Determining the Meaning of Parameters in Multilevel Models for Longitudinal Data

Christof Schuster

University of Michigan, USA

Alexander von Eye

Michigan State University, USA

This paper is concerned with parameter interpretation in longitudinal, multilevel models. Models are described that consider repeated observations nested within individuals. These models typically first estimate subject-specific parameters for growth curves that describe the development of some observed variable over time. Examples of such descriptors include polynomials. It is shown that interpretation of polynomial parameters can be facilitated by linear transformations. Examples of such transformations include centring (i.e. subtracting the mean from raw data). When parameters are specified such that they have no straightforward meaning at the first level of analysis, interpretation problems carry over to the second and higher levels. Therefore, proper specification of models at the first level is of utmost importance. Methods of transformation are introduced. Examples illustrate the method using data that describe children's vocabulary development in the second year of life.

Multilevel models (Goldstein, 1986, 1987) have become a prominent methodological approach in psychology and other social sciences. These models are suitable for analysing data that were sampled in several stages. A simple example for a two-stage sampling scheme is: In the first stage take a random sample of schools and in the second stage take one random sample of pupils from every school sampled in the first stage. Even if data

---

are obtained by simple random sampling, it might be useful to think of them in terms of a hierarchical structure and therefore use a multilevel approach to model them. Additionally, data obtained by repeatedly measuring individuals as it is typically done in, for instance, developmental psychology can also be considered as possessing a hierarchical structure. In context of analysing panel data multilevel models are also known as growth curve models. In cross-sectional data analysis they are also known as hierarchical linear models (Bryk & Raudenbush, 1987, 1992), random effects models (Laird & Ware, 1982; Ware, 1985) as well as random coefficient models (Longford, 1993).

When fitting multilevel models to data, it is important to check whether model assumptions are at least approximately satisfied. However, compared to assumption checking for other types of statistical analysis, checking the assumptions for multilevel models is more difficult because there is not only one model under consideration, but two or more models at different levels. Adequacy of models at higher levels depends on the meaning of the parameters at the lower level. Therefore, when using multilevel models it is important that the lower level model is parameterised such that level-two models have a good substantive interpretation. If this is the case, the formulation of higher level models is facilitated because the model building process focuses directly on the research hypotheses under study. In contrast, poor parameterisation at the lower level model can result in considerable problems in finding an adequate higher level model, for one is trying to model a variable that is not related to substantive hypotheses.

In this paper we explain how a transformation of the model matrix for the lower level model can be used to determine the meaning of the parameters in such a way that they correspond to substantively relevant parameters for which a higher level model is sought. The approach presented is limited to linear transformations. Although this restricts the set of possible meanings a parameter can be assigned by our method we believe it is still general enough to answer many practically important research hypotheses.

In section 1 we review the basic ideas of multilevel models and explain how growth curve models can be considered multilevel models. Section 2 shows how the interpretation of the parameters can be derived from the model formulation. In the following section (3) it is shown how the very common transformation of subtracting a constant from the values of the predictor variables can be formulated using matrices. Section 4 reverses the previously given line of arguments and demonstrates by an example how the reparameterisation can be achieved once the desired meaning of the parameter is defined. In the final section (5), the suggested approach is discussed.

1. Growth Curve Models and Multilevel Models

The most basic idea of multilevel modelling is to use a model for describing the data that accounts for the hierarchical structure of the sampling plan. This can be done by specifying separate models for each level. Because data obtained from a multistage sampling procedure usually result in correlation between observations of level-one units within the same level-two units, multilevel models are natural candidates for modelling this type of data.

However, data from a panel study can also be considered a result of a two stage sampling process. In the first stage individuals are sampled and in the second stage each of these individuals is repeatedly observed. Here, responses are said to be nested within individuals. Note that whereas observations within an individual are assumed to be correlated, observations obtained from different individuals are considered independent. Again, multilevel models should be well suited to describe this kind of data.

The model for the first stage, that is, for observations within the ith individual at the jth point in time, is obtained by expressing the observations y_{ij} as a function of time t_{ij} as well as a function of other time-varying covariates x_{ij} involving parameter θ_i, plus a residual term, ε_{ij},

$$y_{ij} = f(t_{ij}, x_{ij}, \theta_i) + \varepsilon_{ij}$$

where the conditional expectation of the residual term is assumed to be zero, that is, $E(\varepsilon_{ij}|x_{ij}, t_{ij}) = 0$.

Although this model equation is too general to be practically useful, it shows that the dependence of the observations y_{ij} on time and time-varying covariates is allowed to vary across individuals. This is indicated by the subscript i assigned to the parameter θ. Usually a linear function in the parameters is chosen for f.

For concreteness and to illustrate the use of nonlinear functions we write the foregoing equation assuming a quadratic model for the dependence of the observations on time and omitting all time-varying covariates. This yields

$$y_{ij} = \alpha_{i0} + \alpha_{i1}t_{ij} + \alpha_{i2}t_{ij}^2 + \varepsilon_{ij}, \tag{1}$$

as the level-one model, where $\theta_i = (\alpha_{i0}, \alpha_{i1}, \alpha_{i2})$. Note that the expected value of y_{ij} is given as $E(y_{ij}) = \alpha_{i0} + \alpha_{i1}t_{ij} + \alpha_{i2}t_{ij}^2$.

In the second level each regression coefficient that is assumed to vary across individuals, becomes the dependent variable of a linear model. For the quadratic level-one model from above we could write, for instance,

$$\alpha_{i0} = \beta_{00} + \beta_{01}z + \zeta_{i0},$$
$$\alpha_{i1} = \beta_{10} + \beta_{11}z + \zeta_{i1},$$
$$\alpha_{i2} = \beta_{20} + \beta_{21}z + \zeta_{i2}.$$

Although the β-coefficients are typically considered fixed in a two stage model each of the regression equations contains a residual term ζ that is considered random with zero mean. Hence, the α-regression coefficients of the first level are random as well. The observed independent variable z models the dependence of the α-parameters on time-invariant characteristics of the individual. The β parameters can be interpreted in the following way. If z is a dichotomous variable such as gender that is dummy-coded with values, 0 for males and 1 for females, then β_{10} can be interpreted as the mean of the α_{i1}-coefficients for males, and β_{11} indicates the mean shift in the α_{i1}-coefficients between males and females. Similar results hold for the other β parameters.

It is often the main goal of a multilevel analysis to obtain good estimates of the level-two coefficients as they contain the information that relates the covariates, for example, z, to the α-coefficients that are part of the systematic component $E(y_{ij})$ of individual responses. The main purpose of the level-one model is to allow for correlations among the repeated observations within individuals.

Obviously, a correct specification of the level-one model is crucial for the estimates of the level-two model to give useful and valid results. Suppose that an important variable has been omitted from the level-one model. If that variable is correlated with other predictor variables already in the equation it is well known that parameter estimates will be biased (Box, 1966; Miller, 1990). Hence, the level-two model will try to predict a biased estimator. If the bias is considerable then the results of the level-two model that tries to predict this random coefficient will not be interesting as they are also biased (see Bryk & Raudenbush, 1992, p. 203).

2. Interpretation of Level-One Parameters

Assuming a correct specification of a level-one model the ease with which a correct and meaningful formulation for the level-two model can be found depends on the meaning of the level-one parameters. To see how the meaning of parameters can be changed it is important to understand how the interpretation of parameters can be obtained from the model specification.

In general, the meaning of a parameter in a statistical model can be found by taking a model equation that contains the parameter and then try to solve the equation for that parameter in such a way that no other unknown parameters are contained in the resulting equation. Once this is

achieved the interpretation of the parameter can be found from this equation. We will now demonstrate this for the level-one model in equation 1 that describes the response variable as a quadratic polynomial in time by finding the interpretation of all the three parameters, α_{i0}, α_{i1}, and α_{i2}, contained in that equation.

Finding the interpretation of α_{i0} is straightforward. We simply insert zero as the value for t. This yields.

$$E(y(0)) = \alpha_{i0}.$$

Hence, α_{i0} is the y-value that the quadratic polynomial assumes when t equals zero. If $t = 0$ describes the point in time where the observation of y started, α_{i0} can sometimes be considered a baseline. This may be a substantively important parameter for some studies, for others it may be a meaningless value. For instance, if t is set equal to *age* of individuals and $y(t)$ represents IQ-values over time, then α_{i0} has the meaning the expected IQ value for that individual at birth.

Now, let us look at the meaning of α_{i1}. Because $E(y)$ is a smooth function of time, we can find its first derivative with respect to t. This yields

$$\frac{d}{dt}E(y(t)) = \alpha_{i1} + 2\alpha_{i2}t.$$

As this expression still contains two different α parameters we first have to eliminate α_{i2} from the equation to find the meaning of α_{i1}. This can be done by inserting zero for t. We obtain

$$\frac{d}{dt}E(y(t))\bigg|_{t=0} = \alpha_{i1}.$$

As the first derivative of $E(y)$ evaluated at $t = 0$ expresses how fast the function y changes at that point in time, the parameter α_{i1} can be interpreted as the instantaneous rate of change for the time point that is defined for $t = 0$. As with the interpretation of α_{i0} the substantive interest of this parameter may crucially depend on how the time variable is scaled. Thus, it may be the case that researchers have no specific hypothesis concerning an instantaneous change rate and it may therefore be difficult to express a level-two model for that parameter.

Finally, let us see what meaning can be given to α_{i2}. Calculating the second derivative of equation 1 we obtain

$$\frac{d^2}{dt^2}E(y(t)) = 2\alpha_{i2}.$$

Therefore, α_{i2} can be interpreted as half of the second derivative of $E(y)$ with respect to t. The second derivative of a function is often interpreted as

the "acceleration". As the right-hand side of the last equation does not involve the variable t the "acceleration" is constant over the support of t.

Although this explanation concerning α_{i2} may be satisfactory from a mathematical point of view, researchers rarely have substantive hypotheses concerning the "acceleration" of a time-dependent process $y(t)$. Therefore, specifying a level-two model for this parameter is often hard, if not impossible. Perhaps the best that can be done is to not specify a model for that parameter at all and consider a model where α_{i2} is just allowed to vary randomly without specifying any model on how the mean of this random variation depends on covariates, that is, to use a model as

$$\alpha_{i2} = \beta_{20} + \zeta_{i2}.$$

However, this may not be satisfactory because the possibility to model variation in the outcome variable under study is not used.

3. A Simple Example for Determining the Meaning of Parameters

Perhaps the best known change in interpretation of a parameter that is achieved by manipulating the values of the predictor variables is the change in the meaning of the intercept parameter, α_{i0}, that results from subtracting a constant c from the predictor variable. If c equals the mean of t_{ij} one also says that subtracting c from each t_{ij} centres the predictor t. To see how the meaning of the parameters changes consider the quadratic growth curve model

$$y_{ij} = \alpha_{i0} + \alpha_{i1}(t_{ij} - c) + \alpha_{i2}(t_{ij} - c)^2 + \varepsilon_{ij}.$$

Note that we have marked every parameter with a "tilde" as the manipulation of the predictor has changed the meaning of at least some of the parameters. In order to obtain an equation that contains just the intercept parameter α_{i0} from which its meaning can be inferred we insert the value c for t_{ij} instead of zero. Therefore, α_{i0} has now the meaning of the y-value that is expected to be observed if the observation time takes the value c.

Note also, that the interpretation of the parameter α_{i1} changes as well, whereas the meaning of the parameter α_{i2} remains the same. The parameter α_{i1} expresses now the instantaneous change rate if the time of observation takes the value c.

The subtraction of a constant value does not change the most important characteristics of the model under study such as for instance, the fitted curve and, therefore, the fitted values, the residuals, the explained variance in terms of R^2, and so forth. However, the intercorrelation of the estimated parameters might change. This will only be of interest if the subtraction of

a constant value induces considerable multicollinearity among the predictors of the model matrix. However, in contrast, multicollinearity can often be considerably reduced by this operation. For instance, it is well known that centring typically decreases the correlation between the columns of the model matrix in polynomial regression (von Eye & Schuster, 1998).

It is instructive to see how the transformation just described proceeds, using a more general matrix formulation. It is sufficient to consider the level-one model for a single observation of an individual. Let $t'_{ij} = (1, t_{ij}, t^2_{ij})$ be the jth row of the model matrix T_i. The vector of individual regression coefficients is $\alpha'_i = (\alpha_{i0}, \alpha_{i1}, \alpha_{i2})$. The level-one model is given as

$$E(y_{ij}) = (1, t_{ij}, t^2_{ij}) \begin{pmatrix} \alpha_{i0} \\ \alpha_{i1} \\ \alpha_{i2} \end{pmatrix},$$

or expressed in shorthand notation

$$E(y_{ij}) = t'_{ij}\alpha_i. \tag{2}$$

Now, suppose we have a square matrix Q of full rank that transforms t'_{ij} to $t'_{ij}Q = \tilde{\tau}'_{ij}$, where $\tilde{\tau}'_{ij} = (1, (t_{ij} - c), (t_{ij} - c)^2)$. Because by assumption the inverse of Q exists, we can express equation 2 as

$$\begin{aligned} E(y_{ij}) &= t'_{ij}Q Q^{-1}\alpha_i \\ &= \tilde{\tau}'_{ij}Q^{-1}\alpha_i \\ &= \tilde{\tau}'_{ij}\alpha_i. \end{aligned}$$

While $\tilde{\tau}'_{ij}$ contains the predictors from which the constant c is subtracted, the vector α_i contains parameters with changed meaning.

For the operation of subtracting a constant from the predictors the matrix Q is given as

$$Q = \begin{pmatrix} 1 & -c & c^2 \\ 0 & 1 & -2c \\ 0 & 0 & 1 \end{pmatrix}.$$

Simply calculating $t'_{ij}Q$ shows that Q yields the desired transformation from t'_{ij} to $\tilde{\tau}'_{ij}$. Note also, that Q is of full rank. At the moment, the expression $Q^{-1}\alpha_i$ is not of much interest. Here, it merely shows that the transformed parameter vector α_i relates to the original α_i through a linear transformation. However, for completeness we also present Q^{-1}

$$Q^{-1} = \begin{pmatrix} 1 & c & c^2 \\ 0 & 1 & 2c \\ 0 & 0 & 1 \end{pmatrix}.$$

From this matrix it can be seen that although the meaning of α_{i0} and α_{i1} changes as a result of the subtraction of the constant c from the predictors the meaning of α_{i2} is unchanged because of the simple form taken by the third row of Q^{-1}.

Although the preceding mathematics were not necessary to derive the meaning of the new parameters it shows that the change in their meaning results from a transformation of the model matrix T_i to a new model matrix $\tilde{T}_i$. This transformation is defined by the matrix Q. Hence, transforming parameters to a new substantive more meaningful interpretation translates into the technical question of how to set up the matrix Q or equivalently, how to set up Q^{-1}.

4. A More Complex Example of Determining the Meaning of Parameters

In the following example we reverse the line of argumentation. In the last section we determined for a given transformation the new meaning of the parameters. Technically this corresponds to specifying Q. In this example we first define the meaning the parameters are supposed to have and then determine the appropriate transformation. Technically this corresponds to specifying Q^{-1}. Once this matrix is found, calculating its inverse yields Q, and after determining $\tilde{T}_i = T_i Q$ the model can be fitted as usual and the parameters have the desired interpretation.

Suppose data from a panel study are going to be fitted by a two-level model. Assume further that after examining the data it turns out that a quadratic polynomial provides a good description of the individual trajectories over time.

Also assume that—perhaps after subtracting a suitable constant from the predictors—the parameters α_{i0} and α_{i1} have a substantively interesting interpretation and therefore specifying a level-two model for these two parameters is relatively easy. However, the parameter α_{i2} might not correspond to a meaningful research hypothesis. From a substantive viewpoint the most interesting parameter might be the difference of two mean responses that occur at two well-defined points in time. For instance, in developmental psychology the most important variable for predicting a future outcome might be the progress a child makes in a variable between, say, the ages of two and five. If it is possible to replace the "acceleration" parameter α_{i2} by a parameter that expresses the change between these two ages it may be possible to find a good model for that parameter because it is the most meaningful and interesting one.

We will now show how this can be accomplished by a simple line of argumentation. Let r_1 and r_2 represent the two time points or ages. If a quadratic model fits the data then the difference in mean response for the ith individual can be expressed as

$$E(y_i(r_2) - y_i(r_1)) = \alpha_{i1}(r_2 - r_1) + \alpha_{i2}(r_2^2 - r_1^2).$$

From this equation it can be seen that the transformed parameter $\alpha_{i2} = E(y_i(r_2) - y_i(r_1))$ can be obtained by a linear transformation of the original parameters. Suppose that we would like to retain the meaning of the other two parameters α_{i0} and α_{i1} then the matrix Q^{-1} can be specified as follows

$$Q^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & (r_2 - r_1) & (r_2^2 - r_1^2) \end{pmatrix},$$

because $\alpha_i = Q^{-1}\alpha_i$. It follows that the inverse of Q^{-1}, that is Q, is

$$Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -(r_2 + r_1)^{-1} & (r_2^2 - r_1^2)^{-1} \end{pmatrix}.$$

Hence, transforming the model matrices T_i to $\tilde{T}_i$ and fitting the model using $\tilde{T}_i$ instead of T_i allows one to specify the level-two model for α_{i2} that has exactly the desired meaning.

To gain a more intuitive understanding of the suggested method we can solve the matrix expression

$$E(y_{ij}) = \tilde{t}'_{ij}\alpha_i = t'_{ij}Q\alpha_i$$

and write it as a simple equation. This yields

$$E(y_{ij}) = \alpha_{i0} + \alpha_{i1}\left(t - \frac{t^2}{r_2 + r_1}\right) + \alpha_{i2}\left(\frac{t^2}{r_2^2 - r_1^2}\right). \tag{3}$$

We can now consider equation 3 from a naïve point of view and ask about the meaning of the parameters. It comes as no surprise and is easily verified that

$$\alpha_{i0} = E(y(0)) \quad \text{and} \quad \alpha_{i1} = \frac{d}{dt}E(y(t))\Big|_{t=0}$$

as the transformation was intended to leave these parameters unchanged. However, consider the meaning of the parameter α_{i2}. This time calculating the second derivative will not result in an equation where α_{i2} is the only unknown in the resulting equation. Although we already know its

interpretation it is instructive to calculate $E(y_i(r_2) - y_i(r_1))$ and see that the result is indeed α_{i2}. We obtain

$$E(y_i(r_2) - y_i(r_1))$$

$$= \alpha_{i1}\left[\left(r_2 - \frac{r_2^2}{r_2 + r_1}\right) - \left(r_1 - \frac{r_1^2}{r_2 + r_1}\right)\right] + \alpha_{i2}\left(\frac{r_2^2}{r_2^2 - r_1^2} - \frac{r_1^2}{r_2^2 - r_1^2}\right)$$

$$= \alpha_{i1} * 0 + \alpha_{i2} * 1$$

$$= \alpha_{i2}.$$

This calculation shows that the parameter α_{i2} has indeed the meaning it is supposed to have.

The computations for transforming the values in T_i are very simple for the example given and require only a few programming statements in a SAS data step or in other statistical software packages as SPSS, SYSTAT, or S-Plus. From equation 3 it can best be seen what calculations have to be performed. Only the second and third column of the model matrix T_i have to be changed to yield $\tilde{T}_i$.

The values of second column must be changed from t to $t - t^2 = (r_2 + r_1)$ while the values of the third column must be changed from t^2 to $t^2 = (r_2^2 - r_1^2)$. In SAS this can be achieved with the following statements:

```
data new_T;
    input y t t2;
    r1 = 2; r2 = 5;
    t_tilde = t-t**2 / (r2+r1);
    t2_tilde = t**2 / (r2**2 - r1**2);
    cards;
    . . .
    . . .
; run;
```

By setting the values of r1 and r2 to the desired values and using the new variables t_tilde and t2_tilde instead of t and t2 the meaning of $\tilde{\alpha}_{i2}$ changes as desired.

In general, transformations of parameters can be made arbitrarily complex as it is possible, once a first transformation Q_1 is applied, to repeat the procedure over and over again using matrices $Q_2, Q_3, \ldots$. Formally,

$$E(y_{ij}) = t'_{ij}(Q_1 Q_2 \ldots Q_k)(Q_k^{-1} \ldots Q_2^{-1} Q_1^{-1})\alpha_i$$

$$= t'_{ij}Q\tilde{\alpha}_i,$$

where $Q = Q_1 Q_2 \ldots Q_k$ and $\alpha_i = Q_k^{-1} \ldots Q_2^{-1} Q_1^{-1}\alpha_i = Q^{-1}\alpha_i$.

For instance we can transform the parameters using the transformation of the last section to change the meaning of α_{i0} and α_{i1} and after that change the meaning of α_{i2} by applying the transformation suggested in this section. It should be noted that the transformation from this section will retain the change in meaning that has been achieved by the transformation of the last section.

5. Example

To illustrate the outlined approach to analysing repeated measures data with multilevel models we reanalyse data from a study of children's vocabulary development during the second year of life (Huttenlocher, Haight, Bryk, & Seltzer, 1991). The data set is distributed with the HLM (Version 4) computer software (Bryk, Raudenbush, & Congdon, 1996) and has been analysed by Bryk and Raudenbush (1992, p. 141).

It combines the data of two different studies. In each study, the vocabulary size of 11 children was measured at the age of 12 months and on several follow-up studies thereafter. The exact schedule of obtaining measurements as well as the amount of repeated observations differed somewhat between the two studies but this poses no problems when analysing the data using a multilevel model. However, it is important to note that the number of repeated observations ranges from 3 to 8 time points. If one does not want to exclude children with only three repeated measures this restricts the order of level-one polynomials that can be fitted to the second-degrees.

When the children were 16 months old the amount of maternal speech was also recorded. This measure serves as a level-two covariate. The other level-two covariates are gender and study group.

Bryk and Raudenbush considered fitting a multilevel model based on the following decisions:

1. Age is expressed as deviations (in months) from the age of 12 months. Therefore, when speaking of age zero in the following we essentially refer to the age of 12 months.
2. The intercept coefficient for the level-one model, α_{i0}, is constrained to be zero. Consequently, all individual trajectories take on a value of zero vocabulary size at the age of zero. This is equivalent to saying that all individual trajectories are constrained to pass through the origin.
3. The slope coefficient of the level-one model, α_{i1}, expressing the change rate at age zero is a random variable with mean zero. Therefore, no level-two model is necessary for this parameter.

4. The "acceleration" coefficient is modelled as

$$\alpha_{i2} = \beta_{20} + \beta_{21}z_1 + \beta_{22}z_2 + \beta_{23}z_3 + \beta_{24}z_1z_2 + \beta_{25}z_1z_3 + \zeta_{i2},$$

where

$$z_1 = \text{study group,}$$
$$z_2 = \text{gender,}$$
$$z_3 = \text{maternal speech index.}$$

Note that this model equation contains two interaction terms. Fig 1. gives a visual impression of the data from both studies. The left panel shows data from the children of the first group, the right panel shows data from the second group. For the observations of each child a second-order polynomial was fitted by ordinary least squares (OLS). These regression lines are also shown in the plot. Note that all regression lines are constrained to go through the origin and that in the figure the x-axis expresses age as deviations from the age of 12 months.

From Fig. 1 a few details should be noted. First, the data points for each child are usually quite close to the quadratic polynomials. Second, there is no variation of the vocabulary size measure at age zero. Looking at the raw data shows that all children obtained a score of one for vocabulary size at the age of 12 months. Third, although the polynomials capture the overall shape of the growth in vocabulary size quite well, there are also some undesirable characteristics of the polynomials presented in the left panel of Fig. 1. Note that the visual impression is that there might be two different types of children. The first type shows a fast increase in vocabulary size in the time interval under study whereas the second type shows only a moderate increase. However, for the "type one" children the development at the age of 18 months is clearly overestimated by the polynomials whereas for the "type two" children it appears that at least for some children vocabulary size decreases at first which is unreasonable.

Problematic characteristics of fitted polynomials like these are very typical. Polynomials are virtually always fitted to data because they can be easily handled and are very flexible. However, polynomials seldom are believed to be the "true" model but are considered a more or less good approximation to it. For a discussion of undesirable characteristics of polynomial regression models see Royston and Altman (1994).

Consider now the parameter estimates for the level-two model for α_{i2} as reported by Bryk and Raudenbush (1992, p. 147) given in Table 1. First note that neither interaction effect associated with the coefficients β_{24} and β_{25} is significant. With the contrast coding scheme employed for fitting the model to the data the difference in the average "acceleration" between
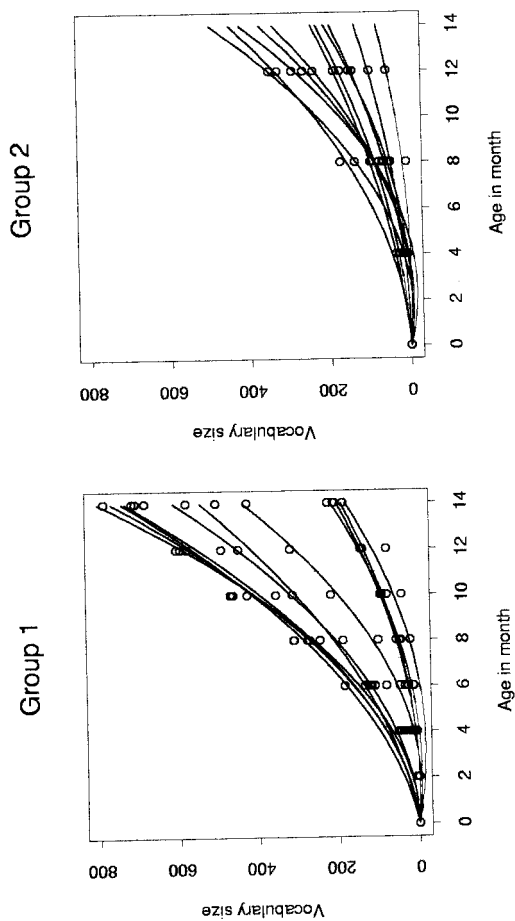
FIG. 1 Raw data and second-degree polynomials fitted by OLS for each individual. All regression curves are forced to go through the origin. The left panel shows data from group 1 and the right panel shows data from group 2. Age is expressed in deviations from 12 months of age.

TABLE 1
Parameter Estimates with Standard Errors (SE) for the Level-Two Model of the
"Acceleration" Parameter α_{i2}, as well as the Variance Components for the Three
Random Effects in the Model

Coefficient	Estimate	SE	t-value
β_{20} Intercept	2.031	0.157	12.887
β_{21} Group	− 0.433	0.157	− 2.747
β_{22} Gender	0.312	0.165	1.891
β_{23} Momspeech	0.793	0.334	2.370
β_{24} Interac.1	0.144	0.165	0.876
β_{25} Interac.2	− 0.158	0.334	− 0.473
Random Effect	Variance Comp.	df	χ^2
ζ_1	18.778	22	39.547
ζ_2	0.282	16	39.864
ε_{ij}	707.156		

Deviance = 1,284.3 with 4 df

Note: These values were reported by Bryk and Raudenbush (1992, p. 147).

boys and girls is estimated to be $2 * \hat{\beta}_{22} = 0.624$ words/month^2. Likewise, the difference in the average "acceleration" between children from the two study groups is estimated to be 0.866 words/month^2.

These results are not easily interpreted as the units in which the differences are expressed are at least in the minds of the present authors not very meaningful. This is due to the fact that the units in which the regression coefficients are expressed include the term "month^2". Evaluating these regression coefficients in a qualitative way one could say that a positive value means a faster growing change rate for one of the groups. This results on average in higher vocabulary size values for one of the groups with the difference between vocabulary size growing over time. Note also, that the "acceleration" rate does not depend on age. Therefore, extrapolating vocabulary size beyond the time period from which data were obtained should be avoided because it might give unreasonable results.

We now use the reparameterisation that was explained theoretically in the last section. From a substantive point of view the change in vocabulary size between 12 months and 26 months of age may be an interesting characteristic. We can obtain a parameter that expresses that difference simply by using equation 3. However, before using this transformation we "centre" the age variable so that age 12 months corresponds to age zero. Therefore, the values r_2 and r_1 in equation 3 are set to the values 14 and 0, respectively. After having transformed the variables in the design matrix we fit the model again and obtain the results that are given in Table 2.

TABLE 2
Parameter Estimates with Standard Errors (SE) for the Level-Two Model after
Reparameterisation as well as the Variance Components for the Three Random
Effects in the Model

Coefficient	Estimate	SE	t-value
β_{20} Intercept	398.44	30.71	12.971
β_{21} Group	– 84.60	30.71	– 2.754
β_{22} Gender	60.72	32.17	1.887
β_{23} Momspeech	155.41	65.34	2.379
β_{24} Interac.1	29.05	32.17	0.903
β_{25} Interac.2	– 30.52	65.34	– 0.467
Random Effect	Variance Comp.	df	χ^2
ζ_1	17.557	22	36.639
ζ_2	24697.871	16	812.888
ε_{ij}	713.062		

Deviance = 1,221.1 with 4 df

Before considering the interpretation of the parameters note that both
interaction terms are, as before, not significant. Hence, the main effects
behave additively. Also note that the t-values of Tables 1 and 2 are
virtually identical. The reparameterisation does not change the validity of
the model. It merely influences the meaning of the parameters.

These have now a very natural interpretation. Consider first the
coefficient for gender, β_{22}. Recall that all fitted polynomials are forced
to go through the origin. Taking the contrast coding scheme for gender
into account (-1 = male, 1 = female) the estimated value of $\hat{\beta}_{22} = 60.72$ as
given in Table 2 means that on average by the time the investigation is
finished the mean difference in vocabulary size between girls and boys is
$1 * 60.72 – (– 1) * 60.72 = 121.44$ after controlling for study group and
maternal speech.

In a similar manner the estimate for the group effect, β_{21} can be
interpreted that after controlling for gender and the maternal speech index
the vocabulary size of the first group (coded as -1) is at the time the
investigation is finished on average 169.2 units higher than the vocabulary
size of the second group (coded as 1).

Finally, the parameter estimate for the maternal speech index, β_{23},
shows that an increase of this variable in logarithmic units is accompanied
on average by a change of vocabulary size of 155.41 over the 14-month
study period after controlling for gender and group.

A look at the variance component estimates of Tables 1 and 2 is also
very interesting. The first and the third variance component are very
similar in both tables meaning that the variation of the slopes at time zero

(12 months of age) is about the same in both models and the random error variance of the level-one model is about equal in both models. However, the variance component for the individual differences in growth of vocabulary size over the 14-month study period is huge. This is a very desirable characteristic of a parameter since it is usually easier to specify a model for a parameter that shows considerable variation than for a parameter that shows only a small variation.

It should be noted that although the example illustrates how the meaning of the parameters can be determined by transforming the model matrix it could not be used to demonstrate how the model building process is influenced by modelling parameters that are substantively meaningful.

DISCUSSION

As with statistical models in general, conclusions drawn from multilevel models depend heavily on an adequate model specification and on at least approximately satisfied model assumptions. When using multilevel models the problem of possible model misspecification is even more problematic than in most of the other statistical models since multilevel models set up not only a level-one model but usually several level-two models as well. It is well known that misspecification of the level-one model not only leads to bias in the level-one but also in the level-two parameter estimates. Additionally, a level-two model for a level-one parameter that holds, will only lead to the right conclusions if there is no or only a small bias in the estimates for the level-one parameter.

Another important aspect of formulating level-two models is the meaning of the level-one parameters. Because it is much easier and much more desirable from a substantive viewpoint to specify models that focus directly on the research hypothesis under study, the possibility to custom tailor the meaning of level-one parameters might considerably improve the statistical analysis. Once an adequate level-one model is found, a reparameterisation of the level-one model formula can lead to more meaningful parameters and therefore to more natural model formulation at the second level. The method suggested by the present authors can be used as long as the new parameters can be expressed as linear functions of the original parameters. It is easy to use and there is no need for specialised software since in most situations it should be possible to yield more meaningful parameters through relatively simple transformations. Furthermore, because a reparameterisation can proceed in several steps the computations can be split up to make calculations with matrices unnecessary.

REFERENCES

Box, G.E.P. (1966). Use and abuse of regression. *Technometrics*, *8*, 625–629.

Bryk, A.S., Raudenbush, S., & Congdon, R. (1996). *HML—hierarchical linear and nonlinear modeling with HLM/2L and HLM/3L programs.* Chicago, IL: Scientific Software International, Inc.

Bryk, A.S., & Raudenbush, S.W. (1987). Application of hierarchical linear models to assessing change. *Psychologicl Bulletin*, *101*, 147–158.

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

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

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

Huttenlocher, J.E., Haight, W., Bryk, A.S., & Seltzer, M. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, *27*, 236–249.

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

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

Miller, A.J. (1990). *Subset selection in regression.* London: Chapman & Hall.

Royston, P., & Altman, D.G. (1994). Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Applied Statistics*, *43*, 1–26.

von Eye, A., & Schuster, C. (1998). *Regression analysis for the social sciences.* San Diego, CA: Academic Press.

Ware, J.H. (1985). Linear models for the analysis of longitudinal studies. *The American Statistician*, *39*, 95–101.