MULTILEVEL MODELLING NEWSLETTER

Vol. 10 No. 2　　　　　　　　　　　　　　**November, 1998**

More forthcoming Workshops

• **6-8 January 1999**, a three-day introductory workshop to multilevel modelling using *MLwiN* will take place in the Institute of Education, University of London. Enquiries to Anne-Lise McDonald at *Health Policy and Practice Unit, UEA, Norwich, NR4* 7TJ. Tel: +44 (0)1603 514867, email**: a.cox@uea.ac.uk**.

• **25-26 March 1999,** a two-day introductory workshop to multilevel modelling using *MLwiN* is taking place in the Institute of Education, University of London. Enquiries to Min Yang at Mathematical Sciences, Institute of Education, 20 Bedford Way, London WC1H 0AL. Tel: +44 (0)171 612 6682, email: **m.yang@ioe.ac.uk**.

New project in applying multilevel analysis for A-level grades :

The *Multilevel Models Project* in London has been funded by the Economic and Social Research Council (ESRC) in the UK for an analysing of A-level exam grades. This study will look into some practical issues of applying multilevel multivariate models as well as ordered categorical data models.

Call for paper proposals

A package of research papers on the application of multilevel modeling (random effects, hierarchical models) to psychological data is to be put together for a special issue of Multivariate Behavioral Research and will be refereed. The goal of the special issue is to introduce MLM, and the special advantages, strengths, and problems involved, to a quantitatively sophisticated audience of psychologists. There is interest in: 1) illustrative papers that apply MLM to address interesting substantive psychological research questions with either longitudinal or cross-sectional data analysis, 2) papers dealing with quantitative issues pertaining to the application of MLM. For example, research reports dealing with power issues, violations of assumptions, computation of effect sizes, or identification of outliers, would be most welcome.

If you have a paper idea for the special issue, please send a 2-page proposal outlining your ideas to Steve Reise (**reise@psych.ucla.edu** or tel:310 794 1594). The deadline for proposals is Feb 1, 1999.

Also In This Issue

A review of *MLwiN* V1.01

Review of 'Statistics in Education'

Fitting multilevel models using SAS

Numerical intergation via high-order, multivariate Laplace approximation for generalised linear multilevel models

Some new references on multilevel modelling in 1997/98

E·S·R·C
ECONOMIC

Multilevel Modeling in Windows; a review of *MLwiN*

(Version 1.01.0001 of May 1998)

Joop Hox, University of Amsterdam, The Netherlands

There are two strategies to move a DOS program, such as *MLn*, to Windows. The first is to keep the program more or less as it is, and to tack on an interface with dialog boxes and buttons to replace the command-driven interface. This is useful, because reading questions and clicking on buttons are much easier than trying to remember commands which then must be typed in without errors. The second strategy is to really use the graphical and multitasking abilities of Windows to enhance the capabilities of the program and to give the user more control. The new Windows version of *MLn*, called *MLwiN*, is clearly trying to accomplish the latter. The old *MLn* is still there, and the old commands can still be used, but *MLwiN* has many new features, some of which use the Windows interface to do things that are difficult to program under DOS.

When *MLwiN* starts, we get an almost empty screen, with a menu bar, and a tool bar. To start working, we must load a raw data file or a worksheet (*MLwiN* can import and export *MLn* worksheets), and open one of the windows.
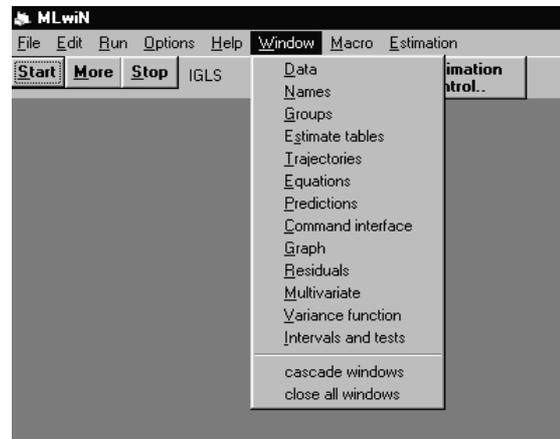
As a user, when you upgrade a program to a radically changed version, there is a natural inclination to keep the old version around, just to make sure that you have something familiar that you know how to work with. In *MLwiN,* the old version is actually available by opening the **Command interface** window and its associated output window. The old *MLn* commands can then be typed in, and the familiar output appears in the Output window. Only the command HELP does not work, in fact I managed to crash *MLwiN* by doing just this [1].

---
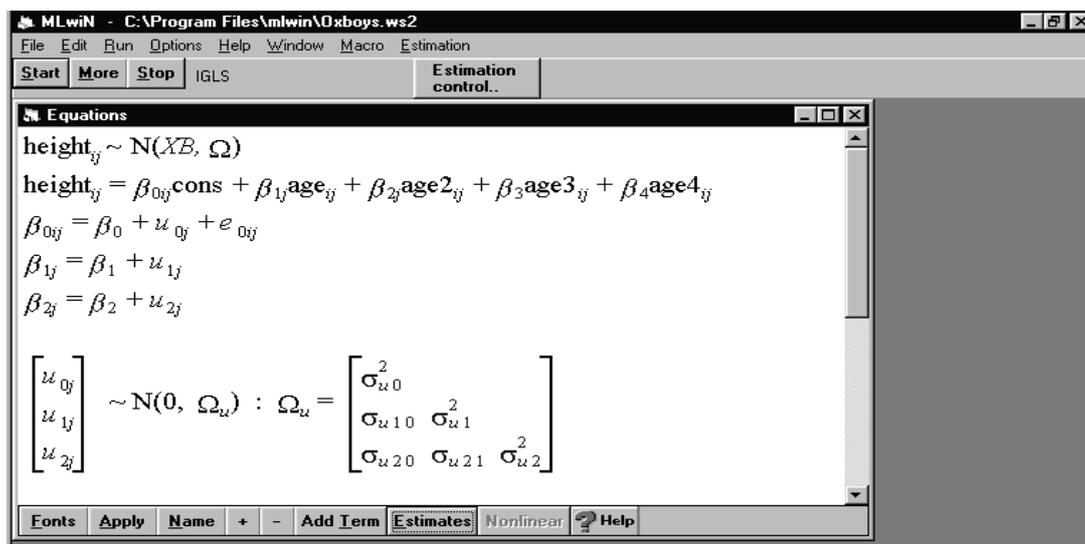
[1] Editor's note: this has been fixed in version 1.02 (June 1998).

The central command centre to specify models is the **Equations** window. Here, the graphical capacities of Windows are used to build up the equation for the model. If we load the example worksheet Oxboys.ws2, and open the equations window, we get the equations for the current model. Clicking on the Names button then produces the equations with variable names instead of X's and Y's.

We now have an algebraic representation of our model. The brief notation ~N(XB, Ω) informs us that we assume a normal distribution for the residual errors. By clicking on the **N**, we can change this, for example into a binomial error structure. Other options are to put estimates and standard errors in place of the beta's and sigma's.

The equation window makes clear what I mean with 'really using the Windows interface.' We can click on various parameters to open a dialog box, to specify if it is in the random part at one of the levels. When the specifications of a parameter are changed, for instance by putting it in the random part at level two, this is immediately reflected in changes in the equations. During the iteration process, we can view the current estimates. Other windows show the estimates in a table, and show the **Trajectories**, that is, a graph of the successive estimates in the iteration process. If the iterations are still running, the trajectories are continually being updated, so we can monitor the iteration process. This is very useful to diagnose convergence problems. The **Residuals** window is used to specify the calculation of residuals, which can then be used in the **Graph** window to construct a variety of plots. Of course, the equations, tables and graphs can be copied to another program via the Windows clipboard.

There are some new features in *MLwiN*. The most remarkable are two new estimation methods: bootstrapping and MCMC. *MLn* already had facilities for bootstrapping. In *MLwiN,* bootstrapping is built into the Windows interface, which makes it easy to use. The method is a *parametric bootstrap*, which means that the program computes bootstrap samples by adding randomly chosen residuals from the appropriate error distributions. This is used to correct estimates for bias, and to construct standard errors and confidence intervals if there are reasons to distrust the asymptotic standard errors, for instance with small samples.

Markov Chain Monte Carlo (MCMC) estimation methods are quite different. To begin with, they are Bayesian, which means that we have to provide a prior distribution for the parameters we want to estimate. By default, *MLwiN* uses a very flat normal distribution for the regression coefficients, and a uniform distribution between 0 and a very large number for the (co)variances. Bayesian statistics combine the prior distribution and the data into a posterior distribution for the parameters. If the posterior distribution is simple, we can calculate its mean or mode, and establish confidence intervals. Often, the posterior distributions are extremely complex. MCMC methods are procedures to sample parameter values from the posterior distribution. Given enough independent samples from the posterior distribution, we can calculate the mean, mode, and confidence intervals on this sample. *MLwiN* has two MCMC procedures: Gibbs sampling and Metropolis-Hastings. MCMC methods typically have two stages: a *burn-in* stage and a *monitoring* or sampling stage. In the burn-in stage, the MCMC computations are run for a large number of iterations, to allow it to converge on the correct posterior distribution. Next, a large number of iterations following that are collected, to be used for estimating parameter values and standard errors. There are two problems here. The first is to ensure that we have enough iterations in the burn-in to converge on the correct posterior distribution. The second is to collect enough independent samples from the posterior distribution to permit good estimates. The key word here is *independent*. The successive draws produced by the MCMC method are correlated, sometimes highly correlated. Thus, we must find out how high these correlations are. Then, we draw either only at every tenth or at every hundredth MCMC iteration, or we simply draw a very large number of MCMC samples, much more than would be required if they were independent.

MLwiN provides some useful diagnostics to inspect the MCMC results, such as parameter plots at the successive iterations and autocorrelations. The example shows the MCMC diagnostics window overlaid on the trajectories window. In the trajectories window, we see the successive values of the parameters. What we like to see there, is a plot with no structure. By clicking on a specific plot, we get the MCMC diagnostics, that help us to decide whether we

have enough MCMC iterations for the burn-in and the monitoring process.

several different methods and comparing the results. If the results vary strongly across different estimation methods, we should probably not trust any of them. It could be that our data do not contain much information on the parameter that we try to estimate, and we must either modify the model or collect more data.

Conclusion

In general, I like *MLwiN* a lot. The program is not a simple port of the DOS version to Windows. It has a many new features, both statistical and in the user interface, and some
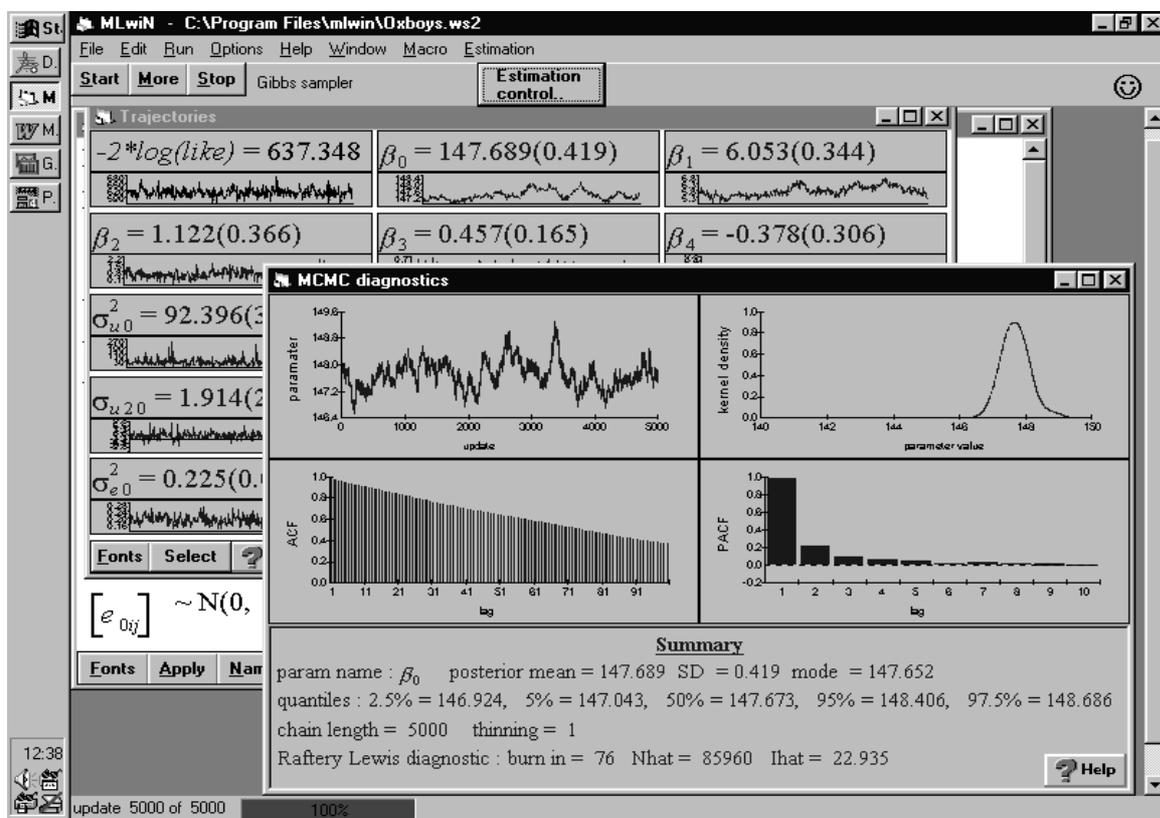
The MCMC diagnostics shown are for the intercept parameter in the Oxboys example. The autocorrelations, shown in the ACF box, are large in this example; the correlation between draws that are separated by 100 iterations is still appreciable. The number NHAT=85960 is the number of MCMC iterations required for precise estimation of the confidence intervals. This is clearly a case where we should use a very large number of iterations, and possibly monitor only every hundredth or so. A nice feature of *MLwiN* is that it makes it easy to switch between different estimation methods. We could do a sensitivity analysis by using

old features are now much easier to use with the new interface. It is the first Windows version of a complex program; it looks well, and is pretty stable. In the course of preparing this review, I have tried it on a variety of data files, and used most of the available features. In doing this, I have managed to crash the system once, which could easily be caused by one of the other three programs that were concurrently running. *MLwiN* provides users with new bootstrap and MCMC estimation methods simply by hitting a button; these certainly are powerful capabilities. With reasonably-sized data sets (two level problems with about 1000 individuals), the

computer-intensive bootstrapping and MCMC methods ran fast enough on a Pentium 233 system. A few features are still only available through the Command window, and sometimes clicking on an object produces a 'feature not yet implemented' message. However, the makers of *MLwiN* aim to put such features into Windows point-and-click interfaces eventually. Program upgrades and macro's can be downloaded from the *MLwiN* website. (**http://www.ioe.ac.uk/mlwin/**)

I have some minor grumblings too. The documentation is, just like earlier versions, rather terse. There is a good tutorial, which runs new users through a number of examples. This is nice,

but it also makes it difficult to find specific information. *MLwiN* has a good Help system, which is expanded regularly and upgraded via the website, but not everybody likes to look up everything online. For instance, all information on constructing macro's is in the Help system. In going through that section, I get the impression that it is now possible to use macros as a plug-in to the program, so they feel like an extension of the program. That is neat. However, it is not very clear how to accomplish this. I would like to have had more examples.

Fitting multilevel models using SAS PROC MIXED

Judith D. Singer
Harvard University, Graduate School of Education, USA

SAS PROC MIXED is a flexible program suitable for fitting many types of multilevel models. Although the documentation is complex and the "defaults" are rarely the ones multilevel modelers desire, the ability to do data reduction, management, and analysis in a single software package already familiar to many researchers makes this approach particularly attractive.

In a forthcoming paper in the *Journal of Educational and Behavioral Statistics,* I provide a step-by-step tutorial for using PROC MIXED to fit three common types of multilevel models: (1) two-level "school effects" models, for data on individuals nested within hierarchies; (2) two-level "individual growth" models, for exploring longitudinal data over time; and, (3) three-level models, which combine these two features. The full text of the paper can be downloaded from **http://hugse1.harvard.edu/~faculty/singer**.
Here, I introduce the basic logic underlying the PROC MIXED syntax, present illustrative output, and briefly describe some of the models that can be fit.

Basic Syntax

The logic behind PROC MIXED syntax can be understood by considering an unconditional means model for a classic two-level "school effects" data set. Suppose you have data at two levels of a hierarchy–say, students within schools--and you would like to examine the behavior of a level-1 outcome, Y_{ij}. Analysis usually begins with an unconditional means model with no predictors, which expresses Y_{ij} as the sum of a level-2 "intercept" (β_{0j}) and a level-1 random error (r_{ij}) associated with the ith student in the jth school. At level-2, we express the school-level intercepts as the sum of an overall mean (γ_{00}) and a series of random deviations from that mean (u_{0j}). We further assume that $r_{ij} \sim N(0, \sigma^2)$ and $u_{0j} \sim N(0, \tau_{00})$. This yields:

$$Y_{ij} = \beta_{0j} + r_{ij}$$
$$\beta_{0j} = \gamma_{00} + u_{0j}$$

which can be rewritten as:

$$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij} \qquad (1)$$

The combined specification in (1) is most helpful for specifying a model in PROC MIXED. As highlighted by the brackets, the model is composed of two parts: a *fixed* part, which contains the single term γ_{00} (for the intercept) and a *random* part, which contains two terms (for the intercept u_{0j} and for the within-

school residual r_{ij}). PROC MIXED syntax reflects this partition:

```
proc mixed;
   class school;
   model y = /solution;
   random intercept/sub=school;
```

After invoking the procedure (using the PROC statement) and identifying categorical variables (using the CLASS statement), the MODEL statement specifies the fixed effects and the RANDOM statement specifies the random effects. This MODEL statement may appear odd because it seems as if it has no predictors. In reality, it has one *implied* predictor, the vector **1**, the intercept, included by default. (The intercept can be suppressed by adding the option /NOINT to the MODEL statement.) The /SOLUTION option asks SAS to print estimates for fixed effects.

The RANDOM statement is crucial and its specification is usually the trickiest part about using PROC MIXED. Like most regression programs, the procedure assumes one random effect for the lowest-level (here, within-school) residual, r_{ij}. By specifying the INTERCEPT on the RANDOM statement, we indicate the presence of a second random effect–that the implied INTERCEPT should be treated not only as a *fixed* effect (γ_{00}) but also as a *random* effect (τ_{00}). The SUB= option on the RANDOM statement specifies the multilevel structure, indicating how level-1 units are divided into level-2 units. Here, the subgroups are designated by the classification variable SCHOOL. Without this RANDOM statement, the model fit would be

$Y_{ij} = \gamma_{00} + r_{ij}$.

Figure 1 presents the results of fitting this model to the High School and Beyond math achievement data included with HLM (Bryk, Raudenbush, & Congdon, 1996). Equivalent results are presented on pages 62-66 of Bryk and Raudenbush (1992). The output is comprised of four sections. The *Iteration History* section provides convergence information. Because the program is efficient, convergence is often quick, as it is here. With more complex models or grossly imbalanced data sets, convergence slows. The *Covariance Parameter* section provides estimates and hypothesis tests for random effects. The *Model Fitting* section provides information useful for comparing the goodness of fit of multiple models with the same *fixed effects* but different *random effects*. The two criteria likely to be the most helpful are Akaike's Information Criterion and Schwarz's Bayesian Criterion. The final section presents parameter estimates and hypothesis tests for fixed effects.

Including level-1 and level-2 predictors

Once you understand the basic syntax, it's easy to include level-1 and level-2 predictors. Additional fixed effects are specified on the MODEL statement; additional random effects are specified on the RANDOM statement. Interactions between predictors (at the same level or at different levels) are specified using VAR1*VAR2 syntax. Categorical predictors can be specified using either user-defined dummy variables or by adding them to the CLASS statement.

Many researchers want to "center" predictors at either the grand-mean or the context mean. Unlike some multilevel programs that allow the user to center when specifying a model, users of PROC MIXED must center in a previous DATA step. (PROC SUMMARY, which computes grand means and context means, makes this easy to do.) Given the misconceptions surrounding centering and its consequences, some might argue that this deliberate approach is preferable to the automatic approach available in other programs.

To illustrate, consider two predictors: MEANSES, the school mean SES and CSES, the difference between each student's SES and the school mean. We fit an "intercepts and slopes as outcomes" model as follows:

```
proc mixed;
     class school;
     model y = cses meanses

cses*meanses/solution;
     random intercept cses/
          sub=school type=un;
```

Figure 1: Results of fitting an unconditional school effects model

```
                    REML Estimation Iteration History

            Iteration  Evaluations     Objective      Criterion

                    0          1    34899.608417
                    1          2    33913.503461     0.00000109
                    2          1    33913.484655     0.00000000

Convergence criteria met.

                    Covariance Parameter Estimates (REML)

     Cov Parm        Ratio       Estimate     Std Error      Z     Pr > |Z|

     INTERCEPT    0.21992178    8.60965741   1.07782320    7.99    0.0001
     Residual     1.00000000   39.14872611   0.66065147   59.26    0.0001

                    Model Fitting Information for MATHACH

     Description                          Value

     Observations                       7185.000
     REML Log Likelihood                -23558.4
     Akaike's Information Criterion     -23560.4
     Schwarz's Bayesian Criterion       -23567.3
     -2 REML Log Likelihood             47116.79
```

Notice the similarity between this syntax and that for the unconditional means model. Three fixed effects are added to the MODEL statement–one for each predictor and a third for the cross-level interaction. A single random effect, CSES, is added to the RANDOM statement to indicate that student CSES slopes are allowed to vary across schools. The TYPE=UN indicates that the variance-covariance matrix for the random effects is completely general, telling SAS to estimate a variance component for the INTERCEPTs and for the CSES slopes as well as the covariance between them.

Individual growth models

Individual growth models can be fit by using a RANDOM statement, which mimics a school-effects analysis or by using a REPEATED statement, which mimics repeated measures analysis of variance. Either way, you must first create a person-period data set in which each individual has one record for every time-period that he or she is observed. I present SAS code for the conversion in Singer (in press).

Consider the basic individual growth model:

$$Y_{ij} = \pi_{0j} + \pi_{1j}(TIME)_{ij} + r_{ij}$$

$$\pi_{0j} = \gamma_{00} + u_{0j}$$

$$\pi_{1j} = \gamma_{10} + u_{1j}$$

(2)

where we assume:

$$r_{ij} \sim N(0, \sigma^2)$$

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix}$$

(3)

The model in (2), which can be rewritten as:

$$Y_{ij} = [\gamma_{00} + \gamma_{10} TIME_{ij}]$$

$$+ [u_{0j} u_{1j} \pi_{1j} TIME_{ij} + r_{ij}]$$

(4)

can be fit using the following code:

```
proc mixed;
   class person;
   model y = time/solution;
   random intercept time/
        subject=person type=un;
```

Notice the similarity to the syntax for the school-effects analysis. The CLASS variable changes from SCHOOL to PERSON to indicate that the data represent multiple observations over time for individuals. By using this CLASS variable on the RANDOM statement (with the SUBJECT=option) we allow both intercepts and slopes to vary across people.

This model places a common, but often unrealistic, assumption on the behavior of the r_{ij}, the within-person residuals over time. In fitting a model in which only the intercepts vary across persons, we assume a compound symmetric error covariance matrix for each person. In fitting a model in which the slopes vary as well, we introduce heteroscedasticity into this matrix. One of the appeals of PROC MIXED is that it allows

the user to compare different structures for the error covariance matrix. Instead of the individual growth model in (2)-(4), consider the following alternative:

$$Y_{ij} = \pi_{0j} + \pi_{1j}(TIME)_{ij} + r_{ij}$$

$$\pi_{0j} = \gamma_{00}, \qquad \pi_{1j} = \gamma_{10} \qquad (5)$$

$$r_{ij} \sim N(0,\Sigma)$$

Here, the intercepts and growth rates are assumed to be constant across people. But the model introduces a different type of complexity: the residual observations *within* persons (after controlling for the linear effect of TIME) are correlated through the within-person error variance-covariance matrix Σ. By specifying alternative structures for Σ (that ideally derive from theory) the user can compare the goodness of fit of different models. SAS has dozens of alternative error-covariance structures built in. If there are only three waves of data, it is worth exploring only a few of these because there is so little data for each person. With additional observations per person, exploring these alternatives for Σ (called R in PROC MIXED) can be profitable. Consult pages 311-313 of the PROC MIXED documentation, pages 92-102 of *The SAS System for Mixed Models*, and Wolfinger (1996).

Three-level models and other extensions

PROC MIXED can be used to fit three- and higher-level models. In the case of "school-effects" analyses, you can include additional RANDOM statements, with appropriate nesting specifications given in the SUB= option. For example, if you have data on students within teachers within schools, you can fit an unconditional means model with the syntax:

```
proc mixed;
  class teacher school;
  model y = /solution;
  random intercept/sub=school;
  random intercept/
        sub=teacher(school);
```

In the case of longitudinal analyses that track individuals nested within groups, the school-effects specification can be added to the individual growth model specification.

Other options are available to users interested in more complex mixed models. Heterogeneity in the error variance-covariance matrix can be introduced using a /GROUP option on the RANDOM statement. Sampling-based Bayesian analysis can be conducted using a PRIOR statement that permits a variety of distributional specifications for the variance components parameters' prior density. SAS also provides two macros--GLMMIX and NLINMIX--that can be used for fitting generalized linear mixed models and nonlinear mixed models that do not involve the normal continuous outcomes treated here.

Because of its flexibility, it takes practice to use PROC MIXED. In addition, because it was not written with multilevel models in mind, the procedure's defaults must usually be overridden. The investment in time can yield a high payoff, however, as once the basic syntax is understood, many diverse types of models can be fit. Users wanting to learn more should consult the SAS documentation for PROC MIXED (SAS Institute, 1996), *The SAS System for Mixed Models* (Littell et al., 1996), and Singer (forthcoming).

References

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models* Newbury Park, CA: Sage.

Bryk, A. S., Raudenbush, S. W., & Congdon, R. T. (1996). *HLM: Hierarchical linear and nonlinear modeling with the HLM/2L and HLM/3L programs.* Chicago, IL: Scientific Software International, Inc.

Latour, D., Latour, K., & Wolfinger, R. D (1994). *Getting started with PROC MIXED*. Cary, NC: SAS Institute, Inc.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS System for Mixed Models*. Cary, NC: SAS Institute, Inc.

SAS Institute (1996). *SAS/STAT Software: Changes and Enhancements through Release 6.11*, Cary, NC: SAS Institute, Inc.

Singer, J. D. (in press). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual

growth models. *Journal of Educational and Behavioral Statistics.*

Wolfinger, R. D. (1996). Heterogeneous variance-covariance structures for repeated measures. *Journal of Agricultural, Biological, and Environmental Statistics, 1*(2), 205-230.

Review of 'Statistics in Education' (*Ian Plewis. London: Arnold, 1997*)

Juan C. Oliver, Rodriguez Universidad Jaime I, Castellon, Spain

The purpose of the book is to review recent developments in methods for the statistical analysis of educational data that can also be of interest to social scientists in general. It is addressed to a reader with some introductory background in statistics, graduate students and researchers. The book focuses specially on identifying different types of statistical models that are useful in frequent kinds of educational questions and research designs. The first half is devoted to continuous responses, while the second half covers categorical data. The emphasis is on substantive aspects of model interpretation and usage, rather than on details of statistical inference or computation. Abundant references are given on these issues for the interested reader.

The multiple regression model and its assumptions are reviewed in chapter two; examples include situations with categorical and continuous predictors, transformed response variables, residual plots, and some of the problems with missing data. Discussion of the ecological fallacy (Freedman et. al, 1991) serves as a link to multilevel models, in a case where different regression lines and processes are observed within and between classrooms for mathematics attainment as a function of curriculum coverage.

Use of realistic models of the hierarchical nature of educational data (pupils within classrooms within schools within local education authorities within governments) is dealt with in chapter three. Two and three level models with random intercepts and slopes are introduced that allow one to quantify the amount of variability in pupil attainment and progress (level one) that is accounted for by differences in higher level units, such as classrooms (level two) or schools (level three). A series of graphs and residuals plots illustrate these ideas graphically, as well as providing diagnostics of model fit. The identification of explanatory variables that may reduce the magnitude of the random effects observed at higher levels is then discussed as a second step in model building; e.g., differences in school funding (level three), teacher styles (level two) or pupil sex (level one) are variables that could explain the above components of the variance in pupil attainment. The importance of taking into consideration for the hierarchical nature of data to avoid biases and improve efficiency in the design and analysis of intervention studies is discussed. A brief overview of software for multilevel modelling is also given.

Chapter four addresses another type of hierarchical structure often found in educational research: repeated measurements nested within pupils. The advantage of using multilevel modelling in this situation versus a more traditional approach is reviewed. Most of the chapter is devoted to an analysis of a set of longitudinal measures of reading attainment, which is used to show how to specify, estimate and interpret some basic two and three level growth curve models. The development of reading is shown to be a nonlinear function of age, with significant components of variation at levels two (pupils) and three (schools). Sex and ethnic groups are described as useful explanatory variables at level two. The possibilities of modelling and testing heteroskedasticity of residuals at level one (reading level at different measurement occasions) is also briefly introduced. Some of the problems and types of missing data are discussed, along with useful

diagnostic techniques. Finally, a word of caution is given on the dependence of certain statistical results on the type of measurement scale used; potential interpretation problems and the strength of the conditional regression model in this respect are reviewed.

Chapters five, six and seven deal with categorical data, reviewing the analysis of contingency tables for dichotomous or polytomous responses, ordered and unordered, logistic regression and log-linear models. Methods for dependent measurements in the analysis of change and agreement are introduced as extensions of the above techniques and they are also distinguished from general association statistics.

Chapter seven is devoted to modelling the probability of discrete educational events in a given time period during the life history of individuals. Occurrences of preschool events (such as nursery or playgroups) and their transitions to and from home are analyzed as a function of explanatory variables such as mother´s educational level or child´s gender. Use of this approach has not been very common up to now; the field is however expanding with the increasing availability of longitudinal data sets. Some elements of multilevel event-history analysis are outlined.

Chapter eight addresses the way in which measurement error can affect model estimates, an important topic that is seldom considered in the practice of educational statistics. Some of the problems encountered with both continuous and categorical data are discussed, along with general correction strategies. Finally, several extensions of multilevel modelling of practical use in education are reviewed: a) Cross classified situations, which occur when a child belongs to more than one higher level unit, such as two different schools; b) Multivariate models, where more than one performance variable is measured and it may be of interest to distinguish within unit (classroom, as an example) from between unit correlation; c) categorical data, whose analysis can result in biased error terms and misleading goodness of fit statistics if the multilevel structure of the data is not observed. Population averaged, and subject specific methods are presented here as two useful general approaches to estimate fixed and random effects in this situation.

The examples data sets discussed in the book are included in a diskette. Several exercises are given at the end of each chapter, along with questions on interpretation of MLn output and some sample code in GLIM and SAS. Answers to some of them are provided at the end of the book. Since the solution of these exercises involves some model fitting, I think interested readers would benefit from an appendix that included basic inference and the notion of deviance in generalized linear models. From the viewpoint of a novice trying to learn multilevel concepts, it would also be good in my opinion to make more explicit links to other standard statistical terms that may be more familiar to the reader. One of them is the notion of nested or hierarchical models (Kirk, 1995; Myers & Well, 1995). Milliken and Johnson (1994) distinguish between nesting in the treatment structure and nesting in the design structure in an experimental setting. In a standard analysis of variance, there is a different error term for each level of experimental units in the design structure, but that does not happen if the nesting occurs in the treatment structure. It appears to me that multilevel models address the first case, in which there are different levels of experimental units. It is possible to have a two-fold fixed effects nested model, with no multilevel structure. In that case, a standard nested ANOVA would apply. Even if the multilevel framework allows for modelling both situations, these would represent two different types of hierarchical models. Other links can be drawn by making its relationship with the analysis of covariance more explicit (Longford, 1993); and also with the identification of simplifying conditions (if there were any besides the intra-unit correlation being equal to zero) under which other standard methods could be used, as happens in the case of some balanced mixed models. In my opinion, such considerations are instructive exercises and may help the reader to build on previous knowledge and broaden his/her appreciation of the general (or generalized) linear model.

Overall, Ian Plewis´s book is a well-documented, very readable book that identifies a broad range

of methodological problems and solution strategies for educational and social researchers. It serves its purpose of reviewing recent advances in the field well, and it is therefore recommended to anybody that wants to update his/her knowledge and analysis skills with newly available statistical techniques. ( )

References

Freedman, D., Pisani R.,.Purves, R. & Adhikari, A. (1991) Statistics (2nd. Ed.). New York: Norton.

Kirk, R. E. (1995). Experimental Design: Procedures for the behavioral sciences. London: Brooks-Cole.

Longford, N. T. (1993) Random coefficient models. Oxford UK: Oxford University Press.

Milliken, G. A. & Johnson, D. E. (1994). Analysis of Messy Data: Vol 1. Designed Experiments. London: Chapman Hall.

Myers, J.L. & Well, A. D. (1995). Research design and statistical analysis. Hilldale NJ: Lawrence Erlbaum.

Numerical Integration via High-Order, Multivariate LaPlace Approximation with Application to Multilevel Models

Stephen W. Raudenbush, University of Michigan
Meng-Li Yang, Nan-Tai Institute of Technology, Taiwan

To obtain maximum likelihood estimates (MLE) in multilevel analysis often requires the evaluation of difficult integrals. In a two-level setting, the model will have a hierarchical structure such that the level-1 outcome, y, given the random effects, b, has some distribution $L(y|b, \theta)$. At level 2, the random effects themselves are assumed distributed according to $p(b|\theta)$ where θ denotes a parameter vector. The likelihood of observing y is then

$$L(y|\theta) = \int L(y|b,\theta)\, p(b|\theta)db. \qquad (1)$$

Thus, the random effects, b, are "integrated out" to yield the marginal density (or "likelihood") of y given θ. Given y, the MLE is the value of θ that maximizes this likelihood.

Evaluation of the integral in Equation 1 is sometimes easy. For example, if both $L(y|b, \theta)$ and $p(b|\theta)$ are normal densities, the integral, which will also be a normal density, can be evaluated analytically. The likelihood is then maximized by standard algorithms such as EM (Dempster, Laird, and Rubin, 1977) or Fisher scoring via iterative generalized least squares (c.f. Goldstein, 1995). In fact, whenever the

distribution of b is conjugate to that of y, integration can be comparatively simple.[1] Conjugate pairs include not only the normal-normal, but also the Poisson-gamma, binomial-beta, and gamma-inverse gamma, for example.

Often, however, non-conjugate pairs are of great interest. For example, in multilevel logistic regression, the outcome y_{ij} for student i in school j is Bernoulli, taking on a value of 1 with probability μ_{ij}, where

$$\log[\mu_{ij}/(1-\mu_{ij})] = \eta_{ij} = X_{ij}^T \beta + Z_{ij}^T b_j.$$

A flexible model for the random effects b_j is the multivariate normal, $b_j \sim N(0,D)$ where D is the variance-covariance matrix of the random effects. The integrand of Equation 1 is thus a Bernoulli-normal product, where θ includes the elements of (β,D). In this case, analytic evaluation of Equation 1 is impossible. Other

---

[1] This assumes that a link function is carefully chosen (McCullagh and Nelder, 1989; Lee and Nelder, 1996).

interesting non-conjugate cases include the Poisson-normal for count data and the exponential-normal for duration-time data.

Researchers have proposed a number of approximations to Equation 1 when analytic evaluation is impossible. These include approximate posterior modal estimation (Stiratelli, Laird, and Ware, 1984; Wong and Mason, 1985) and, closely related, marginal quasi-likelihood or MQL (Goldstein, 1991; Longford, 1993) and penalized quasi-likelihood (Breslow and Clayton, 1993). These "first-order" approaches work well when the diagonals of D are small but produce non-trivial negative asymptotic bias for large variances (Rodriguez and Goldman,1995; Lin and Breslow, 1995). A second-order correction greatly reduces the bias (Goldstein and Rasbash, 1996).

Hedeker and Gibbons (1994; see also Gibbons and Hedeker, 1997) proposed evaluation of Equation 1 via Gauss-Hermite quadrature. The beauty of this approach is that, by adding quadrature points, the approximation becomes as accurate as desired. The key difficulty is that computations can be burdensome, especially for very large data sets and multi-dimensional random effects.

Based on the work of Raudenbush and Yang (1998) and Yang (1998) we now consider an alternative approach, which we shall call "LaPlace6." Our experience to date indicates that LaPlace 6 is computationally fast and produces results equivalent to those based on 20-40

quadrature points, generally regarded as a large number. The approach has three steps.

1. Expand the integrand of Equation 1 in an infinite Taylor series. Defining $f(b) = L(y|b,\theta)p(b|\theta)$, and assuming *f(b)* and all its partial derivatives are continuous in a neighborhood about b, we have

$$f(b) = f(\tilde{b}) + f^{(1)}(\tilde{b})(b - \tilde{b}) + \sum_{k=2}^{\infty} \frac{1}{k!}[\otimes(b - \tilde{b})^T]f^{(k)}(\tilde{b})(b - \tilde{b}) \qquad (2)$$

where $f^{(k)}(\tilde{b}) = \dfrac{\partial vec\, f^{(k-1)}}{\partial b^T}\Big|_{b=\tilde{b}}$, and $\overset{k}{\underset{1}{\otimes}}u = u \otimes u \otimes \ldots \otimes u$ there being k u's in the Kronecker product. For example, $\overset{3}{\underset{1}{\otimes}}u = u \otimes u \otimes u$.

2. Apply LaPlace's method. Suppose f(b) has a maximum at $b = \tilde{b}$. Then

$$L = \int \exp[f(b)]db = (2\pi)^{1/2}|-f^{(2)}(\tilde{b})|^{-1/2}\exp[f(\tilde{b})] \times E\{\exp[\overset{\infty}{\underset{k=3}{\sum}}T_k]\}$$

with $T_k = \dfrac{1}{k!}([\overset{k-1}{\underset{1}{\otimes}}(b - \tilde{b})^T]f^{(k)}(\tilde{b})(b - \tilde{b}))$ \qquad (3)

The proof appears in Raudenbush and Yang (1998).

3. Maximize (3) with respect to θ using approximate Fisher scoring (Green, 1984), which requires only first-order derivatives.

In practice, it is not possible to evaluate the infinite series of Equation 3. Our experience with a variety of data suggests that an approximation

$$E[\exp\{\overset{\infty}{\underset{k=3}{\sum}}T_k\}] \approx 1 + E(T_4) + E(T_6) + E(T_3^2)/2 \qquad (4)$$

is adequate. In the two-level logistic case, we thus have

$$L \approx \prod_{J=1}^{J} |D|^{-1/2} \, |Z_j^T W_j Z_j + D^{-1}|^{-1/2} \, \exp\{\tilde{l}_j - \tilde{b}_j^1 D^{-1} \tilde{b}_j / 2\} \times$$

$$[1 + E(T_{4j}) + E(T_{6j}) + (1/2)E(T_{3j}^2)]$$

(5)

where Z_j contains rows Z_{ij}^T, W_j is diagonal with elements $\mu_{ij}(1-\mu_{ij})$ and computational formulas for T_{4j}, T_{6j} and T_{3j}^2 are provided by Raudenbush and Yang (1998). Making these computations tractable is Yang's (1998) compact representation of all moments of the multivariate normal distribution.

We simulated 100 data sets and compared its results ("Laplace6") with those from PQL and Gaussian quadrature with 10 or 20 quadrature points

("Gauss 10") and "Gauss 20"). The model produced data with asymmetric probabilities having an average conditional expectation equal to 0.14. The structure of the data sets follows Rodriguez and Goldman (1995). These data sets involve 16 hypothetical children nested within each of 161 hypothetical communities, with 2576 children overall. The model reflects the belief that the outcome y_{ij} depends on a child-level covariate ("childcov") and a community-level covariate ("commucov"), and that intercepts and slopes varying across "communities". Thus, we have a model for person i in group j,

$$\eta_{ij} = \log[\mu_{ij}/(1-\mu_{ij})] = \beta_{00} + \beta_{01}(commu\,cov)_j + \beta_{10}(child\,cov)_{ij} + b_{0j} + b_{1j}(child\,cov)_{ij}$$

with b_{0j} and b_{1j} together forming a bivariate normal distribution with means 0, variances D_{00},

D_{11} respectively, and covariance D_{01}, where D_{00} = 1.625, D_{11} = 0.25, and D_{01} = 0.1.

Table 1 Averages of the Estimates for Simulated Data

	PQL	Gauss-10	Gauss-20	Laplace6
D_{00} =1.625	1.2752	1.6532	1.6546	1.6352
D_{01} =.1	.0538	.1003	.0995	.0960
D_{11} =.25	.1614	.2575	.2562	.2667
β_{00} =-1.2	-1.0904	-1.1977	-1.2045	-1.2007
β_{01} =1	.9004	1.0081	1.0148	1.0029
β_{10} =1	.9114	0.9971	.9976	.9975

Table 2 Mean Squared Errors for Simulated Data

	PQL	Gauss-10	Gauss-20	Laplace6
D_{00} =1.625	.1522	.0737	.0633	.0563
D_{01} =.1	.0080	.0115	.0120	.0108
D_{11} =.25	.0113	.0073	.0072	.0075
β_{00} =-1.2	.0271	.0231	.0196	.0190
β_{01} =1	.0236	.0193	.0175	.0164
β_{10} =1	.0116	.0051	.0053	.0051

Tables 1 and 2 present averages and mean squared errors of the estimates across 100 replications. As expected, PQL estimates are negatively biased for all parameters. The underestimation of the variance components range from 22% to 46%, while that of the β's are around 9%. The other methods greatly

reduce the bias. Mean squared errors tend to be a bit smaller under Laplace6 than under Gauss-10 or Gauss-20 in these results. In particular, when Gauss-20 appears non-trivially better than Gauss-10, Laplace replicates the advantage or possibly improves upon it.

In summary, for bivariate random effects, Laplace6 seemed to do as well as Gauss using 10 or 20 quadrature points. To assess computational efficiency, we timed analyses on 6 randomly selected data sets using Pentium233, and found the analysis took on average 10 seconds per data set using Laplace6. These analyses were found to take, on average, 3 minutes on Gauss-10 and 12 minutes on Gauss-20.

We conclude that approximation of difficult integrals via high-order, multivariate Laplace approximation appears to be a promising strategy for evaluating likelihoods in hierarchical models. The approach can be extended to the degree of accuracy required, and can be written for multivariate integration of arbitrary dimension. Of course, much more research is needed: a) on the behavior of this approach in binomial-normal models; b) in other hierarchical generalized models; and c) in a broader class of hierarchical models.

References

Breslow, N., & Clayton, D. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association, 88,* 9-25.

Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*(39), 1-8.

Gibbons, R. D., & Hedeker, D. (1997). Random effects probit and logistic regression models for three-level data. *Biometrics, 53,* 1527-1537.

Goldstein, H. (1991). Nonlinear multilevel models with an application to discrete response data. *Biometrika, 78,* 45-51.

Goldstein, H. (1995). *Multilevel Statistical Models, Second Edition.* New York: John Wiley & Sons.

Goldstein, H., & Rasbash. (1996). Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society, 159*(part 3), 505-513.

Green, P. (1984). Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *Journal of the Royal Statistical Society, B, 46,* 149-192.

Hedeker, D., & Gibbons, R. (1994). A random-effects ordinal regression model for multilevel analysis. *Biometrics,* pp. 993-944.

Lee, Y., & Nelder, J. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society, Series B, 58,* 619-678.

Lin, X., & Breslow, N. (1995). Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of American Statistical Association.*

Longford, N. (1993). *Random Coefficient Models.* Oxford: Clarendon Press.

McCullagh, P., & Nelder, J. (1989). *Generalized Linear Models, 2nd Edition.* London: Chapman and Hill.

Raudenbush, S., & Yang, M.. (1998). *Maximum likelihood for hierarchical models via high-order, multivariate LaPlace approximation.* Paper submitted to the Journal of Computational and Graphical Statistics.

Rodriguez, G., & Goldman, N. (1995). An assessment of estimation procedures for multilevel models with binary responses. *Journal of the Royal Statistical Society, A, 158,* 73-89.

Stiratelli, R., Laird, N., & Ware, J. (1984). Random effects models for serial observations with binary response. *Biometrics, 40,* 961-971.

Wong, G., & Mason, W. (1985). The hierarchical logistic regression model for multilevel analysis. *Journal of the American Statistical Association, 80*(391), 513-524.

Yang, M. (1998). Increasing the efficiency in estimating multilevel Bernoulli models [Diss], East Lansing, MI: Michigan State University.

Applications of Random Effects/Multilevel Models to Categorical Data in Social Sciences and Medicine

A one-day conference with the above title was held at the Royal Statistical Society headquarters on 20 October 1998. It was a joint venture between the Social and Medical Statistics sections of the RSS, the ESRC programme on the Analysis of Large and Complex Datasets (ALCD), and Statistics in Society' (JRSS(A)). Twelve papers were presented in four sessions, and a range of disciplines was covered in each session. One of the aims of the conference was to explore commonalities of application, problems and nomenclature across disciplines as diverse as criminology, education, epidemiology and market research. Bayesian and likelihood-based methods, clinical trials and observational studies were all represented. It is hoped that a substantial proportion of the papers, along with commentaries from discussants of each session and an introductory editorial will appear as a special issue of 'Statistics in Society' in 1999.

Abstracts of all the papers presented at the conference can be found at **http://www.ioe.ac.uk/ms/conf.html** and full versions of some of them can be accessed from this URL. (*Ian Plewis, Institute of Education, University of London*)
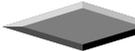
From October 1998, William Browne has joined the Multilevel Models Project as a new member of the team after his PhD study in Bath University of UK. His speciality is MCMC modelling. Bill has programmed the MCMC procedures in MLwiN V1.0, and will be working on fitting a wider range of models using the MCMC techniques. He can be contacted by email **w.browne@ioe.ac.uk** *or telephone +44 (0)171 612 6680.*
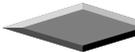
Some Publications in 1997/98 Using Multilevel Models

Cohen, M. P. (1998), Determining sample sizes for surveys with data analyzed by hierarchical linear models. *Journal of Official Statistics* 14:3, 267-276.

Courgeau, D. and Baccaini, B. (1998). Multilevel analysis in the social sciences. *Population* **10**: 39-71.

Deely, J. J. and Smith, A. F. M. (1998). Quantitative refinements for comparisons of institutional performance. *Journal of the Royal Statistical Society, A.* **161**: 5-12.

Diggle, P. J. and Verbyla, A. P. (1998). Nonparametric estimation of covariance structure in longitudinal data. *Biometrics* **54**: 401-15.

Goldstein, H. and Blatchford, P. (1998). Class size and educational achievement: a review of methodology with particular reference to study design. *British Educational Research Journal* **24**: 255-268.

Hedeker D. & Gibbons R.D. 1997, Application of random-effect. Pattern-mixture models for missing data in longitudinal studies, *Psychological Methods*, Vol. 2, No. 1 64-78.

Hill, P. W. and Goldstein, H. (1998). Multilevel modelling of educational data with cross classification and missing identification of units. *Journal of Educational and Behavioural statistics* **23**: 117-128.

Hox, J. 1997, Multilevel Modeling: When and Why, paper presented to the *21st Annual Conference of the Gesellschaft fuer Klassifikation e.V., Potsdam, FRG,*

Huggins, R. M. and Loesch, D. Z. (1998). On the analysis of mixed longitudinal growth data. *Biometrics* **54**: 583-95.

Kasim, R. M. and Raudenbush, S. W. (1998). Application of Gibbs sampling to nested variance components with heterogeneous within-group variance. *Journal of Educational and behavioural Statistics* **23**: 93-116.

Langford, I.H. & Lewis, T. 1998, Outliers in multilevel data, *J.R.Statist. Soc.* **A**, 161:2, 121-160.

Langford, I. H., Bentham, G. and McDonald, A. (1998). Multilevel modelling of geographically aggregated health data: a case study on malignant melanoma mortality and UV exposure in the European community. *Statistics in Medicine* **17**: 41-58.

Lesaffre, E. and verbeke, G. (1998). Local influence in linear mixed models. *Biometrics* **54**: 570-82.

Maritz, J. S., Lombard, C. J. and Morrell, C. H. (1998). Exact group comparisons using irregular longitudinal data. *Journal of the Royal Statistical Society, Series C* **47**: 351-60.

Meng, X. and Dyk, D. (1998). Fast EM implementations for mixed effects models. *Journal of the Royal Statistical Society, B* **60**: 559-78.

Moerbeek, M., Breuelen, G. and Berger, M. (1998). *Design issues for experiments in multilevel populations*. Maastricht, Maastricht university.

O'muircheartaigh, C. and Campanelli, P. (1998). The relative impact of interviewer effects and sample design effects on survey precision. *Journal of the Royal Statistical Society, A.* **161**: 63-78.

Pan, H.Q. & Goldstein, H. 1998, Multi-level repeated measures growth modelling using extended spline functions, *Statistics in Medicine*, 17 (forecoming).

Plewis, I. and Hurry, J. (1998). A multilevel perspective on the design and analysis of intervention studies. *Educational research and evaluation* **4**: 13-26.

Saei, A. and McGilchrist, C. A. (1998). Longitudinal threshold models with random components. *Journal of the Royal Statistical Society, D* **47**: 365-75.

Ten have, T. R., Kunselman, A. R., Pulkstenis, E. P. and Landis, J. R. (1998). Mixed effects logistic regression models for longitudinal binary response data with informative dropout. *Biometrics* **54**: 367-83.

Troxel, A. B., Harrington, D. P. and Lipsitz, S. R. (1998). Analysis of longitudinal data with non-ignorable non-monotone missing values. *Journal of the Royal Statistical Society, C,* **47**: 425-38.

Wang, N., Lin, X., Gutierrez, G. and carroll, R. J. (1998). Bias analysis and SIMEX approach in generalised linear mixed measurement error models. *Journal of the American Statistical Association* **93**: 249-61.

Wright, D.B. 1998. Modelling clustered data in autobiographical memory research: the multilevel approach*, Applied Cognitive Psychology*, vol. 12, 339-357.

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

Contributors

We are grateful to the following who provided articles and news for this issue.

Joop Hox, University of Amsterdam, Faculty of Educational Sciences: **hox@educ.uva.nl**

Juan Oliver-Rodriguez, Universitat Jaume I, Depto. De Psicologia: **oliverr@psi.uji.es**

Ian Plewis, Institute of Education, University of London: **I.plewis@ioe.ac.uk**

Stephen Raudenbush, University of Michigan: **rauden@umich.edu**

Judith Singer, Harvard University, Graduate School of Education: **judith_singer@harvard.edu**

Editorial Questions

The Multilevel Models Project is considering ways of improving dissemination of multilevel analysis, and the future of the newsletter. Your opinion on the following questions would be greatly appreciated. Please send your answers by email to the Editor (**m.yang@ioe.ac.uk**) before Christmas.

Q1: What overall rating would you give to the M.M. Newsletter for its usefulness?

No used at all	1	2	3	4	5	6	7	8	9	Very useful

Q2: Ranking the following is order of usefulness.

Courses/Clinics	Book review	References	Theoretical papers	Applications	Project news

Q3: In what form do you prefer to receive the newsletter?

1. Newsletter by post	2. Newsletter by email attachment	3. Web download

Q4: Would you consider submitting to the Newsletter?

Very dislikely	1	2	3	4	5	6	Very likely

Q5: How could the MM Newsletter be improved?
Please let us have your views.