

A GENERAL MODEL FOR THE ANALYSIS OF MULTILEVEL DATA

HARVEY GOLDSTEIN

UNIVERSITY OF LONDON INSTITUTE OF EDUCATION

RODERICK P. McDONALD

MACQUARIE UNIVERSITY

A general model is developed for the analysis of multivariate multilevel data structures. Special cases of the model include repeated measures designs, multiple matrix samples, multilevel latent variable models, multiple time series, and variance and covariance component models.

Key words: complex surveys, mixed effects models, multivariate linear model, repeated measures, time series, variance component models.

1. Introduction

Methods for estimating and testing the fit of linear models, such as regression models with fixed regressor (explanatory) variables, or general models for linear structural relations, are well understood for the analysis of a simple random sample, or for the simultaneous analysis of random samples from distinct populations (see, for example, Jöreskog & Sörbom, 1979; McArdle & McDonald, 1984). The extension of such methods to multilevel data, as given, for example, by samples of randomly drawn students from randomly drawn classes from randomly drawn schools, has presented both theoretical and practical difficulties. Aitkin and Longford, (1986), describing the application of a 2-level model to educational data, with students nested within schools, show that it is misleading either to aggregate the student variables to form school means for a "means-on-means" analysis, or to ignore the hierarchical structure of the data. It can be expected that the application of a factor model or a model for linear structural relations to multilevel data similarly will need to take into account the structure of the data. For example, existing models and computer software do not cater for the important case in which the factor structure of a set of tests given to random samples of students within a random sample of schools, results from distinct factor structures for the between-school covariance matrix and the between-student-within-school covariance matrix. Similarly, existing models for linear structural relations do not allow causal modeling of relationships between school, class and student characteristics. The general theory to be described covers such cases.

In addition to the contribution from Aitkin and Longford (1986), in recent years a number of articles have described particular models for the analysis of certain kinds of multilevel data structures. Thus, Strenio, Weisberg, and Bryk (1983) show how maximum likelihood estimates can be obtained for growth curve data viewed as part of a two-level structure. Goldstein (1986a) shows how a general p -level multivariate mixed effects model

We would like to acknowledge the helpful comments of Ruth Silver. We also wish to thank the referees for helping to clarify the paper. This work was partly carried out with research funds provided by the Economic and Social Research Council (U.K.).

Requests for reprints should be sent to Harvey Goldstein, Institute of Education, 20 Bedford Way, London, UNITED KINGDOM WC1H 0AL.

can be defined and analyzed with allowance made for possible measurement errors in the explanatory variables.

In the present paper we develop the model in Goldstein (1986a) into a more general one which includes the above models as special cases and among others contains models for multilevel structural relations, possibly with latent variables, generalizability theory and general repeated measures designs. In the next section we describe the basic general model and following that, several special cases of interest.

2. A General Multilevel Model

Let y be a $(n \times 1)$ vector of responses obtained from a hierarchical design with h levels of nesting, and ordered to reflect the hierarchy. The i th response, the i th component of y , may be indexed by writing

$$y_i = y_{i_h i_{h-1} \cdots i_1},$$

to indicate that it corresponds to the i_1 -th first-level unit within the i_2 -th second level unit within $\cdots$ the i_h -th h -level unit. For example, we might have students as first-level units randomly drawn from classes as second level units randomly drawn from schools as third level units, and write

$$y_{i_3 i_2 i_1}$$

for the response of the i_1 -th student in the i_2 -th class in the i_3 -th school. In some applications one level of nesting may correspond to repeated observations as in longitudinal or time series data (see sections 5 and 6). Further, the response vector may include responses defined at any level of the hierarchy. Thus, for example, schools classes and students may all be characterized by one or more response variables. Missing response data at any level require no special treatment; the corresponding component is simply omitted from y .

We suppose that the components of y are ordered in a natural way, such that y may be partitioned in h ways, according to each of the levels of the hierarchy, with the subvectors of each level nested within subvectors of a higher level. Corresponding to the k -th level, y is partitioned into n_k subvectors, and we write s_{mk} for the total number of observations (usually of level one units) in the m -th k -level unit.

We write the linear model

$$y = X\beta \quad (1)$$

where X is a $(n \times p)$ "design" matrix and β is a $(p \times 1)$ vector of unknown coefficients, in a suitable form for the hierarchical design, namely;

$$y = X_0 \beta_0 + \sum_k X_k \beta_k \quad k = 1, \dots, h \quad (2)$$

where β_0 is a $(p_0 \times 1)$ vector of fixed parameters, and β_k consists of n_k random subvectors β_{km} , $m = 1, \dots, n_k$, each of p_k components, that are independently and identically distributed with mean zero and covariance matrix Ω_k , of order p_k , whence

$$\text{cov}(\beta_k) = I_{n_k} \otimes \Omega_k \quad (3)$$

It is assumed that;

$$\text{cov}(\beta_k, \beta_m) = 0, \quad k \neq m \quad (4)$$

That is, random variables at different levels of the hierarchy are supposed uncorrelated.

As illustrations of the model, consider two simple cases. In the first, suppose we have

a two level hierarchy consisting of students within a population of schools. The response y_{ji} of the i -th student in the j -th school, is given by;

$$y_{ji} = \beta_{01} + \beta_{02} x_{ji} + \beta_{03} z_j + \beta_{21j} + \beta_{22j} x_{ji} + \beta_{1ji}, \quad i = 1, \dots, l_j; j = 1, \dots, n_2.$$

The coefficients β_{01} , β_{02} , β_{03} are fixed parameters defining a linear regression of y_{ji} on the student level explanatory variable x_{ji} and the school-level explanatory variable z_j . The coefficients β_{21j} and β_{1ji} are respectively the simple between-school and between-student-within-school residuals, and β_{22j} is a random coefficient of x_{ji} , varying across level-two units, that is between schools. The β_{21j} and β_{22j} are in general correlated but uncorrelated with β_{1ji} . There are n_2 schools, with l_j students from the j th school. In the notation of (2), $\mathbf{y}$ is a $(n \times 1)$ vector where;

$$n = \sum_j l_j, \quad j = 1, \dots, n_2$$

$$\boldsymbol{\beta}_0^T = (\beta_{01} \quad \beta_{02} \quad \beta_{03})$$

and X_0 ($n \times 3$) has $(1 \ x_{ji} \ z_j)$ as its ji -th row; $\boldsymbol{\beta}_1$ ($n \times 1$) has β_{1ji} as its ji -th component and $X_1 = I_n$; $\boldsymbol{\beta}_2$ ($2n_2 \times 1$) has $(\beta_{21j} \ \beta_{22j})^T$ as its j -th subvector, and X_2 is a diagonal block matrix ($n \times 2n_2$) whose j -th diagonal block ($l_j \times 2$) has $(1 \ x_{ji})$ as its i -th row. The β_{1ji} are i.i.d. with variance σ_1^2 and the $(\beta_{21j} \ \beta_{22j})^T$ are i.i.d. with covariance matrix Ω_2 , whence;

$$\text{cov}(\boldsymbol{\beta}_1) = \sigma_1^2 I_n;$$

$$\text{cov}(\boldsymbol{\beta}_2) = I_{n_2} \otimes \Omega_2$$

As an example of a different kind, again with, say, students within schools, suppose we have one measure y_{j1} characterising the j th school, and two response measures y_{ji2} , y_{ji3} , characterising the i th student in the j -th school, $i = 1, \dots, l_j; j = 1, \dots, n_2$. We write;

$$y_{j1} = \beta_{j1}$$

$$y_{ji2} = \beta_{j2} + \beta_{ji2}$$

$$y_{ji3} = \beta_{j3} + \beta_{ji3}.$$

In the notation of (2), $\boldsymbol{\beta}_1$ is a $(2n \times 1)$ vector whose n subvectors, $(\beta_{ji2} \ \beta_{ji3})^T$ are i.i.d. with covariance matrix Ω_1 , and $\boldsymbol{\beta}_2$ is a $(3n_2 \times 1)$ vector whose n_2 subvectors $(\beta_{j1} \ \beta_{j2} \ \beta_{j3})^T$ are i.i.d. with covariance matrix Ω_2 . Table 1 shows, in extenso, a simple case of this model with just two schools, three students in one, two in the other, and missing responses y_{133} , y_{222} .

Table 1

$\begin{bmatrix} y_{11} \\ y_{112} \\ y_{113} \\ y_{122} \\ y_{123} \\ y_{132} \\ y_{21} \\ y_{212} \\ y_{213} \\ y_{223} \end{bmatrix}$	$=$	$\begin{bmatrix} 100000 \\ 010000 \\ 001000 \\ 010000 \\ 001000 \\ 010000 \\ 000100 \\ 000010 \\ 000001 \\ 000001 \end{bmatrix}$	$\begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \beta_{21} \\ \beta_{22} \\ \beta_{23} \end{bmatrix}$	$+$	$\begin{bmatrix} 0000000000 \\ 1000000000 \\ 0100000000 \\ 0010000000 \\ 0001000000 \\ 0000100000 \\ 0000000000 \\ 0000001000 \\ 0000000100 \\ 0000000001 \end{bmatrix}$	$\begin{bmatrix} \beta_{112} \\ \beta_{113} \\ \beta_{122} \\ \beta_{123} \\ \beta_{132} \\ \beta_{133} \\ \beta_{212} \\ \beta_{213} \\ \beta_{222} \\ \beta_{223} \end{bmatrix}$
--	-----	--	--	-----	--	--

Complete Design for Multivariate Example

From this example it may be seen more generally how the design matrix handles multivariate data with randomly missing responses. From another point of view the model can be described as a three-level model with schools as the third level, students as the second level, and (response) variables as the first level, but without terms representing variation between first-level units. This approach (see below), which regards the multivariate model as a special case of a univariate model, is described more fully in Goldstein (1986b).

Since in general n may be very large, procedures for estimation in this model are based on the use of expressions for the covariance matrix of $\mathbf{y}$, its inverse and (for maximum likelihood estimation) its determinant, that take a suitable form for computation. Generally, as in the examples given, the design matrix X_k is block diagonal, partitioned by rows corresponding to the k -level partitioning of $\mathbf{y}$, into diagonal blocks X_{km} , of order $(s_{mk} \times p_k)$. Hence we have;

$$\begin{aligned} V = \text{cov}(\mathbf{y}) &= E\{(\mathbf{y} - X_0 \beta_0)(\mathbf{y} - X_0 \beta_0)^T\} \\ &= \sum_k X_k(I_{n_k} \otimes \Omega_k)X_k^T \\ &= \sum_k \bigoplus_m X_{km} \Omega_k X_{km}^T, \quad m = 1, \dots, n_k \end{aligned} \quad (5)$$

where $\bigoplus$ is the direct sum operator.

Note that each of the k terms in this sum is a diagonal block matrix, with diagonal blocks of order $(s_{mk} \times s_{mk})$, $m = 1, \dots, n_k$; $k = 1, \dots, h$. Note also that this generalizes somewhat the model in Appendix 2 of Goldstein (1986a) where it is proposed that $s_{m1} = 1$. The two models can be made formally identical, however, by considering an $(h+1)$ -level model where level one in Goldstein has no random variation between units and defines a set of multivariate responses using level one (0, 1) dummy explanatory variables, the coefficients of which are random at level two and possibly at higher levels. The number of observations in the m th level- k unit is s_{mk} which may be large in applications.

A convenient recursive expression for V is obtained by writing

$$V = V_h,$$

where

$$V_{g+1} = V_g + X_{g+1}(I_{n_{g+1}} \otimes \Omega_{g+1})X_{g+1}^T, \quad (6)$$

or

$$V_{g+1} = V_g + \bigoplus_m X_{g+1,m} \Omega_{g+1} X_{g+1,m}^T, \quad m = 1, \dots, n_{g+1}. \quad (7)$$

This yields, further, a recursive expression for V^{-1} , namely,

$$V_{g+1}^{-1} = V_g^{-1} - V_g^{-1} X_{g+1} (I_{n_{g+1}} \otimes \Omega_{g+1}) \{I + X_{g+1}^T V_g^{-1} X_{g+1} (I_{n_{g+1}} \otimes \Omega_{g+1})\}^{-1} X_{g+1}^T V_g^{-1} \quad (8)$$

by the use of a well-known identity for the inverse of matrices of this form (see Goldstein 1986a). A computationally convenient recursive expression for the determinant of V is given in the appendix. By the use of these recursive relations, the largest matrix requiring

inversion by numerical methods is of order equal to the largest number of random coefficients at any level.

These expressions provide a basis for practical procedures, applicable to large data sets, for maximum likelihood estimation of the parameters β_0 and Ω_k , $k = 1, \dots, h$ of the general model (2), or of further parameters of which these are restricted to be functions. In effect they reduce an initial numerical problem in (generally) very large sparse matrices to a tractably compact form.

For a wide class of models following Goldstein (1986a), an iterative generalized least squares (IGLS) procedure can be recommended, as it is relatively straightforward and makes it easy to incorporate adjustments for errors in variables not available with standard maximum likelihood (ML) approaches. Nevertheless, as shown in appendix 1 of Goldstein (1986a), IGLS is equivalent to ML for a wide class of models when the random variables have a multivariate normal distribution. The essence of the IGLS algorithm is as follows.

If $\Omega = \{\Omega_1, \dots, \Omega_n\}$ and hence V are known then the GLS estimates are given by

$$\hat{\beta}_0 = (X^T V^{-1} X)^{-1} X^T V^{-1} y \quad (9)$$

When Ω is unknown, the numerical algorithm proceeds by choosing a starting value (typically the ordinary least squares (OLS) value $V = I\sigma^2$). The residuals

$$\tilde{y} = y - X\hat{\beta}_0$$

and hence the matrix

$$Z = \tilde{y}\tilde{y}^T$$

is calculated, and if the model is correctly specified,

$$E(Z) = V.$$

As is evident from (7), V is a known linear function of the parameters of Ω . These parameters are estimated from a GLS regression of $\text{vech}(Z)$ on D , where D is the design matrix relating V to a linear function of the elements of Ω as in (7), and the current estimate of $\text{cov}\{\text{vech}(Z)\}$ is used as the weight matrix. The vector $\text{vech}(Z)$ is formed by stacking the columns of the lower triangle of the symmetric matrix Z under one another. Details of this procedure are given by Goldstein (1986a), who also indicates how algorithms can be developed for efficient use of computer storage. The estimated parameters of Ω are then used to form an updated estimate of V which is used in (9) recursively until convergence is obtained. Extensive numerical work will be needed to compare alternative numerical algorithms, but the principles governing such work now are clear.

In the present paper we develop the model in Goldstein (1986a) into a more general one which includes the above as special cases. We show that this general model contains the following special cases: time series models, models for longitudinal data, multiple matrix sampling models, generalizability theory models, multilevel common factor models and multilevel counterparts of models for linear structural relations (path analysis) with latent variables, and complex sample survey designs.

The general model provides both a theoretical unification of a wide range of models for multivariate and univariate data and general procedures for fitting these models by ML or IGLS. The theory is developed in terms of a hierarchical model, for simplicity of presentation, but in section 11 we show that it readily generalizes to include cross classifications for random factors, thus incorporating the class of variance and covariance component models.

3. Parameter Constraints

From the model specification it is clear that the structures of the fixed and random parts of the model can be defined quite separately. In effect this involves constraining subsets of the parameters to be fixed functions, typically by requiring a set of linear functions to be zero. The simplest example is the well known case of so called "regression through the origin" where

$$y_j = \alpha_j + \beta x_j$$

and

$$E(\alpha_j) = 0, \quad \text{var}(\alpha_j) = \sigma^2.$$

In addition, however, we can jointly constrain the fixed and random parameters with the constraining function being imposed at each cycle of iteration. An example is a "constant coefficient of variation" model where the square of the predicted response is equal to the simple between-individual variance at level 1. An interesting class of models which could be studied is that where a subset of the random parameters is constrained so that the values are known functions of the fixed part of the model, and possibly other random parameters, by virtue of particular distributional assumptions.

4. Conditional Estimates

We can obtain conditional estimates of β_k , $k = 1, \dots, h$ by ordinary regression. These may be referred to as conditional or posterior means or generalized (shrunk) residuals at level k . (See Aitkin & Longford, 1986 for a discussion). By (2), (3) and (4) we have

$$\text{cov}(\mathbf{y}, \beta_k) = X_k(I_{n_k} \otimes \Omega_k) = W_k, \quad (10)$$

whence we obtain an estimate of

$$\beta_k | \mathbf{y} - X_0 \beta_0, \Omega_k$$

as

$$\hat{\beta}_k = W_k^T V^{-1}(\mathbf{y} - X_0 \beta_0)$$

with

$$\text{cov}(\hat{\beta}_k) = W_k^T V^{-1} \{ V - X_0(X_0^T V^{-1} X_0)^{-1} X_0^T \} V^{-1} W_k.$$

A consistent estimator of $\text{cov}(\hat{\beta}_k | \mathbf{y})$ is given by substituting parameter estimates in

$$(I_{n_k} \otimes \Omega_k) - W_k^T V^{-1} W_k.$$

If multivariate normality is assumed we obtain the same estimators. On substituting estimates for V and W_k we can provide standardized shrunk residuals to assess model adequacy, or for the higher level estimates to give estimated residuals for the individual units. These estimates can be ranked or otherwise compared (Aitkin & Longford, 1986), although care needs to be taken to ensure that there is no model misspecification which might alter them appreciably.

5. Time Series Models

Single level time series models typically lead to structures for V which involve nonlinear (product) functions of the parameters to be estimated. Without going into great detail we shall suppose that suitable nonlinear estimation procedures are available, and

we see that if there are level 1 random errors which have a time series structure, then for a multilevel model the estimation procedure follows the lines already described.

Consider, for example, the first order autoregressive model, with a single residual term at Level 1, with the following structure:

$$e_t = \rho e_{(t-1)} + v_t,$$

where

$$E(\mathbf{v}\mathbf{v}^T) = I\sigma^2, \quad E(v_t) = E(e_t) = 0.$$

We have

$$W = E(\mathbf{e}\mathbf{e}^T) = \sigma^2(1 - \rho^2)^{-1}R,$$

where the (i, j) -th element of R is $\rho^{|i-j|}$, and W^{-1} has a simple form.

An alternative to a full nonlinear analysis of such models is to use the estimated residuals $\hat{\mathbf{e}}$ at the current iteration as response variables to obtain estimates of ρ and σ^2 which are used for the next cycle of iteration. Such a procedure will generally give consistent estimates for a wide class of models. One important application is where there is one short time series on each of a number of randomly chosen units. By imbedding the time series model within the general multilevel linear model we can provide efficient estimates of the common parameters, and at the same time incorporate covariates in a straightforward fashion.

6. Longitudinal Data

The usual multivariate models for the analysis of repeated measurement data can be reformulated in terms of a 2-level model, in which individual subjects are at Level 2 and measurements at occasions within subjects are at Level 1. Thus, for example, the 2-level polynomial growth curve model can be written as

$$y_{it} = \sum_j \beta_{ij} X_t^j + \sum_k \alpha_k z_{ik} + \varepsilon_{it}, \quad (11)$$

where

$$\beta_{ij} = \beta_j + \gamma_{ij},$$

where the γ_{ij} are random variables at Level 2, x_t is age or time, ε_{it} is a random variable at Level 1, and $U = \text{cov}(\gamma_{ij})$ is the covariance matrix of the polynomial coefficients between subjects. The second term in (11) allows us to specify further explanatory variables and levels. We can also further model the ε_{it} as a function of time so that the within-subject residual variation varies with time. The principal advantage of this formulation of the model is that it places no restrictions on the number or spacing of occasions, which has been one of the drawbacks of the traditional multivariate models (Goldstein, 1979).

We return to the case of a fixed set of p measurement occasions, and write (11) as

$$y_{ij} = \sum_j \beta_{ij} w_j + \sum_k \alpha_k z_{ijk}, \quad (12)$$

where

$$w_j = \begin{cases} 1 & \text{if the measurement is at occasion } j \\ 0 & \text{if not} \end{cases}$$

There is now no Level 1 variation and (12) is essentially the model (Jones, 1980) for providing efficient estimation of means in a "mixed longitudinal" or "rotated design" study. The model (12) is also more flexible than that of Jones in that further fixed or

random explanatory variables and levels can be incorporated, and any pattern of missing data can be handled easily. Further details with examples are given in Goldstein (1986b).

7. Multivariate Models and Matrix Sample Designs

For the case of complete p -variate data obtained from units sampled at two levels, say p response measures on students in schools, we consider the simple model

$$y = X_0 \beta_0 + X_1 \beta_1 + X_2 \beta_2, \quad (13)$$

where the ji -th subvector of β_1 , β_{1ji} and the j -th subvector of β_2 , β_{2j} , are $(p \times 1)$ vectors with covariance matrices Ω_1 , Ω_2 respectively representing between-student-within-school covariance and between-school covariance.

Alternatively, for the m -th Level 2 unit, we write

$$y_m = X_{m0} \beta_{m0} + X_{m1} \beta_{m1} + X_{m2} \beta_{m2}, \quad (14)$$

where

$$X_{m2} = \mathbf{1}_{t_{1m}} \otimes I_p;$$

$$X_{m1} = I_{t_{1m}} \otimes I_p,$$

β_{m1} is $(p \times 1)$ with covariance matrix Ω_1 , and β_{m2} is $(pt_{1m} \times 1)$ with covariance matrix

$$I_{t_{1m}} \otimes \Omega_2,$$

where t_{1m} is the number of Level 1 units in the m -th level 2 unit. Thus with complete data we have

$$V_m = \text{cov}(y_m) = \mathbf{1}_{t_{1m}} \mathbf{1}_{t_{1m}}^T \otimes \Omega_2 + I_{t_{1m}} \otimes \Omega_1, \quad (15)$$

in which case the inverse and determinant take simple forms. More general cases, with further fixed and random coefficients, with missing observations or with variables at different levels, can be handled in the way illustrated by Table 1.

Other multivariate models can be derived. For example, the multiple discriminant function can be derived from estimates of "within" and "between" group covariance matrices. As in Table 1, designs where the multivariate vector consists of responses defined at several levels of the hierarchy will have applications, for example in path analysis educational models where, say, a second occasion response vector may consist of students' achievements related to previous measurements and (Level 2) teacher attitudes which are also related to previous teacher and student variables. Another application is in repeated measurement studies where, for example, responses are (level 1) growth measurements and (Level 2) measurements such as final growth status with the latter being correlated at Level 2 with the polynomial coefficients.

An application, of importance in surveys of educational achievement, is to provide efficient analysis of multiple matrix designs where individual subjects provide measurements on only a subset of possible response variables. Hence this is a particular case of a multivariate linear model with missing responses. Such designs, typically, are carried out in the context of hierarchical educational or social systems, the characteristics of which can be incorporated as higher levels in the same model, as can further explanatory variables. An example of such an analysis is given by Goldstein and Silver (in press) for a design where students are sampled within schools and each student takes a common core mathematics test and one of four "rotated" forms, with explanatory variables measured at both the student and teacher level.

8. Generalizability Theory

According to Shavelson and Webb (1981) in their review of generalizability theory, estimation has been its "achilles heel". In fact, generalizability models are simply variance or covariance component models (see section 11) where a test score response is related to parameters which vary between subjects and between tests or items within subjects. They are therefore special cases of the general multilevel model. Moreover, the general model provides efficient estimation procedures and flexibility with respect to incorporating extra levels for an educational system and further explanatory variables. It also allows a more complete specification of the covariance matrix V by including covariances between the random parameters; a typical omission in traditional generalizability analyses (Johnson and Bell, 1985). In our view, it should become easier to evaluate the validity of generalizability analysis when it is expressed in terms of the notation and terminology of multilevel models, than when using the specialized notation found in the generalizability literature.

9. Latent Variable Models

We now turn to the application of multilevel models where some or all of the random coefficients depend on further unobserved or "latent" variables. Existing formulations of such latent variable or covariance structure models are concerned exclusively with a single level and we shall demonstrate how multilevel latent variable models can be specified in a natural way.

Consider the general multilevel model given by (1), (2) and (3). We define the multilevel latent variable model as one where some components of the vectors β_k are dependent on a set of unobserved variables. We note that any other fixed or random explanatory variables will already have been incorporated into $X_0\beta_0$. Any of the standard formulations for specifying latent structures may be used. In particular the Reticular Action Model (McArdle & McDonald, 1984; McDonald, 1985), a general latent variable path model which makes no distinction between response and explanatory variables, is readily adaptable to the hierarchical model. Thus, for example, we can define a 2-level common factor model by writing

$$\begin{aligned}\beta_{2j} &= A_{w_j} + u_j \\ \beta_{1ji} &= Bz_{ji} + e_{ji}\end{aligned}\tag{16}$$

in (13) with the usual assumptions defining a common factor model at each level, yielding

$$y_{ji} = X_0\beta_0 + X_{1ji}(Bz_{ji} + e_{ji}) + X_{2j}(A_{w_j} + u_j),\tag{17}$$

where y_{ji} is a p -variate vector corresponding to the i -th level 1 unit in the j -th level 2 unit. By choosing various definitions of X_1 , X_2 in (13), the latent variables at each level can be independent or related in part to the same explanatory variables.

An interesting special case is obtained by choosing X_1 , X_2 as above for (13). Then the ji -th observation vector becomes

$$y_{ji} = X_{0ji}\beta_0 + A_{w_j} + u_j + Bz_{ji} + e_{ji}\tag{18}$$

where in general A ($p \times q$) and B ($p \times r$) may be of different column orders. Equation (18) can be regarded as a generalization of the classical model for simultaneous factor analysis in several groups (Jöreskog and Sörbom, 1979), with the response variables being functions of between group plus within group factors. Further, for some groups a separate

factor structure may be fitted, so allowing a general model which incorporates the classical simultaneous factor model as a special case.

More generally, as in the second example in section 2, this formulation of the multi-level linear model allows response variables at different levels and possibly with missing data for which it provides efficient estimates in the ordinary single-level case. It thus becomes possible to write more general counterparts of existing linear structural relations models. Thus, in the context of educational research we may have causal models relating latent variables associated with schools, with teachers and with students. In the context of econometrics, for example, we may have relationships between latent characteristics of macro systems and those of micro systems. As in section 4 we can obtain conditional estimates for factor scores and residuals.

The efficient estimation of the parameters of a general linear multilevel latent variable model proceeds as follows. At each cycle of the iteration for the basic multilevel model, the current estimates of $\text{cov}(\beta_k)$ are used to estimate the parameters of the latent variable model using a standard efficient algorithm. These parameters are then used to construct new predicted estimates of the $\text{cov}(\beta_k)$ which are then used in the next iteration cycle. Thus there is a double cycle of iterations which takes place. Such a procedure, therefore, may make heavy computing demands and a compromise, less efficient but still consistent, is to use only the final estimates of $\text{cov}(\beta_k)$ from the multilevel fitting in the latent variable estimation. A study of this and other aspects of multilevel latent variable models, will be reported elsewhere (McDonald & Goldstein, in preparation).

10. Sample Surveys

Complex sample surveys with multistage designs are an example of hierarchical structures which can be modeled in the ways already described, where a "super-population" structure is assumed, and where stratification factors can be modeled as fixed effects. The advantage of a direct modeling approach is that efficient parameter estimates can be obtained and a wide variety of models fitted as a standard procedure. We also note that aspects of a hierarchical population structure not included in a sampling scheme can be modeled. Thus, for example, we might sample schools and children within schools, but we can also include classrooms as an extra level in a model.

Often in sample surveys, Level 1 units will have differing selection probabilities and hence, according to standard sample survey theory, require different weights in an analysis. If w_j is the weight of the j -th unit and $\sum_j w_j = 1$, and $W = \text{diag}\{w_j\}$, then the analysis proceeds as described above with y being replaced by $W^{0.5}y$ and X by $W^{0.5}X$. Where there are response variables at several levels in a multivariate model we will define separate sets of weights for each level with each set summing to unity. The weights attached to a higher level unit need not be a function of the weights for the lower level constituent units, but for example could be based on the higher level unit's own selection probability. In addition, a priori weighting considerations may influence the weights assigned. Nevertheless, in models where responses at different levels are all related to the same set of explanatory variables, care will be needed in choosing weights across levels and this is an issue requiring further study.

11. Variance and Covariance Component Models

In previous sections we have dealt with models where the random part has a fully nested structure. In some cases, however, the variation in the random terms between the units at one or more of the levels may itself be structured at that level. Thus, for example,

in a two level model, with students at Level 1, the Level 2 units may be cross classified by school and by neighborhood, with the total variation between the resulting "cells" of the cross classification consisting of the sum of a component between schools and a component between neighborhoods. Suppose that there are r factors classifying the level k units. The contribution to the random variation at level k for a particular level $k + 1$ unit, is given by:

$$\sum_i U_{(k)i} \beta_{(k)i}, \quad i = 1, \dots, r \quad (19)$$

where $U_{(k)i}$ is the $(n_{(k+1)} \times pq_i)$ design matrix for the i -th factor, $\beta_{(k)i}$ is the (pq_i) vector of random terms for the i th factor, p is the number of explanatory variables defining the random variables at level k , $n_{(k+1)}$ is the number of Level 1 observations in the level $k + 1$ unit, and q_i is the number of design variables for factor i .

We have therefore

$$\text{cov}(\beta_{(k)i}) = I \otimes \Omega_{k,i},$$

and we assume;

$$\text{cov}(\beta_{(k)i}, \beta_{(k)j}) = 0$$

where $\Omega_{k,i}$ is the $(p \times p)$ covariance matrix of the random coefficients at level k .

Note that for high order cross classifications, the factors referred to above may comprise "interaction" terms as well as main effects. Furthermore, some of the cells of the cross classification can be missing.

The first term in the brackets on the right hand side of (6) now becomes modified to

$$\bigoplus_h \{V_{(k),m,n}\} + \sum_i U_i(I \otimes \Omega_{k,i})U_i^T, \quad (20)$$

where $V_{(k),m,n}$ now represents the contribution of units at levels $1, \dots, k - 1$.

This is a generalization of the strictly hierarchical model, but can be analyzed using the same general procedures and involving in general the inversion of matrices of order pq_i . All the special cases considered in previous sections can be generalized to incorporate the structure of (20).

In the special case where $p = 1$ in a 2 level model with a simple random term at Level 1, we have the usual r -way variance component or mixed effects model.

12. Discussion

A number of other issues not discussed in detail also arise. We have shown how to deal with missing responses in multivariate response data, but the case of missing explanatory variables in a multilevel model requires further investigation. We have discussed models with particular covariance structures at Level 1. In some applications higher level units also may have such structures, for example a time series structure. In principle the methods we have used can be extended to such cases. There is the issue of assigning weights to units at different levels which we have discussed, and there is the issue of robustness of the models against alternative distributional assumptions, and here it seems we require both theoretical and empirical investigations. The estimates given for the standard errors of the parameter estimates assume that V is known, and hence ignore its sampling variation. Further study of this problem is desirable, especially where there are small numbers of higher level units.

In many scientific areas, but especially in the human sciences, data typically are

obtained from hierarchically structured populations, whether by a deliberate multistage sampling scheme, by a census, or by some other probability method. For such data it is important that statistical model fitting takes account of these hierarchies, since failure to do so can result in inefficient and/or inconsistent parameter estimates. This is true for both observed variable models and latent variable models. The applications described in this paper are intended to show how existing practices can be improved by incorporating the multilevel structure of the data into a range of widely used techniques. The theory identifies the form of an extensive program for the development of computer software. We anticipate that further applications of the ideas in this paper will be found for other models, including further generalized linear models (McCullagh & Nelder, 1983) and nonlinear models. The methodology presented in this paper can be applied generally where data are obtained from replicated surveys or experiments. Thus, so called meta-analysis procedures can be studied in terms of "meta-parameters" representing the variation between individual parameters across studies. These latter parameters might be derived, for example, from regression models or, say, from latent variable models. Finally, the ability of the models to handle random coefficients provides a powerful tool for the detailed exploration of the structure of the random component of linear models, and thus opens up a promising new area of data analysis.

Appendix *Estimating the Determinant of V*

Let $V_{g+1,m}$ be the m -th diagonal block of V_{g+1} . Then,

$$|V_{g+1}| = \prod_m |V_{g+1,m}|, \quad m = 1, \dots, n_{g+1} \quad (\text{A1})$$

where

$$V_{g+1,m} = \bigoplus_t V_{g,t} + X_{g+1,m} \Omega_{g+1} X_{g+1,m}^T, \quad t = 1, \dots, l_{g+1,m} \quad (\text{A2})$$

We partition $X_{g+1,m}$ by rows into $l_{g+1,m}$ submatrices $X_{g+1,m}^{(t)}$ corresponding to the g -level partition of $y_{g+1,m}$, and define also $X_{g+1,m}^{*(t)}$, a matrix partitioned by rows whose submatrices are the first i submatrices, $X_{g+1,m}^{(t)}$, $t = 1, \dots, i$, of $X_{g+1,m}$, and define also

$$V_{g+1,m}^{(i)} = \bigoplus_t V_{g,m}^{(t)} + X_{g+1,m}^{*(i)} \Omega_{g+1} X_{g+1,m}^{(i)T}, \quad t = 1, \dots, i. \quad (\text{A3})$$

Then

$$\begin{aligned} & |V_{g+1,m}^{(i+1)}| \\ &= |V_{g+1,m}^{(i)} + X_{g+1,m}^{*(i+1)} \Omega_{g+1} (I + X_{g+1,m}^{*(i)T} [\bigoplus_t V_{g,m}^{(t)}]^{-1} X_{g+1,m}^{*(i)} \Omega_{g+1})^{-1} X_{g+1,m}^{(i+1)T}| \quad (\text{A4}) \end{aligned}$$

References

- Aitkin, M. A., & Longford, N. (1986). Statistical Modelling in School effectiveness studies. (with discussion). *Journal of the Royal Statistical Society, Series A*, 149, 1-43.
- Goldstein, H. (1979). *The Design and Analysis of longitudinal Studies*. London: Academic Press.
- Goldstein, H. (1986a). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, 73, 43-56.
- Goldstein, H. (1986b). Efficient statistical modelling of longitudinal data. *Annals of Human Biology*, 13, 129-142.
- Goldstein, H., & Silyer (in press). D. Holt & F. Smith (Ed.), Multilevel and multivariate models in survey analysis. In *The Analysis of Complex Surveys*.
- Johnson, S., & Bell, M. (1985). Evaluating and predicting survey efficiency using generalisability theory. *Journal of Educational Measurement*, 22, 107-120.
- Jones, R. G. (1980). Best linear unbiased estimation in repeated surveys. *Journal of the Royal Statistical Society, Series B*, 42, 221-226.

- Jöreskog, K. G. and Sörbom, D. (1979). *Advances in factor analysis and structural equation models*. Cambridge, MA: Abt Books.
- McArdle, J. J., & McDonald, R. P. (1984). Some algebraic properties of the reticular action model for moment structures. *British Journal of the Mathematical and Statistical Psychology*, 37, 234–251.
- McCullagh, P., & Nelder, J. A. (1983). *Generalized linear models*. London: Chapman and Hall.
- McDonald, R. P. (1985). *Factor analysis and related methods*. Hillsdale, NJ: Lawrence Erlbaum.
- McDonald, R. P., & Goldstein, H. (in preparation). Applications of multilevel latent variable models.
- Shavelson, P., & Webb, R. (1981). Generalisability theory: 1973–1980. *British Journal of Mathematical and Statistical Psychology*, 34, 133–166.
- Strenio, T., Weisberg, H. I., & Bryk, A. S. (1983). Empirical bayes estimation of individual growth curve parameters and their relationship to covariates. *Biometrics*, 39, 71–86.

Manuscript received 9/9/86

Final version received 7/13/87