Modeling the residual error variance in Two-Level Random-Coefficient Multilevel Models

George Leckie
Centre for Multilevel Modelling, University of Bristol, UK
E-mail: g.leckie@bristol.ac.uk

Abstract

Multilevel models are a popular method of clustered and longitudinal data analysis in the social, behavioral and medical sciences. The standard two-level random-coefficient model for continuous responses nearly always assumes a constant residual error variance at level-1. However, there is no reason why this homogeneity assumption should hold in practice and in many studies it will be intrinsically interesting to relax it. In this paper, we model the level-1 residual error variance as a function of predictors and we allow a random-intercept and random-coefficients to be included in this function. We illustrate our approach through a real data application to modeling school effects on student achievement.

Key Words: heterogeneous within-group variances; heteroskedasticity; hierarchical linear models; log-linear variance models; mixed-effects models; variance functions

1. Introduction

Multilevel models (Goldstein, 2011; Snijders and Bosker, 2012) – also known as hierarchical linear models (Raudenbush and Bryk, 2002), mixed-effects models or random effects models – are widely applied in the social, behavioral and medical sciences. The simplest and most commonly fitted model is the two-level random-coefficient model for continuous responses. A standard assumption is that the level-1 residual error variance is constant across the units in the data. However, there is no reason why this should be true and in many studies it will be intrinsically interesting to model this parameter as a function of the predictors and level-2 random effects. Such studies include those concerned with notions of inequality, segregation, consistency, similarity, volatility, erraticness, and the predictability of individuals within groups.

In this paper we extend the two-level random-coefficient multilevel model to allow for heterogeneous level-1 residual error variances. Similar extensions have been considered by Hedeker et al. (2008) and Lee and Nelder (2006), among others. Specifically, we model the level-1 variances as a function of the predictors and we allow both a random-intercept and random-coefficients to be included in this variance function. We present these methodological developments in Section 2. Section 3 illustrates our approach through a real data application to modeling school effects on student achievement. We conclude with a discussion in Section 4.

2. Methods

Consider the two-level random-coefficient multilevel model for continuous response y_{ij} on level-1 unit i ($i = 1,2, \dots, n_j$) in level-2 unit j ($j = 1,2, \dots, J$), modeled, for simplicity, in terms of a single level-1 predictor, x_{ij}. This model is written as

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_{0j} + u_{1j} x_{ij} + e_{ij}$$

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim \mathrm{N}\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u0}^2 & \\ \sigma_{u01} & \sigma_{u1}^2 \end{pmatrix} \right\}$$

$$e_{ij} \sim \mathrm{N}(0, \sigma_e^2)$$

where u_{0j} and u_{1j} are the level-2 random-intercept and random-slope effects and e_{ij} is the level-1 random effect or residual error term. The level-2 and level-1 random effects are assumed normally distributed and independent across levels. The level-1 residual error variance is assumed homogenous across of level-1 units.

We can relax the homogenous residual error variance assumption by modeling it as a log-linear function of the predictor

$$\log\left(\sigma_{e_{ij}}^2\right) = \alpha_0 + \alpha_1 x_{ij} + v_{0j} + v_{1j} x_{ij}$$

$$\begin{pmatrix} v_{0j} \\ v_{1j} \end{pmatrix} \sim \mathrm{N}\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{v0}^2 & \\ \sigma_{v01} & \sigma_{v1}^2 \end{pmatrix} \right\}$$

where the v_{0j} and v_{1j} are the level-2 random-intercept and random-slope effects. Note that the residual error variance is the expected variance, not the observed variance, and so there is no residual error in this equation. These variance function random effects are assumed bivariate normally distributed and independent of the mean function random effects. However, we can equally allow the two sets of random effects to be correlated.

We fit the model described above using Markov chain Monte Carlo (MCMC) methods using the eStat estimation engine within the Stat-JR (pronounced "stature") statistics package (Charlton et al. 2012) being developed at the Centre for Multilevel Modelling (CMM). We call Stat-JR from within the Stata general-purpose statistics package using the user-written runstatjr Stata command (Leckie and Charlton, 2013) to facilitate pre- and post-estimation data manipulation and graphics, although this is not a requirement of using Stat-JR. We specify diffuse (vague, flat or minimally informative) prior distributions for all parameters. We fit all models with a burn-in period of 10000 iterations and a monitoring period of 10000 iterations.

3. Application

For our real-data application, we reanalyze the "tutorial" dataset provided with the MLwiN User Manual (Rasbash et al., 2012) and first analyzed by Goldstein et al. (1993). The data are two-level with 4,059 students (level-1 units) nested within 65 schools (level-

2 units). The response is a standardized age 16 end of school examination score. The main predictor variables are a standardized age 11 reading test score and student gender.

Model 1 is a standard two-level random-slope model for students' age 16 scores, but where we allow every school to have their own residual error variance. The model is written as

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + u_{0j} + u_{1j} x_{1ij} + e_{ij}$$

$$\log\left(\sigma^2_{e_{ij}}\right) = \alpha_0 + v_{0j}$$

$$\begin{pmatrix} u_{0j} \\ u_{1j} \\ v_{0j} \end{pmatrix} \sim \mathrm{N}\left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{u0} & & \\ \sigma_{u01} & \sigma^2_{u1} & \\ \sigma_{u0v0} & \sigma_{u1v0} & \sigma^2_{v0} \end{pmatrix}\right\}$$

$$e_{ij} \sim \mathrm{N}\left(0, \sigma^2_{e_{ij}}\right)$$

where y_{ij} is the age 16 exam score for student i ($i = 1, \ldots, n_j$) in school j ($j = 1, \ldots, 65$), x_{1ij} is their corresponding age 11 score when they entered their school, and x_{2ij} is a binary indicator for whether the student is a girl.

Table 1 presents the results for this model. Turning our attention to the mean function, we see that, in the average school, a one standard deviation increase in age 11 scores is associated with a 0.55 standard deviation increase in age 16 scores. However, the strength of this relationship varies between schools as illustrated by the plotted mean functions in Figure 1 (left panel). Having adjusted for age 11 scores, we see that girls score 0.17 of a standard deviation higher than boys at age 16. Now turning our attention to the variance function, we see that the intercept variance appears significant and so students in some schools appear to score more variably at age 16 that students in other schools, even after adjusting for their age 11 scores and gender. This suggests that schools influence the variability in student responses as well as their mean response.

Model 2 adds the age 11 scores and student gender predictor variables to the variance function. The model is written as

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + u_{0j} + u_{1j} x_{1ij} + e_{ij}$$

$$\log\left(\sigma^2_{e_{ij}}\right) = \alpha_0 + \alpha_1 x_{1ij} + \alpha_2 x_{2ij} + v_{0j}$$

$$\begin{pmatrix} u_{0j} \\ u_{1j} \\ v_{0j} \end{pmatrix} \sim \mathrm{N}\left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{u0} & & \\ \sigma_{u01} & \sigma^2_{u1} & \\ \sigma_{u0v0} & \sigma_{u1v0} & \sigma^2_{v0} \end{pmatrix}\right\}$$

$$e_{ij} \sim \mathrm{N}\left(0, \sigma^2_{e_{ij}}\right)$$

The results for the mean function are broadly similar to Model 1. Turning our attention to the variance function, we see that age 11 scores have a significant negative effect on the level-1 residual error variances. Thus, the variability in students' age 16 scores around their school lines decreases with increasing age 11 scores. Put differently, students who were scoring low at intake have more dispersed age 16 responses than students who were scoring high at intake; they are less predictable. Gender also has a significant negative coefficient and so boys are also found to score more erratically than girls.

Table 1 *Results for Models 4, 5 and 6 fitted to the student achievement data*

Parameter		Model 1		Model 2		Model 3	
		Mean	SD	Mean	SD	Mean	SD
Mean function							
β_0	Intercept	-0.11	0.04	-0.11	0.05	-0.11	0.04
β_1	Age 11 scores	0.55	0.02	0.55	0.02	0.55	0.02
β_2	Girl	0.17	0.03	0.17	0.03	0.17	0.03
σ^2_{u0}	Intercept variance	0.10	0.02	0.11	0.02	0.11	0.02
σ^2_{u1}	Slope variance	0.02	0.01	0.02	0.01	0.02	0.01
ρ_{u01}	Intercept-Slope correlation	0.51	0.15	0.53	0.15	0.52	0.14
Variance function							
α_0	Intercept	-0.63	0.04	-0.57	0.05	-0.57	0.05
α_1	Age 11 scores	–	–	-0.07	0.02	-0.07	0.03
α_2	Girl	–	–	-0.10	0.05	-0.10	0.06
σ^2_{v0}	Intercept variance	0.06	0.02	0.06	0.02	0.06	0.02
σ^2_{v1}	Slope variance	–	–	–	–	0.03	0.01
ρ_{v01}	Intercept-Slope correlation	–	–	–	–	-0.15	0.29
Cross-function correlations							
ρ_{u0v0}	Intercept-Intercept correlation	0.40	0.17	0.45	0.15	0.46	0.16
ρ_{u1v0}	Slope-Intercept correlation	0.76	0.14	0.77	0.13	0.77	0.14
ρ_{u0v1}	Intercept-Slope correlation	–	–	–	–	0.03	0.24
ρ_{u1v1}	Slope-Slope correlation	–	–	–	–	-0.04	0.27

Note: The reported means and standard deviations are the posterior means and posterior standard deviations of the corresponding parameter chains.

Model 3 allows the effect of age 11 scores in the variance function to vary randomly across schools. The model is written as

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + u_{0j} + u_{1j} x_{1ij} + e_{ij}$$

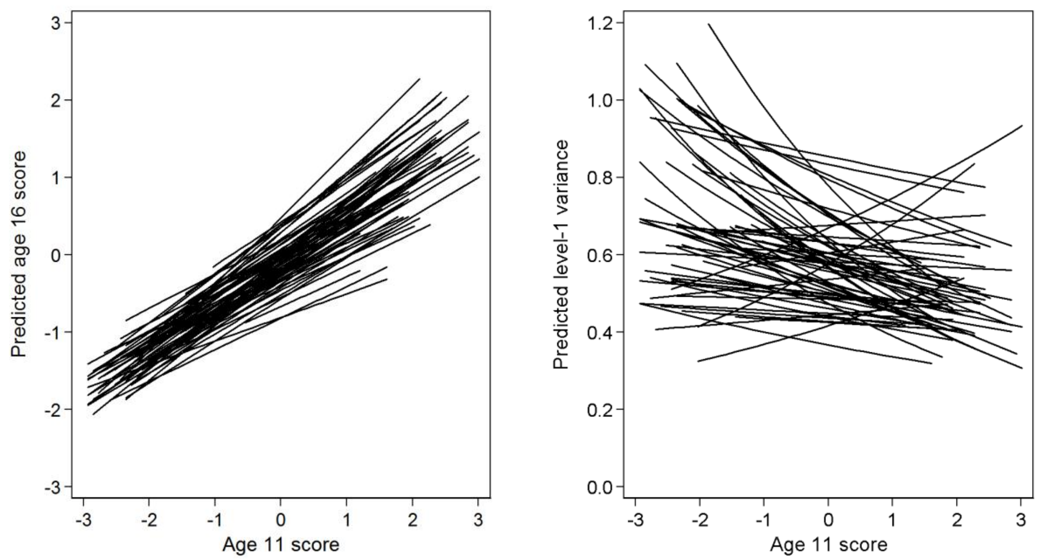
$$\log\left(\sigma^2_{e_{ij}}\right) = \alpha_0 + \alpha_1 x_{1ij} + \alpha_2 x_{2ij} + v_{0j} + v_{1j} x_{1ij}$$

$$\begin{pmatrix} u_{0j} \\ u_{1j} \\ v_{0j} \\ v_{1j} \end{pmatrix} \sim \mathrm{N}\left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{u0} & & & \\ \sigma_{u01} & \sigma^2_{u1} & & \\ \sigma_{u0v0} & \sigma_{u1v0} & \sigma^2_{v0} & \\ \sigma_{u0v1} & \sigma_{u1v1} & \sigma_{v01} & \sigma^2_{v1} \end{pmatrix} \right\}$$

$$e_{ij} \sim \mathrm{N}\left(0, \sigma^2_{e_{ij}}\right)$$

The results for the mean function are again broadly similar to Model 1. Turning our attention to the variance function, we see that the slope variance on age 11 scores appears significant and so there is some suggestion that the relationship between age 16 achievement dispersion and age 11 scores does vary from school to school. Figure 1 (right panel) gives a sense of this variability by plotting the variance function for each school.

Figure 1. Model 1 predicted mean functions (left panel) and Model 3 predicted variance functions (right panel)



4. Conclusion

We have extended the two-level random-coefficient model for continuous responses by modeling the level-1 residual error variance as a log-linear function of predictors and these predictors are allowed to enter into this variance function with random coefficients. Thus, both the response and the level-1 residual error variance are jointly modeled as random-coefficient models where the random effects in each model are allowed to covary. This extension allows for a substantively richer and more realistically complex class of two-level multilevel models. The principle of modeling the residual error variance in two-level models as a function of the predictors and further random effects generalizes to more complex multilevel models, including those with three- and higher-levels, crossed-random effects, and discrete responses.

A full length and extended version of this paper is currently in preparation (Leckie et al., 2013).

References

Browne, W. J., Draper, D., Goldstein, H., & Rasbash, J. (2002). Bayesian and likelihood methods for fitting multilevel models with complex level-1 variation. *Computational statistics & data analysis, 39*, 203-225.

Charlton, C. M. J., Michaelides, D. T., Cameron, B., Szmaragd, C., Parker, R. M. A., Yang, H., Zhang, Z., & Browne, W. J. (2012). Stat-JR software. Center for Multilevel Modelling, University of Bristol and Electronics and Computer Science, University of Southampton.

Goldstein, H. (2011). Multilevel Statistical Models, 4th edition. Chichester, UK: Wiley

Goldstein, H., Rasbash, J., Yang, M., Woodhouse, G., Pan, H., Nuttall, D., & Thomas, S. (1993). A Multilevel Analysis of School Examination Results. *Oxford Review of Education*, *19*, 425-433.

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

Leckie, G., & Charlton, C. (2013).rRunstatjr - A Program to Run the Stat-JR Statistical Modeling Software from within Stata. University of Bristol.

Leckie, G., French, R., Charlton, C. and Browne, W. (2013). Modeling Heterogeneous Variance-Covariance Components in Two-Level Multilevel Models. University of Bristol.

Lee, Y., & Nelder, J. A. (2006). Double hierarchical generalized linear models (with discussion). *Applied Statistics*, *55*, 139–185.

Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical Linear Models: Applications and Data Analysis methods. 2nd edition. Thousand Oaks, CA: Sage.

Snijders, T. A. B., & Bosker, R. J. (2012). Multilevel Analysis: an Introduction to Basic and Advanced Multilevel Modeling. 2nd edition. Sage.