Concepts and Techniques in Modern Geography

Multi-Level Models for Geographical Research

54

Kelvyn Jones BSc PhD

Department of Geography
Portsmouth Polytechnic

1991

ISSN 0306-6142
ISBN 1 872464 04 1

Listing of Catmogs in print

CATMOGS (Concepts and Techniques in Modern Geography) are edited by the Quantitative Methods Study Group of the Institute of British Geographers. These guides are both for the teacher, yet cheap enough for students as the basis of classwork. Each CATMOG is written by an author currently working with the technique or concept he describes.

For details of membership of the Study Group, write to the Institute of British Geographers

1:	Collins, Introduction to Markov chain analysis.	3.00
2:	Taylor, Distance decay in spatial interactions.	3.00
3:	Clark, Understanding canonical correlation analysis.	3.00
4:	Openshaw, Some theoretical and applied aspects of spatial interaction shopping models. (fiche only)	3.00
5:	Unwin, An introduction to trend surface analysis.	3.00
6:	Johnston, Classification in geography.	3.00
7:	Goddard & Kirby, An introduction to factor analysis.	3.00
8:	Daultrey, Principal components analysis.	3.50
9:	Davidson, Causal inferences from dichotomous variables.	3.00
10:	Wrigley, Introduction to the use of logit models In geography.	3.00
11:	Hay, Linear programming: elementary geographical applications of the transportation problem.	3.00
12:	Thomas, An introduction to quadrat analysis (2nd ed.).	3.00
13:	Thrift, An introduction to time geography.	3.00
14:	Tinkler, An introduction to graph theoretical methods in geography.	3.50
15:	Ferguson, Linear regression in geography.	3.00
16:	Wrigley, Probability surface mapping. An introduction with examples and FORTRAN programs. (fiche only)	3.00
17:	Dixon & Leach, Sampling methods for geographical research.	3.00
18:	Dixon & Leach, Questionnaires and interviews in geographical research.	3.50
19:	Gardiner & Gardiner, Analysis of frequency distribution .(fiche only)	3.00
20:	Silk, Analysis of covarience and comparison of regression lines.	3.00
21:	Todd, An introduction to the use of simultaneous-equation regression analysis in geography.	3.00
22:	Pong-wai Lai, Transfer function modelling: relationship between time series variables.	3.00
23:	Richards, Stochastic processes in one dimensional series: an introduction.	3.50
24:	Killen, Linear programming: the Simplex method with geographical applications.	3.00
25:	Gaile & Burt, Directional statistics.	3.00
26:	Rich, Potential models in human geography	3.00
27:	Pringle, Causal modelling: the Simon-Blalock approach.	3.00
28:	Bennett, Statistical forecasting.	3.00
29:	Dewdney, The British census.	3.50

(continued inside back cover)

CONCEPTS AND TECHNIQUES IN MODERN GEOGRAPHY

MULTI-LEVEL MODELS FOR GEOGRAPHICAL RESEARCH

by

Kelvyn Jones BSc PhD
(Department of Geography
Portsmouth Polytechnic)

CONTENTS

I PRELIMINARIES

(i) Summary of argument

Multi-level (ML) models, as their name suggests, operate at more than one level or scale simultaneously. This represents a considerable improvement over the usual single-level model by allowing relationships to vary from place to place, and according to context. ML models are calibrated by shrinkage estimators which make very efficient use of information in comparison to the usual ordinary-least-squares (OLS) estimators. Using a number of practical applications, this CATMOG demonstrates how to specify and interpret ML models.

(ii) Pre-requisites

ML models are sophisticated techniques which were only developed in the 1980's. As such they build on existing procedures. It is therefore essential that the reader is already familiar with:

regression models estimated by OLS (CATMOG 15);
the use of dummy variables in regression (CATMOG's 20, 30);
significance testing using the t and F tests (Silk, 1979).

It would be helpful if the reader also has knowledge of:

properties of estimators (CATMOG 5, p19-21);
autocorrelation (CATMOG 47).

(iii) Organisation of text

Given this demanding set of prerequisites, the aim is to provide a largely intuitive account, relegating derivations and proofs to references. The introduction deals with all the fundamental concepts of ML models in words and pictures. These ideas are then turned into model equations and illustrated with some simulated (but realistic) data on clinic immunisation uptake. A discussion of model estimation follows with comments on appropriate software. The final parts of the text consist of applications of the models to school assessment, and a general consideration of the potential importance of ML models for quantitative geography.

II INTRODUCTION

(i) Hierarchies and levels

'Human society is generally arranged into nested hierarchies'
Moellering and Tobler (1972)

Many data sets have an intrinsically hierarchical nature in which, to use a standard terminology, individuals at the lower level form groups at the higher level. Examples include individual pupils (level 1) who are taught in classes (2) which are organised by schools (3) and which are administered by local education authorities (4). Similarly, citizens (1) vote in wards (2) for politicians to represent them on district (3) councils. In physical geography, there could be soil-erosion plots (1) in fields undergoing specific agricultural activities (2), which are located in different landscape facets (3) within particular drainage basins (4). All the above are cases of 'naturally' occurring hierarchies, but it is also possible to induce such structure by the data collection process. Many large scale surveys (for example the General Household Survey) do not sample respondents at random from all over the country. They use, instead, a cost-saving, multi-stage design (CATMOG 17) in which respondents (1) are selected from certain towns (2) from all standard regions (3).

With both types of hierarchical data, problems stem from ignoring this structure during analysis. Inferential error is likely to occur when inappropriate single-level models are used, and when multi-stage data are pushed through a computer package designed for a random sample. A clear example of this is Aitkin *et al's* (1981) re-analysis of what is called the 'teaching styles' study. The original research (Bennett, 1976), using a single-level model estimated by OLS, had found that progress in test scores for English, Reading and Mathematics during the fourth year of the primary school were significantly influenced by teaching style. These results had a considerable impact, and were credited with influencing Prime Minister, Callaghan, in launching 'the Great Debate' and promoting a return to 'traditional' or formal methods in primary schools. The re-analysis using an improved statistical classification of teaching styles, and ML models to take account of the hierarchical structure of pupils (level 1) within teachers/classes (level 2), found that there were no significant differences due to teaching style.

(ii) Cross-level inference

'The fallacy of the wrong level occurs when properties or relations developed at one level are inappropriately translated to another level ... The fallacy can be committed working downwards or upwards'
(Roberts and Burstein, 1980).

A basic problem facing any geographical research is the choice of appropriate scale. Should the analysis be 'micro', based on the individual level-1 data, or on some groupings of these data at a higher level in a 'macro', ecological analysis? It has been known for some time that results from one scale of analysis cannot be readily transferred to another, the problem of cross-level inference. Pearson (1896) first pointed out that correlations of averages will not give identical results to correlations of ungrouped data. The problem was put clearly by Thorndike (1939) while Robinson (1950) famously demonstrated the ecological fallacy of transferring areal, higher-level results to individuals and Alker (1969) identified the atomistic fallacy of transferring individual-level results to the .aggregate level. Openshaw (1977, and CATMOG 38) demonstrated that an extremely wide range of results could result from different aggregations to a higher level. The ML literature goes one step further in arguing that models that do not simultaneously take account of all the relevant scales of variation are likely to be mis-specified with consequent errors of inference (Jones, 1990a).

Taking the education example, the processes that affect a pupil's performance may operate at several scales simultaneously. It may be postulated that examination performances are affected by individual-level variables (eg ability of the pupil, number of hours of homework undertaken), class-level variables (eg the average ability of peers, size of class), school-level variables (eg teacher/pupil ratios, expenditure per pupil) and 'idiosyncratic' factors for each pupil, class and school. Consequently, working at a single level is likely to lead to a distorted representation of reality. Thus, estimates of the relationship between the average level of attainment in a school and aggregate school characteristics (the so-called ecological 'means on means' analysis) can give very misleading results if individual pupil characteristics are ignored. Similarly, if a model is only estimated at the individual level, the class or school 'contextual' effects are not being taken into account. Being in a class of high achievers may affect an individual student's performance in addition to, or in interaction with, the student's own characteristics.

Such arguments permit a clearer understanding of Robinson's (1950) results. Using data from the 1930 census, he found a weak correlation of 0.20 between being illiterate and being black at the individual level. At the higher level of the state he found a fairly strong correlation of 0.77. Individual blacks are not especially likely to be illiterate, but the states with high levels of black people are the areas with high levels of illiteracy. As Alker (1969) argued, and Hanushek *et al* (1974) showed empirically, there is a missing higher-level variable in Robinson's analysis. Blacks and illiterates in 1930 tended to be concentrated in southern states where the poverty of the blacks provided an inadequate tax base for educational facilities. The lack of these communal facilities increased the likelihood of illiteracy among both whites and blacks. Grouping individuals according to state of residence approximates to grouping to maximise the missing variable of educational provision. From the perspective of **ML** models it is not somehow wrong to model at the ecological scale and right at the individual, it depends on the scale(s) at which the important variables are operating. The choice of appropriate scale becomes in part an empirical question whether or not the postulated relationships exist for the sample under study.

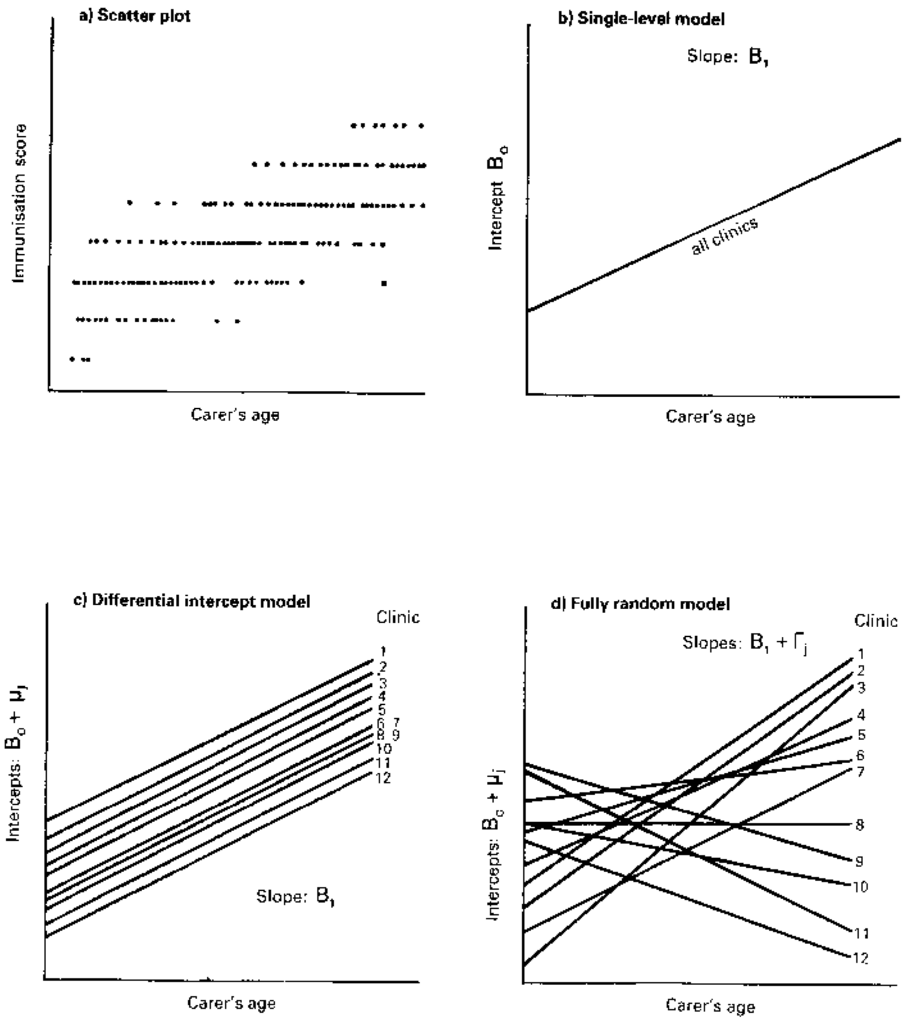
(iii) Relationships in context

'There are no general laws in social science that are constant over time and independent of the context in which they are embedded'
Rein, quoted in King (1976)

The object of statistical modelling is to provide a simplified representation of the underlying population. This is achieved by separating systematic features of the data from random variation. The systematic features are represented in the model by a function (usually linear) involving parameters that relate the response variable to important predictor variables. Random variation is represented by assuming a specific probability distribution (usually the Gaussian, ie the normal) which is then typically captured in a single parameter which measures the average size of this variation.

Figure 1(a) is a plot of a response variable, immunisation scores (total number of immunisations), against the predictor of carer's age. The plot represents data for 600 children (level 1) at 12 clinics (level 2). In a single-level model, this variation is reduced to the single line of (b) to produce one overall relationship, thereby ignoring the higher, clinic level. Thus the systematic component would consist of the two fixed parameters of the equation of the straight line (the

Figure 1 Alternative models



intercept and slope), while the scatter around the line, the random variation, would be reduced to a single measure. In contrast, the remaining two plots show what can be achieved by ML models. Figure 1(c) consists of a set of parallel lines, one for each clinic, which have different intercepts. These random, allowed-to-vary, intercepts permit different clinics to have different levels of uptake around the overall, fixed average for all clinics. Finally in (d), both the intercepts and slopes for each clinic can vary around the overall average for all clinics. Potentially, this allows differing relationships between uptake and parental age for different clinics; there may be clinics where there is little or no relationship, while in others it is quite steep. This last plot represents the simplest fully-random two-level model. As we shall see, it can be extended to have more fixed effects, more random terms, to accommodate complex variation at the lower level, and to higher levels. The elaboration of the models in Figures 1(b) to (d) may appear rather trivial, but it represents a substantial conceptual advance in permitting relationships to vary according to context. Instead of reducing the world to one universal equation, there can be different relationships for different places. To take a pertinent example from electoral geography, the relationship between being a coal-miner and voting Labour (Johnston *et al*, 1988) does not have to be modelled as the same relationship for the entire country. The anticipated differing relationships for the non-NUM coalfields, such as Nottinghamshire, can be accommodated in the model. Problems remain of course: the models are mechanical, the quantitative results obtained from such models do not explain themselves, but the ML procedures do allow the empirical analysis of both the general and the specific in a coherent overall framework.

(iv) Auto-correlation

'Everything is related to everything else, but near things are more related than distant things'

Tobler's(1970) First Law of Geography

It can be anticipated that individuals who are members of the same group at a higher level are more likely to be alike, in some way, than a random sample. That is 'internal correlation' or auto-correlation is to be expected in hierarchical data. Groups are rarely formed at random but rather on the basis of some similarity, and once formed the members of a group may interact with each other to create even greater homogeneity. To take the education example, pupils in the same class are more likely to have similar characteristics than a sample drawn from the entire student population at random. While ML models expect such auto-correlation, single-level models estimated by OLS are troubled by its presence; over-statement of the statistical significance of results will usually result for the critical cut-offs of the t and F tests assume independent observations. The degree of auto-correlation in ML models can loosely be conceived as the ratio of 'variation at the higher level' to the 'total variation of all levels'. A value of zero signifies no auto-correlation, indicating that there is no variation at the higher level; multi-level models will not be needed for all the variability occurs at the individual level. By contrast, a value of one implies that all the relevant variation occurs at the higher level, and that only a higher-level model is needed. Values between the extremes, which are much more likely to be encountered empirically, suggest that a ML model is needed to take account of the variation at both levels. In Aitkin *et al's* (1981) re-analysis of the teaching-styles data, substantial auto-correlation was found between pupils within a class (eg 0.35 for Reading test scores). Such intra-class correlation suggests that the results from a single level model may appear significant when they are not; a speculation confirmed in this case by the ML re-analysis.

(v) Historical context

'It is clear that multilevel modelling is the next great unifying step'
Healy (1989)

It may be useful at this point to summarise the argument by emphasizing the differences between single- and multi-level models. The single-level model assumes that the data does not have a hierarchical structure, that all the relevant variation is at one scale, that there is no auto-correlation, and that there is a single general relationship across time and space. To put it bluntly, this model denies geography and history; everywhere and anytime is basically the same! The single-level model is an impoverished representation of reality, and it is amazing that geographers have been so interested in it. Multi-level models take hierarchical structure into account by modelling the variability at each of the levels, and they allow the individuals that belong to a particular group to be more alike than a random sample. Moreover, within each group in the hierarchy, different relationships are allowed around the overall relationship for all individuals in all groups. The ML model combines the individual with the ecological, the general with the specific.

The basic concepts of multi-level modelling have been known for some time under a variety of labels: variance-component, nested, mixed and hierarchical linear models. Indeed, in geography Chorley *et al* (1966) and Moellering and Tobler (1972) are applications of multi-level models, although they did not use the term. But in these early papers, while the response variable was measured

at a number of different scales (thereby allowing its variability to be disaggregated and attributed to different levels), the models did not include any predictor variables. During the 1970's there was considerable conceptual elaboration of multi-level models in educational sociology (Boyd and Iversen, 1979; Burstein, 1980) and these developments have been presented for a geographical audience by Hauer (1984). But the methods proposed for model estimation by workers in this field were reliant on the inappropriate OLS procedures (Raudenbush and Bryk, 1986). The seminal contributions to a more theoretically justified estimation and a more flexible specification of hierarchical models had in fact been made in Lindley and Smith (1972). Until the 1980's, however, the practical use of these methods had been prohibited by the absence of an efficient computational strategy and associated computer software. So, while the ideas of multi-level models are not new, their actual use is. In the geographical literature, ML models have been used by Davies *et al* (1988), Bondi and Bradford (1990), Jones and Moon (1990). For a more extended account of the history of these statistical developments see Raudenbush (1988).

III SPECIFYING AND INTERPRETING ML MODELS

(i) Specifying the two-level model

Having considered the fundamentals of ML models, discussion now proceeds to specification. This means turning Figure 1 into equations. The most basic of all models is the simple bivariate regression model:

$$y_i = \underbrace{\overbrace{\beta_0 x_0 + \beta_1 x_{1i}}^{\text{systematic}}}_{\text{fixed}} + \underbrace{\overbrace{\varepsilon_i}^{\text{random}}}_{\text{random}} \qquad (1)$$

where the subscript i refers to an individual child. As this model is specified only at the individual level it is known as the micro-model. There are three known terms: y, the response variable is the total number of immunisations completed by the time the child is aged 6 years; x_0, the constant which by convention is set to one for every child; and x_1 is the present age of the parent or guardian who presented the child at the first clinic appointment. To facilitate interpretation, the latter is expressed in deviations from its median so that a figure of 0 represents a 30 year old carer, while a 25 year old, and a 40 year old are given the values of -5 and 10 respectively[1]. There are two fixed parameters: β_0, the intercept, is the average immunisation score of a child with a 30 year-old carer (when x_1 equals 0); β_1, the slope term, is the change in score resulting from the carer being one year older. This equation is the algebraic form of Figure 1(b) in which this single-level model is depicted as a fixed and linear relationship between score and age.

The random or allowed-to-vary element is supposed to be captured by ε. This random term stands for the innumerable, but individually unimportant predictors of immunisation that have not been included in the model as specified[2]. It is typically assumed that the mean or expected value of the random term is zero, that there is constant variability (that is homoscedasticity) and no autocorrelation (that is each element is independent of any other element). Consequently, the random variation can be characterised by a single parameter σ_ε^2, which is the variance of the random term. The complex reality of immunisation variation has been reduced to three parameters: the two fixed terms, namely, β_0, the immunisation score for a child with a parent of average age; and β_1, the score-age relationship; and σ_ε^2, a measure of the variance remaining unaccounted for by the fixed terms.

A more complex model may be developed by letting the average immunisation score vary from clinic to clinic. This can be achieved by allowing the intercept to vary in a higher, level-2, between-clinic, macro-model:

$$\beta_{0j} = \beta_0 + \mu_j \qquad (2)$$

β_{0j}, the average score of a child with a 30 year-old parent in clinic j, is a function of the average score across all clinics, β_0, plus a varying difference, μ_j for each of the clinics. Combining the micro- and macro- models produces a 2-level mixed model:

$$y_{ij} = \underbrace{\beta_0 x_0 + \beta_1 x_{1ij}}_{\text{fixed}} + \underbrace{(\mu_j + \varepsilon_{ij})}_{\text{random}} \qquad (3)$$

[1] This specification of predictor variables in 'deviation' form has two useful properties. First, it improves numerical precision by preventing very large numbers occurring which cannot be adequately stored in the machine. Second, it allows the intercept to be interpreted as the predicted response when the predictors are at their 'average' or typical value. This transformation does not alter the estimates of the other terms in the model.

[2] The term random part is used instead of the more usual description of 'error term' because at the higher level, random terms represent clinic effects or differences. Quite generally in the ML literature, the estimates of the random part are known as residuals.

with the subscript ij denoting child i in clinic j, and the brackets denoting the random part. Figure 1(c) depicts this model as a series of parallel lines with the same fixed score-age relationship (β_1) but varying intercepts ($\beta_0 + \mu_j$) for the clinic average score. The μ_j terms are the clinic-level random terms at level 2; making the same assumptions as for the level-1 random part, they can be summarised in a single variance term, σ_μ^2. The aim of ML modelling is, therefore, to estimate, the fixed intercept value, β_0, which represents the average value, and the variance σ_μ^2 which measures the extent of the variability around this average.

Similarly, the slope can be allowed to vary so that the score-age relationship varies from clinic to clinic. Another level-2 model is needed:

$$\beta_{1j} = \beta_1 + \Gamma_j \qquad (4)$$

that is the clinic slope term is conceived as an average slope for all clinics plus variation from clinic to clinic. Combining equations (1), (2) and (4) produces the full, random-effects model:

$$y_{ij} = \beta_0 x_0 + \beta_1 x_{1ij} + (\Gamma_j x_{1ij} + \mu_j + \varepsilon_{ij}) \qquad (5)$$

in which both the slopes and intercepts are allowed to vary as shown in Figure 1(d). The Γ_j terms are another set of level-2 random terms. Making the usual assumptions, they can be summarised in a single variance term, σ_Γ^2. It is possible that the slopes and intercepts are correlated so that, for example, high clinic average scores are associated with a less steep marginal increase for the score-age relation. This is realised in the model by assuming that the level-2 random terms have a joint distribution with a mean of zero, and a covariance[3] of $\sigma_{\mu\Gamma}$. The correlation between the clinic slope and intercepts is then defined as the ratio of their covariance to the product of their standard deviations:

$$\mathrm{Corr}(\mu_j, \Gamma_j) = \sigma_{\mu\Gamma} / (\sigma_\mu * \sigma_\Gamma) \qquad (6)$$

The random-effects model of equation (5) therefore requires that two fixed coefficients are estimated, three variance/covariances at level 2, and one variance at level 1. The nature of the three random terms can be more readily appreciated if the random part of the model is specified in full:

[3] In general, covariance terms between random coeffficients should always be included in the model; Goldstein (1987, 38) and Davies *et al* (1988, 1072).

$(\Gamma_j x_{1ij} + \mu_j x_0 + \varepsilon_{ij} x_0)$

so that it is clear that the level-2 intercept random term is associated with the constant, as is the level-1 random term. Consequently, the random terms may be depicted in ML models as a variance-covariance matrix at each level in relation to the variables with which they are associated:

LEVEL 2	x_0	x_1	LEVEL 1	x_0	x_1
x_0	σ_μ^2		x_0	σ_ε^2	
x_1	$\sigma_{\mu\Gamma}$	σ_Γ^2	x_1		

If on estimation σ_μ^2 is found to be small in relation to sampling error, there is no tendency for clinics to vary in their average score once age is taken into account; if σ_Γ^2 is small, score-age relationships do not vary greatly from clinic to clinic; if $\sigma_{\mu\Gamma}$ is small, variations in the score-age relationship are unrelated to clinic average score. If all three terms are small, the multi-level equation (5) reduces to the single-level equation (1), the complexity of Figure 1(d) is not needed, and there is no contextual effect for clinics. Any apparent clinic variation is merely the result of the clinic's composition in age of parent. Before calibrating this model, it is useful to make the ML model both simpler (by removing the predictor variable) and more complex by including additional fixed terms. Thus, fitting a constant-only or null, two-level model:

$$y_{ij} = \beta_0 x_0 + (\mu_j + \varepsilon_{ij}) \qquad (7)$$

allows the calculation of the level-1 proportion of residual variability as:

$$\sigma_\varepsilon^2/(\sigma_\varepsilon^2 + \sigma_\mu^2)$$

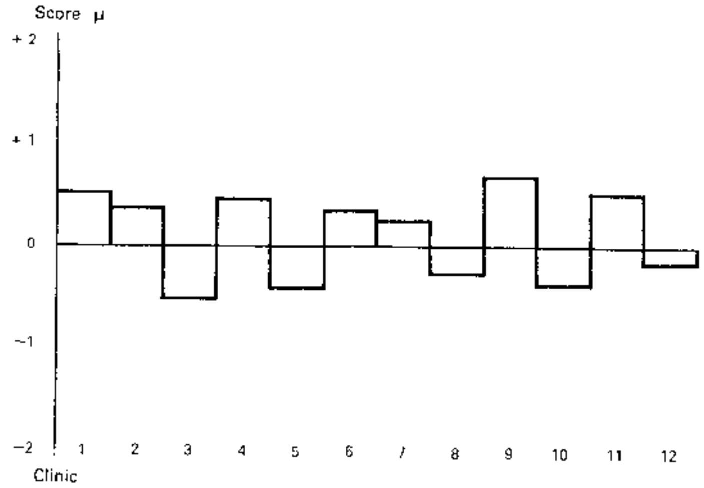
and the level-2 proportion as:

$$\sigma_\mu^2/(\sigma_\varepsilon^2 + \sigma_\mu^2)$$

The latter value represents both the autocorrelation of scores within clinics, and the maximum variability that can be accounted for at the clinic level; if it is zero, a multi-level model is not needed. This proportion cannot be increased for a particular sample, but it can be reduced by including level-1 predictors that vary systematically by clinic, and by including specific level-2 predictors. The level-1 proportion can only be reduced by the inclusion of relevant level-1 predictors, for the level-2 variables are constant within a clinic at the child level.

The null model is also known as the random effects analysis of variance (Searle, 1971), and is equivalent to that used by the geographers Moellering and Tobler (1972) in the early application of ML models. Figure 2 provides a graphic

Figure 2 Random effects analysis of variance model



representation of this model; in effect, differential clinic averages (the μ_j) are allowed to vary around the grand average (β_0) for all clinics. Thus, at the clinic level only the intercepts are allowed to vary:

LEVEL 2	x_0	LEVEL 1	x_0
x_0	σ_μ^2	x_0	σ_ε^2

The second more complex model is specified as follows:

$$y_{ij} = \beta_0 x_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \alpha_1 w_j + (\mu_j + \varepsilon_{ij}) \qquad (8)$$

where there are two extra fixed terms. The variable x_2 is a dummy variable representing parental class; it is coded 1 for non-manual, 0 for manual. Consequently, a positive value for β_2 represents the additional average score for non-manual parents across all clinics holding the other predictor variables statistically constant. The variable w, as denoted by the subscript j, is a level-2 variable that only varies at the clinic level; it represents clinic opening hours.

The fixed term, α_1 is the predicted change in immunisation score if the clinic is open one extra hour. If w is also specified in deviation form, the fixed intercept, β_0 represents the average score for a child with manual parents ($x_2=0$) aged 30, who attends a clinic with the average number of hours set aside for immunisation. In this more general form of the ML model, the higher-level fixed variables can be 'global', as in this case, and only occur at the higher level (Burstein, 1980). Alternatively, they can be 'aggregations' of lower level characteristics such as the proportion of children assigned to a clinic who have non-manual parents. All the elaborations of this model occur in the fixed part, so the random covariance matrices remain:

LEVEL 2		LEVEL 1	
	x_0		x_0
x_0	σ_μ^2	x_0	σ_e^2

Table 1 Immunisation data structure; IMM.DAT

ID1	ID2	IMM	AGE	CLS	HOR	CON
1	1	8	29	0	22	1
2	1	8	34	0	22	1
3	1	8	25	1	22	1
.	.	.	.	.	.	.
.	.	.	.	.	.	.
50	1	9	31	0	22	1
1	2	3	31	0	47	1
2	2	4	23	1	47	1
.	.	.	.	.	.	.
.	.	.	.	.	.	.
50	8	6	29	0	21	1
1	9	7	35	1	35	1
.	.	.	.	.	.	.
.	.	.	.	.	.	.
49	12	6	27	1	51	1
50	12	8	39	0	51	1

(ii) Interpreting the two-level model

Specified ML models have to be estimated and interpreted. The technical aspects of estimation are postponed until a later section, while the present discussion focuses on interpretation. Table 1 shows an extract from a file called IMM.DAT. There are seven variables in the file; the first two provide a unique identifier for each child. ID2 represents the level-2 identifier and goes from 1 to 12 for each of the twelve clinics', while ID1 is the unique code for each

'A sample size of 12 is somewhat small for effective estimation of clinic-level effects but it makes for effective presentation of the models, tables and figures.

child and runs from 1 to 50 for each clinic. The data therefore represent a sample of fifty children all born in the same year at each of twelve clinics. It is not necessary that the clinic sample size should be equal for modern software is not troubled by 'imbalance'. IMM is the codename for the immunisation count for each child, AGE is the parent's age, CLS represents the dummy variable for class (non-manual is 1, manual is 0), HOR is the number of hours per year that the clinic, to which the child has been assigned, provides for immunisation. Clearly, there are only twelve different values for this level-two variable. The final variable is the constant (CON) which is set to 1 for all 600 observations. In the following discussion four alternative models for this data set are specified, evaluated and interpreted. Appendix 1 provides the sequence of commands for one particular software package, ML3 for performing all the analyses discussed here. Both AGE and HOR are not given in deviation form as this can be specified in the program prior to estimation (by the command FMEAN).

Table 2 League tables of clinic performance

CLINIC	MEAN	RANK	MODEL A	RANK	MODEL B	RANK	MODEL C	RANK	MODEL D	RANK
1	8.1	1	1.8	1	1.1	2	0.9	3	0.5	4
2	3.9	12	-1.9	12	-1.9	12	-2.0	12	-2.2	12
3	7.5	2	1.3	2	1.4	1	1.3	2	1.7	2
4	4.3	11	-1.6	11	-1.0	11	-1.1	11	-0.8	9
5	6.9	3	0.7	3	0.7	3	0.7	4	0.6	3
6	5.3	10	-0.7	10	-0.1	7	-0.2	7	-0.1	8
7	6.0	8	-0.1	8	0.6	5	0.5	5	0.3	5
8	6.2	6	0.1	6	-0.8	10	-1.1	10	-0.9	10
9	5.8	9	-0.3	9	-0.4	9	-0.6	9	-1.0	11
10	6.1	7	0.0	7	0.3	6	0.3	6	0.1	6
11	6.7	4	0.5	4	-0.4	8	-0.5	8	-0.0	7
12	6.5	5	0.4	5	0.7	4	1.8	1	1.7	1

Overall mean : 6.10

The second column of Table 2 gives the mean immunisation score for each clinic. There are marked differences between clinics with, for example, clinic 1 having the highest mean at 8.1 signifying the 'best' performance, while clinic 2 appears to have the worst overall uptake with a mean score of only 3.9. This information permits the construction of league-tables of performance; the third column in the table is the rank of each clinic in terms of the average score, with 1 signifying the best performance. Clinic managers could rightly object that this crude performance indicator takes into account neither the characteristics of the client nor the resources of the clinic. It may be argued, for example, that clinic

2's poor performance is the result of receiving a high proportion of young working-class mothers who are thought to be poor attenders. Multi-level model estimates allow an assessment of such claims.

$$\textbf{MODEL A: } y_{ij} = CON + (\mu_j + \varepsilon_{ij}) \quad (9)$$

The first model, A, is a repeat of the null model of equation (7) with the co-variance matrices as specified in Table 3. This constant-only model summarises the data in three statistics: the constant, the between-child variance, and the between-clinic variance(Table 4). The constant of 6.10 is, of course, the overall mean across all children and all clinics. The residual variation about this mean consists of a figure of 0.47 units of variation at the child-level, and a greater variation of 1.36 at the clinic level, that is a total of 1.83. The within-clinic correlation can therefore can be estimated as:

$$\hat{\rho} = 1.36/(0.47 + 1.36) = 0.74 \quad (10)$$

which suggests that children attending each clinic are very alike in their immunisation performance. This figure also gives the maximum variability that can be accounted for at the clinic level (74 percent). In a similar fashion the amount of residual variation at the child-level is 26 percent that is 0.47/(0.47 + 1.36). To summarise, in this model of no fixed parameters except the constant, the fixed terms accounted for 0 percent, the child-level for 26 percent, and the clinic-level for 74 percent of the variation about the mean of the immunisation scores.

Taking the usual cut-off for the ratio of coefficient to standard error of plus or minus two, the pseudo z tests of Table 4 suggest that the level-2 variance of model A is significantly different from zero, thus suggesting that a multi-level model is needed. The data would not be adequately described by a single-level

Table 3 Indicator covariance matrices for Models A to D

		LEVEL 2				**LEVEL 1**
Models A and B		CON				CON
	CON	σ_μ^2			CON	σ_ε^2
Model C		CON	CLS			CON
	CON	σ_μ^2			CON	σ_ε^2
	CLS	$\sigma_{\mu\Gamma}$	σ_Γ^2			
Model D		CON	AGE	CLS		CON
	CON	σ_μ^2			CON	σ_ε^2
	AGE	$\sigma_{\mu\delta}$	σ_δ^2			
	CLS	$\sigma_{\mu\Gamma}$	$\sigma_{\delta\Gamma}$	σ_Γ^2		

Table 4 ML estimates for Models A to C

MODEL	**A**	**B**	**C**	
Fixed terms				
CON	6.10	5.68	5.69	
AGE		0.06(17.3)	0.05(22.6)	
CLS		0.83(22.8)	0.84(4.8)	
HOR		-0.04(2.1)	-0.05(2.4)	
Random terms				
Level 1 : between children				
	CON	CON	CON	
CON	0.47	0.20	0.11	
Level 2 : between clinics				
	CON	CON	CON	CLS
CON	1.36(2.4)	0.85(2.4)	1.1(2.4)	
AGE				
CLS			-0.3(1.6)	0.4(2.4)
Level 2 : clinic level				
CLINIC	CON	CON	CON	CLS
1	1.8	1.1	0.9	0.2
2	-1.9	-1.9	-2.0	0.2
3	1.3	1.4	1.3	0.3
4	-1.6	-1.0	-1.1	0.2
5	0.7	0.7	0.7	0.1
6	-0.7	-0.1	-0.2	0.2
7	-0.1	0.6	0.5	0.1
8	0.1	-0.8	-1.1	0.3
9	-0.3	-0.4	-0.6	0.2
10	0.0	0.3	0.3	0.0
11	0.5	-0.4	-0.5	0.2
12	0.4	0.7	1.8	-1.9
	μ_j	μ_j	μ_j	Γ_j

Figures in brackets refer to ratios of estimates to standard errors

Table 4 (continued) ML estimates for Model D

MODEL

Fixed terms

CON	5.63
AGE	0.05(12.6)
CLS	0.84(4.8)
HOR	-0.05(2.5)

Random terms

Level 1 : between children

	CON
CON	0.11

Level 2 : between clinics

	CON	AGE	CLS
CON	1.2(2.3)		
AGE	0.0(0.8)	0.0(1.7)	
CLS	-0.3(1.4)	0.0(0.4)	0.4(2.4)

Level 2 : clinic level

CLINIC	CON	AGE	CLS
1	0.5	0.0	0.2
2	-2.2	0.0	0.2
3	1.7	-0.0	0.3
4	-0.8	-0.0	0.2
5	0.6	0.0	0.1
6	-0.1	-0.0	0.2
7	0.3	0.0	0.1
8	-0.9	-0.0	0.3
9	-1.0	0.0	0.2
10	0.1	0.0	0.0
11	-0.0	-0.0	0.2
12	1.7	0.0	-1.9
	μ_j	δ_j	Γ_j

Figures in brackets refer to ratios of estimates to standard errors

model for there are different clinic levels of vaccination around the overall average of 6.1. Table 2 gives the clinic-level residuals and their rank in the fourth and fifth columns respectively. Thus clinic l's performance is 1.8 higher than the overall average of 6.10, while clinic 2 is 1.9 down. The ranks of the performance are exactly the same as for the mean score, with clinic 1 achieving the best performance and clinic 2 the worst. This is not surprising for the model has not yet taken into account client and clinic characteristics; it has merely decomposed the variability into two levels.

MODEL B: $y_{ij} = CON + \beta_1 AGE_{ij} + \beta_2 CLS_{ij} + \alpha_1 HOR_j + (\mu_j + \varepsilon_{ij}) \qquad (11)$

This is a repeat of the complex model of equation (8) and as such has the same random structure as model A and therefore the same indicator matrices. The fixed part now includes three predictor variables in addition to the constant term, age and class varying at the individual level, and hours varying at the clinic level. Table 4 shows that at the conventional cut-off, the estimated fixed coefficients for each of the variables are significantly different from zero as is the level-2 residual variance. There is a positive effect for mother's age such that a child of a 40-year old mother in comparison to the child of a mother of 23 can be expected to have about one more immunisation, that is (40-23) * 0.06. The coefficient for class is also positive and implies that the child of non-manual parents can be expected to have 0.83 more immunisations than the child of manual parents. The coefficient of clinic hours (-0.04) is such that a 25 hour increase leads to an (un)expected drop in the immunisation score of 1. This is somewhat against expectations and as the coefficient is only barely significant at 2.1, it should be treated with some caution. It may be, of course, that the district managers have already recognised clinics with poor uptake and allocated them extra resources.

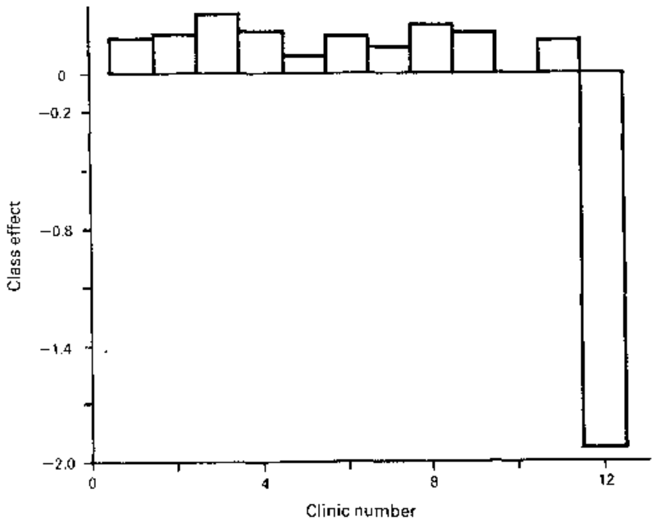
The inclusion of these three fixed terms in comparison to model A has led to substantial reductions in the residual variation, with the between-child variation dropping from 0.47 to 0.20, and the between-clinic variation falling from 1.36 to 0.85. Remembering that the total variability of the scores is 1.83 (that is the total residual variance of the constant-only model), the estimates for model B imply that 11 percent of this variability remains at the individual level (0.20/1.83), and 46 percent remains at the clinic level (0.85/1.83). The fixed terms of the model must therefore account for 43 percent of the variability of the immunisation scores, (1.83-0.20-0.85)/1.83. The clinic-level residuals for model B (Table 2) show some changes of rank when compared to the means-analysis and model A. Clinic 2 still has the worst performance but it is now clinic 3 (and not clinic 1) that is the best; both clinics 8 and 11 have dropped several places in rank. These changes in the size of the clinic residuals can be produced in two ways; either through the inclusion of clinic hours and/or due to the mean of mother's age and the proportion of non-manual classes varying systematically by clinic. Whatever the source, it is clear that the ranks based on means-analysis should not be taken at face value, and a multi-level model is needed to assess overall performance.

$$\text{MODEL C: } y_{ij}=CON + \beta_1 AGE_{ij} + \beta_2 CLS_{ij} + \alpha HOR_j + (\Gamma_j CLS_{ij} + \mu_j + \varepsilon_{ij}) \quad (12)$$

Model C has the same fixed part as model B but an extra random coefficient at the clinic level to reflect differential treatment of parents from different class backgrounds. The more complicated structure is reflected in the covariance matrix which shows that four random variance terms need to be estimated (Table 3). These are the level-1 (σ_ε^2) and level- 2 random parts (σ_μ^2) associated with the constant, the level-2 random effects associated with class (σ_Γ^2), and the covariance between the two level-2 terms ($\sigma_{\mu\Gamma}$).

The estimated fixed terms (Table 4) are all significantly different from zero at the conventional cut-off, and as the estimated values are very close to Model B, there is no change in their interpretation. However, the random effects at level 2, include a significant residual variance of 0.4 for the clinic class terms (σ_Γ^2) around the overall class term of 0.84. It appears that the clinics are not treating parents of the same class in the same way. Looking in detail at the twelve random coefficients for class in Table 4 and the graph of these estimates in Figure 3, it is clear that clinic 12 is operating quite differently. The estimated fixed term for class suggests that, over all the clinics, children of non-manual parents can be expected to have 0.84 immunisations more than manual children. The random coefficients for class show a similar positive effect for the first eleven clinics so that the 'total' class effect for clinic 3 can be conceived as an overall class effect of 0.84 and a specific effect of 0.30. Non-manual children

Figure 3 Varying coefficients for class, Model C



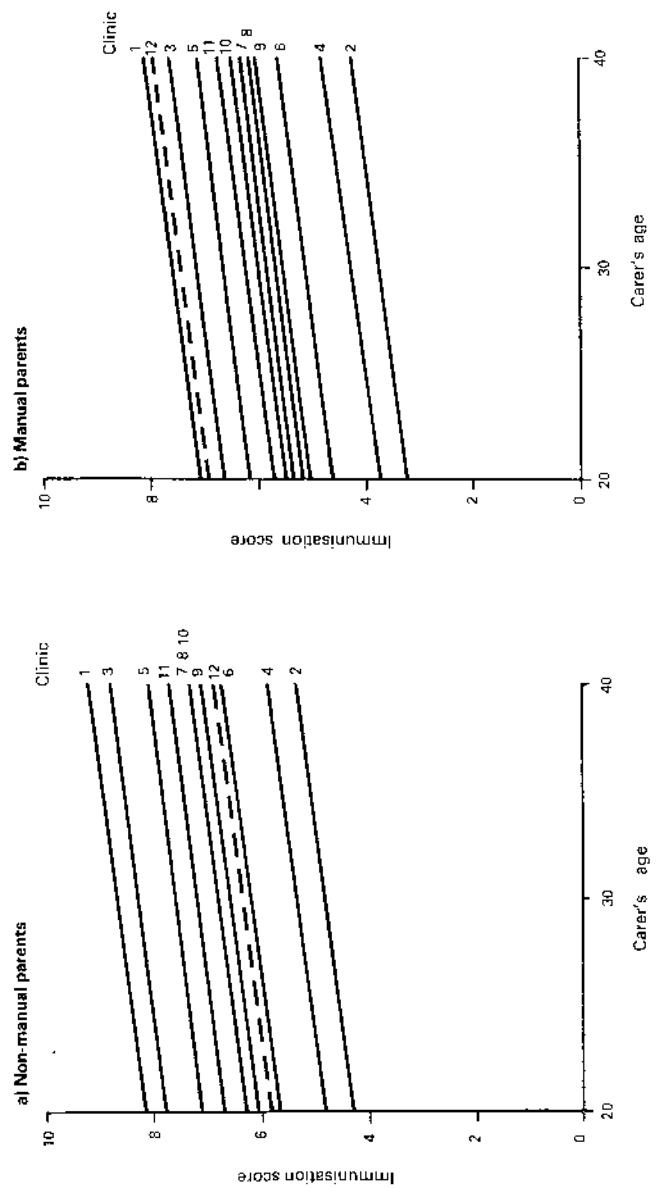
attending clinic 3 have approximately one more immunisation than manual children. For clinic 12, however, the 'total' class effect consists of an overall effect of 0.84 and a specific clinic effect of -1.93, so that it is manual children that have one more immunisation than the non-manual. There is something about this clinic, apart from the age of the parents who are assigned to it, and the hours that it is open for immunisation, that makes it very successful with manual parents in comparison with the other eleven clinics. The clinic-level residuals and their rank are given in Table 2. They should, however, be treated sceptically for with model C, clinics do not just perform in an overall good or poor manner (as reflected in their μ_j values) but differentially according to the characteristics of their clients. The basic findings are well conveyed in Figure 4 where separate graphs are given for each clinic for manual and non-manual children. Clinic 12 has one of the best performances for manual children but is poorish for non-manual. Without a two-level model with quite a complicated residual structure, it would have been impossible to find this very interesting result.

$$\text{MODEL D: } y_{ij}=CON+\beta_1 AGE_{ij}+\beta_2 CLS_{ij}+\alpha HOR_j+(\delta_j AGE_{ij}+\Gamma_j CLS_{ij}+\mu_j+\varepsilon_{ij}) \quad (13)$$

This final model is the most complicated of all for it contains random coefficients for both age and class. As a result, four fixed coefficients need to be estimated (the constant, and terms for age, class and hours) and seven random coefficients (the single level-1 variance term, the three level-2 variance terms and the three level-2 covariance terms for the constant, age and class; Table 3). The estimates of the fixed terms are very similar to models B and C and need no further comment. The random terms for age and all three covariances are not significant but the level-2 residual variances associated with the constant and class are significant. There is a strong suggestion that given the variables specified, model C is sufficiently complicated to capture the underlying order of the immunisation data and the extra terms of model **D** are not needed. For completeness, the clinic residuals associated with the constant are given and ranked in Table 2.

The use of significance tests as a discriminatory tool has therefore led to the choice of model C as an adequate representation of the immunisation data. But this decision is based on the assumption that the relationships are linear, that all the relevant predictor variables have been included, that the residual variance is homogenous, that the residuals have a Gaussian distribution, and that there is not substantial measurement error present. All these assumptions need to be evaluated, particularly as there appears to be substantial residual variation at the clinic level that may possibly be captured by clinic-level variables or (more

Figure 4 Clinic performance based on model C



unlikely?) by individual-level variances that vary systematically over the clinics. For example, the clinics with a poor performance may be relatively inaccessible, or less assiduous and efficient in their appointment and recall system, or there may be a clinic ethos that does not favour immunisation. Diagnostic evaluation of the model is required, and further data collection is necessary. There is a real need to understand why clinic 12 is operating so differently from the rest.

(iii) Complex level 1 random terms

The elaborations of the random part have so far concerned modelling effects at higher levels, but a complex level-1 specification is appropriate if unequal variances (heteroscedasticity) are expected in the the micro-model. For example, it is possible to specify a inicro-model with two level-1 random terms to accommodate heteroscedasticity associated with a dichotomous categorical variable, and thereby perinit more precise estimation of the fixed part. Modifying Model C (equation 12):

$$y_{ij}=CON+\beta_1 AGE_{ij}+\beta_2 CLS_{ij}+\alpha HOR_j+(\Gamma_j CLS_{ij}+\mu_j+\varepsilon_{1ij}CON+\varepsilon_{2ij}CLS_{ij}) \qquad (14)$$

the random part now includes two level-1 random terms, e, and e_2, representing manual and non-manual children. This formulation allows differential variability (greater and lesser) around the fixed terms of the constant and the class dummy. Put simply, holding the other variables constant, non-manual children may not only achieve a higher mean score they may also may be more or less variable than manual children in their scores.

It would appear that the level-1 covariance matrix should consist of two variances and the covariance. Thus, because of the specific coding of a constant and a dummy and following Goldstein (1987, 36) the variance for manual children would be $\sigma_{\varepsilon1}^2$ and non-manual $\sigma_{\varepsilon1}^2+\sigma_{\varepsilon2}^2+2\sigma_{\varepsilon1,\varepsilon2}$. But such a formulation is 'over-parameterised' for, as every child is either of manual class or not, there is only sufficient information to estimate two of these three parameters and the estimates will fail to converge. The obvious solution of constraining the covariance to zero is not a good one for this will force the variance for non-manual children to be at least as great as that of manual children (as variances must be non-negative). Fortunately, the ML software has the unique capacity to solve this problem in two equivalent ways. The first way is to re-formulate the level-1 random part:

$$y_{ij}=CON+\beta_1 AGE_{ij}+\beta_2 CLS_{ij}+\alpha HOR_j+(\Gamma_j CLS_{ij}+\mu_j+\varepsilon_{1ij}MAN_{ij}+\varepsilon_{2ij}CLS_{ij}) \qquad (15)$$

in which, the constant and the dummy have been replaced by two dummies; MAN equals 1 for manual children, and CLS is set to 1 for non-manual

Table 5 ML estimates with complex level 1 structures

Model C	Complex level 1; two dummy method [Equation 15]		Complex level 1; variance and covariance method [Equation 14]	
Fixed terms				
CON 5.69	5.69		5.69	
AGE 0.05	0.05		0.05	
CLS 0.84	0.84		0.84	
HOR-0.05	-0.05		-0.05	
Random terms				
Level 1 : between children				
CON	MAN	CLS	CON	CLS
CON 0.110	MAN 0.108		CON 0.108	
	CLS	0.112	CLS 0.002	
Level 2 : between clinics				
CON CLS	CON	CLS	CON	CLS
CON 1.10	1.10		1.10	
CLS-0.33 0.353	-0.33	0.353	-0.33	0.353

children, both variables being set to zero otherwise. The level-1 random part then consists simply of a variance for each duinmy and no covariance:

LEVEL 2	CON	CLS	LEVEL 1	MAN	CLS
CON	σ_μ^2		MAN	$\sigma_{\varepsilon 1}^2$	
CLS	$\sigma_{\mu r}$	σ_r^2	CLS		$\sigma_{\varepsilon 2}^2$

The estimates for this model are given in Table 5. Also given in the table are estimates derived for equation (14) by the alternative strategy of only fitting the variances associated with the constant and the covariance for level 1:

LEVEL 2	CON	CLS	LEVEL 1	CON	CLS
CON	σ_μ^2		CON	$\sigma_{\varepsilon 1}^2$	
CLS	$\sigma_{\mu r}$	σ_r^2	CLS	$\sigma_{\varepsilon 1,\varepsilon 2}$	

and then deriving the variance associated with non-manual children (CLS) as $\sigma_{\varepsilon 1}^2 + 2\sigma_{\varepsilon 1,\varepsilon 2}$. Thus the non-manual variance is 0.108 +(2*0.0021) which is 0.112 which is exactly the same result derived by using the method based on the the two dummies. Interpreting the results, the variance terms for manual and non-manual children are very similar (at 0.108 and 0.112), and the accommodation of possible heteroscedasticity in the model has not led to even ininor changes in the rest of the estimated coefficients in comparison to Model C. There is no evidence of differential variability between children of different class. This ability to estimate such complex level-1 terms is particularly useful when modelling proportions in a inulti-level log-linear model (Goldstein, 1987, chapter 6).

(iv) Higher and more levels

The models considered so far have been liinited to two levels, but there is no conceptual reason why this cannot be extended to a third level. For example:

$$y_{ijk} = \beta_0 + \beta_1 x_{1ijk} + \beta_2 x_{2ijk} + \alpha_1 w_{jk} + (\phi_k + \mu_{jk} + \varepsilon_{ijk}) \tag{16}$$

where the subscript k refers to a health-authority district. It is the presence of the random term at 3 levels which makes it a 3-level model. Of course, an explicit level-three predictor, such as the district expenditure on community health, could be included, and indeed the model could be extended to higher levels. But it must be remembered that the precision of estimation depends on the number of units at that level. If there were only a few districts in the sample, it is unlikely that a significant effect will be found, and the estimated coefficients will be 'unstable' and unreliable.

The model may also be extended to have two non-hierarchical higher levels. For example, the individual children can be cross-classified by both the clinic to which they are assigned, and the neighbourhood in which they live:

$$y_{i(jk)} = \beta_0 + \beta_1 x_{1i(jk)} + \beta_2 x_{2i(jk)} + \alpha_1 s_k + \alpha_2 w_j + (\phi_k + \mu_j + \varepsilon_{i(jk)}) \tag{17}$$

where k is the neighbourhood, j is the clinic, (jk) is bracketed to indicate that they are at the same level; s is a neighbourhood-level predictor variable such as socio-economic status, and w remains clinic opening hours. Unfortunately, while a computational strategy has been developed for such a cross-classified inodel, (Goldstein, 1987), it has not yet been implemented in any available software.

IV ESTIMATION AND SOFTWARE

(i) OLS and shrinkage estimators

The nature of the ML estimators can be seen by contrasting them with the usual OLS procedures. Returning to Figure 1(d), the different lines could be formulated as twelve separate inodels, one for each clinic:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \varepsilon_{ij} \tag{18}$$

so that the 12 intercepts and 12 slopes need to be estimated by OLS. Indeed, this was a common procedure in educational research (Raudenbush and Bryk, 1986) and the twelve estimated regression slope values could then be specified as the the response variable in a further model with higher-level school predictors to be estimated by OLS:

$$\hat{\beta}_{1j} = \alpha_0 + \alpha_1 w_{1j} + \alpha_2 w_{2j} + \mu_j \tag{19}$$

Such 'slopes as response' models are extremely limited for in the higher level model there can only be one response variable and therefore the micro-level model must be limited to one slope term for one predictor variable. Moreover, not all the variation in the response can be accounted for by school-level predictors as soine should be attributed to sainpling fluctuation.

Another possible way to forinulate Figure 1(c) and (d) in such a way that they can be estimated by OLS is by using sets of dummies. In CATMOG 20, Silk deals with this procedure in detail. To estimate the differential intercept model of Figure 1(c), dummy 1 (D_1) is set to 1 for clinic 1, 0 otherwise, dummy 2 (D_2) is set to 2 for clinic, 0 otherwise, and so on up to and including dummy 11. Clinic 12 is represented by a 1 for the constant and 0 on all the **11** dummies. Thus equation (3) becomes:

$$y_i = \underbrace{\beta_0 x_0 + \beta_2 D_1 + \beta_3 D_2 + \beta_4 D_3 + \ldots \beta_{12} D_{11}}_{\text{differential intercepts}} + \underbrace{\beta_1 x_{1i}}_{\text{fixed slope}} + \varepsilon_i \tag{20}$$

β_0 is the intercept for clinic 12, β_2 is the intercept for clinic 1, β_3 is the intercept for clinic 2 and so on. In a similar fashion, the differential slopes in Figure 1(d) can be accommodated by including a set of dummy/predictor interaction terms where the new variable D_1x_1 is simply formed by multiplying the relevant dummy by the predictor variable. Equation (5) now takes on the horrific dimensions of:

$$y_i = \underbrace{\beta_0 x_0 + \beta_2 D_1 + \beta_3 D_2 + \ldots \beta_{12} D_{11}}_{\text{differential intercepts....}} + \underbrace{\beta_1 x_{1i} + \beta_{13} D_1 x_1 + \ldots \beta_{23} D_{11} x_1}_{\text{differential slopes.....}} + \varepsilon_i \tag{21}$$

There are now 24 terms to be estimated, and quite clearly this procedure is nothing more then the estimation of 2 terms for 12 separate models. Equation (5) is not a complex model in ML terms, and there may be hundreds of clinics not just 12. Nevertheless, this dummy variable approach has been used in the geographical literature with, for example, Johnston *et al* (1988, 179) using a model with over fifty terms to estimate differential regional effects in voting behaviour.

Estimators are like archers; they may be be close to the bulls-eye (precise) or all over the target (imprecise), (Kennedy, 1979, Chapter 2). The separate OLS strategy means that each clinic paraineter is estimated in an 'un-connected' way; the OLS estimates do not 'pool' the information for all clinics and this lack of efficiency in exploiting the data results in relative imprecision. The OLS estimates will have high imprecision when the clinic sample size is small, and when the range of variability in the predictors is limited. Thus, we would have little confidence in the estimated slope for age if the range in a particular clinic is only from 21 to 22 years of age and there are only three parents attending that clinic. Indeed, it is not possible to estimate separate models for each clinic if the number of coefficients (including the intercept) exceeds the size of the sample attending that particular clinic.

By contrast, ML models represent the clinics as a sample of all clinics, and treat any particular parameter, for example, β_{0j} of equation (2), as having a **distribution** with, in this case, a mean of β_0 and a variance of σ_μ^2. In brief, ML models can use information from all clinics simultaneously in estimating each clinic's effect. Thus, when calibrating the fully-random model of equation (5), instead of estimating the 24 separate parameters of equation (21), only five need be estimated, namely the two fixed terms, and the variances of the three random terms. Moreover, this number is unrelated to the number of clinics, and will not increase if there are more than 12 clinics. In fact, the ML parameters are transformed or 'shrunken' versions of the separate OLS parameters (Jones, 1990b). For example, in ML models the fixed or 'average' slope turns out to be the 'weighted' average of all the OLS slopes, weighted in such a way that the unreliable ones (with the largest sampling error) make the smallest contribution to the overall effect. Similarly, the multi-level individual clinic slopes, Γ_j, are also weighted combinations of the OLS slopes for each clinic and the overall weighted average of the fixed slope. In a multi-level model the weight given to a particular OLS slope depends on its sampling variability. If it is not very reliable, the Γ_j will be based primarily on, or 'shrunk' towards the weighted mean slope β_1. As Raudenbush and Bryk (1986, 6) conclude 'the estimates borrow their strength from whatever relationships happen to exist in the data' so that clinic relationships which may be poorly estimated (or indeed, inestimable) on their own, benefit from the data for other clinics. In contrast, a reliable coefficient is substantially immune to the influence of other clinics.

In practice, a nuinber of studies have shown that shrinkage estimates represent a considerable improvement over OLS estimates (see Morris, 1983, table 2). For example, Rubin (1989) reports a study that tried to predict the performance in the first year of eighty-two law schools from a pre-entry test. The usual OLS estimates were used to derive the relationship for each school, but it was found that these were unstable over time and some un-interpretable, negative values were obtained. This problem of 'bouncing beta's' is an outcome of the restricted range of pre-entry scores in any one school; the lack of information in the predictor variable leads to imprecise estimation. In contrast, the shrinkage estimates were both stable over time and sensible. In a cross-validation, the model was calibrated on the 1974 entry and used to predict the performance of the 1975 entry. The shrinkage estimates were on average a staggering forty times better than the OLS ones when measured by the correlation between the predicted and observed performance of the 1975 cohort.

(ii) IGLS estimates and assumptions

Until the 1980's ML modelling was severely limited by the lack of general estimation procedures and suitable software. The problems were twofold: the difficulty of estimating the fixed and random parameters simultaneously, and doing so in a computationally feasible form. Early programs, such as **3V** in **BMDP** and **VARCOMP** in SAS, were extremely limited in the size of the problem they could handle by the need to invert large matrices. Recently, three alternative methods have been developed: the EM algorithm (Dempster *et al* 1981; Mason *et al* 1984, Raudenbush and Bryk, 1986); a Fisher-scoring algorithm (Aitkin and Longford, 1986; Longford, 1987; De Leeuw and Kreft, 1987), and a Iterative Generalised Least-Squares algorithm (Goldstein, 1986a). In theory, all three procedures should give equivalent results, and the choice of technique must to some extent rest on availability and quality of the software. Here, one procedure, that of IGLS will be described. It is obvious from Goldstein and McDonald (1988) that this is a highly flexible estimation strategy, and it is claimed that the software associated with it, ML3, is unique in its capability for fitting complex models.

In broad terms, IGLS estimation proceeds as follows. Ignoring the multi-level nature of the random terms, the OLS estimates of the fixed parameters are derived in the usual way. The squared residuals from this OLS model are then regressed on the set of variables defining the structure of the random part to provide initial estimates of the variance-covariance matrix.These initial estimates are then used to derive generalised least-squares estimates of the fixed coefficients that take the random-part variances into account. The residuals from this estimation are used to update the estimates of the random part, which in turn are used to re-estimate fixed coefficients, and so on until convergence. Crucially, Goldstein's procedure can be expressed in a computationally convenient form so that a very large, and rather empty, covariance matrix is derived in a compact form. A difficult estimation problem is decomposed into a sequence of linear regressions which can be solved efficiently and effectively.

Goldstein (1986a) provides a proof that the IGLS estimates are consistent, and that under certain assumptions they are maximum likelihood estimates. These terms refer to desirable properties of estimators in repeated samples. Maximum likelihood estimates (Edwards, 1972) are unbiased, so that the sampling distribution of the estimates will be centred over the true parameter values; consistent, so that the variability of the estimates around the true value reduces to zero as the sample reaches infinity; and efficient, so that as the sample size increases, the variability of the estimator decreases faster than any other consistent estimator.

These properties depend on two key assumptions. First, given the predictor variables, the response variables have a Gaussian distribution; second, the within-clinic parameters also have a Gaussian distribution. An assessment of the first assumption can be made by examining plots of the residuals for each clinic, but the second assumption is more difficult to assess in that the random coefficient distribution is not directly observable. Raudenbush and Bryk (1986,14) while acknowledging that there has been little empirical work on the consequences of violating these assumptions, suggest that problems are most likely to occur with estimates of the sampling variances and in hypothesis testing. They argue that estimates of the coefficients are more robust to violation of distributional assumptions. Their advice is to treat 'marginal' significance with care.

(iii) Residuals

With the single-level model there is just one set of residuals. These are readily calculated as the difference between the actual and predicted values of the response, and are mainly used for diagnosing model ills (Cox and Jones, 1981). For ML models the situation is inore coinplicated in that there may be several sets of residuals at different levels, which can be used to compare size of effects (which clinic is best?) as well as for diagnosis. With the IGLS procedure the residuals have to be estimated after the estimates have converged. The nature of this process can be illustrated for the random intercepts model of equation (3) in relation to what are known as the composite residuals, $\hat{q}_{ij}$ which are the differences between the actual immunisation score and the sum of the fixed effects:

$$\hat{q}_{ij} = y_{ij} - (\hat{\beta}_0 X_0 + \hat{\beta}_1 x_{1ij}) \quad (22)$$

In this case these residuals are the composite of the level-1 residuals, $\hat{\varepsilon}_{ij}$, and the level-2 residuals, $\hat{\mu}_j$ which have to be estimated separately, the hat (^)signifying estimates. Goldstein (1987, 20) demonstrates that these are transformations of the composite residuals which are 'shrunk' to have smaller variances. The level-1 residuals are then derived as:

$$\hat{\varepsilon}_{ij} = \hat{q}_{ij}(\hat{\sigma}_\varepsilon^2/(\hat{\sigma}_\varepsilon^2 + \hat{\sigma}_\mu^2)) \quad (23)$$

and the level-2 residuals as:

$$\hat{\mu}_j = \hat{q}_j((n_j\hat{\sigma}_\mu^2)/(n_j\hat{\sigma}_\mu^2+\hat{\sigma}_\varepsilon^2) \quad (24)$$

where $\hat{q}_j$ is the composite residual mean for clinic j:

$$\hat{q}_j = (\Sigma_i \hat{q}_{ij})/n_i \quad (25)$$

As the number of children attending a clinic decreases, there is less information about the clinic, and there is greater shrinkage towards the overall mean. Because they are based on expected values these estimates are also referred to as 'conditional' or 'posterior' means.

In the usual way' these residuals can be standardised by dividing by the estimated standard errors, but an added complication is that there are two types of standard error. The total variance of the estimated residuals can be decomposed into two parts: the population variance of the true values about the

fixed part, and the sampling variance of the estimates about their true value (Raudenbush and Bryk, 1986). It is the estimate of the total variance which is used when the residuals are being standardised for diagnosis (such as plotting I', against p to identify unusual level 2 units). When the residuals are being standardised for comparison, the standard error is estimated on the basis of the sampling variance only (Goldstein, 1987, Appendix 3.2). Goldstein (1987) encourages exploratory plotting of residuals and gives a nuinber of examples in his book. Davies *et al* (1988) urge caution due to the lack of experience in interpreting such diagnostic plots.

(iv) ML software

Goldstein and his co-workers have developed general-purpose software to implement these procedures and, as the naine implies, the **ML3** package (Rasbash *et al* 1989) is currently capable of estimating models with up-to three levels. At its core there is a general interactive statistical package which has more than a passing resemblance to **MINITAB** (Ryan *et al* 1986). This can be used to create, transform and manipulate data prior to modelling and to plot the resultant estimates. The ML model is specified as a list of variables in the fixed part, and as an 'indicator' covariance matrix for each level of the random part. Estimation by IGLS coinmonly takes less than five iterations in comparison with 80 iterations of the EM algorithm. Output includes estimates of the parameters and their standard errors, the ratio of the two being judged against a standard Gaussian distribution. More coinplex testing of sets of hypotheses about the random and fixed parts is achieved by using contrast matrices. **ML3** will output the differences between the stated values of the null hypothesis and the estimated values (the contrasts), test statistics for all the contrasts simultaneously and individually (which are evaluated against a chi- square distribution) and individual and simultaneous 95% confidence limits. It must be stressed, however, that all such confirmatory testing is based on large-sample theory and assumptions discussed above; if these conditions are not met, the tests are likely to over-state the significance of the results. **ML3** outputs the predicted values, which are just the sum of the fixed effects, and permits the calculation of residuals at each level. It is possible to standardise these residuals for both comparison and diagnosis. The program has graphics commands for exploratory plotting.

Although **ML3** is highly flexible and powerful, it can currently only handle three levels. Longford's (1986) **VARCL** package consists of one program for three levels of nesting, and another for up to nine levels of nesting, the latter having a more restricted covariance structure in only allowing random intercepts.

Both programs are capable of handling very large problems as they do not store the original data. A particularly useful feature of this package is a set of options for fitting models with non-Gaussian residual structures, such as a binary response; for example, iinmunised or not (Jones *et al* 1990). Such models can be fitted in **ML3** but special code has to be written (Goldstein, 1990). A detailed coinparison of available software **(GENMOD, HLM, ML2, VARCL)** is given by Kreft *et al* (1990); Appendix 2 details how to obtain this software.

V APPLICATIONS

(i) The performance of London schools

One of the most impressive applications of ML modelling is Nuttall *et al's* (forthcoming) analysis of the performance of children in London secondary schools. The response variable is an ILEA score which is an aggregation for each pupil of the number and level of passes in public examinations. Thus an 'A' at **0** level scores 7, a B scores 6 and so on; a very high score of 70 is equivalent to 10 A's, a zero signifies no passes whatsoever. Table 6 presents the estimates for a single ML model for this very 'rich' data set, there being 140 schools (level 3), three year cohorts (2) and 31,623 pupils(1) and a number of background variables. The key to interpretation is to realise that the intercept term represents the ILEA score for 'Mister Average-low-ability'. The model predicts a score of 17.8 in 1985 for a inale pupil who is an ESWI (that is English/ Scottish/ Welsh/ Irish), who was assessed at age 11 to be in the lowest verbal reasoning band (VR3, lowest 25 percent across all schools), and who has attended a inixed, non-denominational school with the ILEA average proportion of those entitled to free-school meals. The fixed effects are then, if positive, additional points on the ILEA score. Thus girls are on average 2.5 points higher, while those 25 percent of pupils in the highest verbal reasoning band, VR1, score 19 more points. At the school level, both single sex and denominational schools do better than mixed, county schools, while lower scores are predicted if the pupil attends a school where a high proportion receive free school meals (fsm). As the proportion of fsm increases from 10 to 30 percent, the pupil score is predicted to drop by 6 points. Perhaps the most striking aspect of the fixed effects, is that ethnic groups do substantially better than the ESWI's with the exception of the Caribbeans who do worse (but not significantly so). The significant effect for year indicates that the results have improved by 1.4 points per year on average.

Table 6 ML estimates for London schools

RESPONSE

ILEA score	7	6	5	4	3	2	1	0
GCE '0' level	A	B	C	D	E	–	–	U
16+	A	B	C	2	3	4	5	U
CSE	-	-	1	2	3	4	5	U

FIXED EFFECTS
(ratio of estimate to standard error in brackets)

Level 1: Pupil		
Intercept	17.8	
Girl	2.5	(12.5)
VR1	19.0	(63.3)
VR2	8.2	(41.0)
African	4.0	(8.0)
Arab	4.4	(4.0)
Bangladeshi	4.7	(6.7)
Caribbean	-0.4	(1.9)
Greek	4.6	(6.6)
Indian	7.3	(14.6)
Pakistani	6.0	(10.0)
SE Asian	8.3	(13.8)
Turkish	3.7	(9.3)
Other	3.8	(9.5)
Level 2: Year		
year	1.4	(7.0)
Level 3: School		
Boys	0.8	(2.7)
Girls	1.4	(4.7)
Church of England	1.2	(3.0)
Roman Catholic	2.4	(8.0)
Prop. free school meals	-0.41	(10.3)
Prop. squared of fsm	.003	(7.5)

RANDOM EFFECTS
(variances and covariances, correlations in brackets)

Level 1: Between pupils 96.7
Level 2: Between years 1.2
Level 3: Between schools

	Intercept	**VR1**	**VR3**	**Sex**	**Carib**	**Year**
Intercept	2.9					
VR1	-1.9 (0.0)	17.4				
VR2	-0.1 (0.0)	6.1 (0.9)	2.8			
Sex	-1.5 (-0.6)	2.4 (0.4)	0.6 (0.2)	2.1		
Carib	-0.4 (-0.2)	-1.8 (-0.4)	-0.5 (-0.3)	-0.3	1.1	
Year	(0.1)					0.5

Source: Nuttall at a/ (forthcoming)

The bulk of the random variation (96.7) is at the pupil level but there is a noticeable variance of 17.4 for the differential school effect for VR1. On average, the pupil with the high prior attaininent of VR1 band gets 19 more points than the VR3 pupil. But there is substantial variation around this fixed effect from school to school with some only achieving 11 points while others manage as much as 26. The size of this variation being interpolated form the overall effect plus or minus twice the standard deviation of 4.2. On the basis of relatively large variances in comparison to fixed effects, there are schools where boys do better than girls and the Caribbeans do better than the ESWI's, and where there has been no improvement in attainment over the three years. The relatively small variance for the intercept (at 2.9) suggests that the variation in performance of VR3 pupils between schools is small. The generally low correlations between the school-level random terms (with the exception of 0.9 between VR1 and VR2) suggests that schools that do well for one type of pupil do not necessarily do well for another. In summary, these results strongly imply that no school is best overall, pupils vary considerably and the school effect appears to depend on the characteristics of the pupil.

(ii) Other applications

Most applications to date of ML modelling are in educational research; a number of these are reviewed in Jones (1990a), while a detailed recent overview, which considers soine eighteen studies is provided by Raudenbush (1988). Particularly worth reading is Aitkin and Longford (1986) which represents an excellent empirical application of the random slopes model; it is accompanied by an account of the lively discussion following their paper.

Non-educational applications include the work of the Population Studies Center, University of Michigan, which as part of the World Fertility Survey has attempted to assess macro paraineters at the country-level and micro parameters at the individual-level (Entwistle *et al*, 1982). It is interesting to note that Lloyd (in Cleland and Scott, 1987, 611) states that 'it remains questionable whether underlying macro-level relationships have a quantifiable structure'. The Michigan group have also modelled the variations in contraceptive use in Thailand using access to clinic as a macro-variable (Entwistle *et al* 1984). Other examples include Fieldsend *et al* (1987) who examine how financial ratios vary by industrial sector, Longford (1985) on house-price variations, and Davies *et al* (1988) who examine variations in earnings in the engineering industry.

VI CONCLUSIONS

ML modelling offers three broad advantages over single-level modelling. In *methodological terms* it is extremely useful to handle data at several scales simultaneously; not doing so may frequently result in a mis-specified model and invalid inferences. In *technical terms* the multi-level estiinates offer improved precision over the OLS equivalents and this may lead to different decisions over whether a particular result is an outcome of sampling error or not. In *substantive terms* it is a major advance to be able to postulate models which can incorporate 'contextuality'. Multi-level models with varying parameters allow different relationships in different places and, because the higher level variables do not have to be aggregates of lower-level variables, an element of structure or situation can be incorporated, however mechanically, into the model. The action of groups is not completely atomised to the behaviour of individuals.

Another benefit of these models is that they represent a narrow bridge between two major alternative research designs: extensive and intensive (Sayer, 1984). ML modelling is firmly based within the extensive framework, using quantitative techniques to seek patterns and generalisations. But the this may expose specific 'groups' that need to be studied by intensive, qualitative research to explain their distinctiveness. This opportunity afforded by multi-level modelling is well captured by Plewis (1988) writing on ethnicity and educational attainment.

> *'There is a danger that too much emphasis is put on the characteristics of individuals in multiple regression models and not enough on the social and political context in which those individuals find themselves. 1 do not mean to imply that individual characteristics are unimportant when seeking to explain variation in progress, or that linear models can only be specified at the individual level and hence are, in an ideological sense, necessarily biased. Indeed, the development of multi-level models seems to me to be of great importance, both statistically and because it brings schools and classrooms explicitly into view as social institutions. It thus brings statistics and social science a little closer and might, for example, enable us to use statistical arguments to choose particularly unusual or successful schools. These schools could then be studied more intensively as case studies using ethnographic methods.'*

The models used in the immunisation case-study allow each clinic to be a 'social institution' employing qualitatively different policies. The anomaly revealed by this approach, clinic 12, clearly needs to be studied intensively to discover why the clinic appears to be operating so differently.

At the same time, we must not be over-optimistic and expect too much from these methods. Technical problems still remain. The maximum-likelihood estimates represent highly desirable properties for very large samples, but the small-sample characteristics are not really known. Nor do we know a great deal about the robustness of the estimates to non-fulfilment of assumptions. Little research has been conducted on the requirements of sample sizes that are needed at each level for effective estimation, although (following Goldstein, 1984) the clustered multi-stage sampling design is the most appropriate means of collecting information. A major methodological problem facing the geographical application of these models, is recognising and defining the appropriate hierarchies. Multi-level models assume that the groups at the higher levels are soinehow 'naturally' given. While in many cases this will be the case (eg schools, clinics), in others it will not for the groupings may reflect the rather arbitrary criteria of the data collection agency (eg census enumeration districts). While it may be possible to use Openshaw's (1977) automatic zoning procedure in tandein with inulti-level models, this is not a problem that can be overcome by a technical fix. Any research using these methods will have to define, on the basis of theory, what are the appropriate levels and units in relation to the particular entities being studied.

In terms of the conceptual usefulness of the models, it must be agreed with Hauser (1970) that the models remain too mechanical and are only distantly related to actual processes. In no way do they manage to *'recapture the flow of human agency as a series of situated events in space and time'* as a contextual theory should (Thrift, 1983). On balance, it is possible to agree with Eulau in his foreward to Boyd and Iversen (1979,vi):

> *'the analysis of social contexts through multilevel treatment is an enormous step forward in the social sciences, yet it remains essentially aggregative in the sense that it discovers mathematical rather than real social structures. That the individual or groups located in social contexts actually interact with and behave in structured ways towards each other is an essential assumption of contextual analysis, but it is an assumption that cannot be proved as such by contextual analysis.'*

Nothing much has been written in this text about the usefulness of multi-level modelling in physical geography. It may be possible to come to grips in at least a partial sense with both the immanent (the underlying general processes) and the configurational (the specifics which are local, and unique in space and time) (Simpson, 1963), using these contextualising procedures. However, it must be recognised that ready-made, sharply demarcated hierarchies are not the norm in physical geography (Kennedy, 1979, 551). Therefore, perhaps the greatest usefulness of the technique will come when the hierarchical structure has been imposed by the researcher during data collection. For example, in a time-dependent, longitudinal study, the lower level could be measurement occasion (Bryk and Raudenbush, 1987; Ware, 1985), while in multi-stage sampling designs, the higher level would be cluster membership. In both these examples, multi-level modelling allows the particularities of the research design to be incorporated and allowed for in the modelling process. Finally, it must be remembered that multi-level models are recent developments and novel uses are still being developed. For example, Raudenbush and Bryk (1985), have used the procedures in meta-analysis, that is quantitative research synthesis in which the estimates of effects in previous research are the response variable, and the predictors are study designs and contexts of the research.

VII FURTHER READING

It is important when reading in this area to remember that effective estimation procedures have only readily been available since the mid-1980's. Literature written before this time may represent good conceptual accounts of the problem, but the proposed ways of dealing with it have now been effectively superseded. Good overviews that consider concepts, estimation and applications are Mason *et al* (1984), Raudenbush and Bryk (1986), Raudenbush (1988) and the short book, by Goldstein (1987). Blalock (1984) and Hauser (1970) consider the problems of specifying a realistic and meaningful model. A good introduction to the benefits of multi-level models is given in articles by Burstein (1980), Cooley *et al* (1981), and the book by Boyd and Iversen (1979). In terms of making comparisons between schools or between areas, the highly readable account of Goldstein (1984) which contains no inathematics, is recommended. The important articles on estimation are: Lindley and Smith (1972), Harville (1977), Dempster *et al* (1981), Laird (1982), Kackar and Harville (1984), Goldstein (1986), and Longford (1987). The problems of analysing sample surveys with a multistage design are considered in Goldstein and Silver (1989). The use of inulti-level modelling in the analysis of longitudinal designs is discussed in Goldstein (1986b, 1987, Chapter 4), Bryk and Raudenbush (1987), Laird and Ware (1982) and Ware (1985).

APPENDIX 1 ML3 COMMANDS

The following is a guide to the actual commands that have been used to perform the analyses of the Models A to D. The data have to be read in; the model has to be specified in terms of levels, predictor and response variables, fixed and random parts; the model has to be calibrated and choices made concerning what output is required. The left-hand side of the page gives the commands, the right-hand side provides comments.

SESSION 1

	ACTIVATE THE SESSION
ML3	
LOGO	Call the program
IMM.LIS	Keep a listing of all that appears on the screen in IMM.LIS
	INPUT THE DATA
FDIN C1-C7	Read in file, 1 line at a time into 7 Columns from file IMM.DAT;
IMM.DAT	
(10,10,10,10,10,10,10)	fixed field of 10 for each variable

NAME C1 'ID1' C2 'ID2'	Naming the variables for easier reference
NAME C3 'IMM'	
NAME C4 'AGE' C5 'CLS'	
NAME C6 'HOR' C7 'CON'	
NAME C1-C6	On its own gives number of observations, max. and min.
	Use utilities to display, tabulate or summarise data
	BEGIN MULTI-LEVEL MODELLING
SETT	A multi-level screen will appear
IDEN 2 'ID2'	Specify codes for levels 2 and 1
IDEN 1 'ID1'	
SUMM	gives a summary of level 1 units for each level 2 group
RESP 'IMM'	Declare the response variable
EXPL 'CON' 'AGE"CLS"HOR'	List candidate predictor variables
SETT	Use SETT at any time to see what has been specified on the multi-level screen
	MODEL SPECIFICATION

Model A	Model B	Model C
Constant only	plus three predictors	plus random co-efficient for class
	Fixed part	
FPAR	FPAR	FPAR
FPAR 'CON'	FPAR 'CON' 'AGE' 'CLS' 'HOR'	FPAR 'CON' 'AGE' 'CLS' 'HOR'
	FMEAN 'AGE' 'HOR'	FMEAN 'AGE' 'HOR'
	Random part at each level	
SETV 1 'CON'	SETV 1 'CON'	SETV 1 'CON'
SETV 2 'CON'	SETV 2 'CON'	SETV 2 'CON' 'CLS'

SETT	Shows requested model specification and indicator matrices for co-variances
	MODEL ESTIMATION
	Default Settings
TOLER 2	Tolerance for convergence

MAXI 5	Number of iterations
BATCH OFF	Stops after each cycle of convergence
METHOD IGLS	Iterative generalised least squares estimation
	Running the models
START	Begin estimation with iteration 1
FIXE	Print fixed estimates
RAND	Print random estimates
NEXT	Begin next iteration
FIXE	Check for convergence
RAND	Check for convergence
	Repeat last three commands until convergence
	RESIDUALS
	Model A and B: same residual structure
RESI 1 1 'CON' C50	One level-1 residual term associated with the constant put in C50
RESI 1 2 'CON' C52	One level-2 residual term associated with the constant put in C52
PRED C80	Predicted values of response variable
NAME C80 'PRED'	
PRINT C50 C80	Print residuals and predicted values for inspection and interpretation
PRINT C52	
	Model C: residual structure includes random coefficients for class
RESI 1 1 'CON' C50	One level-1 residual in c50
RESI 2 2 'CON"CLS' C52 C92	Two level-2 residuals; one associated with constant in C52; other for random coefficients for class in C92
PRED C60	
PRINT C50 C60	Print residuals and predicted values
PRINT C52 C92	
SAVE	Save data, names and structures in
IMM.SAV	File imm.sav
STOP	End first session

SESSION 2

LOGO	ACTIVATE SESSION
IMM.LIS	Keep a log of all activity
RETRIEVE	Retrieve data and previous
IMM.SAV	settings
SETT	Show current settings
CLRV 2	Clear level-2 random part
SETV 2 'CON"AGE"CLS'	Specify three random terms and their covariates
SETT	Check new settings
START	Begin estimation
NEXT	
NEXT	Repeat until convergence
FIXE	Print estimates
RAND	
RESI 1 1 'CON' c50	Estimate level-1 residuals
RESI 3 2 'CON"AGE"CLS' C51, C52, C53	Estimate level-2 residuals
PRED C90	Predicted values
PRINT C3,C90,C50	Print out original values,
PRINT C51,C52 C53	fitted values and residuals
STOP	End of session

APPENDIX 2 OBTAINING SOFTWARE

Unfortunately, no multi-level programs are available as part of existing statistical systems, but there are a number of free-standing programs undergoing active development.

(i) GENMOD was one of the earliest programs to be developed; it uses the EM algorithm which has relatively slow convergence properties. Details are available from:
Professor William Mason,
Dept. of Sociology, University of Michigan, USA

(ii) HLM2 and HLM3 have been developed by Raudenbush and Bryk (1986) for 2- and 3-level analyses using the EM algorithm. Residuals have to be graphed and analysed with external software. It is available for mini- and mainframes and has an interactive user interface. It is available from:
Scientific Software, 1369 Neitzel Road,
Mooresville Indiana, 46158-9312, USA.

(iii) VARCL3, VARCL9. These are Longford's (1986) programs for 3 and 9 levels respectively. The entire user interface is organised into a 'structured conversation' with yes/no questions. It is only the 3-level version that permits random slopes as well as random intercepts. It is available from
Dr N T Longford, Educational Testing Service,
21-T Rosedale Road, Princeton, NJ08540
BITNET: NLT6600@rosedale
or in the UK
Centre for Applied Statistics
University of Lancaster
E-MAIL: Statistics@uk.ac.lancaster

(iv) ML3 has been used in the present study and it uses Goldstein's (1986a) generalised-least squares estimator. Enquires should be addressed to:
Bob Prosser, Multilevel Models Project
Mathematics, Statistics and Coinputing Dept.,
Institute of Education, University of London,
20 Bedford Way, London. WC1H OAL.
E-MAIL: rprosser%UK.ac.lon.educ.isis

This project also produces a **Multilevel Modelling Newsletter** which has details of conferences, articles, research, program developments, and book reviews.

BIBLIOGRAPHY

Aitken, M, Anderson, D and Hinde, J (1981) 'Statistical modelling of data on teaching styles', **Journal of the Royal Statistical Society, Series A,** 144, 419-61.

Aitkin, M and Longford, N (1986) 'Statistical modelling in school effectiveness studies (with discussion)' **Journal of the Royal Statistical Society, Series A,** 149, 1-43.

Alker, H S (1969) A typology of ecological fallacies in Dogan, M and Rokkan, S (eds) **Quantitative Ecological Analysis,** Mass, MIT Press.

Bennett, S N (1976) **Teaching styles and pupil progress,** London, Open Books.

Blalock, H M (1984) 'Contextual-effects models: theoretical and methodological issues', **Annual Review of Sociology,** 10, 353-372.

Bondi, E and Bradford, M G (1990) 'Applications of multi-level modelling to geography' **Area** (to appear)

Boyd, L H Jr. and Iversen G R (1979) **Contextual Analysis: Concepts and Statistical Techniques,** Belmont, Calif., Wadsworth.

Bryk, A S and Raudenbush, S W (1987) 'Application of hierarchical linear models to assessing change' **Psychological Bulletin,** 101, 147-158.

Burstein, L (1980) 'The analysis of multilevel data in educational research and evaluation' in Berliner, D C (ed.), **Review of Research in Education 8,** Washington, American Educational Research Association.

Chorley, R J, Stoddart, D R, Haggett, P and Slaymaker, H O (1966) 'Regional and local components in the areal distribution of surface sand facies in the Breckland, E England' **Journal of Sedimentary Petrology,** 36, 209-20.

Cleland, J and Scott, C (1987) **The World Fertility Survey: An Assessment,** Oxford, Oxford University Press.

Cooley, W M, Bond, L and Mao, B J (1981) 'Analysing multilevel data' in Berk, R **A (ed.), Educational Evaluation Methodology: The State of the Art,** Baltimore, John Hopkins Press.

Cox, N J and Jones, K (1981) 'Exploratory data analysis' in Wrigley and Bennett, R J (eds.), **Quantitative Geography,** London, Routledge.

Davies, R **B,** Martin, A M and Penn, R (1988) 'Linear modelling with clustered observations: an illustrative example of earnings in the engineering industry' **Environment and Planning Series A,** 20,1069-84.

De Leeuw, J and Kreft, I. (1986) 'Random coefficient models for multilevel analysis' **Journal of Educational Statistics,** 11,57-85.

Dempster, A P, Rubin, D B and Tsutakawa, R K (1981) 'Estimation in covariance components models' **Journal of the American Statistical Association,** 76, 341-353.

Edwards, A W T (1972) **Likelihood,** Cambridge, Cambridge University Press.

Entwistle, B et al (1984) 'A inultilevel model of family planning availability and contraceptive use in rural Thailand' **Demography,** 21, 559-574.

Entwistle, B, Hermann, A I and Mason, W M (1982) **Socioeconomic determinants of fertility behaviour in developing countries,** National Academy Press, Washington DC.

Fieldsend, S, Longford, N T and McLeary, S (1987) 'Industry effects and the proportionality assumption in ratio analysis: a variance component analysis' **Journal of Business Finance and Accounting,** 14, 497-517.

Goldstein, H (1984) 'The methodology of school comparisons' **Oxford Review of Education,** 10, 69-74.

Goldstein, H (1986a) 'Multilevel mixed linear model analysis using iterative generalised least squares' **Biometrika,** 73, 43-55.

Goldstein, H (1986b) 'Efficient statistical modelling of longitudinal data' **Annals of Human Biology,** 13, 129-42.

Goldstein, H (1987) **Multilevel Models in Educational and Social Research,** London, Charles Griffin.

Goldstein, H (1990) 'Fitting loglinear multi-level models' **Multilevel Modelling Newsletter,** 2,3.

Goldstein, H and McDonald, R P (1988) 'A general model for the analysis of multilevel data' **Psychometrika,** 53,455-67.

Goldstein, H and Silver, R (1989) 'Multilevel and multivariate models in survey analysis' in Skinner, C, Holt, D and Smith, F (eds.), **The Analysis of Complex Surveys,** New York, Wiley.

Hanushek, E A, Jackson, J E and Kain, J F (1974) 'Model specification, use of aggregate data, and the ecological fallacy' **Political Methodology,** 1, 89-107.

Harville, D A (1977) 'Maximum likelihood approaches to variance component estimation and to related problems' **Journal of the American Statistical Association,** 72, 320-340.

Hauer, J (1984) 'Towards multilevel analysis: general and theoretical considerations' in Bahrenberg, G, Fischer, M M, Nijkamp, P (eds.) **Recent Developments in Spatial Data Analysis,** Aldershot, Gower.

Hauser, R M (1970) 'Context and consex: a cautionary tale' **American Journal of Sociology,** 75, 645-664.

Healy, M R (1989) 'Book review' **Multilevel Modelling Newsletter** 1(3),3.

Johnston, R J, Pattie, C J and Allsopp, J G (1988) **A Nation Dividing?,** Longman, London.

Jones, K (1990a) 'Variations in school examination performance: the application of multi-level modelling' Dept. of Geography, Portsmouth Polytechnic, **mimeo.**

Jones, K (1990b) 'Specifying and estimating multi-level models for geographical research', **Transactions of the Institute of British Geographers,** to appear.

Jones, K and Moon, G (1990) 'A multi-level approach to immunisation uptake' **Area,** to appear.

Jones, K Moon, G and Clegg, A (1990) 'Ecological and individual effects in childhood immunisation uptake; a generalised linear multi-level approach', paper presented to 4th International Symposium in Medical Geography, Norwich; submitted to **Social Science and Medicine.**

Kackar, R N and Harville, D A (1984) 'Approximations for standard errors of estimators of fixed and random effects in mixed linear models' **Journal of the American Statistical Association,** 79, 853-862.

Kennedy, B A (1979) 'A naughty world' **Transactions of the Institute of British Geographers,** 4, 550-558.

Kennedy, P (1979) **A guide to econometrics,** Oxford, Martin Robertson.

King, L **J** (1976) 'Alternatives to a positive economic geography' **Annals of the Association of American Geographers,** 66, 293-308.

Kreft, I, De Leeuw, J and Kim, K-S (1990) **Comparing four different packages for hierarchical linear regression,** Los Angeles, UCLA Statistics Series No. 50.

Laird, N M (1982) 'Coinputation of variance components using the EM algorithm' **Journal of Statistical Computing and Simulation,** 14, 295-303.

Laird, N and Ware, **J** (1982) 'Random effects models for longitudinal research' **Biometrics,** 38, 963-974.

Lindley, D V and Smith, A F M (1972) 'Bayes estimates for the linear model' **Journal of the Royal Statistical Society,** B, 34, 1-41.

Longford, N T (1985) 'Mixed linear models and an application to school effectiveness' **Computing and Statistical Quarterly,** 2, 109-118.

Longford, **N T** (1986) 'VARCL: interactive software for variance components analysis' **Professional Statistician,** 5, 28-32.

Longford, N T (1987) 'A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects' **Biometrika,** 74, 817-827.

Mason, W M, Wong, G Y and Entwistle, B (1984)' The multilevel model: a better way to do contextual analysis', **Sociological Methodology,** Jossey Press, London.

Moellering, H and Tobler, W R (1972) 'Geographical variances' **Geographical Analysis,** 4, 34-50.

Morris, C (1983) 'Parametric empirical Bayes' **Journal of the American Statistical Association,** 78, 47-65.

Nuttall, D L, Goldstein, H, Prosser, R and Rasbash, J (forthcoming) 'Differential school effectiveness' **International Journal of Education Research,** to appear.

Openshaw, S (1977) 'A geographical solution to scale and aggregation problems in region-building,partitioning and spatial modelling' **Transactions of the Institute of British Geographers,** NS2, 459-72.

Pearson, K (1896) 'Mathematical contributions to the theory of evolution' **Philosophical Transactions of the Royal Society, Series A,** 187, 253-318.

Plewis, I (1988) 'Assessing and understanding the education progress of children from different ethnic groups' **Journal of the Royal Statistical Society, Series A,** 151, 315-326.

Rasbash, J, Prosser, R and Goldstein, H (1989) **ML2: Software for Two-level Analysis, User's Guide(with ML3 supplement),** University of London, Institute of Education.

Raudenbush, S W (1988) 'Educational applications of hierarchical linear models: a review', **Journal of Educational Statistics,** 13, 85-116.

Raudenbush, S W and Bryk, A S (1985) 'Empirical Bayes meta-analysis' **Journal of Educational Statistics, 10,** 75-98.

Raudenbush, S W and Bryk, A S (1986) 'A hierarchical model for studying school effects', **Sociology of Education,** 59, 1-17.

Roberts, K H and Burstein, L (1980) 'Editors' notes' **New Directions for Methodology of Social and Behavioral Science,** 6,vii-ix.

Robinson, W S (1950) 'Ecological correlations and the behaviour of individuals', **American Sociological Review,** 15, 351-7.

Rubin, D B (1989) 'Some applications of multilevel models to educational data', in Bock, R D (ed.) **Multilevel analysis of educational data,** San Diego, Academic Press.

Ryan, T A, Joiner, B L and Ryan, B F (1986) **MINITAB Manual,** Duxbury, Mass.

Sayer, A (1984) **Method in Social Science: A Realist Approach,** London, Hutchinson.

Searle, S (1971) 'Topics in variance component estimation' **Biometrics,** 27, 1-76.

Silk, J A (1979) **Statistical Concepts in Geography,** Allen and Unwin, London.

Simpson, G G (1963) 'Historical science' in Albritton, C C (ed.) **The Fabric of Geology,** Stanford, New York, 24-48.

Thorndike, E L (1939) 'On the fallacy of imputing the correlations found for groups to the individuals or smaller groups composing them' **American Journal of Psychology,** 52, 122-124.

Thrift, N J (1983) 'On the determination of social action in space and time' **Environment and Planning, Series D,** 1, 23-57.

Tobler, W R (1970) 'A computer movie siinulating urban growth in the Detroit region' **Economic Geography,** 46,234-240.

Ware, J H (1985) 'Linear models for the analysis of longitudinal studies' **American Statistican,** 39, 95-101.

Listing of Catmogs in print

30:	Silk, The **analysis of variance.**	**3.50**
31:	Thomas, **Information statistics in geography.**	**3.00**
32:	Kellerman, **Centrographic measures in geography.**	**3.00**
33:	Haynes, **An introduction to dimensional analysis for geographers.**	**3.00**
34:	**Beaumont & Gatrell, An introduction to Q-analysis.**	**3.50**
35:	The agricultural **census - United Kingdom and United States.**	**3.00**
36:	**Aplin, Order-neighbour analysis.**	**3.00**
37:	Johnston & Semple, Classification using information **statistics.**	**3.00**
38:	Openshaw, The modifiable areal unit problem.	**3.00**
39:	Dixon & Leach, Survey research in underdeveloped **countries.**	**5.00**
40:	Clark, Innovation diffusion: contemporary geographical **approaches.**	**3.00**
41:	Kirby, Choice in field surveying.	**3.00**
42:	Pickles, An introduction to likelihood analysis.	**4.00**
43:	Dewdney, the UK census of population 1981.	**5.00**
44:	Pickles, Geography and humanism.	**3.00**
45:	Boots, Voronoi (Thiessen) polygons.	**3.50**
46:	Fotheringham & Knudsen, Goodness-of-fit statistics.	**3.50**
47:	Goodchild, Spatial autocorrelation.	**3.50**
48:	Tinkler, Introductory matrix algebra.	**4.00**
49:	Sibley, Spatial applications of exploratory data analysis.	**3.00**
50:	Coshall, The application of nonparametric statistical tests in geography	**7.50**
51:	O'Brien, The statistical analysis of contigency table designs	**3.50**
52:	Bracken, Higgs, Martin and Webster, A classification of geographical information systems literature and applications	**5.00**
53:	Beaumont, An introduction to market analysis	**3.50**
54:	Jones, Multi-level models for geographical research	**3.00**

Further titles in preparation

Order (including standing orders) from:

Environmental Publications,
University of East Anglia,
Norwich NR4 7TJ.

Prices include postage