Specification of random effects in multilevel models: a review

Post-Print copy of Grilli L., Rampichini C. (2015) Specification of random effects in multilevel models: a review. Quality & Quantity. 49 (3), pp 967-976.

Specification of random effects in multilevel models: a review

Leonardo Grilli and Carla Rampichini
Dipartimento di Statistica, Informatica, Applicazioni "G. Parenti"
Univerisità di Firenze
grilli@disia.unifi.it, rampichini@disia.unifi.it

Abstract: The analysis of highly structured data requires models with unobserved components (random effects) able to adequately account for the patterns of variances and correlations. The specification of the unobserved components is a key and challenging task. In this paper, we first review the literature about the consequences of misspecifying the distribution of the random effects and the related diagnostic tools; we then outline the main alternatives and generalizations, also considering some issues arising in Bayesian inference. The relevance of suitably structuring the unobserved components is illustrated by means of an application exploiting a model with heteroscedastic random effects.

Keywords: finite mixture; heteroscedasticity; misspecification; mixed model; prior distribution

1 Introduction

Random effects models are a key tool for the analysis of multilevel data in a wide range of fields. These models are also known as mixed (Demidenko 2013) or multilevel models (Raudenbush and Bryk 2002, Goldstein 2011, Snijders and Bosker 2012). Multilevel data can also be analysed with methods avoiding random effects, such as fixed effects models, marginal models via generalized estimating equations (GEE), and cluster-robust standard errors derived from sandwich estimators or bootstrap (Scott et al. 2013). However, in this review we explicitly focus on the random effects approach.

Multilevel structures include both hierarchical cross-sectional data and repeated measurements, thus our review is not limited to a specific setting. The case study concerns the assessment of school effectiveness (Grilli and Rampichini 2009), where multilevel models aim to determine the contribution of each school to the achievement of its students, usually measured through test scores. The random effects are interpreted as value-added measures, and their predictions are used to compare the schools (Leckie and Goldstein 2009).

To introduce the terminology, let us consider the simple case of a linear random intercept model:

$$y_{ij} = \boldsymbol{\beta}'\mathbf{x}_{ij} + \boldsymbol{\gamma}'\mathbf{z}_j + u_j + e_{ij} \tag{1}$$

where y_{ij} is the response variable for the i-th level 1 unit of the j-th level 2 unit (cluster), $\mathbf{x}_{ij}$ is the vector of level 1 covariates, $\mathbf{z}_j$ is the vector of level 2 covariates, e_{ij} are level 1 errors, and u_j are level 2 errors or random effects. The vectors $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ collect the coefficients of the covariates, also called fixed effects.

The level 1 and level 2 errors are assumed independent, with level 1 errors following a normal distribution with variance σ_e^2. In this review we focus on the assumptions for the random effects u_j, whose distributions is usually assumed to be

$$u_j \overset{iid}{\sim} N(0, \sigma_u^2) \tag{2}$$

where σ_u^2 is the level 2 variance. In other words, the standard assumptions for random effects state that they are independent and identically distributed (thus homoscedastic) across level 2 units, with a normal distribution.

A further assumption, often not explicitly stated, is the mean independence of the random effects on the covariates (level 2 exogeneity), namely

$$E(u_j \mid \mathbf{x}_{1j}, \mathbf{x}_{2j}, \ldots, \mathbf{x}_{n_j j}, \mathbf{z}_j) = 0 \tag{3}$$

Exogeneity is needed for unbiased estimation (Ebbes et al. 2004, Kim and Frees 2007, Grilli and Rampichini 2011). In this review exogeneity is assumed and not considered anymore.

In most cases of practical interest, the distribution of the random effects is technically identifiable (Alonso et al. 2010), but the data usually carry little information for assessing the appropriateness of the assumed distribution. A gross misspecification of the distribution of the random effects can have severe consequences on the properties of the estimators and it can even obscure some key features of the phenomenon under investigation. In response to those concerns, several papers analyzed the consequences of misspecifying the random effects distribution and proposed alternatives to relax standard assumptions.

The paper is organized as follows. Section 2 summarizes the findings about the consequences of misspecifying the random effects distribution, whereas Section 3 briefly describes the corresponding diagnostic tools. Section 4 reviews the main proposals to relax standard assumptions, focusing on flexible specifications of the random effects distribution. Section 5 is devoted to special issues arising in the Bayesian framework. In Section 6 a case study illustrates the relevance of a suitable specification accounting for heteroscedasticity. Section 7 offers some final remarks.

2 Consequences of misspecification of random effects

In linear models, a wrong specification of the random effects distribution has modest consequences on maximum likelihood estimators: Verbeke and Lesaffre (1997) and Maas and Hox

(2004) showed that the estimators of fixed effects and variance components with normality assumption are consistent and asymptotically normally distributed even if the true random effects do not follow a normal distribution, though their asymptotic covariance matrix is biased (see also Jacqmin-Gadda et al. 2007, McCulloch and Neuhaus 2011a). However, there can be serious consequences on the Empirical Bayes predictions of the random effects (Verbeke and Lesaffre 1996; McCulloch and Neuhaus 2011b).

In some fields, like school effectiveness, empirical Bayes predictions are routinely used to rank the level 2 units. Arpino and Varriale (2010) studied the robustness of the ranking based on empirical Bayes level 2 residuals for a linear random intercept model under several random effects distributions. In a situation with 100 clusters of size 50, with an ICC of 0.2 and normally distributed homoscedastic random effects, they found that the Spearman correlation among true and estimated rankings is 0.945. The correlation is lower in case of misspecification of the random effects distribution, the worst case being that of $\chi^2(1)$ random effects yielding a correlation of 0.817.

In generalized linear models, the consequences of misspecified random effects are more serious and difficult to evaluate. Heagerty and Kurland (2001) performed a simulation study for a random intercept logit model, generating random effects with a Gamma distribution and fitting the model with maximum likelihood assuming normal random effects. This kind of misspecification yields seriously biased regression estimators for level 2 covariates when the true distribution of the random effects is highly skewed and the level 2 standard deviation is high (i.e. $\sigma_u \geq 2$, a value common in longitudinal data, but not realistic in cross-sectional data). Moreover, the simulation study shows that serious biases may occur in case of heteroscedastic random effects. Litière et al. (2007, 2008, 2011) investigated the issue further, also considering type I and type II error rates for the fixed effects, though the conclusions are somewhat controversial.

The consequences of misspecified random effects are investigated also by Agresti et al. (2004) for binary response models and survival models. In addition to the bias of the estimators, they considered the loss of efficiency, which is relevant when the assumed (normal) distribution is substantially different from the true distribution (a two-point mixture with a large variance).

3 Checking the assumptions on the random effects

The usual approach to check the distribution of random effects is based on the analysis of level 2 residuals (Langford and Lewis 1998, Eberly and Thackeray 2005, Snijders and Berkhof 2008, Loy and Hofmann 2013). A widely used diagnostic tool is a normal probability plot of the level 2 standardized Empirical Bayes (EB) predictions. However, such a tool is sensitive to both deviations from normality of the random effects and misspecification of other parts of the model, in particular the fixed effects. In general, it is advisable to carefully check the level 1 model specification before checking the random effects distribution.

McCulloch and Neuhaus (2011a) raise serious doubts on the ability of the normal probability plot of the EB residuals to detect deviations from normality. Indeed, they note that the shape of the distribution of the EB predictions may not reflect the true underlying shape of the distribution of the random effects, but instead the assumed distribution. In other words, EB predictions tend to look normal even when normality is violated.

A general misspecification test for nonlinear mixed effects models has been proposed by Huang (2009, 2011). The test compares ML estimates on original data with ML estimates based on reconstructed coarsened data.

Recently, Verbeke and Molenberghs (2013) proposed the gradient function as an exploratory diagnostic tool to assess misspecification of the random effects distribution. This method is applicable to a wide range of mixed models and it is easy to implement because it only needs ML estimates of the current model and the corresponding marginal likelihood function. The gradient function is plotted alongside with confidence bands, pointing out intervals of values of the random effects for which the distribution is locally misspecified.

As a cautionary note, we remind that all the procedures for checking the random effects distribution suffer from a fundamental limitation: in fact, any discrepancy can be attributed to a misspecification of the random effects only by assuming that the other parts of the model are correctly specified.

4 Specifying the random effects in multilevel models: beyond standard assumptions

In addition to level 2 exogeneity stated in equation (3), the standard assumptions for random effects are: (i) independence across level 2 units; (ii) identical distribution across level 2 units; (iii) normal distribution. We now review several proposals to relax assumptions (i)-(iii).

The assumption that the random effects are independent across level 2 units is questionable when such units are adjacent geographical areas. Indeed, in fields such as disease mapping (Besag et al. 1991) and small area estimation (Rao 2003) the models have spatially correlated random effects. In the multilevel literature, correlated random effects are uncommon. Nonetheless, Browne and Goldstein (2010) considered multilevel models where the higher-level random effects are linked by a suitable correlation structure to be estimated. This is relevant in educational effectiveness, where the performances of nearby schools may be correlated.

The assumption that the random effects have identical distribution across level 2 units implies constant level 2 variance (homoscedasticity), which is too restrictive in some settings. Heteroscedasticity across strata of level 2 units (e.g. private vs public schools) is handled by inserting stratum-specific random effects, as shown in Section 6. Heteroscedasticity depending on continuous covariates is accounted by a parametric regression model for the level 2 variance. This can be accomplished in several ways, for instance by adding random coefficients to the level 2 covariates responsible for the heteroscedasticity (Snijders and Bosker 2012, Sect. 8.2), or by specifying a linear model for the logarithm of the level 2 variance, as in the mixed location scale model of Hedeker et al. (2008, 2012).

The assumption of normal distribution for the random effects can be overcome in several ways ranging from two extremes: (i) a continuous parametric non-normal distribution, and (ii) an arbitrary discrete distribution with locations and masses to be estimated.

Examples of continuous parametric non-normal distributions are the multivariate Student's t of Pinheiro et al. (2001), which yields results robust to outliers, and the skewed parametric family of Liu and Day (2008).

More flexible approaches rely on mixtures, such as finite mixtures of Gaussian distributions (Verbeke and Lesaffre 1996), or penalized Gaussian mixtures (Ghidey et al. 2004, Komárek and Lesaffre 2008). Alternatively, the random effects density can be approximated by a semi-non-parameteric (SNP) representation (Zhang and Davidian 2001, Papageorgiou and Hinde 2012).

The random effects can even be modelled through an arbitrary discrete distribution. For a given number of mass points, the likelihood can be maximized through the EM algorithm (e.g. Aitkin 1999). Nonparametric maximum likelihood (NPML) is obtained by increasing the number of mass points. Two techniques to achieve NPML are the directional derivative (e.g. Rabe-Hesketh et al. 2003) and the direct search method based on the gradient function of Lesperance et al. (2014). NPML estimates can also be obtained by iterating until convergence the smoothing by roughening (SBR) method of Shen and Louis (1999). Recently, Azzimonti et al. (2013) proposed an EM algorithm that includes the selection of the optimal number of mass points.

Nonparametric estimators of the random effects distribution based on Fourier inversion have been proposed by Hall and Yao (2003) and Comte and Samson (2012).

Antic et al. (2009) compared some nonparametric methods in nonlinear mixed effects models, with focus on pharmacokinetics applications. Ghidey et al. (2010) compared four flexible methods for estimating the random effects distribution of a linear mixed model: their simulations indicate that the penalized Gaussian mixtures approach of Ghidey et al. (2004) often has the smallest integrated mean squared error for estimating the random effects distribution. Nonetheless, the finite mixture approach (Verbeke and Lesaffre 1996) and the SNP approach (Zhang and Davidian 2001) seem to perform better when the true distribution of the random effects is a mixture.

From a different perspective, a model with random effects having a discrete distribution can be interpreted as a *latent class* multilevel model (Vermunt 2003), where the level 2 units are assumed to belong to latent classes with common unobserved components. In this framework, the choice of the number of latent classes is a difficult task (Lukociene et al. 2010). A promising procedure to classify level 2 units with data-driven selection of the number of classes is represented by Dirichlet process mixtures. This approach was originally proposed in the Bayesian framework (see Section 5), but it has been recently implemented in a frequentist setting by Heinzl and Tutz (2013) via a penalized EM algorithm.

In the context of repeated measures, discrete random effects or latent classes are the core of Growth Mixture Models (Muthén 2004, Palardy and Vermunt 2010) and Latent Markov Models (Bartolucci et al. 2011).

5 Specification issues in the Bayesian framework

In Bayesian hierarchical modelling (Gelman and Hill 2007) the distribution of the random effects is specified conditionally on the level 2 variance. The implied marginal distribution of the random effects is usually not mentioned, but it should be considered in order to make comparisons with non-Bayesian approaches. The most common specification is $u_j \mid \sigma_u^2 \sim N(0, \sigma_u^2)$ with a Gamma prior distribution for the level 2 precision $(\sigma_u^2)^{-1} \sim \Gamma(a,b)$, where a is the shape parameter and b is the rate parameter (so that the mean is a/b and the variance is a/b^2). It follows (Fong et al. 2010, Lemma 1) that $u_j \sim t_{2a}(0, \sqrt{b/a})$, namely the marginal distribution of the random effects is Student's t with $2a$ degrees of freedom, location 0 and scale $\sqrt{b/a}$. Note that $a = 0.5$ yields one degree of freedom, namely a Cauchy distribution, whereas $a < 0.5$ yields a distribution whose tails are much heavier than even a Cauchy. Prior distributions with $a < 0.5$ are common practice, for example the popular BUGS software (Lunn et al. 2012) has default setting $a = b = 0.001$. Fong et al. (2010) argued that it is difficult to justify for the random effects a marginal distribution such as a t with 0.002 degrees of freedom and propose to choose a prior distribution for $(\sigma_u^2)^{-1}$ that implies a reasonable distribution for u_j. For example, in logistic regression they chose $\Gamma(0.5, 0.0164)$ because it implies a marginal Cauchy distribution such that $e^{u_j} \in [0.1, 10]$ with probability 0.95. For the same model, Grilli et al. (2014) performed a simulation study showing that, in case of few clusters (e.g. 10 clusters), the $\Gamma(0.5, 0.0164)$ prior outperforms the $\Gamma(0.001, 0.001)$ prior for the bias on the level 2 variance and for the coverage of the intervals for the fixed effects.

Gelman (2006) pointed out some drawbacks of the Gamma prior and suggested an alternative prior distribution based on the half-t family.

In order to avoid the specification of the random effects distribution, the Dirichlet process can be exploited to devise a semi-parametric Bayesian approach (Kleinmann and Ibrahim 1998, Ohlssen et al. 2007). This approach is used by White et al. (2012) to classify patients in Parkinson disease, and by Guglielmi et al. (2013) to cluster hospitals on the basis of patients survivals. As noted in Section 4, Heinzl and Tutz (2013) proposed a penalized likelihood approach based on Dirichlet process mixtures.

6 Accounting for heteroscedasticity: a case study on the value added of Italian schools

In educational effectiveness, the homoscedasticity assumption is often inappropriate since the variance between schools is typically quite different across geographical areas or school types (e.g. public vs private). Differential variability is a negative feature in educational systems aiming at guarantee equal opportunities. Indeed, important research questions in educational effectiveness are related to the pattern of variability, thus involving tests of hypothesis on random-part parameters.

Sani and Grilli (2011) analyzed the performance of fifth-grade pupils attending Italian schools using data collected by INVALSI (the Italian national institute for the evaluation of the school system). In particular, they considered the mathematics test administered at the end of the 2008/2009 year, along with a pupil's questionnaire for measuring socio-economic factors. The analysis exploited a random intercept linear model on the Rasch score of the mathematics test, with pupil-level errors depending on gender and school-level errors depending on the geographical area:

$$y_{ij} = \boldsymbol{\beta}'\mathbf{x}_{ij} + \boldsymbol{\gamma}'\mathbf{z}_j + u_{j(k)} + e_{ij(m)} \tag{4}$$

where i indexes pupils (level 1 units) and j indexes schools (level 2 units). Model (4) has the same structure of model (1), except for the errors which are assumed to be normally distributed with zero mean and stratum-specific variances:

$$e_{ij(m)} \overset{iid}{\sim} N(0, \sigma^2_{e(m)}), \quad m = 1, \ldots, M \tag{5}$$

$$u_{j(k)} \overset{iid}{\sim} N(0, \sigma^2_{u(k)}), \quad k = 1, \ldots, K \tag{6}$$

The strata of level 1 units (pupils) are defined by gender (M=2), while the strata of level 2 units (schools) are defined by the geographical area (K=5).

The results show a considerable increase in the residual variance among schools when going from North to South, pointing out a serious issue of fairness in Southern Italy. Specifically, with reference to males, the estimated values of the residual Intraclass Correlation Coefficient (proportion of residual variance due to the schools) are 4.7% in North-West, 6.7% in North-East, 13.5% in Center, 35.1% in South and 23.1% in South-Isles. The high values of the residual ICC in the Southern regions imply that in those regions the pupil achievement is strongly affected by unobserved school-level factors, possibly including social segregation effects not accounted for by socio-economic covariates (Leckie et al. 2012).

7 Final remarks

In multilevel modelling, the specification of the random effects is a key and challenging task. An appropriate specification is crucial for valid predictions of the random effects, which is an important goal in some applications. As for parameter estimation, the consequences of misspecification are usually minor in linear models and potentially serious in nonlinear models. However, the bias on parameter estimates should not be the only concern, since a proper modelling of the variance-covariance structure may be essential for the research aims, as illustrated by the heteroscedastic structure adopted in the case study of Section 6.

In general, the consequences of a wrong specification of the random effects distribution depend on the degree of departure from the true distribution. A dangerous situation is when the assumed distribution has a single mode (as usual), but the true distribution has several sharp modes.

In the last fifteen years, there has been a lot of research on flexible specifications of the random effects distribution. The proposed solutions range from heavy-tailed or skewed parametric families to nonparametric approaches, with several intermediate semi-parametric and mixture-based approaches. In general, when moving from parametric to nonparametric methods one has to face potential losses of efficiency and computational problems. An advantage of flexible methods is their ability to provide relevant information on the nature of heterogeneity of the population of interest. If the analysis aims at classifying level 2 units, the approaches based on the Dirichelet process are especially suitable as they allow a data driven selection of the number of latent classes.

A suitable specification of the random effects distribution is relevant regardless of the inferential approach. However, the standard specification is different in frequentist and Bayesian approaches. In fact, in Bayesian modelling the normality assumption is usually placed on the distribution of the random effects conditional on the level 2 variance: therefore, with the usual Gamma prior, the marginal distribution of the random effects is Student's t, which is more robust to outliers. Nevertheless, the specification of the prior for the level 2 variance entails several open issues as discussed in Section 5.

Recently proposed diagnostic tools are promising. Yet, their usefulness is limited since any discrepancy can be attributed to a misspecification of the random effects only by assuming that the other parts of the model are correctly specified. Therefore, after a decade we still subscribe the Agresti et al. (2004) suggestion to rely on sensitivity analysis by fitting the model using both a parametric specification and a flexible approach.

References

[1] 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)

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

[3] Alonso, A., Litire, S., Laenen, A.: A Note on the Indeterminacy of the Random-Effects Distribution in Hierarchical Models. The American Statistician **64**, 318–324 (2010)

[4] Antic, J., Laffont, C.M., Chafaï, D., Concordet, D.: Comparison of nonparametric methods in nonlinear mixed effects models. Computational Statistics and Data Analysis **53**, 642–656 (2009)

[5] Arpino, B., Varriale, R.: Assessing the quality of institutions' rankings obtained through multilevel linear regression models. Journal of Applied Economic Sciences **5**, 7–22 (2010)

[6] Azzimonti, L., Ieva, F., Paganoni, A.M.: Nonlinear nonparametric mixed-effects models for unsupervised classification. Computational Statistics **28**, 1549–1570 (2013)

[7] Bartolucci, F., Pennoni, F., Vittadini, G.: Assessment of School Performance Through a Multilevel Latent Markov Rasch Model. Journal of Educational and Behavioral Statistics **36**, 491–522 (2011)

[8] Besag, J., York, J., Mollié, A.: Bayesian image restoration, with two applications in spatial statistics. Ann. Inst. Statist. Math. **43**, 1–59 (1991)

[9] Browne, W., Goldstein, H.: MCMC Sampling for a Multilevel Model With Nonindependent Residuals Within and Between Cluster Units. Journal of Educational and Behavioral Statistics **35**, 453–473 (2010)

[10] Comte, F., Samson, A.: Nonparametric estimation of random-effects densities in linear mixed-effects model. Journal of Nonparametric Statistics **24**, 951–975 (2012)

[11] Demidenko, E.: Mixed Models: Theory and Applications with R, 2nd Edition. New York: Wiley (2013)

[12] Ebbes, P., Bockenholt, U., Wedel, M.: Regressor and random-effects dependencies in multilevel models. Statistica Neerlandica **58**, 161–178 (2004)

[13] Eberly, L.E., Thackeray, L.M.: On Lange and Ryan's plotting technique for diagnosing non-normality of random effects. Statistics & Probability Letters **75**, 77–85 (2005)

[14] Fong, Y., Rue, H., Wakefield, J. Bayesian inference for Generalized Linear Mixed Models. Biostatistics **11**, 397–412 (2010)

[15] Ghidey, W., Lesaffre, E., Eilers, P.: Smooth Random Effects Distribution in a Linear Mixed Model. Biometrics **60**, 945–953 (2004)

[16] Ghidey, W., Lesaffre, E., Verbeke, G.: A comparison of methods for estimating the random effects distribution of a linear mixed model. Statistical Methods in Medical Research **19**, 575–600 (2010)

[17] Gelman, A.: Prior distributions for variance parameters in hierarchical models. Bayesian Analysis **1**, 515–533 (2006)

[18] Gelman, A., Hill, J.: Data analysis using regression and multilevel/hierarchical models. Cambridge: Cambridge University press (2007)

[19] Goldstein, H.: Multilevel statistical models. John Wiley & Sons (2011)

[20] Grilli, L., Rampichini, C.: Multilevel models for the evaluation of educational institutions: a review. In: Monari P., Bini M., Piccolo D., Salmaso L. (eds.) Statistical Methods for the Evaluation of Educational Services and Quality of Products, pp. 61–80. Physica-Verlag (2009)

[21] Grilli, L., Rampichini, C.: The role of sample cluster means in multilevel models: a view on endogeneity and measurement error issues. Methodology **7**, 121–133 (2011)

[22] Grilli, L., Metelli, S., Rampichini, C.: Bayesian estimation with INLA for logistic multilevel models. Journal of Statistical Computation and Simulation, to appear (2014)

[23] Guglielmi, A., Ieva, F., Paganoni, A.M., Ruggeri, F., Soriano, J.: Semiparametric Bayesian models for clustering and classification in presence of unbalanced in-hospital survival. To appear in Journal of the Royal Statistical Society C (2013)

[24] Hall, P., Yao, Q.: Inference in Components of Variance Models with Low Replication. The Annals of Statistics **31**, 414–441 (2003)

[25] Heagerty, P.J., Kurland, B.F.: Misspecified Maximum Likelihood Estimates and Generalised Linear Mixed Models. Biometrika **88**, 973–985 (2001)

[26] Hedeker, D., Mermelstein, R.J., Demirtas, H.: An Application of a Mixed-Effects Location Scale Model for Analysis of Ecological Momentary Assessment (EMA) Data. Biometrics **64**, 627–634 (2008)

[27] Hedeker, D., Mermelstein, R.J., Demirtas, H.: Modeling between-subject and within-subject variances in ecological momentary assessment data using mixed-effects location scale models. Statistics in Medicine **31**, 3328–3336 (2012)

[28] Heinzl, F., Tutz, G.: Clustering in linear mixed models with approximate Dirichlet process mixtures using EM algorithm. Statistical Modelling **13**, 41–67 (2013)

[29] Huang, X.: Diagnosis of random-effect model misspecification in generalized linear mixed models for binary response. Biometrics **65**, 361–368 (2009)

[30] Huang, X.: Detecting random-effects model misspecification via coarsened data. Computational Statistics and Data Analysis **55**, 703–714 (2011)

[31] Jacqmin-Gadda, H., Sibillot, S., Proust, C., Molina, J.-M., Thiébaut, R.: Robustness of the linear mixed model to misspecified error distribution. Computational Statistics and Data Analysis **51**, 5142–5154 (2007)

[32] Kim, J.S., Frees, E.W.: Multilevel Modeling with Correlated Effects. Psychometrika **72**, 505–533 (2007)

[33] Kleinman, K., Ibrahim, J.: A semi-parametric Bayesian approach to generalized linear mixed models. Statistics in Medicine **17**, 2579–96 (1998)

[34] Komárek, A., Lesaffre, E.: Generalized linear mixed model with a penalized Gaussian mixture as a random effects distribution. Computational Statistics and Data Analysis **52**, 3441–3458 (2008)

[35] Langford, I. H., Lewis, T.: Outliers in multilevel data. Journal of the Royal Statistical Society A **161**, 121–160 (1998)

[36] Leckie, G., Goldstein, H.: The limitations of using school league tables to inform school choice. Journal of the Royal Statistical Society A **172**, 835–851 (2009)

[37] Leckie, G.B., Pillinger, R.J., Jones, K., Goldstein, H.: Multilevel modelling of social segregation. Journal of Educational and Behavioral Statistics **37**, 3–30 (2012)

[38] Lesperance, M., Saab, R., Neuhaus, J.: Nonparametric estimation of the mixing distribution in logistic regression mixed models with random intercepts and slopes. Computational Statistics and Data Analysis **71**, 211–219 (2014)

[39] Litière, S., Alonso, A., Molenberghs, G.: Type I and Type II Error Under Random-Effects Misspecification in Generalized Linear Mixed Models. Biometrics **63**, 1038–1044 (2007)

[40] Litière, S., Alonso, A., Molenberghs, G.: The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. Statistics in medicine **27**, 3125–3144 (2008)

[41] Litière, S., Alonso, A., Molenberghs, G.: Rejoinder to “A note on Type I and Type II Error Under Random-Effects Misspecification in Generalized Linear Mixed Models”. Biometrics **67**, 656–660 (2011)

[42] Liu, J., Dey, D.K.: Skew random effects in multilevel binomial models: an alternative to nonparametric approach. Statistical Modelling **8**, 221–241 (2008)

[43] Loy A., Hofmann H.: Diagnostic tools for hierarchical linear models. WIREs Comput Stat **5**, 48–61 (2013)

[44] Lukociene, O., Varriale, R., Vermunt, J.K.: The simultaneous decision(s) about the number of lower- and higher-level classes in multilevel latent class analysis. Sociological Methodology **40**, 247–283 (2010)

[45] Lunn, D., Jackson, C., Best, N., Thomas, A., Spiegelhalter, D.: The BUGS Book - A Practical Introduction to Bayesian Analysis published. Boca Raton, FL: CRC Press / Chapman and Hall (2012).

[46] Maas, C.J.M., Hox, J.J.: Robustness issues in multilevel regression analysis. Statistica Neerlandica **58**, 127–137 (2004)

[47] McCulloch, C.E., Neuhaus, J.M.: Misspecifying the Shape of a Random Effects Distribution: Why Getting It Wrong May Not Matter. Statistical Science **26**, 388–402 (2011a)

[48] McCulloch, C.E., Neuhaus, J.M.: Prediction of Random Effects in Linear and Generalized Linear Models under Model Misspecification. Biometrics **67**, 270–279 (2011b)

[49] Muthén, B.: Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In: Kaplan D. (ed.) Handbook of quantitative methodology for the social sciences, pp. 345–368. Sage (2004)

[50] Ohlssen, D.I., Sharples, L.D., Spiegelhalter, D.J.: Flexible random-effects models using Bayesian semi-parametric models: Applications to institutional comparisons. Statistics in Medicine **26**, 2088–2112 (2007)

[51] Palardy, G., Vermunt, J.K.: Multilevel growth mixture models for classifying groups. Journal of Educational and Behavioral Statistics **35**, 532–565 (2010)

[52] Papageorgiou, G., Hinde, J.: Multivariate generalized linear mixed models with semi-nonparametric and smooth nonparametric random effects densities. Statistics and Computing **22**, 79–92 (2012)

[53] Pinheiro, J.C., Liu, C., Wu, Y.N.: Efficient Algorithms for Robust Estimation in Linear Mixed-Effects Models Using the Multivariate t Distribution. Journal of Computational and Graphical Statistics, **10**, 249–276 (2001)

[54] 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)

[55] Rao, J.N.K.: Small area estimation. LinkHoboken (N.J.): John Wiley & sons (2003)

[56] Raudenbush, S.W., Bryk, A.S.: Hierarchical linear models: Applications and data analysis methods. Sage (2002)

[57] Sani, C., Grilli, L.: Differential variability of test scores among schools: a multilevel analysis of the fifth-grade Invalsi test using heteroscedastic random effects. Journal of Applied Quantitative Methods **6**, 88–99 (2011)

[58] Scott, M.A., Simonoff, J.S., Marx, B.D. (eds): The Sage Handbook of Multilevel Modeling. Sage (2013)

[59] Shen, W., Louis, T,A.: Empirical Bayes estimation via the smoothing by roughening approach. Journal of Computational and Graphical Statistics **8**, 800–823 (1999)

[60] Snijders, T.A.B., Berkhof, J.: Diagnostic checks for multilevel models. In: de Leeuw, J., Meijer E. (eds.) Handbook of Multilevel Analysis, pp. 141–175. Springer (2008)

[61] Snijders, T.A.B., Bosker, R.J.: Multilevel analysis: an introduction to basic and advanced multilevel modelling. 2nd ed., Sage (2012)

[62] Verbeke, G., Lesaffre, E.: A linear mixed-effects model with heterogeneity in the random-effects population. Journal of the American Statistical Association **91**, 217–221 (1996)

[63] Verbeke, G., Lesaffre, E.: The effects of misspecifying the random-effects distribution in linear mixed models for longitudinal data. Computational Statistics and Data Analysis **23**, 541–56 (1997)

[64] Verbeke, G., Molenberghs, G.: The gradient function as an exploratory goodness-of-fit assessment of the random-effects distribution in mixed models. Biostatistics **14**, 477–490 (2013)

[65] Vermunt, J.K.: Multilevel latent class models. Sociological methodology **33**, 213–239 (2003)

[66] White, N., Johnson, H., Silburn, P.A.: Dirichlet process mixture models for unsupervised clustering of symptoms in Parkinson's disease. Journal of Applied Statistics **39**, 2363–2377 (2012)

[67] Zhang, D., Davidian, M.: Linear Mixed Models with Flexible Distributions of Random Effects for Longitudinal Data. Biometrics **57**, 795–802 (2001)