MULTILEVEL MODELLING NEWSLETTER

Centre for Multilevel Modelling

Bedford Group for Lifecourse and Statistical Studies
Institute of Education, University of London
20 Bedford Way, London WC1H 0AL, ENGLAND
Web site: http://multilevel.ioe.ac.uk/
Enquiries about newsletter to Ian Plewis
E-mail: i.plewis@ioe.ac.uk
Tel: +44 (0) 20 7612 6688
Fax: +44 (0) 20 7612 6572

Vol. 14 No. 2 **December, 2002**

Forthcoming Workshops

7-9 April 2003, a three-day introductory workshop to multilevel modelling using *MLwiN* will take place at the University of Bristol.

Enquiries to Jean Flowers at Graduate School of Education, 35 Berkeley Square, Bristol BS8 1HJ, United Kingdom. Tel: +44 (0) 117 928 7059; Fax: +44 (0) 117 925 5412. Email: jean.flowers@bristol.ac.uk

If you plan to run any workshops using *MLwiN*, please notify Amy Burch and she will advertise these workshops on the multilevel web site.

Multilevel Multiprocess Models for Partnership and Childbearing Event Histories project

Fiona Steele, Harvey Goldstein and Heather Joshi, within the Bedford Group at the Institute of Education, have been awarded a grant for a two-year research project under the ESRC Research Methods programme. The project is due to start in early 2003 and will involve the development of multilevel simultaneous equations models for correlated event histories. The research is motivated by a study of the interrelationships between partnership (marriage or cohabitation) durations and decisions about childbearing, using event history data from the 1958 and 1970 British birth cohort studies. Methodology developed under the project will be implemented in *MLwiN* and aML.

Details of the project, including work in progress and training materials, will be made available on the Centre for Multilevel Modelling website at: http://multilevel.ioe.ac.uk/team/mmmpceh.html

E·S·R·C
ECONOMIC & SOCIAL RESEARCH COUNCIL

Fourth International Amsterdam Conference On Multilevel Analysis

The Fourth International Amsterdam Conference on Multilevel Analysis will be held in Amsterdam on 28-29 April 2003, followed on 30 April 2003 by a half-day course taught by Tom Snijders (ICS, University of Groningen) treating Sample Size Determination in Multilevel Analysis.

The conference will be about all aspects of statistical multilevel analysis: theory, software, methodology, and innovative applications. The conference and course will be in an informal style, with much room for discussion.

Further information can be obtained from t.a.b.snijders@ppsw.rug.nl or at http://www.siswo.uva.nl/congressen/congressen/folders/multilevel%20analyse%202003.htm

Multilevel Modelling Portal from UCLA

The UCLA ATS Statistical Computing Group invites researchers to visit the UCLA Statistical Computing Portal at: http://statcomp.ats.ucla.edu/
The portal links to sites with information about commonly used statistical packages such as SAS, Stata, and SPSS and it has a search engine that searches across these sites, saving you the effort of visiting sites individually for the information you seek. The portal has a section specialising in multilevel modelling at:
http://statcomp.ats.ucla.edu/mlm/
with links on multilevel modelling, and multilevel modelling packages such as *MLwiN* and HLM. The multilevel modelling portal has its own search engine that allows you to search across hundreds of pages from around the world on multilevel modelling. Included in this search is the UCLA multilevel modelling resources page at: http://www.ats.ucla.edu/stat/mlm/
This is a modest but growing site featuring links to downloadable papers on multilevel modelling, frequently asked questions on *MLwiN* and HLM, and textbook examples (web pages that show how to solve examples from chapters of books using packages like *MLwiN*, HLM, SAS Proc Mixed, SPSS MIXED, etc.).

But the UCLA Stat Computing portal is more than a set of links about statistical computing or a search engine that allows you to search across these pages for information about statistical computing. This is a site supporting collaborations with other consulting centres and researchers involved in statistical computing and multilevel modelling, see
http://statcomp.ats.ucla.edu/propcollaboration.htm
for more information.

We invite researchers to visit these pages and hope that researchers will find them useful for searching among the most informative statistical computing pages/sites, drawing upon the collective efforts of people from all over the world.

Michael Mitchell
Statistical Consulting Group
UCLA Academic Technology Services
http://www.ats.ucla.edu/stat/

An Expectation-Maximization Algorithm for Generalised Linear Three-Level Models

Jeroen K. Vermunt
Department of Methodology and Statistics, Tilburg University
j.k.vermunt@uvt.nl

Introduction

A popular estimation method for generalised linear mixed models is maximum likelihood (ML). With nonnormal dependent variables the likelihood function is approximated by means of Gauss-Hermite quadrature. Software packages implementing this method include the MIXOR family of programs (Hedeker and Gibbons, 1996), the SAS NLMIXED procedure, and the STATA GLLAMM routine (Rabe-Hesketh et al., 2001). MIXOR is a two level program; the other two programs can also handle other types of mixed models.

If the mixed model of interest is a two level model, ML estimation can be performed by means of the EM algorithm, which is a natural approach to estimation problems with missing data (Agresti et al., 2000). The standard EM algorithm cannot, however, be used for other types of mixed models because the number of entries in the relevant posterior distribution is huge, making the method impractical. This is a pity because EM is a very stable and quite a fast algorithm, especially if one realises that NLMIXED and GLLAMM maximise the log-likelihood using Newton-type algorithms with numerical derivatives, which may make these procedures somewhat unstable and slow. Lesaffre and Spiessens (2001) reported difficulties with Newton-type algorithms in finding the global ML solution in non-linear mixed models: different routines may give different solutions given a certain number of quadrature points. Computation of numerical first and second derivates is computationally intensive if a model contains more than a few parameters.

In this paper it is shown that with nested random effects as in multilevel models, implementation of the EM algorithm is possible by making use of the conditional independence assumptions implied by a multilevel model. Although for simplicity of exposition we only deal with the three level case, the proposed method can easily be generalised to any number of levels.

The next section describes the three level model of interest. Subsequently, attention is paid to parameter estimation by maximum likelihood (ML) and an application using an empirical data set is presented. We end with a short discussion.

The generalised linear three level model

Let i denote a level-1 unit, j a level-2 unit, and k a level-3 unit. The total number of level-3 units is denoted by K, the number of level-2 units within level-3 unit k by n_k, and the number of level-1 units within level-2 unit jk by n_{jk}. Let y_{ijk} be the response of level-1 unit ijk on the outcome variable of interest, and let $\mathbf{x}_{ijk}$, $\mathbf{z}_{ijk(2)}$, and $\mathbf{z}_{ijk(3)}$ be the design vectors associated with S fixed effects, $R_{(2)}$ level-2 random effects, and $R_{(3)}$ level-3 random effects, respectively. It is assumed that the conditional densities of the responses given covariates and random effects are from the exponential family. Denoting the link function by g[..], the generalised linear three level model (GLTM) can be defined as

$$g[E(y_{ijk} \mid \mathbf{x}_{ijk}, \mathbf{z}_{ijk(2)}, \mathbf{z}_{ijk(3)}, \boldsymbol{\beta}_{jk(2)}, \boldsymbol{\beta}_{k(3)})] =$$
$$\eta_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\alpha} + \mathbf{z}_{ijk(2)}^T \boldsymbol{\beta}_{jk(2)} + \mathbf{z}_{ijk(3)}^T \boldsymbol{\beta}_{k(3)}.$$

Here, $\boldsymbol{\alpha}$ is the vector of unknown fixed effects, $\boldsymbol{\beta}_{jk(2)}$ is the vector of unknown random effects for level-2 unit jk and $\boldsymbol{\beta}_{k(3)}$ is the vector of unknown random effects for level-3 unit k.

As usual, we assume the distribution of the random effects $\boldsymbol{\beta}_{jk(2)}$ and $\boldsymbol{\beta}_{k(3)}$ to be multivariate normal with zero mean vector and covariance matrices $\boldsymbol{\Sigma}_{(2)}$ and $\boldsymbol{\Sigma}_{(3)}$. For parameter estimation, it is convenient to standardise and orthogonalise the random effects. For this, let $\boldsymbol{\beta}_{jk(2)} = \mathbf{C}_{(2)}\ \boldsymbol{\theta}_{jk(2)}$, where $\mathbf{C}_{(2)}$ is the Cholesky decomposition of $\boldsymbol{\Sigma}_{(2)}$. Similarly, we define $\boldsymbol{\beta}_{k(3)} = \mathbf{C}_{(3)}\ \boldsymbol{\theta}_{k(3)}$. The reparameterised GLTM is then

$$\eta_{ijk} = \mathbf{x}_{ijk}^T \boldsymbol{\alpha} + \mathbf{z}_{ijk(2)}^T \mathbf{C}_{(2)} \boldsymbol{\theta}_{jk(2)} + \mathbf{z}_{ijk(3)}^T \mathbf{C}_{(3)} \boldsymbol{\theta}_{k(3)}.$$

Maximum likelihood estimation
Log-likelihood function

The parameters of the GLTM described in the previous section can be estimated by maximum likelihood (ML). The likelihood function is based on the probability densities of the level-3 observations, denoted by $P(\mathbf{y}_k|\mathbf{x}_k,\mathbf{z}_{k(2)},\mathbf{z}_{k(3)})$. To simplify notation, the conditioning on the design vectors is replaced by an index corresponding to the unit concerned, yielding the short-hand notation $P(\mathbf{y}_k)$ for the probability density of unit k. The log-likelihood to be maximised equals

$$\log L = \sum_{k=1}^{K} \log P_k(\mathbf{y}_k),$$

where

$$P_k(\mathbf{y}_k) = \int_{\Theta_{(3)}} P_k(\mathbf{y}_k \mid \boldsymbol{\theta}_{(3)}) f(\boldsymbol{\theta}_{(3)}) d\boldsymbol{\theta}_{(3)} = \int_{\Theta_{(3)}} \left\{ \prod_{j=1}^{n_k} P_{jk}(\mathbf{y}_{jk} \mid \boldsymbol{\theta}_{(3)}) \right\} f(\boldsymbol{\theta}_{(3)}) d\boldsymbol{\theta}_{(3)}, \qquad (1)$$

and

$$P_{jk}(\mathbf{y}_{jk} \mid \boldsymbol{\theta}_{(3)}) = \int_{\Theta_{(2)}} P_{jk}(\mathbf{y}_{jk} \mid \boldsymbol{\theta}_{(2)}, \boldsymbol{\theta}_{(3)}) f(\boldsymbol{\theta}_{(2)}) d\boldsymbol{\theta}_{(2)} = \int_{\Theta_{(2)}} \left\{ \prod_{i=1}^{n_{jk}} P_{ijk}(y_{ijk} \mid \boldsymbol{\theta}_{(2)}, \boldsymbol{\theta}_{(3)}) \right\} f(\boldsymbol{\theta}_{(2)}) d\boldsymbol{\theta}_{(2)}, \qquad (2)$$

As can be seen, the responses of the n_k level-2 units within level-3 unit k are assumed to be independent of one another given the random effects $\boldsymbol{\theta}_{(3)}$, and the responses of the n_{jk} level-1 units within level-2 unit jk are assumed to be independent of one another given the random effects $\boldsymbol{\theta}_{(2)}$ and $\boldsymbol{\theta}_{(3)}$.

The integrals on the right-hand side of equations (1) and (2) can be evaluated by the Gauss-Hermite quadrature numerical integration method (Stroud & Secrest, 1966; Rabe-Hesketh, et al., 2001) in which the multivariate normal mixing distribution is approximated by a limited number of discrete points. More precisely, the integrals are replaced by summations over M and T quadrature points

$$\begin{aligned}P_k(\mathbf{y}_k) &\approx \sum_{m=1}^{M} P_k(\mathbf{y}_k \mid \boldsymbol{\theta}_{m(3)})\pi(\boldsymbol{\theta}_{m(3)}) \\ &= \sum_{m=1}^{M}\left[\prod_{j=1}^{n_k} P_{jk}(\mathbf{y}_{jk} \mid \boldsymbol{\theta}_{m(3)})\right]\pi(\boldsymbol{\theta}_{m(3)}) \\ &= \sum_{m=1}^{M}\left[\prod_{j=1}^{n_k}\sum_{t=1}^{T}\prod_{i=1}^{n_{jk}} P_{ijk}(y_{ijk} \mid \boldsymbol{\theta}_{t(2)},\boldsymbol{\theta}_{m(3)})\pi(\boldsymbol{\theta}_{t(3)})\right]\pi(\boldsymbol{\theta}_{m(3)}).\end{aligned}$$

Here, $\boldsymbol{\theta}_{t(2)}$ and $\boldsymbol{\theta}_{m(3)}$ are quadrature nodes and $\pi(\boldsymbol{\theta}_{t(2)})$ and $\pi(\boldsymbol{\theta}_{m(3)})$ are quadrature weights corresponding to the (multivariate) normal densities of interest. Because the random effects are orthogonalised, the nodes and weights of the separate dimensions are equal to those of the univariate normal density, which can be obtained from standard tables (see, for example, Stroud & Secrest, 1966). Suppose that each dimension is approximated by Q quadrature nodes. The $T = Q^{R(2)}$ and $M = Q^{R(3)}$ weights are then obtained by multiplying the weights of the separate dimensions. The integral can be approximated to any practical degree of accuracy by setting Q sufficiently large.

Implementation of the EM algorithm

ML estimation can be performed by an EM algorithm with an E step that is especially adapted to the problem at hand. This adaptation is necessary because a standard implementation of the E step would involve computing the joint conditional expectation of n_k• $R_{(2)} + R_{(3)}$ random effects; that is, the joint posterior distribution $P_k(\boldsymbol{\theta}_{t1(2)},\boldsymbol{\theta}_{t'2(2)},\boldsymbol{\theta}_{t''n(k)(2)},\boldsymbol{\theta}_{m(3)} \mid \mathbf{y}_k)$ with M• $T^{n(k)}$ entries. This is only possible for very small n_k.

Because of the model structure, the next step after obtaining the posterior probabilities would be to compute the marginal posterior probabilities for each level-2 unit, $P_k(\boldsymbol{\theta}_{tj(2)},\boldsymbol{\theta}_{m(3)} \mid \mathbf{y}_k)$, by collapsing over the random effects of the other level-2 units. In other words, in the E step we only need the n_k marginal posterior probability distributions containing M• T entries. This can be seen from the form of the (approximate) complete data log-likelihood, which is defined as

$$\log L_c = \sum_{m=1}^{M}\sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{j=1}^{n_k}\sum_{i=1}^{n_{jk}} P_k(\boldsymbol{\theta}_{tj(2)},\boldsymbol{\theta}_{m(3)} \mid \mathbf{y}_k)\log P_{ijk}(y_{ijk} \mid \boldsymbol{\theta}_{t(2)},\boldsymbol{\theta}_{m(3)}). \tag{3}$$

It turns out that it is possible to compute $P_k(\boldsymbol{\theta}_{tj(2)},\boldsymbol{\theta}_{m(3)}\,|\mathbf{y}_k)$ without going through the full posterior distribution by making use of the conditional independence assumptions associated with the density function defined in equations (1) and (2). In that sense, our procedure is similar to the forward-backward algorithm for the estimation of hidden Markov models with large numbers of time points (Baum et al., 1970; Juang & Rabiner, 1991). Our procedure could be called an upward-downward algorithm. First, random effects are integrated out going from the lower to the higher levels. Subsequently, the relevant marginal posterior probabilities are computed going from the higher to the lower levels.

The marginal posterior probabilities $P_k(\boldsymbol{\theta}_{tj(2)},\boldsymbol{\theta}_{m(3)}\,|\mathbf{y}_k)$ can be decomposed as follows:

$$P_k(\boldsymbol{\theta}_{tj(2)},\boldsymbol{\theta}_{m(3)}\,|\,\mathbf{y}_k) =$$
$$P_k(\boldsymbol{\theta}_{m(3)}\,|\,\mathbf{y}_k)P_k(\boldsymbol{\theta}_{tj(2)}\,|\,\mathbf{y}_k,\boldsymbol{\theta}_{m(3)}).$$

Our procedure makes use of the fact that in the GLTM

$$P_k(\boldsymbol{\theta}_{tj(2)}\,|\,\mathbf{y}_k,\boldsymbol{\theta}_{m(3)}) = P_{jk}(\boldsymbol{\theta}_{tj(2)}\,|\,\mathbf{y}_{jk},\boldsymbol{\theta}_{m(3)});$$

i.e., $\boldsymbol{\theta}_{tj(2)}$ is independent of the observed and latent variables of the other level-2 units within the same level-3 unit given $\boldsymbol{\theta}_{m(3)}$. This is the result of the fact that level-2 observations are mutually independent given the level-3 random effects, as is expressed in the density function described in equation (1). Using this important result, we get the following slightly simplified decomposition

$$P_k(\boldsymbol{\theta}_{tj(2)},\boldsymbol{\theta}_{m(3)}\,|\,\mathbf{y}_k) =$$
$$P_k(\boldsymbol{\theta}_{m(3)}\,|\,\mathbf{y}_k)P_{jk}(\boldsymbol{\theta}_{tj(2)}\,|\,\mathbf{y}_{jk},\boldsymbol{\theta}_{m(3)}).$$

The computation of the marginal posterior probabilities therefore reduces to the computation of the two terms on the right-hand side of this equation. The term $P_k(\boldsymbol{\theta}_{m(3)}\,|\mathbf{y}_k)$ is obtained by

$$P_k(\boldsymbol{\theta}_{m(3)}\,|\,\mathbf{y}_k) = \frac{P_k(\mathbf{y}_k,\boldsymbol{\theta}_{m(3)})}{P_k(\mathbf{y}_k)} \qquad (4)$$

where

$$P_k(\mathbf{y}_k,\boldsymbol{\theta}_{m(3)}) = \pi(\boldsymbol{\theta}_{m(3)})\prod_{j=1}^{n_k} P_{jk}(\mathbf{y}_{jk},\boldsymbol{\theta}_{m(3)})$$

$$P_k(\mathbf{y}_k) = \sum_{m=1}^{M} P_k(\mathbf{y}_k\,|\,\boldsymbol{\theta}_{m(3)}).$$

The other term, $P_{jk}(\boldsymbol{\theta}_{tj(2)}|\mathbf{y}_{jk},\boldsymbol{\theta}_{m(3)})$ is computed by

$$P_{jk}(\boldsymbol{\theta}_{tj(2)}\,|\,\mathbf{y}_{jk},\boldsymbol{\theta}_{m(3)}) = \frac{P_{jk}(\mathbf{y}_{jk},\boldsymbol{\theta}_{tj(2)}\,|\,\boldsymbol{\theta}_{m(3)})}{P_{jk}(\mathbf{y}_{jk}\,|\,\boldsymbol{\theta}_{m(3)})}$$

where

$$P_{jk}(\mathbf{y}_{jk},\boldsymbol{\theta}_{tj(2)}\,|\,\boldsymbol{\theta}_{m(3)}) = \pi(\boldsymbol{\theta}_{t(2)})\prod_{i=1}^{n_{jk}} P_{ijk}(y_{ijk},\boldsymbol{\theta}_{tj(2)},\boldsymbol{\theta}_{m(3)})$$

$$P_{jk}(\mathbf{y}_{jk}\,|\,\boldsymbol{\theta}_{m(3)}) = \sum_{t=1}^{T} P_{jk}(\mathbf{y}_{jk},\boldsymbol{\theta}_{tj(2)}\,|\,\boldsymbol{\theta}_{m(3)}).$$

These equations show that computer storage and time increases only linearly with the number of level-2 observations instead of exponentially, as would be the case with a standard EM algorithm. It can also be seen that the method can easily be generalised to more than three levels. For example, with four levels, one would have to compute the three terms $P_l(\boldsymbol{\theta}_{o(4)}\ |\mathbf{y}_l)$, $P_{kl}(\boldsymbol{\theta}_{mk(3)}\ |\mathbf{y}_{kl},\boldsymbol{\theta}_{o(4)})$, and $P_{jkl}(\boldsymbol{\theta}_{tj(2)}\ |\mathbf{y}_{jkl},\boldsymbol{\theta}_{mk(3)},\boldsymbol{\theta}_{o(4)})$.

A practical problem in the implementation of the E step is that underflows may occur in the computation of $P_k(\boldsymbol{\theta}_{m(3)}\ |\mathbf{y}_k)$. More precisely, the numerator of equation (4) may become equal to zero for each m because it may involve multiplication of a large number, $(n_k + 1)(n_{jk} + 1)$, of probabilities. Such underflows can, however, be prevented by working on a log scale. Letting

$$a_{mk} = \log \pi(\boldsymbol{\theta}_{m(3)}) + \sum_{j=1}^{n_k} \log P_{jk}(\mathbf{y}_{jk}, \boldsymbol{\theta}_{m(3)})$$

and $b_k = \max(a_{mk})$, $P_k(\boldsymbol{\theta}_{m(3)}\ |\mathbf{y}_k)$ can be obtained by

$$P_k(\boldsymbol{\theta}_{m(3)} \mid \mathbf{y}_k) = \frac{\exp(a_{mk} - b_k)}{\sum_{p=1}^{M} \exp(a_{mk} - b_k)}$$

In the M step of the EM algorithm, the (approximate) complete data log-likelihood described in equation (3) is improved by standard complete data algorithms for the ML estimation of generalised linear models.

Standard errors

Contrary to Newton-like methods, the EM algorithm does not provide standard errors of the model parameters as a by-product. Estimated asymptotic standard errors can be obtained by computing the observed information matrix, the matrix of second-order derivatives of the log-likelihood function for all model parameters. The inverse of this matrix is the estimated variance-covariance matrix. For the example presented in the next section, we computed the necessary derivatives numerically.

The information matrix can also be used to check identifiability. A sufficient condition for local identification is that all the eigenvalues of this matrix are larger than zero.

Application to attitudes towards abortion data

To illustrate the GLTM, we obtained a data set from the data library of the Multilevel Models Project, at the Institute of Education, University of London.
http://multilevel.ioe.ac.uk/intro/datasets.html
The data consist of 264 participants in 1983 to 1986 yearly waves from the British Social Attitudes Survey (McGrath and Waterton, 1986). It is a three-level data set: individuals are nested within districts and time points are nested within individuals.

The dependent variable is the number of yes responses on seven yes/no questions as to whether it is a woman's right to have an abortion under a specific circumstance. Because this variable is a count with a fixed total, it is most natural to work with a logit link and a binomial error function. Individual level predictors in the data set are religion, political preference, gender, age, and self-assessed social class. In accordance with the results of Goldstein (1995), we found no significant effects of gender, age, self-assessed social class, and political preference. Therefore, we did not use these predictors in the further analysis. The predictors that were used are the level-1 predictor year of measurement (1=1983; 2=1984; 3=1985; 4=1986) and the level-2 predictor religion (1=Roman Catholic, 2=Protestant; 3=Other; 4=No religion). Because there was no evidence for a linear time effect, we included time as a set of dummies in the regression model.

Table 1. Fit measures, parameter estimates and standard errors for the estimated models

	Model I	Model II	Model III
Fit measures			
LL	-2188.38	-1711.76	-1708.72
# parameters	7	8	9
BIC	4425.5	3479.21	3480.09
Fixed effects			
Intercept	1.50 (0.07)	1.97 (0.13)	2.09 (0.18)
Time			
1983	-0.13 (0.08)	-0.16 (0.08)	-0.16 (0.08)
1984	-0.55 (0.07)	-0.68 (0.08)	-0.68 (0.08)
1985	-0.22 (0.08)	-0.27 (0.08)	-0.27 (0.08)
Religion			
Catholic	-1.08 (0.10)	-1.07 (0.21)	-1.59 (0.32)
Protestant	-0.38 (0.06)	-0.49 (0.19)	-0.71 (0.21)
Other	-0.82 (0.08)	-1.12 (0.17)	-1.32 (0.24)
Random intercepts			
Level-2 standard deviation		1.20 (0.05)	1.21 (0.07)
Level-3 standard deviation			0.47 (0.33)

Table 1 reports the results obtained with the three models that were estimated: a model without random effects (Model I), a model with a level-2 random intercept (Model II), and a model with level-2 and level-3 random intercepts. We approximated the integrals in the log-likelihood function using 10 quadrature nodes per dimension. The reported fit measures show that the level-2 variance is clearly significant. Based on a log-likelihood difference test between Models II and III, one would conclude that the level-3 variance is just significant. The BIC, on the other hand, indicates that Model II is somewhat better than Model III.

The lower part of Table 1 contains the parameter estimates for Models I, II, and III. As far as the fixed part is concerned, the substantive conclusions would be similar in all three models. The attitudes are most positive at the last time point (reference category) and most negative at the second time point. Furthermore, the effects of religion show that people without religion (reference category) are most in favour and Roman Catholics and Others are most against abortion. Protestants have a position that is close to the no-religion group. As can be seen, introducing a level-2 variance term increases the time effects and introducing a level-3 variance term increases the religion effects.

A natural manner to quantify the importance of the random intercept terms is by their contribution to the total variance. The level-1 variance can be set equal to the variance of the logistic distribution ($\pi^2/2$=3.29), yielding a total variance equal to $3.29+1.21^2+0.47^2$=4.98. Thus, after controlling for the time and religion effects, the level-2 and level-3 variances equal 29% ($1.21^2/4.98$) and 4% ($0.49^2/4.98$) of the total variance, respectively.

Discussion

An EM algorithm was presented for the ML estimation of GLTMs. This upward-downward method prevents the need to process the full posterior distribution, which becomes infeasible with more than a few level-2 units per level-3 unit. The relevant marginal posterior distributions can be obtained by making use of the conditional independence assumptions underlying the GLTM. As was shown, it is straightforward to generalise the method to models with more than three levels.

A limitation of the GLTM is that the numerical integration to be performed for parameter estimation can involve summation over a large number of points when the number of random effects is increased. Despite the fact that the number of points per dimension can be somewhat reduced with multiple random effects, computational burden becomes enormous with more than five or six random coefficients. There exist other methods for computing high-dimensional integrals, like Bayesian simulation and simulated likelihood methods, but these are also computationally intensive. As shown by Vermunt and Van Dijk (2001), these practical problems can be prevented by using a nonparametric random-effects model in which the mixing distribution is approximated with a small number of nodes whose locations and weights are unknown parameters to be estimated (Laird, 1978). The proposed EM algorithm can also be used for the estimation of such nonparametric GLTMs.

References

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

Baum, L.E., Petrie, T., Soules, G. & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, **41**: 164-171.

Goldstein, H. (1995) *Multilevel Statistical Models*. New York: Halsted Press.

Juang, B.H. & Rabiner, L.R. (1991). Hidden Markov models for speech recognition. *Technometrics*, **33**: 251-272.

Hedeker, D. and Gibbons. R.D. (1996). MIXOR: A computer program for mixed effects ordinal regression analysis. *Computer Methods and Programs in Biomedicine,* **49**: 157-176.

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

Lesaffre, E. and Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random-effects model: an example. *Applied Statistics*, **50**: 325-335.

McGrath, K and Waterton, J. (1986). *British Social Attitudes, 1983-1986 Panel Survey*. London: Social and Community Planning Research, Technical Report.

Rabe-Hesketh, S., Pickles, A. and A. Skrondal (2001). GLLAMM: A general class of multilevel models and a Stata program. *Multilevel Modelling Newsletter,* **13**: 17-23.

Stroud, A.H. & Secrest. D. (1966). *Gaussian Quadrature Formulas*. Englewood Cliffs, NJ: Prentice Hall.

Vermunt, J.K. & Van Dijk. L. (2001). A nonparametric random-coefficients approach: the latent class regression model. *Multilevel Modelling Newsletter*, **13**, 6-13.

Review of 'Modeling Intraindividual Variability with Repeated Measures Data: Methods and Applications'. Moskowitz, D. S., and Hershberger, S. L. (Eds.). Mahwah NJ: Lawrence Erlbaum Associates Inc.

ISBN: 0-8058-3125-8, pp. 276.

Ian Plewis

Institute of Education, University of London

Increasingly, researchers are seeing the value of a multilevel approach to the analysis of longitudinal data, especially when those data are repeated measures of the same underlying construct. The approach offers an appealing way of modelling change, of relating change parameters across variables and of using all the available data in analyses. The links between multilevel models and structural equations models (SEMs) are becoming recognised, giving researchers a wider range of models that will account for the observed variances and covariances at different levels. Many of these approaches are represented in this curate's egg of a collection of nine chapters.

The first chapter (Kenny et al.) shows how simple two level models for repeated measures relate to traditional

ANOVA methods and how Ordinary Least Squares (OLS), WLS and iterative least squares methods can be used for estimation. Their choice of an example to motivate the chapter is a little odd, not least because they average over some of their repeated measures. There then follows a solid chapter by Raudenbush, firmly located within the multilevel tradition and showing how different model specifications for growth, each making different assumptions about the within subject covariances, can be tested against a full multivariate model. Chapter 3 (Curran and Hussong) develop models for relating changes in anti-social behaviour to changes in reading attainment within an SEM framework although the specifications of some of these models are somewhat curious. Ramsay then provides a nicely written overview of these three chapters in which he steps outside the multilevel/SEM frameworks to show how they relate to ideas in functional data analysis.

There is not a lot in the remaining five chapters to interest multilevel modellers. Wallace and Green and Singer focus on estimating mixed models using PROC MIXED in SAS, the latter chapter essentially a reprise of her useful article in the Journal of Educational and Behavioral Statistics. Duncan et al. compare different estimation procedures for analysing family and adolescent data on alcohol use but their decision to base these comparisons on complete data over a four year period reduces its value. The final two chapters are on time series methods (Hillmer) and factor analysis (Nesselroade et al.).

The collection is restricted to models for outcomes assumed to be measured on a continuous scale and to nested models. It suffers from what might charitably be described as a light touch from the editors in that there is no consistency in notation and little attempt to link the chapters to create a coherent whole. The production quality is poor – there are too many typos and the figures, especially those in Chapters 7 and 9, are so bad that anyone spending money on this book could reasonably ask for a refund.

Some Recent Publications Using Multilevel Models

Fielding, A. (2002). Ordered Category Responses and Random Effects in Multilevel and Other Complex Structures. In Reise, S., and Duan, N. (Eds.), *Multilevel Modeling: Methodological Advances, Issues and Applications*, pp 181-208. Mahwah NJ: Lawrence Erlbaum Associates Inc.

Goldstein, H., Browne, W. J. and Rasbash, J. (2002). Multilevel modelling of medical data. *Statistics in Medicine*, **21**: 3291-3315.

Moerbeek, M., Van Breukelen, G. J. P., and Berger, M. P. F. (2001). Optimal experimental designs for multilevel models with covariates.

Communications in Statistics - Theory and Methods, **30**: 2683-2697.

Moerbeek, M., & Wong, W. K. (2002). Multiple-objective optimal designs for the hierarchical linear model. *Journal of Official Statistics*, **18**: 291-303.

Rabe-Hesketh, S., Skrondal, A. and Pickles, A. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, **2**: 1-21.

Reise, S., and Duan, N. (Eds.) (2002). *Multilevel Modeling: Methodological Advances, Issues and Applications*. Mahwah NJ: Lawrence Erlbaum Associates Inc.

Rothman, S. (2002). School absence and student background factors: a multilevel analysis. *International Education Journal*, **2** (1): 59-68. http://iej.cjb.net.

Please send us your new publications in multilevel modelling for inclusion in this section in future issues.

MLwiN Clinics in London

Tuesday 17 December 2002
Tuesday 7 January 2003
Tuesday 4 February 2003
Tuesday 4 March 2003
Tuesday 8 April 2003
Tuesday 6 May 2003

at

Centre for Multilevel Modelling
11 Woburn Square, London WC1H 0NS

Contact Min Yang for appointments
Tel: +44 (0) 20 7612 6682
m.yang@ioe.ac.uk
http://multilevel.ioe.ac.uk/support/clinics.html