MULTILEVEL MODELLING NEWSLETTER

The Multilevel Models Project
Mathematical Sciences
Institute of Education, University of London
20 Bedford Way, London WC1H 0AL, ENGLAND
Web sites: http://www.ioe.ac.uk/multilevel/
http://multilevel.ioe.ac.uk
Enquiries about newsletter to Ian Plewis
E-mail: i.plewis@ioe.ac.uk
Tel: +44 (0) 20 7612 6688 *Fax: +44 (0) 20 7612 6572*

Vol. 12 No. 1 **August, 2000**

Forthcoming Workshop

6-8 September 2000, 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 *CSERGE, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, England*. Tel: +44 (0) 1603 593314 or +44 (0) 1692 538196, Fax: +44 (0) 1603 593739, e-mail: a.cox@uea.ac.uk.

8-10 January 2001, a three-day introductory workshop to multilevel modelling using *MLwiN* will take place in the Institute of Education, University of London.
Enquiries to Amy Burch at Mathematical Sciences, Institute of Education, 20 Bedford Way, London WC1H OAL. Tel: +44 (0) 20 7612 6688, Fax: +44 (0) 20 7612 6572, e-mail: a.burch@ioe.ac.uk.

Bootstrapping Measurement Error Project

Dougal Hutchison at the National Foundation for Educational Research has received ESRC funding for a 10-month project which started in January, aimed at developing bootstrapping techniques for correcting for measurement error in dependent and explanatory variables in multilevel models. Jo Morrison and Rachel Felgate of the NFER will be working with him on the project. It is aimed to implement developments in forthcoming versions of *MLwiN*.

Also in this issue

Version 1.10 of *MLwiN* now available

The new version has the following enhancements:

Worksheet and data management

- Comprehensive menu-based data manipulation facilities
- Missing data now handled automatically under full user control
- New macro editor with diagnostic facilities
- Variable categories can have names attached
- Windows for viewing the data structure
- Pasting data sets directly from other applications via windows clipboard
- Variable subscripts can be named

Graphics

- Enhanced graphics facilities with flexible control over formats
- Trellis plots for displaying multiway graph layouts
- Identification of points and lines in different formats and colours and linking across graphs

Modelling

- Improvements to efficiency of MCMC methods and inclusion of general binomial and Poisson models with enhanced diagnostics
- Non-parametric bootstrapping based on model residuals
- A range of model diagnostics, including leverage and influence measures
- Graphically identify outliers at any level and automatically omit or "dummy" them out of the model
- Powerful procedures for including interaction terms
- Weighting at any level

Existing users can download an upgrade from http://multilevel.ioe.ac.uk/mlwin/software-upgrade.html

New users can purchase *MLwiN* using the secure on-line ordering facility at https://multilevel.ioe.ac.uk/ordering/orderform.html

Free teaching version of *MLwiN* available from the TRaMSS web site.

A version of *MLwiN* along with two extended tutorial examples is available for downloading from http://tramss.data-archive.ac.uk/

The software is limited so that it will only work with the tutorial data sets but otherwise has full functionality. The teaching materials are designed to introduce users to the basic concepts of multilevel modelling through worked examples using *MLwiN*. The materials can be used to supplement undergraduate and post-graduate courses on multilevel modelling, as a learning aid for individual researchers, or for users to evaluate *MLwiN* before purchasing a full version.

In addition to the materials on multilevel modelling the TRaMSS (Teaching Resources and Materials for the Social Sciences) site contains material designed to:

- Teach users how to find data using the Essex Data Archive's catalogue
- Promote learning about event history analysis using worked examples from the SABRE modelling package

The TRaMSS web site was funded by the ALCD (Analysis of Large and Complex Datasets) training materials programme. ALCD is an ESRC initiative.

Free student version of HLM 5

A free student version of HLM 5 is available from

http://www.ssicentral.com/other/hlmstu.htm

The student edition contains the following:

- All the examples distributed with the full HLM 5 version. These examples may be run with the student edition.
- An on-line helpfile, as provided for the full version too. The helpfile includes most of the new HLM 5 manual and a complete tutorial showing the use of HLM.

The student edition can run all the analyses the full version can in terms of models selected, statistical options and output. Restrictions are, however, placed on the data used and the size of the model selected.

An Instrumental variable consistent estimation procedure to overcome the problem of endogenous variables in multilevel models

***Neil H Spencer*[1], *Antony Fielding*[2]**
[1]University of Hertfordshire, [2]University of Birmingham
e-mail: n.h.spencer@herts.ac.uk

Introduction

It is not unusual for a multilevel model to contain fixed effect explanatory variables that can be regarded as endogenous. This will happen if the variables are subject to the same unmeasured influences as the dependent response. These influences will be incorporated in the random effect disturbance terms. Thus, these variables are no longer exogenous and are not independent of the random effects in the model. In such circumstances, a basic model assumption is not met and obtaining consistent estimators of the parameters is not straightforward. Many standard multilevel procedures (e.g. Iterative Generalised Least Squares: IGLS) rely on the independence of regressors and model disturbances for their consistency properties (as, indeed, do single level procedures). Here, we present a modelling strategy based on instrumental variables and introduce an *MLwiN* macro that provides consistent estimators.

For exemplification, we consider a simple two level model with endogenous variables. Fielding (1998) introduces a dataset drawn from children in primary schools of the City of Birmingham Local Education Authority. Data are available on a range of school and pupil characteristics. The responses are the results of Key Stage 1 (KS1) tests, and we wish to relate one of these test results to gender, age of the child in months, and the results of baseline tests carried out when the child entered the reception classes in the school. A model such as this may be used to examine the progress children are making between reception and KS1 in different schools. For pupil i in school j and where we have just one baseline test we may write:

$$\begin{aligned} KS1TEST_{ij} &= \beta_0 + \beta_1 GENDER_{ij} + \\ &\beta_2 AGE_{ij} + \beta_3 BASELINE_{ij} + u_j + \varepsilon_{ij} \end{aligned} \quad (1)$$

The term u_j in model (1) represents a random school effect and ε_{ij} is a within school random pupil effect. The endogeneity in this model arises because the baseline test may be supposed to be related to the random pupil effect through the existence of important unmeasured and unmeasureable influences acting at this lowest level of the hierarchy (e.g. home circumstances). These influences are incorporated in the disturbance ε_{ij} but may also influence baseline test performance. It is also possible that there are some influences that make the baseline test related to the school effect u_j. The common influences may be such things as the locality in which the school is situated and from which the pupils generally come.

Overcoming the problem of endogeneity

Solutions to the problem of inconsistency caused by endogenous regressors, particularly when they are thought to relate to higher level effects such as u_j, have been proposed by Kiviet (1995) and Rice et al. (1998). Kiviet uses a bias corrected version of the least squares dummy variable estimator (LSDV). Rice et al. use conditioned iterative generalised least squares (CIGLS). The first of these approaches suffers from a problem that the bias correction applied may actually increase the bias in some circumstances. Neither approach can easily cope with the case where the level 1 (pupil) random effect is correlated with regressors. It is this latter situation on which we mainly focus. The difficulties caused by endogenous regressors in the context of generalised linear models for count data are also discussed by Crouchley and Davies (1999).

A frequently used method of overcoming such endogeneity problems in single level models is to use instrumental variable techniques. We adapt these techniques to cover multilevel random effects models within the framework of the *MLwiN* IGLS estimation procedures. The possibility is mentioned briefly and independently by Rice et al. (1998). Spencer (1997) also suggests such an approach for repeat testing in educational situations where explanatory variables are lagged versions of the response. A supplementary multilevel model for the endogenous explanatory variable is constructed using fixed effect explanatory variables that are assumed exogenous and independent of the random part of model (1). We stress that the existence of such variables and the adequate collection of data on them are a necessary pre-requisite. Predictions of the endogenous variable values for each child are then obtained from the fixed parts of the supplementary models. These predicted values, being independent of the random part of the model of interest (1), are then used as instruments.

Armed with data on the original set of regressors of model (1) and the set of instruments (being the original regressor set with endogenous variables replaced by their instruments), we estimate the fixed effect parameters in model 1 (see, e.g., Bowden & Turkington (1984)). This provides us with consistent estimates of the fixed parameters but at this stage adequate estimates of their standard errors are not available.

The next stage, then, is to obtain estimates of the random part of model (1). This is done by using *MLwiN* procedures to create constraints on the fixed parameters. They are forced to be equal to those calculated from the instrumental variable procedure. The resulting estimates of the random part of the model will be consistent (Goldstein, 1986) and can then be used to obtain standard errors of the instrumental variable fixed effects estimates.

MLwiN macros called IV have been written to implement this procedure. They are available from the authors on request or can be downloaded from the Birmingham web page www.bham.ac.uk/economics/staff/tony.htm

Application

We now use the example of Fielding (1999) discussed above to demonstrate the use and performance of the instrumental variable method embodied in the macros. Simulation results are also available (Spencer and Fielding, 2000). The particular response variable used is the Mathematics test at Key Stage 1, standardised to have mean zero and unit variance.

The seven baseline tests available in the data (various forms of Mathematics and English tests) are inevitably highly correlated and so the first principal component (accounting for 60% of the variation) was used. The supplementary multilevel model for the endogenous principal component score had a similar structure to model (1) with intercept random effects for school and pupils. Fixed effect explanatory variables included pupil's ethnicity, first language and attendance at nursery school. The ones used were, on investigation, related to ability and therefore to baseline test scores. However, none appeared to have an influence on progress. It is unlikely, therefore, that they are correlated with disturbances in target model (1). Predictions of the principal component of baseline scores from this model were thus thought to provide an instrument that was free of the problem of dependence on the disturbances in the original progress model of interest.

Table 1 shows estimates of the fixed parameters (and estimated standard errors) of the adapted model (1) obtained with and without the consistent instrumental variable estimation procedure (IV). It is noticeable that the influence of gender and baseline testing decreases and that of age increases (indeed almost doubles) when the consistent procedure is applied.

Table 1: Results with and without instrumental variable procedures

	Without IV		**With IV**	
Coefficient for	**Estimate**	**s. e**	**Estimate**	**s. e**
Intercept	-0.0671	0.0520	0.0353	0.0611
Male gender dummy	0.102	0.0244	0.0758	0.0335
Centred age in months	0.0145	0.00379	0.0281	0.00828
Baseline 1st Principal Component	0.314	0.00775	0.211	0.0540

It is well known that if good instruments for the endogenous variables cannot be found, then the resulting estimates, although consistent, may be quite imprecise. In some cases standard errors can become so large as to make results uninterpretable.

The estimated standard errors produced by the IV procedures are substantially higher than those produced without it. It is a matter of judgement whether the price in imprecision is worth paying to secure the promise of consistency.

Conclusion

A solution to the problem of inconsistency caused by the presence of endogenous variables in a multilevel model has been proposed, based on instrumental variable procedures. The implementation of the consistent estimation method suggested has been made possible using the flexible macro facilities of *MLwiN*. An illustration of the method has been presented and the results contrasted with those produced when the problem of heterogeneity is ignored. It is possible that the availability of further background data might have further improved the precision of the estimates. Sound planning in data collection is therefore important.

References

Bowden, R.J. and Turkington, D. A. (1984). *Instrumental Variables*, Cambridge, CUP.

Crouchley, R. and Davies, R. B. (1999). A Comparison of Population Average and Random Effects Models for the Analysis of Longitudinal Count Data with Baseline Information, *Journal of the Royal Statistical Society*, Series A, **162** (3), 331-348.

Fielding, A. (1999). Why Use Arbitrary Points Scores? Ordered Categories in Models of Educational Progress. *Journal of the Royal Statistical Society*, Series A, **162** (3), 303-328.

Goldstein, H. (1986). Multilevel Mixed Linear Model Analysis using Iterative Generalised Least Squares. *Biometrika* **73**, 43-56.

Kiviet, J. F. (1995). On Bias, Inconsistency and Efficiency of Various Estimators in Dynamic Panel Data Models. *Journal of Econometrics*, **68**, 53-78.

Rice, N., Jones, A. and Goldstein, H. (1998). Multilevel Models where the Random Effects are Correlated with the Fixed Predictors: A Conditioned Iterative Generalised Least Squares Estimator (CIGLS). *Multilevel Modelling Newsletter*, **10** (1), 7-11.

Spencer, N. H. (1998) Consistent Parameter Estimation for Lagged Multilevel Models, *University of Hertfordshire Business School Technical Report 1*, UHBS 1998:19.

Spencer, N. H. and Fielding, A. (2000). A Comparison of Modelling Strategies for Value-Added Analyses of Educational Data, *University of Hertfordshire Business School Technical Report 2*, UHBS 2000:7.

Extra-binomial variation in logistic multilevel models - a simulation

Marita Jacob
Max-Planck-Institut for Human Development, Berlin
e-mail: mjacob@mpib-berlin.mpg.de

Introduction

Multilevel models for binary responses are widely used. The usual recommendation is to constrain the level one variance to 1.0 under the assumption of a binomial (or Bernoulli) distribution. It is also suggested that the variance at level 1 should be estimated from the data in a second step, thus allowing *extra*-binomial variation. In many applications, only the coefficients of the fixed effects are compared if there are considerable changes from the constrained to the unconstrained model. The estimate of the level 1 variance is often assessed by a "rule-of-thumb" that an estimate close to 1.0 indicates conformity to the binomial assumption (e.g. Yang et al., 2000).

We concentrate on a binary dependent variable Y_{ij} distributed as $(1, \pi_{ij})$ and conduct an examination of the estimate of binomial variation under different circumstances, using simulated data.

Extra-binomial variation

Assuming that the binary variable Y_{ij} is Bernoulli distributed, its mean and variance are given by $E(Y_{ij}) = \pi_{ij}$ and $Var(Y_{ij}) = \pi_{ij} (1-\pi_{ij})$. In applications, it can often be observed that the variance, given the theoretical mean, is greater or smaller than the expected value $\pi_{ij} (1-\pi_{ij})$. This is called extra-binomial variation, more precisely underdispersion if the variance is smaller and overdispersion if it is greater. Extra-binomial variation can occur for a number of reasons, unobserved heterogeneity and positive correlations between individual responses being the most common. Neglecting a relevant predictor, or insufficiently modelled cluster effects, may have serious effects on the estimate of the level 1 variance (Goldstein, 1995). Both mis-specifications are expected to result in overdispersion.

Extra-binomial variation can be estimated by introducing a scale factor. A variable z_{ij} is defined by $z_{ij}=[\pi_{ij} (1-\pi_{ij})]^{1/2}$. This variable z_{ij} is included in the random part as a scale factor for the level 1 residual e_{ij}. Constraining the level 1 variance, σ^2_e, to be 1, then leads to binomial variance for Y_{ij}. To estimate extra-binomial variation, σ^2_e is no longer constrained to 1, and the level 1 variance is estimated from the data.

Method and datasets

Simulation methods are used to distinguish the effects of small sample size and two model misspecifications: omitting a relevant predictor and neglecting a clustering variable.

Five models are compared in two separate simulations. As the estimation of the level 1 variance is of main interest, three models are estimated using a true variance components model. We concentrate on two level models.

Table 1 lists the different models specified in the first simulation for estimation. Model 1 is correctly specified, in model 2 the predictor X_1 is dropped and in model 3 a single level model is specified, neglecting the multilevel structure.

Table 1 Different simulation models I

Model	Logit function	Description
1	$logit_1(\pi_{ij})=\beta_0 X_{0ij}+\beta_1 X_{1ij}+u_{0j}$	correctly specified
2	$logit_2(\pi_{ij})=\beta_0 X_{0ij}+u_{0j}$	neglecting X_1
3	$logit_3(\pi_{ij})=\beta_0 X_{0i}+\beta_1 X_{1i}$	single level model

The fourth and fifth simulation models are based on a logistic multilevel model introducing a second random effect, a random slope at level 2. The fourth simulation model is correctly specified. A single level model is estimated as an alternative specification in the second simulation.

Table 2 Different simulation models II

Model	Logit function	Description
4	$logit_4(\pi_{ij})=\beta_0 X_{0ij}+\beta_1 X_{1ij}+u_{0j}+X_{1ij}u_{1j}$	correctly specified
5	$logit_5(\pi_{ij})=\beta_0 X_{0i}+\beta_1 X_{1i}$	single level model

The two sets of simulations were performed using MLn and *MLwiN*, with 500 datasets generated for each true model. The macros for generating, estimating and storing the results are available from the author. They are similar to those of Wright (1995), who kindly gave me access to his macros. For parameter estimation the non-linear macros of *MLwiN* discrete were used. Parameter values of β_0 and β_1 were set to 1.0 in data generating. The level 2 random effects u_{0j} and u_{1j} were assumed to be normally distributed residuals with a centred mean of 0 and a variance of 0.5. To avoid interpretational problems, the predictor X_1 has a standard-Normal distribution. X_0 is a column of 1s. Adding the fixed part and level 2 random effects gives the logit (π_{ij}). Finally, the values of the binary Y_{ij} variable are generated by a binomial distribution with parameter π_{ij}.

Different sample sizes were chosen. There are 20 balanced designs, varying four different numbers of level 2 units (J=10,25,50,100) by five different numbers of level 1 units (n=5,10,25,50,100) (see Table 3). (Also 20 moderately unbalanced designs were examined, leaving N and J constant but randomly distributing the number of level 1 units. The results do not differ substantially and they are not presented here.)

Table 3 Sampling design (balanced)

Sample	1	2	3	4	5	6	7	8	9	10
J	10	10	10	10	10	25	25	25	25	25
n	5	10	25	50	100	5	10	25	50	100
N	50	100	250	500	1000	125	250	625	1250	2500

Sample	11	12	13	14	15	16	17	18	19	20
J	50	50	50	50	50	100	100	100	100	100
n	5	10	25	50	100	5	10	25	50	100
N	250	500	1250	2500	5000	500	1000	2500	5000	10000

J number of level 2-units
$n = n_j$ level 1-units per level 2-unit
N sample size

To estimate the parameters, MQL1 followed by PQL2 was used (MQL2 for the single level models 3 and 5).

Results

For small samples ($N \leq 500$) non-convergence or improper solutions were observed for models 1, 2 and 4. In models 1 and 2 the lowest percentage of convergence was 60%, in model 4 only 41% datasets of the smallest sample (sample 1) converged. For $N \geq 500$ in all models more than 95% of the 500 datasets converged. Non-convergent estimations and improper solutions were omitted from the analysis.

In all two level models (model 1, 2 and 4) the level 1 variance is underestimated whereas in the single level models the estimates are very close to 1.0. In Table 4, the model-specific results are summarised by the mean of all estimates per model.

Table 4 Means of the level 1 variance estimates

Model	Mean
1	0.9459
2	0.9622
3	1.0024
4	0.8789
5	1.0146

Boxplots of the distributions of the 500 estimates (or less in the case of non-convergence or improper solutions) for each sample size are shown for models 1, 3 and 4.

Figure 1 Distribution of level 1 variance estimates, model 1

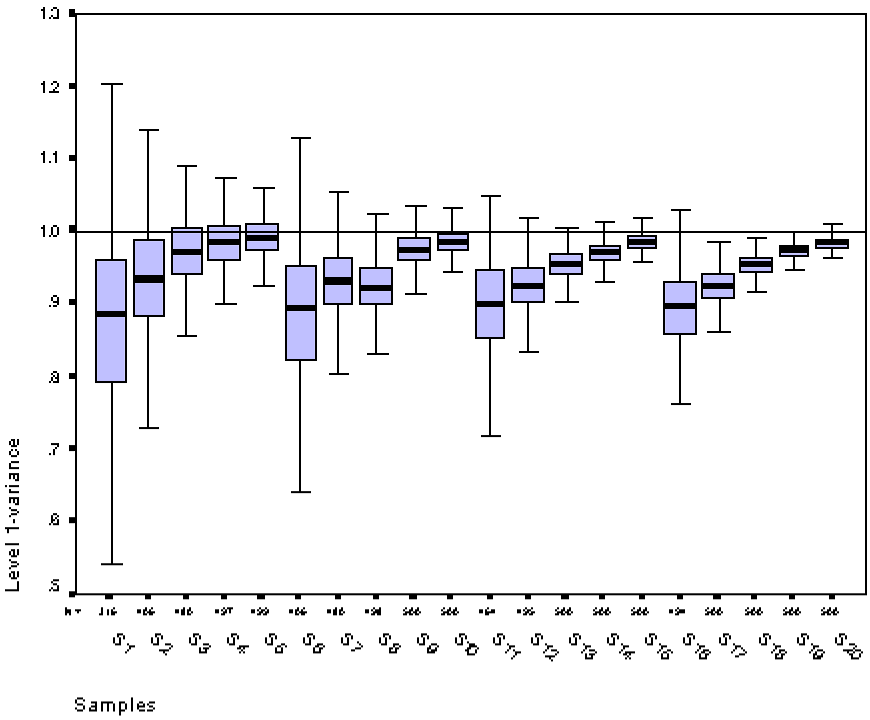


Figure 2 Distribution of level 1 variance estimates, model 3

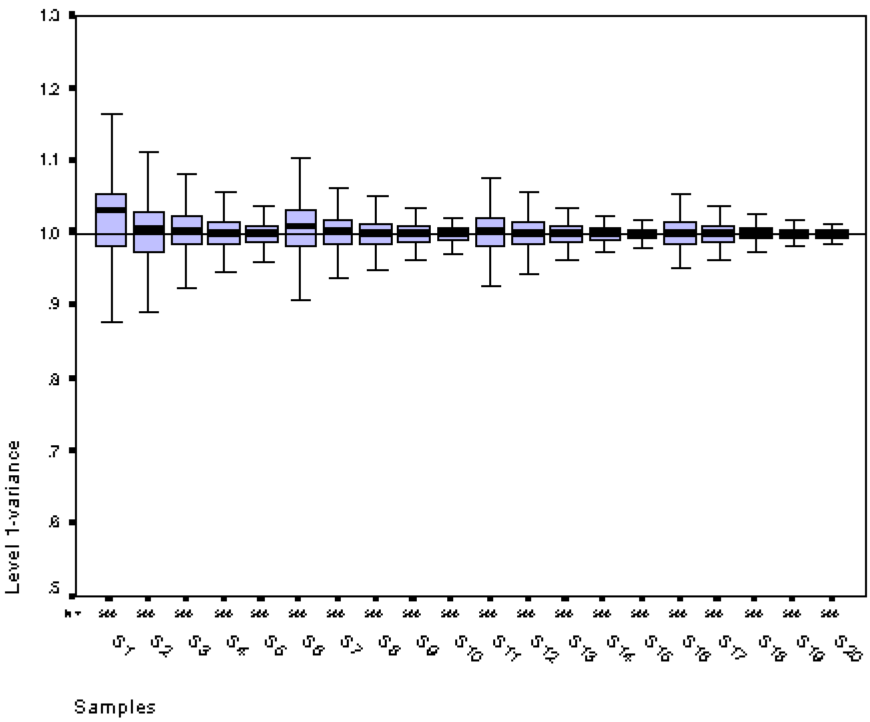
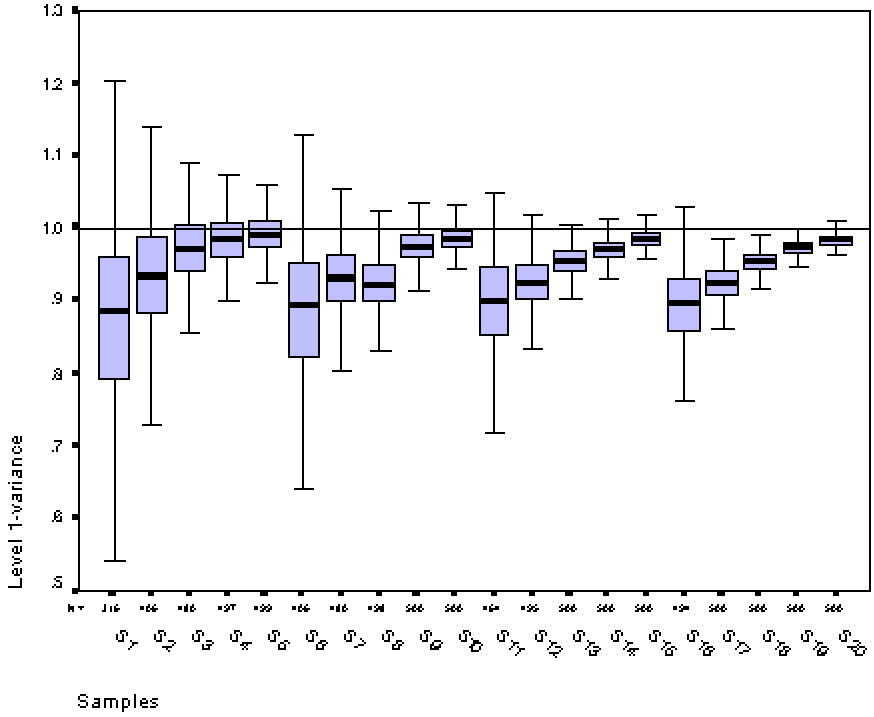


Figure 3 Distribution of level 1 variance estimates, model 4



Comparing these three models, the greatest deviation from 1.0 can be seen in the correctly specified model 1 (Figure 1). In samples with few level 1 units per level 2 unit the median is about 0.90. Model 3 gives estimates very close to 1.0, although the multilevel structure in the data is neglected (Figure 2). Neglecting the level 2 random effects results in a level 1 variance almost equal to 1.0. The estimates of the level 1 variance in model 2 are closer to 1.0 and the variation decreases considerably. Neglecting a predictor leads to better results because overdispersion raises the estimates closer to 1.0. Thus, an estimate very close to 1.0 seems to be an indication for the incorrect specification rather than for a correct binomial assumption.

In all models, increasing sample size improves the median of the simulations, and the variation of the estimates decreases. Given the number of level 1 units, by increasing the number of level 2 units the variation decreases. For J = 100 (samples 16 to 20) in model 1 almost all boxplots are completely below the line of reference, i.e. only possible outliers of the 500 estimates are above 1.0. Given the number of level 2 units, by increasing number of level 1 units the deviation of the median and the range of the estimates decreases.

The boxplots of the fourth and fifth model do not differ substantially from the others. Only the boxplots of model 4 are presented here (Figure 3). The deviation from 1.0 and also the range are considerably greater than in the variance components models. The median in samples with few level 1 units (n = 5) is less than 0.8. The greater the sample size, the lower the range. For large samples ($N \geq 1000$) the reference line at 1.0 is again not included in the boxplots. Model 5 is the only model in which overdispersion emerges but in absolute terms, the difference from 1.0 is only marginal. In small samples the median is about 1.05.

Summary

The objective of this study was to investigate the effects of sample size and model misspecification on estimating the level 1 variance for binary responses. The small simulation study shows that:

(1) in small samples, estimates around 0.8 can appear although the model is correctly specified under the binomial assumption, and

(2) estimates close to 1.0 do not allow us to conclude correct model specification. Estimating the level 1 variance in logistic multilevel models cannot be treated as a simple indication to test the assumed binomial distribution. In large sample sizes, the estimate is usually close to 1.0, but it cannot be used as an assessment of the distribution by statistical testing.

Model mis-specification leads to "improved" estimates, because overdispersion reduces the downward bias. The estimation procedure PQL2 gives downwardly biased results, so that "too close" estimates are rather an indication of an incorrect model. A variance larger than 1.0 emerges only when cluster effects are neglected. Then, at least, overdispersion can be used to detect model mis-specifications.

A further analysis of more complex models including more predictors, varying the level 2 variance and several random effects is necessary, to be sure if underdispersion is always found under a given binomial distribution.

Acknowledgement

This article is part of the author's diploma thesis, supervised by Prof. S.-M. Kuehnel and Prof. D. Krebs at the Institute of Sociology at the University of Giessen.

References

Busing, F. (1993) Distribution Characteristics of Variance Estimates in Two-Level Models, *Leiden Psych. Reports*, PRM 93-04, University of Leiden

Goldstein, H. (1995) *Multilevel Statistical Models*, 2nd ed., Arnold, London

Goldstein, H. & Rasbash, J. (1996) Improved Approximations for Multilevel Models with Binary Responses, *Journal of the Royal Statistical Society*, Series A, **159**, 175-177

Rodriguez, G. & Goldman, N. (1995) An Assessment of Estimation Procedures for Multilevel Models with Binary Responses, *Journal of the Royal Statistical Society*, Series A, **158**, 73-89

Wright, D. B. (1997) Extra-binomial variation in multilevel logistic models with sparse structures, *British Journal of Math. and Stat. Psychology*, **50**, 21-29

Yang, M., Heath, A. & Goldstein, H. (2000) Multilevel models for repeated binary outcomes: attitudes and vote over the electoral cycle, *Journal of the Royal Statistical Society*, Series A, **163**, 49-62

Project report: Application of advanced multilevel modelling methods for the analysis of examination data

Min Yang, Geoff Woodhouse
Institute of Education, London
m.yang@ioe.ac.uk

Supported by the Economic and Social Research Council and directed by Professor Harvey Goldstein at the Institute of Education, University of London, this project has entered the last year of its three-year period.

The main aims of the project are to develop existing multilevel methodology to handle efficiently ordered categorical responses and measurement error, and to provide substantive conclusions from an analysis of a very large cohort of students with GCSE and A/AS level examination results. The exam data were provided by the Department for Education and Employment.

Following are summaries of the main outcomes of the project.

Study I: Progress from GCSE to A and AS level: institutional and gender differences, and trends over time

In this part we study the relationship between results obtained in examinations for the General Certificate of Education at Advanced and Advanced Supplementary (A/AS) level and those obtained by the same students two years earlier in examinations for the General Certificate of Secondary Education (GCSE). Using comprehensive data on four cohorts examined between 1994 and 1997 totalling 696,600 students from 2,794 educational institutions, we build a multilevel, longitudinal model of student progress which takes into account the age and gender of the students and the type and location of the establishment they attended. We find that progress differs between men and women, and between students of different ages, and that the size of these effects depends upon prior performance. The average GCSE performance of the students in an establishment is a significant predictor of individual progress. Once establishments are matched on this measure, and students are matched on their own GCSE performance, the effects of most establishment types are substantially reduced: in particular, the average progress of students in maintained grammar schools does not differ significantly from that of students in maintained comprehensive schools. Using the new model, we find less stability over time in the usual residual estimates of the relative effectiveness of institutions than has been found in earlier studies. Despite the apparent simplicity of the measures used, the relationship between them is complex, and we argue that this complexity must be respected when judging institutions using value added procedures.

Study II: Multilevel ordinal models in the modelling of examination grades

In this part we concentrate on A level outcome in 1997 for the two subjects of Chemistry and Geography for substantive and methodological reasons. Both the point scores and exam grades of 30,910 students from 2,409 institutions for Chemistry and 33,276 students from 2,317 institutions for Geography are analysed. We fit normal multilevel models to the point scores, and ordinal multilevel models with a logit link to the grades for each subject. Comparing the precision of the parameter estimates associated with the effects of GCSE scores, gender, age of students, type of institutions and board of examination, we find virtually no difference between the two types of models. The estimated random effects of schools are comparable. The ordinal multilevel model is extended further to estimate the conditional distribution of the grade for each institution, assuming variable cut-points of the grades across schools. Several institutions are used to illustrate how the extended ordinal model gives us insights into the grade distribution, which is useful for school effectiveness research.

Study III: Analysing A/AS level mathematical scores using multilevel multivariate models

In this part we present the complexity of the data of A/AS level results for mathematics using students from 2,592 institutions in 1997. Multiple entries on different type of maths courses by a single student create multiple responses with strong dependencies between some of them, typically Main and Further maths, Pure and Applied maths. The exam entry of 59,369 from 52,587 students with between 1 and 4 entries for each, results in a highly unbalanced design. The AS level maths scores, with different distributions from the A-level scores, form another set of outcome variables. Students taking different combinations of mathematical subjects have different distributions on their A/AS level scores for the same mathematics course. We fit multilevel multivariate models to the data, showing how the complexity of the data can be modelled by explicitly specifying variance and covariance terms for different type of maths entry by A and AS results at student and school levels. We also show how to identify the student groups according to their choice on course combinations, and how problematic the model could be if failing to do so.

Study IV: Adjusting for measurement error

It has long been known that institutions can be differentially effective for students with different levels of prior achievement. In the context of A/AS levels this differential effectiveness may be modelled by random coefficients of (functions of) GCSE scores. But GCSE scores are subject to measurement error, and adjusting for this in the presence of a random coefficient is tricky.

We now have a method of adjustment, using MCMC estimation, which appears from simulations to be both practical in terms of processing, and effective in retrieving unbiased estimates of parameters in simple normal models. The next step is to apply this method to a subset of the A/AS level data set, assuming a normal response and using a function of GCSE scores derived from our earlier work as the GCSE predictor. In this way we shall show the effect on inferences of different assumptions about the extent of measurement error in the GCSE predictor. After that, we shall incorporate institutional averages as well as the student-level scores.

Papers from the studies will be made available by the end of the project in early 2001.

Applying multilevel models to university admissions grades: a note

Anna Cuxart-Jardí
Universitat Pompeu Fabra, Barcelona, Catalonia, Spain
e-mail: anna.cuxart@econ.upf.es

High-school graduates in Spain who want to enrol at a public university must take a set of examinations called *Pruebas de Aptitud para el Acceso a la Universidad* (PAAU). In Catalonia, the examinations are prepared, administered and scored by *Coordinació de les PAAU*, an institution funded by the seven public universities in Catalonia. *Coordinació* also maintain an extensive database of PAAU and high-school examinations grades and scores. The PAAU examinations have nine components in Catalonia. Five are compulsory: Philosophy, Spanish, Catalan, a Foreign language (English for the 90% of the students) and an Essay on a prescribed topic.

This note summarises an application of multilevel modelling to the joint analysis of these five compulsory subjects. The aim is to explore the correlational structure of the PAAU scores at student and school level. The data come from a random sample of 26 schools (1619 students, June 1993). More details can be found in Cuxart (1998) and Cuxart and Longford (1998).

A multivariate multilevel model for the analysis of association between subjects in PAAU examinations and explanatory variables like gender, curriculum and type of the school is specified:

$$y_{ij} = \beta_{1ij} z_{1ij} + \beta_{2ij} z_{2ij} + \beta_{3ij} z_{3ij} + \beta_{4ij} z_{4ij} + \beta_{5ij} z_{5ij}$$

$$\beta_{kij} = \mu_k + \sum_{r=1}^{m_k} \gamma_r x_{rij} + u_{kj} + \varepsilon_{kij}$$

z_{kij} are dummy variables (0/1) indicators of the responses y_{ij} (subject score); x_{rij} are explanatory variables; β_{kij} are random variables varying among students and among schools. Subscripts *i* and *j* stand for the student and school, respectively.

Table 1. Decomposition of overall variation of the five compulsory subjects' records in PAAU exams. Variances are on the diagonal and correlations are below.

	Catalan	Spanish	Philosophy	Foreign L.	Essay
Between schools					
Catalan	0.69				
Spanish	0.45	0.44			
Philosophy	0.23	0.40	0.45		
Foreign L.	0.41	0.62	0.27	0.33	
Essay	0.11	0.04	-0.08	0.21	0.51
Within schools					
Catalan	2.74				
Spanish	0.32	2.05			
Philosophy	0.23	0.27	2.57		
Foreign L.	0.25	0.32	0.22	3.24	
Essay	0.23	0.23	0.19	0.25	1.87

We find that (Table 1):

- Some covariate effects in the two-level model (scores and students) are not present in the three-level model. They were caused by characteristics of specific schools.
- There is a significant variation between schools in subject means for all subjects.
- The between school correlations between Essay and the rest of the subjects are low.

Acknowledgement

The study was supported by research grant DGES PB96-0300, Spanish Ministry of Education.

References

Cuxart, A. (1998). *Models estadístics en avaluació educativa: les proves d'accés a la universitat.* PhD thesis, Universitat Politècnica de Catalunya, Barcelona.

Cuxart, A. & Longford, N. T. (1998). Monitoring the university admissions process in Spain. *Higher Education in Europe*, **XXIII**, 385-396.

Review of 'An Introduction to Multilevel Modelling Techniques'

R H Heck & S L Thomas. Pp xiii & 209.

Mahwah, NJ: Lawrence Erlbaum Associates, 2000

M J R Healy

Institute of Education, London

With the growing popularity of multilevel modelling methodology, introductory textbooks are starting to appear. Those by Kreft and DeLeeuw and by Snijders and Bosker are becoming well-known and the present book may be regarded as a competitor in the same field. It has two distinguishing features – the authors work in the field of organisational research rather than in education, and as much or more emphasis is placed upon aspects of multivariate modelling as upon the more familiar univariate methods.

The book opens with an introductory chapter specifically aimed at organisational research, a field rich in hierarchies ready-made for a multilevel approach. This chapter classifies data as univariate or multivariate and as single-level or multilevel. It is followed by a second more detailed account of multilevel methods, again contrasting the multilevel approach with less satisfactory single level approaches. This chapter contains an account of shrinkage, attributed to a Bayesian approach (the qualifier 'empirical' is smuggled in without explanation). It also contains a discussion of power (concluding little more than that its calculation in a multilevel setting is difficult) and of structural equation modelling including multilevel path and factor analysis.

Chapter 3 is simply an account of OLS regression. This is presented at considerable length although many of the practical pitfalls remain unmentioned. The example studied contains a very obvious and influential outlier on which no comment is made.

Chapter 4 is the only one devoted to the standard univariate multilevel model with potentially random intercepts and slopes. The computer program used is HLM (*MLwiN* and other packages are mentioned in the introduction to the book) and space is given to describing the input requirements for this. An example is given relating to a sixth-grade reading test score using as predictors gender, socio-economic status (as a 0-1 variable) and a previous third grade score. The model is two-level (students nested within schools), and the predictors are initially taken to be fixed. The slope for previous test score is 0.15 and, surprisingly, this is interpreted as implying that 'on average, students' sixth grade test scores are 15% higher than their score at third-grade'. Subsequently previous score is allowed to become random and a school level variable of median parental income is introduced. The authors cannot be given responsibility for the vagaries of HLM, but it is still a little daunting to find in their text that the overall mean test score given to no fewer than 9 significant figures. The program provides estimates of the random parameters (labelled 'variance components') without standard errors. The estimated variances are accompanied by chi-squared values giving p = 0.000 (sic). The interpretation of these in the text appears to me to be dubious.

The next three chapters (more than half the book) are devoted to structural equation methods. Chapter 5 is an overview of single level methodology, chapter 6 deals with confirmatory factor analysis and chapter 7 with multivariate structural equation modelling. I have to confess that as one unfamiliar with this rather specialised aspect of multivariate analysis I found much of the material (including many pages of LISREL input and output) almost incomprehensible. It was not encouraging to find what should be the routine calculation of within-group and between-group covariance matrices described as 'somewhat problematical'. It is regarded as a virtue of LISREL that it routinely provides ten different indicators of goodness of fit, and the fact that the value of chi squared reflects the sample size is taken to be an 'undesirable property'.

A strong impression left on reading this book is that the authors have not given enough thought to defining their readership. Readers requiring an elementary introduction to simple regression or needing to be told that 'researchers often set the α region for rejection of a null hypothesis at 0.05 or 0.01' are unlikely to make much of the matrix algebra and multiple suffix notation in the later chapters. It is worrying to read such statements as 'the sample data may depart from normality and therefore may not represent the population accurately' or 'sufficient sample sizes are required to determine whether parameters are indeed significant'. There are a number of printers' errors, some of them in simple mathematical formulae.

Statistics is a difficult subject and producing a useful introductory text is not an easy task, much the less so when the special topic is as complex as multilevel analysis. In spite of its non-standard coverage, this book cannot be recommended.

Some Recent Publications Using Multilevel Models

Barbosa, M. F. (2000). Discrete Response Multilevel Models for Repeated Measures: An Application to Voting Intentions Data. *Quality & Quantity*, **34**, 323-330.

Burton, P., Gurrin, L. & Sly, P. (1998). Extending the simple linear regression model to account for correlated responses: an introduction to generalised estimating equations and multilevel mixed modelling. *Statistics in Medicine*, **17,** 1261-1291.

Goldstein, H., Yang, M., Omar, R., Turner, R. & Thompson, S. (2000). Meta-analysis using multilevel models with an application to the study of class size effects. *Applied Statistics*, **49**, 399-412.

Hox, J. (2000). Multilevel Analyses of Grouped and Longitudinal Data, in: Little, T. D., Schnabel, K. U. & Baumert, J. (eds.), *Modeling longitudinal and multilevel data*. Mahwah, NJ, Lawrence Erlbaum Associates.

Kef, S., Hox, J.J. & Habekothe, H. T. (2000). Social networks of visually impaired and blind adolescents. Structure and effect on well-being. *Social Networks*, **22**, 73-91.

Lindsey, J. K. & Lambert, P. (1998). On the appropriateness of marginal models for repeated measurements in clinical trials. *Statistics in Medicine*, **17,** 447-469.

Loonen, M. J. J.E., Bruinzeel, L. W., Black, J. M. & Drent R. H. (1999). The benefit of large broods in barnacle geese: a study using natural and experimental manipulations. *Journal of Animal Ecology*, **68**, 753-768.

Mealli, F. & Rampichini, C. (1999). Estimating binary multilevel models through indirect inference. *Computational Statistics & Data Analysis*, **29**, 313-324.

Plewis, I. (2000). Evaluating Educational Interventions Using Multilevel Growth Curves: The Case of Reading Recovery. *Educational Research and Evaluation*, **6**, 83-101.

Rampichini, C. & Schifini, S. (1998). A Hierarchical Ordinal Probit model for the analysis of life satisfaction in Italy. *Social Indicators Research*, **44**, 5-39.

Spikman, J.M., Timmerman, M.E., van Zomeren, A.H. & Deelman, B.G. (1999). Recovery versus retest effects in attention after closed head injury. *Journal of Clinical and Experimental Neuropsychology*, **21** (5), 585-605.

Tymms, P., Merrell, C. & Henderson, B. (2000). Baseline Assessment and Progress during the First Three Years at School. *Educational Research and Evaluation*, **6** (2), 105-129.

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

We have used the format from previous newsletters for this electronic version. If, however, you would prefer a single column layout please let us know. Any other comments on the Newsletter would be welcomed.

MLwiN Clinics in London 2000

Tuesday 5 September 2000
Tuesday 10 October 2000
Tuesday 7 November
Tuesday 5 December 2000

at

Multilevel Models Project
11 Woburn Square, London WC1A 0SN

Contact Min Yang for appointments
Tel: +44 (0) 20 7612 6682/6688
Email: m.yang@ioe.ac.uk

If you would like to be added to the e-mail list, please send your e-mail address stating that you wish to subscribe to the newsletter, to a.burch@ioe.ac.uk