Multilevel Non-Linear Models

Introduction to Multilevel Modeling

The analysis of data with a hierarchical structure has been described in the literature under various names. It is known as hierarchical modeling, random coefficient modeling, latent curve modeling, growth curve modeling or multilevel modeling. The basic underlying structure of measurements nested within units at a higher level of the hierarchy is, however, common to all. In a repeated measurements growth model, for example, the measurements or outcomes are nested within the experimental units (second level units) of the hierarchy.

Ignoring the hierarchical structure of data can have serious implications, as the use of alternatives such as aggregation and disaggregation of information to another level can induce high collinearity among predictors and large or biased standard errors for the estimates. Standard fixed parameter regression models do not allow for the exploration of variation between groups, which may be of interest in its own right. For a discussion of the effects of these alternatives, see Bryk and Raudenbush (1992), Longford (1987) and Rasbash (1993).

Multilevel or hierarchical modeling provides the opportunity to study variation at different levels of the hierarchy. Such a model can also include separate regression coefficients at different levels of the hierarchy that have no meaning without recognition of the hierarchical structure of the population. The dependence of repeated measurements belonging to one experimental unit in a typical growth curve analysis, for example, is taken into account with this approach. In addition, the data to be analyzed need not be balanced in nature. This has the advantage that estimates can also be units for which a very limited amount of information is available.

Multilevel Non-Linear Regression Models

It was pointed out by Pinheiro and Bates (2000) that one would want to use nonlinear latent coefficient models for reasons of *interpretability*, *parsimony*, and more importantly, *validity* beyond the observed range of the data.

By increasing the order of a polynomial model, one can get increasingly accurate approximations to the true, usually nonlinear, regression function, within the range of the observed data. High order polynomial models often result in multicollinearity problems and provide no theoretical considerations about the underlying mechanism producing the data.

There are many possible nonlinear regression models to select from. Examples are given by Gallant (1987) and Pinheiro and Bates (2000). The Richards function (Richards, 1959) is a generalization of a family of non-linear functions and is used to describe growth curves (Koops, 1988). Three special cases of the Richards function are the logistic, Gompertz and Monomolecular functions, respectively. Presently, one can select curves of the form

$$y = f_1(x) + f_2(x) + e$$

where the first component, $f_1(x)$ may have the form:

- logistic: $\frac{b_1}{(1 + s\exp(b_2 - b_3 x)}$
- Gompertz: $b_1 \exp(-b_2 \exp(-b_3 x))$
- Monomolecular: $b_1(1 + s\exp(b_2 - b_3 x))$
- power: $b_1 x^{b_2}$
- exponential: $b_1 \exp(-b_2 x)$

The second component, $f_2(x)$, may have the form:

- logistic: $\frac{c_1}{(1 + s\exp(c_2 - c_3 x)}$
- Gompertz: $c_1 \exp(-c_2 \exp(-c_3 x))$
- Monomolecular: $c_1(1 + s\exp(c_2 - c_3 x))$
- power: $c_1 x^{c_2}$
- exponential: $c_1 \exp(-c_2 x)$

In the curves above *s* denotes the sign of the term $\exp(b_2 - b_3 x)$ and is equal to 1 or -1.

Since the parameters in the first three functions above have definite physical meanings, a curve from this family is preferred to a polynomial curve, which may often be fitted to a set of responses with the same degree of accuracy. The parameter b_1 represents the time asymptotic value of the characteristic that has been measured, the parameter b_2 represents the potential increase (or decrease) in the value of the function during the course of time t_1 to t_p, and the parameter b_3 characterizes the rate of growth.

The coefficients $b_1, b_2, \ldots, c_3$ are assumed to be random, and, as in linear hierarchical models, can be written as level-2 outcome variables where

$$
\begin{aligned}
b_1 &= \beta_1 + u_1 \\
b_2 &= \beta_2 + u_2 \\
b_3 &= \beta_3 + u_3 \\
c_1 &= \beta_4 + u_4 \\
c_2 &= \beta_5 + u_5 \\
c_3 &= \beta_6 + u_6.
\end{aligned}
$$

It is assumed that the level-2 residuals $u_1, u_2, \ldots, u_6$ have a normal distribution with zero means and covariance matrix $\mathbf{\Phi}$. In LISREL, it may further be assumed that the values of any of the random coefficients are affected by some level-2 covariate so that, in general,

$$
\begin{aligned}
b_1 &= \beta_1 + \gamma_1 z_1 + u_1 \\
b_2 &= \beta_2 + \gamma_2 z_2 + u_2 \\
b_3 &= \beta_3 + \gamma_3 z_3 + u_3 \\
c_1 &= \beta_4 + \gamma_4 z_4 + u_4 \\
c_2 &= \beta_5 + \gamma_5 z_5 + u_5 \\
c_3 &= \beta_6 + \gamma_6 z_6 + u_6.
\end{aligned}
$$

where z_i denotes the value of a covariate and γ_i the corresponding coefficient.

It is usually sufficient to select a single component $(f_1(x))$ to describe a large number of monotone increasing (or decreasing) growth patterns.

To describe more complex patterns, use can be made of two-component regression models. LISREL allows the user to select any of the 5 curve types as component 1 and to combine it with any one of the 5 curve types for the second component.

Valid choices are, for example,

- Monomolecular + Gompertz
- logistic
- exponential + logistic
- logistic + logistic

The unknown model parameters are the vector of fixed coefficients $(\boldsymbol{\beta})$, the vector of covariate coefficients $(\boldsymbol{\gamma})$, the covariance matrix $(\mathbf{\Phi})$ of the level-2 residuals and the variance (σ^2) of the level-1 measurement errors. See Chapter 5 for an example of fitting of a multilevel nonlinear model. Additional examples are given in the **nonlinex** folder.

Estimation Procedure for Multilevel Non-Linear Regression Models

In linear multilevel models, *y* has a normal distribution, since *y* is a linear combination of the random coefficients. For example, the intercept-and-slopes-as-outcomes model is

$$y = b_1 + b_2 x + e$$

where $b_1 = u_1$, $b_2 = u_2$ and (u_1, u_2) is assumed to be normally distributed.

A multilevel nonlinear model is a regression model which cannot be expressed as a linear combination of its coefficients and therefore *y* is no longer normally distributed. The probability density function of *y* can be evaluated as the multiple integral

$$f(y) = \int_{b_1} \dots \int_{c_3} f(y, b_1, \dots, c_3)\, db_1 \dots dc_3$$

that, in general, cannot be solved in closed form.

To evaluate the likelihood function

$$L = \prod_{i=1}^{n} f(y_i),$$

one has to use a numerical integration technique. We assume that *e* has a $N(0, \sigma^2)$ distribution and that $(b_1, b_2, b_3, c_1, c_2, c_3)$ has a $N(\mathbf{0}, \mathbf{\Phi}^2)$ distribution.

In the multilevels procedure, use is made of a Gauss quadrature procedure to evaluate the integrals numerically. The ML method requires good starting values for the unknown parameters. The estimation procedure is described by Cudeck and du Toit (in press).

Starting values

Once a model is selected to describe the nonlinear pattern in the data, for example as revealed by a plot of *y* on *x*, a curve is fitted to each individual using ordinary non-linear least squares.

In step 1 of the fitting procedure, these OLS parameter estimates are written to a file and estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\Phi}$ are obtained by using the sample means and covariances of the set of fitted parameters. Since observed values from some individual cases may not be adequately described by the selected model, these cases can have excessively large residuals, and it may not be advisable to include them in the calculation of the $\boldsymbol{\beta}$ and $\boldsymbol{\Phi}$ estimates.

In step 2 of the model fitting procedure, use is made of the MAP (Maximum Aposterior) estimator of the unknown parameters.

Suppose that a single component $f_1(b, x)$ is fitted to the data. Since

$$f(b \mid y) = f(b) f(y \mid b) / f(y),$$

where $f(b \mid y)$ is the conditional probability density function of the random coefficients *b* given the observations, it follows that

$$\ln f(b \mid y) = \ln f(b) + \ln f(y \mid b) + k,$$

where $k = -\ln f(y)$.

The MAP procedure can briefly be described as follows:

Step a:

Given starting values of $\boldsymbol{\beta}$, $\boldsymbol{\Phi}$ and σ^2, obtain estimates of the random coefficients $\hat{b}_i$ from

$$\frac{\partial}{\partial b_i} \ln f(b_i \mid y_i) = 0, \quad i = 1, 2, \ldots, n.$$

Step b:

Use the estimates $\hat{b}_1, \hat{b}_2, \ldots, \hat{b}_n$ and $Cov(\hat{b}_1), \ldots, Cov(\hat{b}_n)$ to obtain new estimates of $\boldsymbol{\beta}$, $\boldsymbol{\Phi}$ and σ^2 (see Herbst, A. (1993) for a detailed discussion).

Repeat steps a and b until convergence is attained.

For many practical purposes, results obtained from the MAP procedure may be sufficient. However, if covariates are included in the model, parameter estimates are only available via the ML option, which uses the MAP estimates of $\boldsymbol{\beta}$, $\boldsymbol{\Phi}$ and σ^2 as starting values.

References

Bryk, A.S., and Raudenbush, S.W. (1992). *Hierarchical Linear Models*. Sage: Newbury Park, CA.

Gallant, A.R. (1987). *Nonlinear Statistical Models*. Wiley: New York.

Herbst, A. (1993). *The Statistical Modelling of Growth*. Unpublished Ph.D. dissertation, University of Pretoria: Pretoria.

Koops, W.J. (1988). Multiphasic Analysis of Growth Curves in Chickens. *Poultry Science*, **67**, 33-42.

Longford, N.T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested effects. *Biometrika*, **741(4)**, 817-827.

Pinheiro, J.C., and Bates, D.M. (2000). *Mixed-Effects Models in S and S-Plus*. Springer-Verlag: New York.

Rasbash, J. (1993). *ML3E Version 2.3 Manual Supplement*. University of London: Institute of Education.

Richards, F.J. (1959). A flexible growth function for empirical use. *Journal of Experimental Botany*, **10**, 29(290-300).