A review of random effects modelling in SAS (release 8.2)

Min Yang
Queen Mary, University of London
m.yang@qmul.ac.uk

1. Introduction

1.1 Background

For random effects analysis of variance, SAS has had procedures such as PROC VARCOMP and PROC GLM. Both procedures produce expected mean squares leading to the traditional ANOVA estimates of variance components. The GLM procedure (see below) also fits standard linear models accounting for random effects. For modelling general covariance structure, nested structures and time series, the procedures CALIS, NESTED and ARIMA as well as AUTOREG can be used. Although each of these procedures worked well in doing its own job, their capacity limitations make the extension to random effects modelling difficult. For example, the VARCOMP procedure does not produce estimates of covariate effects; the GLM procedure treats the effects specified in the RANDOM statement as fixed; the NESTED and CALIS procedures do not allow fixed effects in the model; and the ARIMA as well as AUTOREG procedures do not allow random components. The launch of PROC MIXED in SAS release 6 overcomes many of the limitations of these procedures. This procedure mixes functionality of the fixed effects models of procedures such as GLM and the random effects models of VARCOMP and NESTED together. It is a generalisation of PROC GLM with much flexibility in random effects modelling. After two more releases of SAS, the MIXED procedure has been improved and used widely. In addition to the MIXED procedure for fitting linear regression models to continuous responses, PROC NLMIXED is also available to fit mixed models for non-linear relationships between the response and covariates together with the NLINMIX and GLIMMIX macros that are based on different estimation methods. PROC NLMIXED is used to fit Binomial and Possion models as well as more general models which may be specified by the users, based on directly maximizing an approximate integrated likelihood.

Our review of SAS for random effects modelling focuses on the MIXED and NLMIXED procedures in release 8.2 (2001). A few other relevant procedures are briefly explored.

1.2 Software and hardware requirements

Supplied with the software package is a 29-page document, System Requirements with full details on all aspects of the system requirements. The System Requirements Wizard in the System Setup program is useful for ensuring that your computer meets the minimum system requirements before installing the SAS System. In brief, release 8.2 of the SAS System is supported on Microsoft Windows 95, 98 and ME operating systems. Windows 95 must be updated with additional Year 2000 software updates by Microsoft. Windows 98 must be updated with Service Pack I or to Windows 98, 2nd Edition. It is also supported on Windows 2000 Professional and all Windows 2000 Service Editions. It can also run under Windows NT, Version 4.0 updated with Service Pack 4 or higher (excluding Service Pack 6), plus additional Year 2000 software updates by Microsoft.

The minimum hardware requirement includes an Intel or Intel-compatible Pentium with floating point unit or Maths Coprocessor, 64 MB or more of RAM on PCs (256 or more on Servers) and a CD ROM disc drive.

1.3 Data input/output functionality

As one of the biggest statistical packages worldwide, the SAS System has good facilities under windows for data input and output.

The standard choice for data input is the Open tool in the File dropdown list which brings ASCII data files or other text files into the Program Editor. The data file can be delimited by spaces, commas, tabs or an other user-specified format. Variable names and missing values coded as dot or other symbols can be read in without problems. The second approach, the Import tool in the File list, will turn an ASCII data file delimited by spaces, commas or tabs into a SAS System file. One can view the data using PROC PRINT. Variable names and other characters in the data file can be imported at the same time. The third approach would be the standard Copy/Paste tools for Windows. One can copy text or ASCII data from Excel directly into the Program Editor. The SAS System can work with multiple data files at the same time as long as these files are named differently. The fourth approach is to use the INFILE statement under the DATA procedure, which is useful for working on very large data sets and running analysis in a batch file based on SAS syntax. Using this approach, one only needs to specify the full path to where the ASCII data file is stored, followed by variable names. For small data, one can always use the fifth approach which is to key in the data directly into the Program Editor.

For data output, the Save As tool can be used to output files with the file types LIST, SAS, Log, Data or RTF from any of the SAS interface windows, Log, Program Editor and Output. Using the Export tool in the File list, one can output data either in standard data formats with space, comma or tab delimitors, or in user-defined formats. Another way of outputting data is to use the statements OUT or OUTFILE available under some procedures. One can output parameter estimates, predictions as well as data. Results in the form of texts or tables which appear in the Output window can be copied/cut and pasted into a Word or other word processor files.

1.4 Interface features: data manipulation, commands, menus

A statistical task such as model fitting can be carried out through SAS syntax consisting of procedures. In SAS, a statistical *procedure* comprises a collection of *statements*, where each statement can be followed by many *options*. This review will focus on the use of syntax for fitting a variety of random effect models.

Within the system, one mainly works with three windows named Program Editor, Log and Output. The Program Editor allows users to input data and to write syntax for the statistical tasks to be submitted. Once a task has been run, the syntax executed will appear in the Log window. If errors in the syntax are detected by the system, they will be reported in this window too. Results of analysis will be displayed in the Output window. The contents of any of the three windows can be copied into Word or Notepad for editing. They can also be saved as text files.

Before carrying out statistical analysis, the DATA procedure should be called to read and manipulate data. Programming statements are available in the DATA step which enables complex data processing or data reconstruction. In this review, we use only the necessary and basic statements in this step in order to prepare data for specific random effect models to be fitted.

2. Standard tools for random effects modelling

2.1 A brief check of models

The current SAS System can fit many types of models in the random effects modelling domain. Other terms for random effects models that are also used in this review are multilevel models

(*MLwiN*, *SYSTAT, LISREL*) or hierarchical models (*HLM*). The following are other terms used in SAS documents which differs from those used in other packages.

Other terms	SAS terms
Level identification	Subject
Random parameters	Covariance parameters
Categorical variable	Class variable
Residuals at higher levels	Random effects
High level residual var-covariance matrix	G matrix
Level 1 residual var-covariance matrix	R matrix

In Table 1 attached, we list random effect models that can be fitted by SAS. As described above, two standard procedures, MIXED and NLMIXED, are used in such models. In this review, we shall explore mainly these two procedures by models.

2.2 Tools for statistical inference and model fit

Table 2 Tools available for inference and model fit

Type of response	Statistics for overall fit	Tests for fixed effects	Tests for covariance parameters	Other specific features
Normal	-2 log likelihood AIC AICC BIC	t-test (default) F-test (default) Chi-squared (request)	Wald test (default)	An option to compute a specific inflation factor to account for the downward bias by using approximate F and t tests
Generalised Linear model	-2 log likelihood AIC AICC BIC	t-test (default) Confidence interval (default)	t-test Confidence interval (default)	t-test for difference of fixed parameters (request)

There are three-types of F-test. By default type III results are reported. It is not clear from the manual when the other two should be considered as appropriate. For all tests, the degrees of freedom (d.f.) for each parameter estimate in the model may not be the same. For example, the cluster level variable would have fewer degrees of freedom than that of an individual level variable, and balanced data would have different degrees of freedom to unbalanced data. The two procedures have statements which allow users to specify the d.f. for each parameter estimate in the model. For Normal responses the MIXED procedure also has five methods to calculate approximate degrees of freedom. The choice of methods may affect the speed of model convergence.

3. Model specifications — Basic models

In this section, we explore some basic multilevel models that can be fitted using standard procedures. The models include Normal models, logistic models for binary data, the Poisson models for count data and repeated measures models for data with a two- or three-level hierarchical structure. As far as capacity and efficiency of estimation procedures are concerned, we shall explore the syntax used to specify the models, information on model estimates and convergence times. All models are run on a PC computer with Window NT system, 433 Mhz processor and 398 Mb RAM.

3.1 Two-level Normal models

The data set to be used is the example which appears in the user's guide to MLwiN (Rasbash et al. 2000). It consists of 4,059 students (level 1 units) nested within 65 schools (level 2 units). The outcome is their examination score at age 16 (EXAM). A key covariate is the prior London Reading Test score (STANDLRT, x_1) taken at age 11. Both the outcome and the reading scores were standardised with zero mean and unit variance and in addition the outcome score was Normalised.

Another student level variable is gender (GENDER, x_2: code 1 for girl and 2 for boy). Also considered is the school level variables are school gender (SCHGEND, coded 1 for mixed schools, 2 for boys and 3 for girls schools respectively). Before fitting models, the data are assumed to have been read into SAS and the named data file is EXAM.

Five models are fitted, each one on extension of another.

Model **A** is a variance components model with fixed effects for all three covariates. For the variable school gender of three categories, two parameters are estimated in the model. Only the intercept is allowed to have random effects u_{0j} among schools. The model can be written as

$$y_{ij} = \beta_{0ij} + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3j} + \beta_4 x_{4j} \qquad (1)$$

$$\beta_{0ij} = \beta_0 + u_{0j} + e_{0ij}$$

The β s in (1) are fixed effects, and var(u_{0j}) and var(e_{0ij}) are two covariance parameters to be estimated. This model shows the usage of the basic statements in the procedure MIXED for fitting the simple two-level model.

Model **B** is an extension of Model **A,** including an interaction between the two student level variables, London Reading Score and student gender. This illustrates the ease with which the SAS tool can be used to extend the fixed part of the model.

Model **C** extends Model **B** further by allowing β_1 the effect of a student's intake, to vary randomly among the population of schools:

$$y_{ij} = \beta_{0ij} + \beta_{1j} x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3j} + \beta_4 x_{4j} + \beta_5 (x_{1ij} \times x_{2ij}) \qquad (2)$$

$$\beta_{0ij} = \beta_0 + u_{0j} + e_{0ij}$$

$$\beta_{1j} = \beta_1 + u_{1j}$$

At the school level, a full variance - covariance structure for the two sets of random effects u_{0j} and u_{1j} is fitted, leading to a 2×2 covariance matrix. This type of covariance structure is termed as *Unstructured* (UN) in SAS, which by default assumes a multivariate Normal distribution for the two random variables. SAS also allows independent distributions for the two sets of random effects by specifying the *Banded Main Diagonal* (UN1) structure, where no covariance between the intercepts and slopes is estimated. In addition, there are more than 20 covariance structures including 1st order *Auto-Regression* AR(1), 1st order *Auto-Regression and Moving Average* ARMA(1,1), *Toeplitz*, and *Spatial Power,* which can be fitted under the TYPE option. We will explore some of these features in later sections of the review.

Model **D** extends **C** by changing the assumption about the residual variance at level 1. Instead of estimating a single variance term for residuals e_{0ij}, we assume in this model a non-constant residual variance difference between boys and girls, and fit separate variances for each gender. This model is fitted by using the REPEATED statement jointly with the GROUP option. Under the REPEATED statement, the TYPE option can also be used to allow the level 1 residual variance to be constructed as functions of other variables.

Model **E** is another version of Model **D**. Instead of partitioning the residual variance by gender, **E** models the residual variance as an exponential function of the student intake variable as follows

$$\text{var}(e_{ij}) = \sigma_e^2 diag[\exp(x_{1ij}\delta)] \qquad (3)$$

where δ is a parameter to be estimated.

The syntax and model estimates, as well as convergence times are listed in Table 3.

As we see, the basic statements in PROC MIXED for fitting those models are CLASS, MODEL, RANDOM, TYPE (option), REPEATED and SUBJECT (option). CLASS declares the categorical variables including the identifications for levels. Any covariate specified as a class variable with integer codes $1, 2, ... g$ and appearing in the MODEL statement will be fitted as $g - 1$ dichotomous variables, contrasted with a reference group. In the MIXED procedure, the reference group is the last code by default, i.e. code 2 (boys) for the individual gender and code 3 (boys school) for the school gender in our case. The MODEL statement specifies the fixed part of the model; the RANDOM statement with the TYPE option specifies the covariance structure at level 2; the REPEATED statement with the TYPE option specifies the covariance structure at level 1. The SUBJECT option indicates the block or cluster identifier. The RANDOM statement allows fitting of random slope models, either structured or unstructured, and the REPEATED statement allows modelling of the level 1 residuals, using many types of structures defined by the TYPE functions. We explain how PROC MIXED can be used to fit 3-level models below.

There are other useful features in PROC MIXED. The following are some examples:

- Random parameters at any level can be constrained to equal a known value using the PARMS statement with the HOLD option. For example, to fix the variance for girls and boys in model D in Table 3 to equal 0.55, one extra syntax line would be for example

    PARMS (0.08369) (0.02065) (0.01545) (0.55) (0.55) / HOLD = 4,5;

- In some situations where the variance matrix may become negative definite, one can set a lower or upper boundary on the random parameters to avoid computational problems. This can be done using options LOWERB=<value list> or UPPERB=<value list> under the PARMS statement.

- To carry out custom hypothesis tests on any fixed parameters or random parameters, the statement CONTRAST with options for F or χ^2 tests is available.

- To obtain the empirical best linear unbiased predictor of level 2 residuals or random effects, one can use the option SOLUTION (or S in short) under the RANDOM statement. Together with each residual estimate, the comparative estimates of the residual variance (using the terminology of MLwiN) is reported.

- There are options for outputting the empirical sandwich estimator for the variances of fixed effects, for displaying the variance-covariance matrix of the fixed and the random effects, for information on each subject (block), and for specifying the degrees of freedom for hypotheses tests.

- Up to 8 spatial covariance structures for level 2 or level 1 random effects can be specified using the TYPE option.

In combination, these statements and options provide easy-to-use tools for fitting Normal multilevel models to data with a two-level structure.

3.2 Three-level models for a Normal response

It is possible to fit three- or more than three-level Normal models using PROC MIXED, although the computational methods, which use the sweep operator and the Cholesky decomposition, can be rather inefficient when dealing with large datasets. This issue is clearly explained in the SAS manual in the chapter for PROC MIXED, with estimates of the computational speed of algorithms given numbers of variables, observations, random parameters, clusters and maximum observations per cluster. The covariance parameters at levels 3 and 2 can be defined using the RANDOM statement repeatedly, or the RANDOM and REPEATED statements jointly.

Sometimes one may be only interested in fitting a variance components model with many nested levels, unbalanced data, and no covariates. In this case, PROC NESTED is a useful tool to obtain estimates for each component of the total variance. This procedure runs very fast even on a large data set.

For illustration, we fitted models on A/AS-level examination data that consist of 1997 Chemistry on 31022 individuals from 2280 schools in 131 Local Education Authorities (LEA) in England (Yang et al. 2002). The explanatory variables are an intake score, the average GCSE score of the individual, their gender and age.

We have fitted only three variance component models with and without the fixed effects of the covariates. The syntax, model estimates and convergence times are given in Table 4.

For the model with variance components only, PROC NESTED is very quick, taking less than 1 second to converge. To add the fixed effect of the intake score to the model, PROC MIXED is used, since PROC NESTED does not allows covariates. Having added one fixed effect to the model, the MIXED procedure took more than 3 minutes to estimate the model using the RANDOM/REPEATED combination. The REPEATED statement specifies the middle-level random terms. However it is not straightforward to obtain the full variance-covariance structure for both random intercepts and slopes. To do this you replace REPEATED by the RANDOM statement in order to obtain the full variance-covariance structure on the random effects of intercepts and slopes at both LEA and school levels. In doing so, however, the results in Table 4 show how slow the procedure is, compared to the previous model fitted using the statement REPEATED.

It may be possible to write SAS program macros to improve the efficiency of the estimation procedure for three-level problems on large data sets. But we shall not explore this in the review.

3.3 Two-level models for binary/binomial data

Two-level binary and binomial data can be modelled using PROC NLMIXED.

The procedure uses adaptive Gaussian quadrature, which allows a number of choices for the first or second order optimisation. The default is the first order optimization. Overall, this procedure allows 73 options that are related to computational aspects in fitting nonlinear models. There are clear descriptions and definitions of the options in the online help system (Chapter 46, Statistical procedures). On the one hand, these options could be useful in gaining access to the core of the algorithm, since they allow specifications of optimization and derivatives, convergence criterion functions as well as method of approximation. On the other hand, it can be very difficult for the majority of users to make a choice without knowledge of the computational details and techniques in modelling nonlinear data. The procedure also consists of 11 statements for model specification.

To illustrate how the procedure works, we fitted models to binary data from the 1989 Bangladesh Fertility Survey (Huq & Cleland, 1990). The data are a sub-sample of 1934 women grouped into 60

districts. The outcome variable is use of contraception (y_{ij}) which equals 1 for using contraception and 0 otherwise. Three covariates are considered: age at survey centered at the sample mean (x_{1ij}); type of region of residence (x_{2ij}) which equals 1 for urban and 0 for rural; and number of living children (0=none, 1=one, 2=two, 3=three or more), represented by three dummy variables for the last three categories (x_{3ij}, x_{4ij} and x_{5ij} respectively).

Data have to be clustered according to the SUBJECT variable, districts in this example, by sorting the data by district prior to calling PROC NLMIXED. Unlike PROC MIXED, there is no CLASS statement available within this procedure, and the categorical covariate 'number of living children' is handled by creating three dummy variables in the DATA procedure beforehand. Good starting values, specified by the PARMS statement, are crucial to achieve convergence.

We firstly fitted a variance components model allowing only random intercepts at level 2, using both logit and probit link functions. We then fitted a model which allows both random intercepts and random slopes for the effect associated with the type of region of residence at level 2, using the default settings which produces 1st order ML estimates. The results are in Table 5. The SAS syntax for fitting these models is presented at the bottom of the table.

PROC NLMIXED allows only one RANDOM statement, which limits nonlinear models to data with two levels. For advanced users, the programming language SAS/IML could be used to write macros to fit models to data with more than two levels, but this would not be an easy task for most users. Alternatively, the GLIMMIX macro may be used. It is essentially PQL1 estimation (see summary tables).

To evaluate overall goodness of fit, the default Fit Statistics include -2log-likelihood, AIC, AICC and BIC values. For each parameter estimate in the model a t-test value and a confidence interval based on the tail probability for given degrees of freedom are displayed.

We can also see that SAS, aML and STATA produced similar results for the -2log-likelihood values for these models, since these packages all use the Gaussian Quadrature integration method that produces maximum likelihood estimates. The likelihood values produced by MLwiN using the IGLS/RIGLS estimation procedure of quasi-likelihood are less accurate than others for binary models.

For binomial data where the outcome is a proportion p based on n observations for each case, the distribution BINOMIAL (n, p) can be used for. Models can be fitted to count data using POISSON (λ) distribution, where λ is the mean. The general function GENERAL(ll) specifies a general log-likelihood function for other nonlinear models such as an ordinal model with random effects, and a Normal distribution NORMAL(m, v) can be used to fit general nonlinear models to a continuous response. We shall review these functions in the following sections.

3.4 A two-level model for count data

For count data with a two-level hierarchy, Poisson models can be fitted using PROC NLMIXED in the same way as two-level logistic models. Here we illustrate the use of this procedure on the malignant melanoma mortality data from 354 counties within 78 regions within 9 European countries. The data consist of observed deaths and expected deaths due to malignant melanoma, which produce the standardised mortality rate (SMR), *(observed deaths) / (expected Deaths)*. One important environmental variable that might be associated with the mortality rate is the county level UV radiation exposure. A detailed analysis is presented in Langford, Bentham and McDonald,

(1998). We treat the data as having a two-level structure: counties nested within regions, due to the small number of countries.

The simple variance components model is

$$\frac{y_{ij}}{E_{ij}} = \lambda_{ij}, \qquad \lambda_{ij} = \exp(\beta_0 + u_{0j} + \beta_1 x_{1ij}), \qquad u_{0j} \sim N(0, \sigma_u^2), \tag{4}$$

where y_{ij} is the observed death count in the ith county from the jth region which is assumed to follow a Poisson distribution with mean λ_{ij}, E_{ij} is the expected death count, x_{1ij} is the UV exposure measure, and u_{0j} are the random effects for regions.

This model is easy to fit, and the SAS syntax is as follows.

```
Proc nlmixed data=MMM;
Parms b0=-0.01 b1=-0.03 sigma2=0.1;
y = offset +b0*intercept+b1*uvb+u;
Lambda = exp(y);
Model death ~ Poisson (lambda);
Random u ~ normal (0,sigma2) subject=region;
Run;
```

The term *offset* is the logarithm of the expected number of deaths. Starting values are given under the PARMS statement. The results for the NLMIXED and other packages give similar results.

Random slopes models can be fitted in a similar way to that illustrated for a binary outcome in Table 5.

It is not clear from the manual whether any facility is available for detecting over or under dispersion for generalised models such as this and the models in the previous section, although the SAS macro GLIMMIX provides such a tool (Guo & Zhao, 2000)

3.5 Models for repeated measures data

Fitting models to repeated measures data, with either balanced or unbalanced observations among subjects, is straightforward using PROC MIXED with SUBJECT = individuals at level 2 and occasions at level 1. Time series models can be fitted to the level 1 residuals over occasions within subjects, specified by the TYPE option under the REPEATED statement. The type of time series functions include the 1st order *Autoregressive AR(1), Heterogeneous AR(1)* called ARH(1) and *First-Order Autoregressive Moving Average ARMA(1,1)*. The data have to be balanced with a fixed set of time intervals.

The estimation of time series models using PROC MIXED is illustrated using the data consisting of the height (cm) of 26 boys between the ages of 11~13 in Oxford over 9 occasions approximately 0.25 year apart. The data are balanced. The same data were used by Goldstein, Healy and Rasbash (1994) to describe some general multilevel time series models which allow for the autocorrelation structure between the height measurements. Here an AR(1) model is fitted and the results are compared to those produced by MLwiN macros.

Model A fits a 4th degree polynomial function to the overall growth, with quadratic function fitted to each individual's growth. In Model B, the AR(1) structure is fitted to the level 1 residuals, and sin and cosine terms are added to the fixed part of the model.

The model specifications and syntax are as follows. In model B, sine and cosine functions are fitted in the fixed part. Model estimates are listed in Table 6.

Models   Basic Syntax

A: $ht_{ij} = \beta_{0j} + \beta_{1j}t_{ij} + \beta_{2j}t_{ij}^2 + \beta_{3j}t_{ij}^3 + \beta_{4j}t_{ij}^4 + e_{ij}$

$\qquad \beta_{0j} = \beta_0 + u_{0j}, \ \beta_{1j} = \beta_1 + u_{1j}, \ \beta_{2j} = \beta_2 + u_{2j}$

$$\begin{pmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{pmatrix} \sim MN(0, \Omega_2), \ e_{ij} \sim N(0, \sigma_e^2)$$

```
Proc mixed method=ML;
Class boys occasion;
Model ht=t t*t t*t*t t*t*t*t/s;
Random int t t*t/type=un sub=boys;
Run;
```

B: $ht_{ij} = \sum_{h=0-2} \beta_{hj}t_{ij}^h + \sum_{h=3,4} \beta_h t_{ij}^h + \beta_5 \sin_{ij} + \beta_6 \cos_{ij} + e_{ij}$

$\qquad \beta_{hj} = \beta_h + u_{hj}, \ h = 0,1,2$

$$\begin{pmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{pmatrix} \sim MN(0, \Omega_2), \ e_{ij} \sim N(0, \Omega_e^2), \ \Omega_e = \sigma^2 \rho^{|i-i'|}$$

```
Proc mixed method=ML;
Class boys occasion;
Model ht=t t*t t*t*t t*t*t*t sin cos
/s;
Random int t t*t/type=un sub=boys;
Repeated /type=ar(1) sub=occasion;
Run;
```

4. Model specifications — other random effects models

4.1 Random effect models for a multiple categorical response

In this section we consider ordinal multinomial models for ordered responses and nominal multinomial models for unordered responses.

As an example, a sub-sample of the British Social Attitudes Survey, a panel study, is considered (McGrath, K. and Waterton, J, 1986). The data consist of 264 adult respondents in 54 districts who completed interviews in 1983, 1984, 1985 and 1986. The outcome of interest is a total score, ranging from 1 to 7, calculated as the number of positive answers to seven questions about whether they supported or opposed a woman's right to have an abortion under different circumstances. The score is arranged to reflect attitudes from the most restrictive (code 1) to the most permissive (code 7) towards to a woman's right to have an abortion. The data are structured as years nested within respondents within districts. There were many covariates measuring the background and socio-economic status as well as the political party preference of the respondent. For the purpose of illustrations, we consider only the religion of the respondent: Roman Catholic=1, protestant/Church of England=2, others=3 and none=4 for illustration purpose.

We shall ignore the district level as the NLMIXED procedure can only fit random effect models to two-level data.

Denote the cumulative probability by $\gamma_{ij}^s = p(y_{ij} \le s)$ ($\sum_{s=1}^{7} \gamma^{(s)} = 1$), the simplest model for ordinal data is the proportional odds model with fixed effects for the three religion groups (x_l) contrasting to Roman Catholic, and random effects (u_j) for respondents. The possible association among residuals of the outcome over time is ignored in this model. The model can be written as

$$y_{ij}^{(s)} = \log\left(\frac{\gamma_{ij}^{(s)}}{1-\gamma_{ij}^{(s)}}\right) = \alpha^{(s)} - \left(\sum_{l=1}^{3} \beta_l x_{lij} + u_j\right) \tag{5}$$

$$u_j \sim N(0, \sigma_u^2), \qquad \text{cov}\left(\gamma_{ij}^{(s)}, \gamma_{ij}^{(r)}\right) = \gamma_{ij}^{(s)}(1-\gamma_{ij}^{(r)})/n_{ij} \qquad s \leq r$$

The six cutoff points $\alpha^{(s)}$ will have values between $-\infty$ and $+\infty$ on the logit scale, and $\gamma_{ij}^{(s)}$ are the tail probabilities corresponding to the cutoff points, conditional on the covariates and random effects.

Using following syntax, we firstly bring the ASCII data into SAS, then name variables and generate dummy variables for the three religion groups. Roman Catholic is treated as the reference group. We have specified arrays c(i), y(i) and gamma(i) (i=1,2,…6) for the cut points, the logit dependent variables and cumulative probabilities respectively, corresponding to the intervals between cut points.

```
data socatt;
  infile 'c:\socatt.txt';
  input year yvar religion person;
  if (religion=2) then rel2=1;
    else rel2=0;
  if (religion=3) then rel3=1;
    else rel3=0;
  if (religion=4) then rel4=1;
    else rel4=0;
  run;

proc nlmixed data=socatt;
  parms ss=1 c1=-3 c2=-2 c3=-1 c4=0 c5=0.1 c6=0.5 b1-b3=-0.5;
  array c[6] c1-c6;
  array y[6] y1-y6;
  array gamma[6] gamma1-gamma6;

  do i=1 to 6;
  y[i] = c[i]+b1*rel2+b2*rel3+b3*rel4 + u;
  gamma[i]=exp(-y[i])/(1+exp(-y[i]));
  end;

  If (yvar=1) then z=1-gamma1;
  else if (yvar=2) then z=gamma1 - gamma2;
  else if (yvar=3) then z=gamma2 - gamma3;
  else if (yvar=4) then z=gamma3 - gamma4;
  else if (yvar=5) then z=gamma4 - gamma5;
  else if (yvar=6) then z=gamma5 - gamma6;
  else z=gamma6;
  if (z>1e-8) then ll=log(z);
  else ll=-1e100;

  model yvar ~ general(ll);
  random u ~ normal(0,ss) subject=person;
  run;
```

It can be seen the proportional odds model may be specified in PROC NLMIXED using the general distribution function with the log-likelihood function specified by Z. The RANDOM statement allows more than one random parameter.

Suppose that the response is an unordered categorical variable ($s = 1,…,7$), and we fit a multinomial model with a series of odds ratios comparing each of $s-1$ categories with a base category, $t = 7$ in our case. The simplest model with common random effects for each odds ratio would be

$$y_{ij}^{(s)} = \log\left(\frac{p_{ij}^{(s)}}{p_{ij}^{(t)}}\right) = \beta_0^{(s)} + u_j + \sum_{l=1}^{3}\beta_l^{(s)} x_{lij}^{(s)}, \qquad u_j \sim N(0, \sigma_u^2) \tag{6}$$

$$p_{ij}^{(s)} = \exp\left(\beta_0^{(s)} + u_j + \sum_{l=1}^{3}\beta_l^{(s)} x_{lij}^{(s)}\right) \times \left(1 + \sum_{h=1}^{6}\exp\left(\beta_0^{(h)} + u_j + \sum_{l=1}^{3}\beta_l^{(h)} x_{lij}^{(h)}\right)\right)^{-1}$$

The syntax for this model is as follows. For simplification, we have ignored the covariates x in the fixed part of the model.

```
proc nlmixed data=socatt;
  parms b1-b2=-1 b3-b6=0.5 ss=5;
  array eta[6] y1-y6;
  array p[6] p1-p6;
  array b[6] b1-b6;

  do i=1 to 6;
  y[i] = b[i] + u;
  Psum=Psum + exp(y[i]);
  end;

  do i=1 to 6;
  p[i]=exp(y[i])/(1+Psum);
  if (yvar=i) then z=p[i];
  end;
  if (yvar=7) then z=1-p1-p2-p3-p4-p5-p6;

  if (z>1e-8) then ll=log(z);
  else ll=-1e100;

  model yvar ~ general(ll);
  random u ~ normal(0,ss) subject=person;
  run;
```

ML estimates for both models are given in Table 7. For aggregated proportional data, the REPLICATE statement can be used to specify the model.

In SAS poor starting values are often responsible for a failure to estimate the model, the error message ' no valid parameter points were found' is given. Currently, users have to find their own way of obtaining reasonable starting values, which can be difficult. Having an option in the procedure for generating good starting values using a non-iterative method that does not depend on users input, would be extremely useful.

4.2 Cross-classified random effects model

Fitting a two-way cross classification model using the MIXED procedure is no different to fitting a random effects model to data with a three-level hierarchy. The exam scores of 3435 16 years old students who attended 148 primary schools and 19 secondary schools in Fife Scotland were analyzed. The attainment score is the outcome variable, and separating out the variability in the attainment score due to primary and secondary schools is of particular interest. We consider the following simple model

$$y_{i(jk)} = \beta_0 + \beta_1 x_{1ij} + u_j + u_k + e_{i(jk)} \tag{7}$$

$$u_j \sim N(0, \sigma_{u_j}^2), \ u_k \sim N(0, \sigma_{u_k}^2), \ e_{i(jk)} \sim N(0, \sigma_e^2),$$

where i denotes students at level 1, j and k denote primary and secondary schools respectively at level 2, and the covariate is gender. Random effects for primary and secondary schools are

represented by u_j and u_k respectively. Typically, no correlation structure between the two sets of random effects is assumed. This model can be fitted using the following syntax

```
proc mixed data=xc;
  class primary secondary sex;
  model attain= sex/s;
  random int /sub=primary;
  random int /sub=secondary;
  run;
```

Alternatively, one can use the REPEATED statement to replace the second RANDOM statement to obtain the same results but at much slower speed (1'45" v.s. 2"). Another way is to set the classification with the most units as the subject for the REPEATED statement. In our example there are many more primary schools than secondary, so the model is specified using

```
random int /sub=secondary;
repeated /sub=primary;
```

This reduces the dimension of the design matrix associated with the higher level random effects from 148 to 19 between the two usages of the repeated statement. Therefore, the convergence time of 4" has been much improved.

The results for Model (7) in Table 8 are comparable to other packages.

To allow random effects of gender across primary schools, we only need to change the syntax line for the first RANDOM statement:

```
random sex /type=un sub=primary;
```

This model takes 41 seconds to converge. At the primary school level, we have the variance for gender=1 and the variance for gender=2 as well as a covariance between them. The log likelihood ratio statistic is 2.2 on 2 degree of freedom, suggesting no significant difference in the variability by gender.

4.3 Multivariate Normal response model

Modelling multiple responses, accounting for the dependence between them, is of interest to many researchers. By treating the responses as repeated measurements nested within subject, the multivariate model fits neatly into the multilevel framework. The example consists of scores on a science examination obtained on both a traditional written paper (Y_1) and coursework (Y_2) from 1905 16 years old students in 73 schools in England. The data have a three-level structure of exam marks nested within students within schools. Interest in these data centres on the relationship between the two components at both school and student levels, whether there are gender (X =1 for girls and 2 for boys) differences in this relationship and whether the variability differs for the two components. The bivariate Normal model was considered for this purpose. The responses are both *N(0,1)*.

Let $i = 1, 2$ indicate the responses at level 1, j the students at level 2, and k the schools at level 3. An index variable z (1 for written paper, 2 for coursework) is required. The basic model fitted is as follows.

$$y_{ijk} = z y_{1jk} + (1-z) y_{2jk} \qquad (8)$$

$$y_{1jk} = \beta_0 + \beta_1 x_{jk} + v_{1k} + u_{1jk}$$

$$y_{2jk} = \alpha_0 + \alpha_1 x_{jk} + v_{2k} + u_{2jk}$$

$$\begin{pmatrix} v_{1k} \\ v_{2k} \end{pmatrix} \sim MVN(0, \Omega_v): \Omega_v = \begin{pmatrix} \sigma_{v_1}^2 & \\ \sigma_{v_{12}} & \sigma_{v_2}^2 \end{pmatrix}$$

$$\begin{pmatrix} u_{1jk} \\ u_{2jk} \end{pmatrix} \sim MVN(0, \Omega_u): \Omega_u = \begin{pmatrix} \sigma_{u_1}^2 & \\ \sigma_{u_{12}} & \sigma_{u_2}^2 \end{pmatrix}$$

In this model, β_1 and α_1 are the gender effects on written paper and coursework respectively. At the school level, the unstructured covariance matrix Ω_v estimates the variability of mean marks of the two components among schools, with the correlation coefficient calculated as $\rho_v = \sigma_{v_{12}} / \sigma_{v_1} \sigma_{v_2}$. Similarly, the unstructured covariance matrix Ω_u is obtained at the student level.

The SAS syntax for fitting this model is straightforward. However, a few tricks in preparing the data are required in order to obtain correct estimates. Failure to do so, results in non-convergence or incorrect results.

(1) Data should be structured as a three-level hierarchy with the responses stacked in one column.

(2) The level 2 identification should be unique for each subject (each student in this case), no matter which level 3 unit the student comes from. Otherwise, the procedure would incorrectly classify the blocks of students based on the subject id on which some students in different schools might be coded the same.

(3) Students with missing values on one or more of the outcomes should be kept in the data set. Their missing values should be coded as full stops (periods) and placed at the end of the level 1 units within each level 2 unit. For example, student number 10 from school 2 did not take the written paper exam, so the data segment for the person should look like

school	student	index	Y	gender
2	10	2	y2	1
2	10	1	•	1

Provided the data are structured correctly, the procedure will fit the multivariate model in the same way as other conventional three-level models. It can be very slow for large data sets. Fitting Model (8) in our example takes 14'27" to converge to the following REML estimates, with standard errors in brackets:

```
Proc mixed data=bivariate;
    Class school student index gender;
    Model y=index index*gender/s noint;
    Random index /type=un sub=school;
    Repeated /type=un sub=student;
    Run;
```

$\beta_0 = 49.41$ (0.939), $\beta_1 = -2.500$ (0.561)

$\alpha_0 = 69.62$ (1.179), $\alpha_1 = 6.752$ (0.671)

$\sigma_{v_1}^2 = 47.61$ (9.597), $\sigma_{v_2}^2 = 76.41$ (14.98)

$\sigma_{v_{12}} = 25.29$ (9.197)

$\sigma_{u_1}^2 = 124.69$ (4.346), $\sigma_{u_2}^2 = 180.19$ (6.256)

$\sigma_{u_{12}} = 73.02$ (4.160)

Time to converge is 14'27" on a computer of Pentium 433 with 398 RAM under Window NT system.

To obtain correlation coefficients instead of covariances between the two outcomes at both school and student levels, one can use the option TYPE=UNR at both levels.

Effects for other covariates can be fitted easily by adding terms "index*var" in the Model statement.

4.4 A Model for Meta-analysis

In a meta analysis means and standard deviations are collected for a set of studies together with some variables such as the control/treatment group at the study level. The means are regressed on some covariates X given the standard deviations or the sizes of groups of the studies. The between-study variation is of interest.

Let j indicate studies at level 2 and $\bullet$ denote the means or some other aggregated measure, the model may be written as

$$y_{\bullet j} = (X\beta)_{\bullet j} + u_j + e_{\bullet j} \qquad (9)$$

$$u_j \sim N(0, \sigma_u^2), \qquad e_{\bullet j} \sim N\left(0, \frac{\sigma_{e_j}^2}{n_j}\right),$$

where σ_u^2 is the variance parameter to be estimated for the between-study variability, $\sigma_{e_j}^2$ is the standard deviation known for the jth study, and X is any explanatory variable at the study level. In SAS terminology, we say that the R matrix for level-1 residuals is known. There is no need to estimate the residual variance as long as we supply this information in the model and constrain R to be fixed at this value while estimating other parameters in the model. This requirement can be implemented through the WEIGHT and PARMS statements in the MIXED procedure.

To illustrate the procedure, we use the example from the meta-analysis of teacher expectancy effects. The effect-size and standard deviation, together with an explanatory variable for 19 clusters, are available. The syntax for fitting this model is given below.

```
Data meta;
    Infile 'c:\meta.txt';
    Input effect sd week studyID;
    Invsd=1/(sd);
    Run;

Proc mixed covtest data=meta;
    Class studyID;
    Weight invsd;
    Model effect = week/s;
    Random int /type=un sub=studyID;
    Parms (0.1) (1)/hold=(2);
    Run;
```

Table 9 Model estimates (REML) based on 19 effect-sizes data in meta analysis

Parameters	SAS
β_0	0.409 (0.087)
β_1	-0.158 (0.036)
σ_u^2	0.000 (0.000)
-2LogLikelihood	-2.700

If the sample size of the control/treatment groups, n_{gj} (g=1, 2 for control and treatment), is provided, instead of the standard deviations of the studies, the weight variable in the SAS syntax should be $\sqrt{n_{gj}}$.

If the study output is a proportion such as an odds-ratio from case-control studies, we can fit the same model using the log odds-ratio. In some cases, we may have individual level data of

binary/binomial outcomes from all studies, in which case an ordinary two-level logistic model with study level random effects can be fitted as described in Section 3.3 of the review.

4.5 Spatial models

Models with spatial structures can be specified using the option TYPE. There are eight choices available in the current version. Factor analysis of the covariance structures can be fitted in the same way with three choices. We have not explored these models in this review.

5 Documentation and user support

SAS on-line documents are provided on a CD ROM with the package. One can either install the document on the computer as an on-line help system or read it directly from the CD drive when needed. Detailed information on SAS Procedures is included in the Statistics part of the document.

For the majority of multilevel models reviewed in this article, three procedures are involved: MIXED, NLMIXED and NESTED. They are described in Chapters 41, 46 and 44 respectively. Each chapter starts with an overview or introduction of the procedure, followed by a brief comparison between that procedure and other related procedures, and an informative bibliography. This background information is particularly useful for users to gain familiarity with the general features of the procedure before analysing data and to make a choice among many procedures that may share similar functions. The documentation also includes general notation of the basic models which can be fitted, syntax format, discussion of computational issues and some examples.

SAS users receive a free monthly e-newsletter 'Your SAS Technology Report'. It contains news on SAS conferences, user group updates, latest development in the documentation and new software linked to the main package, new statements in procedures, and new frequently asked questions, etc. Downloading new documents is also available through links provided. The newsletter reflects the international network of SAS users.

For a major package like SAS with many users worldwide, user support implies a large input of human resources. On a couple of occasions, we made enquires on general issues, and these were effectively dealt with by the provider of the review version. We found the information provided on the SAS web site (www.sas.com/service/techsup/) regarding technical support policies and regional contacts to be very convenient and helpful. One can also find information on training program was all over the world from the web site.

6 Conclusion

Generally speaking, the MIXED and NLMIXED procedures in SAS Release 8.2 are powerful tools that handle many types of multilevel or random effect models, as shown in the review. For small samples, the MIXED procedure can be used to fit models with more than three levels. Statistical inferences based on REML and ML are available. For large data sets, however, with more than two levels, the procedure can be rather inefficient.

The NLMIXED procedure is intuitive for nonlinear models and user friendly, although it only handles data with a two-level hierarchy. However, obtaining suitable starting values is an issue in fitting complicated nonlinear models. The choice on functions for the covariance structure is an excellent feature compared to other packages that have much less choice.

For repeated measures data, SAS handles data with autocorrelation over time, dependent response structures and spatial covariance structures, although only for discrete time points. A few examples or more explanation on how to use these features would be helpful in promoting their application.

The documents for general computational issues, statistical inference and syntax details are comprehensive, informative and helpful for statisticians. But for sociologists or applied researchers, this information could be difficult to digest. As the procedures are still command driven, the syntax can be complicated according to the type of model. Also a statement in one procedure may operate differently in another procedure. For these reasons, new users accustomed to a Windows environment would have a steep learning curve. The SAS Institute provides training at different levels in a number of regions to help such users.

For more details on the current release, contacts to distributors in your own region, and on-line order for products, The SAS web site is a good resource:
http://www.sas.com/products/sassystem/release82/.

Acknowledgement

Reference

Bryk, A., & Raudenbush, S.W. 1992: *Hierarchical Linear Models for Social and Behavioral Research: Applications and Data Analysis Methods*. Newbury Park, CA: Sage.

Goldstein H., Healy, M.J.R. and Rasbash, J. 1994: Multilevel time series models with applications to repeated measures data. *Statistics in Medicine* **13**, 1643-55.

Goldstein, H. 1995: *Multilevel Statistical Models*, Second Edition, Edward Arnold, London.

Guo, G & Zhao HX. 2000: Multilevel modeling for binary data, *Annu. Rev. Sociol.* 26: 441-62.

Huq, N.M. and Cleland, J., 1990: Bangladesh Fertility Survey 1989, Main Report. Dhaka: National Institute of Population Research and Training (NIPORT).

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

McGrath, K. and Waterton, J, 1986: British Social Attitudes, 1983-1986 Panel Survey, Technical Report. London, Social and Community Planning research.

McLeod, A. 2001: Multivariate Multilevel Models, in Leyland, A. H. and Goldstein, H. Ed. *Multilevel Modelling of Health Statistics*, Wiley & Sons, Chichester.

Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I. etc, 2000: *A user's guide to MLwiN*, Multilevel Models Project, Institute of Education, University of London.

Rasbash, J. and Browne, W. 2001: Modelling Non-Hierarchical Structure, in Leyland, A. H. and Goldstein, H. Ed. *Multilevel Modelling of Health Statistics*, Wiley & Sons, Chichester.

Rijmen, F., Tuerlinckx, F. and Boeck, P.D., 2001*: A nonlinear mixed model framework for IRT models*, presented to Belgium Statistics Society Conference, October, 2001, Oostende.

The MIXED/NLMIXED Procedure 1999: SAS/STAT User's Guide, SAS On-Line Document. SAS Institute Inc.

Yang, M., Rasbash, J., Goldstein, H. and Barbosa, M., 1999*: MLwiN macros for advanced multilevel modelling*, Multilevel Models Project, Institute of Education, University of London.

Yang, M., Fielding, A. and Goldstein, H. 2002: Multilevel ordinal models for examination grades, *submitted for publication.*

Table 1 Multilevel models that can be fitted by SAS procedures

Model type	Algorithms	Estimating methods	Covariates	Random slopes	Weighting	Fitting variance function (any level)	Procedures
Normal response	Newton-Raphson	REML ML MIVQUE0*	Yes	Yes	level 1	Yes	MIXED fits most two-level models efficiently; NESTED fits data with more than 2 levels for simple variance component model very efficiently.
Binary/Binomial	Methods A and B	ML	Yes	Yes	none	Yes (level 2)	NLMIXED. two level hierarchy only
Poisson	Methods A and B	ML	Yes	Yes	none	Yes (level 2)	NLMIXED. two level hierarchy only
Nominal multinomial	Methods A and B	ML	Yes	Yes	none		NLMIXED, two-level hierarchy only
Ordered multinomial	Methods A and B	ML	Yes	Yes	none		NLMIXED, two-level hierarchy only
Cross-classified	Newton-Raphson	REML/ML MIVQUE0*	Yes	Yes	Yes	Yes	MIXED
Repeated measures with time series	Newton-Raphson	REML/ML MIVQUE0*	Yes	Yes	Yes	Yes	MIXED
Multivariate Normal	Newton-Raphson	REML/ML MIVQUE0*	Yes	Yes	Yes	Yes	MIXED
Meta-analysis for normal response	Newton-Raphson	REML/ML MIVQUE0*	Yes	Yes	level 1	Yes (level 2)	MIXED
Model with general nonlinear functions	Methods A and B	ML	Yes	Yes	none		NLMIXED, two level hierarchy only
Spatial model	Newton-Raphson	REML/ML MIVQUE0*	Yes	Yes	Yes		Mixed
Multiple membership							N/A
Survival data							N/A
Mixed responses model							N/A
	• MIVQUE0 stands for Minimum Variance Quadratic Unbiased Estimation of the random parameters • Method A: Adaptive Gaussian quadrature and dual quasi-Newton (1st order optimisation) • Method B: Adaptive Gaussian quadrature and Newton-Raphson (2nd order optimisation)						

Model building	Fixed and random parameters.	Syntax specification to fit the model	REML Estimates (SE)	Seconds to convergence
A: Variance component with covariates 'standlrt' (x1), 'girl' (x2) and 'schgend' (x3, x4) *Note: By default, SAS treats the last category code in the class variable as the reference code.*	*Fixed* $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$ *Random* Level 2: $(\sigma_{u_0}^2)$ Level 1: $(\sigma_{e_0}^2)$	`Proc mixed noclprint covtest; Class school girl schgend; Model exam=standlrt girl schgend /s ddfm=bw; Random int /type=un sub=school; Run;`	$\hat{\beta}_0$ =-0.0094 (0.0779) $\hat{\beta}_1$ =0.5598 (0.01245) $\hat{\beta}_2$ =0.1674 (0.0341) $\hat{\beta}_3$ =-0.1590 (0.0894) $\hat{\beta}_4$ =0.0187 (0.1260) $\hat{\sigma}_{u_0}^2$ =0.0858 (0.0178) $\hat{\sigma}_{e_0}^2$ =0.5625 (0.0126) -2LogL=9347.67	0.38
B: Variance component with interaction, 'girl'*'standlrt' (x5)	*Fixed* $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$ *Random* Level 2: $(\sigma_{u_0}^2)$ Level 1: $(\sigma_{e_0}^2)$	`Proc mixed noclprint covtest; Class school girl schgend; Model exam=standlrt girl schgend gender*standlrt /s ddfm=bw; Random int /type=un sub=school; Run;`	$\hat{\beta}_0$ =-0.0094 (0.0779) $\hat{\beta}_1$ =0.5626 (0.01837) $\hat{\beta}_2$ =0.1673 (0.0341) $\hat{\beta}_3$ =-0.1588 (0.0894) $\hat{\beta}_4$ =0.0189 (0.1261) $\hat{\beta}_5$ =0.0051 (0.0246) $\hat{\sigma}_{u_0}^2$ =0.0858 (0.0178) $\hat{\sigma}_{e_0}^2$ =0.5627 (0.0126) -2LogL=9353.20	0.38
C: Random slopes of 'standlrt' effect'	*Fixed* $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$ *Random* Level 2: $(\sigma_{u_0}^2, \sigma_{u_{01}}, \sigma_{u_1}^2)$ Level 1: $(\sigma_{e_0}^2)$	`Proc mixed noclprint covtest; Class school girl schgend; Model exam=standlrt girl schgend girl*standlrt /s ddfm=bw; Random int standlrt /type=un sub=school; Run;`	$\hat{\beta}_0$ =-0.01197 (0.0742) $\hat{\beta}_1$ =0.5503 (0.0257) $\hat{\beta}_2$ =0.1686 (0.0338) $\hat{\beta}_3$ =-0.1779 (0.0821) $\hat{\beta}_4$ =-0.0004 (0.1162) $\hat{\beta}_5$ =0.0069 (0.0295) $\hat{\sigma}_{u_0}^2$ =0.0837 (0.0174) $\sigma_{u_{01}}$ =0.0205 (0.0070) $\sigma_{u_1}^2$ =0.0153 (0.0047) $\hat{\sigma}_{e_0}^2$ =0.5504 (0.0124) -2LogL=9308.24	0.55
D: Level 1 variance heterogeneity by 'girl' (1=boy, 0=girl)	*Fixed* $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$ *Random* Level 2: $(\sigma_{u_0}^2, \sigma_{u_{01}}, \sigma_{u_1}^2)$ Level 1: $(\sigma_{e_1}^2, \sigma_{e_2}^2)$	`Proc mixed noclprint covtest; Class school girl schgend; Model exam=standlrt girl schgend girl*standlrt /s ddfm=bw; Random int standlrt /type=un sub=school; Repeated /type=un(1) group=girl; Run;`	$\hat{\beta}_0$ =-0.0119 (0.0741) $\hat{\beta}_1$ =0.5502 (0.0262) $\hat{\beta}_2$ =0.1687 (0.0340) $\hat{\beta}_3$ =-0.1782 (0.0816) $\hat{\beta}_4$ =-0.0005 (0.1166) $\hat{\beta}_5$ =0.0069 (0.0297) $\hat{\sigma}_{u_0}^2$ =0.0837 (0.0175) $\sigma_{u_{01}}$ =0.0206 (0.0070) $\sigma_{u_1}^2$ =0.0155 (0.0048) $\hat{\sigma}_{e_1}^2$ =0.5880 (0.0210) $\sigma_{e_2}^2$ =0.5253 (0.0153) -2LogL=9302.21	1.86

E: Level 1 variance as an exponential function of 'standlrt'	*Fixed*
	 $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$
	 Random
	 Level 2: $(\sigma_{u_0}^2, \sigma_{u_{01}}, \sigma_{u_1}^2)$
	 Level 1: (σ_e^2, δ)
		`Proc mixed noclprint`
		`covtest; Class school`
		`girl schgend; Model`
		`exam=standlrt girl`
		`schgend girl*standlrt`
		`/s ddfm=bw; Random`
		`int standlrt /type=un`
		`sub=school; Repeated`
		`/local=exp(standlrt);`
		`Run;`
			$\hat{\beta}_0$ =-0.0115 (0.0744)
			
			 $\hat{\beta}_1$ =0.5514 (0.0257)
			
			 $\hat{\beta}_2$ =0.1687 (0.0338)
			
			 $\hat{\beta}_3$ =-0.1782 (0.0822)
			
			 $\hat{\beta}_4$ =-0.0004 (0.1166)
			
			 $\hat{\beta}_5$ =0.0073 (0.0296)
			
			 $\hat{\sigma}_{u_0}^2$ =0.0839 (0.0175)
			
			 $\sigma_{u_{01}}$ =0.0211 (0.0070)
			
			 $\sigma_{u_1}^2$ =0.01492 (0.0047)
			
			 $\hat{\sigma}_e^2$ =0.5498 (0.0124)
			
			 δ =-0.0540 (0.0231)
			 -2LogL=9302.79
				2.16

Note: All models were run on a computer of Pentium 433 with 398 RAM under Window NT system.

Table 4 PROC MIXED specifications for 3-level Normal models
(31022 students nested within 2279 schools nested within 130 local educational authorities)

Model	Fixed and random parameters.	Syntax specification to fit models	REML Estimates (SE)	time to convergence

F: Variance component without any covariate	*Fixed*
	Intercept: (β_0)
	
	Random
	Level 3: ($\sigma^2_{v_0}$)
	
	Level 2: ($\sigma^2_{u_0}$)
	
	Level 1: ($\sigma^2_{e_0}$)
		`Proc nested data=chem;`
		`Class lea school;`
		`Var Ascore;`
		`Run;`
			β_0 = 5.81 (0.051)
			
			$\sigma^2_{v_0}$ =0.0075
			
			$\sigma^2_{u_0}$ =2.521
			
			$\sigma^2_{e_0}$ =8.491
				0.13"

G: Variance component with covariates 'GCSE' (x1)	*Fixed*
	(β_0, β_1)
	Random
	Level 3: ($\sigma^2_{v_0}$)
	
	Level 2: ($\sigma^2_{u_0}$)
	
	Level 1: ($\sigma^2_{e_0}$)
		`Proc mixed noclprint covtest`
		`data=chem;`
		`Class school lea;`
		`Model Ascore=gcse /s;`
		`Random int /type=un sub=lea;`
		`Repeated /type=cs sub=school;`
		`Run;`
			β_0 = 5.621 (0.031)
			β_1 = 2.473 (0.017)
			$\sigma^2_{v_0}$ = 0.015 (0.014)
			$\sigma^2_{u_0}$ = 1.166 (0.056)
			$\sigma^2_{e_0}$ = 5.154 (0.043)
			-2LogL=141,696.89
				3'18"

		`Proc mixed noclprint covtest`
		`data=chem;`
		`Class school lea;`
		`Model y=gcse /s;`
		`Random int /type=un sub=lea;`
		`Random int /type=un`
		`sub=school;`
		`Run;`
			β_0 = 5.621 (0.031)
			β_1 = 2.473 (0.017)
			$\sigma^2_{v_0}$ = 0.015 (0.014)
			$\sigma^2_{u_0}$ = 1.166 (0.056)
			$\sigma^2_{e_0}$ = 5.154 (0.043)
			-2LogL=141,696.89
				2:10'44"

Note: All models were run on a computer of Pentium 433 with 398 RAM under Windows NT4.

(Note: The results of Model G in the table are based on the raw GCSE average uncentred. They will be rerun)

Table 5 PROC NLMIXED specifications for 2-level binary response models

Parameter	Logit link	Probit link
	Logit link	Probit link
β_0	-1.692(0.148)	-1.029(0.087)
β_1	0.732(0.119)	0.449(0.072)
β_2	-0.027(0.008)	-0.016(0.005)
β_3	1.109(0.158)	0.670(0.095)
β_4	1.377(0.175)	0.835(0.105)
β_5	1.346(0.180)	0.815(0.107)
$\sigma^2_{u_0}$	0.216(0.073)	0.080(0.027)
-2LL	2413.4	2412.8
Time to run	47"	35"
β_0	-1.712(0.160)	-1.042(0.095)
β_1	0.816(0.171)	0.500(0.104)
β_2	-0.027(0.008)	-0.016(0.005)
β_3	1.126(0.160)	0.682(0.096)
β_4	1.368(0.177)	0.831(0.106)
β_5	1.355(0.183)	0.825(0.109)
$\sigma^2_{u_0}$	0.388(0.129)	0.143(0.047)
$\sigma_{u_{01}}$	-0.403(0.175)	-0.150(0.064)
$\sigma^2_{u_1}$	0.663(0.321)	0.247(0.119)
-2LL	2398.4	2397.6
Time to run	3'31"	3'25"

SAS Syntax
 	
Random intercept model
	`Proc nlmixed data=use;`
	`Parms b0=-1 b1=0.4 b2=-0.02 b3=0.7 b4=1 b5=1 sigma=0.4;`
	`y=b0+u+b1*urban+b2*age+b3*lc2+b4*lc3+b5*lc4;`
	`p=1/(1+exp(-y));`
	`Model use ~ binary(p);`
	`Random u ~ normal(0,sigma)`
		`Proc nlmixed data=use;`
		`Parms b0=-1 b1=0.4 b2=-0.02 b3=0.7 b4=1 b5=1 sigma=0.4;`
		`y=b0+u+b1*urban+b2*age+b3*lc2+b4*lc3+b5*lc4;`
		`p=probnorm(y);`
		`Model use ~ binary(p);`
		`Random u ~ normal(0,sigma) subject=district;`
		`Run;`

	subject=district; Run;	

SAS Syntax
 	
Random intercept and random slope model
	Proc nlmixed data=use;
	Parms b0=-1 b1=0.4 b2=-0.02 b3=0.7 b4=1 b5=1 ss1=0.4 s1s2=-0.01 ss2=0.2;
	y=b0+u0+(b1+u1)*urban+b2*age+b3*lc2+b4*lc3+b5*lc4;
	p=1/(1+exp(-y));
	Model use ~ binary(p);
	Random u0 u1 ~ normal([0,0],[ss1,s1s2,ss2]) subject=district;
	Run;
		Proc nlmixed data=use;
		Parms b0=-1 b1=0.4 b2=-0.02 b3=0.7 b4=1 b5=1 ss1=0.4 s1s2=-0.01 ss2=0.2;
		y=b0+u0+(b1+u1)*urban+b2*age+b3*lc2+b4*lc3+b5*lc4;
		p=probnorm(y);
		Model use ~ binary(p);
		Random u0 u1 ~ normal([0,0],[ss1,s1s2,ss2]) subject=district;
		Run;

Table 6 Model estimates for 2-level repeated measures data

Parameters	SAS (REML)
Fixed	
β_0	148.98 (1.570)
β_1	6.166 (0.357)
β_2	1.091 (0.353)
β_3	0.468 (0.164)
β_4	-0.340 (0.302)
β_5	
β_6	
Random at level 2	
$\sigma^2_{u_0}$	64.012 (18.121)
$\sigma^2_{u_1}$	2.875 (0.830)
$\sigma^2_{u_2}$	0.660 (0.238)
$\sigma_{u_{01}}$	8.312 (3.206)
$\sigma_{u_{02}}$	1.416 (1.491)
$\sigma_{u_{12}}$	0.910 (0.363)
Random at level 1	
σ^2_e	0.220 (0.025)
ρ	
-2log-likelihood	629.825
Time to converge	2.0"

Table 7 Estimates of two-level ordinal and nominal models for Social Attitudes data

Parameters	Ordinal category	Parameters	nominal category
	SAS (ML)		SAS (ML)
Fixed effects		Fixed effects	
$\alpha^{(1)}$	-3.858(0.498)	$\beta_0^{(1)}$	-1.165 (0.280)
$\alpha^{(2)}$	-2.104(0.459)	$\beta_0^{(2)}$	-0.377 (0.235)
$\alpha^{(3)}$	0.692(0.451)	$\beta_0^{(3)}$	1.144 (0.201)
$\alpha^{(4)}$	1.628(0.454)	$\beta_0^{(4)}$	0.463 (0.211)
$\alpha^{(5)}$	2.521(0.460)	$\beta_0^{(5)}$	0.421 (0.212)
$\alpha^{(6)}$	3.571(0.469)	$\beta_0^{(6)}$	0.629 (0.208)
β_1	2.243(0.507)		
β_2	0.976(0.535)		
β_3	3.010(0.509)		
Random effect		Random effect	
σ_u^2	5.162(0.667)	σ_u^2	5.342 (0.540)
Time to converge	58.7"		1'47"

Table 8 Estimates of two-level cross-classification model

Parameters	SAS	
	REML	ML
β_0	5.754 (0.185)	5.756 (0.181)
β_1	0.499 (0.098)	0.499 (0.098)
σ_p^2	1.110 (0.204)	1.104 (0.202)
σ_s^2	0.370 (0.173)	0.346 (0.161)
σ_e^2	8.055 (0.199)	8.053 (0.199)
-2LL	17129.9	17123.5
Time to converge	3"	2"