table 4 presents the percentage of simulation replications (out of 1000) in which a particular number of latent classes was selected using the semi-parametric approach. As can be seen, the number of latent classes is often underestimated with the smaller level-2 sizes, and this tendency is stronger when also the level-1 sample size is small. It can also be observed that the semi-parametric specification never overestimates the number of latent classes, which confirms that BIC is a somewhat conservative measure when deciding about the number of classes (see, for example, Dias (2004)).

As indicated earlier, in the nonparametric approach one increases the number of classes till a saturations point in reached, which seemingly lead to severely biased and much less efficient estimates. The NPML solution often consisted of a larger number of latent classes than the true discrete distribution even for the smallest level-2 sample size of 30. Such solutions contained nodes with small weights but very extreme locations, with explains the bias and inefficiency of this approach. In contrast, the semi-parametric approach will not accept such classes in the final solution because they do not yield a significantly better description of the data according to the BIC.

Table 4 Percentage of replications selecting a particular number of latent classes based on BIC in semi-parametric approach

n	n_j	True distribution	2 classes	3 classes	4 classes	5 classes
1000	50	Discrete with 4 classes			100	
		Discrete with 5 classes				100
	10	Discrete with 4 classes			100	
		Discrete with 5 classes				100
100	50	Discrete with 4 classes		19	81	
		Discrete with 5 classes		1	70	29
	10	Discrete with 4 classes	79	20	1	
		Discrete with 5 classes		1	70	29
30	50	Discrete with 4 classes	1	37	62	
		Discrete with 5 classes	5	34	58	3
	10	Discrete with 4 classes	90	9	1	
		Discrete with 5 classes	3	36	58	3

5 Conclusions

The two questions that we wished to answer based on the simulation study are 1) whether the NPML and/or semi-parametric approaches perform better in terms of bias and efficiency compared to the parametric model when the latter is misspecified, and 2) whether the NPML and/or semi-parametric approaches perform equally well in terms of bias and efficiency compared to the parametric model when the latter is correctly specified. This was studied for small and large level-1 and level-2 sample sizes and different types of random effects distributions (with a moderate ICC value). We are now able to answer these two questions for the two-level logistic regression model.

Our study showed that the NPML method gives the worst results in terms of bias and relative efficiency when compared to the parametric and semi-parametric methods, and this applies irrespective of the true random effects distribution. The semi-parametric approach performs best when the true underlying distribution of random effects is discrete. When the assumptions of the parametric model hold, the parametric approach is the best for the fixed effects estimation, but the semi-parametric approach is the preferred one for the random effects estimation. When the true distribution is exponential (continuous but not normal), the parametric model is still preferred with a small level-1 sample size, but the semi-parametric model is better with a larger level-1 sample size.

We may finally compare our conclusions with those derived from the study by Lukočienė (2008) on multilevel logistic regression with only a random intercept. One important difference concerns the performance of the NPML method. Whereas this earlier study found that the NPML approach performs rather well as long as the level-1 sample size is not too small, here we have to conclude that it is by far the worst approach. In fact, the NPML method should not be used with multidimensional random effects. Another new element compared to this earlier study is that we also looked at the semi-parametric method which turned out to perform much better than the NPML method. As far as the parametric approach is concerned, similarly to the previous study it can be concluded that it is the preferred method when the normal distribution assumption holds, as well as when the distribution is continuous but not normal and the level-1 sample size is small.

One limitation of our study is that it concerned two-level regression models, and it is not clear whether our findings can be generalized to models containing more hierarchical lev-

els. Another limitation is that we focussed on models for binary responses. The suggestion for the future research would be to look at other models from the generalized linear modeling family, as well as at models with more than two levels; that is, at the class of models described by (Vermunt, 2004).

In our study we investigated three different specifications for the random effects distribution: a parametric approach with an underlying normal distribution, as well as nonparametric and semi-parametric approaches using an unspecified discrete mixing distribution. As a possible alternative one may use a combination of these, namely a finite mixture of normal distributions (Magder and Zeger, 1996; Verbeke and Molenberghs, 2000). Whereas such an approach may have particular advantages, such as that contrary to the nonparametric and semi-parametric approaches it yields nondiscrete random effects, Agresti *et al.* (2004) obtained somewhat disappointing results with this approach in the context of a log linear model for an odds ratio. Nevertheless, we believe that this hybrid approach may be promising in other situations, especially when the aim of the study is to obtain interpretable latent classes (Magidson and Vermunt, 2007) .

References

Agresti, A., Booth, J.G., Hobert, J.P., Caffo, B., Random-effects modeling of categorical response data, Sociological Methodology 30, 27–80 (2000).

Agresti, A., Caffo, B., Ohman-Strickland, P., Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies, Computational Statistics and Data Analysis 47, 639–653 (2004).

Aitkin, M., A general maximum likelihood analysis of variance components in generalized linear models, Biometrics 55, 117–128 (1999).

Bock, R.D., Aitkin, M., Marginal maximum likelihood estimation of item parameters, Psychometrika 46, 443–459 (1981).

Böhning, D., Computer-assisted analysis of mixtures and applications. Meta-analysis, disease mapping and others. London: Chapman & Hall (2000).

Breslow, N.E., Clayton, D.G., Approximate inference in generalized linear mixed models, Journal of the American Statistical Association 88, 9–25 (1993).

Bryk, A.S., Raudenbush, S.W., Hierarchical linear models. Newbury Park: Sage (1992).

Busing, F., Distribution characteristics of variance estimates in two-level models, Unpublished manuscript. Department of Psychometrics and Research Methodology, Leiden University (1993).

Dempster, A.P., Laird, N.M., Rubin, D.B., Maximum likelihood estimation from incomplete data via the EM algorithm, Journal of the Royal Statistical Society 39(1), 1–38 (1977).

Dias, J.G., Finite mixture models: review, applications and computer intensive methods. Doctoral dissertation, SOM Research School, University of Groningen, Netherlands (2004).

Galindo-Garre, F., Vermunt, J.K., Bergsma, W., Bayesian posterior estimation of logit parameters with small samples, Sociological Methods and Research 39(33), 88–117 (2004).

Hartzel, J., Agresti, A., Caffo, B., Multinomial logit random effects models, Statistical Modelling 1(2), 81–102 (2001).

Heagerty, P.J., Kurland, B.F., Misspecified maximum likelihood estimates and generalized liner mixed models, Biometrica 88, 973–985 (2001).

Heckman, J.J., Singer, B., A method for minimizing the impact of distributional assumptions in econometric models of duration, Econometrica 52, 271–320 (1984).

Hox, J., Multilevel analysis: techniques and applications. Mahwah, NJ: Lawrence Erlbaum (2002).

Hox, J.J., Applied multilevel analysis. Amsterdam: TT (1994).

Hox, J.J., Maas, C.J.M., The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples, Structural Equation Modeling 8, 157–174 (2001).

Kreft, I.G.G., de Leeuw, J., Introducing multilevel modeling. Sage, Newbury Park, CA (1998).

Laird, N., Nonparametric maximum likelihood estimation of a mixture distribution, Journal of the American Statistical Association 73, 805–811 (1978).

Leroux, B.G., Maximum likelihood estimation for hidden Markov models. Stoch. Proc. and their appl. 40, 127–143 (1992).

Leroux, B.G., Puterman, M.L., Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models, Biometrics 48, 545–558 (1992).

Lindsay, B.G., The geometry of mixture likelihoods: a general theory, The Annals of Statistics 11, 86–94 (1983).

Lindsay, B.G., Mixture models: theory, geometry and applications. NSF-CBMS Regional Conference Series in Probability and Statistics, vol.5. Hayward, CA: Institute of Mathematical statistics (1995).

Longford, N.T., A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects, Biometrika 74, 817–827 (1987).

Longford, N.T., Random coefficient models. Oxford: Clarendon (1995).

Lukočienė, O., Vermunt, J.K., A Comparison of multilevel logistic regression models with parametric and nonparametric random intercepts, Manuscript submitted for publication (2008).

Maas, C.J.M., Hox, J.J., The influence of violations of assumptions on multilevel parameter estimates and their standard errors, Computational Statistics and Data Analysis 46, 427–440 (2004).

Magder, L.S., Zeger, S.L., A smooth nonparametric estimate of mixing distribution using mixtures of Gaussians, Journal of the American Statistical Association 91, 1141–1151 (1996).

Magidson, J., Vermunt, J.K., Use of a random intercept in latent class regression models to remove response. Bulletin of the International Statistical Institute, 56th Session, paper 1604, 1–4 (2007).

Neuhaus, J.M., Hauck, W.W., Kalbfleisch, J.D., The effects of mixture distribution misspecification when fitting mixed effects logistic models, Biometrika 79, 755–762 (1992).

Pan, J.X., Thompson, R., Gauss-Hermite quadrature approximation for estimationin in generalised linear mixed models, Computational Statistics 18, 57–78 (2003).

Rabe-Hesketh, S., Pickles, A., Skrondal, A., Correcting for covariate measurement error in logistic regression using nonparametric maximum likelihood estimation, Statistical Modelling 3, 215–232 (2003).

Rabe-Hesketh, S., Pickles, A., Skrondal, A., Generalized multilevel structural equation modeling, Psychometrica 69, 167–190 (2004).

Schwarz, G., Estimating the dimension of a model, The Annals of Statistics 6(2), 461–464 (1978).

Singer, J.D., Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models, Journal of Educational and Behavioral Statistics 24, 323–355 (1998).

Skrondal, A., Rabe-Hesketh, S., Generalized latent variables modeling: multilevel, longitudinal, and structural equation models. Boca Raton, FL: Chapman & Hall/CRC (2004).

Snijders, T.A.B., Bosker, R.J., Multilevel analysis. London: Sage Publications (1999).
Verbeke, G., Molenberghs, G., Linear mixed models for longitudinal data. Springer, Berlin (2000).
Vermunt, J.K., Log-linear models for event histories. Advanced Quantitative Techniques in the Social Sciences Series 8. Sage Publications (1997).
Vermunt, J.K., van Dijk, L., A nonparametric random-coefficients approach: the latent class regression model, Multilevel Modelling Newsletter 13, 6–13 (2001).
Vermunt, J.K., An EM algorithm for the estimation of parametric and nonparametric hierarchical nonlinear models, Statistica Neerlandica 58, 220–233 (2004).
Vermunt, J.K., Magidson, J., Technical guide to Latent GOLD: basic and advanced. Belmont, MA: Statistical Innovations Inc (2005).
Wedel, M., DeSarbo, W.S., A review of recent developments in latent class regression models. in Advanced Methods of Marketing Research, R.P. Bagozzi, ed. Cambridge: Blackwell Publishers, 352–388 (1994).
Wolfinger, R., O'Connell, M., Generalized linear mixed models: a pseudo-likelihood approach, Journal of Statistical Computation and Simulation 48, 233–243 (1993).
Wood, A., Hinde, J., Binomial variance component models with a non-parametric assumption concerning random effects. In: Crouchley R, ed. Longitudinal data analysis. Avebury, Aldershot: Hants, 110–128 (1987).