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. 14 No. 1 **July, 2002**

Forthcoming Workshops

11-13 September 2002, 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 0AL, United Kingdom. Tel: +44 (0) 20 7612 6688, Fax: +44 (0) 20 7612 6572, email: a.burch@ioe.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.

Royal Statistical Society Conference, Plymouth, 3-6 September 2002

The following four papers will be presented by members of the Centre for Multilevel Modelling, Institute of Education. For further details about the conference, go to: http://www.tech.plym.ac.uk/maths/research/stats/rssprogramme.html

1. **Multiple membership models for complex multilevel data structures**
Harvey Goldstein and William Browne
h.goldstein@ioe.ac.uk

In the standard multilevel model lower level units are uniquely classified within one higher level unit, for example students within schools or people within households. In many kinds of data, however, individuals will 'belong' to more than one such unit.

Thus, in a longitudinal study students may move between schools so that the school 'effect' must be considered as shared among the schools attended. Likewise, over time, individuals may experience more than one household.

In order to model such data multiple membership models have been developed. The talk will describe such models, introducing a new notation and examples of applications. An application to the analysis of poultry salmonella outbreaks will be presented. The lowest level unit is a poultry flock whose members are derived from several parent flocks and there is also a cross classification of the parent flocks with the poultry farms. The modelling approach provides estimates of the contributions of each parent flock to the infection probability.

The methodology can also be used for fitting spatial models. An interesting application arises in education and other areas where measurements are available for individuals, for example achievement scores, and also for groups of individuals where there is a single group response. Under suitable conditions, the joint analysis of individual and group responses allows the estimation of the effective contribution each individual makes towards the group response together with the relationship between that and the individual's separate response.

2. **An MCMC algorithm for problems involving 'constrained' variance matrices with applications in multilevel modelling**

William Browne
w.browne@ioe.ac.uk

General-purpose Bayesian software packages, for example WinBUGS, that utilize MCMC methods are now being used widely by quantitative researchers. To make such software as flexible as possible, MCMC methods that can be adapted to fit the widest range of statistical models have been preferred. Originally, Gibbs sampling algorithms primarily through the AR sampler were used to fit models using univariate updates with the restriction that all conditional posteriors be log concave. More recently this restriction has been removed by using adaptive (random-walk) Metropolis samplers for parameters without log concave distributions. These samplers are used (where necessary) in both the WinBUGS and *MLwiN* software packages.

In this talk, we discuss another feature of certain statistical models - 'constrained' variance matrices - which cannot currently be dealt with in general purpose packages. By 'constrained' we mean that the variance matrix is subject to some additional constraints (as well as the positive definite constraint). For example, two elements of the matrix could be constrained to be equal, or an element could equal a constant or be a function of predictor variables.

3. **Exploring differential parental treatment of siblings using multilevel models**

Jon Rasbash with Thomas G. O'Connor, Institute of Psychiatry, University of London and Jenny Jenkins, University of Toronto
j.rasbash@ioe.ac.uk

In family studies a question of great interest is the extent to which children in the same family get treated differently by their parents. There is much professional theorising and lay speculation about the existence and effect of differential parental treatment. In this paper we use multilevel modelling techniques to explore the nature and extent of differential parenting. Data from 8,476 children, between 4 and 11 years old, in 3,762 families were analysed. The data came from a large-scale population study in Canada. We currently have data for the first wave in the longitudinal study. We only have one time point, but do have data on multiple children per family. This means that our models can describe the extent to which differential parental treatment exists, but they cannot address questions about the subsequent effects of differential parenting. Multilevel modelling provides a natural and powerful framework for handling family data. In the study parents rate their relationship with each of their children using a number of items. These items are then combined into a single score, which serves as the response variable in our models. In these models the level one variance (between child within family) gives us a direct measure of differential parenting. The paper shows how multilevel modelling can provide some important descriptions of differential parenting by modelling the level one variance as a complex function of predictors.

4. **A Multilevel Multistate Competing Risks Model for Event History Data**

Fiona Steele and Harvey Goldstein
f.steele@ioe.ac.uk

Event history data are collected in many social surveys, providing a longitudinal record of events such as changes in partnerships and employment. Most of these events can be experienced more than once over an individual's lifetime, and durations between events may be correlated. Multilevel event history models have been developed to analyse repeated events. Another common feature of event history data is that there are sometimes several 'competing' events that may occur, or an event can be experienced for one of several reasons. For example, a cohabiting partnership may end in marriage or separation. Multilevel event history models have been developed to handle such competing risks. Another extension to the basic repeated events model is the multistate model which has been developed for situations where an individual may be in one of several possible states at a given point in time, for example in or out of a partnership. In this paper, we propose a general model for the analysis of repeated events which allows for both competing risks and multiple states, where the competing risks may be state-dependent. The multilevel multistate competing risks model is applied in an analysis of transitions in and out of contraceptive use in Indonesia, using retrospective contraceptive histories collected in a national survey.

Modelling ordinal data using *MLwiN*

Ian Plewis
Institute of Education, University of London.
i.plewis@ioe.ac.uk

Introduction

The literature contains a number of ways of representing the ordering in an ordered categorical response. Perhaps the most popular is to use cumulative logits in a proportional odds model. For a variable with *M* categories, the cumulative logits are:

$$\log \sum_{c=1}^{m} \pi_c / (1 - \sum_{c=1}^{m} \pi_c)\text{; } m = 1..M\text{-}1\text{; } \pi_c \text{ is the probability in category } c,$$

and for a two level model (for example, occasions i within subjects j), we might write:

$$\log \sum_{c=1}^{m} \pi_{cij} / (1 - \sum_{c=1}^{m} \pi_{cij}) = \beta_{moj} + \sum_{h=1}^{2} \beta_{hj} a_{ij}^{h}\text{; } i = 1..t_j\text{; } j = 1..n. \qquad (1)$$

Here, the right-hand side of the model is a quadratic in age with the variation in the coefficients across subjects assumed to be multivariate Normal. The MULTICAT macro in *MLwiN* (Yang, 1997) is available to estimate this model, using quasi-likelihood methods. See Fielding (1999) and Ribaudo et al. (1999) for applications.

An alternative representation of the ordering is to use log continuation odds (or continuation ratios), defined as:

$$\log \sum_{c>m} \pi_{cij} / \pi_{mij} \text{ - } m = 1..M\text{-}1$$

These *M*-1 odds are asymptotically independent binary functions (Fienberg, 1980) and can therefore be treated as a set of multivariate binary responses. It is relatively straightforward to set up this model in *MLwiN*, essentially by constructing a series of cumulative responses and treating all the odds for $m^{/} < m$, $m > 1$ as missing. As each successive log odds is based on a smaller and smaller part of an assumed underlying continuous distribution, the number of missing responses increases as m increases. The model can be set up using the multivariate window in *MLwiN* with the level one covariance matrix set to an identity matrix of order *M*-1, assuming Bernoulli variation. It is possible to allow the model coefficients to vary with m, i.e. to have β_{mhj}, $h \geq 0$ on the right-hand side of the model. (Note that *MLwiN* uses subscripts j and k rather than i and j because level one (i) is used to define the multivariate structure of the *M*-1 responses.)

The purpose of this article is not to argue for the superiority of one representation of ordering over another but to show that there is more than one way of specifying multilevel models for ordinal data. The choice between the different specifications should be based on the nature of the response and on goodness-of-fit criteria.

An application

Ten Have and Uttal (1994) published data from a study of young children's search strategies. Children were randomly assigned to two conditions and observed for ten trials so i = 1..10 and n = 1..89 in equation (1). The response was the number of attempts up to three a child needed to find a hidden toy given a map that was either rotated (treatment = 1) or non-rotated (treatment = 0), so the response was an ordered 4-category variable (immediately successful through to failure, the latter being common). There were no missing data. Ten Have and Uttal used a continuation odds approach within a Bayesian (MCMC) framework but quasi-likelihood is used for most of the analyses presented here.

To be consistent with Ten Have and Uttal, the following effects are included in the initial model: linear and quadratic terms for trial (centred at zero), both varying randomly across children, and a treatment effect that interacts with these linear and quadratic terms. The main effect of treatment is allowed to be different for each continuation odds.

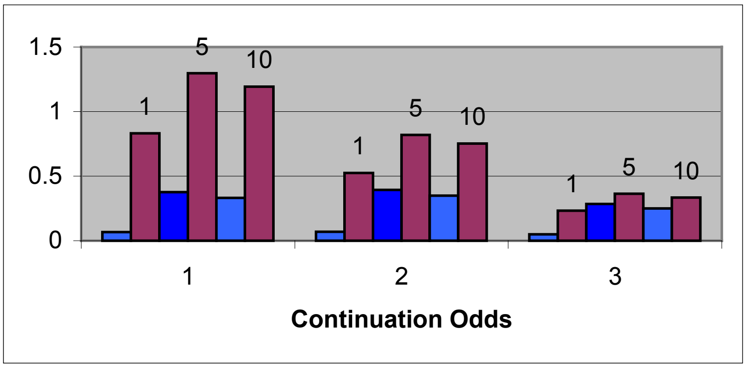
Four models were fitted using second-order penalized quasi-likelihood (PQL(2)): two continuation odds models with different assumptions about the random effects, a proportional odds model and a binary model for which the response was just success or failure at each trial. The fixed effects are shown in Table 1.

Table 1: Fixed effects and standard errors for four models

	Cont. Odds (1)	Cont. Odds (2)	Prop. Odds	Binary
Cons.1	-0.95 (0.22)	-0.98 (0.19)	-0.85 (0.20)	0.50 (0.25)
Cons.2	-0.85 (0.23)	-0.93 (0.21)	0.027 (0.19)	-
Cons.3	-0.91 (0.26)	-1.26 (0.24)	0.46 (0.20)	-
Trt.,cons1	1.28 (0.30)	1.24 (0.26)	1.17 (0.27)	1.17 (0.37)
Trt.,cons2	0.96 (0.32)	0.73 (0.30)	"	-
Trt.,cons3	0.37 (0.38)	0.25 (0.36)	"	-
Trial, linear	0.25 (0.040)	0.18 (0.032)	0.19 (0.035)	0.24 (0.051)
Trial, quad.	-0.069 (0.015)	-0.046 (0.012)	-0.053 (0.013)	-0.061 (0.019)
Trt.*linear	-0.18 (0.049)	-0.14 (0.042)	-0.14 (0.046)	-0.21 (0.074)
Trt.*quad.	0.049 (0.019)	0.033 (0.016)	0.034 (0.018)	0.067 (0.028)

'Cons. m' (m = 1, 2, 3) are the three cut-points; the second and third of these have different interpretations for the continuation odds and proportional odds models. The MULTICAT macro makes the usual proportionality assumption so that the fixed effects are constrained to the same for each cut-point. Following Ten Have and Uttal, the trial main effects and the treatment*trial interactions are the same for each cut-point in the continuation odds model. However, whereas the proportional odds model assumes a constant treatment main effect, the continuation odds models allow it to vary with cut-point and the estimates suggest that the effect is stronger in terms of influencing early success. This is illustrated in Figure 1 which is based on the estimates for the second continuation odds model in Table 1. Otherwise, the estimates are fairly consistent across models (including the binary model) and tell a similar story to Ten Have and Uttal although direct comparisons with the published estimates are difficult as they use a different parameterisation.

Figure 1: Treatment and trial effects



Notes

The numbers above the bars are trial numbers. The odds for the non-rotated group are coloured blue, the rotated group brown. The treatment differences become smaller across the continuation odds. The odds initially rise with trial but later start to decline.

We turn now to the random effects. In the first continuation odds model, a full covariance matrix at level two is estimated; in other words, a model with nine variances (three each for the cut-points and for the linear and quadratic trial terms) and hence 36 covariances. The algorithm converges but, not surprisingly with just 89 subjects, a number of the parameters have zero estimates and one of the correlations is less than minus 1. The level two covariance matrix for the second continuation odds model in Table 1 is the same as the one used by Ten Have and Uttal - single variances for the intercept, linear and quadratic terms and the three associated covariances. The variance for the quadratic term, and its associated covariances, are estimated to be zero. Table 2 gives the results for this model, and also for the proportional odds model (for which only the constant term has a positive variance) and for the model with a binary response. Those random effects for which it was possible to get a non-zero estimate are all, apart from the intercept, small. The posterior distribution for the variance component for the quadratic trial effect obtained by Ten Have and Uttal is not, however, massed near zero although the linear effect is. For the binary model, estimates from running a MCMC model with 100K iterations, a thinning rate of 10% and non-informative Wishart priors for the random effects produced similar estimates to those obtained by PQL(2).

Table 2: Random effects, standard errors and *correlations*

	Cont. Odds (2)	Prop. Odds	Binary
Intercept	0.65 (0.13)	0.80 (0.17)	1.42 (0.44)
Trial, linear	0.0021 (0.0081)	0	0.028 (0.018)
Trial, quadratic	0	0	0.0030 (0.0024)
Cov (int., linear)	0.017 (0.015) *(0.44)*	0	0.082 (0.063) *(0.42)*
Cov. (int., quad.)	0	0	-0.0035 (0.025) *(-0.05)*
Cov. (lin., quad.)	0	0	-0.0052 (0.0046) *(-0.57)*

Discussion

Using continuation odds for ordered responses can sometimes be more natural than using proportional odds, for example when subjects can only move up (or down) a scale over time - educational qualifications is one example. There is also an advantage within the current version of *MLwiN* of not having to use a macro to estimate the model. On the other hand, the assumption of proportionality leads to a more parsimonious model, although, for the data used here, perhaps a less realistic one.

One of the issues to be considered when estimating a continuation odds model is how much complexity to put into the random part of the model. Although the odds are asymptotically independent, it could be more valid to allow the Bernoulli variates at level one to covary when there are only a small number of level-one units per level-two unit. With repeated measures data, it is unusual to have more than the 10 occasions (trials) analysed here. The level-one correlations were small with these data but their introduction into the model led to some level-two correlations becoming greater than one. It is worth noting that, for the full covariance matrix at level two, some of the variances and covariances associated with the quadratic trial term were non-zero but for the less complex model, this was not so. Further exploration of these issues within a MCMC framework could be fruitful. However, the model's flexibility in terms of allowing higher level variances to vary with the cut-points is likely to be useful only with much larger datasets than the one used here.

Rather than a logit link, a complementary log-log link might be more appropriate especially as this strengthens the connection with hazard models. However, its application led to zero estimates for the variances of the trial terms with these data. Another extension that might be particularly appropriate with repeated measures data of the kind analysed here, with little time elapsing between trials, would be to introduce some kind of time series structure at level one (Barbosa and Goldstein, 2000), to allow for the possibility that the residuals from successive trials could be correlated.

References

Barbosa, M., and Goldstein, H. (2000). Discrete response multilevel models for repeated measures: An application to voting intentions data. *Quality and Quantity*, **34**: 323-330.

Fielding, A. (1999). Why use arbitrary points scores?: ordered categories in models of educational progress. *Journal of Royal Statistical Society, A,* **162**: 303-328.

Fienberg, S. E. (1980). *The Analysis of Cross-Classified Categorical Data (2nd. Ed.)*. Cambridge, Mass.: MIT Press.

Ribaudo, H. J., Bacchi, M., Bernhard, J., and Thompson, S. G. (1999). A multilevel analysis of longitudinal ordinal data: evaluation of the level of physical performance of women receiving adjuvant therapy for breast cancer. *Journal of Royal Statistical Society, A,* **162**: 349-360.

Ten Have, T. R. and Uttal, D. H. (1994). Subject-specific and population-averaged continuation ratio logit models for multiple discrete time survival profiles. *Applied Statistics*, **43**: 371-384.

Yang, M. (1997). Multilevel models for multiple category responses by *MLn* simulation. *Multilevel Modelling Newsletter*, **9** (1): 9-15.

Fitting multilevel models under informative probability sampling

Danny Pfeffermann[1], Fernando Moura[2] and Pedro Nascimento Silva[3]

[1]Hebrew University, Israel and University of Southampton, UK

[2]Federal University of Rio de Janeiro, Brazil

[3]IBGE, Brazil

msdanny@socsci.soton.ac.uk

Introduction

Classical theory underlying the use of multilevel models assumes implicitly either that all the population groups (clusters) at all levels are represented in the sample or that they are selected by simple random sampling. This assumption may not hold in a typical sample survey where the clusters or the final sampling units or both are often selected with unequal selection probabilities. When the selection probabilities are related to the values of the response variable even after conditioning on the model covariates, the sampling process becomes *informative* and the model holding for the sample data is then different from the population model. Ignoring the sampling process in such cases may yield biased estimators and distort the analysis.

As an example, consider a study of pupils' proficiencies with schools as the second level units and pupils as first level units, and suppose that the schools are selected with probabilities proportional to their sizes. If the size of the school is related to the school's achievements, say the large schools are mostly in areas with low achievements, and the size of the school is not included among the model covariates, the sample of schools will tend to contain large schools with low achievements, and hence no longer represent the population of schools.

As suggested by the example, a possible way to deal with the problem of informative sampling is to include among the model covariates all the design variables that define the selection probabilities at the various levels. This approach is often not practical, however.

In a recent article, Pfeffermann et al. (1998a) propose *probability weighting* of first and second level units to adjust for the effect of informative sampling on the estimation of the multilevel model parameters. The authors also develop appropriate variance estimators. The use of this procedure is justified based on asymptotic arguments but it is shown to perform well in a simulation study even with moderate sample sizes. Nonetheless, the use of the sampling weights (inverse of sample inclusion probabilities) for bias correction has four important limitations:

1. The variances of the weighted estimators are generally larger than the variances of the corresponding unweighted estimators.

2. Inference is restricted primarily to point estimation. Probabilistic statements require asymptotic normality assumptions. The exact distribution of weighted point estimators is generally unknown.

3. The use of the randomisation distribution for inference (the distribution over all possible samples from the target population) does not permit in general to condition on the selected sample of clusters (second and higher level units) or values of the model explanatory variables.

4. It is not clear how to predict second and higher level random effects under informative sampling with this approach; for example, how to predict the mean school achievement for schools not represented in the sample.

The purpose of this article is to describe a model dependent approach for multilevel modelling under informative sampling, as proposed by Pfeffermann et al (2001). The idea behind the approach is to extract the hierarchical model holding for the sample data as a function of the population model and the first order sample inclusion probabilities, and then fit the *sample model* using classical techniques. An important implication of the use of this approach is that the selection probabilities feature in the analysis as additional outcome values that are used to strengthen the estimators. Evidently, if the sample model is specified correctly, the use of this approach overcomes the limitations mentioned with respect to the probability weighting approach. We restrict ourselves to a two level model and apply the full Bayesian paradigm by use of Markov Chain Monte Carlo (MCMC) integration, but the approach can be extended to higher level models and different inference procedures.

Population model, sampling design and sample model

Consider the following two-level hierarchical model:

First level: $$y_{ij} \mid \beta_{0j} = \beta_{0j} + x_{ij}'\beta + \varepsilon_{ij} \quad ; \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \ , \quad i = 1...M_j \qquad (1)$$

Second level: $$\beta_{0j} = z_j'\gamma + u_j \quad ; \quad u_j \sim N(0, \sigma_u^2) \ , \quad j = 1...N \qquad (2)$$

This model is often referred to in the literature as the *random intercept regression model* and it contains as unknown parameters the vectors of coefficients (β, γ) and the first and second level variances $(\sigma_\varepsilon^2, \sigma_u^2)$. Note that the intercepts are modelled as linear functions of known regressor values z_j. In the simulation experiment described in Section 4 the outcome y_{ij} is the test score of pupil i in school j, x_{ij} defines the sex, age and parents' education of the same pupil and z_j consists of two dummy variables defining geographical regions. The second level random effects u_j account for the variation of the intercept terms not explained by the variables z_j.

We assume a two stage sampling process. In the first stage $n < N$ second level units are selected with probabilities $\pi_j = \Pr(j \in s)$ that could be correlated with the random effects u_j. In the second stage m_j first level units are sampled from second level unit j selected in the first stage, with probabilities $\pi_{i|j} = \Pr(i \in s_j \mid j \in s)$ that may be correlated with the residuals ε_{ij}. In Section 3 we elaborate on the sampling process used for the simulation study.

Following Pfeffermann et al. (1998), the sample distributions of the first and second level units are respectively,

$$f_s(y_{ij} \mid x_{ij},\theta_j) = f(y_{ij} \mid x_{ij},\theta_j, i \in s_j) = \frac{E_p(\pi_{i|j} \mid y_{ij}, x_{ij}, \theta_j) f_p(y_{ij} \mid x_{ij},\theta_j)}{E_p(\pi_{i|j} \mid x_{ij},\theta_j)} \quad (3)$$

$$f_s(\beta_{0j} \mid z_j,\lambda) = f(\beta_{0j} \mid z_j,\lambda, j \in s) = \frac{E_p(\pi_j \mid \beta_{0j}, z_j, \lambda) f_p(\beta_{0j} \mid z_j,\lambda)}{E_p(\pi_j \mid z_j,\lambda)} \quad (4)$$

where s_j defines the first level sample from second level unit j, $\theta_j = (\beta_{0j}, \beta', \sigma_\varepsilon^2)$, $\lambda = (\gamma', \sigma_u^2)$ and $f_p(\cdot)$ and $f_s(\cdot)$ are the population and sample distributions with expectations $E_p(\cdot)$ and $E_s(\cdot)$.

Equations (3) and (4) define another two level model, holding for the sample data. This model depends on the population model and the first order sample selection probabilities of first and second level units. The expectations in (3) and (4) can be modelled based on knowledge of the sampling process and the sample data. See Pfeffermann and Sverchkov (1999) for discussion and examples and also the section below.

Monte Carlo simulation experiment

The purpose of the simulation experiment is to study the performance of the model dependent approach introduced in the previous section and compare it with the probability weighting procedure described in Pfeffermann et al. (1998a), with both approaches compared to the common practice of ignoring the sampling process in the estimation process. The model and sampling design underlying this experiment follow the Basic Education Evaluation study carried out in 1996 in Rio de Janeiro, Brazil. The outcome data in that study are the proficiency scores of 14,831 pupils in 392 schools located in three different regions. In what follows we use schools to define the second level units and pupils to define the first level units. The experiment consists of generating 400 populations from the model defined by (1) and (2) and selecting four samples from each population using four different sampling designs, defined by combining each of two methods for the selection of schools with two different methods for the sampling of pupils within the selected schools. Schools were selected using either Method A1 - simple random sampling without replacement (SRSWR) or Method A2 - probability proportional to the school size (PPS) using the Sampford method, with the school size generated as,

$$\log(M_j) \sim N(\alpha' z_j + \alpha_3 \beta_{0j}, \sigma_M^2). \quad (5)$$

Note that Method A2 is informative since the school sizes depend on the intercepts β_{0j}. Students within the selected schools were sampled by either Method B1- SRSWR or Method B2- disproportionate stratified sampling with the strata defined by the values of propensity scores,

$$p_{ij} = b_0 + b_1 y_{ij} + \zeta_{ij}, \; \zeta_{ij} \sim N(0, \sigma_\zeta^2). \quad (6)$$

Method B2 is again informative since the propensity scores depend on the proficiency scores y_{ij}. One sample of 40 schools and 10 pupils from each selected school were drawn from each population using each of the 4 sampling schemes. For the stratified sample selection (Method B2) we sampled four pupils from strata one and two and two pupils from stratum three.

The sample models for general two-stage sampling schemes are defined by (3) and (4). The expectations defining these models under the informative sampling schemes A2 and B2 considered in the present simulation are given in equations (8), (9), (10) and (11) of Pfeffermann et al. (2001).

The population model defined by (1) and (2) has a Bayesian hierarchical structure with β, γ, σ_{ε}^2 and σ_u^2 as hyper-parameters. The MCMC algorithm applied for the present experiment (see Section 1) consists of sampling alternately from the conditional posterior distribution of each of the unknown parameters, given the data and the remaining quantities. Our proposed adjustment for the effect of informative sampling consists of replacing the population densities $f_p(y_{ij} \mid x_{ij}, \theta_j)$ and $f_p(\beta_{0j} \mid z_j, \lambda)$ defined by (1) and (2) by their sample counterparts defined by (3) and (4). In the simulation study we assigned non-informative priors to the hyper-parameters as described in Pfeffermann et al. (2001). The MCMC computations have been implemented using version 1.3 of the WinBUGS program (Spiegelhalter et al. 2000), generating 5000 values from each posterior distribution after discarding the first 5000 values as 'burn in'. The conditional posterior distributions of the various parameters given the data and the remaining parameter values, needed for application of the MCMC simulation are defined in Pfeffermann et al. (2001).

Simulation results

The full set of simulation results is presented in Pfeffermann et al. (2001). Here we present only a subset of these results, focusing on the estimation bias. The simulation results are based on 400 replications, where each replication consists of generating a new population and selecting a single sample of 40 schools and 10 pupils from each school by each of the four sampling methods described above.

We start by showing the results obtained when ignoring the sample selection and fitting the population model. These results serve as benchmarks for assessing the performance of the alternative approaches of probability weighting and the use of the sample model to deal with the effects of informative sampling. Table 1 shows the true hyper-parameter values, the percent absolute bias (PAB) and the *p*-values (P-V) of the conventional *t*-tests of bias as obtained under the four sampling methods. Note that the parameter estimates from each sample are the empirical means of the 5000 values drawn from the corresponding posterior distribution after discarding the first 5000 values as 'burn in'.

Table 1. Percent Absolute Bias (PAB) and P-Values (P-V) of tests of bias when ignoring the sampling process

	Selection of Schools							
	Non Informative, Method A1				Informative, Method A2			
Selection of Students	Non Informative Method B1		Informative Method B2		Non Informative Method B1		Informative Method B2	
Parameter	PAB	P-V	PAB	P-V	PAB	P-V	PAB	P-V
γ_0= 86.9	0.0	97.3	3.2	0.0	7.0	0.0	1.0	0.5
γ_1= -6.8	4.3	40.5	18.2	0.0	1.8	71.4	25.9	0.0
γ_2= -13.8	0.6	84.0	25.9	0.0	4.1	13.6	29.3	0.0
β_1= -10.9	0.4	81.4	6.2	0.0	1.2	40.0	4.3	0.5
β_2= -16.0	0.6	58.6	6.9	0.0	0.2	83.1	5.1	0.0
β_3= -36.5	0.1	84.4	4.9	0.0	0.4	57.8	3.9	0.0
β_4= -7.2	3.7	11.5	2.2	36.6	1.0	65.6	2.9	21.2
σ_u^2	23.6	0.0	50.7	0.0	19.3	0.0	53.9	0.0
σ_ε^2	1.3	0.0	8.8	0.0	0.6	0.0	9.1	0.0

The results in Table 1 illustrate the kind of biases that can be encountered when ignoring an informative sample selection process. For the present simulation, informative sampling of pupils within the schools (Method B2) has a much stronger biasing effect than informative selection of schools (use of PPS sampling, Method A2). In particular, large biases are observed when estimating the between schools variance σ_u^2 and the two region coefficients γ_1 and γ_2. (The estimator $\hat{\sigma}_u^2$ is biased under all four sampling schemes considered, as discussed below.) The P-values for the significance of the bias indicate that almost all the estimators are biased under informative selection of pupils (Method B2). There is also a statistically significant bias in the estimation of the intercept, γ_0 and the within school variance, σ_ε^2 under informative selection of the schools, even with non-informative selection of pupils, although the relative bias is rather low in these cases.

Next we compare the two approaches that account for the sample selection process. Table 2 shows the bias estimates obtained under probability weighting (PW) and the use of the sample model (SM). As clearly seen, the bias estimates are generally much smaller with the use of these approaches than in Table 1, particularly under Method B2, but under probability weighting large and significant biases still persist in the estimation of σ_u^2 with all four sampling schemes and the estimation of γ_1 under informative selection of schools and non-informative selection of pupils. The use of the sample model yields biased estimates for σ_u^2 under non-informative selection of pupils within the schools (Method B1) with both sampling methods for the schools. The latter bias may seem surprising particularly since the bias is much smaller under informative selection of pupils but this is not unique to our study (see below). Except for these cases, the use of probability weighting and the sample model yield similar biases under the four sampling methods.

Table 2. Percent Absolute Bias (PAB) when accounting for the sampling process by use of Probability Weighting (PW) and the Sample Model (SM)

	Selection of Schools							
	Non Informative, Method A1				Informative, Method A2			
Selection of Students	Non Informative Method B1		Informative Method B2		Non Informative Method B1		Informative Method B2	
Parameter	PW	SM	PW	SM	PW	SM	PW	SM
γ_0=86.9	0.1	0.0	0.6	0.9	1.9	0.1	1.3	0.2
γ_1=-6.8	5.1	4.3	1.1	2.8	14.9	1.7	5.7	5.3
γ_2=-13.8	0.9	0.6	1.6	3.0	5.8	4.1	0.6	0.2
β_1=-10.9	0.3	0.4	2.5	3.8	1.6	1.5	0.6	2.9
β_2=-16.0	0.6	0.6	2.9	4.0	0.5	0.1	1.4	2.1
β_3=-36.5	0.1	0.1	2.7	3.4	1.0	0.3	2.3	3.2
β_4=-7.2	3.8	3.7	0.7	0.0	2.6	1.1	0.8	1.2
σ_u^2=132.2	10.2	23.6	25.3	6.2	19.6	18.4	34.7	8.3
σ_ε^2=963.0	0.8	1.3	1.0	2.3	1.8	0.2	0.9	4.0

Empirical percentage coverage of nominal 95% confidence intervals as obtained when ignoring the sampling design and by use of probability weighting and the sample model was also examined in the simulation study. The use of the sample model yields almost perfect coverage percentages for all the parameters under all the sampling schemes, including for the between schools variance, σ_u^2 where the point estimators have a large relative bias (Table 1). This is not the case with PW or when ignoring the sampling process. See Pfeffermann et al. (2001). The occasionally bad performance of the PW approach even in cases where the corresponding point estimators have a small bias suggests that with the sample sizes considered in this study the use of the normal approximation is not valid. These results illustrate the possible advantage of the use of the posterior distribution for the construction of confidence (credibility) intervals. We mention in this regard that we also computed the means and standard deviations of the lengths of the confidence intervals over the 400 replications and none of the three approaches dominates the others in this respect.

We noted earlier the large bias of the posterior mean of σ_u^2 in the case of non-informative selection of pupils within the schools. Browne and Draper (2002) observed a similar bias in a different context and compare the use of different prior distributions for σ_u^2. The use of probability weighting yields biased estimators for this variance under all four sampling schemes. It is important to mention in this regard that, unlike the estimation of the within schools variance, σ_ε^2 , that uses all the 40×10 pupil observations, the effective sample size for the estimation of σ_u^2 is 40, the number of selected schools. In order to study the effect of the number of schools on the behaviour of the posterior mean of σ_u^2 we are currently repeating the same simulation study increasing the number of schools in the sample from 40 to 80. Preliminary results obtained so far show bias reductions in the order of 35%-50% under both probability weighting and when using the sample model.

Summary and outline of future work

An important message reinforced in the present study is that ignoring an informative sample selection scheme and fitting the population model may yield large biases of point estimators and distort the analysis. We compare two approaches for controlling the bias. The first approach uses probability weighting to obtain approximately unbiased and consistent estimators for the corresponding census estimators under the randomisation distribution. The census estimators are the hypothetical estimators computed from all the population values, which for large populations are expected to be sufficiently close to the true model parameters. The second approach attempts to identify the parametric model holding for the sample data as a function of the population model and the first order sample selection probabilities and then fits the sample model. The two approaches have been shown in the simulation experiment to remove the bias of the point estimators except in the case of the between school variance where with a small number of schools the use of probability weighting produces large biases under all sampling schemes considered, including the non-informative scheme where both the selection of schools and the selection of pupils within the selected schools is by simple random sampling. The use of the sample model likewise produces biased estimators for this variance and the bias depends also in this case on the choice of the corresponding prior distribution.

Probability weighting has two important advantages over the use of the sample model. First and foremost, it does not require any additional assumptions beyond the specification of the population model, although the validation of the model under this approach is an open problem. The second advantage of this approach is that it is very simple and requires minimal computation resources, including the estimation of the variances of the point estimators. However, the use of probability weighting has some serious limitations already discussed in the introduction.

The use of the sample model is much more flexible and, with the specification of appropriate prior distributions, it makes it possible to simulate from the posterior distribution of the target parameters. This major advantage of the use of the sample model is demonstrated when comparing the percentage coverage of confidence intervals produced by the two approaches. Inference based on the sample model requires, however, the specification of the conditional expectations of the sample selection probabilities at the various levels of the model hierarchy given the values of the corresponding response and explanatory variables. As illustrated in the present study, these expectations may depend on a large number of additional unknown parameters that need to be estimated along with the population parameters (see Pfeffermann et al. 2001 for details). Application of this approach by MCMC is very computationally intensive and with small sample sizes, the performance of variance estimators may depend on the specification of the prior distributions even if restricted to non-informative priors. Nonetheless, with correct specification of the sample model the use of this approach overcomes the inference limitations of probability weighting noted in the introduction.

The major question underlying the use of the sample model for inference is its robustness to wrong specification of the conditional expectations of the sample selection probabilities that determine the sample model. This issue is currently under investigation.

References

Browne, W. J., and Draper, D. (2002). A comparison of Bayesian and Likelihood-based methods for fitting multilevel models. *Submitted*.

Pfeffermann, D., Skinner, C. J., Holmes, D. J., Goldstein, H., and Rasbash, J. (1998a). Weighting for unequal selection probabilities in multilevel models. *Journal of the Royal Statistical Society*, B, **60**: 23-40.

Pfeffermann, D., Krieger, A. M., and Rinott, Y. (1998b). Parametric distributions of complex survey data under informative probability sampling. *Statistica Sinica*, **8**: 1087-1114.

Pfeffermann, D., and Sverchkov, M. (1999). Parametric and semi-parametric estimation of regression models fitted to survey data. *Sankhya*, *B*, **61**: 166-186.

Pfeffermann, D., Moura, F. A. S. and Nascimento Silva, P. L. (2001). Multilevel modelling under informative probability sampling. *Proceedings of invited paper sessions organised by the International Association of Survey Statisticians (IASS) for the 53rd session of the International Statistical Institute*.

Spiegelhalter, D., Thomas, A., and Best, N. G. (2000). *Bayesian Inference using Gibbs Sampling. WinBUGS version 1.3, User manual*. MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge, UK.

Review of 'Multilevel Analysis, Techniques and Applications'. Joop Hox. x + 304 pages (2002). Mahwah NJ: Lawrence Erlbaum Associates Inc

M J R Healy

Institute of Education, University of London

Here is another introductory textbook, aimed mostly at the social sciences. It includes both regression and variance structure models and it assumes a basic knowledge of the corresponding techniques for single level data, including structural equation modelling.

The first part of the book follows quite closely the plan of the *MLwiN* manual. The first introductory chapter aims at motivating the use of multilevel methods. Chapter 2 introduces the usual two level model and this is followed by discussions of estimation methods and tests of significance. One chapter deals with repeated measurements, another with binary data and proportions using a logistic model. Methods for cross-classifications, meta-analyses and multivariate data are also described. A useful chapter is devoted to power calculations and sample size calculations. Advanced methods include robust standard errors, bootstrapping and Bayesian techniques such as MCMC. The remaining three chapters are devoted to factor, path and latent-curve models. A substantial number of datasets are used as examples, some genuine, some simulated, and most of the tabulated results are in a form which will be familiar to *MLwiN* users.

There are by now several introductory texts on multilevel modelling (there is even one competing text from the same publisher) and it has to be asked whether the book under review presents particularly attractive features. I found its approach quite difficult in several respects. The initial (artificial) example relates to a number of classes with girl and boy pupils in each. The y-variate is a 0-10 popularity score (possibly somewhat non-Normal?) and the predictors are gender at level 1 and teacher's length of experience at level 2. This immediately introduces a categorical variable so that the 'slopes' at level 1 (even when actually plotted as straight lines) are somewhat metaphorical. The earliest two level model to be considered has both intercepts and slopes random so that the discussion has to plunge immediately into cross-level interactions and heteroscedastic errors. On the other hand, no provision is made for different variances for boys and girls. As is not uncommon, the discussion of centering is needlessly elaborate. A reader who struggles with the explanation of the third 'dummy' level in cross-classified analysis on page 126 is not likely to be encouraged by the *MLwiN* equations window that appears on the same page.

The publishers rather blandly announce that camera-ready copy was supplied by the author. This means that the rather numerous misprints must be laid at the author's door.

Explaining multilevel modelling to an audience of social scientists is no easy task. The pioneers, Goldstein and Bryk & Raudenbush did their best, and I am not sure that this book or their other successors have improved upon them.

MLwiN/MLn activities: Summary of an electronic survey

Min Yang

Institute of Education, University of London

The survey of *MLwiN/MLn* user activity was designed to collect feedback from users on what purposes they used the program for, what difficulties they had in using it, and what new features they would like the program to have in future.

The questionnaire was placed on the web site of the Centre for Multilevel Modelling. Invitation to participate in the survey was made through the multilevel email discussion list and the Centre newsletter mailing list. Over six weeks, 100 users replied. A brief summary of the survey data follows.

The majority of respondents are from universities and research institutions in the higher education sector. A minority are from government organisations or independent institutions, and a few are self-employed statisticians or analysts. They come from the following areas: Europe (32%), North America (26%), United Kingdom (21%), Australia (9%), Other (12%).

Their experiences in using *MLwiN/MLn* go back to before 1995 with 39% having at least five years experience although the same percentage have two years experience or less.

Only 15% of users said that they used the program for teaching on courses or workshops. There was praise for the Equation window, and for the examples and manuals that are suitable for teaching. The main drawbacks are occasional program crashes and lack of good facilities for data preparation. It was found difficult to explain to students the reasons for creating columns of constants and the denominators for modelling binary outcomes.

Eighty two percent of respondents made remarks on new features that they would like to see in future versions of *MLwiN*. In brief, their suggestions can be grouped into 10 classes in the following table. For each suggested new feature, the Project team has responded in terms of possible solutions, both short-term and long-term.

Summary for new features requested by *MLwiN* users

Description of the new features required	**% of respondents**	**Response by the Team**
1. Better data import and export. Better facilities for data exchange between other packages such as SAS, SPSS, SPLUS, Excel, ACCESS, Fox, STATA	39	The development version of *MLwiN** can import categorical alphanumeric variables. We currently have no plans to build direct data exchange with other packages via non-ASCII data formats.
2. New tools (Spline /local regression), estimation procedures, automatic inclusion of test statistics and p-values on model output, Heckman models, maximum likelihood estimation for GLMM, more development for survival models, factor analysis & IRT model, automatic inclusion of ICC	18	We are currently researching simulated maximum likelihood and quadrature as two possible estimation procedures for producing direct likelihood estimation for GLMMs. A basic factor analysis model, estimated via MCMC, is included in the development version of *MLwiN*. We have just received research funding to further develop modelling of multivariate multilevel survival data.
3. More elaborate manuals or on-line help for data management, for commands/macros, for graphics, and for troubleshooting. More examples for different models	18	New manuals are currently being written. Also, we are commissioning other training materials to be written. These materials will appear in a training archive of our web site.
4. Make the program more stable, crashing less	12	We are working on it.

5. Having a GUI for more complex models such as multinomial models, survival models, time series models, cross-classified and multiple membership models, getting rid of BCONS/PCONS and better performance on GLMM	11	The development version of *MLwiN* has a setup window for cross-classified and multiple membership models using an MCMC engine for estimation. The models are easier to specify in the new version and the MCMC estimation algorithm is inherently more efficient than IGLS for non-hierarchical models. BCONS/PCONS etc will be removed from the next release of the development version which will result in a more direct representation for GLMM in *MLwiN*. Also multinomial models will be available from the equations window.
6. Better graphical facilities	10	
7. Version for student, for OS platforms such as Linux and Mac	5	We are not considering producing MAC or LINUX versions of *MLwiN*. We do not produce a student version. However, there is a training version of *MLwiN* available from tramss.data-archive.ac.uk. This training version will be extended and we will implement a training archive on the *MLwiN* website.
8. More user-friendly in general	5	We are trying. We are currently developing software features and documentation to make the move from single level modelling to multilevel modelling easier.
9. Allowing string variables	2	The development version can now read alphanumeric data.
10. Macro facilities fitting models one after another in batch mode	1	This can already be done using the macro language.

*See companion note in this newsletter entitled "*MLwiN* Version 1.2 – Development version"

MLwiN Version 1.2 – Development version

William Browne and Jon Rasbash
Institute of Education, University of London

On 12th April 2002 a new 'development' version of the *MLwiN* software package was made available to the *MLwiN* user community. This version of the software contains many improvements and new features including the ability to fit many new models. It will be regularly updated with bug fixes and additional features until the end of the *MLwiN* project team's current ESRC funded project in February 2003 when it is envisaged that it will replace *MLwiN* 1.1 as the official released version of the software.

The software is available to download (for existing *MLwiN* users) from the website at http://multilevel.ioe.ac.uk/dev/develop.html The software was last updated on 15th May 2002.

Data input has been improved greatly in *MLwiN* 1.2. The software can now read in alphanumeric data via the 'Paste' option and automatically treats such data as categorical. There is also an improved interface for fitting categorical parameters as main effects and interaction terms into a model.

The MCMC estimation methods have been greatly extended and a new MCMC manual is available to download from the web page. MCMC can now be used to fit many models previously only available using likelihood-based methods e.g. multivariate response models and complex level one variation. There are also many other models that are either only available using MCMC or are easier to fit using this method. New models available using MCMC include multilevel factor analysis models, measurement error models, cross-classified, multiple membership and spatial models. There is also the opportunity to use the DIC diagnostic for model comparison for some types of models.

Some Recent Publications Using Multilevel Models

Issue on Multilevel Modelling of Statistical Methods in Medical Research, **10** (6). (2001).
Thompson, S. G., Turner, R. M., and Warn, D. E., Multilevel models for meta-analysis, and their application to absolute risk differences, 375-392.
van den Oord, E. J. C. G., Estimating effects of latent and measured genotypes in multilevel models, 393-407.
Rabe-Hesketh, S., Yang, S. and Pickles, A., Multilevel models for censored and latent responses, 409-427.
Longford, N. T., Multilevel analysis with messy data, 429-444.

Barthelemy, K. D. (2001). Modelling hierarchically clustered longitudinal survival processes with applications to childhood mortality and maternal health. *Canadian Studies in Population* (Special Issue on Longitudinal Methodology), **28** (2): 535-561.

Berkhof, J., and Snijders. T. A. B. (2001). Variance component testing in multilevel models. *Journal of Educational and Behavioural Statistics*, **26** (2): 133-152.

Blatchford, P., Goldstein, H., Martin, C. and Browne, W. (2002). A study of class size effects in English school reception year classes. *British Educational Research Journal*, **28**: 169-185.

Browne, W. J., Draper, D., Goldstein, H. and Rasbash, J. (2002). Bayesian and Likelihood methods for fitting multilevel models with complex level-1 variation. *Computational Statistics and Data Analysis*, **39**: 203-225.

Carlin, J. B., Wolfe, R., Hendricks Brown, C., and Gelman, A. (2001). A case study on the choice, interpretation and checking of multilevel models for longitudinal binary outcomes. *Biostatistics*, **2** (4): 397-416.

Fielding, A. (2002). Teaching groups as foci for evaluating performance in cost-effectiveness of GCE Advanced Level provision: Some practical methodological innovations. *School Effectiveness and School Improvement*, **13** (2): 225-246.

Goldstein, H., and Browne, W. J. (2002). Multilevel factor analysis modelling using Markov Chain Monte Carlo estimation. *Latent Variable and Latent Structure Models*. Marcoulides, G., and Moustaki, I. (Eds.). London, Lawrence Erlbaum.

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

Oliver, J.C., Rosel, J. and Jara, P. (2000). Modelos de regresión multinivel: aplicación en psicología escolar. *Psicothema*, **12**: 487-494.

Plewis, I. (2002) Modelling impact heterogeneity. *Journal of the Royal Statistical Society*, *A*, **165**: 31-38.

Pituch, K. (2001). Using multilevel modeling in large-scale planned variation educational experiments: Improving understanding of intervention effects. *Journal of Experimental Education*, **69**: 347-372.

Pituch, K. (1999). Describing school effects with residual terms: Modeling the interaction between school practice and student background. *Evaluation Review*, **23**: 190-211.

Pituch, K., and Miller, J. (1999). Strengthening multisite education interventions: An illustration with multilevel modeling. *Educational Research and Evaluation*, **5**: 62-75.

Rabe-Hesketh, S., and Skrondal, A. (2001). Parameterization of multivariate random effects models for categorical data. *Biometrics*, **57**: 1256-1264.

Spencer, N. H., and Fielding, A. (2002). A comparison of modelling strategies for value-added analysis of educational data. *Computational Statistics*, **17** (1): 103-116.

Yang, M., Goldstein, H., Browne, W. J., and Woodhouse, G. (2002). Multivariate multilevel analysis of examination results. *Journal of the Royal Statistical Society*, *A*, **165** (1): 137-153.

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

MLwiN Clinics in London

Tuesday 1 October 2002
Tuesday 5 November 2002
Tuesday 3 December 2002

at

Multilevel Models Project
11 Woburn Square, London WC1H 0NS

Contact Min Yang for appointments
Tel: +44 (0) 20 7612 6682
m.yang@ioe.ac.uk