

A Newton procedure for conditionally linear mixed-effects models

SHELLEY A. BLOZIS

University of California, Davis, California

This article reviews Newton procedures for the analysis of mean and covariance structures that may be functions of parameters that enter a model nonlinearly. The kind of model considered is a mixed-effects model that is conditionally linear with regard to its parameters. This means parameters entering the model nonlinearly must be fixed, whereas those entering linearly may vary across individuals. This framework encompasses several models, including hierarchical linear models, linear and nonlinear factor analysis models, and nonlinear latent curve models. A full maximum-likelihood estimation procedure is described. Mx, a statistical software package often used to estimate structural equation models, is considered for estimation of these models. An example with Mx syntax is provided.

Mixed-effects models, also called multilevel models and hierarchical models, provide a framework for the study of clustered data when randomly selected units of study are nested within higher-order units that are also randomly selected. Hierarchical data structures occur in many domains, including psychology, sociology, business, nursing, and medicine. A classic example is education, where students are “nested” within classrooms. This general methodology addresses the dependencies of scores within random clusters by allowing coefficients defining a data model to vary across clusters. Many books are devoted to the study of these models, including Davidian and Giltinan (1995), Hox (2002), Kreft and de Leeuw (1998), Longford (1993), Raudenbush and Bryk (2002), Singer and Willett (2003), and Snijders and Bosker (1999). A mixed-effects model allows for both the study of variables at multiple levels and for interactions between variables at different levels. Estimation of effects at the different levels is simultaneous. The methodology is flexible, allowing for unbalanced data, such as unequal group sizes.

Given the adaptable nature of this approach, mixed-effects models are increasingly used in areas where stratified sampling is common. This is also due in part to the widespread availability of statistical software packages for estimating these models. These include HLM (Raudenbush, Bryk, Cheong, & Congdon, 2004), LISREL (Jöreskog, Sörbom, Du Toit, & Du Toit, 2001), MLwiN (Goldstein et al., 1998), Mplus (Muthén & Muthén, 1998–2004), SAS PROC MIXED (Littell, Milliken, Stroup, Wolfinger, & Schabenberger, 2006) and SAS PROC NL-MIXED (Wolfinger, 1999), the mixed-effects procedures in S-PLUS (MathSoft, Inc., 1997), SPSS (SPSS, 2001), and STATA (Stata, 1997), SUDAAN (Shah, Barnwell, & Bieler, 1997), VARCL (Longford, 1990), and Donald

Hedeker’s programs for a variety of mixed-effects models (available for downloading at tigger.uic.edu/~hedeker/mix.html).

These models are also applied to repeated measures and longitudinal data in which repeated measures of the same variable are taken, for example, within a single experimental session or across some specified time period. Such data are hierarchically structured because the repeated measures are nested within individuals. Because of the flexibility of the methodology, individuals need not be assessed the same number of times or measured (in the context of time-structured data) according to the same measurement occasions (Jennrich & Schluchter, 1986). Thus, unbalanced data and unequal spacing between assessments (for time-structured data) are easily handled. This means the methodology has greater utility than classic methods, such as a repeated measures ANOVA, which requires complete, time-balanced data. In cases of time-structured data in particular, taking into account accurate measures of time may be important in many studies in which change is a primary focus (e.g., Collins & Graham, 2002).

In the context of longitudinal data in particular, the models are often called latent curve models (Meredith & Tisak, 1984, 1990). In a latent curve model, an individual’s response is typically studied as a function of time with the addition of time-specific error. Coefficients characterizing the response over time, such as an intercept and linear time effect when change is linear, may be individual-specific; for instance, each individual may have a unique intercept and time effect. These individual-specific effects, called random coefficients or random effects, are assumed to be random deviates about a corresponding population value. In this way, the model provides a means of relating re-

S. A. Blozis, sablozis@ucdavis.edu

sponses at the individual level to responses at the level of the general population.

In a latent curve model, variation is commonly considered at two levels. At the first level, errors are time-specific and relate to the deviations of scores about an individual's curve. At the second level, the random effects describing an individual response as a systematic function of time may vary across individuals. Like mixed-effects models for cross-sectional data, latent curve models may include predictors at both levels. Predictors may be time-varying (Level 1) or time-invariant, in which case a variable varies only between individuals (Level 2). Estimation of a latent curve model typically yields estimates of fixed effects (e.g., fixed intercept, fixed time effect) and variances of the random effects and their covariances at the second level, in addition to variance and possibly covariance estimates relating to the time-specific errors at the first level. Random effects at the second level are not estimated directly in the process of fitting a model to data; however, estimates may be obtained via least squares or Bayesian procedures based on an estimated model (Raudenbush & Bryk, 2002).

While linear forms of change are often appropriate in describing behavioral trajectories, there are many situations in which a behavior follows a nonlinear trend. Linear models can handle some nonlinear forms of change. Polynomial functions in which various powers of the predictors (e.g., time) are included can approximate some nonlinear change forms. In other cases, a nonlinear function may be more appropriate. In a nonlinear mixed-effects model, at least one parameter enters in a nonlinear way. Some examples include functions in which a parameter appears in an exponent, such as in an exponential function. In a fully nonlinear mixed-effects model, a parameter that enters the model nonlinearly may vary at random across individuals (Davidian & Giltinan, 1995, chap. 4). This may be useful in longitudinal studies, for example, where a variable changes nonlinearly and a nonlinear change rate varies across individuals. A limited number of available software packages, such as SAS PROC NL MIXED, may handle nonlinear mixed-effects models (Wolfinger, 1999). Applications of these models are often limited, however, due to difficulties in, and computational demands of, their estimation. In many cases, a model may be limited to one, possibly two, random effects that enter nonlinearly in the second level because of these estimation difficulties. In the social and behavioral sciences, this may be problematic where there is growing interest in treating multiple variables studied longitudinally (Blozis, 2004; MacCallum, Kim, Malarkey, & Kiecolt-Glaser, 1997). In these applications, to study the interrelationships among change characteristics of different variables—such as the study of how the change rate in one variable is related to the change rate in another variable—multiple longitudinal variables are considered simultaneously. Fully nonlinear mixed-effects models will be limited in such cases, to the extent that the models for the individual variables include parameters that enter nonlinearly.

An alternative to a fully nonlinear mixed-effects model is a conditionally linear mixed-effects model (CLMM)

(Blozis & Cudeck, 1999). In a CLMM, parameters entering a model in a nonlinear manner are fixed. Those entering linearly may be fixed or random. Consequently, estimation of the model may proceed using standard procedures for linear models without the computational difficulties inherent to a fully nonlinear mixed-effects model (Davidian & Giltinan, 1995). Although restricted to cases in which the relationship between the response and predictor may be nonlinear but where nonlinear parameters are fixed across individuals, the approach still represents a flexible means of specifying models and may serve as a superior alternative to the standard approach typically based on polynomial functions.

This article reviews a method of estimating a class of models based on mean and covariance structures that may be functions of parameters entered nonlinearly. We also examine how this estimation procedure, typically employed for strictly linear models, may be used to handle this class of models, which usually includes a CLMM—a model in which nonlinear parameters must be fixed and linear parameters may be random (Blozis & Cudeck, 1999). Other models that may be considered under this general framework include a hierarchical linear model (Raudenbush & Bryk, 2002), a linear factor analysis model (Lawley & Maxwell, 1971), a nonlinear factor analysis model, and a nonlinear latent curve model for longitudinal data (Browne, 1993; Browne & Du Toit, 1991). Newton–Raphson and quasi-Newton algorithms are reviewed for estimation of CLMMs to obtain maximum-likelihood estimates (MLEs) of model parameters. Relatively few statistical software packages typically used by psychologists are commercially available for estimating CLMMs. Blozis and Cudeck (1999) consider different programming languages (e.g., SAS IML, a matrix-based programming language) to fit such models. In practice, such programming languages may be unfamiliar to many researchers in the social and behavioral sciences. Mx (Neale, Boker, Xie, & Maes, 2003), a program commonly used for estimating structural equation models, may be used for estimating CLMMs because it allows for some nonlinear constraints on model coefficients. Mx uses a quasi-Newton procedure to obtain MLEs.

The intentions behind this effort were threefold: First, to show how a CLMM is general, encompassing several models, including a hierarchical linear model, linear and nonlinear factor analysis models, and a nonlinear latent curve model; second, to provide a pedagogical guide for understanding steps underlying estimation procedures that may be used to fit these models; third, to demonstrate by way of an example how the software program Mx may be used to carry out estimation of these models. An example is provided.

MODEL SPECIFICATION

A CLMM is a generalization of a linear mixed-effects model for repeated measures (Blozis & Cudeck, 1999). An individual response set is given by $\mathbf{y}_i = (y_{1i}, \dots, y_{T_i})'$, where i denotes an individual in a sample of $i = 1, \dots, n$, and T_i is the number of repeated measures for the individ-

ual. The subscript T_i indicates that the number of assessments may vary across individuals. A CLMM for \mathbf{y}_i is

$$\mathbf{y}_i = \mathbf{\Lambda}_i(\boldsymbol{\lambda}, \mathbf{X}_i)\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i, \tag{1}$$

where $\mathbf{\Lambda}_i(\boldsymbol{\lambda}, \mathbf{X}_i)$ is a $T_i \times p$ matrix that, for individual i , may be a function of fixed or random variables contained in \mathbf{X}_i , such as measures of time, and which may be a function of a set of fixed, unknown parameters, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_q)'$, with elements that may enter the model in a linear or nonlinear manner. The coefficient set $\boldsymbol{\eta}_i = (\eta_{1i}, \dots, \eta_{pi})'$ that enters the model linearly contains effects that may vary across individuals. The regression errors are denoted by $\boldsymbol{\varepsilon}_i = (\varepsilon_{1i}, \dots, \varepsilon_{T_i})'$. For time-structured data, the error in Equation 1, $\boldsymbol{\varepsilon}_i$, is considered time-specific. An important model feature is that the coefficient $\boldsymbol{\lambda}$ is fixed, whereas $\boldsymbol{\eta}_i$ may be random. That is, elements of $\boldsymbol{\eta}_i$ may be combinations of fixed and random effects. For example, assuming two elements in $\boldsymbol{\eta}_i$ where the first is fixed and the second a combination of a fixed and random effect, $\boldsymbol{\eta}_i$ may be specified as

$$\boldsymbol{\eta}_i = \begin{bmatrix} \eta_{1i} \\ \eta_{2i} \end{bmatrix} = \begin{bmatrix} \eta_1 \\ \eta_2 + b_{2i} \end{bmatrix},$$

where η_1 and η_2 are fixed and b_{2i} corresponding to η_2 is random.

Within individuals, the error, $\boldsymbol{\varepsilon}_i$, is assumed to be independently and identically distributed as

$$\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Theta}_i), \tag{2}$$

where $\boldsymbol{\Theta}_i$ is a function of a parameter set $\boldsymbol{\theta} = (\theta_1, \dots)'$, containing unique elements to represent the variances and possibly covariances of the errors at the first level. For time-structured data, the errors are often assumed to be independent with constant variance across time, such that $\boldsymbol{\Theta}_i = \theta \mathbf{I}_{T_i}$, where θ is a common error variance, and \mathbf{I}_{T_i} is an identity matrix of order T_i . The matrix $\boldsymbol{\Theta}_i$ generally varies across individuals only with regard to its dimensions. Other structures for $\boldsymbol{\Theta}_i$ may be considered, such as an autocorrelation structure (Browne, 1993). Between individuals the coefficient $\boldsymbol{\eta}_i$ is assumed to be independent and identically distributed as

$$\boldsymbol{\eta}_i \sim N(\boldsymbol{\alpha}, \boldsymbol{\Psi}), \tag{3}$$

where $\boldsymbol{\alpha}$ is the expected value of $\boldsymbol{\eta}_i$, and $\boldsymbol{\Psi}$ is a function of a parameter set $\boldsymbol{\psi} = (\psi_1, \dots)'$ containing unique elements that represent the variances and covariances of the random effects at the second level. Often, the covariance matrix $\boldsymbol{\Psi}$ is assumed to be symmetric and unstructured so that the variances of the random effects $\boldsymbol{\eta}_i$ and their covariances are freely estimated. Finally, it is assumed that the random effects at the second level are independent of the errors at the first.

Given the model in Equation 1 with distributional assumptions in Equations 2 and 3, the expected value and covariance matrix for \mathbf{y}_i are

$$\boldsymbol{\mu}_i = \mathbf{\Lambda}_i(\boldsymbol{\lambda}, \mathbf{X}_i)\boldsymbol{\alpha}$$

and

$$\boldsymbol{\Sigma}_i = \mathbf{\Lambda}_i(\boldsymbol{\lambda}, \mathbf{X}_i)\boldsymbol{\Psi}\mathbf{\Lambda}_i(\boldsymbol{\lambda}, \mathbf{X}_i)' + \boldsymbol{\Theta}_i,$$

respectively. The orders of $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are both T_i .

For convenience, the model parameters described above are specified as a set, $\boldsymbol{\tau} = (\boldsymbol{\alpha}', \boldsymbol{\lambda}', \boldsymbol{\psi}', \boldsymbol{\theta}')'$. In some cases, elements contained in $\boldsymbol{\lambda}$ may also be contained in $\boldsymbol{\alpha}$, as will become evident in one application reviewed below. Assuming $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$ contain unique elements and no restrictions are placed on the elements contained in $\boldsymbol{\Psi}$ or $\boldsymbol{\Theta}_i$, the order of $\boldsymbol{\tau}$ is $w = p + q + j + k$, where p is the number of coefficients in $\boldsymbol{\alpha}$, q is the number of coefficients in $\boldsymbol{\lambda}$, j is the number of elements in $\boldsymbol{\psi}$, and k is the number of elements in $\boldsymbol{\theta}$. For convenience, it is assumed that elements in $\boldsymbol{\eta}_i$ have corresponding elements in $\boldsymbol{\alpha}$, so that the length of $\boldsymbol{\eta}_i$ and $\boldsymbol{\alpha}$ are equal to p . This assumption may be relaxed, however, as described earlier.

NONLINEAR FUNCTIONS

Many nonlinear functions may be handled by a CLMM, in contrast to linear latent curve models that are restricted to functions with parameters that enter linearly. A CLMM is restricted in that parameters that enter nonlinearly must be fixed. Still, this includes a variety of possibilities. Some nonlinear functions that can be handled are, for example, an exponential function

$$y_{it} = \eta_i \exp(\lambda t_{it}) + \varepsilon_{it},$$

in which η_i varies across individuals and λ is fixed. In this case, the matrix $\mathbf{\Lambda}_i = \mathbf{\Lambda}_i(\lambda, \mathbf{t}_i)$, where \mathbf{t}_i is a set of values denoting times of measurement ($\mathbf{t}_i = 1, \dots, T_i$) for the i th individual, would have the m th row defined as $[\exp(\lambda t_{mi})]$. The coefficient η_i is linearly combined with $\mathbf{\Lambda}_i(\lambda, \mathbf{t}_i)$ to yield the predicted value of \mathbf{y}_i . Another possibility is a power function,

$$y_{it} = \eta_{1i} + \eta_{2i}t_{it}^\lambda + \varepsilon_{it},$$

in which $\boldsymbol{\eta}_i = (\eta_{1i}, \eta_{2i})'$ varies across individuals and λ is fixed. For this model, the matrix $\mathbf{\Lambda}_i = \mathbf{\Lambda}_i(\lambda, \mathbf{t}_i)$ would have the m th row defined as $[1, t_{mi}^\lambda]$. Similarly, the coefficient $\boldsymbol{\eta}_i$ would be linearly combined with $\mathbf{\Lambda}_i(\lambda, \mathbf{t}_i)$ to yield the predicted value of \mathbf{y}_i .

In a CLMM, the parameters making up the matrix $\mathbf{\Lambda}_i$, $\boldsymbol{\lambda}$, may enter linearly or nonlinearly but are fixed across individuals in a population. In contrast, those contained in $\boldsymbol{\eta}_i$ enter the model linearly and may vary across individuals. This restriction on model parameters means that estimation is generally straightforward using standard methods such as the Newton–Raphson algorithm (Jennrich & Schluchter, 1986) for strictly linear models. This is because the random effects at the second level and the time-specific errors at the first level enter the model in an additive and linear manner, and are assumed to be independent of each other. Due to the ease of estimation of a CLMM, the approach offers a flexible alternative to polynomial functions that are often used in behavioral research settings while relying on the same estimation techniques.

SPECIAL CASES

The CLMM is a flexible framework that encompasses a number of models, including a hierarchical linear model for data not structured according to time, a linear factor

analysis model, a nonlinear factor analysis model, and a nonlinear latent curve model for longitudinal data. That is, many commonly applied models can be seen as special cases of a CLMM. These cases are described below.

Hierarchical Linear Model

A special case of a CLMM is one in which the matrix Λ_i is a function of a set of known variables, \mathbf{X}_i , that may include fixed predictors or random variables (treated as fixed; Raudenbush & Bryk, 2002), but unlike a CLMM, is not a function of a fixed parameter λ . With this restriction, the model in Equation 1 is modified to represent a hierarchical linear model:

$$\mathbf{y}_i = \Lambda_i(\mathbf{X}_i)\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i,$$

where $\Lambda_i(\mathbf{X}_i)$ is an $n_i \times p$ matrix that is a function of fixed or random variables measured for the i th Level 2 unit, where n_i denotes the number of observations for unit i ; $\boldsymbol{\eta}_i = (\eta_{1i}, \dots, \eta_{pi})'$ is a set of effects that may vary across Level 2 units (e.g., across random groups in a stratified sampling design); and $\boldsymbol{\varepsilon}_i$ is the within-unit error. The random coefficient $\boldsymbol{\eta}_i$ enters linearly. No parameters enter the model nonlinearly. The model may be applied to time-structured or time-unstructured repeated measures data, as well as to cross-sectional and multilevel data. Like the CLMM, a hierarchical linear model allows the number of observations in \mathbf{y}_i to vary across Level 2 units by having the rows of Λ_i and $\boldsymbol{\varepsilon}_i$ and elements of \mathbf{X}_i correspond to the observed values in \mathbf{y}_i . Given a hierarchical linear model for \mathbf{y}_i and assuming the Level 1 errors have mean equal to zero and symmetric covariance matrix Θ , the random effects at Level 2 are independent and identically distributed across Level 2 units with mean equal to $\boldsymbol{\alpha}$ and symmetric covariance matrix Ψ and independence between the errors at Level 1 and the random effects at Level 2, the expected value and covariance matrix of \mathbf{y}_i are

$$\boldsymbol{\mu}_i = \Lambda_i(\mathbf{X}_i)\boldsymbol{\alpha}$$

and

$$\boldsymbol{\Sigma}_i = \Lambda_i(\mathbf{X}_i)\Phi\Lambda_i(\mathbf{X}_i)' + \Theta_i,$$

respectively. The orders of $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are n_i .

Linear Factor Analysis Model

Readers not familiar with a linear factor analysis model may refer to Lawley and Maxwell (1971). In linear factor analysis models, sets of manifest variables are considered linear functions of sets of latent variables. As in a hierarchical linear model, the parameters of a linear factor analysis model enter the model linearly. The model includes a factor matrix patterned to characterize the relationships between the manifest and latent variables. The parameters of the factor matrix are commonly referred to as factor loadings and enter the matrix linearly. Given the specific patterns between the manifest and latent variables, a linear factor analysis model may be expressed as

$$\mathbf{y}_i = \Lambda_i(\lambda)\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i,$$

where, for individual i , $\Lambda_i(\lambda)$ is an $m_i \times p$ matrix of fixed and known elements (typically either 1s or 0s) and a set of fixed and unknown parameters, $\lambda = (\lambda_1, \dots, \lambda_q)'$, where

m_i is the number of indicator variables. Missing data are handled for cases in which individuals have a subset of the m indicator variables, where m is the total number of possible indicators, as denoted by the notation m_i (Finkbeiner, 1979). When data are incomplete, the rows in $\Lambda_i(\lambda)$ and $\boldsymbol{\varepsilon}_i$ correspond to those in \mathbf{y}_i (e.g., see Blozis & Cudeck, 1999). The coefficient $\boldsymbol{\eta}_i = (\eta_{1i}, \dots, \eta_{pi})'$ represents a set of factors that vary across individuals, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{1i}, \dots, \varepsilon_{mi})'$ is the set of measurement errors. Given a linear factor analysis model for \mathbf{y}_i and assuming the factors are independent and identically distributed across individuals with mean equal to $\boldsymbol{\alpha}$ and symmetric covariance matrix Ψ , the measurement errors relating to the manifest variables have means equal to zero and symmetric covariance matrix Θ (typically assumed to be a diagonal matrix), and independence between factors and measurement errors, the expected value and covariance matrix of \mathbf{y}_i are

$$\boldsymbol{\mu}_i = \Lambda_i(\lambda)\boldsymbol{\alpha}$$

and

$$\boldsymbol{\Sigma}_i = \Lambda_i(\lambda)\Psi\Lambda_i(\lambda)' + \Theta_i,$$

respectively. The orders of $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are both m_i .

Nonlinear Factor Analysis Model

A nonlinear factor analysis model may be specified in which manifest variables may be nonlinear functions of latent variables. As with the CLMM in Equation 1, a nonlinear factor analysis model may be specified in which the matrix Λ_i is a function of a fixed, unknown parameter λ with elements that may enter the matrix linearly or nonlinearly. Similar to a linear factor analysis model, the elements of Λ_i , λ , characterize the relationships between the manifest and latent variables. In this case, the relationships may be nonlinear. This model represents a restricted case of a nonlinear factor analysis model in Amemiya (1993; see also Yalcin & Amemiya, 2001) where both the factor $\boldsymbol{\eta}_i$ and the parameters in λ may enter the model in a nonlinear manner. The model presented here is different from a nonlinear factor analysis model considered by McDonald (1962, 1965, 1967a, 1967b) and Etezadi-Amoli and McDonald (1983), where it is the factor $\boldsymbol{\eta}_i$ that may enter the model in a nonlinear manner and the parameter λ that may only enter linearly.

Given the specific patterns among the manifest and latent variables, a nonlinear factor analysis model may be expressed as

$$\mathbf{y}_i = \Lambda_i(\lambda)\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i,$$

where Λ_i is an $m_i \times p$ matrix that contains for individual i fixed and known elements (typically either 1s or 0s) and may be a function of a fixed and unknown parameter set $\lambda = (\lambda_1, \dots, \lambda_q)'$ whose elements may enter the matrix nonlinearly. The coefficients in $\boldsymbol{\eta}_i = (\eta_{1i}, \dots, \eta_{pi})'$ are factors that vary across individuals, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{1i}, \dots, \varepsilon_{mi})'$ is the set of measurement errors. As in the previous models, some missing data can be handled. That is, an individual may have valid responses for a subset of indicators, as denoted by m_i . Given a nonlinear factor analysis model for \mathbf{y}_i and assuming the factors are independent and identically distributed across individuals with mean equal

to α and symmetric covariance matrix Ψ , the measurement errors relating to the manifest variables have mean equal to zero and symmetric covariance matrix Θ (typically assumed to be a diagonal matrix), and independence between the factors and measurement errors, the expected value and covariance matrix of y_i are

$$\mu_i = \Lambda_i(\lambda)\alpha$$

and

$$\Sigma_i = \Lambda_i(\lambda)\Phi\Lambda_i(\lambda)' + \Theta_i,$$

respectively. The orders of μ_i and Σ_i are m_i .

Nonlinear Latent Curve Model

Browne (1993; see also Browne & Du Toit, 1991) proposed a structured latent curve model (SLCM), a latent curve model in which a score at a given occasion t is assumed to be the sum of a common score, c_{it} , and time-specific error, ε_{it} , and where the mean common score across occasions, denoted by μ , is assumed to follow a function $f(\lambda, t)$ of a fixed coefficient set $\lambda = (\lambda_1, \dots, \lambda_q)'$, with elements that may enter linearly or nonlinearly, and time, $t = (1, \dots, T)'$. This is different from the CLMM in Equation 1, in which the individual-level response is assumed to follow a particular function. Although the theories underlying the SLCM and the CLMM of Blozis and Cudeck (1999) differ, the models share the same technical aspect: Parameters that enter the model in a nonlinear manner are fixed, whereas those that enter linearly may be either random or fixed.

The common scores at the individual level, c_i , are assumed to be linear combinations of a common factor matrix Λ and a set of weights, η_i , that vary across individuals. A factor matrix Λ is formed by first-order partial derivatives of the common function $f(\lambda, t)$, taken with respect to the elements in λ and evaluated across time:

$$\Lambda(\lambda, t) = \begin{bmatrix} \frac{\partial f(\lambda, 1)}{\partial \lambda_1} & \dots & \frac{\partial f(\lambda, 1)}{\partial \lambda_q} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\lambda, T)}{\partial \lambda_1} & \dots & \frac{\partial f(\lambda, T)}{\partial \lambda_q} \end{bmatrix},$$

where

$$\frac{\partial f(\lambda, t)}{\partial \lambda_r}$$

is the first-order partial derivative of the function $f(\lambda, t)$ with respect to the r th coefficient in λ , evaluated at time t . For instance, if the mean common response follows an exponential function such as

$$\mu_i = f(\lambda, t) = \lambda_1 - \exp\{\lambda_2(t)\},$$

then $\Lambda = \Lambda(\lambda, t)$ would be specified as

$$\Lambda(\lambda, t) = \begin{bmatrix} 1 & -t_1 \exp\{\lambda_2 t_1\} \\ \vdots & \vdots \\ 1 & -t_T \exp\{\lambda_2 T\} \end{bmatrix}.$$

Columns of Λ represent basis curves, common to all individuals, relating to different aspects of change in the mean response. An individual's response, y_{it} , is assumed to be a linear, weighted combination of the basis curves based on a first-order Taylor polynomial taken with regard to the mean function, $f(\lambda, t)$, plus error, ε_{it} :

$$y_{it} = f(\lambda, t) + z_{1i}f'_1(\lambda, t) + \dots + z_{si}f'_s(\lambda, t) + \varepsilon_{it},$$

where z_{1i}, \dots, z_{si} are individual-specific weights and $f'_1(\lambda, t), \dots, f'_s(\lambda, t)$ are first-order partial derivatives of the mean function with regard to the coefficients in λ . Specification of a SLCM is possible when the mean function as evaluated according to time, $f(\lambda, t)$, can be composed by a linear combination of its basis curves in $\Lambda(\lambda, t)$ and a set of coefficients in α , where α is obtained by solving the linear equation $f(\lambda, t) = \Lambda\alpha$. Thus, elements in α may contain one or more of the elements in λ . Functions that can be composed in this manner are those that are invariant to a constant scaling factor (see Browne, 1993, p. 177), such as an exponential, logistic, and Gompertz function (Browne & Du Toit, 1991, pp. 178–179). Assuming $f(\lambda, t) = \Lambda\alpha$, the individual response evaluated across time, $y_i = (\gamma_{1i}, \dots, \gamma_{Ti})$, may be expressed using matrix notation as

$$y_i = \Lambda(\lambda, t)\alpha + \Lambda(\lambda, t)z_i + \varepsilon_i,$$

where $z_i = (z_{1i}, \dots, z_{si})'$ and $\varepsilon_i = (\varepsilon_{1i}, \dots, \varepsilon_{Ti})'$. Assuming a random coefficient η_i is the sum of a fixed coefficient set α and random weight set z_i , the model above may be simplified to

$$y_i = \Lambda(\lambda, t)\eta_i + \varepsilon_i.$$

Due to individual-weighting of the basis curves, individual responses depend on the basis curves at varying levels. So although the mean response form is specified by the function $f(\lambda, t)$, no assumptions are made about the form of the individual curves. The model has also been extended for simultaneous study of multiple variables and to handle unique time points across individuals (Blozis, 2004). For unbalanced or missing data, the factor matrix may be a function of time with values that vary across individuals (i.e., $t = 1, \dots, T_i$). In these cases, Λ_i with the added subscript i allows flexibility in how the matrix is evaluated with regard to time. Based a SLCM for y_i and assuming the random weight z_i is independent and identically distributed across individuals with mean equal to zero and symmetric covariance matrix Ψ , the time-specific errors relating to the observed scores have mean equal to zero and symmetric covariance matrix Θ_i , and independence between the random weights at the second level and time-specific errors at the first, the expected value and covariance matrix of y_i are

$$\mu_i = \Lambda_i(\lambda, t_i)\alpha$$

and

$$\Sigma_i = \Lambda_i(\lambda, t_i)\Psi\Lambda_i(\lambda, t_i)' + \Theta_i,$$

respectively. The orders of μ_i and Σ_i are T_i .

MAXIMUM-LIKELIHOOD ESTIMATION OF MODEL PARAMETERS

MLE of model parameters for the CLMM in Equation 1 may be obtained using the Newton–Raphson algorithm, a procedure often used for estimation of linear models. Assuming \mathbf{y}_i is a longitudinal response set for individual i , a response set across individuals is denoted by $\mathbf{y}'_1, \dots, \mathbf{y}'_N$. Assuming \mathbf{y}_i is multivariate normal with mean $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$, it is hypothesized that $\boldsymbol{\mu}_i = \boldsymbol{\mu}_i(\boldsymbol{\tau})$ and $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i(\boldsymbol{\tau})$, where $\boldsymbol{\tau}$ is an unknown fixed parameter vector with elements defined by the model in Equation 1, with distributional assumptions given in Equations 2 and 3. That is, under the model, the mean vector and covariance matrix are assumed to be functions of the parameters contained in $\boldsymbol{\tau}$.

Parameter estimates may be obtained by optimizing a log-likelihood function corresponding to the marginal density of a response, \mathbf{y}_i , specified by structures hypothesized for $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$. A log-likelihood function based on parameter set $\boldsymbol{\tau}$ and observations across individuals, $\mathbf{y}'_1, \dots, \mathbf{y}'_N$, is

$$\ln L(\boldsymbol{\tau}; \mathbf{y}'_1, \dots, \mathbf{y}'_N) \propto C - \frac{1}{2} \sum_{i=1}^N \left(\ln |\boldsymbol{\Sigma}_i| + \mathbf{q}'_i \boldsymbol{\Sigma}_i^{-1} \mathbf{q}_i \right), \quad (4)$$

where $\mathbf{q}_i = (\mathbf{y}_i - \boldsymbol{\mu}_i)$, C is a constant, and N is the number of individuals, that is, a log-likelihood is a function of the parameter vector $\boldsymbol{\tau}$ and the observed responses across individuals, $\mathbf{y}'_1, \dots, \mathbf{y}'_N$. When the repeated measures are both complete and observed at the same times for all individuals, a different form of a log-likelihood function based on sufficient statistics (e.g., a mean vector and covariance matrix) may be preferable to reduce computational demands (see Browne & Du Toit, 1991). The method described here is most flexible in allowing unbalanced and missing data (Jennrich & Schluchter, 1986).

The goal of the estimation process is to obtain an estimate of $\boldsymbol{\tau}$ that maximizes the log-likelihood function in Equation 4, given sample data. With different values of $\boldsymbol{\tau}$, the log-likelihood will take on different values. The expectation is that across different values of $\boldsymbol{\tau}$, a point in the curve of log-likelihood values will peak. Using ideas from calculus, it may be possible to find values for elements in $\boldsymbol{\tau}$ for which the change in the log-likelihood is equal to zero, and thus, a log-likelihood is at a maximum. This occurs when the first-order partial derivatives of the log-likelihood function with respect to the elements in $\boldsymbol{\tau}$ are equal to zero. The goal of the Newton–Raphson algorithm in this context is to find values for the parameters in $\boldsymbol{\tau}$ where these first-order partial derivatives are equal to zero.

The Newton–Raphson algorithm requires a gradient vector, denoted by $\mathbf{g}(\boldsymbol{\tau})$, and a Hessian matrix, denoted by $\mathbf{H}(\boldsymbol{\tau})$, both of which correspond to the log-likelihood function given in Equation 4 and are of order w . The gradient vector contains the set of first-order partial derivatives of the log-likelihood function with respect to the parameter vector $\boldsymbol{\tau}$:

$$\mathbf{g}(\boldsymbol{\tau}) = \frac{\partial \ln L}{\partial \boldsymbol{\tau}}.$$

The Hessian matrix is a matrix of second-order partial derivatives of the log-likelihood function taken with respect to the parameter vector $\boldsymbol{\tau}$:

$$\mathbf{H}(\boldsymbol{\tau}) = \frac{\partial^2 \ln L}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}'}$$

Elements of the gradient vector and Hessian matrix may be obtained by different methods, such as by calculating the derivatives analytically or by numerical approximation. When both the gradient and Hessian are calculated analytically, the algorithm is referred to as *Newton–Raphson*. A less complicated but reliable approach is to compute the derivatives for the gradient vector analytically and approximate the Hessian matrix by numerical differentiation of the gradient vector (Dennis & Schnabel, 1986; Jamshidian & Jennrich, 1997); in this case, the procedure is termed *quasi-Newton*. The latter approach is considered here. Elements of the gradient vector obtained by calculating the derivatives analytically for the model given in Equation 1 are provided in Appendix A, where a numerical approximation to the Hessian matrix is also described.

A Newton procedure maximizes the log-likelihood function in Equation 4 with respect to the parameter vector $\boldsymbol{\tau}$ by solving a linear system of equations so that elements of the gradient vector are equal to zero. First, let $\boldsymbol{\tau}_m$ represent a set of current values for $\boldsymbol{\tau}$. The algorithm works by computing new values of the parameters, denoted by $\boldsymbol{\tau}_{m+1}$, based on current values of the parameters, $\boldsymbol{\tau}_m$, and an adjustment based on a ratio of the gradient vector to the Hessian matrix:

$$\boldsymbol{\tau}_{m+1} = \boldsymbol{\tau}_m - \mathbf{H}_m^{-1} \mathbf{g}_m,$$

where the gradient vector, \mathbf{g}_m , and the inverse of the Hessian matrix, \mathbf{H}_m^{-1} , are calculated using the current value of the parameter vector, $\boldsymbol{\tau}_m$. The product of the inverse of the Hessian matrix and the gradient vector, $\mathbf{H}_m^{-1} \mathbf{g}_m$, determines the direction of the update. The update for the parameter vector, $\boldsymbol{\tau}_{m+1}$, is evaluated and then updated. The process continues until the absolute value of the elements contained in the gradient vector are smaller than some specified constant, such as $\text{TOL} = 10^{-5}$:

$$\max(|\mathbf{g}(\boldsymbol{\tau}_m)|) < \text{TOL}. \quad (5)$$

In cases where the updated value of the parameter vector does not increase the log-likelihood function in Equation 4 due to what is commonly termed *overstepping*, a step-half parameter, $0 < \varepsilon_v < 1$, may be introduced. Overstepping may occur when the updated parameter estimate does not improve evaluation of the log-likelihood function value but could have, if the change in parameter values had been scaled to be relatively small. The parameter vector $\boldsymbol{\tau}_{m+1}$ may be successively rescaled by this amount in the sequence $\varepsilon_v = 1, 1/2, 1/4, \dots$, until the new version improves the value of the log-likelihood function:

$$\ln L(\varepsilon_v \boldsymbol{\tau}_{m+1}) > \ln L(\boldsymbol{\tau}_m).$$

The update $\boldsymbol{\tau}_{m+1}$ is set equal to $\varepsilon_v \boldsymbol{\tau}_{m+1}$ and the steps repeated.

When the stopping rule in Equation 5 is satisfied, the likelihood function is at least at a local maximum, and the MLEs are often taken to be the corresponding maximizer:

$$\hat{\boldsymbol{\tau}}_{\text{MLE}} = \boldsymbol{\tau}_{m+1},$$

where $\hat{\boldsymbol{\tau}}_{\text{MLE}}$ represents the set of parameter estimates for $\boldsymbol{\tau}$ at this maximum. It is also worth noting that during the estimation process, it is possible to maximize the marginal distribution of \mathbf{y}_i using in place of $\boldsymbol{\Psi}$ the Cholesky factor of $\boldsymbol{\Psi}$:

$$\mathbf{L}\mathbf{L}' = \boldsymbol{\Psi},$$

where \mathbf{L} is lower triangular. This step ensures that $\boldsymbol{\Psi}$ is positive-definite (Lindstrom & Bates, 1988). A similar strategy may be implemented for $\boldsymbol{\Theta}_i$.

Estimated standard errors corresponding to the parameter estimates may be obtained using the multivariate delta method (Rao, 1973, pp. 385–389). These are obtained from the estimated asymptotic variances of the estimates $\hat{\boldsymbol{\tau}}$. The information matrix, denoted by \mathbf{F} , is the inverse of the expected value of the negative of $\mathbf{H}(\boldsymbol{\tau})$. When $\boldsymbol{\tau}$ is identified and sample size is large, $\hat{\boldsymbol{\tau}}$ is approximately normal with mean vector $\boldsymbol{\tau}$ and covariance matrix \mathbf{F} . Standard errors are obtained by taking the square root of the diagonal of the matrix \mathbf{F} .

Missing Data

Statistical inference of MLE for a mixed-effects model is valid under the assumption that missing data are missing at random, such that the missingness depends only on observed data (Rubin, 1976; Little & Rubin, 2002). For a mixed-effects model based on data where some responses are missing due to subject attrition—for example, if the probability of nonresponse was dependent on previous values of the outcome variable and observed covariates—nonresponse may be ignorable (Laird, 1988). A more restrictive type of missing data is that which arises when the missingness is independent of both covariates and the observed values of the outcome variable. This kind of missing data is considered missing completely at random and represents a greater restriction on the type of missing data permitted under a mixed-effects model. When data are not missing at random, a random-effects pattern-mixture model may be adopted (Little, 1995), in which individuals are classified by their pattern of missing data, and indicator variables are created to represent the different classes. These indicator variables are then included in the model and their effects studied along with covariates and possible interactions between them. Under this model, nonresponse may be ignorable, therefore, inferences based on MLE may be considered valid (Hedeker & Gibbons, 1997).

Software

Many different computer programming languages (e.g., Fortran, C++), including matrix-based programs (e.g., Gauss, SAS IML) may be used to implement the procedures described here to obtain MLE for a CLMM. A high-level language that implements the procedures described here and is available for download is Mx (Neale et al.,

2003). The user need specify only the mean and covariance structures for a particular model, not the code for the algorithm used to obtain parameter estimates. Mx uses a quasi-Newton procedure for optimization. Thus, depending on the resources of the researcher, many options are possible. Mx, in particular, is commonly used to estimate structural equation models but is flexible, in that nonlinear constraints may be specified, making it possible to estimate models that are conditionally linear with regard to the random effects at the second level. Mx has replaced estimates of standard errors with confidence interval estimates. An example presented below is supplemented by Mx syntax (see Appendix B).

Example

Data are presented from the National Longitudinal Survey of Youth (NLSY) of Labor Market Experience in Youth supported by the U.S. Department of Labor. The data considered here concern the children of female respondents of the original sample and represent a subset of 405 cases described in Curran (1997). Scores for children who were between 6 and 8 years old at the first assessment were considered. Reading performance was measured by the Peabody Individual Achievement Test (PIAT) Reading Recognition subtest, on 84-item test designed to assess word recognition and pronunciation ability. Test scores represent a sum of items correctly answered, possibly ranging from zero to 84. As in Curran, scores were divided by 10. Children assessed approximately every other year up to four times are considered here. Some reading data were missing.

Reading scores were studied as a function of age. Age was based on the child’s reported age at the first assessment and subsequent ages calculated as expected ages equal to the starting age plus increments of 2, 4, or 6 years following the first assessment. A plot of scores by age for a random set of 30 children with complete data for four occasions suggests scores increase over the years with change in scores possibly slowing as the children grow older (see Figure 1). Age was centered to 10 years by computing $age_i = age_i^* - 10$, where age_i^* was a child’s age before centering. The set of ages for an individual across the study period is denoted by $\mathbf{age}_i = (age_{i1}, \dots, age_{iT_i})'$, where $t_i = 1, \dots, T_i$, and T_i is the total number of observations for individual i .

A series of models were considered for these data with estimation carried out using Mx version 1.5. Deviance values ($-2\ln L$) and an index of fit provided by the Akaike Information Criterion (AIC) for comparing the different models considered are provided in Table 1. The AIC is based on deviance values and penalizes models with additional parameters. Smaller values among competing models are preferred. The AIC was calculated as $AIC = -2\ln L + 2p$.

The first model, Model 1A, was a quadratic growth model in which three regression coefficients were assumed to be random across individuals. As specified at the first level for individual i , the model was

$$y_{it} = \eta_{0i} + \eta_{1i}age_{it} + \eta_{2i}age_{it}^2 + \epsilon_{it},$$

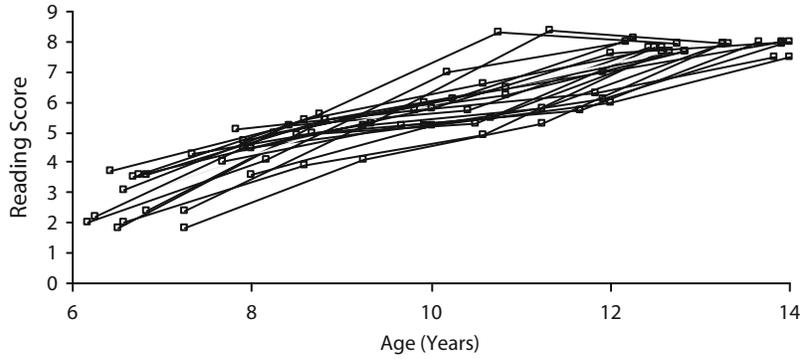


Figure 1. Random subsample of reading scores for 30 children with complete data by approximate age at time of measurement.

where, for individual i , y_{it} was a reading score at time t , and the coefficients η_{0i} , η_{1i} , and η_{2i} were the expected reading level at age 10 (due to centering age at 10), the instantaneous linear rate of change in reading at age 10 (also due to centering age at 10), and the deceleration rate in scores over time, respectively. Across multiple time points, the model may be expressed in matrix notation:

$$y_i = \Lambda_i(\mathbf{age}_i)\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i, \tag{6}$$

where $\Lambda_i(\mathbf{age}_i)$ was specified as

$$\Lambda_i(\mathbf{age}_i) = \begin{bmatrix} 1 & age_{i1} & age_{i1}^2 \\ \vdots & \vdots & \vdots \\ 1 & age_{iT_i} & age_{iT_i}^2 \end{bmatrix}.$$

At the first level of the model, the time-specific errors were assumed to be independent and normally distributed with mean equal to zero and constant variance across age:

$$\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Theta}_i), \tag{7}$$

with $\boldsymbol{\Theta}_i = \theta \mathbf{I}_{T_i}$, where $\theta = var(\boldsymbol{\varepsilon}_{it})$ was the variance of the errors at any time point and \mathbf{I}_{T_i} an identity matrix of order T_i . Thus, $\boldsymbol{\Theta}_i$ varied across individuals only with regard to its dimensions. At the second level, the random coefficients $\boldsymbol{\eta}_i = (\eta_{0i}, \eta_{1i}, \eta_{2i})'$ were assumed to be independent and normally distributed as

$$\boldsymbol{\eta}_i \sim N(\boldsymbol{\alpha}, \boldsymbol{\Psi}), \tag{8}$$

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2)$ represents population values of reading level at age 10, the instantaneous change rate in scores at age 10, and the deceleration rate, respectively. The covariance matrix for the random coefficients, $\boldsymbol{\Psi}$, was assumed to be symmetric with no constraints on its elements:

$$\boldsymbol{\Psi} = \begin{bmatrix} \psi_{00} & & \\ \psi_{10} & \psi_{11} & \\ \psi_{20} & \psi_{21} & \psi_{22} \end{bmatrix}.$$

The diagonal elements of $\boldsymbol{\Psi}$ are the variances of the random coefficients (i.e., ψ_{00} , ψ_{11} , and ψ_{22}) and characterize the extent of individual differences in change features. The off-diagonal elements are the covariances between change characteristics (i.e., ψ_{10} , ψ_{20} , and ψ_{21}). Given the model for the individual-level response in Equation 6, with distributional assumptions in Equations 7 and 8, the expected value and covariance matrix of y_i under Model 1A were

$$\boldsymbol{\mu}_i = \Lambda_i(\mathbf{age}_i)\boldsymbol{\alpha}$$

and

$$\boldsymbol{\Sigma}_i = \Lambda_i(\mathbf{age}_i)\boldsymbol{\Psi}\Lambda_i(\mathbf{age}_i)' = \boldsymbol{\Theta}_i,$$

respectively. Model 1A required 10 parameters in total. Although not required by Mx, starting values were supplied to fit the model. For the mean function, starting values were obtained using ordinary least squares and by fitting a quadratic curve to the mean scores at each wave.

**Table 1
Growth Models for Reading Scores by Age**

Model	Data Model	Level 1 Error			
		Structure	p	$-2\ln L$	AIC
1A	$y_{it} = \eta_{0i} + \eta_{1i}age_{it} + \eta_{2i}age_{it}^2 + \varepsilon_{it}$	I	10	1970	1990
1B	$y_{it} = \eta_{0i} + \eta_{1i}age_{it} + \eta_{2i}age_{it}^2 + \varepsilon_{it}$	I	7	2006	2020
1C	$y_{it} = \eta_{0i} + \eta_{1i}age_{it} + \eta_{2i}age_{it}^2 + \varepsilon_{it}$	II	11	1969	1991
2A	$y_{it} = f(\boldsymbol{\lambda}, age_{it}) + f'_1z_{1i} + f'_2z_{2i} + f'_3z_{3i} + \varepsilon_{it}$	I	10	1958	1977
2B	$y_{it} = f(\boldsymbol{\lambda}, age_{it}) + f'_1z_{1i} + f'_2z_{2i} + \varepsilon_{it}$	I	7	1979	1993
2C	$y_{it} = f(\boldsymbol{\lambda}, age_{it}) + f'_1z_{1i} + f'_2z_{2i} + f'_3z_{3i} + \varepsilon_{it}$	II	11	1958	1980

Note—Level 1 error structure: I assumes independent errors with constant variance across age, II assumes homogeneous error variances across age and covariances between errors at temporally adjacent time points (with a 2-year interval) with constant covariance across time. For each model, the covariance matrix for the random effects at the second level was unstructured so that the variances of each effect and their covariances were unconstrained. AIC, Akaike information criterion.

Table 2
Maximum-Likelihood Estimates (MLEs) Based on an Exponential Function for Reading Scores by Age

	MLE	95% CI
Quadratic Growth Model—Parameter		
Reading level at 10 years, α_0	4.66	(4.53, 4.79)
Instantaneous change rate at age 10 years, α_1	.537	(.515, .558)
Deceleration rate, α_2	-.0470	(-.0544, -.0397)
Exponential Growth Model—Parameter		
Potential reading score, λ_0	7.53	(7.16, 7.98)
Reading score at age 10 years, λ_1	4.66	(4.55, 4.77)
Initial change rate, λ_2	.178	(.156, .200)

Note—Age is centered at 10 years. Time-specific errors (Level 1) are assumed to be independent with constant variance across age. CI, confidence interval.

Starting values for the variances and covariances at both levels were set to 1.

To consider the possibility that a simpler model could reasonably describe the data, a second model, Model 1B, was specified in which the quadratic effect was assumed to be constant across individuals. A likelihood ratio test was performed by calculating the difference in deviances for Models 1A and 1B. Model 1B had three fewer parameters because of the restriction that ψ_{20} , ψ_{21} , and ψ_{22} were equal to zero. Assuming the difference in deviances was distributed as chi-square, a difference of 36.1 with 3 *df* suggested that the more complex model, Model 1A, was preferable. This was also supported by the relatively small value for the AIC under Model 1A.

In a third model, Model 1C, the assumption that the time-specific errors were independent with constant variance across age was relaxed and an autocorrelation structure considered in which temporally adjacent errors were allowed to covary, assuming constant covariation across time. Such an error structure was viewed as reasonable because—although individuals were measured at unique ages—the intervals between assessments was assumed to be two years for everyone. Given the additional parameter for the autocorrelation, Model 1C required a total of 11 parameters. A likelihood ratio test suggested the additional parameter was not necessary in improving the overall model fit [$\chi^2(1) = 1, p = .317$]. Relative to the value obtained under Model 1C, the AIC was smallest under Model 1A. MLEs with corresponding 95% confidence intervals for the fixed effects relating to mean change under Model 1A are given in the upper part of Table 2. None of the confidence intervals contained zero as interior points.

While a quadratic growth model may be reasonable in describing reading scores across age in the observed study period, the quadratic function predicts a decrease in performance levels at greater ages. Theoretically, reading performance is generally expected to level off at greater ages. Thus, the quadratic function provides a mismatch to the theoretical expectations of this behavior (e.g., Fran-

cis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996). In a second set of models, a structured latent curve model was considered for the reading data in which the population-level response μ_i was assumed to follow an exponential function across time (Meredith & Tisak, 1990, p. 117):

$$\mu_i = f(\lambda, t) = \lambda_0 - (\lambda_0 - \lambda_1) \exp(-\lambda_2 \text{age}_i),$$

where λ_0 and λ_1 represent the population potential reading score and reading score at age 10 years, respectively, and λ_2 denotes the population change rate in reading scores. Thus, the model assumed mean reading scores approached an asymptote. Based on the function for the mean response, $f(\lambda, t)$, the matrix $\Lambda_i = \Lambda_i(\lambda, \text{age}_i)$ (adapted from Browne & Du Toit, 1991) was specified as shown in the equation at the bottom of the page, where the columns of Λ_i are first-order partial derivatives of $f(\lambda, t)$, with regard to the elements in λ and evaluated across time.

The individual-level response was assumed to be a linear combination of the basis curves in $\Lambda_i(\lambda, \text{age}_i)$ and random weights, η_i , plus error:

$$y_i = \Lambda_i(\lambda, \text{age}_i) \eta_i + \epsilon_i, \tag{9}$$

assuming $\eta_i = \alpha + z_i$, where α , a fixed parameter set, was obtained by solving the set of linear equations $\mathbf{f}(\lambda, \text{age}_i) = \Lambda_i \alpha$. For this function, $\alpha = (\lambda_0, \lambda_1, 0)'$. The time-specific error, ϵ_i , was assumed to be distributed normally with mean equal to zero with symmetric covariance matrix Θ_i . In this first model, Model 2A, the time-specific errors were assumed to be independent with constant variance across time:

$$\epsilon_i \sim N(\mathbf{0}, \Theta_i), \tag{10}$$

as had been done in Model 1A. The random weights in z_i were assumed to be distributed normally with mean equal to zero and symmetric covariance matrix Ψ :

$$\eta_i \sim N(\alpha, \Psi),$$

with α defined earlier. The diagonal elements of Ψ are the variances of the random weights, representing variances

$$\Lambda_i(\lambda, \text{age}_i) = \begin{bmatrix} 1 - \exp\{-\lambda_2 \text{age}_{i1}\} & \exp\{-\lambda_2 \text{age}_{i1}\} & (\lambda_0 - \lambda_1) \text{age}_{i1} \exp\{-\lambda_2 \text{age}_{i1}\} \\ \vdots & \vdots & \vdots \\ 1 - \exp\{-\lambda_2 \text{age}_{iT_i}\} & \exp\{-\lambda_2 \text{age}_{iT_i}\} & (\lambda_0 - \lambda_1) \text{age}_{iT_i} \exp\{-\lambda_2 \text{age}_{iT_i}\} \end{bmatrix},$$

of the characteristics defining change at the individual level. Specifically, ψ_{00} and ψ_{11} are the variances of potential reading scores and scores at age 10 across individuals, respectively, and ψ_{22} is the variance of the change rates across individuals. Under Model 2A, the expected value and covariance matrix of \mathbf{y}_i were

$$\boldsymbol{\mu}_i = \boldsymbol{\Lambda}_i(\boldsymbol{\lambda}, \text{age}_i)\boldsymbol{\alpha}$$

and

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}_i(\boldsymbol{\lambda}, \text{age}_i)\boldsymbol{\Psi}\boldsymbol{\Lambda}_i(\boldsymbol{\lambda}, \text{age}_i)' + \boldsymbol{\Theta}_i,$$

respectively, where $\boldsymbol{\Psi} = \boldsymbol{\Psi}(\boldsymbol{\psi})$ and $\boldsymbol{\Theta}_i = \boldsymbol{\Theta}_i(\boldsymbol{\theta})$. The model required an estimation of 10 parameters.

Following a procedure similar to that in the first set of models, a simpler version of Model 2A was specified, in which the change rate in reading scores was assumed to be constant across individuals. Given this restriction, a modified model, Model 2B, required an estimation of 7 parameters. Assuming the difference in deviances for Models 2A and 2B was distributed as chi-square, a difference of 20.5 with 3 *df* suggested that the more complex model, Model 2A, provided a better fit. Further, the AIC under Model 2A was lower than that under Model 2B. In the last model considered, the assumption that the time-specific errors were independent between occasions was relaxed and an autocorrelation structure considered. In Model 2C, the time-specific errors between temporally adjacent time points were allowed to covary, assuming the covariances were constant across age. One additional parameter was needed to fit Model 2C. Assuming the difference in deviances for Models 2A and 2C was distributed as chi-square, a difference of 0.4 with 1 *df* suggested that the correlation structure for the time-specific errors was not necessary to improve model fit. MLEs and corresponding 95% confidence intervals for parameters of Model 2A are given in the lower part of Table 2. None of the confidence intervals contain zero as interior points.

Fitted mean trajectories based on MLEs for the quadratic and exponential growth models are displayed in Figure 2. Over the study period, the trajectories based on the two models are closely matched. For children aged 6, for ex-

ample, the quadratic growth model estimates a higher performance level than that of the exponential growth model, whereas for children 14 years old, the exponential growth model estimates a higher performance level relative to the quadratic growth model, although these differences are small. The discrepancies between the two models is greatest outside of the observed range. Assuming either model is valid beyond the age span considered, the quadratic growth model would predict an average performance level for an 18-year-old, for example, of 5.33 points, whereas the exponential growth model would predict an average performance level of 6.84 points. To consider the validity of the two models further, future work should consider data that cover a wider age span.

DISCUSSION

The CLMM provides an alternative to the linear latent curve model commonly applied in the social and behavioral sciences. In a linear latent curve model, parameters, fixed or random, enter linearly. Thus, polynomial functions are typically used to characterize longitudinal responses. In contrast, a CLMM allows for fixed parameters to enter a model nonlinearly. Random effects at the second level may only enter linearly. Thus, the CLMM provides a means of specifying a variety of nonlinear models that extend beyond models commonly based on polynomials. Further, the model is quite general, encompassing several common models as special cases, including a hierarchical linear model, linear and nonlinear factor analysis models, and a nonlinear latent curve model. The fact that the CLMM is restricted to situations in which only linear parameters may be random means that standard approaches typically used to estimate linear models may be used. This is a computational benefit, as opposed to a fully nonlinear mixed-effects model that becomes increasingly difficult to estimate with an increase in the number of random effects that enter the model nonlinearly. This is particularly relevant to studies involving multiple longitudinal variables, in which there is a focus on studying the associations between change characteristics in different variables

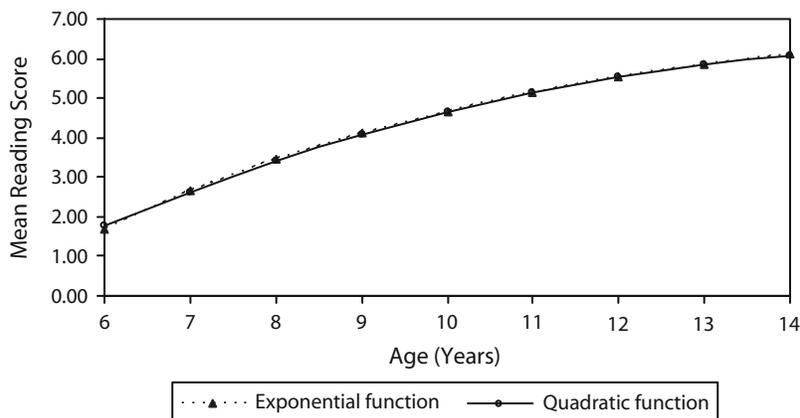


Figure 2. Estimated mean trajectories based on quadratic and exponential growth models.

whose forms of change may be approximated by nonlinear functions (e.g., Blozis, 2004).

The limiting feature of a CLMM that gives the model its computational advantage affects the model's applicability, despite which many nonlinear functions may be specified by the user. That is, many different nonlinear functions may be used to describe a longitudinal response. For instance, in a CLMM described in Blozis and Cudeck (1999), a parameter that appears in the exponent of an exponential growth model must be fixed across individuals. This implies that the growth rate is constant across individuals in the population. There are instances of behaviors that show such patterns (e.g., Blozis & Cudeck); so, although there is a wide range of possibilities in how the individual-level response is characterized, there is this trade-off: Specific parameters must be fixed. A model technically related to the CLMM is the SLCM, a nonlinear latent curve model, under which framework it is possible to accommodate individual differences in change characteristics that enter the model nonlinearly. This is possible because the individual-level response is assumed to be a linear combination of basis curves—that, defined with respect to the mean function, may be nonlinear with regard to their parameters—and weights that vary across individuals. Although limited with regard to the functions that may be specified for the mean response, a SLCM can handle several commonly applied functions, including forms of logistic and exponential functions. Because the SLCM shares the same technical aspect of the CLMM, it is also estimable using procedures commonly applied to strictly linear models.

AUTHOR NOTE

Correspondence concerning this article should be addressed to S. A. Blozis, Psychology Department, One Shields Avenue, University of California, Davis, CA 95616 (e-mail: sablozis@ucdavis.edu).

REFERENCES

- AMEMIYA, Y. (1993). Instrumental variable estimation for nonlinear factor analysis. In C. M. Cuadras & C. R. Rao (Eds.), *Multivariate analysis: Future directions 2* (pp. 113-129). Amsterdam: Elsevier.
- BLOZIS, S. A. (2004). Structured latent curve models for the study of change in multivariate repeated measures. *Psychological Methods, 9*, 334-353.
- BLOZIS, S. A., & CUDECK, R. (1999). Conditionally linear mixed-effects models with latent variable covariates. *Journal of Educational & Behavioral Statistics, 24*, 245-270.
- BROWNE, M. W. (1993). Structured latent curve models. In C. M. Cuadras & C. R. Rao (Eds.), *Multivariate analysis: Future directions 2* (pp. 171-197). Amsterdam: Elsevier.
- BROWNE, M. W., & DU TOIT, S. H. C. (1991). Models for learning data. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 47-68). Washington: American Psychological Association.
- COLLINS, L. M., & GRAHAM, J. W. (2002). The effect of the timing and spacing of observations in longitudinal studies of tobacco and other drug use: Temporal design considerations. *Drug & Alcohol Dependence, 68*, S85-S96.
- CURRAN, P. J. (1997, April). The bridging of quantitative methodology and applied developmental research. In P. J. Curran (Chair), *Comparing three modern approaches to longitudinal data analysis: An examination of a single developmental sample*. Symposium conducted at the 1997 meeting of the Society for Research on Child Development, Washington, DC.
- DAVIDIAN, M., & GILTINAN, D. M. (1995). *Nonlinear models for repeated measurement data*. New York: Chapman & Hall.
- DENNIS, J. E., & SCHNABEL, R. B. (1986). *Numerical methods for unconstrained optimization and nonlinear equations*. Englewood Cliffs, NJ: Prentice Hall.
- ETEZADI-AMOLI, J., & McDONALD, R. P. (1983). A second generation nonlinear factor analysis. *Psychometrika, 48*, 315-342.
- FINKBEINER, C. (1979). Estimation for the multiple factor model when data are missing. *Psychometrika, 44*, 409-420.
- FRANCIS, D. J., SHAYWITZ, S. E., STUEBING, K. K., SHAYWITZ, B. A., & FLETCHER, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology, 88*, 3-17.
- GOLDSTEIN, H., RASBASH, J., PLEWIS, I., DRAPER, D., BROWNE, W., YANG, M., ET AL. (1998). *A user's guide to MLwiN* (Tech. Rep.). London: University of London, Institute of Education.
- HEDEKER, D., & GIBBONS, R. D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods, 2*, 64-78.
- HOX, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.
- JAMSHIDIAN, M., & JENNRICH, R. I. (1997). Acceleration of the EM algorithm using quasi-Newton methods. *Journal of the Royal Statistical Society: Series B, 59*, 569-587.
- JENNRICH, R. I., & SCHLUCHTER, M. D. (1986). Unbalanced repeated measures models with structured covariance matrices. *Biometrics, 42*, 805-820.
- JÖRESKOG, K. G., SÖRBOM, D., DU TOIT, S., & DU TOIT, M. (2001). *LISREL 8: New statistical features*. Lincolnwood, IL: Scientific Software International.
- KREFT, I., & DE LEEUW, J. (1998). *Introducing multilevel modeling*. Thousand Oaks, CA: Sage.
- LAIRD, N. M. (1988). Missing data in longitudinal studies. *Statistics in Medicine, 7*, 305-315.
- LAWLEY, D. N., & MAXWELL, A. E. (1971). *Factor analysis as a statistical method*. New York: Elsevier.
- LINDSTROM, M. J., & BATES, D. M. (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association, 83*, 1014-1022.
- LITTELL, R. C., MILLIKEN, G. A., STROUP, W. W., WOLFINGER, R. D., & SCHABENBERGER, O. (2006). *SAS for mixed models* (2nd ed.). Cary, NC: SAS Institute.
- LITTLE, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association, 90*, 1112-1121.
- LITTLE, R. J. A., & RUBIN, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York: Wiley.
- LONGFORD, N. [T.] (1990). *VARCL: Software for variance component analysis of data with nested random effects (maximum likelihood)*. Princeton, NJ: Educational Testing Service.
- LONGFORD, N. T. (1993). *Random coefficient models*. Oxford: Oxford University Press.
- MACCALLUM, R. C., KIM, C., MALARKEY, W. B., & KIECOLT-GLASER, J. K. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research, 32*, 215-253.
- MATHSOFT, INC. (1997). *S-PLUS user's guide*. Seattle: Author.
- McDONALD, R. P. (1962). A general approach to nonlinear factor analysis. *Psychometrika, 27*, 397-415.
- McDONALD, R. P. (1965). Difficulty factors and nonlinear factor analysis. *British Journal of Mathematical & Statistical Psychology, 18*, 11-23.
- McDONALD, R. P. (1967a). Factor interaction in nonlinear factor analysis. *British Journal of Mathematical & Statistical Psychology, 20*, 205-215.
- McDONALD, R. P. (1967b). Numerical methods for polynomial models in nonlinear factor analysis. *Psychometrika, 32*, 77-112.
- MEREDITH, W., & TISAK, J. (1984). *Tuckerizing curves*. Paper presented at the annual meeting of the Psychometric Society, Santa Barbara, CA.
- MEREDITH, W., & TISAK, J. (1990). Latent curve analysis. *Psychometrika, 55*, 107-122.
- MUTHÉN, L. K., & MUTHÉN, B. O. (1998-2004). *Mplus user's guide* (3rd ed.). Los Angeles: Authors.
- NEALE, M. C., BOKER, S. M., XIE, G., & MAES, H. H. (2003). *Mx: Statistical modeling* (6th ed.). Richmond, VA: Virginia Commonwealth University, Department of Psychiatry.

- RAO, C. R. (1973). *Linear statistical inference and its applications* (2nd ed.). New York: Wiley.
- RAUDENBUSH, S. W., & BRYK, A. S. (2002). *Hierarchical linear models*. Thousand Oaks, CA: Sage.
- RAUDENBUSH, S. W., BRYK, A. S., CHEONG, Y. F., & CONGDON, R. (2004). *HLM 6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.
- RUBIN, D. B. (1976). Inference and missing data. *Biometrika*, **63**, 581-592.
- SHAH, B. V., BARNWELL, B. G., & BIELER, G. S. (1997). *SUDAAN user's manual* (Version 7.5). Research Triangle Park, NC: Research Triangle Institute.
- SINGER, J. D., & WILLETT, J. B. (2003). *Applied longitudinal data analysis*. Oxford: Oxford University Press.
- SNIDERS, T., & BOSKER, R. (1999). *Multilevel analysis*. London: Sage.
- SPSS FOR WINDOWS, REL. 11.0.1. (2001). Chicago: SPSS.
- STATA CORPORATION (1997). *STATA reference manual*. College Station, TX: Stata Press.
- WOLFINGER, R. (1999). Fitting nonlinear mixed models with the new NLMIXED procedure. In *Proceedings of the twenty-fourth annual SAS Users Group International Conference* (paper 287), 1-10. (www2.sas.com/proceedings/sugi24/Stats/p287-24.pdf).
- YALCIN, I., & AMEMIYA, Y. (2001). Nonlinear factor analysis as a statistical method. *Statistical Science*, **16**, 275-294.

APPENDIX A

The gradient vector contains the first-order partial derivatives of a log-likelihood function with respect to parameters in τ :

$$\mathbf{g}(\boldsymbol{\tau}) = \frac{\partial \ln L}{\partial \boldsymbol{\tau}} = (\mathbf{g}'_{\alpha}, \mathbf{g}'_{\lambda}, \mathbf{g}'_{\psi}, \mathbf{g}'_{\theta})',$$

where \mathbf{g}_{α} is the gradient with respect to the coefficient vector $\boldsymbol{\alpha}$, \mathbf{g}_{λ} is the gradient with respect to the coefficient vector $\boldsymbol{\lambda}$, \mathbf{g}_{ψ} is the gradient with respect to the coefficient vector $\boldsymbol{\psi}$, and \mathbf{g}_{θ} is the gradient with respect to the coefficient vector $\boldsymbol{\theta}$. Analytic first-order partial derivatives based on the data model in Equation 1 with mean and covariance structures in Equations 4 and 5, respectively, are provided here. Particular elements of each part of the gradient vector \mathbf{g} require special notation and so are denoted below by subscripts. Specifically, as shown below, $[\mathbf{g}_{\lambda}]_q$ represents the gradient with respect to element q in the parameter vector $\boldsymbol{\lambda}$, $[\mathbf{g}_{\psi}]_v$ represents the gradient with respect to element v in the parameter vector $\boldsymbol{\psi}$, and $[\mathbf{g}_{\theta}]_w$ represents the gradient with respect to element w in the parameter vector $\boldsymbol{\theta}$.

$$\begin{aligned}\mathbf{g}_{\alpha} &= \sum_{i=1}^N \boldsymbol{\Lambda}'_i \mathbf{w}_i, \\ [\mathbf{g}_{\lambda}]_q &= \frac{1}{2} \sum_{i=1}^N \left\{ 2\dot{\boldsymbol{\lambda}}'_q \boldsymbol{\alpha}'_i \mathbf{w}_i + \mathbf{w}'_i \sum_q \dot{\mathbf{w}}_i - \text{tr} \left(\sum_i \sum_q \dot{\mathbf{w}}_i \right) \right\}, \\ [\mathbf{g}_{\psi}]_v &= \frac{1}{2} \sum_{i=1}^N \left\{ \mathbf{w}'_i \dot{\mathbf{w}}_v \mathbf{w}_i - \text{tr} \left(\sum_i \dot{\mathbf{w}}_v \right) \right\}, \\ [\mathbf{g}_{\theta}]_w &= \frac{1}{2} \sum_{i=1}^N \left\{ \mathbf{w}'_i \dot{\boldsymbol{\theta}}_w \mathbf{w}_i - \text{tr} \left(\sum_i \dot{\boldsymbol{\theta}}_w \right) \right\},\end{aligned}$$

where

$$\mathbf{w}_i = \sum_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i), \quad \dot{\boldsymbol{\lambda}}_q = \frac{\partial \boldsymbol{\Lambda}_i}{\partial [\boldsymbol{\lambda}]_q}, \quad \dot{\boldsymbol{\theta}}_w = \frac{\partial \boldsymbol{\Theta}}{\partial [\boldsymbol{\theta}]_w}, \quad \dot{\mathbf{w}}_v = \boldsymbol{\Lambda}_i \frac{\partial \boldsymbol{\Psi}}{\partial [\boldsymbol{\psi}]_v} \boldsymbol{\Lambda}'_i,$$

and

$$\dot{\boldsymbol{\Sigma}} = \dot{\boldsymbol{\Lambda}}_u \boldsymbol{\Psi} \boldsymbol{\Lambda}'_i + \boldsymbol{\Lambda}_i \boldsymbol{\Psi} \dot{\boldsymbol{\Lambda}}'_u.$$

The Hessian matrix,

$$\mathbf{H}(\boldsymbol{\tau}) = \frac{\partial^2 \ln L}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}'},$$

may be approximated by numerical differentiation of $\mathbf{g}(\boldsymbol{\tau})$ (Dennis & Schnabel, 1986; Jamshidian & Jennrich, 1997). This eases the technical burden of specifying the matrix. Let $\kappa_j = \kappa^* \max(|\tau_j|, 1)$, be the scaling term for the j th parameter, where κ^* is a small positive value such as $\kappa^* = 10^{-5}$. The approximation is

$$\mathbf{H}(\boldsymbol{\tau}) \approx \frac{1}{2} [\mathbf{G}(\boldsymbol{\tau}) + \mathbf{G}(\boldsymbol{\tau})'],$$

where columns of $\mathbf{G}(\boldsymbol{\tau})$ are computed from forward differences

$$[\mathbf{G}(\boldsymbol{\tau})]_j = \kappa_j^{-1} [\mathbf{g}(\boldsymbol{\tau} + \kappa_j \mathbf{v}_j) - \mathbf{g}(\boldsymbol{\tau})],$$

where \mathbf{v}_j is a vector with all elements equal to zero except for a single value of unity in the j th position.

APPENDIX B

```

! Mx syntax for fitting a structured latent curve model to reading scores measured up to four
! times over an 8-year period. Mean reading scores are assumed to follow an exponential
! function that includes three parameters, one of which enters the model nonlinearly.
! Age is centered at 10 years
! Age varies by person but equal intervals of 2 years between assessments is assumed for all
#ngroups 3
Group 1:
Calculation
Matrices;
L zero 1 1
Q full 1 1 free
R full 1 1 free
S full 1 1 free
U unit 4 1
X full 4 1
Y symm 3 3 free
Z symm 4 4
End Matrices
! Mean and covariance structures with starting values
! Alpha
ST 6 Q 1 1
ST 1 R 1 1
ST .5 S 1 1
! Phi
ST 1 Y 1 1
ST 1 Y 2 1
ST 1 Y 2 2
ST 1 Y 3 1
ST 1 Y 3 2
ST 1 Y 3 3
! Theta Epsilon
FR Z 1 1 Z 2 2 Z 3 3 Z 4 4
EQ Z 1 1 Z 2 2 Z 3 3 Z 4 4
ST 1 Z 1 1 Z 2 2 Z 3 3 Z 4 4
End Group
Group 2:
Data NI=8 NO=405
VLENGTH file=\srcd_read.dat
Labels read1 read2 read3 read4 age1 age2 age3 age4
Definition age1 age2 age3 age4;
Begin Matrices = Group 1;
End Matrices;
Specify X age1 age2 age3 age4;
Begin Algebra;
G = U - (\exp(-S@X));
H = (\exp(-S@X));
I = (Q-R)@X.(\exp(-S@X));
J = G|H|I;
C = Q_R_L;
End Algebra;
Mean J*C /
Covariance J*Y*J' + Z /
End Group
Group 3:
Data Calculation
MATRIX
P FULL 10 1
COMPUTE P /
SPECIFY P
1 2 3 4 5 6 7 8 9 10
INTERVAL@95 P 1 1
INTERVAL@95 P 2 1
INTERVAL@95 P 3 1
INTERVAL@95 P 4 1

```

APPENDIX B (Continued)

INTERVAL@95 P 5 1
INTERVAL@95 P 6 1
INTERVAL@95 P 7 1
INTERVAL@95 P 8 1
INTERVAL@95 P 9 1
INTERVAL @95 P 10 1
end group

(Manuscript received April 3, 2005;
revision accepted for publication June 11, 2006.)