A Review of Two Different Approaches for the Analysis of Growth Data Using Longitudinal Mixed Linear Models

Rien van der Leeden, Karen Vrijburg
Dept. of Psychometrics and Research Methodology
University of Leiden

Jan de Leeuw
Depts. of Psychology and Mathematics, UCLA

Subtitle:

Comparing Hierarchical Linear Regression (ML3, HLM) and Repeated Measures Designs with Structured Covariance Matrices (BMDP5V)

UCLA Classification: Theory and Methods

Abstract

In this paper we review two approaches for the analysis of growth data by means of longitudinal mixed linear models. In these models the individual growth parameters, (most often) specifying polynomial growth curves, may vary randomly across individuals. This variation may in turn be accounted for by explaining variables.

The first approach we discuss, is a type of multilevel model in which growth data are treated as having a hierarchical structure: measurements are 'nested' within individuals. The second is a version of a MANOVA repeated measures model employing a structured (error)covariance matrix. Of both approaches we examine the underlying statistical models and their interrelations. Apart from this theoretical comparison we review software by which they can be applied for real data analysis: two multilevel programs, ML3 and HLM, and one repeated measures program, BMDP5V. The programs are described and discussed with respect to several more general criteria, such as data setup and handling, implemented numerical routines and user friendliness, and, in particular, with respect to their application in longitudinal situations, i.e. their capabilities for the analysis of data on growth. Two data sets are used to compare the results of analyses performed by the three programs.

Although both ways of specifying growth curve models show some shortcomings, each appears to be a fruitful method to handle growth data, theoretically, as well as in a practical sense. For the most part, shortcomings are induced by the accompanying software, developed within different scientific traditions. Applied to comparable problems, the three programs produce equivalent results.

Preface

This paper has been written for anyone who is concerned with longitudinal research and the analysis of growth data in particular.

The drafts for this paper have their origin in 1989 and although data on growth have been here since the beginning of our times, the models and the analysis of these data have been subjected to several different viewpoints.

Traditionally, data on growth are conceived as 'repeated measures' and treated within a MANOVA framework. With the introduction of multilevel theory and with that the concept of 'random coefficient models' a whole new class of growth curve models emerged.

It was our intention to investigate these alternative approaches for the analysis of growth data, not only conceptually, but also practically by applying several popular computer programs to different datasets.

1. Introduction

In social science one frequently encounters research situations in which individuals are measured repeatedly over a period of time. The main goal of such longitudinal research can be considered as the assesment of change. Often, in social science research, this change concerns a kind of growth rate, e.g. biological, learning, etc. In that case, we could call the measurements 'growth data'.

Growth data, more specifically, can be considerd as measurements on one variable for the same (groups of) individuals on a number of consecutive points in time. In this paper we will be concerned with the analysis of individual growth as well as with the analysis of mean growth for a group of subjects.

Because data on growth reflect a developmental process as a function of time, a suitable method to model them is describing the expected values of the observations as functions of time. It has become very common to use polynomial functions for this purpose, and to present the results in the form of (growth) curves.

Generally, the analysis of longitudinal data is more complicated than that of cross sectional data, since repeated measurements for the same subjects are likely to be serially dependent. A useful property for growth curve models could therefore be the possibility of handling more complex, e.g. autoregressive, within-subject covariance structures.

Models for growth data

Many authors have considered the analysis of growth data from several points of view and a variety of techniques for fitting polynomial growth curves has been proposed. An overview of these developments, as well as a number of useful references, is given by Visser (1982) and Van der Leeden (1990). All the techniques are variations of a general linear model. Nevertheless, two major approaches concerning models for growth data can be distinguished. On the one hand, a (more traditional) class of models can be observed based on the MANOVA repeated measures design. In that case the consecutive measures on the growth variable are treated as separate dependent variables corresponding with the different timepoints. The work of

Potthoff and Roy (1964) has inspired much research on this class of models. Some references include Rao (1959, 1965, 1966, 1967), Khatri (1966), Grizzle and Allen (1969), Lee and Geiser (1972, 1975), Timm (1980) and Geisser (1970, 1980).

On the other hand there is a class of growth curve models emerging within the framework of multilevel analysis, which technique is also known as hierarchical linear regression analysis. In the application of multilevel models, data are viewed as having a hierarchical structure. Longitudinal data can be viewed as such: the measurements at multiple timepoints are 'nested' within the individuals. In longitudinal multilevel models, each individual is allowed to have his/her own growth curve, whereas the individual growth parameters vary across individuals. For more details concerning the multilevel treatment of longitudinal data we refer to Laird and Ware (1982), Goldstein (1986a, 1986b, 1987), Bryk and Raudenbush (1987, 1989) and Bock (1989).

Longitudinal multilevel models can be considered very appropriate for the modelling of growth data, conceptually as well as statistically. It can be shown that they incorporate two important features of real growth (cf. Bryk and Raudenbush, 1987; Prosser, Rasbash and Goldstein, 1990): observations on the different timepoints are correlated and the variance of the observations is a function of time (i.e. the between-subject variation is a function of time). Longitudinal multilevel models allow for the description of the mean growth trajectory and for the estimation of individual variation around this mean. If grouping variables are involved (e.g. sex group), these can be used as explanatory variables to account for the between-subject variation of individual growth parameters. In longitudinal multilevel models, differences in growth rate between groups of individuals will therefore be revealed not only in different mean levels of growth rate. Moreover, differences between groups will be modelled as different distributions of growth parameters within each group. Further, longitudinal multilevel models can treat unbalancedness due to missing observations very well, since it is not necessary for each individual to have the same number of measurements.

Compared to longitudinal multilevel models, repeated measures models appear to be less attractive for the analysis of growth data. Models of this kind are mostly concentrated upon (MANOVA-like) between-group designs and thus upon the estimation of (fixed) group effects.

Hence, they are not as flexible as one should wish, concerning the modelling of random variation of individual growth parameters around the mean growth trajectory. Generally the results consist of mean growth curves estimated for specified groups, found as deviations from the mean growth curve for the whole sample. In the ultimate case, groups may be reduced to individuals, and random error components could be interpreted as the most basic individual variation. But since estimation procedures for repeated measures model are generally not suited for N=1 groups and assumptions for (in that case) within-subject (co)variances are fairly simple and these (co)variances can not be modelled, this does not provide an appealing approach. However, recent developments include repeated measures models with structured covariance matrices, which are suited to handle incomplete data too. Jennrich and Schluchter (1986) present a very general model formulated at the level of the individual response vectors. Their model may be considered the most advanced repeated measures model as it includes all other previously suggested approaches as special cases, such as univariate mixed-model ANOVA and (versions of) MANOVA (for details see Jennrich and Schluchter, 1986). In the model of Jennrich and Schluchter, the within-subject (co)variances can be modeled as arbitrary functions of a set of unknown (co)variance parameters. The mean structure is specified by a set of regression parameters defined by a (fixed) between-subject design matrix. The general parametrization of the within-subject covariance matrix makes it possible to fit a numer of different structures, among which random effects and time-series structures are the most interesting for the analysis of growth data.

Purpose of the paper

In this paper we will review both approaches for the analysis of growth data described above. At first, we will examine the underlying statistical models and their interrelations. Although they are originating from different fields of application, their elementary concepts are basically alike. Secondly, we will give a review and comparison of several computer programs with which these models can be applied for data analysis. Concerning longitudinal multilevel models we will discuss ML3 and HLM. ML3 is software for two or three level analysis, written by Prosser, Rabash and Goldstein. HLM is a program for two level analysis written by Bryk,

Raudenbush and Congdon. A third program we will discuss is BMDP5V, designed to fit repeated measures models with structured covariance matrices and written by Jennrich and Schluchter.

We will focuss our discussion upon models in which the individual growth parameters may vary randomly across individuals and in which this variation may be, partially, explained by one or more variables on the group level. Models of that kind go under various names, such as 'random coefficient regression models', 'longitudinal random effects models', etc. Because it can be shown that these models can be split up into a random part and a fixed part (cf. Bryk and Raudenbush, 1987) we prefer the general term 'longitudinal mixed linear models'.

Limitations

The comparison of these three programs will be as detailed as possible. It will, however, be limited: not all programs could be run on the same computer system or in similar computer and user environments: ML3 and HLM were available as MS-DOS applications and ran on a 80286 based PC, but BMDP5V was run on an AMDAHL 5870 mainframe. So, one aspect which is not going to be used in our evaluation is the consumption of CPU time.

This paper deals with techniques for analyzing growth data. Therefore, our discussion is restricted to those features of the computer programs relevant for the modeling of these data. Clearly, these programs are meant to be more general and capable of computing various other (related) models for data analysis. ML3 and HLM can handle a variety of hierarchical linear models. BMDP5V is designed as a general procedure for the analysis of repeated measures data. A more complete discussion and comparison of both multilevel programs ML3 and HLM is given in Kreft, De Leeuw and Kim (1990) and in Kreft, De Leeuw and Van der Leeden (1991). BMDP5V is extensively discussed in Dixon (1988).

2. Comparison of models

Individual growth can often be modeled as a polynomial function of time. Suppose a subject j is measured on some response variable Y on T consecutive occasions, indicated by i. Then, the basic model to fit these responses with a polynomial growth curve looks like

$$Y_{ij} = \beta_{0j} + \beta_{1j} t_{ij} + \beta_{2j} t_{ij}^2 + \ldots + e_{ij} . \tag{2.1}$$

Using more convenient matrix notation, model (2.1) is written as

$$y_j = T \smallint_j + e_j , \tag{2.2}$$

where y_j is the (T x 1) response vector containing the repeated measurements for subject j, T is a (T x r, r ≤ T) matrix of known constants, $\smallint_j$ is a (r x 1) vector of individual parameters specifying the shape of the growth curve for person j and e_j is a (T x 1) vector of random error components. The matrix T can be defined as

$$T = \begin{bmatrix} 1 & t_1 & t_1^2 & \ldots & t_1^{T-1} \\ 1 & t_2 & t_2^2 & \ldots & t_2^{T-1} \\ . & . & . & \ldots & . \\ 1 & t_T & t_T^2 & \ldots & t_T^{T-1} \end{bmatrix}. \tag{2.3}$$

Since T is of order (T x r, r ≤ T), the degree of the polynomial fitted for person j is r-1, and T-1 at most. Note that in this formulation, it is not necessary for each individual to have T measurements. Matrix T may contain a set of polynomial vectors (see e.g. De Lury, 1950), or it may contain an explanatory variable, as well as powers of that variable, corresponding to the T occasions, such as 'weight' or 'age'.

Model (2.2) is well known in the literature. It can be considered as a version of the polynomial

regression model with the special feature that the observations, i.e. the repeated measurements contained by y_j, are not indepently distributed. The problem of estimating $\hat{T}_j$ can, for instance, be solved using generalized least squares (cf. Rao, 1973; Timm, 1975; Visser, 1982).

In the sequel we will take this model for individual growth as a common basis to compare the formulation of longitudinal mixed linear models within a multilevel framework and within a generalized MANOVA repeated measures design. In this kind of models, the individual growth parameters vary randomly across a group of subjects and this variation may be accounted for by variables on the group level.

Longitudinal mixed linear models in a multilevel framework

Multilevel models are designed for the analysis of data with hierarchical structure. The term 'multilevel' points out that we study nested membership relations among the units of observation. Models for the analysis of hierarchical data and their corresponding estimation procedures have been extensively discussed in the literature. Among others we mention Goldstein (1986a; 1987), Raudenbush (1988) and Bock (1989).

Usually, in multilevel model building, individuals are considered as the lowest (level 1) level units. For instance, students are members of classes, and classes (level 2) are nested within schools. Growth data can be interpreted as having a hierarchical structure too. Data from individuals that are measured at a number of consecutive points in time can be interpreted as having a two-level structure: individuals are considered as level 2 units and observations as level 1 units, that is, measurements are nested within individuals. In that way, a longitudinal mixed linear model is easily formulated within a multilevel framework. Equation (2.2) serves as the *within-subject* (level 1) model for person j (which is a level 2 unit). Clearly, each person has its own set of growth parameters, i.e. each person's growth curve is unique. The variation of the individual growth parameters across the total group is modeled by the *between-subject* (level 2) model (2.4b). Hence, the complete two-level model for growth data is given by

$$y_j = T\!\!\int_j + e_j \,, \qquad\qquad (2.4a)$$

$$\smallint_j = \mathbf{Z}© + \mathbf{u}_j \, . \tag{2.4b}$$

In equation (2.4b), $\mathbf{Z}$ is a (r x q) between-subject design matrix with known, fixed elements, © is a (q x 1) vector of fixed coefficients and $\mathbf{u}_j$ is a (r x 1) vector of random error components. The *between-subject* model (2.4b) is formulated in a general way. For instance, if $\mathbf{Z} = \mathbf{I}$, we have a model of simple variation in which © is a vector containing parameters indicating the mean growth curve over all individuals, whereas the elements of $\mathbf{u}_j$ denote the departures from this mean curve for each individual j. More elaborated models emerge if $\mathbf{Z}$ contains dummy variables coding subgroups of individuals and/or other explanatory variables that could account for the variation of the $\smallint_j$'s. In that way it is possible to study the variability of the growth curve coefficients as a function of the differences between the individuals on some characteristic that remains fixed across occasions (e.g. sex). Another extension to the model would be to include time varying covariates to $\mathbf{T}$ at level 1.

The *within-* and *between-subject* models can be combined into one single equation model that is given by

$$\mathbf{y}_j = \mathbf{TZ}© + \mathbf{Tu}_j + \mathbf{e}_j \, . \tag{2.5}$$

Usually it is assumed that the elements of $\mathbf{e}_j$ are independently distributed as $\mathbf{e}_j \sim N(0, \sigma^2 \mathbf{I})$, that the level 2 random terms in $\mathbf{u}_j$ are distributed as $\mathbf{u}_j \sim N(0, \text{„})$, and that the level 1 random terms, the elements of $\mathbf{e}_j$, are distributed independently from the level 2 random terms in $\mathbf{u}_j$.

In the combined model (2.5) the term $\mathbf{TZ}©$ can be called the fixed part and the term $\mathbf{Tu}_j + \mathbf{e}_j$ the random part. This explains why we can view the multilevel (longitudinal) model as a 'mixed' model. Some useful references concerning the multilevel treatment of longitudinal data can be found in section 1.

The variance of the random part, i.e. the (co)variances among the elements of $\mathbf{y}_j$ that are not accounted for by the fixed part, is given by

$$\text{Var } (Tu_j + e_j) = \text{Var } (y_j|TZ\copyright) = \hat{f}_j = T\Omega T' + \sigma^2 I \hspace{2cm} (2.6)$$

The assumptions described above, especially that $e_j \sim N(0, \sigma^2 I)$, lead to a relatively simple covariance structure on the first (individual) level: the error terms are equal and uncorrelated across the timepoints. Although it is possible to formulate a more general model incorporating a complex level 1 covariance structure, Bryk and Raudenbusch (1987) argue that this simple structure is sufficient for most longitudinal applications with a limited number of timepoints.

The multilevel software that we compare and review in this paper, which implements the longitudinal mixed model, includes ML3 and HLM. Both programs are designed for multilevel analysis in general. In HLM one can specify two-level models, in ML3 two or three-level models. For details concerning these programs we refer to a later section.

Longitudinal mixed linear models in a MANOVA repeated measures context

There exists a huge amount of literature concerning MANOVA and MANOVA repeated measures models. They are discussed extensively e.g. in Kirk (1982) in an experimental framework, and in many textbooks on multivariate data analysis, such as Timm (1975). Relevant for our present discussion is the following. Typically, MANOVA models are focussed upon the *between-group/subject* side of the analysis. The total variance to be explained in the dependent variables, is accounted for by differences in group membership as good as possible. Main effects, as well as interaction effects are estimated by looking at differences between the overall mean and the group means. Usually, this takes place under fairly simple assumptions for the *within-group/subject* covariances.

MANOVA models therefore typically emphasize the fixed part of the modeling. An 'explaining' structure of the means is estimated and the *within-group/subject* covariances are considered as random error, that is, they are the final part of the variance in the dependent variables that can not be explained by group membership.

In repeated measures applications the fixed part of the model is expanded. To fit polynomial growth curves of a certain degree, the set of explanatory variables, first indicating group

. membership alone, is supplemented with within-subject variables (and powers of that variables) corresponding to the different timepoints, such as 'age'. On the individual level, we can write for such a model

$$y_j = X\!\int + v_j \,, \tag{2.7}$$

Again, y_j contains the T repeated measurements for a single subject j, X is a known (T x q) design matrix and $\int$ is a (q x 1) vector of regression coefficients. The v_j are vectors of random error components. Since the repeated measures are likely to be dependent, it is assumed that the v_j are independently distributed as $N(0,\!\int_j)$. The columns of X correspond to the different terms in the model, i.e. the elements of $\int$. Normally the first column is a vector of ones, neccesary to specify the intercept. Further columns code (between-subject) grouping variables and (within-subject) explaining variables (and powers of that variables). The fixed parameters in $\int$ can be combined so that we find the parameters of a mean growth curve for the whole group and of separate (mean) curves for the subgroups, found as deviations from this mean curve.

Clearly, model (2.7) is conceptually not aimed at the modeling of individual growth. Only if X would be specified in such way that each individual forms a subgroup, we could estimate individual growth curves. However, this option is not very attractive from a statistical point of view, not in the least because we would be confronted with the problem of incidental parameters: each new individual adds another set of parameters to the model (see e.g. Kendall & Stuart, 1979; Van der Leeden, 1990).

So, even though the fixed part of the repeated measures model contains both between- and within-subject elements it still remains a fixed part. To include random variation of the individual growth parameters in model (2.7), we have to give the random part of it a more complex structure. This can be accomplished by specifying

$$v_j = Z b_j + r_j \,, \tag{2.8}$$

where v_j is as before and Z is a known (T x r) matrix. The (r x 1) b_j and (T x 1) r_j are

independent random vectors. It is assumed that the b_j and r_j are independently, identically distributed as $b_j \sim N(0, \ddot{I})$ and $r_j \sim N(0, \sigma^2 I)$, respectively. If Z is specified accordingly, b_j contains the (random) growth parameters for subject j. This holds if its columns contain the same (within-subject) explaining variables (and powers of that variables). The r_j contain the remaining random error which can not be accounted for by the fixed part modeled with X and the random part modeled with Z. As in the previous section both models (2.7) and (2.8) can be combined into one single equation model

$$ y_j = X \int + Z b_j + r_j . \qquad (2.9) $$

As it holds for model (2.5), this model can be split up into a fixed part and a random part too. The variance of the random part in (2.9) is given by

$$ \mathrm{Var}\,(Z b_j + r_j) \; = \mathrm{Var}\,(y_j | X \int) = \acute{I}_j = Z \ddot{I} Z' + \sigma^2 I . \qquad (2.10) $$

A comparison of models (2.5) and (2.6) with (2.9) and (2.10) shows that both approaches, i.e. the multilevel and the MANOVA repeated measures framework, if specified accordingly, may lead to the same longitudinal mixed linear model (see also Vrijburg (1991) for a comprehensive discussion on this comparison). If we let $X = X_F$, $\int = ©$, $Z = T_R$, $b_j = u_j$ and $r_j = e_j$ in (2.9) and if we let $TZ = X_F$ and $T = T_R$ in (2.5), we can write this longutidinal mixed linear model as

$$ y_j = X_F © + T_R u_j + e_j , \qquad (2.11) $$

where the u_j and e_j are independently, identically distributed as $u_j \sim N(0, \text{,,})$ and $e_j \sim N(0, \sigma^2 I)$, respectively. The variance of the random part, $\mathrm{Var}\,(T_R u_j + e_j)$, is then written as $T_R \text{,,} T_R' + \sigma^2 I$. In the literature, models employing this relatively simple within-subject covariance structure generally go under the name of 'random coefficient models' (De Leeuw and Kreft, 1986; Prosser et al., 1990).

3. Description of the programs

In this paper we review two different approaches for the analysis of growth data using longitudinal mixed linear models. In the previous section we have examined the underlying statistical models and their interralations. In this section we review and compare three computer programs with which these models can be applied for data analysis. The programs are ML3, HLM and BMDP5V. They are described on the basis of the following topics:

- Design philosophy.
- Implementation details.
- Implemented model.
- Routines.
- Data setup and data handling.
- Output and results.
- User friendliness.
- Special features.

In this paper we are directing our attention to longitudinal applications, i.e. to the analysis of growth data. In the descriptions of the programs, we have tried to balance the emphasis given to aspects relevant for those applications, and to more general features. A certain quantity of general remarks is very valuable and unavoidable too. Nonetheless, some aspects are treated briefly. For a very detailed, extensive discussion of ML3 and HLM we refer to Kreft, De Leeuw and Kim (1990). In Kreft, De Leeuw and Van der Leeden (1991), the application of BMDP5V for longitudinal mixed models is discussed too.

Program ML3

A. Design philosophy

The ML3 program is designed for multilevel analysis with two or three levels. One is able to fit mixed linear models to data with a two or three-level hierarchical structure, in which the coefficients of the explanatory variables may be random. The program is written by Prosser,

In the sequel we will adopt the notation described above. Hence, fixed effects will be indicated with ©'s and the variances of the random parameters with Σ's, the elements of „. The variance of the elements of e_j (which are constant in our case) will be denoted by σ^2.

Rasbash and Goldstein. It is described and documented in Prosser, Rasbash and Goldstein (1990), the ML3 user's guide. ML3 is combined with the more general statistical package NANOSTAT written by Healy (Healy, 1989), which makes a variety of utility functions available to the program. The theory on which ML3 is based is described in Goldstein (1987) and Goldstein and Rasbash (1989). Models that can be estimated using the program include random coefficient regression models, growth curve models and hierarchical logit models. Fitting growth curve models with ML3 is basically a two-level analysis of longitudinal data. It means that the repeated measurements (level 1) are viewed as being nested within individuals (level 2). If necessary, the individuals may be grouped on level 3. The handling of longitudinal data in a multilevel framework is discussed in Goldstein (1986a, 1987, 1989) and in Bryk and Raudenbush (1987).

B. Implementation details

ML3 is available as an executable program only. That is, it is shipped without the program source. It runs on 80286- and 80386-based personal computers under DOS/OS2. Proper operation of the program requires 640 Kb of RAM and a hard disk. A CGA, EGA or Hercules graphics card is required if one wishes to use the program's high-resolution graphics capabilities. ML3 is available either as a math coprocessor version or as a non-coprocessor version.. It is shipped on a single diskette in compressed form. Installation is simple as the files are easily expanded and tranfered to a subdirectory on the hard disk. According to the ML3 User's guide, insufficient memory is the most common problem that new users experience. For this paper we ran ML3 Version 2.1 on a 80286 PC with math coprocessor.

There are two versions of ML3 available for analyses involving very large data sets. ML3/E is especially compiled for 386-based personal computers with expanded RAM. It lets you work with all the extended memory available. The very recent ML3/V is a special version of ML3 for VAX computers running under VMS.

C. Implemented model

In ML3 the basic model implemented is the mixed linear model for data with a two or three-level

nested structure. At any level of the hierarchy, the coefficient of any explanatory variable may be defined random or fixed. The random coefficients at each level may have any (co)variance pattern. Concerning models for longitudinal data, ML3 can fit polynomial growth curve models, as well as growth models fitting functions with free coefficients. In the manual, the latter models go under the name of fixed occasion models.

D. Routines

In ML3 the model parameters are estimated using an iterative generalized least squares (IGLS) algorithm. This procedure yields consistent estimates of model parameters and maximum likelihood estimates if the assumption of multivariate normality is valid. The program can also compute unbiased or restricted (RIGLS) estimates. Under the assumption of multivariate normality, these estimates are called restricted maximum likelihood (REML) estimates. IGLS is described by Goldstein (1986b, 1987). RIGLS is described by Goldstein and Rasbash (1989). The RIGLS procedure is comparable to REML in the other multilevel program discussed in this paper, HLM by Bryk, Raudenbush and Congdon. The RIGLS procedure adds bias correction terms to the IGLS estimates on each iteration. This implies that during the estimation of the random parameters, the fixed coefficients are treated as quantities which have uncertainty "built in". In IGLS, the fixed coefficients are treated as known when estimating the random parameters. Starting values for both IGLS and RIGLS procedures are provided by an ordinary least squares (OLS) regression using all cases and ignoring grouping.

According to the ML3 User's guide, RIGLS and IGLS estimates may differ considerably for snall data sets. IGLS appears to be the default method of estimation. However, neither the reason why, nor further information concerning these estimation procedures, are given in the manual.

The number of iterations may range from 1 to 999, but the default number is set to 5. According to the manual this number is sufficient to reach convergence if several criteria are met, such as a number of observations per unit large enough for stable estimation, a small number of parameters to be estimated, and so on (further details are given in the ML3 manual).

E. Data setup and data handling

ML3 stores data in a memory segment called the worksheet. This worksheet is divided into 100 columns which can be refered to as C1, C2, etc. When a data file is read in, these columns will hold the variables. On the worksheet there are 100 boxes too, refered to as K1, K2, etc., which are capable of holding single numbers, for instance, the result of some computation. A third section of the worksheet is meant for the specification of the model. For convience, the data in a worksheet can be stored in a file for use in future work sessions.

The input file needs to be a standard, numeric ASCII file. File reading may be unformatted or formatted, using a FORTRAN format statement. Small quantities of data, for instance starting values, may be entered by keyboard.

Records in the input file are considered individual cases as usual. In regular multilevel analysis, individuals are nested within groups. Hence, ML3 expects each case to be labeled with a group identifier (ID2 -> level 2) and a case identifier (ID1 -> level 1). However, when analyzing repeated measures data within a two-level context, the individuals are considered as level 2 units, and the measurements-within-individuals are the level 1 units. Hence, the usual structure of repeated measures data needs to be transformed accordingly: each occasion should be given a separate record. For the Potthoff-Roy data (Potthoff and Roy, 1964; see also section 4), in the case of a quadratic model where growth is modeled as a function of age (and age^2), with one explanatory grouping variable (sex), this may look like:

1	01	21	8	64	1
2	01	20	10	100	1
3	01	21.5	12	144	1
4	01	23	14	196	1
1	02	21	8	64	1
2	02	21.5	10	100	1
.	...	.....	....		.
.	...	.....	....		.
3	27	23.5	12	144	-1
4	27	25	14	196	-1

occasion	person	response	age	age^2	sex
number	ID code	variable	(explanatory variables)		

Fortunately, ML3 contains a utility for converting a "one record per individual file" into a "one record per occasion" file.

On the whole, data handling is very flexible in ML3. A variety of data operations can be performed within the NANOSTAT package, such as computational and statistical transformations, coding and recoding, generation of dummy variables, selection and sorting of cases, and so on. Auxiliary files can be used. Even vectors and matrices can be inputted.

Missing data can be handled if they are indicated with a numerical code chosen by the user. Several options are available including the imputation of estimated values to replace missing data and listwise deletion.

F. Output and results

The default output is none, program convergence is the only information reported. All special output can be acquired by commands. "Rand" and "fixe", for instance, generates the estimates of the random coefficients and the fixed coefficients, respectively, as well as their accompanying standard errors. Although not documented in the manual, "like" produces the - 2 log likelihood value (also called "deviance") of the estimated model. Commands enable the user to compute predicted values and residuals. This is of particular interest in the analysis of growth curve data if one is interested in the individual growth trajectories.

All output may be printed on the screen or written to an output file specified by the user.

The NANOSTAT package provides the user with several graphic commands. Plots such as stem-and-leaf plots, boxplots ansd scatterplots may be obtained.

For a more elaborate discussion of the options mentioned above, as well as for further features, we refer to the report of Kreft, De Leeuw and Kim (1990).

G. User friendliness

The program is extensively documented, the manual is clearly written and complete. A whole chapter with examples is included in which the log files of six example runs (with different submodels) are given and discussed. The manual provides a nice introduction to multilevel modelling theory too, with many references. This introduction may be of great help to a novice in this kind of data analysis.

The interactive capabilities of ML3 are very powerful. Users have much freedom in making adjustments during the run. The specification of a model is very flexible. One proceeds by entering commands that tell the program the roles of the input, or created, variables and which parameters are to be estimated. A screen displaying all current specifications and choices may be called up at any time entering the command SETT. An online help program is built in too, which shows the commands and their syntax.

The ML3 program offers many possibilities to advanced users of multilevel analysis. There are many options, not in the least because of the extensive NANOSTAT package which is merged with the program. The program is probably not an easy one to start out with for an unexperienced user.

H. Special features

There are a number of special features. We mention the possibility to enter starting values for parameters other than default OLS estimates. Further, estimation of certain, slow converging, parameters can be 'freezed' during the estimation proces, to speed up convergence. Special is the ability to fit simple multilevel logit and log-linear models. This allows the analysis of data with proportions or binary variables as dependent variables.

Program HLM

A. Design philosophy

The abbreviation HLM refers to the Hierarchical Linear Model. The program HLM, implementing this model is written by Bryk, Raudenbusch and Congdon (1988) and is designed for the analysis of multi-level data with a two level hierarchical structure. The program is

described and documented in Bryk, Raudenbush, Seltzer and Congdon (1988), the HLM user's guide. According to this manual the program is aimed at two broad classes of applications: the familiar analysis of nested data, as in the study of school effects, for instance, and the analysis of growth curve data. In both cases we have a within-unit and between-unit representation: within and between class analysis and within-subject (time-series observations) and between-subject analysis, respectively.

Recently, a beta-version of HLM 3 has been released. This program can deal with three-level data.

B. Implementation details

HLM is available in two versions, one for workstations or mainframes and one for the PC operating under MS-DOS. The PC version of the program is an executable only, i.e. there is no source code available. The other version comes with the source code too (the source of HLM used to be in the public domain, but this appears to have come to an end with version 2.1). PC users may run HLM on machines ranging from simple 8086 XT's to fast 80386 computers with math coprocessor. The computational kernel of the program is written in FORTRAN 77. Some additional routines have been written in C. For the PC version, current program limitations (see the manual, p. 39) are: a maximum of 10 within-unit variables (either fixed or random) in a model, a maximum of 15 between-unit variables and a maximum of 35 fixed effects in all equations. HLM uses the input data to create a file of "sufficient statistics". The current maximums for this sufficient statistics file are 25 within-unit variables, 25 between-unit variables, and a total of 300 units. For the non-PC version, these maximums are slightly higher, although they can all be changed if the parameter statements in the source code are modified.

For this paper we used HLM version 2.1 running on a 80286 PC with math coprocessor.

C. Implemented model

The basic model implemented in HLM is the two-level hierarchical model. By default, the program assumes that the coefficients of the within-unit model are to be treated as random, but this can be changed. In the interactive process during the run, the user may apply additional

constraints to the coefficients. Either the variance or the mean of one or more of the within-unit coefficients can be set to zero. The result is that we can have models in which a within-unit variable has a fixed effect or a random effect only, or both.

Also by default, both the within and between model have a base term (intercept). This can be overruled too. For instance in growth curve analysis it may be of interest to fit models without a within-unit intercept.

Concerning model for growth data, HLM has roughly the same possibilities as ML3, except for a third level.

D. Routines

HLM implements an EM algorithm for its parameter estimation (Dempster, Laird and Rubin, 1977). The algorithm is supplemented by an Aitkin accelerator (Laird, Lange and Strom, 1987). According to the manual, using this accelerator a 30 to 40 percent reduction in computation time can be achieved. The number of iterations is optional. The manual suggests that 10 iterations are often sufficient for explanatory analysis, but recognizes the fact that generally this number will not be high enough to reach convergence in practice. In fact, the EM algorithm is well known for its notoriously slow convergence, especially when variance estimates are close to zero (close to a boundary of the parameter space). In those cases, a number of over 100 iterations is easily reached.

A second algorithm provided by HLM is the so called "V-known" routine. This program is in fact a very general, multivariate regression routine for cases in which the sampling variances and covariances are (assumed) known for each unit. The V-known routine is mainly used in meta-analysis. Further details concerning this topic are given by Raudenbush and Bryk (1985).

E. Data setup and data handling

If a user wants to analyze a particular data set for the first time, HLM needs two input files: a within-unit and a between-unit file. From these files, the program computes a file of sufficient statistics, which can be used in all subsequent analyses of the same data. For growth data the within-unit file has the familar "one record per occasion" structure (see the section on ML3).

Subjects are identified by an ID number in the first field of each record, which must be read in text format. This also holds for the between-unit file which uses these ID numbers to link the between-unit information for a unit to its within-unit data. ID numbers in both files must therefore be completely consistent. For the Potthoff-Roy data (see section 4), in the case of a linear model where growth is modeled as a function of age, these files may look like:

within-unit file				between-unit file	
01	21	8		01	1
01	20	10		02	1
01	21.5	12		03	1
01	23	14		...	...
02	21	8		16	1
02	21.5	10		17	-1
...	....			18	-1
...	....			...	...
27	23.5	12		27	-1
27	25	14			

person	response	age		person	sex
ID code	variable			ID code	

A HLM utility, PRINTSSM, can be used to print and thereafter inspect the sufficient statistics file when it is first created.

HLM will also accept a SYSTAT system file as input (SYSTAT is a statistical package for use on personal computers). Using the V-known routine requires a different input file structure. In that case the input includes estimates of within-unit sample variances and covariances (see the manual, p. 57 for further details).

The program can handle missing data in the within-unit file only. It is assumed that the between-

unit file contains no missing data. If so, the user has to take action such as delete the case or impute a value. Handling of missing data can be listwise or pairwise deletion.

F. Output and results

By default HLM produces a complete and comprehensive output file, for which name the user is prompted. Standard errors for the fixed parameter estimates are given as well as a table with t-statistics and p-values. The variance estimates (the random parameters) are tested in a chi-square table with which one can decide if a sufficient part of variance is explained by variables on the second level. Reliability estimates of these parameters are provided too. Model fit can be assessed with a deviance statistic. Options control the printing of OLS estimates for each unit. Either for ten or for all units the within-unit regression parameters are printed. For growth curve applications this is of interest: it provides the user with the individual growth trajectories, wheras the average growth trajectory for the whole sample is given too. The output concerning the history of iterations is controlled by the user too. The value of the likelihood function is given either for all iterations or only for the first and the last one.

For an extensive discussion of the options mentioned above, as well as for further features, we refer to the report of Kreft, De Leeuw and Kim (1990).

Upon request HLM produces a so called residual file, providing a means of checking the fit and distributional assumptions of the between-unit model. This file can be read in into SAS, SPSS or SYSTAT if the user wants to compute various diagnostics.

G. User friendliness

The program is extremely interactive: it starts by asking the user the first of a long series of questions, which guides the user step-by-step through the model specification process. This makes the program fairly easy to use.

The manual is definitely of great help to an unexperienced user. It is build up along the lines of the annotated output of a number of example runs. A serious disadvantage of this approach is that relevant information is scattered throughout the manual. Specific information is mixed with example output, instead of organized under specific headings. It is therefore not always easy to

find.

H. Special features

Compared to the other two programs, HLM is the only one that provides a variety of tests. We mention: a t-test for significance of the fixed parameters, a chi-square test for residual unexplained variance in the first level parameters, a reliability estimate of the first level variables, a multivariate hypothesis test involving the fixed effects, a homogeneity of variance test for the residual variation in each unit and a likelihood-ratio test for the variance-covariance components model.

Although the program is meant to be totally interactive, it can also be executed in a batch mode using an extended default file (or "personal" default file). This can be of great use to bypass irrelevant prompts and other cumbersome interactive features as one becomes a more experienced user.

Program BMDP5V

A. Design philosophy

The BMDP5V program is designed by Jennrich and Schluchter and described in Schluchter (1988). The theory on which the program is based is given by Jennrich and Schluchter (1986). Being a procedure of the BMDP software package, the program is extensively documented and illustrated in the BMDP user's guide (Dixon, 1988).

BMDP5V is especially meant for the analysis of repeated measures with special emphasis on unbalanced situations, including unbalance caused by missing data. Although the program has many possible applications, it is developed with small experimental data sets in mind, especially in a biological context. When there are many observations and/or a large number of parameters to be estimated, the program becomes expensive to use.

B. Implementation details

BMDP5V is provided at any site where the BMDP package is running. Usually this package has

been implemented on mainframes and mini computers, so there are versions working under different operating systems such as MVS, VM, VAX/VMS and so on. More recently, PC versions of the package have become available for use under MS-DOS or OS/2.

The BMDP procedures can be operated either in batch mode, or in interactive mode, depending on the user's preferences and/or the available computer facilities.

BMDP5V shows no limitations concerning the number of individuals and variables. However, the memory needed for data storage will increase with these numbers, and will also depend on the specified model. Probable implementation restrictions could therefore arise dependent on the computer system used for running BMDP.

C. Implemented model

The models that can be fitted in BMDP5V belong to a general class of multivariate linear models. In these models a set of regression parameters describes the structure of the expected values of the observations, and a set of covariance parameters provide for a general parametrization of the within-subject covariances. A number of special structures for these covariances are built-in. User-defined covariance structures can be specified too.

When repeated measures data are analyzed using a generalized multivariate linear model such as in BMDP5V, the individuals are considered as data records and the observations on multiple occasions are interpreted as separate dependent variables. In a multilevel context, this makes individuals the second level within which the (first level) observations are nested.

D. Routines

BMDP5V uses three different algorithms to compute maximum likelihood estimates for all model parameters: a Newton-Raphson, a Fisher scoring and a generalized EM algorithm. Producing identical results, these methods differ with respect to the number of iterations required and the costs per iteration. Generally, EM will converge slowly, but with low costs per iteration, whereas Newton-Raphson will require the smallest number of iterations, but has the highest cost. BMDP advises to use Newton-Raphson if the number of covariance parameters is 15 or less, otherwise EM is preferable. BMDP5V offers two different algorithms to compute

restricted maximum likelihood estimates for the covariance parameters: generalized EM and a quasi-scoring algorithm. Results will be identical but convergence is faster for the scoring method with higher cost and vice versa for the restricted ML method. For any of these procedures, the maximum number of iterations is set to 15 by default. If neccessary, this number can be altered by the user.

E. Data setup and data handling

The data input file must be organized so that individuals are the records and the repeated measurements are the variables. Note that this is different with respect to the way in which hierarchical data are usually organized.

We have run the program in batch mode. The BMDP instructions are given by keyword-statements, which look much alike usual SPSS statements. Input data and BMDP instructions can be stored in the same file, or we may use separate files for both. Keywords control the names of the variables, their number, format, tranformations, and so on. The actual modeling of the data is controlled by three keywords-statements: DESIGN, MODEL and STRUCTURE.

The DESIGN statement specifies the structure of the data. It identifies variables classifying subjects and variables measured repeatedly. The models itself are specified in the MODEL statement, using variable names. The STRUCTURE statement specifies the structure of the within-subject covariances. An example of a set of BMDP instructions for the Potthoff-Roy data, is given below.

```
/PROBLEM   TITLE = 'POTTHOFF-ROY DATA'.

/INPUT     FILE = DENTAL.DAT.

/DESIGN    GROUPING = SEX.

           REPEATED = TIME.

           LEVEL = 4.

           DPNAME = DISTANCE.

           DPVAR = DIST1, DIST2, DIST3, DIST4.

           CONTRAST (TIME) = AGE.
```

```
                 AGE = 8, 10, 12, 14.
/MODEL           DISTANCE = 'SEX + AGE + SEX.AGE'.
/STUCTURE        TYPE=UNSTRUC.
```

In this example (see section 4), data are defined and modeled consisting of measurements of some dental distance for a group of children, obtained at ages 8, 10, 12 and 14 (DIST1, DIST2, DIST3 and DIST4). As a characteristic of interest, remaining fixed over occassions, we have the grouping variable sex (SEX). In the example setup, the within-subject factor is defined to be TIME. The CONTRAST statement creates a variable, AGE, taking the values 8, 10, 12 and 14 respectively, for each of the values of TIME. The model specified here fits linear growth curves to the data. The curves are found as a regression of distance on age. The coefficients of the individual curves vary randomly across the individuals. A mean growth curve for the whole group, as well as mean growth curves for boys and girls separately, are easily produced by combining the fixed coefficients of the model in the correct way. The expression SEX.AGE adds an interaction term to the model. We will return to this data in section 4 where the results of the different programs are compared.

In the STRUCTURE statement one of the built-in structures for the within-subject covariances can be specified. In the example setup above, these covariances are contained in an arbitrary matrix (fully parametrized). Other built-in special structures include first order autoregressive models, compound symmetry and banded or general autoregressive structures. For some structures additional input is needed. A factor analytic structure requires the number of factors to be specified. Random effects requires the input of a (known) matrix containing the variables that are to be used for the specification of random coefficients. For the Potthoff-Roy data the additional input in that case may look like:

```
/STUCTURE        TYPE=RANDOM.
                 ZCOL1 = 1, 1, 1, 1.
                 ZCOL2 = 8, 10, 12, 14.
```

A general linear structure requires the number of parameters to be specified as well as a set of known matrices. Finally a user-defined covariance structure which is not built-in and not linear, can be specified by adding a FORTRAN 77 subroutine to the program input. This routine becomes a part of the program, called in each iteration.

F. Output and results

The output of BMDP5V is controlled by the PRINT statement and can be very comprehensive. There are no special output files. Program instructions, model specifications and (default) values of various program parameters are extensively listed. An extended output presents, for each iteration, a table with the log likelihood of the specified model, the values of the regression parameters and the values of the covariance parameters. Akaike's information criterion is provided to check the appropriateness of a chosen covariance structure. Asymptotic standard errors and z-values are given for all maximum likelihood estimates of the model parameters. For the terms of the regression model Wald (chi-square) tests of significance are included. Within-subject and all-pairs within-subject covariance and correlation matrices are given. A very useful option, especially when analyzing individual growth trajectories, is a listing of individual responses (at the different timepoints), their predicted values, residuals and standardized residuals. With this listing, Mahalanobis distances are provided. These distances can be helpful in detecting cases which may be outliers.

G. User friendliness

The BMDP user's guide is written very clearly. BMDP5V, as any BMDP procedure, is extensively discussed and commented. The different options are very well illustrated presenting a variety of real data examples. Most technical details can be found in the appendix.

The program is easy to use. As far as the keyword command structure, and the ability to run BMDP in interactive mode is considered user-friendly, the program is too.

When repeated measures are to be analyzed, the built-in autoregressive structures for the within-subject covariances can be very useful. However, to include a user-defined covariance structure by adding a FORTRAN 77 subroutine to the program may be quite harassing for regular users.

H. Special features

A unique option of BMDP5V is the ability to analyze incomplete data. Using an EM algorithm, values for missing data can be imputed. In that case it is assumed that there is an underlying set of complete data from which the observations of each subject form a subset. This property is very useful because it is been known to improve the reliability of the data analysis (cf. Little and Rubin, 1987). An option like this is certainly not very common in statistical software.

Another special feature is that BMDP5V can deal with time-varying covariates. For instance, in the case that we have multivariate repeated observations, that is two or more dependent variables measured repeatedly, we could take one (or more) of these variables as a covariate. This feature is built-in and easy to use.

4. Comparison of results

Introduction

In this section we present comparisons of the outcomes of growth curve analyses with ML3, HLM and BMDP5V. In doing this, two different datasets will be used: a small data set called DENTAL data, and a large data set called SCHOOL data (for a description see below).

First, a description of the data will be given, followed by the models which we have analysed. Each model will be accompanied by a table with results comparing the three programs. For ML3 we have included results from two estimation methods available within this program: full information and restricted maximum likelihood, indicated by ML3F and ML3R respectively. As mentioned before we have concentrated upon models in which the individual growth parameters vary randomly across individuals and in which this variation is (partially) explained by variables on the group level.

DENTAL data

The DENTAL data were already briefly mentioned in Section 3 where data setup and handling in the three programs were discussed. The data are taken from Potthoff and Roy (1964). They were collected at the University of North Carolina Dental School and concern the measurements of the distance (DIST) from the centre of the pituitary to the pteryomaxillary fissure for eleven girls and sixteen boys, at ages of 8, 10, 12 and 14.

Both approaches for the analysis of growth data using a random coefficient model have been applied to these data: a multilevel analysis and a generalized MANOVA. Conceptually, the random coefficient model as a general case of the multilevel model is equivalent to the random coefficient model as a special case of the generalized MANOVA-model of Jennrich and Schluchter (Jennrich and Schluchter, 1986). Consequently, we will compare the results of the two multilevel programs, ML3 and HLM, with the results of the MANOVA program of Jennrich en Schluchter, BMDP5V.

The analysis consists of fitting linear growth curves found as a regression of distance in age. In a multilevel framework the repeated measurements are considered nested within the subjects. On

the first level the individual growth curve coefficients are treated as random variables. On the second level (SEX), defined as a dummy variable with values -1 for girls and 1 for boys, is used as an explanatory variable of random variation across the indiviual growth trajectories. The within-subject model can be written as

$$(DIST)_{ij} = \beta_{0j} + \beta_{1j} (AGE)_j + e_{ij} ,$$

and the between-subject model is

$$\beta_{0j} = \gamma_{00} + \gamma_{01} (SEX)_j + u_{0j} ,$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} (SEX)_j + u_{1j} .$$

The single equation version of the model is written as

$$(DIST)_{ij} = \gamma_{00} + \gamma_{01} (SEX)_j + \gamma_{10} (AGE)_j + \gamma_{11} (AGE)_j(SEX)_j +$$
$$+ u_{ij}(AGE)_j + u_{0j} + e_{ij} .$$

Jennrich and Schluchter define the random coefficient model in a slightly different way. According to the MANOVA tradition one typically wants to research whether one or more dependent variables (or the mean score on it) differs between several subpopulations. The dimensions along which these subpopulations are divided, are considered the independent variables. In this example the independent variables are (SEX) and (AGE). In other words, one wants to study the joint effect of the independent variables on one (here DIST) or more dependent variables. In the MANOVA tradition these are all *fixed* effects. Random variation of the individual growth parameters can be achieved by a modeling of the error structure, as was explained in Section 2. The within-subject covariance matrix, Σ_j, is given the 'random coefficient' structure $X\Omega X' + \sigma^2 I$. In that way we can derive at a single equation model which can be interpreted in the same way as that of the multilevel model described above.

For the DENTAL data model, Σ_j is thus build up from four covariance terms: the variances of the intercept and the linear component, their covariance and the variance of the (remaining) within-subject error. The former three are the same elements that are contained in the second level variance-covariance matrix (Ω) of the multilevel model. The latter term is equal to the only element of which the first level variance-covariance matrix consist in the multilevel model (σ^2). Results of the analysis are presented in Table 1.

- Table 1 about here -

Refering to the model described above, we find the following entries in Table 1. First we have the fixed coefficients, the gammas; γ_{00} and γ_{10} denote the fixed intercept and slope of the mean growth curve for the whole group. Deviations from this mean intercept and slope for the girls' group and for the boys' group, are given by γ_{01} and γ_{11} respectively. Hence, the fixed intercept and slope of the mean curve for the girls are given by (γ_{00} - γ_{01}) and (γ_{10} - γ_{11}). Equivalently, the fixed intercept and slope of the mean curve for the boys are given by (γ_{00} + γ_{01}) and (γ_{10} + γ_{11}).

The random coefficients ω_{00} and ω_{11} denote the variances of the random (within-subject) intercept and the random (within-subject) slope, i.e. the variance not accounted for by (AGE) and (SEX). ω_{01} gives the covariance between both random coefficients. The variance of the first level disturbances is contained in σ^2. Note that regardless of the fact that the measurements are time related here, this random coefficient model does not include a special covariance structure for the error components.

Comparing solutions in Table 1, we see that the three programs produce similar, though not identical results. Estimates for fixed and random coefficients are the same for BMDP5V and ML3F. However, results from HLM differ from these to a certain extent. This holds for the estimates of the random parameters in the ML3R solution too. The major part of the differences in the solutions can be explained by the use of different estimation methods: HLM and ML3R use restricted maximum likelihood, whereas BMDP5V and ML3F use full information maximum likelihood. Differences between both likelihood procedures are strengthened in this

case because we use a small data set. Results from analyses on larger data sets reported in the sequel as well as results from Kreft, De Leeuw and Kim (1990) support this finding.

If we compare the number of required iterations, Table 1 clearly demonstrates the (well known) slow convergence of the EM algorithm employed by HLM, when parameter estimates come close to zero.

SCHOOL data

For the next illustration we use part of the results stemming from an educational study. De data are derived from a longtitudinal study concerning the individual development of technical reading (TR). Technical reading is measured as a score on a so-called "One-Minute Test". This test measures the speed at which a pupil can decode words. Scores of 436 pupils are used which are measured after respectively 1, 8, 11, 18 and 22 months, on 25 Dutch Primary Schools (for further details we refer to Guldenmund, 1991). Two explanatory variables will be used on the second level, the variable (SEX) and a score on a test measuring fonemic analysis ability, (FA). The first model to be analysed is a random coefficient model in which (TR) is regressed on (TIME) as a quadratic function. (SEX) is used as an explanatory variable on the second level. This model looks similar to the model which we used to analyse the DENTAL data. However, now we have a much larger sample and a quadratic instead of a linear function is fitted.

The within-subject model can be written as

$$(TR)_{ij} = \beta_{0j} + \beta_{1j} (TIME)_j + \beta_{2j} (TIME)_j^2 + e_{ij} ,$$

and the between-subject model is

$$\beta_{0j} = \gamma_{00} + \gamma_{01} (SEX)_j + u_{0j} ,$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} (SEX)_j + u_{1j} .$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21} (SEX)_j + u_{2j} \; .$$

A single equation version of this model looks like

$$(TR)_{ij} = \gamma_{00} + \gamma_{01}(SEX)_j + \gamma_{10}(TIME)_j + \gamma_{20}(TIME)_j^2 +$$
$$+ \gamma_{11}(SEX)_j(TIME)_j + \gamma_{21}(SEX)_j(TIME)_j^2 +$$
$$+ u_{0j} + u_{1j}(TIME)_j + u_{2j}(TIME)_j^2 + e_{ij} \; .$$

The analysis of these data were troubled by some program limitations. First, HLM appeared to run only if the independent variable (TIME) was centered, otherwise singularity problems occured. In order to make an as adequate comparison of the results we had to center the data using ML3 and BMDP5V too. In ML3 this requires some data manipulation, in BMDP5V however, this appeared to be impossible. It is caused by the fact that the program does not allow the input of noninteger values in its CONTRAST statement (see Section 3), necessary for the specification of centered variables[1].

Secondly, we were forced to analyse the data using only 300 cases. This number is the default maximum number of cases in HLM when individuals are treated as level 2 units (see also Section 3). Since we did not have immediate acces to the HLM Fortran source, we were not capable of extending this maximum.

The results of our analysis are shown in table 2.

- Table 2 about here -

The entries in Table 2 are interpreted in a similar way as in Table 1. Again, the two multilevel programs produce similar though not identical results. Comparing ML3F and ML3R, the estimates of the random parameters are relatively more alike than those in the first example analysis. This is explained by the fact that we use a much larger data set here.

The more general differences between HLM and ML3 can be attributed to their different estimation procedures. The slow convergence of the EM algorithm employed by HLM is shown

dramatically here.

The parameter estimates of the noncentered solution of BMDP5V are not quite comparable to the estimates of the other solutions. As might be expected, the largest difference in parameter value can be observed regarding the intercept term γ_{00}. Centering of the measurement times, i.e. the within-unit independent variable, can be computationally convenient, since it causes the sampling covariance of the ordinary least squares intercept and slope to be equal to zero. At the same time it changes the way in which the within-unit coefficients are interpreted. It is therefore probably the most informative to look at the deviance only. Comparing values of this statistic, we observe considerable differences among the programs. According to the results in both Table 1 and 2, it seems as if HLM overestimates the deviance. It is not quite clear to us if the differences in deviance value among all programs in general, and between BMDP5V and ML3(F and R) in particular, are acceptable.

Notice the negative variance estimates in the solutions of HLM and BMDP5V. Apparently, these parameters are at the boundary of the parameter space (equal to zero). Both programs do not provide a proper solution for this problem. ML3, however, does, by simply setting those estimates to zero.

In the next analysis, instead of (SEX) we use another explanatory variable, fonemic analysis (FA). Now, the within-subject model can be written as

$$(TR)_{ij} = \beta_{0j} + \beta_{1j} (TIME)_j + \beta_{2j} (TIME)_j^2 + e_{ij} \,.$$

and the between-subject model is

$$\beta_{0j} = \gamma_{00} + \gamma_{01} (FA)_j + u_{0j} \,,$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} (FA)_j + u_{1j} \,.$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21} (FA)_j + u_{2j} \,.$$

The single equation version of this model equals

$$(TR)_{ij} = \gamma_{00} + \gamma_{01}(FA)_j + \gamma_{10}(TIME)_j + \gamma_{20}(TIME)_j^2 +$$
$$+ \gamma_{11}(FA)_j(TIME)_j + \gamma_{21}(FA)_j(TIME)_j^2 +$$
$$+ u_{0j} + u_{1j}(TIME)_j + u_{2j}(TIME)_j^2 + e_{ij}.$$

Obviously, this model is similar to the model using (SEX) as a second level explanatory variable, only now we use (FA) for this purpose. However, besides the afore mentioned program limitations we now stumble upon what can be seen as a fundamental difference between the analysis of variance and regression analysis. This difference lies in the assumption of the measurement level which is allowed for the independent variable. While in the analysis of variance only independent variables of no more then nominal level are assumed, in regression analysis the independent variables are assumed to be of at least interval level. Since (FA) consists of a scale with scores ranging from 1 to 29, it can be considered as being measured on an interval scale. Hence, different results will occur if multilevel analysis considers the variable (FA) as numerical, whereas BMDP5V considers it a categorical variable. As a consequence, the only way this program can deal with such explanatory variable, is to code the categories by means of a set of 28 dummy variables. In other words, in BMDP5V this problem is considered as one in which experimental conditions or groups are involved: the scores on (FA) are conceived as a groupcode with 29 possible effects.

This problem is highlighted using BMDP5V versus ML3 only. Results are given in Table 3

- Table 3 about here -

In Table 3, the entries indicated by FA_k, FA_k* Lin and FA_k* Qua, each consists of 28 separate values denoting interactions of (FA) with the intercept, linear and quadratic component, respectively. Hence, the amount of parameters to be estimated in BMDP5V is much larger . A comparison with the parameter values of ML3 is therefore problematic.

Concerning ML3F and ML3R results, Table 3 shows that the differences between the random parameters are even less substantial here, with $N = 384$, as in the previous analysis.

So far, we have been comparing random coefficient models to analyse data on growth, using ML3, HLM and BMDP5V. As mentioned earlier one can expect growth data to be serially dependent. If this is the case, one can consider the modeling of a characteristic structure for the within-subject covariance matrix. In multilevel theory the aspect of serial dependency is not reflected in the modeling of more complex level-1 covariance structures. Usually the assumption of uncorrelated error components with equal variance is made. Although the manual describes the possibility to model complex level-1 covariance structures, in fact, it is not clear how this should be accomplished for serial dependent structures.

Contrary to both multilevel programs, BMDP5V explicitly provides several possibilities to model the within-subject covariance structure. A set of frequently encountered special structures are build into the program, including autoregressive structures (see also Section 3).

In the next analysis, an illustration is given of a model in which several within-covariance strucures are fitted: banded, or general autoregressive, first-order autoregressive, completely unconstrained (unstructured) and the already familiar random coefficient structure. Again the school data have been analyzed. The fitted model is identical to the first model described in this section, with (SEX) as a grouping variable and employing a quadratic growth curve. The scores of 384 pupils were used. The results are given in Table 4. Only the fixed terms of the model and the deviance are listed. Estimates of the random terms are omitted because there number varies across the different covariance structures.

- Table 4 about here -

If we assume that the two models applying autoregressive covariance structures, as well as that with the random coefficient structure, are 'nested' within the model with unconstrained covariance matrix, likelihood-ratio tests can be computed to decide which model is the most appropriate. Differences between relevant deviance values are chi-square distributed with

degrees of freedom equal to the difference between the number of independent parameters in the corresponding models. Hence, from Table 4 it can be concluded that the models applying autoregressive covariance structures are neither favorable over the random coefficient model (x^2 = 447.2, df = 5 for the first-order autoregressive model and x^2 = 18.9, df = 2 for the general autoregressive model), nor over the unconstrained model (x^2 = 589.8, df = 13 for the first-order autoregressive model and x^2 = 161.4, df = 10 for the general autoregressive model). In fact, Table 4 shows that the random coefficient model is less appropriate than the completely unconstrained model (x^2 = 142.5 df = 8).

Notes

[1] Personal communication with BMDP has indicated that there is indeed a prohibition against the input of other than integer parameters in the CONTRAST statement. However, BMDP suggested a so-called work-around approach based on the use of covariates to make use of centered variables. Moreover, it was stated that in the PC-90 version of the BMDP statistical software one has the possibility of introducing non-integer elements for the matrix Z (see ZCOL1 and ZCOL2 in the example on p. 19), and for the independent variables.

5. Conclusions

In this paper we have reviewed two approaches for the analysis of growth data by means of longitudinal mixed linear models. The first is a type of multilevel model in which growth data are treated as hierarchically structured. The second approach is a version of a MANOVA repeated measures model employing a structured (error)covariance matrix. From our theoretical and practical comparisons, it can be concluded that both ways of specifying 'random coefficient models' are fruitful for the analysis of growth data. Although they originate from different statistical and data-analytic traditions, the growth curve models they provide are basically alike. Nonetheless, both approaches show a few shortcomings which are not really on a theoretical level, but most strongly connected to the associated computer programs. Several of these limitations are listed below.

In multilevel modeling, and within the accompanying multilevel programs, one has optimal flexibility specifying the coefficients either as random or as fixed on any level. One has also much freedom in choosing which explanatory variable explains what variation on which level of the model. In BMDP5V however, either all or non of the (level 1) individual growth coefficients are specified as random. Coefficients on the second, between-subject, level are always fixed. Explanatory variables can therefore only operate on the second level to account for individual growth curve coefficient variation. Further, in BMDP5V there is no build in way to specify interaction between fixed variables at level 2 and random coefficients at level 1. Inflexibilities of this kind in BMDP5V are caused by the fact that the program is in fact a variance component model, decomposing the total variance into a between part and a within part. The between model uses fixed coefficients always. The random coefficient growth curve model emerges by modeling the within-subject error components with an appropriate covariance structure, which brings up the limitations mentioned above.

The smooth way in which the within-subject error covariances can be modeled in BMDP5V is a very useful feature analyzing time dependent data. Especially when the number of time points enlarges, the modeling of serial dependency becomes increasingly valuable. In BMDP5V this is easily accomplished: several relevant covariance structures are build into the program and ready

for use. However, the user has to be aware of the fact that applying those structures not neccessarily implies a random coefficient structure for the individual growth parameters. Reversely, the multilevel programs do not offer easily applicable ways in which the within-subject (level 1) covariances can be modeled. Although in theory, complex level 1 covariance structures can be formulated, it is unclear how this should be achieved in practice.

Centering of the within-unit explanatory variables brings up another point of difference. Although it changes the way in which the within-unit coefficients are interpreted, as was mentioned before, centering can be computationally convenient. Concerning the multilevel programs, in HLM centering is offered as an option, whereas in ML3 it is realized by simple data manipulation using NANOSTAT commands. However, HLM refused to run with uncentered variables, because of singularity problems. ML3 on the other hand, provided centered as well as uncentered solutions. In contrast, it seemed impossible to compute a centered solution in BMDP5V. For this purpose, the program requires additional input of non-integer values. Unfortunately, in the BMDP version currently available to us, this option was not available. Meanwhile we now from personal communication with BMDP that this feature is implemented in the PC-90 version of the BMDP software. As an alternative a 'work-around' approach using covariates was suggested but we were not able to further elaborate on this topic. Another major point of difference is the handling of numerical, explanatory variables in the between-subject model. In HLM and ML3, numerical variables at this second level are treated as predictor variables measured at an interval level, just as is the case in common regression analysis. BMDP5V, on the other hand, is developed as a general version of the analysis of variance model. Hence, variables acting in the between-subject model are considered as categorical 'grouping' variables. The categories of these variables are coded by a set of dummy variables, as if they are indicating separate experimental conditions. It means that for each dummy variable, or separate group code, an additional model parameter is created. Hence, the variance that can be explained by the model is decomposed into single degree of freedom terms. Therefore, in those cases, the results of BMDP5V versus those of the multilevel programs are somewhat difficult to compare, although the amount of variance accounted for by the model, should be identical for both approaches.

As far as parameter estimates are concerned, we have observed that the three programs produce similar, though not identical results. Differences are mainly caused by the different estimation procedures employed. In HLM and BMDP5V variance estimates may become slightly negative in case these parameters are at the boundary of the parameter space (equal to or close to zero). ML3 simply sets these estimates to zero. The programs show a large difference between EM and the other implemented estimation procedures, in the number of iterations required for convergence. It is clearly illustrated that when parameter estimates come close to zero, the EM algorithm is slowly converging.

Minor points of concern include the following. When growth data are analyzed in HLM, the number of cases is limited to 300. This restriction is caused by the fact that the number of *groups* is limited analyzing growth data. Groups refer to the second level, so in a usual hierarchically organized data set, where individuals are considered level 1 units, this is not a problem in particular. In practice, we will seldom want to make use of more than 300 groups. However, when growth data are treated as hierarchical, measurements are the level 1 units and individuals become the second level. Hence, we have a restriction on the number of cases. It may be overcome by changing the defaults in the HLM source code and having it recompiled. As far as we know one has access to this source code in the non-PC versions of HLM only.
In ML3 it was unclear at first how output of the deviance statistic could be accomplished. It appeared that one has to use a 'mysterious' NANOSTAT command, which is not documented in the manual. This NANOSTAT package at the same time, makes the program a very flexible tool for data analysis.

6. References

Bock, R.D. (Ed.). (1989). *Multilevel analysis of educational data*. San Diego, CA: Academic Press.

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

Bryk, A.S., and Raudenbush, S.W. (1989). Towards a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model. In: R.D. Bock (Ed.), *Multilevel Analysis of Educational Data* (pp. 159-204). San Diego (CA): Academic Press.

Bryk, A.S., Raudenbush, S.W., Seltzer, M., and Congdon, R.T. (1988). *An Introduction to HLM: Computer Program and User's Guide*. University of Chicago.

Demster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B, 39*, 1-38.

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

De Lury, D.B. (1950). *Values and integrals of the orthogonal polynimials up to n = 26*. Toronto: University of Toronto Press.

Dixon, W.J. (Ed.) (1988). *BMDP Statistical software manual, Vol. 2*. Berkeley, CA: University of California Press.

Geisser, S. (1970). A Bayesian analysis of growth curves. *Sankhyā A, 32*, 53-64.

Geisser, S. (1980). Growth curve analysis. In: P.R. Krishnaiah (Ed.), *Handbook of statistics, Vol. 1* (pp. 89-115). Amsterdam: North Holland.

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

Goldstein, H. (1986b). Efficient statistical modelling of longitudinal data. *Annals of Human Biology, 13*, 129-141.

Goldstein, H. (1987). *Multilevel models in educational and social research*. London: Griffin.

Goldstein, H. (1989). Models for multilevel response variables with an application to growth curves. In: D. Bock (Ed.), *Multilevel Analysis of Educational Data* (pp. 107-125). San Diego (CA): Academic Press.

Goldstein, H., and Rasbash, J. (1989). *Efficient computational procedures for the estimation of parameters in multilevel models, based on iterative generalized least squares*. Manuscript submitted for publication.

Grizzle, J.E., and Allen, D.M. (1969). Analysis of growth and dose response curves. *Biometrika, 25*, 357-382.

Guldenmund, F.W. (1991). *Single and two step approaches to modeling individual psychological growth*. Unpublished MA thesis. Department of Psychometrics and Research Methodology, University of Leiden.

Healy, M. (1989). *Nanostat users' guide*. London: AlphaBridge Ltd.

Jennrich, R.I., and Schluchter, M.D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics, 42*, 805-820.

Kendall, M.G., and Stuart, A. (1979). *The advanced theory of statistics, Vol. 2, Inference and relationship*, 4th ed. London: Griffin.

Khatri, O.G. (1966). A note on a MANOVA model applied to problems in growth curve. *Annals of the Institute of Statistical Mathematics, 18,* 75-86.

Kirk, R.E. (1982). *Experimental design: Procedures for the behavioral sciences,* 2nd ed. Monterey, CA: Brooks/Cole.

Kreft, I.G.G., De Leeuw, J., and Kim, K-S. (1990). *Comparing four different statistical packages for hierarchical linear regression VARCL, HLM, GENMOD, and ML2*. UCLA Statistics Series #50.

Kreft, I.G.G., De Leeuw, J., and Van der Leeden, R. (1991). Review of five multilevel analysis programs for the IBM PC. *The American Statistician, in press.*

Laird, N.M., Lange, N., and Strom, D. (1987). Maximum likelihood estimation with repeated measures: applications to the EM algorithm. *Journal of the American Statistical Association, 82*, 97-105.

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

Lee, J., and Geisser, S. (1972). Growth curve prediction. *Sankhyā A, 34*, 393-412.

Lee, J., and Geisser, S. (1975). Applications of growth curve prediction. *Sankhyā A, 37*, 239-256.

Little, R.J.A., and Rubin, D.B. (1987). *Statistical analysis with missing data.* New York: John Wiley.

Potthoff, R.F., and Roy, S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems, *Biometrika, 51,* 313-326.

Prosser, R., Rasbash, J., and Goldstein, H. (1990). *ML3 Software for Three-level Analysis User's Guide.* Institute of Education, University of London.

Rao, C.R. (1959). Some problems involving linear hypotheses in multivariate analysis. *Biometrika, 46,* 49-58.

Rao, C.R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika, 52,* 447-458.

Rao, C.R. (1966). Covariance adjustment and related problems in multivariate analysis. In P.R. Krishnaiah (Ed.), *Multivariate analysis II*, pp.87-103. New York: Academic Press.

Rao, C.R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics, 1,* 355-372.

Rao, C.R. (1973). *Linear statistical inference and its applications,* 2nd ed. New York: John Wiley.

Raudenbush, S.W. (1988). Educational applications of hierarchical linear models: A review. *Journal of Educational Statistics, 13,* 85-118.

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

Schluchter, M.D. (1988). Unbalanced repeated measures models with structured covariance matrices. In: W.J. Dixon (Ed.), *BMDP Statistical software manual, Vol. 2* (pp. 1081-1114). Berkeley, CA: University of Clifornia Press.

Timm, N.H. (1975). *Multivariate analysis with applications in education and psychology.* Belmont, CA: Wadsworth.

Timm, N.H. (1980). Multivariate analysis of variance of repeated measurements. In: P.R. Krishnaiah (Ed.), *Handbook of statistics, Vol. 1* (pp.41-87). Amsterdam: North Holland.

Van der Leeden, R. (1990). *Reduced rank regression with structured residuals.* Leiden: DSWO Press.

Visser, R.A. (1982). *On quantitative longitudinal data in psychological research.* PhD dissertation: University of Leiden. Published in 1985 as: Analysis of longitudinal data in behavioural and social research. Leiden: DSWO Press.

Vrijburg, K.E. (1991). *Het modelleren van groeidata* [The modeling of growth data]. Unpublished MA thesis. Department of Psychometrics and Research Methodology, University of Leiden.

Ware, J.H. (1985). Linear models for the analysis of longitudinal studies. *Journal of the American Statistical Association, 80,* 95-101.

Table 1. Results of the analyses of the DENTAL data, N = 27.

	BMDP5V	HLM	ML3F	ML3R
γ_{00}	16.8567	16.8993	16.8570	16.8570
γ_{10}	0.6320	0.6287	0.6320	0.6320
γ_{01}	-0.5161	-0.3824	-0.5161	-0.5161
γ_{11}	0.1524	0.1419	0.1524	0.1524
σ^2	1.7162	1.7651	1.7162	1.7162
ω_{00}	4.5569	5.7509	4.5569	5.7861
ω_{10}	-0.1983	-0.3082	-0.1983	-0.2896
ω_{11}	0.0238	0.0356	0.0238	0.0325
# iterations	2	60	2	4
deviance	427.80	431.73	427.81	427.96

Table 2. Results of the analyses of the SCHOOL data, N = 300, explanatory variable is (SEX), measurement times are centered, except for BMDP5V.

	BMDP5V	HLM	ML3F	ML3R
γ_{00}	45.9339	60.2410	60.2800	60.2800
γ_{10}	0.7522	0.8009	0.7522	0.7522
γ_{20}	0.0288	0.0258	0.0288	0.0288
γ_{01}	1.6653	1.4756	1.4580	1.4580
γ_{11}	0.0187	0.0291	0.0187	0.0187
γ_{21}	-0.0023	-0.0025	-0.0023	-0.0023
σ^2	37.2859	36.9274	31.6500	31.7300
ω_{00}	158.9831	179.5018	182.0000	183.2000
ω_{10}	7.6650	2.4385	5.4620	5.4710
ω_{11}	-1.0276	0.0520	0.0000	0.0000
ω_{20}	-0.3977	-0.1551	-0.2824	-0.2831
ω_{21}	0.0426	-0.0019	0.0016	0.0016
ω_{22}	-0.0017	0.0002	0.0000	0.0000
# iterations	3	158	6	6
deviance	10511.88	10574.60	10549.30	10549.30

Table 3. Results of the analyses of the SCHOOL data, N = 384, explanatory variable is (FA), measurement times not centered.

	ML3F	ML3R	BMDP5V
γ_{00}	38.0600	38.0600	44.8670
γ_{10}	0.9890	0.9890	0.7067
γ_{20}	0.0247	0.0247	0.0320
γ_{01}	0.4459	0.4459	FA_k
γ_{11}	-0.0147	-0.0147	FA_k* Lin
γ_{21}	0.0003	0.0003	FA_k* Qua
σ^2	32.1200	32.1800	38.1900
ω_{00}	155.1000	156.1000	130.0900
ω_{10}	4.5250	4.5280	5.8400
ω_{11}	0.0000	0.0000	-0.8200
ω_{20}	-0.2742	-0.2749	-0.0300
ω_{21}	0.0015	0.0015	0.0300
ω_{22}	0.0000	0.0000	-0.0010
# iterations	5	5	4
deviance	13547.10	13547.1	13327.84

Table 4. Results of the analyses of the SCHOOL data using BMDP5V, N = 384, explanatory variable is (SEX), different structures for $\hat{\Sigma}$.

$\hat{\Sigma}$ structure	$\hat{\Sigma}$ unconstrained	$\hat{\Sigma}=T_{F_n}T_F'+\sigma^2I$	$\sigma_{ij}=\varsigma_k, k=|i\text{-}j|+1$	$\sigma_{ij}=\sigma^2 \circledR^{|i\text{-}j|}$
γ_{00}	49.5493	46.2603	44.7326	48.1164
γ_{10}	0.8338	0.6877	1.0413	0.5681
γ_{20}	0.0163	0.0314	0.0152	0.0304
γ_{01}	-1.0327	-0.9486	-1.0553	-0.9652
γ_{11}	-0.0308	-0.0455	-0.0196	-0.0349
γ_{21}	0.0022	0.0027	0.0017	0.0021
# parameters	15	7	5	2
deviance	13270.098	13412.618	13431.496	13859.856