Efficient Analysis of Mixed Hierarchical and Cross-Classified Random Structures Using a Multilevel Model

Jon Rasbash; Harvey Goldstein

Stable URL:

Journal of Educational and Behavioral Statistics is currently published by American Educational Research Association.

Efficient Analysis of Mixed Hierarchical and Cross-Classified Random Structures Using a Multilevel Model

Jon Rasbash
Harvey Goldstein
Institute of Education, University of London

An efficient and straightforward procedure is described for specifying and estimating parameters of general mixed models which contain both hierarchical and crossed random factors. This is done using a model formulated for purely hierarchically structured data and generalizes the results of Raudenbush (1993). The exposition is for the continuous response linear model with natural extensions to generalized linear, nonlinear, and multivariate models.

A considerable literature exists on random coefficient models where population units are hierarchically clustered. There have been applications in many areas such as data on students nested within schools (Aitkin & Longford, 1986) and repeated measures grouped within individual subjects (Laird & Ware, 1982). Bock (1989) contains a range of applications, and the books by Goldstein (1987b), Bryk and Raudenbush (1992), and Longford (1993) give detailed expositions of estimation theory and applications.

There are many cases, however, where units at the same level of a hierarchy are simultaneously classified by more than one factor. There are many instances of such structures, especially in the biological and social sciences. For example, school pupils may be classified by the school they attend as well as the neighborhood they live in, where both classifications are regarded as random. This is a two-level structure, with students at Level 1 nested within the cells of the Level 2 cross-classification. If the students were repeatedly measured then the measurement occasions would constitute the lowest Level 1, the students Level 2, and the crossing of schools and neighborhoods Level 3. In the present article we show how quite general models of this kind can be analyzed efficiently in a straightforward manner using

procedures designed for purely hierarchical or multilevel structures. A flexible algorithm for obtaining estimates is the iterative generalized least squares procedure (Goldstein, 1986), which allows complete generality in specifying random structures. It leads to maximum likelihood or restricted maximum likelihood estimates in the normal case.

Hartley and Rao (1967) and Patterson and Thompson (1971) discuss maximum likelihood estimation for cross-classified variance component structures, and Goldstein (1987a) describes a general procedure for modeling data containing random classifications for general random coefficient models which can be hierarchical or cross-classified, but does not discuss estimation problems. Raudenbush (1993) considers the two-level, two-way additive cross-classification and describes an EM algorithm to carry out the estimation together with worked examples. This is extended by Kang (1993), who introduces interaction terms into the model. The present article generalizes these results to p-way cross-classifications at any number of levels and provides an efficient computational procedure.

We now review briefly the basic two-level variance components model and then show how this can be generalized to more complex structures involving further hierarchical levels and random cross-classifications.

The Basic Hierarchical Model

A simple two-level hierarchical variance components model can be written as

$$y_{ij} = X_{ij}\beta + u_j + e_{ij} \tag{1}$$

$$\mathrm{var}(y_{ij}) = \sigma_u^2 + \sigma_e^2$$

$$\mathrm{cov}(y_{ij}y_{i'j}) = \sigma_u^2$$

$$\mathrm{var}(e_{ij}) = \sigma_e^2, \qquad \mathrm{var}(u_j) = \sigma_u^2$$

$$i = 1, \ldots, n_j; \qquad j = 1, \ldots, J$$

for the ith Level 1 unit within the jth Level 2 unit. The e_{ij} are mutually independent, all the other covariances are zero, X_{ij} is the ijth p element row vector of the total design matrix X, and β is a $(p \times 1)$ vector of coefficients for the fixed part of the model. Additionally, some of these coefficients may be modeled as random variables at any of the levels (Goldstein, 1986) giving a general random coefficients model.

Writing Y for the vector of responses and $\tilde{Y} = Y - X\beta$ for the vector containing the random variables, we have

$$E(\tilde{Y}\tilde{Y}^T) = V_2$$

which is an ($n \times n$) matrix where n is the sample size (that is, the number of Level 1 units), and V_2 has the following block-diagonal structure:

$$V_2 = \oplus_j (V_{1(2)j} + V_{2(2)j}) = V_{1(2)} + V_{2(2)} \tag{2}$$

$$V_{1(2)j} = \oplus_i (Z_{ij}^{(1)} \Omega_1 Z_{ij}^{(1)T})$$

$$\Omega_1 = \sigma_e^2$$

$$V_{2(2)j} = Z_j^{(2)} \Omega_2 Z_j^{(2)T} \tag{3}$$

$$\Omega_2 = \sigma_u^2$$

where $\oplus$ is the direct sum operator and $V_{m(t)}$ is the level m contribution to V for a t-level model. More generally, we will write $V_{m_1, m_2 \ldots m_r(t)}$ for the sum of the contributions to V from levels $m_1, m_2 \ldots m_r$ for a t-level model.

The design matrix for the Level 1 random coefficients is $Z_{ij}^{(1)}$, and for the Level 2 random coefficients $Z_j^{(2)}$. This formulation is general in allowing also for random coefficient models. For the simple variance components model without random coefficients, $Z_{ij}^{(1)}$, $Z_j^{(2)}$ are vectors with elements all equal to unity, so that

$$V_{1(2)} = \sigma_e^2 I_n, \qquad V_{2(2)} = \sigma_u^2 \oplus_j J_{(n_j)}$$

where $J_{(n_j)}$ is an ($n_j \times n_j$) matrix of ones. The matrices Ω_1 and Ω_2 are scalars in the variance components model but, in general, will be of order equal to the number of random coefficients at each level.

The V matrices for higher levels are constructed in a similar recursive fashion; for example, for a three-level model we have

$$V_3 = V_{12(3)} + \oplus_k Z_k^{(3)} \Omega_3 Z_k^{(3)T} \tag{4}$$

with inverse defined recursively as

$$V_3^{-1} = V_{12(3)}^{-1} - V_{12(3)}^{-1} Z^{(3)} \Omega_3 \{I + Z^{(3)T} V_{12(3)}^{-1} Z^{(3)} \Omega_3\}^{-1} Z^{(3)T} V_{12(3)}^{-1} \tag{5}$$

where $Z^{(3)}$ is the design matrix for the Level 3 random coefficients. Computational details are given by Goldstein and Rasbash (1992).

A Cross-Classified Model

Consider now the two-level additive variance components model with crossing at Level 2 and simple random variation at Level 1. We refer to the

two types of Level 2 unit as (a) Type 1 with subscript j_1, $j_1 = 1, \ldots, q_1$ and (b) Type 2 with subscript j_2, $j_2 = 1, \ldots, q_2$. Parentheses group classifications at the same level. The subscript i refers to the Level 1 unit, with each Level 1 unit having a unique identification. Thus, in terms of the above example, the data are structured as individual students (Level 1 units) nested within the cells of the School $\times$ Neighborhood classification (the Level 2 units). Write

$$y_{i(j_1 j_2)} = X_{i(j_1 j_2)}\beta + u_{1 j_1} + u_{2 j_2} + e_{i(j_1 j_2)}. \tag{6}$$

The Level 1 random structure is defined by the independently distributed variables $e_{i(j_1 j_2)}$ with variance σ_e^2, and we obtain the following Level 2 covariance structure:

$$\text{cov}(y_{i(j_1 j_2)} y_{i'(j_1 j'_2)}) = \sigma_1^2$$

$$\text{cov}(y_{i(j_1 j_2)} y_{i'(j'_1 j_2)}) = \sigma_2^2$$

$$\text{var}(y_{i(j_1 j_2)}) = \text{cov}(y_{i(j_1 j_2)} y_{i'(j_1 j_2)}) = \sigma_1^2 + \sigma_2^2$$

$$\text{cov}(y_{i(j_1 j_2)} y_{i'(j'_1 j'_2)}) = 0$$

This can be written as

$$E(\tilde{Y}\tilde{Y}^T) = V_{1(2)} + Z_1^{(2)}\Omega_{(1)2}Z_1^{(2)^T} + Z_2^{(2)}\Omega_{(2)2}Z_2^{(2)^T}$$

$$\Omega_{(1)2} = \sigma_1^2 I_{(q_1)}, \qquad \Omega_{(2)2} = \sigma_2^2 I_{(q_2)} \tag{7}$$

$Z_1^{(2)}$ and $Z_2^{(2)}$ are $(n \times q_1)$, $(n \times q_2)$ design matrices with component vectors

$z_{1mij_1} = 1$ for the j_1th Type 1 Level 2 unit

$\quad = 0$ otherwise

$\qquad\qquad m = 1, \ldots, q_1$

$z_{2mij_2} = 1$ for the j_2th Type 2 Level 2 unit

$\quad = 0$ otherwise

$\qquad\qquad m = 1, \ldots, q_2$

For example, Figure 1 shows a hypothetical structure for (6) with two schools and two neighborhoods, with one or two students within each cell.

Thus, (7) is a special case of (4) with a single Level 2 unit nested within a single Level 3 unit. The q_1 columns of $Z_1^{(2)}$ define q_1 variables (indexed by m) with coefficients random at Level 2, each with an associated variance parameter, zero covariances, and variances constrained to be equal to σ_1^2. The q_2 columns of $Z_2^{(2)}$ define q_2 coefficients random at Level 3 with associated variance parameters, zero covariances, and variances constrained to be equal to σ_2^2.

FIGURE 1. *A random design matrix for Model 6*

School	Neighborhood	Z_{11}	Z_{12}	Z_{21}	Z_{22}
1	1	1	0	1	0
1	2	1	0	0	1
1	2	1	0	0	1
2	1	0	1	1	0
2	1	0	1	1	0
2	2	0	1	0	1
2	2	0	1	0	1

We also note that (7) can be written alternatively as

$$E(\tilde{Y}\tilde{Y}^T) = V_{1(2)} + Z^{(2)}\Omega_2 Z^{(2)^T}$$

$$Z = (Z_1^{(2)}, Z_2^{(2)})$$

$$\Omega_2 = \begin{bmatrix} \Omega_{(1)2} & 0 \\ 0 & \Omega_{(2)2} \end{bmatrix}$$

which is a special case of (3), with a single Level 2 unit with corresponding $q_1 + q_2$ variables having coefficients random at Level 2 and with associated variance parameters, where the first set is constrained to be equal to σ_1^2 and the second set to σ_2^2. In this case, we will need to invert a matrix of order $(q_1 + q_2)$, whereas in the three-level formulation, the size of the largest matrix to be inverted is max (q_1, q_2) (Goldstein and Rasbash, 1992).

The generalization to further ways of classification is straightforward, and the general p-way cross-classification can be specified as a $(p + 1)$-level hierarchical model with a single unit at each of Levels 2, . . . , $p + 1$. From (5) we see that the order of the largest matrix to be inverted is max (q_i) where q_i is the number of categories of the ith classification.

Model 6 is the basic cross-classification model without interactions (see, for example, Searle, Casella, & McCulloch, 1992). In the present formulation, unbalanced data are handled automatically and, in addition, heterogeneous Level 1 structures, where the Level 1 variance depends on further explanatory variables, can be specified as in the standard hierarchical model (Goldstein, 1987b). Level 3 and higher level structures can also be added above the levels already specified with crossing within any of these further levels. As in the purely hierarchical case, we can introduce further explanatory variables with random coefficients so that $\Omega_{(1)2}$ and $\Omega_{(2)2}$ become general covariance matrices.

We now show how the hierarchical model formulation allows a more efficient estimation procedure requiring the inversion of matrices of order no larger than the number of categories in the second largest classification.

This corresponds to a generalization of Raudenbush's (1993) procedure for the two-way classification.

An Efficient Three-Level Specification for the Level 2 Crossed Design

Write the following three-level model, which is equivalent to Example 1 of Raudenbush (1993):

$$y_{i(j_1 j_2)} = X_{i(j_1 j_2)}\beta + u_{10j_1} + \sum_{m=1}^{q_2} u_{20mj_2} z_{2mij_2} + e_{i(j_1 j_2)} \tag{8}$$

where the third term defines the Type 2 coefficients random at Level 3 as before. The notation u_{10j_1} refers to the j_1th category for the intercept random variable for the Type 1 classification. The notation u_{20mj_2} refers to the j_2th category for the mth intercept random variable for the Type 2 classification.

The Type 1 units are now regarded as constituting a simple Level 2 classification, and if we order the data on these Type 1 units as in a two-level hierarchical model, we see that we may write

$$Z_1^{(2)}\Omega_{(1)2}Z_1^{(2)^T} = \bigoplus_j Z_{1j_1}^{(2)}\Omega_{(1)2}Z_{1j_1}^{(2)^T}$$

where $Z_{1j_1}^{(2)}$ is $(n \times r_1)$, and r_1 is the number of random coefficients associated with the Type 1 units. Thus, (8) leads to the form (4) and the model now becomes a three-level model with q_1 Level 2 units within a single Level 3 unit and q_2 variables defining the Type 2 classification with coefficients random at Level 3 as before.

For the p-way classification, therefore, an efficient procedure is to choose that classification with the greatest number of units and model this as random at Level 2 of the hierarchy, with the remaining ways of classification modeled at Levels 3, . . . , $p + 1$. This is also the approach for the two-level model adopted by Raudenbush (1993).

Random Coefficients, Interactions, and Classifications Across Higher Level Units

We can extend the basic variance components model by introducing further explanatory variables with coefficients random across units. Consider the Level 2 cross-classification with the intercept as before plus a further explanatory variable with a random coefficient, say x^*. The above procedure can be extended as follows. We choose the Type 1 units as random at Level 2 as before, with a random intercept and coefficient of x^*. We define sets of dummy variables, one set for each Type 2 unit defining both the intercept and the coefficient of x^*, and we make these intercepts and coefficients random at Level 3. Thus, the largest matrix to be inverted is now of order $r_h q_h$ where r represents the number of random coefficients associated with a

classification and where h refers to the classification where this product is the second largest. For each intercept-coefficient pair, we fit two variances and a covariance. For each type, the variances are constrained to equality, as are the covariances. Thus, we would fit the following extension of (8):

$$y_{i(j_1j_2)} = X_{i(j_1j_2)}\beta + u_{10j_1} + u_{11j_1}x^*_{1ij_1}$$
$$+ \sum_{m=1}^{n_{j_2}} (u_{20mj_2}z_{2mij_2} + u_{21mj_2}z_{2mij_2}x^*_{2ij_2}) + e_{i(j_1j_2)} \tag{9}$$

where the random coefficients of x^* have the second subscript taking the value 1.

Until now we have considered models which are additive in the random effects. Consider the following two-level model extension of (6):

$$y_{i(j_1j_2)} = X_{i(j_1j_2)}\beta + u_{j_1} + u_{j_2} + u_{(j_1j_2)} + e_{i(j_1j_2)} \tag{10}$$

The usual specification for the random interaction term $u_{(j_1j_2)}$ is that it has a simple variance $\sigma^2_{(12)}$ across the Level 2 cells (Searle et al., 1992). To fit such a model, we would define each cell of the cross-classification as a Level 2 unit with the simple variance term $\sigma^2_{(12)}$ and a single Level 3 unit with Level 3 random coefficients for dummy variables for the Type 1 classifications, and a single Level 4 unit for the Type 2 classification. We can test the adequacy of such a model against an additive model using a likelihood ratio test criterion.

Other specifications for the random interaction in (10) are possible. For example, it may be structured by further explanatory variables. Consider a two-level model with students at Level 1 and their primary school attended crossed with their secondary school attended at Level 2. In addition to modeling the simple additive variances across the two types of school, we may suppose that teacher experience (t), suitably measured, affects the between-school variation. We would add this to the model with an extra random coefficient at Level 2 as follows:

$$y_{i(j_1j_2)} = X_{i(j_1j_2)}\beta + u_{j_1} + u_{j_2} + u_{0(j_1j_2)} + u_{1(j_1j_2)}t_{i(j_1j_2)} + e_{i(j_1j_2)} \tag{11}$$

This extends (10) by adding a random coefficient as part of the interaction and, correspondingly, an extra variance and covariance parameter.

In general, we could define a full interaction model by fitting a separate variance term for each cell $u_{(j_1j_2)}$; that is, the model

$$y_{i(j_1j_2)} = X_{i(j_1j_2)}\beta + u_{(j_1j_2)} + e_{i(j_1j_2)}. \tag{12}$$

This would be fitted as a two-level model with a single Level 2 unit and a

dummy variable for each cell of the classification with coefficients random at Level 2 leading to $m_1 m_2$ variances. The adequacy of the additive or other reduced parameter model can again be tested using a suitable likelihood ratio test criterion.

A special case occurs with a Level 1 cross-classification. By definition, we have a design with only one unit per cell. We note also that this design is formally equivalent to a two-level classification with exactly one Level 1 unit per cell. If an additive model does not fully describe the between-unit variation, we have several possibilities for elaborating the model, analogous to the two-level case. We can seek explanatory variables with coefficients random at Level 1, or, for example, attempt to fit the full interaction model as above. Another approach is to regard the interaction as within-individual variation. The model then formally becomes the two-level model

$$y_{i(j_1 j_2)} = X_{i(j_1 j_2)} \beta + u_{j_1} + u_{j_2} + e_{i(j_1 j_2)} \tag{13}$$

with just one Level 1 unit per Level 2 cell. It would seem, in general, that (13) is a satisfactory formulation for such a Level 1 randomly cross-classified model.

It is possible to have cross-classifications which extend across more than one level. For example, in the Level 1 case we may have individual students in different schools who are assessed by the same random sample of raters. Consider the two-level model where within each school, which is a Level 2 unit, we have a Level 1 cross-classification of students by raters. Since it is the same set of raters, the contributions to V from the raters extends across all the Level 2 units. The model is therefore specified as having students random at Level 1, schools at Level 2, and raters with dummy explanatory variables and coefficients random at Level 3.

This model will also be appropriate for so-called *generalizability theory* models. In a simple case, instead of raters we have test items or groups of test items belonging to a test which is administered to samples of students in different schools. In this case, and with appropriate assumptions about response independence, we can use the above procedures. Where the response is discrete—for example, binary—we can fit a loglinear model. This is discussed further below. All generalizability models can be analyzed within the framework of this article.

Cross-Unit Membership Models

An interesting special case of a random cross-classification arises when units can be members of more than one higher level unit simultaneously. Thus, for example, students may attend more than one school, or children may belong to more than one extended family, where such a family is defined in terms of aunts and uncles. We suppose that, for each lower level unit, we have a weight associated with each higher level unit. In the case of students

and schools, this could be the proportion of time spent in each school. Another application might arise in a case where knowledge about unit membership is imprecise, so that for each lower level unit we have a known or estimated probability of belonging to each higher level unit, with the probabilities used as weights. The weights will sum to one, and generally there will be individuals, often the majority, with a single weight of unity and other weights zero.

To illustrate this, we consider again the simple two-level variance components model given by (1) with each Level 1 unit belonging to, at most, two Level 2 units. We write

$$y_{i(j_1 j_2)} = X_{i(j_1 j_2)}\beta + w_{ij_1} u_{j_1} + w_{ij_2} u_{j_2} + e_{i(j_1 j_2)}$$
$$w_{ij_1} + w_{ij_2} = 1 \tag{14}$$

so that the overall contribution at Level 2 is the weighted sum over the Level 2 units to which the Level 1 unit belongs. This leads to the following covariance structure:

$$\begin{aligned}
\text{var}(y_{i(j_1 j_2)}) &= (w_{ij_1}^2 + w_{ij_2}^2)\sigma_u^2 \\
\text{cov}(y_{ij_1} y_{i'j_1}) &= w_{ij_1} w_{i'j_1}\sigma_u^2 \\
\text{cov}(y_{ij_1} y_{i'j_2}) &= w_{ij_1} w_{i'j_2}\sigma_u^2
\end{aligned} \tag{15}$$

This has just the structure described by (7) with the additional constraint

$$\sigma_1^2 = \sigma_2^2 = \sigma_u^2$$

and with Z_{j_1}, Z_{j_2} replaced by the sets of weights W_{j_1}, W_{j_2}. The cross-classification at Level 2 is formally the set of Level 2 units crossed by themselves.

As before, this cross-unit model can be extended to incorporate random coefficients and general p-unit membership.

Relationship With Other Approaches

Engel (1990) discusses various approaches to the estimation of general variance component models, without random coefficients. He considers the method of *absorption* for avoiding the problem of handling matrices based upon a single factor with a very large number of categories. This is analogous to our procedure and that of Raudenbush (1993) for treating such a factor as the primary hierarchical one. He also deals with the situation where the data set can be partitioned into independent blocks according to a higher level unit (see below).

The software package SAS (1993) has a procedure "MIXED" which allows a two-level hierarchical model to be combined with further random factors or coefficients. This can deal, therefore, with a single factor having a large number of categories by treating it as the primary factor at either of the two

levels. Where there are three or more cross-classifications, those beyond the first have to be specified in full, using dummy variables equal in number to the total number of categories. This can result in having to handle very large matrices. Likewise, Engel (1990) deals only with the procedure for absorbing one factor.

Computing Considerations

All these computations can be carried out within ML3 (Prosser, Rasbash, & Goldstein, 1991), which is a general statistical package for fitting multilevel or hierarchical models. Its successor due in 1995, MLn, will allow an effectively unlimited number of levels and so be able to fit general p-way classifications. This software has flexible procedures for specifying random structures, including linear constraints, which allows the models of this article to be fitted. In some cases, the highest level units can be separated into disjoint sets with no cross-unit memebership. For example, we may have a cross-classification of schools by neighborhoods in a number of separate cities. We may then form a further level of the hierarchy in which each city is a Level 3 unit, and then fit a three-level model, possibly incorporating variation among cities. In such a case, the software is able to search for such sets and construct the necessary higher level units. The calculation of the log likelihood values in ML3 is slow and inefficient, but is rapid in MLn.

Care needs to be taken in choosing which classifications are assigned dummy explanatory variable vectors. In general, it is most efficient to choose the classification with the largest number of units as that which is treated as random at the level for which it is defined, taking into account the total number of random coefficients for each classification.

Example

This example uses examination scores for 16-year-old students. The secondary school a child attends is cross-classified by the primary school the child attended. The model is of the form given in Equation 6, where u_{1j_1} refers to the secondary school and u_{2j_2} refers to the primary school. The data are on 3,435 children who attended 148 primary schools and 19 secondary schools in Fife, Scotland. Table 1 shows the extent of the crossing by listing the distribution of primary school sizes for each secondary school.

A two-level variance components model with primary school as the Level 2 unit is specified, with a single Level 3 unit spanning the entire data set. Dummy variables, one for each secondary school, are created with coefficients random at Level 3 and variances constrained to be equal. We fit two models; the first (Model A) contains only the overall mean for the fixed part, and the second (Model B) adds a verbal reasoning (VR) test score, obtained by the students just prior to transfer to secondary school, to the fixed part.

The model takes four iterations to converge; on a 33-MHz 486 PC each iteration takes 81 seconds. Table 2 gives the results.

TABLE 1
Number of primary schools with different student numbers for each secondary school

Secondary school	Number of students in each primary school				
	1	2	3–5	6–10	11+
1	8	0	3	2	5
2	4	0	2	1	6
3	6	2	0	0	6
4	6	1	2	1	4
5	9	2	1	0	5
6	6	0	3	0	6
7	1	1	0	1	4
8	8	0	1	0	3
9	5	0	0	1	4
10	10	2	1	0	4
11	14	0	8	3	7
12	5	4	3	4	7
13	4	1	2	0	7
14	8	4	2	4	8
15	5	3	0	2	4
16	2	1	1	1	5
17	6	2	2	1	5
18	6	1	2	1	8
19	7	0	0	1	5

This analysis shows that the variation in achievement at age 16 attributable to primary school is 3 times greater than the variation attributable to secondary school. After fitting VR, the between-secondary-school variance becomes very small. We may interpret the analysis that fits VR as concerned with progress through secondary school, and the predominant influence of the primary school on attainment raises some important interpretational issues, which we shall not pursue here.

In this analysis, there are 120 cells of the cross-classification which contain

TABLE 2
Analysis of secondary by primary school examination attainment

Parameter	Estimates (*SE*)	
	Model A	Model B
σ_{u1}^2 (primary)	1.12 (0.20)	0.27 (0.06)
σ_{u2}^2 (secondary)	0.35 (0.16)	0.011 (0.021)
σ_e^2 (between students)	8.1 (0.2)	4.25 (0.10)
β_0 (intercept)	5.50 (0.18)	5.98 (0.07)
β_1 (VR)		0.16 (0.003)

only one student each. If we eliminate these from the analysis, we obtain two disjoint subsets containing 14 and 5 secondary schools. There are a further 23 cells containing two students each, and if these are removed, we obtain six disjoint subsets with the largest containing 8 secondary schools. Table 3 shows the results of fitting the models for these two reduced data sets. The first analysis took 32 seconds and the second took 13 seconds per iteration.

The only substantial difference is in the values of the between-secondary-school variance, but this is no greater than its standard error and imprecisely estimated. Thus, even where there are no proper disjoint subsets, it may be acceptable to reduce the computational time, and the storage overheads, by eliminating cells with small sample sizes from the analysis.

Discussion

We have shown how quite general designs involving crossing and nesting of random classifications can be handled efficiently within the general framework of a purely hierarchical model. In principle, it appears that designs of quite general complexity can be specified in a relatively straightforward manner. The estimation procedure, for Normally distributed data, can be either full or restricted maximum likelihood as shown by Goldstein (1989).

These procedures can be extended to all the other kinds of models which have been developed as extensions of the basic hierarchical or multilevel model. These include nonlinear models (Goldstein, 1991; Breslow & Clayton, 1993) with the special cases of binary, Poisson, and multinomial response models. Time series data (Goldstein, Healy, & Rasbash, 1994) where there is an autocorrelation structure at Level 1 can also be embedded within higher crossed and nested structures. An important extension is to multivariate data. Goldstein (1986, 1987b, 1991) showed how the general multivariate linear (and nonlinear) model could be formulated as a multilevel model with an extra lowest level which specified the multivariate structure. Such models can therefore be accommodated within the present framework so that multivariate general crossed and nested random designs are readily specified.

TABLE 3
Secondary by primary school examination attainment omitting small cells

Parameter	Estimates (SE)	
	Omitting cells with <2 students	Omitting cells with <3 students
σ_{u1}^2 (primary)	0.27 (0.06)	0.25 (0.06)
σ_{u2}^2 (secondary)	0.004 (0.021)	0.028 (0.030)
σ_e^2 (between students)	4.28 (0.11)	4.29 (0.11)
β_0 (intercept)	6.00 (0.06)	6.00 (0.07)
β_1 (VR)	0.16 (0.003)	0.16 (0.003)

References

Aitkin, M., & Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society, 149,* 1–43.

Bock, R. D. (Ed.). (1989). *Multilevel analysis of educational data.* New York: Academic Press.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalised linear models. *Journal of the American Statistical Association, 88,* 9–25.

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

Engel, B. (1990). The analysis of unbalanced linear models with variance components. *Statistica Nederlandica, 44,* 195–219.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalised least squares. *Biometrika, 73,* 43–56.

Goldstein, H. (1987a). Multilevel covariance component models. *Biometrika, 74,* 430–431.

Goldstein, H. (1987b). *Multilevel models in educational and social research.* New York: Oxford University Press.

Goldstein, H. (1989). Restricted unbiassed iterative generalised least squares estimation. *Biometrika, 76,* 622–623.

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

Goldstein, H., Healy, M. J. R., & Rasbash J. (1994). Multilevel time series models with applications to repeated measures data. *Statistics in Medicine, 13,* 1643–1655.

Goldstein, H., & Rasbash, J. (1992). Efficient computational procedures for the estimation of parameters in multilevel models based on iterative generalised least squares. *Computational Statistics and Data Analysis, 13,* 63–71.

Hartley, H. O., & Rao, J. N. K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika, 54,* 93–108.

Kang, S. J. (1993). *A covariance components model with two-way crossed random effects and ML estimation via the EM algorithm.* Unpublished doctoral dissertation, University of Michigan.

Laird, N. M., & Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics, 38,* 963–974.

Longford, N. T. (1993). *Random coefficient models.* Oxford: Clarendon Press.

Patterson, H. D., & Thompson, R. (1971). Maximum likelihood estimation of variances. *Biometrika, 58,* 545–554.

Prosser, R., Rasbash, J., & Goldstein, H. (1991). *ML3 software for three-level analysis: Users' guide.* London: Institute of Education.

Raudenbush, S. W. (1993). A crossed random effects model for unbalanced data with applications in cross sectional and longitudinal research. *Journal of Educational Statistics, 18,* 421–49.

SAS Institute, Inc. (1993). *SAS (Version 7): Users' guide.* Cary, NC: Author.

Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance components.* New York: Wiley.

Authors

JON RASBASH is Research Officer, Institute of Education, 20 Bedford Way, London, WC1HOAL, England. He specializes in statistical computing.

HARVEY GOLDSTEIN is Professor, Institute of Education, 20 Bedford Way, London, WC1HOAL, England. He specializes in the modeling of hierarchical data structures.

LINKED CITATIONS

- Page 1 of 2 -

You have printed the following article:

Efficient Analysis of Mixed Hierarchical and Cross-Classified Random Structures Using a Multilevel Model

Jon Rasbash; Harvey Goldstein

Journal of Educational and Behavioral Statistics, Vol. 19, No. 4. (Winter, 1994), pp. 337-350.

Stable URL:

http://links.jstor.org/sici?sici=1076-9986%28199424%2919%3A4%3C337%3AEAOMHA%3E2.0.CO%3B2-S

This article references the following linked citations. If you are trying to access articles from an off-campus location, you may be required to first logon via your library web site to access JSTOR. Please visit your library's website or contact a librarian to learn about options for remote access to JSTOR.

References

Statistical Modelling Issues in School Effectiveness Studies

M. Aitkin; N. Longford

Journal of the Royal Statistical Society. Series A (General), Vol. 149, No. 1. (1986), pp. 1-43.

Stable URL:

http://links.jstor.org/sici?sici=0035-9238%281986%29149%3A1%3C1%3ASMIISE%3E2.0.CO%3B2-6

Approximate Inference in Generalized Linear Mixed Models

N. E. Breslow; D. G. Clayton

Journal of the American Statistical Association, Vol. 88, No. 421. (Mar., 1993), pp. 9-25.

Stable URL:

http://links.jstor.org/sici?sici=0162-1459%28199303%2988%3A421%3C9%3AAIIGLM%3E2.0.CO%3B2-T

Multilevel Mixed Linear Model Analysis Using Iterative Generalized Least Squares

H. Goldstein

Biometrika, Vol. 73, No. 1. (Apr., 1986), pp. 43-56.

Stable URL:

http://links.jstor.org/sici?sici=0006-3444%28198604%2973%3A1%3C43%3AMMLMAU%3E2.0.CO%3B2-5

Multilevel Covariance Component Models

H. Goldstein

Biometrika, Vol. 74, No. 2. (Jun., 1987), pp. 430-431.

Stable URL:

http://links.jstor.org/sici?sici=0006-3444%28198706%2974%3A2%3C430%3AMCCM%3E2.0.CO%3B2-R

Restricted Unbiased Iterative Generalized Least-Squares Estimation

Harvey Goldstein

Biometrika, Vol. 76, No. 3. (Sep., 1989), pp. 622-623.

Stable URL:

http://links.jstor.org/sici?sici=0006-3444%28198909%2976%3A3%3C622%3ARUIGLE%3E2.0.CO%3B2-9

Nonlinear Multilevel Models, with an Application to Discrete Response Data

Harvey Goldstein

Biometrika, Vol. 78, No. 1. (Mar., 1991), pp. 45-51.

Stable URL:

http://links.jstor.org/sici?sici=0006-3444%28199103%2978%3A1%3C45%3ANMMWAA%3E2.0.CO%3B2-J

Maximum-Likelihood Estimation for the Mixed Analysis of Variance Model

H. O. Hartley; J. N. K. Rao

Biometrika, Vol. 54, No. 1/2. (Jun., 1967), pp. 93-108.

Stable URL:

http://links.jstor.org/sici?sici=0006-3444%28196706%2954%3A1%2F2%3C93%3AMEFTMA%3E2.0.CO%3B2-7

Random-Effects Models for Longitudinal Data

Nan M. Laird; James H. Ware

Biometrics, Vol. 38, No. 4. (Dec., 1982), pp. 963-974.

Stable URL:

http://links.jstor.org/sici?sici=0006-341X%28198212%2938%3A4%3C963%3ARMFLD%3E2.0.CO%3B2-G

Recovery of Inter-Block Information when Block Sizes are Unequal

H. D. Patterson; R. Thompson

Biometrika, Vol. 58, No. 3. (Dec., 1971), pp. 545-554.

Stable URL:

http://links.jstor.org/sici?sici=0006-3444%28197112%2958%3A3%3C545%3AROIIWB%3E2.0.CO%3B2-N

A Crossed Random Effects Model for Unbalanced Data with Applications in Cross-Sectional and Longitudinal Research

Stephen W. Raudenbush

Journal of Educational Statistics, Vol. 18, No. 4. (Winter, 1993), pp. 321-349.

Stable URL:

http://links.jstor.org/sici?sici=0362-9791%28199324%2918%3A4%3C321%3AACREMF%3E2.0.CO%3B2-C