MULTILEVEL MODELLING NEWSLETTER

Multilevel Models Project
Mathematical Sciences
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. 13 No. 1 **June, 2001**

Forthcoming Workshop

5-7 September 2001, a three-day introductory workshop to multilevel modelling using *MLwiN* will take place in the Institute of Education, University of London.

This workshop can be booked on-line: http://multilevel.ioe.ac.uk/support/workshop.html.

Enquiries to Amy Burch at Mathematical Sciences, Institute of Education, 20 Bedford Way, London WC1H OAL. Tel: +44 (0)20 7612 6688, Fax: +44 (0)20 7612 6572, email: a.burch@ioe.ac.uk.

International Statistical Institute, 53rd Biennial Session

This conference is to be held in Seoul, South Korea on 22-29 August 2001. The following papers have been invited for presentation in the session 'Multilevel Models for Survey Design and Analysis':

1. Cohen, M. P. *Sample Size Considerations for Multilevel Surveys*. michael.cohen@bts.gov

2. Steele, F. *Selection Effects in an Analysis of Contraceptive Discontinuation in Morocco: An Application of a Multiprocess Multilevel Model*. f.steele@lse.ac.uk

3. Pfeffermann, D., Moura, F., and do Nascimento Silva, P. L. *Multilevel Modelling under Informative Sampling*. msdanny@mscc.huji.ac.il

aLCD analysis OF LARGE AND COMPLEX DATASETS

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

Third International Amsterdam Conference on Multilevel Analysis

The above conference was held in Amsterdam, Netherlands between 9 and 10 April. The following papers were presented:

1. McLeod, A. *Methods of inference for random parameters in generalized multilevel models.* a.mcleod@psychology.bbk.ac.uk

2. Fielding, A. *The scaling of residual variance components in generalised multilevel models for ordinal responses*. a.fielding@bham.ac.uk

3. Grilli, L., and Rampichini, C. *Specification issues in variance component ordinal response models*. carla@ds.unifi.it

4. Leyland, A. H. *Approximating likelihood functions for multilevel Poisson models*. a.leyland@msoc.mrc.gla.ac.uk

5. van Dijk, L. A., and Vermunt, J. K. *Modelling dependent observations; comparison of parametric and non-parametric random-coefficient approaches.* l.a.vandijk@kub.nl

6. Plewis, I. *Multilevel event history analysis applied to nursing careers*. i.plewis@ioe.ac.uk

7. Rabe-Hesketh, S. *Generalising the generalized linear model*. sophia@moonlite.demon.co.uk

8. Mazzolli, B. *A multilevel structural equation model with polytomous and dichotomous data*. mazzolli@ds.unifi.it

9. Herrin, J., Scheel, I., Hagen, K. B., and Oxman, A. *A cluster randomized three-arm trial of strategies to encourage use of active sick leave among back pain patients*. jherrin@online.no

10. Stryhn, H., Dohoo, I. R., and Tillard, E. *The use of multilevel models to evaluate sources of variation in reproductive performance in dairy cattle.* hes@svs.dk

11. Lewsey, J. D., Gilthorpe, M. S., and Gulabivala, K. *Meta-regression of success of root canal treatment*. j.lewsey@eastman.ucl.ac.uk

12. Stevenson, M. A., Perkins, N. R., McKay, B. J., Hitchcock, L., and Marchant, R. *Multilevel models to evaluate variation in pericalving mastitis risk in New Zealand dairy cattle*. m.stevenson@massey.ac.nz

13. Fox, J-P. *Bayesian modeling of measurement error using item response theory*. foxj@edte.utwente.nl

14. Cheong, Y. F. *Detecting differential item functioning in problem behavior items via multilevel and multidimensional Rasch models*. ycheong@emory.edu

15. Ravanera, Z. R., Rajulton, F., and Burch, T. K. *Multilevel influences on early life transitions of Canadian youth*. ravanera@julian.uwo.ca

16. Rajulton, F., and Ravanera, Z. R. *Community characteristics and family support in Canada*. fernando@julian.uwo.ca

17. Goldstein, H., and Browne, W. J. *Fitting models to complex data involving hierarchical, crossed and multiple membership structures*. h.goldstein@ioe.ac.uk

18. Ecob, R., and Der, G. *An iterative method for the detection and elimination of outliers in longitudinal data using multilevel models*. russell_ecob@hotmail.com

19. Bell, J. F. *Exploratory data analysis of multilevel models using XLSIP-STAT*. bell.j@ucles.org.uk

20. Moerbeek, M. *Design issues for multilevel experiments*. mirjam.moerbeek@rivm.nl

21. Hox. J. J. *Power analysis in multilevel regression models*. j.hox@fss.uu.nl

22. Ma, X. *A multivariate, multilevel model for estimating the stability of socioeconomic gaps in academic achievement across multiple school subjects*. xin.ma@ualberta.ca

23. Riether, M. M. *Possible effects of informative probability sampling on modeling*. mriether@inep.gov.br

24. Bosker, R. J., Rekers-Mombarg, L. T. M. and Béguin, A. A. *Which school caused the effect? Using weights in multilevel multiple membership models*. r.j.bosker@edte.utwente.nl

25. Hutchison, D., Morrison, J., and Felgate, R. *Correcting for measurement error in complex survey data using bootstrapping procedures*. d.hutchison@nfer.ac.uk

26. Barbosa, M. E., Beltrão, K. Fernandes, C., Santos, D., and Suarez, M. *Multilevel models applied on educational data assessment for comparative analysis of Brazilian regions*. mariabarbosa@ibge.gov.br

27. Ahouagi Vaz de Magalhães, D. J. *A multilevel approach to the analysis of the residential location choice in Belo Horizonte metropolitan area (Brazil)*. david@etg.ufmg.br

28. Maas, C. J. M., and Hox, J. J. *The influence of violations of model assumptions on multilevel parameter estimates*. c.maas@fss.uu.nl

29. Candel, M. J. J. M., and Winkens, B. *Performance of empirical Bayes estimators of level-2 random parameters in multilevel analysis: a Monte Carlo study for longitudinal designs*. math.candel@stat.unimaas.nl

30. Bellio, R., and Gori, E. *Impact evaluations based on multilevel modelling with application to job training programmes*. gori@dss.uniud.it

31. Jordan, K., Jones, P. W., Dziedzic, K., Ong, B. N., and Dawes, P. T. *Multilevel modelling of movements of the neck and of the shoulder*. k.p.jordan@cphc.keele.ac.uk

An MCMC algorithm for adjusting for errors in variables in random slopes multilevel models

William J. Browne, Harvey Goldstein, Geoff Woodhouse and Min Yang
Institute of Education, University of London

Introduction

The problem of adjusting for errors in measurement of predictor variables when fitting a statistical model affects virtually all statistical analysis. Errors in predictor variables can throw up spurious significant relationships or, more commonly, mask the extent of effects. Classical frequentist approaches for adjusting for measurement error can be found in, for example, Fuller (1987). In the case of multilevel modelling with random effects these techniques become mathematically very complex (see Woodhouse et al., 1996) and consequently computationally slow. In particular, in the case where there are errors of measurement in a predictor that also has a random coefficient, likelihood and moment based techniques become intractable (Woodhouse, 1996) and an alternative approach is required.

With the advent of faster computers, computationally intensive statistical techniques like Markov Chain Monte Carlo (MCMC) estimation algorithms can be routinely used to fit complex model structures to large datasets. Bayesian measurement error models based on MCMC estimation methods have been described by several authors for particular applications. Richardson (1996) provides a review of much of this research.

In this paper we look at some of the empirical properties of classical Bayesian measurement error models when applied to a random coefficients regression model. We will fit 1,000 simulation datasets based on an underlying dataset taken from education. Our example dataset (Rasbash et al., 2000) consists of data on 4059 pupils from 65 schools. For these pupils we have a response that is a (normalized) total exam score at age 16. As a predictor variable we have a reading test that each child took at age 11 before entering secondary school. In Chapter 2 of Rasbash et al. (2000) a random coefficient regression model is fitted where, for each school, we fit a random intercept and random coefficient for the reading test score. We will consider here the problem of 'known' measurement error in this reading test score and how this affects the fixed effect and the variance parameter estimates we obtain.

We will use this design to create simulated responses that are based on our model formulation and compare the performance of our approach with approaches that do not account for measurement error, with comparisons both in terms of bias and coverage properties. For the MCMC method we use a 'burn-in' of 500 iterations and a main run of 10,000 iterations for each dataset as this seemed adequate from initial investigations. Before looking at the simulation experiment in more detail we will give a Gibbs sampling algorithm to fit this model with classical measurement error.

A Gibbs sampling algorithm for a two level random slopes regression model with measurement error in the predictor

Browne and Draper (2000) give Gibbs sampling algorithms for fitting a two level random slopes regression model. Here we will consider a general two level random coefficient regression model with measurement error in the single *X* variable and this can then be easily generalised to any *N* level multilevel Normal model in a similar way to the case with no measurement error which is detailed in Browne (1998).

Our model is as follows

$$y_{ij} = X_{0ij}\beta_0 + X_{1ij}\beta_1 + X_{0ij}u_{0j} + X_{1ij}u_{1j} + e_{ij}$$
$$e_{ij} \sim N(0,\sigma_e^2), u_j \sim MVN(0,\Omega_u),$$
$$X_{1ij}^o \sim N(X_{1ij},\sigma_m^2), X_{1ij} \sim N(\theta,\phi^2)$$

Here we have a vector of 2 fixed effects $\beta = (\beta_0,\beta_1)^T$, a vector of random effects (residuals) for each of *j* level 2 units $u_j = (u_{0j},u_{1j})^T$, and a level 1 residual e_{ij} for each observation. The observed predictor variable we will refer to as X_{1ij}^o; the observations of this variable contain independent errors with error variance σ_m^2. The true value without measurement error we will denote by X_{1ij}. For ease of notation the intercept explanatory variable for each individual will be referred to as $X_{0ij} = X_{0ij}^o = 1 \forall i, j,$ and we will then write $X_{ij} = (X_{0ij}, X_{1ij})$ to denote all our 'true' predictor values. The Gibbs sampling procedures for fitting multilevel models proceed most smoothly by treating both the level 2 residuals and the true predictor values as latent variables when forming the full conditional posterior distributions.

We will assume that the following general priors are used: for the level 1 variance a general scaled inverse χ^2 prior with parameters ν_e and s_e^2, for the level 2 variance matrix Ω_u an inverse Wishart prior with parameters ν_u and S_u, and for the fixed effects $\beta \sim N(\mu_P, S_p)$. The parameter θ has a Uniform prior while ϕ^2 has a general scaled inverse χ^2 prior with parameters, ν_ϕ and s_ϕ^2.

The algorithm then has seven Gibbs sampling steps. The first 4 steps are identical to the algorithm in Browne and Draper (2000) for the case with no measurement error:

Step 1 - The fixed effects, β

$$p(\beta \mid y, X, u, \Omega_u, \sigma_e^2) \sim MVN(\hat{\beta}, \hat{D})$$

where

$$\hat{D} = \left[\frac{\sum_{ij} (X_{ij})^T X_{ij}}{\sigma_e^2} + S_P^{-1} \right]^{-1} \text{ and}$$

$$\hat{\beta} = \hat{D} \left[\frac{\sum_{ij} (X_{ij})^T (y_{ij} - X_{ij} u_j)}{\sigma_e^2} + S_P^{-1} \mu_p \right].$$

Step 2 - The level 2 residuals, u_j

$$p(u_j \mid y, X, \beta, \Omega_u, \sigma_e^2) \sim MVN(\hat{u}_j, \hat{D}_j)$$

where

$$\hat{D}_j = \left[\frac{\sum_{i=1}^{n_j} (X_{ij})^T X_{ij}}{\sigma_e^2} + \Omega_u^{-1} \right]^{-1} \text{ and}$$

$$\hat{u}_j = \frac{\hat{D}_j}{\sigma_e^2} \sum_{i=1}^{n_j} (X_{ij})^T (y_{ij} - X_{ij}\beta)$$

Step 3 - The level 1 variance, σ_e^2

$$p(1/\sigma_e^2 \mid y, X, \beta, u, \Omega_u) \sim Gamma(a_e, b_e)$$

where

$$a_e = \frac{N + \nu_e}{2}, \text{ and } b_e = \frac{1}{2}(\nu_e s_e^2 + \sum_{i,j} e_{ij}^2).$$

Step 4 - The level 2 variance, Ω_u

$$p(\Omega_u^{-1} \mid y, X, u, \beta, \sigma_e^2) \sim Wishart_{n_2}[\nu_{pos}, S_{pos}]$$

$$\nu_{pos} = J + \nu_u, S_{pos} = (\sum_{j=1}^{J} u_j (u_j)^T + S_u)^{-1}$$

where n_2 is the number of random variables at level 2. If a uniform prior is instead required then set S_u = 0 and $\nu_u = -n_2 - 1$.

Step 5. The true predictor value, X_{1ij}

$$p(X_{1ij} \mid y, X_{1ij}^o, u, \beta, \sigma_e^2, \Omega_u, \sigma_m^2, \theta, \phi^2) \sim N(\hat{X}_{1ij}, \hat{V}_{1ij})$$

where

$$\hat{V}_{1ij} = \left[\frac{(\beta_1 + u_{1j})^2}{\sigma_e^2} + \frac{1}{\sigma_m^2} + \frac{1}{\phi^2} \right]^{-1} \text{ and}$$

$$\hat{X}_{1ij} = \hat{V}_{1ij} \left[\frac{(\beta_1 + u_{1j})(y_{ij} - \beta_0 - u_{0j})}{\sigma_e^2} + \frac{X_{1ij}^o}{\sigma_m^2} + \frac{\theta}{\phi^2} \right].$$

Step 6. The mean parameter for X_1, θ

$p(\theta \mid X_{1ij}, \phi^2) \sim N(\hat{\theta}, \hat{V}_\theta)$ where

$\hat{V}_\theta = \frac{\phi^2}{N}$ and $\hat{\theta} = \frac{\sum_{ij} X_{1ij}}{N}$.

Step 7. The variance parameter for X_1, ϕ^2

$$p(1/\phi^2 \mid X_{1ij}, \theta) \sim Gamma(a_\phi, b_\phi)$$

where $a_\phi = \frac{N + \nu_\phi}{2}$ and

$$b_\phi = \frac{1}{2}(\nu_\phi s_\phi^2 + \sum_{i,j}(X_{1ij} - \theta)^2).$$

Note that this algorithm will be quite a bit slower than the analogous algorithm for the model without measurement error as quantities such as $\sum_{ij}(X_{ij})^T X_{ij}$ will not be constant through iterations and will have to be evaluated at each iteration. Of course if we have more predictors without measurement error then some elements of this matrix may be constant and so we may only have to partially recalculate this matrix. In the example that follows running for 10,000 iterations without adjusting for measurement error takes 51 seconds on a 733MHz Pentium; adjusting for measurement error takes 2 minutes 31 seconds, a roughly 3-fold increase in time.

Simulation Design and Results

For our simulation experiment we consider estimates for both the fixed effect and variance parameters that are similar to the true dataset. We will use 'diffuse' Uniform priors for the fixed effects parameters and a *Gamma(ε,ε)* prior for $1/\sigma_e^2$. For the level 2 variance matrix we follow the example of Spiegelhalter et al. (2000) in their 'birats' example and use a vaguely informative inverse Wishart prior with parameters, $\nu = 2$ and $S = \frac{1}{10} * I_2$. Here I_2 is the 2*2 identity matrix.

To run the simulation experiment 1000 datasets were generated. For each dataset we generated a response vector using the true predictor variables and the true parameter values and generating random level 1 and 2 residuals from their respective distributions. We then generated an observed predictor for each individual by generating a measurement error from the distribution $N(0, \sigma_m^2)$ and adding this to the true predictor to complete the observed dataset. The true predictor variable has variance 1.0 and we will use for our simulation a measurement error variance of 0.2 that represents a reliability of 0.83 which is not unusual for educational data.

Each of the 1000 datasets are fitted using both a model that assumes no measurement error and the Bayesian measurement error model described above in a development version of *MLwiN* (Rasbash et al., 2000). The results of the simulations are shown in Table 1. We have used the restricted maximum likelihood based RIGLS methods in the 'no measurement error' case. This will give slightly different estimates and coverage intervals to the equivalent Gibbs sampling method (see Browne and Draper, 2000) but these will be small when compared to the effect of ignoring measurement error seen in Table 1.

In Table 1 we see that not accounting for measurement errors in this problem gives estimates with large biases. Fixed effect parameters have smaller coefficients. The school level variance matrix also has smaller values as the school effects are also diluted by the unaccounted for measurement error. On the other hand the additional measurement error is captured at level 1 leading to an overestimate of the level 1 variance.

Table 1
Results of 1000 datasets generated with known measurement error from a random slopes regression model.

Parameter	True Value	Assumed no Measurement Error	Accounted for Measurement Error
β_0 (Intercept)	0.0	-0.0035 (0.0014)	-0.0035 (0.0014)
β_1 (Reading Test)	0.6	0.4914 (0.0006)	0.5912 (0.0007)
Ω_{u00}	0.09	0.0875 (0.0007)	0.0934 (0.0006)
Ω_{u01}	0.02	0.0150 (0.0003)	0.0189 (0.0003)
Ω_{u11}	0.02	0.0130 (0.0002)	0.0241 (0.0002)
σ^2_e	0.6	0.6632 (0.0006)	0.6002 (0.0005)

Actual Coverage of Nominal 90%/95% Coverage Intervals

β_0 (Intercept)	-	86.9%/92.5%	88.6%/93.9%
β_1 (Reading Test)	-	0.0%/0.1%	89.1%/94.4%
Ω_{u00}	-	15.5%/18.1%	90.2%/95.1%
Ω_{u01}	-	15.4%/18.3%	89.5%/95.5%
Ω_{u11}	-	18.9%/22.8%	90.6%/95.1%
σ^2_e	-	8.2%/15.3%	91.6%/96.2%

When the (assumed Gaussian) coverage intervals are considered we see that apart from the intercept parameter all the other parameters have very poor coverage. In fact for the predictor with measurement error, β_1, only one of the thousand datasets has a 95% interval that covers the true value.

When the Bayesian measurement error model is fitted we see vast improvements. The biases in the fixed effects and the level 1 variance are far smaller. The level 2 variances have (slight) positive biases but this is due to the estimates quoted being posterior means and not modes (see Browne and Draper, 2000). The coverage interval properties are very impressive with all parameters having interval coverage approximately equal to the designated interval coverage.

Discussion

In this paper we have shown how, through the addition of three extra steps to a standard Gibbs sampling algorithm for a random coefficient regression model, we can account for measurement errors in predictors with a known variance. The benefits, both in terms of decreased estimate bias and far superior interval estimates, are evident from the results in Table 1. It should be noted that we used errors in the predictor that would not be unusual in educational data, and even such errors produce estimates with extremely poor coverage when they are not accounted for. Our method relies on the ability of the researcher to estimate the measurement error variance and for that estimate to be close to the actual error variance (in the example the two are equal). In theory the measurement error variance can be estimated by taking repeated measurements of the predictor in question, for a sample of individuals, and estimating the measurement error within this sample. In practice, however, it is difficult to get an accurate estimate and other approaches such as classical test item analysis are used.

It should be noted that the above algorithm can easily be modified to include additional predictors with or without error (assuming independent measurement errors) and additional levels at which the predictor can be random. This ease of modification gives the Bayesian approach a distinct advantage over standard frequentist approaches which are mathematically cumbersome for these complex models. We intend to consider MCMC algorithms for this and additional multilevel modelling based measurement error problems in a subsequent paper.

References

Browne, W. J. (1998). *Applying MCMC Methods to Multilevel Models*. PhD dissertation, Department of Mathematical Sciences, University of Bath, UK.

Browne, W. J., and Draper, D. (2000). Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models. *Computational Statistics*, **15**: 391-420.

Fuller, W.A. (1987). *Measurement Error Models*. New York: Wiley.

Rasbash, J., Browne, W. J., Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., Draper, D., Langford, I., and Lewis, T. (2000). *A User's Guide to MLwiN*, Version 2.1, London: Institute of Education, University of London.

Richardson, S. (1996). Measurement Error. In W.R. Gilks, S. Richardson and D.J. Spiegelhalter (Eds.), *Markov chain Monte Carlo in practice*, pp 401-417., London: Chapman and Hall.

Spiegelhalter, D. J., Thomas, A., and Best, N.G. (2000). *WINBUGS Version 1.3: User Manual.* Cambridge: Medical Research Council Biostatistics Unit.

Woodhouse, G., Yang, M., Goldstein H., and Rasbash, J. (1996). Adjusting for measurement error in multilevel analysis. *Journal of the Royal Statistical Society, Series A*, **159**: 201-212.

Woodhouse, G. (1996). *Errors in Variables in Multilevel Models*. PhD dissertation. Institute of Education, University of London, UK.

Random coefficients models for control parameters in dynamical systems

Steven M. Boker[1] and Paolo Ghisletta[2]

[1]The University of Notre Dame

[2]Max Planck Institute for Human Development

Introduction

Dynamical systems models have been gaining increasing acceptance as testable representations for psychological, sociological and physiological processes. The basic idea behind these models is that some evolving process is at work which can be measured from repeated observations of one or more variables. This *intraindividual variability* is thus presumed to have a structure which can be modelled (Baltes and Nesselroade, 1979). Early models in psychology for this within person variation, such as P–Technique did not include any time dependency, but did recognise the need for consideration of the contributions of within person versus between person variance without specifically solving the problem of multilevel nesting (Cattell, 1963).

More recently, a number of types of models for intraindividual variability have been developed that explicitly incorporate time dependency. For instance, the dynamic factor model (Molenaar, 1985) and similar variants (e.g. McArdle, 1988) use time–lagged covariance matrices to test for time dependent structure in data. These models tend to need to make assumptions of homogeneity across individuals in order to build their occasion–lagged data sets from only a few occasions of measurement per individual.

Differential equations models are another approach to dealing with time dependence in data (Arminger, 1986). These models propose simultaneous relationships between the values of variables and their derivatives. These models are appealing because they can model complex behaviour with few parameters and the resulting parameters often have concrete and understandable interpretations in terms of how a given system regulates its behaviour (Boker and Graham, 1998).

Consider one of the simplest differential equation models, a linear oscillator with dampening (see e.g. Hubbard and West, 1991). This equation describes the motion of an idealised pendulum where the second derivative (the acceleration or curvature) of the pendulum is a linear combination of the first derivative (the velocity or slope) and the offset from equilibrium (the displacement or value of the variable after centering)

$$\ddot{x} = b_1\dot{x} + b_2 x + e \qquad (1)$$

where x is displacement about an equilibrium point, $\dot{x}$ and $\ddot{x}$ are the first and second derivatives of x with respect to time, and e is error of measurement. The two parameters in this equation, b_1 and b_2, are called *control parameters* and have specific interpretations about how the system evolves in time. In this case, b_1 indicates how rapidly the system comes back to equilibrium (if negative) or moves away from equilibrium (if positive) and b_2 indicates how rapidly a system oscillates (if negative) or how sharp is the difference between two competing endpoints (if positive). In the case of the idealised pendulum with friction both parameters are negative and b_2 is indicative of the amount of friction in the pivot and b_1 is indicative of the length of the pendulum. In this case, the motion described is similar to that shown in Figure 1.

Consider the model in Equation 1 as it applies to a self–regulating process. If the two parameters are negative, they have appealing interpretations. The second parameter, b_2, suggests that as the displacement from a desired equilibrium point becomes larger, there is a proportional tendency to produce a negative curvature that turns the system's evolving trajectory around and moves it back towards equilibrium. But near equilibrium, the displacement is near 0, so this parameter doesn't have much effect. In this way, the more negative b_2 becomes, the faster the trajectory will turn around and thus, the faster the system will oscillate around its equilibrium point.

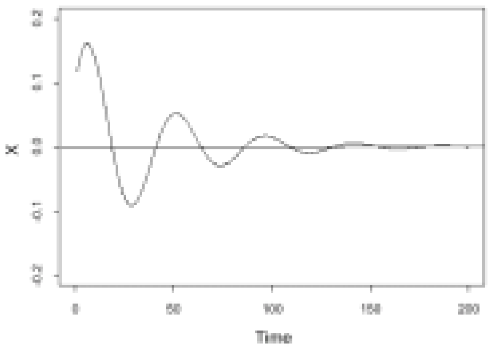

Figure 1
A dampened oscillator. As the value of x diverges from zero, the curvature increases until x finally switches direction, but also as the slope becomes large, the curvature increases resulting in dampening.

Notice that in Figure 1 when the system's trajectory passes 0 the slope is at a maximum value, either positive or negative. Since the parameter b_1 will have its largest effect when the slope is the greatest, the system will tend to slow down as it nears the equilibrium point. Thus a negative value of b_1 will produce dampening, tending to reduce the amplitude of the oscillations until the system tends to remain close to equilibrium. Negative values for the b_1 parameter might be considered to be something like resiliency in psychological terms or as friction in the pendulum example above.

It seems reasonable that there might be individual or group differences in these two parameters. For instance, young individuals might be expected to show greater resiliency than older individuals. Or women might be expected to show different frequency of oscillation in a variable than men. Social groups might tend to show differing levels of response to negative events such that resiliency covaried with some measure of social cohesion. Thus it seems reasonable that random coefficients techniques with hierarchical grouping might be usefully applied to differential equations models (Boker and Nesselroade, 2000). A concrete illustration of how this may be accomplished is described below, using data from an experiment in the development of postural control for sitting in infants. These data were chosen for two reasons: (a) there is a strong hypothesis predicting linear oscillations in the data due to a manipulation of the visual environment of the infants, and (b) infant data are extremely noisy and so if we can estimate reliable predictors from these data, other noisy data may be tractable as well.

Methods

An experiment was performed in which infants' postural adjustments were measured in response to a moving visual stimulus (Bertenthal et al., 1997). Infants were seated in a plastic bicycle seat that was placed on a fixed forceplate such that changes in centre of pressure (COP) of the infants postural adjustments could be continuously recorded. Around the infants was a 1.2 m x 1.2 m x 2.1 m open–ended enclosure, which could be moved in a sinusoidal manner along the anterior/posterior (AP) axis of the infant. AP COP was recorded from the forceplate at 50 Hz (observations per second) and stored to a computer along with synchronously recorded movement of the room.

Forty infants participated in the study, 10 in each of 4 age groups: 5, 7, 9, and 13 months. During each trial in the data analysed here, the room was moved at one of two frequencies, 0.3 Hz or 0.6 Hz. Each trial lasted approximately 10 seconds and each infant received 4 exposures to each frequency. Data were trimmed to remove the first 2 seconds of each trial so as to eliminate the transient movement associated with the beginnings of trials. The next 256 data points (5.1 seconds) were selected for analysis thus eliminating transient movements associated with the ends of trials.

Analysis

In order to fit a model to first and second derivatives of a variable, these derivatives must first be estimated. Several methods for derivative estimation are in use including state–space embedding (Sauer et al., 1991), nonparametric smoothing (Cleveland and Devlin, 1988), and linear interpolation. Here we will use linear interpolation which is a simple calculation and which simulations have shown to be an effective method for use with differential models (Boker and Nesselroade, in press). The first and second derivatives, $\dot{x}(t)$ and $\ddot{x}(t)$ for time t can be estimated from a time series $X = \{x_1, x_2, ..., x_N\}$ as

$$\dot{x}(t) = \frac{x(t+\tau) - x(t-\tau)}{2\tau s} \quad (2)$$

$$\ddot{x}(t) = \frac{x(t+\tau) - 2x(t) - x(t-\tau)}{(\tau s)^2} \quad (3)$$

where τ is a number of intervals skipped between observations used to estimate the derivatives and s is the interval between successive measurements. Appropriate choice of τ will tend to minimise bias in parameter estimates (Boker and Nesselroade, in press). We used τ =17 which is long enough to remove high frequency noise while being short enough to track movements of frequencies in the range of interest (Boker, 2001). The measurement interval s = 0.02 was determined by the recording rate of 50 Hz.

The random coefficients model was then specified such that the parameters in the linear oscillator differential equation were allowed to vary across individuals. Age and frequency were then set up as fixed predictors of the differential equation model's coefficients as follows:

$$\ddot{x}_{ij} = b_{1j}\dot{x}_{ij} + b_{2j}x_{ij} + e_{1ij} \quad (4)$$

$$b_{1j} = c_0 + c_1 a_j + c_2 f_j + e_{2j} \quad (5)$$

$$b_{2j} = d_0 + d_1 a_j + d_2 f_j + e_{3j} \quad (6)$$

where x_{ij} is the ith observation of the jth infant trial, a_j is the age of the jth infant trial and f_j is the target room frequency during the jth infant trial. The frequency and age predictors were centered prior to fitting the model so that the interaction terms would not produce spurious effects. We only grouped by individual, so we are making the assumption of homogeneity across trials within individual and stimulus condition. Note that in Equation 4 there is no intercept term. This is mandated by the theoretical relationship between the derivatives and displacement that defines the linear oscillator model: at equilibrium the curvature is zero. Also note that the relationship between the derivatives is instantaneous and contains no time interval. This means that at this stage in the analysis these data are not time–ordered. While it is beyond the scope of this article, we expect this fact will lead to useful properties of differential models with respect to missing data.

We have chosen to fit the model using the *R* software, which is a free, open–source implementation of the S statistical language available for most computers and operating systems (see Hornik, 2001; Ihaka and Gentleman, 1996). In R there is a library of mixed effects models routines called *lme* (Pinheiro and Bates, 2000) which can be used to fit multilevel models as illustrated by Singer (1998). Thus, Equations 5 and 6 must be substituted into Equation 4 and multiplied out in order to create a mixed model formula appropriate for fitting with *lme* or SAS PROC MIXED:

$$\ddot{x}_{ij} = (c_0 + c_1 a_j + c_2 f_j + e_{2j})\dot{x}_{ij} + (d_0 + d_1 a_j + d_2 f_j + e_{3j})x_{ij} + e_{1ij} \quad (7)$$

$$= c_0 \dot{x}_{ij} + c_1 a_j \dot{x}_{ij} + c_2 f_j \dot{x}_{ij} + e_2 \dot{x}_{ij} + d_0 x_{ij} + d_1 a_j x_{ij} + d_2 f_j x_{ij} + e_{3j} x_{ij} + e_{1ij} \quad (8)$$

Results

Equation 8 was fit using the R procedure *lme* and restricted maximum likelihood to the 81,920 observations in the data set and the fixed effects results are presented in Table 1. The standard deviations for the random effects were $\sigma(b_1)$ = 0.00604 and $\sigma(b_2)$ = 0.00082, while their correlation was $r(b_1, b_2)$= -0.09. The dampening parameter is close to zero (c_0 = 0.00259) at the mean values of the frequency (.45 Hz) and age (8.5 months). Frequency is a significant negative predictor of dampening (c_2 = -0.02350), but age does not predict dampening. The faster frequency of room oscillation (.6 Hz) produces dampening such that infants are tending to reduce their amplitude of sway during the trial. However, the slower frequency of room oscillation (.3 Hz) is associated with infants having larger amplitudes of sway during the course of a trial.

Table 1

Fixed effects for infant AP COP postural sway fit as a multilevel linear oscillator grouped by subject (DF=81,875, AIC=-289,959, Log Likelihood=144,989).

	Value	SE	t–value	p–value
c_0	0.00259	0.00118	2.199	0.0279
c_1	0.00028	0.00037	0.758	0.4484
c_2	-0.02350	0.00277	-8.489	$< .0001$
d_0	-0.00539	0.00014	-39.836	$< .0001$
d_1	0.00013	0.00005	2.945	0.0032
d_2	-0.00155	0.00012	-13.342	$< .0001$

There is a significant negative value for the frequency (d_0 = -0.00539) at the means for frequency and age, thus significant oscillation is observed in these infants. Age has a significant positive effect on the frequency parameter (d_1 = 0.00013), so older infants tend to oscillate more slowly than younger infants. Frequency has a negative effect on the frequency parameter (d_2 = -0.00155) such that when the room oscillates at a faster frequency the frequency parameter is a larger negative number and thus the infants are swaying at a faster frequency.

Discussion

The model reproduced results found in previous work, in that significant infant sway in response to the movement of the room was found using Fourier analytic methods (Bertenthal et al., 1997). However, this analysis was much more powerful in detecting that effect than previous methods used for these data. This analysis also provided new results relating changes in amplitude over a trial to frequency of the room and relating age of the infant to changes in frequency independent of the frequency of the room.

Often, in physiological and sociological data the number of observations can be quite large. Here we modelled 81,920 observations. We ran this model both in R and in SAS PROC MIXED. SAS was able to handle a subset of 20,000 observations and provided parameter estimates that were identical to those provided by R on the same set of 20,000 observations. We were unable to successfully run SAS PROC MIXED on the full data set, while *lme* in R had no problems in running these data. Parameter estimates and fit statistics for the full data set were, of course, very close to those of the subset of 20,000 and for these data no results changed from non-significant to significant.

Conclusions

The random coefficients model can be used to estimate parameters in differential equations models of dynamical systems. We focused here on the linear oscillator model for two reasons, (a) it is an appropriate first step in modelling these data given the sinusoidal room movement and a hypothesis of linear relationships between the room movement and infant response, and (b) it is a simple system as an illustration. Many other differential equation models may prove to be useful in understanding psychological, sociological and physiological phenomena. In many, if not most, of these systems it will be necessary to take into account either individual differences or time–dependent modulations in the coefficients of these models. Multilevel modelling provides an appropriate, powerful, and relatively easy to accomplish method for estimating and predicting patterns in the coefficients of dynamical systems models.

References

Arminger, G. (1986). Linear stochastic differential equation models for panel data with unobserved variables. In N. Tuma (Ed.), *Sociological Methodology 1986* (pp. 187–212). San Francisco: Jossey Bass.

Baltes, P. B., and Nesselroade, J. R. (1979). History and rationale of longitudinal research. In J. R. Nesselroade and P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 1–39). New York: Academic Press.

Bertenthal, B. I., Rose, J. L., and Bai, D. L. (1997). Perception–action coupling in the development of the visual control of posture. *Journal of Experimental Psychology: Human Perception and Performance*, **23** (6): 1631–1643.

Boker, S. M. (2001). Differential structural modeling of intraindividual variability. In L. Collins and A. Sayer (Eds.), *New methods for the analysis of change.* Washington, DC: APA.

Boker, S. M., and Graham, J. (1998). A dynamical systems analysis of adolescent substance abuse. *Multivariate Behavioral Research*, **33** (4): 479–507.

Boker, S. M., and Nesselroade, J. R. (2000). Multilevel modelling of dynamical systems: random coefficients and order parameters. In *Proceedings of the Fifth International Conference on Social Science Methodology.* International Sociological Association.

Boker, S. M., and Nesselroade, J. R. (in press). A method for modeling the intrinsic dynamics of intraindividual variability: Recovering the parameters of simulated oscillators in multiwave panel data. *Multivariate Behavioral Research*.

Cattell, R. B. (1963). The structuring of change by P–technique and incremental R–technique. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 167–198). Madison, WI: University of Wisconsin Press.

Cleveland, W. S., and Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association,* **83:** 596–610.

Hornik, K. (2001). *The R FAQ.*

Hubbard, J. H., and West, B. H. (1991). *Differential equations: A dynamical systems approach.* New York: Springer–Verlag.

Ihaka, R., and Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistic*s, **5** (3): 299–314.

McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In J. R. Nesselroade and R. B. Cattell (Eds.), *The handbook of multivariate experimental psychology, volume 2* (pp. 561–614). New York: Plenum Press.

Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, **50:** 181–202.

Pinheiro, J. C., and Bates, D. M. (2000). *Mixed–effects models in S and S–plus.* New York: Springer–Verlag.

Sauer, T., Yorke, J., and Casdagli, M. (1991). Embedology. *Journal of Statistical Physic*s, **65** (3,4): 95–116.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models and individual growth models. *Journal of Educational and Behavioral Statistic*s, **24** (4): 323–355.

Acknowledgements

Funding for this work was provided in part by NIH Grant No. 1R29 AG14983. We would like to thank Bennett Bertenthal for the use of the example data presented here. This article was written while the first author was a visiting scientist at the Max Planck Institute for Physics of Complex Systems and its support of this work is gratefully acknowledged.

Correspondence may be addressed to Steven M. Boker, Department of Psychology, The University of Notre Dame, Notre Dame Indiana 46556, USA; email sent to sboker@nd.edu; or browsers pointed to http://www.nd.edu/~sboker/.

GLLAMM: A general class of multilevel models and a Stata program

***Sophia Rabe-Hesketh*[1], *Andrew Pickles*[2] *and Anders Skrondal*[3]**

[1]Department of Biostatistics and Computing, Institute of Psychiatry, London

[2] School of Epidemiology and Health Science, University of Manchester

[3]Department of Epidemiology, National Institute of Public Health, Oslo

Introduction

We describe a class of multilevel models, which we call generalised linear latent and mixed (GLLAMM) models, and a Stata program (*gllamm*) to fit these models by maximum likelihood. The models are generalisations of generalised linear mixed models to include many types of multilevel factor and structural equation models as well as models with discrete random effects (or latent variables). The responses can be of mixed type (including continuous, categorical and ordinal) and all of the common links and families used in generalised linear models are available. Similarly to other multilevel programs, *gllamm* treats the variables of a multivariate response as level 1 units and therefore provides maximum likelihood estimates when data are missing at random (MAR). The program uses Stata's standard modelling interface making it easy to use. We concentrate here mostly on describing the class of GLLAMM models. For more information on the Stata program, see Rabe-Hesketh et al. (2000) and the web-site at http://www.iop.kcl.ac.uk/IoP/Departments/BioComp/programs/gllamm.html from where the program and a manual can be downloaded.

Generalised linear latent and mixed models

Generalised linear mixed models assume that the conditional densities of the responses (y) given the covariates and random effects are from the exponential family with conditional expectation specified via:

$$g(E[y|x,u,z]) = \beta' x + u' z = \eta \qquad (1)$$

where g is the link function, x are covariates with fixed effects β, z are covariates with random effects u, η is the linear predictor and subscripts have been omitted to simplify notation.

In the present formulation we require that, if the data have L levels of nesting, the covariates and random effects can be specific to the units at any of the levels as long as any z varies at a lower level than its coefficient u. The random effects may be correlated within a given level of nesting but not between levels. To make the multilevel structure more specific, we will denote the M_l random effects at level l by $\boldsymbol{u}^{(l)} = \left(u_1^{(l)}, u_2^{(l)}, \ldots u_{M_l}^{(l)}\right)'$ and the corresponding covariates by $z^{(l)}$ so that $u = (u^{(2)\prime}, u^{(3)\prime}, \ldots u^{(L)\prime})'$ and $z = (z^{(2)\prime}, z^{(3)\prime}, \ldots z^{(L)\prime})'$, each with a total of $M = \sum_{l=1}^{L} M_l$ elements.

The generalised linear mixed model can then be written as

$$\eta = \boldsymbol{\beta}'\boldsymbol{x} + \sum_{l=2}^{L} \sum_{m=1}^{M_l} u_m^{(l)} z_m^{(l)} \qquad (2)$$

GLLAMMs provide four extensions to generalised linear mixed models (where we refer to u as latent variables; including random coefficients, latent traits, etc.):

1. Multilevel factor structures
2. Multilevel structural equations regressing the latent variables on other latent and observed variables
3. Discrete latent variables
4. Mixed responses

1. Multilevel factor structures

In (2) each random effect multiplies a single covariate. GLLAMMs allow each random effect to multiply a linear combination of covariates:

$$\eta = \boldsymbol{\beta}'\boldsymbol{x} + \sum_{l=2}^{L} \sum_{m=1}^{M_l} u_m^{(l)} \boldsymbol{\lambda}_m^{(l)\prime} \boldsymbol{z}_m^{(l)} \qquad (3)$$

where $\boldsymbol{z}_m^{(l)}$ is now a vector of $P_m^{(l)}$ covariates with corresponding vector of coefficients $\lambda_m^{(l)}$. For identification, the first coefficient, $\lambda_{m_1}^{(l)}$ is set to one so that the model reduces to a generalised linear mixed model if $P_m^{(l)} = 1$ for all l and m.

This extension of the generalised linear mixed model allows factor models to be incorporated in multilevel models. Here (some of) the level one units are the response variables of the factor model and the z_m are dummy variables that assign factor loadings to the appropriate responses. For example, consider a number of subjects i providing true/false responses for a set of J test items j. By regarding the items as nested in subjects, a factor model relating the probability of a correct response to a single factor, interpretable as the subjects' unobserved ability, can be defined as

$$\text{logit } E([\, y_{ij} \,|\, \boldsymbol{x}_j, u_{1i}^{(2)}, \boldsymbol{z}_{1j}^{(2)} \,])$$
$$= \boldsymbol{\beta}'\boldsymbol{x}_j + u_{1i}^{(2)} \boldsymbol{\lambda}_1^{(2)\prime} \boldsymbol{z}_{1j}^{(2)} = \beta_j + u_{1i}^{(2)} \lambda_{1j}^{(2)} \qquad (4)$$

where x_j and $\boldsymbol{z}_{1j}^{(2)}$ are vectors of length J whose jth element equals 1 and all other elements equal 0. The 'factor' $u_{1i}^{(2)}$ represents the ability of the *i*th subject, $-\beta_j$ represents the difficulty of the *j*th item and $\lambda_{1j}^{(2)}$, the factor loading, reflects how well the *j*th item discriminates between subjects of different abilities. This model is also known as a two parameter (β_j and $\lambda_{1j}^{(2)}$) logistic item response model. If the subjects are also nested in classes, further random effects can be added at level 3.

2. Multilevel structural equations regressing the latent variables on other latent and observed variables

Equation (3) relates the observed response variables to the latent variables and observed explanatory variables and is analogous to the measurement part of a structural equation model. In analogy to the structural part of a structural equation model, GLLAMMs allow latent variables to be regressed on other latent variables and explanatory variables using structural equations of the form

$$u = Bu + \Gamma w + \zeta \qquad (5)$$

where B is an $M\times M$ matrix of regression coefficients for the regressions of latent variables on other latent variables, Γ is an $M\times p$ matrix of regression coefficients for the regressions of latent variables on p observed covariates w and ζ is an M dimensional vector of residuals.

Note that the u vector includes latent variables at different levels of clustering. Since we cannot regress a higher level latent variable on a lower level one, the B matrix is restricted to be upper diagonal (the random effects at a given level must be specified in an appropriate order). This restriction implies that simultaneous effects among the latent variables (e.g. u_1 regressed on u_2 and u_2 regressed on u_1) are not permitted. The elements of the residual vector ζ correspond to the same level of nesting as the corresponding elements of u.

As an example of the use of structural equations, consider the item response model in (4). If the subjects are pupils in schools, their abilities may depend on pupil-specific variables (e.g. socioeconomic status, w_{1ij}) and school-level variables (e.g. private versus state school, w_{2i}) and there may be residual heterogeneity between schools. Using indices *i, j, k* for schools, pupils and test items, respectively, we could write the model as

$$\begin{aligned}&\text{logit}\left(E\left[y_{ijk}\,\middle|\,\boldsymbol{x}_k, u_{1ij}^{(2)}, \boldsymbol{z}_{1k}^{(2)}\right]\right)\\ &= \boldsymbol{\beta}'\boldsymbol{x}_k + u_{1ij}^{(2)}\boldsymbol{\lambda}_1^{(2)\prime}\boldsymbol{z}_{1k}^{(2)}\end{aligned} \qquad (6)$$

$$u_{1ij}^{(2)} = \Gamma_{11}w_{1ij} + \Gamma_{12}w_{2i} + u_{1i}^{(3)} + \zeta_{1ij}^{(2)} \qquad (7)$$

where $\boldsymbol{z}_{1k}^{(2)}$ are dummy variables for the items. Note that this specification of the model using two equations, one with observed responses and one with latent responses, is similar to that used by e.g. Bryk and Raudenbush (1992) and the multilevel modelling package HLM. For ordinary generalised linear mixed models, all equations can be substituted into the first and the resulting model estimated (as is done in *MLwiN*). In GLLAMMs this is generally not possible without parameter constraints. For example, substituting (7) into (6) yields a model of the form of (3) where the coefficients of $w_{1ij}z_{1rk}^{(2)}$, $w_{2ij}z_{1rk}^{(2)}$, $u_{1i}^{(3)}z_{1rk}^{(2)}$ and $\zeta_{1ij}^{(2)}z_{1rk}^{(2)}$ are constrained to be constant multiples of each other for all elements $z_{1rk}^{(2)}$ of $\boldsymbol{z}_{1k}^{(2)}$.

A more general 'two level' (here three level due to variables being considered level 1 units) factor model would allow different factor structures at the two levels and include specific factors at both levels:

$$\begin{aligned}\eta_{ijk} &= \boldsymbol{\beta}'\boldsymbol{x}_k + u_{1ij}^{(2)}\boldsymbol{\lambda}_1^{(2)\prime}\boldsymbol{z}_{1k}^{(2)} \\ &+ u_{K+1,i}^{(3)}\boldsymbol{\lambda}_{K+1}^{(3)\prime}\boldsymbol{z}_{K+1,k}^{(3)} + \sum_{m=1}^{K}u_{mi}^{(3)}z_{mk}^{(3)} \qquad (8) \\ &= \boldsymbol{\beta}'\boldsymbol{x}_k + u_{1ij}^{(2)}\lambda_{1k}^{(2)} + u_{K+1,i}^{(3)}\lambda_{K+1}^{(3)} + u_{ki}^{(3)}\end{aligned}$$

where $\boldsymbol{z}_{K+1,k}^{(3)} = \boldsymbol{z}_{1k}^{(2)}$, $z_{mk}^{(3)}$ is equal to the *m*th element of $\boldsymbol{z}_{1k}^{(2)}$ and the $u_{mi}^{(3)}, m = 1,\cdots,K+1$ are mutually independent so that $u_{mi}^{(3)}$ is the specific factor at the school level for the *m*th item.

3. Discrete latent variables

For two level models, the latent variables can have a discrete distribution with non-zero probability on a finite number of points (of dimensionality equal to the number of random effects). This is useful if the level 2 units are believed to fall into a number of groups or 'latent classes' within which the latent variables do not vary. If the number of points, or masses, is chosen to maximise the likelihood, the nonparametric maximum likelihood estimator (NPML) can be achieved (Lindsay et al., 1991).

4. Mixed responses

The conditional distribution of each response variable given the explanatory variables and random effects is specified via a family (Gaussian, binomial, Poisson or gamma) and a link function (identity, log, reciprocal, logit, probit, etc.). For polytomous responses, ordinal logit and probit or multinomial logit can be specified. Different links and families can be combined for responses of mixed type and offsets can be included. For the Gaussian and gamma distributions, the variance parameter can be allowed to differ between groups of observations or to depend on explanatory variables.

Implementation

gllamm uses Stata's maximum likelihood functions to maximise the marginal likelihood using a modified Newton Raphson algorithm based on numerical first and second derivatives of the log marginal likelihood. For multivariate normally distributed random effects, quadrature is used to approximate the marginal likelihood. Here, correlated random effects are represented by a linear combination of independent random effects via the Cholesky decomposition so that the multivariate integrals become nested univariate integrals.

Transformations are used to ensure that parameter estimates are valid where applicable. For example, the Cholesky decomposition of the variance covariance matrix of the random effects is estimated to ensure positive semi-definiteness of the covariance matrix. Standard errors for the back-transformed parameters are obtained using the delta method. The program allows linear constraints to be specified for the transformed versions of the parameters.

Applications

Here we describe a number of applications of GLLAMM models referring to published papers for more details. (The examples do not exhaust the list of possible model types.) The available links and families allow many different response processes to be modelled including first choice and ranking data using multinomial logit, multivariate survival data and multilevel and multivariate discrete time to event data using either proportional hazards or proportional odds models and their non-proportional generalisations (Rabe-Hesketh et al., 2001).

Models with the same random part as the item response model in (4) can be used to structure the covariance matrix of multivariate random effects models for categorical responses thereby reducing the dimensionality of integration (Rabe-Hesketh and Skrondal, 2001). The model can also be used to fit errors in variables models, for example logistic regression with errors in a continuous explanatory variable, by treating both the dichotomous response and the measurements of the explanatory variable as dependent variables stacked into the response vector y and specifying a mixed response model with different links and families for different responses (Rabe-Hesketh and Pickles, 1999). The normality assumption for the true explanatory variable can be relaxed by using a discrete latent variable and increasing the number of masses until no further increase in the maximised likelihood can be achieved, giving the NPML solution.

If a set of binary variables is available at two time points, and at each time subjects fall into a number of groups (e.g. disease present or absent), a latent transition model can be set up using two discrete random effects, one measured by the time 1 variables and the other by the time 2 variables (Rabe-Hesketh and Pickles, 1999).

With responses over several time-points growth curve models can be estimated, the discrete random effects formulation being interpretable as alternative latent developmental trajectories (Maughan et al., 2000). Multicentre test-retest data have been modelled, allowing both the measurement error variance and the true score variance to differ between centres (Leese et al., 2001).

Discussion

Generalised linear latent and mixed models are extremely flexible. The drawback of this flexibility is that programming derivatives of the log marginal likelihood for the general case is difficult. Our implementation therefore uses numerical derivatives which take a long time to compute, particularly if there are a large number of latent variables, a large number of quadrature points are used, and the sample is large. The numerical derivatives are computed by Stata's excellent maximum likelihood functions and numerical precision does not appear to be a problem. One way we hope to speed up the program is by converting some of the code to internal code with the help of Stata Corporation. A great advantage of not having to program the derivatives is that new features can be added to the program extremely easily. New models can therefore easily be developed and tested, and if useful, more efficient programs written to implement them.

A definite advantage of this maximum likelihood method is that likelihood ratio tests can be performed for the random parameters. The quadrature approximation appears to work better than MQL/PQL for binary, ordinal, or polytomous responses when the level 2 clusters are small (Rabe-Hesketh et al., 2001). However, for large level 2 clusters or Poisson distributed responses, where PQL tend to works well, the quadrature approximation can be inaccurate (Albert and Follmann, 2000). The same problem can occur with normally distributed responses for which other software would generally be recommended (e.g. *Mplus* by Muthén and Muthén, 1998, can be used for multilevel structural equation modelling of continuous data). Fortunately, it is easy to diagnose problems with the quadrature approximation by assessing the change in parameter estimates associated with an increase in the number of quadrature points.

References

Albert, P. S., and Follmann, D. A. (2000). Modeling repeated count data subject to informative dropout. *Biometrics*, **56**: 667-677.

Bryk, A. S., and Raudenbush, S. W. (1992). *Hierarchical Linear Models*. Newbury Park, CA: Sage.

Leese, M. N., White, I. R., Schene, A. H., Koeter, M. W., Ruggeri, J., Gaite, M. L., and the EPSILON Study Group (2001). Reliability in multi-site psychiatric studies. *International Journal of Methods in Psychiatric Research*, **10**: In Press.

Lindsay, B. G., Clogg, C. C., and Grego, J. (1991). Semiparametric estimation in the Rasch model and related exponential models, including a simple latent class model for item analysis. *Journal of the American Statistical Association*, **86**: 96-107.

Maughan, B., Pickles, A., Rowe, A., Costello, R., and Angold, A. (2000). Developmental trajectories of aggressive and non-aggressive conduct problems. *Journal of Quantitative Criminology*, **16**: 199-221.

Muthén, L. K., and Muthén, B. O. (1998). *Mplus User's Guide*. Los Angeles: Muthén and Muthén.

Rabe-Hesketh, S., and Pickles, A. (1999). Generalised linear latent and mixed models. In *14th International Workshop on Statistical Modeling*, pp 332-339, Graz, Austria.

Rabe-Hesketh, S., Pickles, A., and Taylor, C. (2000). SG129: Generalized linear latent and mixed models. *Stata Technical Bulletin*, **53**: 47-57.

Rabe-Hesketh, S., and Skrondal, A. (2001). Parameterisation of multivariate random effects models for categorical data. *Biometrics*. In Press.

Rabe-Hesketh, S., Touloupulou, T., and Murray, R.M. (2001). Multilevel modeling of cognitive function in schizophrenics and their first degree relatives. *Multivariate Behavioral Research*. In Press.

Rabe-Hesketh, S., Yang, S., and Pickles, A. (2001). Multilevel models for censored and latent responses. *Statistical Methods in Medical Research*. In Press.

Review of 'Generalized, Linear, and Mixed Models'. Charles E. McCullogh and Shayle R. Searle. New York, Wiley. ISBN 0-471-19364-X. £64.50 (UK); $89.95 (US)

Harvey Goldstein
Institute of Education

This book is aimed at graduates and practising statisticians. Chapter 1 introduces the distinction between crossed and nested designs, interactions and the distinction between random effects and fixed categories or values of a variable. It has a useful discussion of when effects are to be considered random as opposed to fixed and looks briefly at various, non-Bayesian, estimation methods. It is clearly written and a good introduction to the topic.

Chapter 2 deals with the Normal response linear variance components model and derives ML estimators for the balanced and unbalanced cases together with confidence intervals and hypothesis tests. It then looks at the beta binomial model and the logit-Normal model, for the latter mentioning quadrature estimation. Finally there is mention of the probit model. The chapter gives several detailed descriptions of estimators with associated algebra. Much of this seems unnecessary, especially separate derivations for balanced and unbalanced data. It is difficult to see who would need to wade thorough all of this when a few lines can indicate the general results.

In Chapter 2 the model was assumed to consist of a single classification with the group means either fixed or random effects. Chapter 3 extends the models of Chapter 2 by adding a single covariate to the model, again dealing separately with balanced and unbalanced data.

Chapter 4 is nothing more than the classical general linear (regression) model as set out in myriad basic statistical texts and Chapter 5 is just the standard exposition for the generalised linear model.

Chapter 6 introduces the general linear mixed model $Y = XB + ZU$. Incidentally, throughout the book the models are written in terms of y_{ij} so that for Normal models the lowest level residual terms are never explicitly recognised, which may cause some confusion to those familiar with multiple regression. Shrinkage estimators of random effects are discussed and the likelihood equations for the model with simple level 1 variance are presented, but without discussion of estimation in the general case. Again the balanced case is treated separately from the general case.

Chapter 7 looks at longitudinal repeated measures data. It deals mainly with the balanced case, remarking that the general case is 'much more difficult to deal with'. This chapter makes very heavy weather of what is essentially just a version of the 2-level structure which was dealt with in previous chapters.

Chapter 8 starts off with a useful discussion of marginal and conditional (unit specific) generalised linear models and sets out the likelihood equations. Quasilikelihood estimation (PQL1 and MQL1) is discussed and somewhat summarily dismissed, but unfortunately, apart from an oblique reference in Chapter 10, the authors seem unaware of much of the literature (e.g. Goldstein and Rasbash, 1996) that provides evidence for where such approaches and extensions are adequate.

Chapter 9 deals with prediction, especially of random effects and their properties.

Chapter 10 has a brief but useful discussion of some algorithms, including quadrature, EM, MCMC, stochastic approximation, simulated maximum likelihood and quasilikelihood.

Chapter 11. This very briefly introduces non-linear models.

In summary, it is difficult to recommend this book, especially with this price tag. While there are some useful parts, it is very limited in scope and unnecessarily repetitive. It ignores almost all the relevant multilevel literature, does not even mention multivariate models, and there is no illustration of MCMC methods nor a discussion of Bayesian models. One might say that it fills a much needed gap in the literature!

Reference

Goldstein, H., and Rasbash, J. (1996). Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society, A,* **159**: 505-13.

Some Recent Publications Using Multilevel Models

Boyle, M. H., and Willms, J.D. (2001). Multilevel modelling of hierarchical data in developmental studies. *Journal of Child Psychology and Psychiatry*, **42**: 141-162.

Browne, W. J., and Draper, D. (2000). Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models. *Computational Statistics*, **15**: 391-420.

Cunningham, S. J., Gilthorpe, M. S., and Hunt, N. P. (2000). Are orthognathic patients different? *European Journal of Orthodontics*, **22**: 195-202.

Gilthorpe, M. S., and Cunningham, S. J. (2000). The application of multilevel, multivariate modelling to orthodontic research data. *Community Dental Health*, **17**: 236-242.

Gilthorpe, M. S., Griffiths, G. S., Maddick, I. H., and Zamzuri, A. T. (2000). The application of multilevel models to periodontal research data. *Community Dental Health*, **17**: 227-235.

Gilthorpe, M. S., Maddick, I. H., and Petrie, A. (2000). Introduction to multilevel modelling in dental research. *Community Dental Health*, **17**: 222-226.

Ireson, J., Hallam, S., and Plewis, I. (2001). Ability grouping in secondary schools: Effects on pupils' self-concepts. *British Journal of Educational Psychology*, **71:** 315-326.

Merlo, J., Östergren, P.-O., Broms, K., Bjorck-Linné, A., and Liedholm, H. (2001). Survival after initial hospitalisation for heart failure: a multilevel analysis of patients in Swedish acute care hospitals. *Journal of Epidemiology and Community Health*, **55:** 323-329.

Snijders, T. A. B. (2001). Two-level non-parametric scaling for dichotomous data in Boomsma, A., van Duijn, M. A. J., and Snijders, T. A. B. (eds.), Essays on Item Response Theory. *Lecture Notes in Statistics*, **157**: 319-338. New York: Springer.

Subramanian, S. V., Duncan, C., and Jones, K. (2000). 'Illiterate people' and 'illiterate places': the Indian evidence. *Indian Social Science Review*, **2** (2): 237-274.

Subramanian, S. V., Duncan, C., and Jones, K. (2001). Multilevel perspectives on modelling census data. *Environment and Planning A*, **33** (3): 399-417.

Subramanian, S. V., Kawachi, I., and Kennedy, B. P. (2001). Does the state you live in make a difference? Multilevel analysis of self-rated health in the US. *Social Science and Medicine*, **53** (1): 9-19.

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