MULTILEVEL MODELLING NEWSLETTER

The Multilevel Models Project
Mathematical Sciences
Institute of Education, University of London
20 Bedford Way, London WC1H 0AL, ENGLAND
E-mail: temsmya@ioe.ac.uk
Web site http://www.ioe.ac.uk/multilevel/
Tel: +44 (0)171 612 6682 Fax: +44 (0)171 612 6686

Vol. 9 No. 1 **June, 1997**

Centre in Hong Kong: As from May 1 1997 it has been agreed to set up a Multilevel Models in Education Center (MMEC) at the Department of Education, University of Hong Kong. The aims of the Centre are:

(1) to facilitate the cooperation of international networks in research on multilevel analysis;

(2) to enhance international comparisons between educational systems using multilevel analysis;

(3) to disseminate knowledge of multilevel analysis in Asia and the Pacific rim region;

(4) to negociate and facilitate scholastic exchanges of staff and researchers between the London University Institute of Education and MMEC in the Department of Education, University of Hong Kong.

Raymond Lam and *Wong Kam-Cheung* are the co-directors of the centre. Anyone with queries regarding the work of the centre may contact: *KC Wong*, Department of Education, University of Hong Kong, Pokfulam Road, Hong Kong. Tel: *(852) 28592518*, Fax: *(852) 25406360*, Email: kcwong@hkucc.hku.hk

Amsterdam Conference on Multilevel Analysis: On April 1-2, 1997, an international conference about multilevel analysis was held in Amsterdam. The program was organised *by Tom Snijders* (ICS, University of Groningen; Multilevel research Group, NOSMO). This conference was one of the Social Science Information Technology workshops of ProGAMMA. There were about 60 participants from 10 countries. The conference was followed by a 1-day course taught by *Donald Hedeker* (University of Illinois at Chicago) about Multilevel Analysis of Categorical Outcomes. The software discussed in this course can be downloaded from his internet site http://www.uic.edu/~hedeker/mix.html. or the Multilevel Models Project site shown above. A list of invited speakers and contributed papers is presented on page 15.

MLn Clinics in London 1997

Tuesday June 3
Tuesday July 8
Tuesday August 5
Tuesday September 9
Tuesday October 7
at
Multilevel Models Project
11 Woburn Square, London WC1A 0SN
Contact Min Yang for appointment
Tel: (0)171 612 6682
Email: temsmya@ioe.ac.uk

Also In This Issue

Terminology and Definition in Multilevel Models Analysis

Ian Plewis
Institute of Education, University of London

INTRODUCTION

For some time now, there have been questions about the place of multilevel modelling within statistical modelling more generally. There has also been a spate of papers and books over the last few years, especially in biostatistics, on the specification and estimation of models with random effects. To the outsider, and to anyone wanting to know more about statistical modelling of this kind, the ways in which different terms are used to describe essentially the same thing, and, perhaps worse, the ways in which the same words are sometimes used to describe different things, is surely confusing.

The purpose of this note is to set out a list of terms used in this area of applied statistics, with just a little commentary. It is probably incomplete, and will no doubt generate some disagreement and dispute. My hope is that it will be seen at least to start to clarify some issues, and that subscribers to the Newsletter, and others, will fill in gaps, add to it and suggest amendments. These could eventually be put together to produce a more rounded and comprehensive document, more a dictionary and thesaurus than a list, useful to students and researchers and also, most importantly, to non statisticians who are using, or who want to use, these powerful methods of statistical analysis.

The organisation of the material is, at present, as follows. There are two main sections: (A) vocabulary, and (B) modelling. Section A has three sub-sections: structures, models, and heterogeneity. Section B also has three sub-sections: model specification, estimation method, and software. This division could no doubt be improved.

A. VOCABULARY

A.1 Structures

A.1.1 Populations

a) Hierarchical
b) Nested
c) Cross-classified
d) Multilevel

Comment

a) and b) are interchangeable;
d) incorporates each of a), b) and c).

A.1.2 Designs

a) Repeated measures
b) Split and split-split plot
c) Cluster randomisation
d) Crossover
e) Complex/cluster/multi-stage samples

A.2 Models

a) Generalised linear mixed (GLMM)
b) Hierarchical generalized linear (HGLM)
c) Hierarchical linear (HLM)
d) Marginal
e) Mixed linear
f) Mixture
g) Multilevel (MLM)
h) Population average (PAM)
i) Random coefficients
j) Random effects
k) Random intercept
l) Slopes as outcomes
m) Subject/unit specific
n) Variance (and covariance) components
o) Varying coefficient

Comment

a), b), c), e), g), i), l) and o) are all more general than k) and n).

d) and h) are interchangeable. However, the use of d) in this context is confusing (see B.2) and should, I believe, be dropped.

A contrast is often drawn between d) and h) (and also f)) with g) and m) (see B.1).

A.3 Heterogeneity

This is often generated by A.1

a) Extra variation
b) Frailty (and, in demography, fecundity)
c) Over (and under) dispersion
d) Self selection (in non-randomised studies)
e) Unobserved heterogeneity

Comment

a) and b) are interchangeable;
e) and b) are essentially the same.

B. MODELLING

B.1 Model Specification

Here the main issues are:

a) the distinction between models for which only the fixed effects are of interest and the random effects are nuisances (A.2; d), h) and f)), and where both the fixed and the random effects are of interest, and hence the variability of the latter is also modelled.

b) whether or not the distributions of the random effects are assumed to be Normal or, more generally, multivariate Normal.

B.2 Estimation Method

There is a burgeoning variety of approaches, particularly for the more difficult situation of a non-Normal response. I give one recent reference for each.

a) Best linear unbiased prediction (BLUP) - McGilchrist (1994) - allows a general distribution for the response but the random effects are assumed to be Normal. This has been modified by Kuk (1995) to allow for non-Normal random effects.

b) Conditional likelihood - McCullagh and Nelder (1989) - takes out the random intercept by conditioning on it to give estimates with a subject specific interpretation.

c) Estimating a scale factor from the residual deviance to adjust the standard errors of the fixed effects (Francis et al., 1994).

d) Generalized estimating equations (GEE) - Diggle, Liang and Zeger (1994) - for population average models.

e) Hierarchical (h-) likelihood - Lee and Nelder (1996) - for non-Normal models with conjugate distributions of the random effects. The random effects are assumed to be independent.

f) Iterative (and restricted iterative) generalised least squares (IGLS/RIGLS) - Goldstein (1995).

g) Marginal maximum likelihood (MML) - Hedeker and Gibbons (1994) - Normal random effects estimated using numerical integration.

h) Marginal quasi likelihood (MQL) - Goldstein (1995) - Taylor series expansion (first and second order/linear and quadratic) of a non-linear function with Normal random effects.

i) Maximum (or restricted maximum) likelihood (ML/REML) with Fisher scoring - Longford (1993).

j) Maximum (or restricted maximum) likelihood (ML/REML) with EM algorithm - Bryk and Raudenbush (1992).

k) Mixture methods - Lindsey (1995) - e.g. Poisson response plus gamma random intercept gives negative binomial population average model.

l) Monte Carlo Markov chain (MCMC)/Bayes with Gibbs sampling - Best et al. (1996). m) Non parametric estimation (Aitkin, 1996) - no assumptions about the distribution of the random effects.

n) Penalized (or predictive) quasi likelihood (PQL) - Breslow and Clayton (1993) - as MQL but including the estimated random effects at each iteration; second order PQL - Goldstein (1995).

Comment

f), i) and j) are equivalent for Normal responses.

B.3 Software

B.3.1 Specialist packages

BUGS - MCMC/Gibbs sampling

EGRET - conditional likelihood and mixture models.

HLM - general multilevel modelling using EM.

MIXOR - ordered categorical data using MML.

MIXREG - continuous responses

MIXGSUR - for grouped time survival data

MLn - general multilevel modelling using IGLS/RIGLS/MQL/PQL.

PCCARP - for sampling errors from complex designs.

REML - for variance and covariance components.

OSWALD - S+ macros for GEEs

SABRE - for repeated binary outcomes

SUDAAN - for sampling errors from complex designs

VARCL - general multilevel modelling using Fisher scoring.

B3.2 Special features within statistics packages

BMDP5V

GENSTAT macros

GLIM macros

SAS - PROC MIXED for Normal responses; PROC GLIMMIX for non-Normal responses.

REFERENCES

Aitkin, M. (1996) A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and Computing* (in press).

Best, N.G., Spiegelhalter, D.J., Thomas, A. and Brayne, C.E.G. (1996) Bayesian analysis of realistically complex models. *J. Roy. Statist. Soc*., **A**, 159, 323-342.

Breslow, N.E. and Clayton, D.G. (1993) Approximate inference in generalized linear mixed models. *J. Am. Statist. Ass*., 88, 9-25.

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

Diggle, P.J., Liang, K-Y, and Zeger, S.L. (1994) *Analysis of Longitudinal Data*. Oxford: Oxford Science.

Francis, B., Green, M. and Payne, C. (1994) *The GLIM System. Release 4*. Oxford: OUP.

Goldstein, H. (1995) *Multilevel Statistical Models* (2nd. Edn.) London: Edward Arnold.

Hedeker, D. and Gibbons, R.D. (1994) A random-effects ordinal regression model for multilevel analysis. *Biometrics*, 50, 933-944.

Kuk, A.Y.C. (1996) Asymptotically unbiased estimation in generalized linear models with random effects. *J. Roy. Statist. Soc*., **B**, 57, 395-407.

Lee, Y. and Nelder, J.A. (1996) Hierarchical generalized linear models. *J. Roy. Statist. Soc*., **B**, 58, 619-678.

Lindsey, J. K. (1995*) Modelling Frequency and Count Data*. Oxford: Oxford Science.

Longford, N.T. (1993*) Random Coefficient Models*. Oxford: OUP.

McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models* (2nd. Edn.) London: Chapman and Hall.

McGilchrist, C.A. (1994) Estimation in generalized mixed models. *J. Roy. Statist. Soc*., **B**, 56, 61-69.

Note

This a modified and updated version of a piece circulated last year through 'multilevel', the electronic discussion list. I would like to thank everyone who commented on the original version. Further comments and discussion are welcome.

Contributors

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

Alastair Leyland, Public Health Research Unit, University of Glasgow. (a.leyland@udcf.gla.ac.uk)

Huiqi Pan, Mathematical Sciences, Institute of Education University of London. (teuephq@ioe.ac.uk)

Ian Plewis, Mathematical Sciences, Institute of Education University of London. (tesp102@ioe.ac.uk)

Tom A.B. Snijders, Dept. of Statis. & Measur. Theory, Univ. of Groningen (t.a.b.snijders@ppsw.rug.nl)

Sally Thomas, International School Effectiveness & Improvement Centre, Institute of Education University of London. (s.thomas@ioe.ac.uk)

Min Yang, Mathematical Sciences, Institute of Education University of London. (m.yang@ioe.ac.uk)

Multivariate spatial analysis using MLn

Alastair H Leyland[1,3], Ian H Langford[2,3], Jon Rasbash[3] and Harvey Goldstein[3]

1 Public Health Research Unit, University of Glasgow, 1 Lilybank Gardens, Glasgow G12 8RZ
2 CSERGE, University of East Anglia
3 Multilevel Models Project, Institute of Education, University of London

Introduction

When analysing rates in small areas there is typically great variation between areas arising largely due to instability in both the numerator (the number of events) and the denominator (the population at risk). It is not unusual, for example, to have no events observed in a particular area leading to an estimated rate of zero. A common assumption is that the observed values for each area are Poisson realisations of an expected number of events; for the ith area the observed events are distributed as

$$O_i \sim P(\lambda_i) \tag{1}$$

It is then quite conceivable to observe no events in an area with few expected events. The Poisson distribution is handled by the addition of an error term with imposed constraints:

$$\begin{aligned} O_i &= \lambda_i + e_i \\ E(e_i) &= 0 \\ Var(e_i) &= \lambda_i \end{aligned} \tag{2}$$

The expected number of events in the ith area is related to an expected number based on the age/sex composition of the area and population specific rates (E_i) and an area specific random term (u_i) through a link function:

$$\begin{aligned} \log(\lambda_i) &= \log(E_i) + u_i \\ u_i &\sim N(0, \sigma_u^2) \end{aligned} \tag{3}$$

This is then a straight forward random effects model; interest in the analysis focuses on both the estimation of the multipliers u_i - shrunken estimates, having taken the Poisson errors into account, which show more stability - and on the variability in these rates (σ_u^2). When modelling mortality a key measure is the standardised mortality ratio or SMR given by O_i / E_i (usually multiplied by 100); from equations (2) and (3) it is apparent that

$$E(O_i / E_i) = \exp(u_i) \tag{4}$$

Spatial analysis

The first model to consider is a multiple membership model whereby the random effects for each area are also used in the estimation of the effects of other areas through the use of a proximity measure - z_{ij} which denotes a measure of the proximity of the jth area to the ith. In this manner equation (3) is replaced by

$$\log(\lambda_i) = \log(E_i) + u_i + \sum_{j \neq i} z_{ij} u_j \tag{5}$$

As an extension consider the use of a spatial term for each area v_i. This is defined as the effect of each area on other areas and is modified by the proximity measure z_{ij}. Equation (3) then becomes:

$$\begin{aligned} \log(\lambda_i) &= \log(E_i) + u_i + \sum_{j \neq i} z_{ij} v_j \\ \begin{bmatrix} u_i \\ v_i \end{bmatrix} &\sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix} \right) \end{aligned} \tag{6}$$

The covariance term models the association between the random effect of an area (u_i) and its effect on other areas (v_i). Equation (5) is a special case of equation (6) in which the correlation between the u_i and the v_i is assumed to be one. It is also necessary only to give the relative weights of each of the areas in the spatial part in equation (6) - the z_{ij} - as opposed to equation (5) where it is necessary also to consider the total

weight given to the spatial part $\sum_{j \neq i} z_{ij}$ relative to that given to the random part (one). This is because the weighting in equation (6) is determined by the variances, σ_u^2 and σ_v^2. In theory it is possible to fit equation (6) as a random effects model using a series of vectors $\left\{z_{ij}\right\}_j$; in practice this usually proves impractical because *N* areas will require *N - 1* constraints to ensure equality of the variance (σ_v^2) and a further *N - 1* constraints on the covariance (σ_{uv}). Instead it is possible to specify the entire design matrix $\left\{z_{ij}\right\}$ in MLn using the SETDesign command, which means that the random parameters σ_v^2 and σ_{uv} can be estimated explicitly without the need for constraints. To fit a PQL model requires the estimation of residuals at each iteration; at present this must be done directly from the matrices.

Multivariate spatial analysis

It is possible to extend the above to include multivariate outcomes; in this manner covariances between outcomes may be modelled explicitly and the ability to "borrow strength" from other outcomes, particularly given the rarity of the events in question, will provide improved results over independent univariate analyses. Consider an extension to a bivariate model with the outcomes being deaths from two causes, cancers and circulatory disease (denoted A and B). (We will assume that the z_{ij} represent physical proximity and are thus identical for the two causes, although it is straight forward to see how this assumption could be relaxed.) The spatial model in equation (6) may be rewritten:

$$\log\left(\lambda_{i,A}\right) = \log\left(E_{i,A}\right) + u_{i,A} + \sum_{j \neq i} z_{ij} v_{j,A}$$

$$\log\left(\lambda_{i,B}\right) = \log\left(E_{i,B}\right) + u_{i,B} + \sum_{j \neq i} z_{ij} v_{j,B}$$

$$\begin{bmatrix} u_{i,A} \\ v_{i,A} \\ u_{i,B} \\ v_{i,B} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{u,A} & \sigma_{uv,A} & \sigma_{u,A,B} & \sigma_{uv,A,B} \\ \sigma_{uv,A} & \sigma^2_{v,A} & \sigma_{uv,B,A} & \sigma_{v,A,B} \\ \sigma_{u,A,B} & \sigma_{uv,B,A} & \sigma^2_{u,B} & \sigma_{uv,B} \\ \sigma_{uv,A,B} & \sigma_{v,A,B} & \sigma_{uv,B} & \sigma^2_{v,B} \end{bmatrix} \right) \quad (7)$$

although, in practice, it may not make sense to estimate all of the covariances.

Example

As an illustration of this method consider deaths from two causes (cancers and circulatory disease) within all 143 small areas (postcode sectors) in Greater Glasgow Health Board in 1993. There were a total of 3189 cancer deaths (average = 22.3; range = 0 to 61) and 5734 deaths from circulatory disease (average = 40.1; range = 0 to 124) during this year. The log of the expected (standardised) deaths $\log\left(E_i\right)$ were centred around zero so that two constants were also estimated in the fixed part. Table 1 gives these parameter estimates**.**

Table 1: Estimates of fixed parameters

Parameter	Estimate	Standard error
A constant	2.820	0.03104
B constant	3.397	0.03774
$\log\left(E_{i,A}\right)$	1.000	0.00000
$\log\left(E_{i,B}\right)$	1.000	0.00000

The proximity measure used was based on whether two areas bordered each other, giving $z_{ij} = 1/n_i$ if areas i and j are neighbours and 0 otherwise where n_i is the number of neighbours of the i[th] area. Other choices could be made including basing the proximity measure on physical distances between area centroids; one of the authors (IHL) has been considering the use of decay functions with a further, non-linear term in the design matrix for v_i with Z_v to be estimated by the model. Three of the covariances, $\sigma_{uv,B,A}$, $\sigma_{uv,A,B}$ and $\sigma_{v,A,B}$, were set to zero meaning that no correlation was assumed between the random effect for one cause and the spatial effect for another or between the spatial effects for the two causes. Table 2 presents the random parameter estimates; the covariance between the random and spatial effects for cancers ($\sigma_{uv,A}$) was constrained so that the correlation between the two was equal to one to prevent the estimation of a

correlation greater than one and hence has no associated standard error.

The correlation between the random and spatial effects for circulatory disease is also large and positive (0.9899) meaning that, for both causes, the neighbours of areas with high age/sex standardised rates are also likely to have high rates. The correlation between the random effects for the two causes is 0.72, so an area with a high standardised mortality for one cause is also likely to have high standardised mortality for the other.

Table 2: Estimates of random parameters

Parameter	Estimate	S.E.
$\sigma^2_{u,A}$	0.002053	0.007545
$\sigma_{u,A,B}$	0.002239	0.003742
$\sigma^2_{u,B}$	0.004722	0.006157
$\sigma_{uv,A}$	0.01115	-
$\sigma_{uv,B,A}$	0	-
$\sigma^2_{v,A}$	0.06059	0.03891
$\sigma_{uv,A,B}$	0	-
$\sigma_{uv,B}$	0.02374	0.01680
$\sigma_{v,A,B}$	0	-
$\sigma^2_{v,B}$	0.1218	0.0439

Memory requirements

The memory requirements for these models are considerable; a number of large matrices are called up by the program or by the macros. The SETDesign command requires ZZ^T for each random parameter in the model (stored as half of a symmetric matrix; also needed are the original Z design matrices, matrices of the residuals $\{\hat{p}_2\}$ and their full covariance matrix (stored in block form as half of a symmetric matrix). We consider a generalisation of the above model to p outcomes for N areas. The $2p$ design matrices (one Z_u and one Z_v for each area) are each of size $Np \times Np$. $2p(2p+1)/2$ SETD objects (ZZ^T) are required, one for each random parameter to be estimated (10 in the above example, when $p=2$). Each of these half symmetric matrices contains $Np(Np+1)/2$ elements. There are $2p$ design matrices containing N^2p^2 elements and $2p$ residuals containing N elements. The residual variances and covariances are stored in $2p(2p+1)/2$ blocks, each with N^2 elements. The total number of cells required for the storage alone is then

$$N^2p^4 + 5N^2p^3/2 + 2N^2p^2 + N^2p + Np^3 + Np^2/2 + 2Np$$

This expression is dominated by the N^2p^4 term and it is easy to see how memory requirements can get out of hand; in the example above, with $N=143$ and $p=2$ a total of approximately 940,000 worksheet cells but increasing this to 5 outcomes would require over 20,000,000 cells. The N^2p^4 term comes from the storage of the SETD objects required by MLn; in practice these are often repetitive. For example, the four design matrices in equation (12) which must be stored are $\begin{bmatrix} Z_u \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ Z_u \end{bmatrix}, \begin{bmatrix} Z_v \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ Z_v \end{bmatrix}$. Their full products must then also be stored (in half symmetric form) i.e. matrices such as $\begin{bmatrix} Z_u Z_u^T & 0 \\ 0 & 0 \end{bmatrix}$.

This is clearly wasteful and current research is exploring ways of using the structure and likely replication in multivariate models to change these storage requirements.

Acknowledgements

Two of the authors (AHL and IHL) received visiting fellowships from the ESRC as part of the Analysis of Large and Complex Datasets programme. The Public Health Research Unit is supported financially by the Chief Scientist Office of the Scottish Office Department of Health. The views and opinions expressed in this paper are not necessarily those of the Chief Scientist Office.

Software for longitudinal growth norms

Huiqi Pan,
Institute of Education, University of London.

Conventional growth norms are almost exclusively 'cross sectional' in nature; that is they allow the practitioner to judge the 'typicality' of an individual child at a single age. Sometimes, the reference population is desegregated, so that judgements can be made for subgroups, for example based upon parental height, and the most common example is the provision of separate charts for boys and girls.

Very often, however, a sequence of measurements is available and we would like to use these efficiently to judge the typicality of the growth *pattern*. The use of so called 'velocity' norms or standards is the most common example. These, however, are constrained by being limited to just two measurements taken a fixed interval (usually 1 year) apart, whereas in reality a variable number of measurements may be available with varying time intervals.

The aim of this software system (LGROW) is to allow the user to construct norms for any set of measurements taken on a growing child and it seems to be the first serious attempt to do so. Principally it produces two types of norms.

The first is an estimate of the average rate of growth or average acceleration of growth for the set of measurements at the ages they are taken. Thus, for example, if there are three measurements $x_1, x_2, x_3,$ at ages t_1, t_2, t_3 then the average growth rate will be estimated by the slope of the least squares line fitted to these measurements. Using a standardising population sample the software will estimate this value and at the same time derive an estimate of the population distribution from which the sample value can then be assigned a percentile value. Since the data are all transformed to Normality a 'z' score is also computed.

The second type of norm is a 'conditional' one where the distribution of the latest (oldest) measurement is estimated. Thus, in the above example, a percentile position and 'z' score are computed for x_3 conditional on the previous measurements at the previous times. Currently, the system operates with height and weight, and conditional norms can be produced for height given previous heights and for weight given previous heights and/or weights, based on a standardising sample.

The procedure used to derive the norms operates in two stages. The first carries out a transformation of the growth data to Normality (Cole and Green, 1992) and the second stage fits a 2-level repeated measures model to these transformed data. The parameters from the model are used as the basis of the longitudinal norms. Pan and Goldstein (1997) give a full description of the methodology.

The work has been funded by the Medical Research Council and the Child Growth Foundation. A trial version of the software for Windows is available and it is planned to produce a final version in early 1998. Anyone interested in further information should contact Pan Huiqi - teuephq@ioe.ac.uk.

References

Cole, T. J. and Green, P. J., 1992, Smoothing reference centile curves: the LMS method and penalized likelihood. *Statistics in Medicine* 11, 1305-1319.

Pan, H. and Goldstein, H., 1997, Multilevel models for longitudinal growth norms. (*Statistics in Medicine* to appear*)*.

Multilevel models for multiple category responses – a simulation

Min Yang
Institute of Education, University of London

Multiple categorical response data frequently occur in the social sciences, for example attitude scales or choice of party for voting. These responses can either be unordered categories or ordered.

Some literature exists on random-effect models for ordered responses (Harville & Mee, 1984; Janson, 1990; Ezzet & Whitehead, 1991). Recently Hedeker & Gibbons (1994) proposed a random-effect ordinal regression models using Maximum Marginal Likelihood estimation (MML). A FORTRAN program MIXOR (Hedeker & Gibbons, 1996), using a Fisher scoring algorithm, handles either unordered or ordered categories for a two level hierarchy. Goldstein (1991, 1995) extended the multilevel logistic model for binary response to the case of multiple categories, using an iterative generalised least square procedure (IGLS), to obtain approximate Quasi-likelihood estimates. This method has been built into the program MLn in the form of macros MULTICAT (Yang, Goldstein & Rasbash, 1996). It can handle either unordered or ordered category data with many levels with a logit or log-log link.

A brief simulation study has been carried out to examine the estimates given by MLn, on both unordered and ordered multilevel multinomial models, in particular to look at the efficiency of the estimates and convergence properties of the algorithm. This article reports the major findings. Some practical suggestions are provided.

Models

For simplicity we consider a outcome with 3 categories, for individuals grouped within communities. Assuming a true two-level structure with the outcome a set of *3* proportions, $\pi_{ij}^{(t)}$ ($t = 1,2,3$), for the i^{th} individual from the j^{th} community, where category 3 is chosen as the base. The following model is considered

$$\log\left(\frac{\pi_{ij}^{(s)}}{\pi_{ij}^{(3)}}\right) = \beta_0^{(s)} + \beta_1^{(s)} x_{ij} + u_j^{(s)} \qquad (1)$$

$$s = 1,2, \qquad \sum_{t=1}^{3} \pi_{ij}^{(t)} = 1$$

where X is a dichotomous variable, and residuals $u_j^{(s)}$ are random variables at community level (level 2) with a bivariate Normal distribution,

$$\begin{pmatrix} u_j^{(1)} \\ u_j^{(2)} \end{pmatrix} \sim N(0, \Omega_u), \qquad \Omega_u = \begin{bmatrix} \sigma_{u_1}^2 & \\ \sigma_{u_{12}} & \sigma_{u_2}^2 \end{bmatrix}$$

At individual level the response proportions follow a multinomial distribution with variance-covariance matrix

$$n_{ij}^{-1}\begin{bmatrix} \pi_{ij}^{(1)}(1-\pi_{ij}^{(1)}) & \\ -\pi_{ij}^{(1)}\pi_{ij}^{(2)} & \pi_{ij}^{(2)}(1-\pi_{ij}^{(2)}) \end{bmatrix} \qquad (2)$$

The denominator n_{ij} is one in the case where each individual responds just once.

Based on the model above, four sample datasets are simulated (see Table 1).

Sets 1 and 2 are balanced samples with the same sample size and the same parameters. The different between them is that for set 1 there are more level 2 units with overall smaller size than for set 2. Sets 3 and 4 are both unbalanced samples with the same fixed parameters but different random effects at level 2. The level 2 units are sized from 8 to 85. The total number of level 1 units is 1313. These simulations are designed to examine (a) how the sample size at different levels may affect the estimation and convergence; (b) the efficiency of estimates for unbalanced data; (c) the impact of the correlation between categories at a higher level on the estimation procedure.

For the ordered response with a 2-level structure, the response is a set of cumulative proportions, $\gamma_{ij}^{(t)}$ ($t = 1,2,3$), and the following model is used.

$$\log\left(\frac{\gamma_{ij}^{(s)}}{1-\gamma_{ij}^{(s)}}\right) = -(\alpha^{(s)} - \beta x_{ij} + u_j^{(s)}) \qquad (3)$$

$$s = 1,2, \qquad \sum_{h=1}^{s} \pi_{ij}^{(h)} = \gamma_{ij}^{(s)}$$

The residuals $u_j^{(s)}$ have the same distribution as those in model (1), and the covariance matrix at the individual level as

$$n_{ij}^{-1}\begin{bmatrix}\gamma_{ij}^{(1)}(1-\gamma_{ij}^{(1)}) & \\ \gamma_{ij}^{(1)}(1-\gamma_{ij}^{(2)}) & \gamma_{ij}^{(2)}(1-\gamma_{ij}^{(2)})\end{bmatrix} \qquad (4)$$

Four sample sets are used for this model (Table 1) with the same design as for the unordered responses. For set 2, a total of 230 runs was made.

Table 1 Sample size, true parameters and percentage of convened samples (balanced sample=B, unbalanced sample=UB)

	Set 1 (B)	Set 2 (B)	Set 3 (UB)	Set 4 (UB)
Unordered responses				
No. lev. 1 units per lev. 2 U	30	60	8-85	8-85
No. level 2 units	150	75	48	48
Total level 1 units	4,500	4,500	1,313	1,313
% converged, total. runs	100.0%, 200	98.5%, 200	98.0%, 200	68.0%, 200
$\beta_0^{(1)}, \beta_0^{(2)}, \beta_1^{(1)}, \beta_1^{(2)}$	-1.5 ,-2.0, 0.8, 1.2	-1.5 ,-2.0, 0.8, 1.2	-1.5, -2.0, 0.8, 1.2	-1.5, -2.0, 0.8, 1.2
$\sigma_{u_1}^2$, $\sigma_{u_{12}}$, $\sigma_{u_2}^2$	1.0, 0.5, 1.0	1.0, 0.5, 1.0	1.0, 0.5, 1.0	0.6, 0.5, 0.6
Ordered responses				
No. lev. 1 units per lev. 2 U	30	30	8-85	8-85
No. level 2 units	150	150	48	48
Total level 1 units	4,500	4,500	1,313	1,313
% converged, Total runs	96.0%, 200	97.8%, 230	98.0%, 200	89.5%, 200
$\alpha^{(1)}, \alpha^{(2)}, \beta$	-1.5, 2.5, 2.0	-1.5, 2.5, 2.0	-1.5, 2.5, 2.0	-1.5, 2.5, 2.0
$\sigma_{u_1}^2$, $\sigma_{u_{12}}$, $\sigma_{u_2}^2$	1.0, 0.5, 1.0	1.0, 0.8, 1.0	1.0, 0.5, 1.0	0.6, 0.5, 0.6

Using MLn, the Penalised Quasi-likelihood (PQL) procedure (Breslow & Clayton, 1993) with either 1st or 2nd order approximation (Goldstein & Rasbash, 1996) is compared with the simplest Marginal Quasi-likelihood (MQL) procedure for both unordered and ordered models.

Simulation Results

Convergence

From Table 1 we see that set 4 for unordered responses with an unbalanced design, small numbers of level 1 units per level 2 unit and a high correlation between category residuals at level 2 has convergence problems. The sample set 4 for ordered responses has less convergence problems than the unordered sample set 4, although these two sets have the same data structure and same random parameters at level 2. Set 2 for ordered responses with a balanced design and relatively large numbers of level 1 per level 2 unit has few convergence problems, although the samples are drawn from a population with highly correlated level 2 effects.

It is found that nearly all failures in achieving convergence occur in using the PQL procedures, either the 1st or 2nd order approximation. By checking the iteration procedure more closely before the estimation failure happens, it is found that often an estimated correlation at level 2 is greater than one, which gives the convergence problem.

Experimenting with some of these samples suggests some simple ways to get around the convergence problems. First one can choose different procedures from the four options of the macro settings to start the first iteration instead of always choosing the first order MQL to begin with. In particular starting with the procedure of PQL+1st order appeared to perform well on many samples, which failed to achieve convergence on the same procedure when the iteration started with MQL+1st. Secondly adding or removing parameters can help. Thirdly, in some cases where a correlation at higher level is estimated greater than 1.0 in absolute value, we can constrain such a correlation say between u_{z_j} and $u_{z_{j'}}$ to be 1.0 by applying the following constraint after each iteration: $C_1\sigma_{u_{zz'}} - C_2\sigma_{u_{z'}}^2 = 0$

where C_1 is the estimate of $\sigma_{u_{zz'}}$ from the previous iteration, C_2 is the estimate of $\sigma^2_{u_z}$ also from the previous iteration. These should be updated after each iteration.

Bias and Efficiency

In Table 2, we see that the MQL procedure underestimates all parameters, especially the random ones, for the two-level multinomial model.

Table 2 Mean values of simulations: MSE, Empirical standard error and Mean of estimated standard errors in brackets 1st, 2nd and 3rd respectively for unordered response samples

Param.	True values	MQL + 1st	PQL + 1st	PQL + 2nd
Set 1				
$\beta_0^{(1)}$	-1.5	-1.28 (0.056) (0.09)(0.09)	-1.44 (0.013) (0.10)(0.10)	-1.46 (0.011) (0.10)(0.10)
$\beta_0^{(2)}$	-2.0	-1.75 (0.074) (0.10)(0.09)	-1.92 (0.017) (0.10)(0.11)	-1.93 (0.015) (0.10)(0.10)
$\beta_1^{(1)}$	0.8	0.68 (0.020) (0.07)(0.07)	0.81 (0.007) (0.08)(0.08)	0.79 (0.006) (0.08)(0.08)
$\beta_1^{(2)}$	1.2	1.06 (0.027) (0.09)(0.08)	1.22 (0.010) (0.10)(0.08)	1.19 (0.009) (0.10)(0.08)
$\sigma^2_{u_1}$	1.0	0.66 (0.126) (0.09)(0.10)	0.96 (0.020) (0.13)(0.14)	0.94 (0.021) (0.13)(0.13)
$\sigma_{u_{12}}$	0.5	-0.08 (0.339) (0.06)(0.07)	0.52 (0.011) (0.10)(0.10)	0.50 (0.011) (0.10)(0.10)
$\sigma^2_{u_2}$	1.0	0.65 (0.134) (0.10)(0.10)	0.93 (0.023) (0.14)(0.14)	0.91 (0.025) (0.13)(0.14)
Set 2				
$\beta_0^{(1)}$	-1.5	-1.28 (0.061) (0.11)(0.11)	-1.46 (0.017) (0.13)(0.13)	-1.47 (0.016) (0.13)(0.13)
$\beta_0^{(2)}$	-2.0	-1.76 (0.073) (0.12)(0.12)	-1.97 (0.019) (0.13)(0.13)	-1.98 (0.018) (0.13)(0.13)
$\beta_1^{(1)}$	0.8	0.67 (0.020) (0.07)(0.07)	0.80 (0.006) (0.08)(0.08)	0.78 (0.006) (0.07)(0.08)
$\beta_1^{(2)}$	1.2	1.06 (0.027) (0.08)(0.08)	1.21 (0.009) (0.09)(0.09)	1.19 (0.009) (0.09)(0.09)
$\sigma^2_{u_1}$	1.0	0.65 (0.138) (0.12)(0.12)	0.95 (0.035) (0.18)(0.17)	0.93 (0.035) (0.17)(0.17)
$\sigma_{u_{12}}$	0.5	-0.08 (0.345) (0.08)(0.09)	0.50 (0.022) (0.15)(0.14)	0.48 (0.022) (0.15)(0.13)
$\sigma^2_{u_2}$	1.0	0.67 (0.125) (0.13)(0.13)	0.96 (0.036) (0.18)(0.18)	0.94 (0.036) (0.18)(0.18)
Set 3				
$\beta_0^{(1)}$	-1.5	-1.30 (0.067) (0.17)(0.15)	-1.45 (0.036) (0.19)(0.18)	-1.48 (0.034) (0.18)(0.18)
$\beta_0^{(2)}$	-2.0	-1.77 (0.093) (0.20)(0.17)	-1.93 (0.056) (0.22)(0.19)	-1.94 (0.049) (0.21)(0.19)
$\beta_1^{(1)}$	0.8	0.70 (0.029) (0.13)(0.14)	0.83 (0.022) (0.15)(0.14)	0.80 (0.020) (0.14)(0.14)
$\beta_1^{(2)}$	1.2	1.05 (0.048) (0.16)(0.16)	1.20 (0.034) (0.18)(0.16)	1.17 (0.031) (0.17)(0.16)
$\sigma^2_{u_1}$	1.0	0.63 (0.169) (0.18)(0.18)	0.94 (0.084) (0.28)(0.25)	0.90 (0.079) (0.26)(0.24)
$\sigma_{u_{12}}$	0.5	-0.07 (0.343) (0.12)(0.13)	0.51 (0.053) (0.23)(0.19)	0.47 (0.048) (0.22)(0.18)
$\sigma^2_{u_2}$	1.0	0.61 (0.186) (0.19)(0.18)	0.90 (0.086) (0.27)(0.25)	0.86 (0.088) (0.26)(0.24)
Set 4				
$\beta_0^{(1)}$	-1.5	-1.41 (0.031) (0.15)(0.13)	-1.49 (0.025) (0.16)(0.15)	-1.51 (0.024) (0.16)(0.15)
$\beta_0^{(2)}$	-2.0	-1.89 (0.044) (0.18)(0.15)	-1.97 (0.038) (0.19)(0.17)	-1.99 (0.037) (0.19)(0.17)
$\beta_1^{(1)}$	0.8	0.71 (0.027) (0.14)(0.14)	0.82 (0.021) (0.14)(0.14)	0.81 (0.020) (0.14)(0.14)
$\beta_1^{(2)}$	1.2	1.10 (0.038) (0.17)(0.16)	1.21 (0.038) (0.17)(0.16)	1.20 (0.033) (0.18)(0.16)
$\sigma^2_{u_1}$	0.6	0.32 (0.090) (0.11)(0.11)	0.57 (0.031) (0.17)(0.17)	0.57 (0.032) (0.18)(0.17)
$\sigma_{u_{12}}$	0.5	0.14 (0.137) (0.07)(0.08)	0.49 (0.026) (0.16)(0.14)	0.48 (0.027) (0.16)(0.14)
$\sigma^2_{u_2}$	0.6	0.30 (0.106) (0.12)(0.12)	0.58 (0.038) (0.19)(0.18)	0.58 (0.039) (0.19)(0.18)

The procedures PQL+1st and PQL+2nd both improve estimates almost equally well, producing reasonably unbiased estimates for the fixed effects βs and slightly biased parameter estimates for the random effects at level 2. The estimated standard errors appear to be accurate. The estimates on sample set 1 with small numbers of level 1 units

per level 2 unit but more level 2 units are more efficient than those of set 2, in particular for the random parameter estimates. The estimates for unbalanced sample sets 3 and 4 are almost as same good as those for balanced samples.

For ordered categorical data both the fixed and random parameter estimates using MQL procedure in Table 3 appear to be less biased than those for the unordered category data, but are still underestimating. Again PQL procedures provide good estimates. Comparing estimates between the 1st and 2nd order approximations of PQL suggests that the latter one produces estimates closer to their true values than the former, although the differences of MSE between them are rather small on all sample sets. The standard errors are estimated accurately.

Table 3 Mean values of simulations: MSE, Empirical standard error and Mean of estimated standard errors in brackets 1st, 2nd and 3rd respectively for ordered response samples

Params.	True values	MQL + 1st	PQL + 1st	PQL + 2nd
Set 1				
$\alpha^{(1)}$	-1.5	-1.27 (0.059) (0.09)(0.09)	-1.45 (0.012) (0.10)(0.10)	-1.50 (0.010) (0.10)(0.10)
$\alpha^{(2)}$	2.5	2.17 (0.118) (0.10)(0.10)	2.33 (0.041) (0.11)(0.10)	2.49 (0.016) (0.13)(0.11)
β	2.0	1.69 (0.101) (0.07)(0.07)	1.93 (0.010) (0.07)(0.07)	1.99 (0.006) (0.08)(0.07)
$\sigma^2_{u_1}$	1.0	0.71 (0.092) (0.08)(0.10)	0.93 (0.021) (0.13)(0.13)	1.00 (0.020) (0.14)(0.14)
$\sigma_{u_{12}}$	0.5	0.37 (0.024) (0.09)(0.10)	0.44 (0.017) (0.12)(0.12)	0.52 (0.018) (0.13)(0.13)
$\sigma^2_{u_2}$	1.0	1.00 (0.049) (0.22)(0.18)	0.82 (0.055) (0.16)(0.17)	1.02 (0.044) (0.21)(0.21)
Set 2				
$\alpha^{(1)}$	-1.5	-1.27 (0.060) (0.08)(0.08)	-1.45 (0.010) (0.09)(0.10)	-1.50 (0.008) (0.09)(0.10)
$\alpha^{(2)}$	2.5	2.17 (0.116) (0.10)(0.10)	2.39 (0.023) (0.11)(0.11)	2.51 (0.014) (0.12)(0.11)
β	2.0	1.69 (0.099) (0.07)(0.07)	1.94 (0.008) (0.07)(0.07)	2.00 (0.005) (0.07)(0.07)
$\sigma^2_{u_1}$	1.0	0.70 (0.095) (0.08)(0.10)	0.93 (0.023) (0.13)(0.13)	1.00 (0.021) (0.15)(0.14)
$\sigma_{u_{12}}$	0.8	0.57 (0.057) (0.08)(0.11)	0.74 (0.019) (0.13)(0.13)	0.81 (0.019) (0.14)(0.14)
$\sigma^2_{u_2}$	1.0	0.99 (0.057) (0.24)(0.18)	0.89 (0.043) (0.18)(0.18)	1.01 (0.047) (0.22)(0.21)
Set 3				
$\alpha^{(1)}$	-1.5	-1.29 (0.073) (0.16)(0.15)	-1.45 (0.034) (0.18)(0.17)	-1.51 (0.035) (0.19)(0.18)
$\alpha^{(2)}$	2.5	2.18 (0.145) (0.20)(0.19)	2.33 (0.075) (0.21)(0.19)	2.50 (0.062) (0.25)(0.21)
β	2.0	1.69 (0.110) (0.13)(0.12)	1.93 (0.024) (0.14)(0.13)	2.00 (0.020) (0.14)(0.14)
$\sigma^2_{u_1}$	1.0	0.68 (0.122) (0.14)(0.18)	0.90 (0.069) (0.24)(0.23)	0.98 (0.075) (0.27)(0.25)
$\sigma_{u_{12}}$	0.5	0.36 (0.050) (0.17)(0.19)	0.44 (0.064) (0.25)(0.21)	0.52 (0.082) (0.29)(0.24)
$\sigma^2_{u_2}$	1.0	0.95 (0.163) (0.40)(0.34)	0.82 (0.143) (0.33)(0.33)	1.01 (0.047) (0.46)(0.40)
Set 4				
$\alpha^{(1)}$	-1.5	-1.38 (0.038) (0.15)(0.14)	-1.49 (0.024) (0.15)(0.15)	-1.54 (0.027) (0.16)(0.15)
$\alpha^{(2)}$	2.5	2.28 (0.080) (0.18)(0.17)	2.42 (0.046) (0.20)(0.17)	2.51 (0.050) (0.22)(0.18)
β	2.0	1.81 (0.055) (0.14)(0.13)	1.97 (0.020) (0.14)(0.13)	2.03 (0.021) (0.14)(0.13)
$\sigma^2_{u_1}$	0.6	0.47 (0.032) (0.12)(0.15)	0.56 (0.032) (0.17)(0.16)	0.61 (0.037) (0.19)(0.17)
$\sigma_{u_{12}}$	0.5	0.39 (0.029) (0.13)(0.15)	0.48 (0.034) (0.18)(0.17)	0.52 (0.040) (0.20)(0.18)
$\sigma^2_{u_2}$	0.6	0.58 (0.100) (0.32)(0.27)	0.55 (0.071) (0.26)(0.26)	0.60 (0.107) (0.33)(0.29)

Extra-multinomial variation

By removing constraints on the covariance matrix (2) or (4), we examine the distributional assumption on the responses. For the estimation procedure, we would expect no significant extra-multinomial variation to be estimated from the simulated samples. This investigation is carried out only on sample set 1 for both the unordered and ordinal models with 200 runs for each set. Results from the converged samples of 200 from the unordered and 192 from the ordered response samples are listed in Table 4.

Table 4 Estimates of models allowing extra-multinomial variation: MSE, Empirical standard error and Mean of estimated standard errors in brackets 1st, 2nd and 3rd respectively

Params.	True values	MQL + 1st	PQL + 2nd
Unordered			
$\beta_0^{(1)}$	-1.5	-1.28 (0.056) (0.09)(0.08)	-1.46 (0.011) (0.10)(0.10)
$\beta_0^{(2)}$	-2.0	-1.74 (0.077) (0.11)(0.09)	-1.95 (0.015) (0.11)(0.10)
$\beta_1^{(1)}$	0.8	0.68 (0.020) (0.07)(0.07)	0.79 (0.007) (0.08)(0.07)
$\beta_1^{(2)}$	1.2	1.05 (0.030) (0.09)(0.08)	1.19 (0.010) (0.10)(0.08)
$\sigma^2_{u_1}$	1.0	0.66 (0.121) (0.09)(0.10)	0.96 (0.021) (0.14)(0.14)
$\sigma_{u_{12}}$	0.5	-0.08 (0.337) (0.06)(0.07)	0.52 (0.012) (0.11)(0.11)
$\sigma^2_{u_2}$	1.0	0.68 (0.114) (0.11)(0.10)	0.97 (0.024) (0.15)(0.14)
Extra-multinomial	1.0	0.89 (0.013) (0.01)(0.01)	0.90 (0.009) (0.01)(0.01)
Ordered			
$\alpha^{(1)}$	-1.5	-1.27 (0.061) (0.08)(0.08)	-1.50 (0.009) (0.09)(0.10)
$\alpha^{(2)}$	2.5	2.17 (0.118) (0.11)(0.10)	2.52 (0.018) (0.13)(0.11)
β	2.0	1.69 (0.099) (0.07)(0.06)	2.01 (0.006) (0.08)(0.07)
$\sigma^2_{u_1}$	1.0	0.73 (0.083) (0.08)(0.10)	1.03 (0.026) (0.16)(0.14)
$\sigma_{u_{12}}$	0.5	0.38 (0.023) (0.09)(0.10)	0.53 (0.020) (0.14)(0.13)
$\sigma^2_{u_2}$	1.0	1.08 (0.065) (0.24)(0.19)	1.10 (0.068) (0.24)(0.22)
Extra-multinomial	1.0	0.87 (0.017) (0.03)(0.01)	0.88 (0.015) (0.03)(0.01)

For the unordered responses known multinomial distribution, MLn gives estimates of variation about 10% reduced, with the fixed estimates unchanged and the random effects slightly larger compared to the sample set 1 in Table 2.

For the ordinal responses, all convergence failures occurred to the PQL+1st order procedure when it was applied to estimates using MQL+1st. However they all achieved convergence when the iterations started from the PQL+1st order procedure. Again an under-estimation of about 10% on the variance is observed, while the random parameters at the individual level were slightly over-estimated.

MIXOR and MLn - an example

Hedeker and Gibbons [1996] fitted a mixed-effect ordinal regression model to the data from the Television School and Family Smoking Prevention and Cessation Project (TVSFP) with students nested within classrooms in schools. It was designed to test independent and combined effects of a school-based social-resistance curriculum and a television-based program in terms of tobacco use prevention and cessation. The original design forms a 2×2 classification of social-resistance classroom curriculum (CC = yes or no) by mass-media intervention (TV = yes or no) as explanatory variables. The outcome, a tobacco and health knowledge scale (THKS) after intervention, scores ordinally from 1 to 4. The

variable pre-intervention THKS scores from 0 to 7 and is treated as a continuous baseline variable.

The authors modelled the classroom effect using one random term at level 2 based on 1,600 students from 135 classes, and ignored the school clustering because of software limitations. Results are listed in Table 5.

Their model may be written as

$$\log\left(\frac{\gamma_{ij}^{(s)}}{1-\gamma_{ij}^{(s)}}\right) = -(\alpha^{(s)} + \beta_1(prethks)_{ij} + \beta_2(CC)_j + \beta_3(TV)_j + \beta_4(CC*TV)_j + u_j) \quad (5)$$

$s = 1,2,3$

The single residual term u_j has a zero mean and variance σ_u^2. Fitting this model using MLn gives the estimates in Table 5. All estimates are comparable for the two programs with standard errors from MLn estimates slightly smaller than those from MIXOR. Estimates from the MQL procedure compared to those from PQL in this case do not differ as large as those for the simulated samples This may due to the relatively small level 2 variance estimate.

Table 5 Estimates on TVSFP data by MIXOR and MLn (standard error in bracket)

	MIXOR	MLn , PQL+2nd	MLn, MQL+2nd
Fixed effect			
PRETHKS	0.415 (0.041)	0.419 (0.039)	0.417 (0.039)
CC	0.861 (0.187)	0.870 (0.175)	0.872 (0.168)
TV	0.206 (0.168)	0.206 (0.171)	0.209 (0.164)
CC*TV	-0.301 (0.252)	-0.300 (0.246)	-0.315 (0.236)
THKS-1	0.076 (0.154)	0.084 (0.147)	0.069 (0.143)
THKS-2	1.273 (0.063)	1.295 (0.059)	1.199 (0.145)
THKS-3	2.479 (0.080)	2.522 (0.076)	2.401 (0.154)
Random effect			
Class level	0.189 (0.076)	0.197 (0.058)	0.164 (0.054)
*-log-likelihood**	2115.38	2079.21	2182.5

** The MLn log-likelihood value is very approximate as shown here.*

The correlation coefficients of the fixed estimates in Table 6 suggest that both programs give similar results when high correlation is involved but some discrepancy with weak correlation.

Table 6 Correlation of fixed effect estimates by MIXOR and MLn (in brackets)

	THKS-1	THKS-2	THKS-3	PRETHKS	CC	TV
THKS-2	-0.02(-0.16)					
THKS-3	0.02(-0.10)	0.76(0.70)				
PRETHKS	-0.53(-0.52)	0.32(-0.13)	0.31(-0.21)			
CC	-0.56(-0.57)	0.11(-0.07)	0.05(-0.10)	0.03(0.06)		
TV	-0.60(-0.58)	0.08(-0.02)	0.00(-0.03)	-0.01(0.03)	0.50(0.49)	
CC*TV	0.36(0.40)	-0.02(0.02)	0.06(0.02)	0.09(-0.02)	-0.72(-0.71)	-0.66(-0.69)

With both MIXOR and MLn, model (5) can be further extended to include many more random terms at class level. In addition, MLn can handle more than two hierarchical levels.

Summary

This small scale simulation study shows that, implemented by MLn macros, the non-linear approximation procedure for multilevel binary response under performs satisfactorily for multilevel multiple categorical responses especially with ordinal data. The procedure PQL produces much less biased estimates for both fixed and random effects than the MQL procedure, which underestimates parameters. The standard errors of all parameter estimates are estimated accurately .

The PQL procedure may not always achieve convergence especially in the case with an estimated high correlation. In practice, one can constrain such correlation to be one before each

iteration, or remove a parameter with a small estimate from the random part of the model, or start the iteration with a different procedure. As an example all four of the samples from set 3 of the ordinal model which failed to converge on PQL+1st after MQL+1st procedure, achieved convergence when we fitted them using PQL+1st to begin with. Changing the base category in the case of unordered responses and link function may help to achieve convergence.

Estimates for the example with multilevel ordinal responses using MLn are comparable with those obtained from MIXOR based on marginal maximum likelihood, although there are some small discrepancies in the standard errors in this example. For more comprehensive comparison between MIXOR and MLn, fitting models with more random parameters at level 2 on different data structures is required.

(Acknowledgements to the Economic & Social Research Council in UK funded the ALCD research programme and to Don Hedeker who provided with TVSFP data as well as those who commented on the draft).

References

Breslow, N.E. and Clayton, D.G. 1993: Approximate inference in generalised linear models. *J. Amer. Statist. Assocn.* 88: 9-25.

Ezzet, F. & Whitehead, J. 1991: A random effects model for ordinal responses from a crossover trial, *Statistics in Medicine*, **10**, 901-907.

Goldstein, H. 1991: Non-linear multilevel models with an application to discrete response data. *Biometrika*, **73**, 43-51.

Goldstein, H. 1995: *Multilevel Statistical Models*, 2nd edition, London: Edward Arnold.

Goldstein, H. and Rasbash, J. 1996: Improved approximations for multilevel models with binary responses. *JRSS*, A **159**, 505-514.

Harville, D.A. & Mee, R.W. 1984: A mixed-model procedure for analyzing ordered categorical data, *Biometrics* **40**, 393-408.

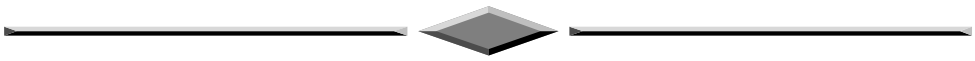
Hedeker, D and Gibbons, R.D. 1994: A random-effects ordinal regression model for multilevel analysis, *Biometrics* **50**, 933-944.

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

Janson, J. 1990: On the statistical analysis of ordinal data when extravariation is presented, *Applied Statistics* **39**, 75-84.

McCullagh, P. and Nelder, J.A. 1992: *Generalized Linear Models*, 2nd Edition, London, Chapman & Hall.

Yang, M., Goldstein, H. and Rasbash, J. 1996: *MLn Macros for Advanced Multilevel Modelling*, **V1.1**, Institute of Education, University of London.



Amsterdam Conference on Multilevel Analysis
– Invited speakers and contributed papers

Invited speakers	**Topic**
Uwe Engel	Handling non-response in multilevel analysis
Donald Hedeker	A partial proportional odds model for multilevel ordinal response
Toby Lewis & Ian Langford	Outliers in multilevel models
Rien van der Leeden, Frank Busing & Erik Meijer	Applications of bootstrap methods for two-level models

Contributed papers	**By**

Score tests for a random slope	Hans Berkhof
Wage differentials across firms: an application of multilevel modelling	Ana Rute Cardoso
Residual analysis in multilevel models: the investigation of independence, variance homogeneity and normality using order statistics evaluated via posterior predictive p-values	Herbert Hoijtink
Modeling growth data: a comparison of multilevel regression and latent growth models	J.J. Hox
Multivariate spatial analysis using MLn	Alastair H. Leyland, Ian H. Langford, Jon Rasbash & Harvey Goldstein
Application of indirect inference for multilevel models with binary response	Fabrizia Mealli & Carla Rampichini
Comparison of three different models for the analysis of multilevel data	M. Moerbeek, M.P.F. Berger & G.J.P. van Breukelen
Random effects models for event data: a study of university dropout	Massimo Montagni & Enrico Gori
A multilevel logistic regression model for the analysis of a cluster randomised trial on breast screening	Rumana Omar, Simon Thompson, John Robson, Maggie Falshaw & Jyoti Atri, Ros Gregg
Modeling interviewer effects in panel surveys	Jan Pickery & Geert Loosveldt
Non-linear models for ordinal repeated measures data	Ian Plewis
Drinking partners out of house and home: an empirical analysis of the identification, estimation and interpretation of higher-level effects	Nigel Rice & Matthew Sutton
Change in friendship and family ties: an example of multilevel analysis on data from ego-centered networks	Jooske van Busschbach
Time- and respondent-related causes of unit and item non-response on the CES-D depression scale: a multilevel model	Pieter van den Eeden, Johannes Smit & Aart-Jan Beekman
Optimal scaling in a multilevel model	L.B. van der Brugge, M. Gerritsma & P. Houweling
Using multilevel analysis in an economic analysis of blood collection	Marjon van der Pol, & John Cairns
Individual and contextual characteristics in the analysis of reproductive behavior in Italy	Susanna Zaccarin