MULTILEVEL MODELLING NEWSLETTER

Centre for Multilevel Modelling

Bedford Group for Lifecourse and Statistical Studies
Institute of Education, University of London
20 Bedford Way, London WC1H 0AL, ENGLAND
Web site: http://www.mlwin.com/
Enquiries about newsletter to Ian Plewis
E-mail: i.plewis@ioe.ac.uk
Tel: +44 (0) 20 7612 6238
Fax: +44 (0) 20 7612 6572

Vol. 17 No. 1 **September, 2005**

Editorial

This is the last issue of the newsletter to be edited by me. I have done the job for the last six years but, to coincide with the move of the Centre for Multilevel Modelling from the Institute of Education in London to Bristol (see Jon Rasbash's article in this issue), I am handing over the reins to Harvey Goldstein. I would like to thank all of you who have contributed articles, news and reviews and also to encourage everyone to continue to think about the newsletter as an outlet for material on multilevel modelling whose importance continues to grow, not only as a set of statistical techniques but also as a valuable way of thinking about many aspects of the world.

Ian Plewis

Fifth International Amsterdam Conference

The Fifth International Amsterdam Conference on Multilevel Analysis was held in Amsterdam on 21-22 March 2005. The following papers were presented:

Carlos A. Q. Coimbra and Tom A. B. Snijders
Estimation of Non-Linear Models by Stochastic Approximation

G. W. Jacobusse, S. Van Buuren and C. G. M. Groothuis-Oudshoorn
Multiple Imputation of Missing Data in a Multilevel Setting

Tom A.B. Snijders
MLwiN Macros for Nonlinear Transformations of Independent Variables

S. Teerenstra, R. J. F. Melis, P. G. M. Peer and G.F. Borm
Pseudo Cluster Randomization

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

Learning Environment for Multilevel Methodology and Applications (LEMMA)

Jon Rasbash
University of Bristol
j.rasbash@bristol.ac.uk

The Centre for Multilevel Modelling is moving from the Institute of Education in London to the University of Bristol. The centre staff, with other Bristol academics, were successful in an application for a new research project called LEMMA. The LEMMA project is one of a set of six nodes funded under the ESRC's National Centre for Research Methods (NCRM). The brief of the NCRM (www.ncrm.ac.uk) is to provide a step change in the quality of social science research in the UK. This article describes the LEMMA project.

The project has three inter-related elements: (i) statistical methodology; (ii) flagship substantive research projects; (iii) sets of materials and systems for training and capacity building.

Methodology

The usual distributional assumption of Normality for higher level random effects can be overly restrictive. To give two examples:

(1) Nagin (1999) describes a formulation for growth curves where a discrete set of latent groups is posited and each individual has a membership distribution across the groups. Muthén (2004) has also implemented models in this area.

(2) In binary response and event history models, many higher level units have response patterns of all zero or all one and this leads to the class of mover/stayer models, which also utilize latent categories. Such models have been implemented in SABRE (http://www.cas.lancs.ac.uk/).

The existing work in the area of latent categorical effects in multilevel models has focused on fitting latent categorical distributions to hierarchical models. (Rabe-Hesketh et al., 2004, Chapter 5; Vermunt and Magidson, 2005). The LEMMA project will build on existing work so that latent categorical distributions can be fitted to any level in multilevel models which can contain mixtures of nested, crossed and multiple membership classifications.

Flagship substantive research projects

A number of projects are planned that will demonstrate how multilevel models can be applied to substantive social science problems.

Geography of school effects

This project addresses the relationship between school effectiveness, school choice and parental relocation, thereby addressing current debates about the effects of quasi-markets in education. This will be tackled by using Pupil Level Annual School Census (PLASC) data and also possibly the exceptionally detailed Avon Longitudinal Study of Parents and Children (ALSPAC) data. These data have a highly complex structure including multiple membership and crossed classifications; repeated measures on individuals within areas; movement of individuals between areas and schools; repeated measures on individuals within primary year cohort within primary school; and repeated measures within secondary year cohort within schools. Spatial models will also be used to model 'competition' between the higher level units, such as schools with overlapping catchments, differentiated by school type.

Modelling the duration of episodes in hospital

The effects of consultants, hospitals and geographical areas on the length of time a patient stays in hospital will be explored using data from the Hospital Episodes Statistics dataset. These data have a highly complex non-hierarchical

structure. A patient might have more than one stay in hospital, leading to repeated episodes nested within individuals. Episodes and consultants have a multiple-membership structure since a patient might see more than one consultant during an episode and different consultants between episodes. In addition, consultants are crossed by hospitals, as consultants can work at more than one hospital, and patients are nested within a cross-classification of hospitals and geographical areas.

Voting choice

The substantive issue focuses on the individual, household and neighbourhood determinants of voting abstention and party choice. Using the British Household Panel Study (BHPS) we have repeated binary measures on voting intention for individuals within households within areas at a variety of scales. Normally distributed individual level random effects are unrealistic as part of the mover/stayer problem as are Normal household effects. We will compare the autocorrelation approach of Goldstein and Barbosa (2000) and the discrete latent-effects model, in particular a doubly-nested model, with latent classifications at the individual and household level.

Mental health and psychosocial development

This uses measures on mental well-being from two data sets. The first is the BHPS and will compare the use of Normal random effects and discrete latent effects to describe between individual variations in patterns of change over time with a multiple-membership model to take account of changes in household composition. The second dataset is the Avon Brothers and Sisters Study with repeated measures on psychosocial adjustment, for multiple children within families. Again we will explore the use of continuous and discrete latent effects to describe between individual patterns of variation in psycho-social development.

Modelling group diversity

Traditional statistical models have concentrated on modelling mean effects as functions of predictor variables. Multilevel models allow us to model the variation for any classification as a function of further variables. For example Goldstein and Noden (2003) model the between school and between LEA variation in the percentage of pupils eligible for free school meals as functions of LEA characteristics. Such models, applied for example to measures of poverty or service delivery, are highly relevant to current debates about diversity since they avoid certain arbitrary features of traditional index measures, and provide efficient and objective estimates of between-unit variation. A further development is to construct models that use such estimates of diversity, for example estimated for each unit in a classification, as predictors in a further model where outcomes are a function of area level measures of diversity.

Training and capacity building

The resources we plan to develop will be useful for solo and group learning.

The resources are aimed at social scientists with a wide spectrum of statistical expertise.

Our planned training materials will be designed to give users the necessary skill to carry out quantitative research on data with complex structure. They will be

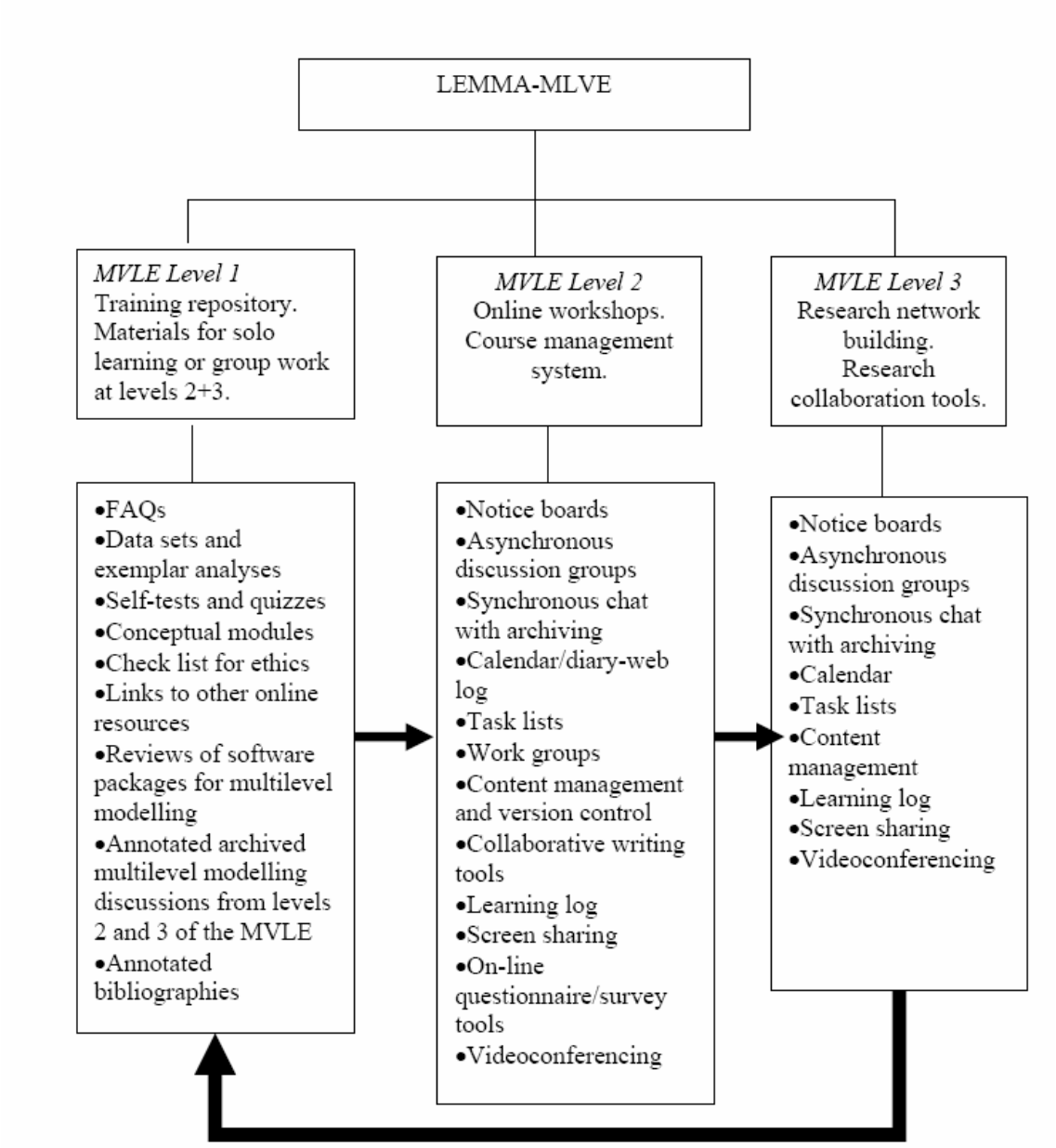
- Carefully graded
- Model-based
- Realistic
- Authentic
- Contextualised

The project will be running a mixture of conventional face-to-face training workshops and clinics. These modes of training have limitations in that workshops are over-subscribed and there are insufficient resources to provide follow-up support. This lack of follow up support often prevents researchers from converting the methodological insights and practical skills gained at a workshop into routine use of these skills in their day to day research work. The LEMMA project aims to address these training shortfalls with a web based learning environment which will support solo and group learning and provide support to try and foster self-supporting sustainable groups of researchers.

Web-based training

A substantial and ambitious new venture that builds on our web experience will be the provision of a range on online resources, including a repository of training materials, a series of collaborative and moderated online workshops, as well as a number of research networks which together form a multilevel modelling virtual learning environment (MVLE) designed to initiate, develop, and support dispersed researchers.

The anticipated architecture of the Information and Communications infrastructure that will be adopted as the MVLE is sketched out in the diagram below:



The proposed ICT architecture as shown maps on to the pedagogical design for the whole system. This anticipates three levels of activity, which feed into each other:

Level 1
Repository of Training Materials
This is essentially a database of teaching and learning materials which

can be used by solo learners or tutors for delivering their own training.

Level 2
Online Workshops
Moderators use materials provided in level one to facilitate group learning in online courses. These are designed to promote online group formation which will be carried through into level three after the end of formal teaching.

Level 3
Online Research Communities
This forms the core of collaborative knowledge building in intensely focused groups. Computer-supported, collaborative work-tools will be used to help form and sustain online networks. It is anticipated that some of the outputs of these networked activities will result in learning objects being deposited in MVLE level one such as new exemplars and annotated archived discussions.

We regard the feedback of new knowledge from level three back into levels one and two of the MVLE as our mechanism for the spread of knowledge, concepts and new practices into the wider research community. The above learning architecture is no more than a representation of what the applicants envisage as the final pedagogical structure as every level of the MVLE will go through an iterative and participatory process of design, delivery, evaluation and re-design.

The first release of the LEMMA MVLE will be in 2006. Following the link to LEMMA on the National Centre for Research Methods site (www.ncrm.ac.uk) will provide progress reports on the LEMMA node's activities.

References

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

Jones, K (2005). An introduction to statistical modeling. In Somekh, B. and Lewin, B. (eds). *Research Methods in the Social Sciences*, Sage.

Goldstein, H., and Noden, P. (2003) Modelling social segregation. *Oxford Review of Education*, **29**: 225-237.

Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (ed.), *Handbook of Quantitative Methodology for the Social Sciences*.

Nagin, D. (1999) Analyzing developmental trajectories: A semi-parametric, group-based approach. *Psychological Methods*, **4**: 139-177.

Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2004). GLLAMM Manual (http://www.bepress.com/ucbbiostat/paper160/). U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 160.

Vermunt, J.K, and Magidson, J. (2005). Hierarchical mixture models for nested data structures. In C. Weihs and W. Gaul (eds), *Classification: The Ubiquitous Challenge*. Heidelberg: Springer.