Psychological Methods
2002, Vol. 7, No. 1, 41–63

1082-989X/02/$5.00 DOI: 10.1037//1082-989X.7.1.41

Multiphase Mixed-Effects Models for Repeated Measures Data

Robert Cudeck
University of Minnesota

Kelli J. Klebe
University of Colorado, Colorado Springs

Behavior that develops in phases may exhibit distinctively different rates of change in one time period than in others. In this article, a mixed-effects model for a response that displays identifiable regimes is reviewed. An interesting component of the model is the change point. In substantive terms, the change point is the time when development switches from one phase to another. In a mixed-effects model, the change point can be a random coefficient. This possibility allows individuals to make the transition from one phase to another at different ages or after different lengths of time in treatment. Two examples are reviewed in detail, both of which can be estimated with software that is widely available.

A wide variety of repeated measures studies and longitudinal investigations are conducted in psychology. The general purpose of these projects is to observe subjects over several occasions to document whether change in behavior occurs, to determine whether change is greater for one subpopulation than another, or to observe whether change can be predicted from background variables. Often the goal of a longitudinal study is not simply to ascertain if subjects change, but instead to gain information about the change process: Is improvement uniform across the period of investigation? Do critical periods exist when development is especially rapid? Is acquisition of a skill during training appreciably different in form for certain subgroups? Describing the change process in quantitative terms has been the subject of a vast amount of methodological and statistical literature.

Many developmental processes exhibit phases, in particular when the time interval in an investigation is not restricted to a brief term. In a learning experiment, for example, there is often a period when subjects show relatively fast improvement and another period when the gain in performance is much slower. The characteristic learning curve for the skill is then composed of an early, rapid acquisition phase that gives way to a later leveling off in improvement as the task is mastered and peak performance is achieved. In the majority of situations the entire behavior can be described in a straightforward algebraic way, even though rate of learning varies over the course of the study. Skill acquisition, as it is investigated in laboratory tasks of great diversity, virtually always demonstrates identifiable phases. Because there are different rates at different periods, a regression model used to account for the response must be nonlinear. Yet because the transition from one phase to another is generally gradual and without abrupt changes, most learning behavior has been summarized with a few functional forms. Lewis (1966, especially chapters 3 and 4) provided a diverse collection of concrete examples. Other developmental behaviors such as language facility, which increases during childhood and levels off in adolescence, can also be effectively described with traditional nonlinear regression models with a few parameters (e.g., Burchinal & Appelbaum, 1991). These processes go through phases, with periods where change is first rapid and then slow, but the functional tools needed to represent the behavior are familiar.

The kind of behavior addressed in this article develops in phases in a way that cannot be summarized satisfactorily by a standard nonlinear regression model of low complexity. This can occur, for ex-

Robert Cudeck, Department of Psychology, University of Minnesota; Kelli J. Klebe, Department of Psychology, University of Colorado, Colorado Springs.

This research was partially supported by National Institute on Alcohol Abuse and Alcoholism Grant AA13458 and National Institute of Mental Health Grant MH5-4576. We are grateful to John J. McArdle for allowing use of data from the Bradway–McArdle longitudinal project, to Paul W. Fox and Thomas J. Bouchard for data from their rotary pursuit experiment, and to Scott Chaiken for data from his experiment with the visual search procedure.

Correspondence concerning this article should be addressed to Robert Cudeck, Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, Minnesota 55455. E-mail: cudeck@umn.edu

ample, because the phases are discontinuous, or because the rate of change in one period is appreciably different than in another, as when a positive trend in performance rapidly changes to negative. To address problems of this kind, a multiphase regression model (Gallant & Fuller, 1973; A. F. M. Smith & Cook, 1980; Seber & Wild, 1989, chapter 9) can be effective. A multiphase model is a combination of two or more submodels, each of which is associated with a distinct subinterval of time. The sections are pieced together to form a more complex whole. Typically the place of transition between sections, called *knots*, or *change points*, are one of the most interesting features of this approach: They mark the time when behavior changes in a notable fashion. Because each piece describes local features of the response, the composite piecewise model can handle shifts in behavior, or even major discontinuities, that are difficult to treat using alternative approaches.

To account for the individual differences that are frequently observed in repeated measures data, the multiphase model can be extended to a more general mixed-effects process (Bryk & Raudenbush, 1992; Crowder & Hand, 1990; Davidian & Giltinan, 1995; Vonesh & Chinchilli, 1997). A mixed-effects model includes a model that pertains to the population, plus distinct submodels for individuals. Individual functions are related to the model for the population profile by using the same response function but with different coefficients for each person. Several authors (Bryk & Raudenbush, 1992, p. 148; Diggle, Liang, & Zeger, 1994, p. 96; Snijders & Bosker, 1999, chapter 12; Verbeke & Molenberghs, 1997, chapter 4; Verbeke & Molenberghs, 2000, chapter 24) have illustrated the use of two-phase linear or quadratic random coefficient models with known change points. An extended version of the same model where the change points may be unknown and estimated in the analysis is reviewed here. The idea, described by Bock (1991); Cudeck (1996); Morrell, Pearson, Carter, and Brant (1995); Pauler and Laird (2000); and Skates, Pauler, and Jacobs (2001), has several noteworthy characteristics. First, the contemporary approach to estimation of mixed-effects models is based on individual data records rather than sufficient statistics such as the mean vector and covariance matrix. Consequently, many forms of missing data, as well as measurement schemes in which different persons are assessed at different times, can be handled efficiently. Participants who would have to be excluded from more traditional analyses because of too few measurements can be included in a mixed-effects model. Even if data for an individual are sparse, multiparameter functions generally can be estimated. Second, longitudinal data may display autocorrelation over the repeated measures or may have heterogeneous variances or other complicated covariance structures. A number of suitable submodels for the variances and covariances can be studied so that realistic covariance patterns can be examined. Third, multiphase functions can be explored based on various combinations of fixed and random effects. Individual differences often are a striking feature of longitudinal studies. Allowing for varying coefficients, even varying change points between phases, seems especially appropriate in the study of behavior.

The nonlinear mixed-effects model is a very general framework for repeated measures data. It is made up of three interconnected subsystems. The *model for the response* describes the change process for a typical subject, specifying algebraically the particular pattern observed in the experiment. Dozens of candidate response models are reviewed in standard textbooks on nonlinear regression (Bates & Watts, 1988; Gallant, 1987; Seber & Wild, 1989). The multiphase model reviewed in this article is one of many formulations for the change in response. The *variability model* sets out the way that data are scattered around an individual's particular function and also specifies how persons vary between one another. This aspect includes assumptions about the form of the distribution and the nature of the variances and possible patterns of correlation, and again a great many possibilities have been considered. Verbeke and Molenberghs (2000, chapters 9 and 10) provided a good review. So-called *level-two covariates* are background measurements on subjects that are included in an analysis to clarify how individuals vary in their development. Basically, the response model and variability model describe individual differences in the change process. Covariates, such as type of experimental treatment, personality or ability measures, features of the environment, or family characteristics, are added as a way of documenting just how the individual differences arise.

Each of these elements—response model, variance and correlation structure, and type and form of covariates—are important in a thorough analysis. Furthermore, decisions regarding one aspect of the model influence and interact with decisions that must be considered in other aspects. For example, even after a particular response model has been specified, different possibilities for the covariance structure usually fit the data quite differently. This makes a thorough setup of a model a complex enterprise. In addition to these scientific components, there are important technical issues concerned with the method of estimation as well as the computational scheme by which the final version is actually implemented.

In this article, the emphasis is on the first item in the above list, the form of the model. Other aspects are treated briefly. This emphasis is not because the other elements are unimportant, but rather because the response model is usually the primary focus of attention. More thorough coverage of other features of the model are reviewed in several recent texts where related matters are covered at length (Brown & Prescott, 1999; Crowder & Hand, 1990; Diggle et al., 1994; du Toit, 1993; Goldstein, 1995; Jones, 1993; Kreft & de Leeuw, 1998; Leyland & Goldstein, 2001; Lindsey, 1999; Longford, 1993; McCulloch & Searle, 2001; Pinheiro & Bates, 2000; Snijders & Bosker, 1999; Verbeke & Molenberghs, 2000; Vonesh & Chinchilli, 1997). The presentation by Davidian and Giltinan (1995) of nonlinear mixed-effects models is especially complete.

Examples of Development in Phases

Three examples are reviewed here to give an idea of data that exhibit phaselike development. They represent problems that arise in rather different fields.

Example 1: Nonverbal Performance Over the Life Span

One of the longest running longitudinal studies ever conducted was begun by Bradway (Bradway & Thompson, 1962) and extended by McArdle (McArdle & Hamagami, 1996). In 1931, 138 children were administered a preliminary version of the Stanford-Binet scale as part of the standardization population for the 1937 test version (Terman & Merrill, 1937). They were retested on as many as five more occasions up to 1992 on both the Stanford-Binet and Wechsler (1958) scales. The Stanford-Binet was administered on Occasions 1 to 4; the Wechsler was given on Occasions 3 to 6. This analysis is based on 74 participants who had at least one score on Occasions 4, 5, or 6. The measurement design is summarized in the left panel of Table 1. Thirty of these individuals were measured on all six occasions. The other 44 had one or two missing scores in six different missing data patterns. For example, the ages of 4 of the children with missing data are listed in the right panel of Table 1. Although there are six occasions in the study, the primary interest is in the relation between age and intelligence. Age was recorded exactly at each occasion; values on the predictor differ for each person, and the independent variable is random.

The response variable is a weighted sum of nonverbal items from the Stanford-Binet plus the performance subscale scores from the Wechsler (Block Design, Picture Completion, Picture Assembly, Object Completion). On the resulting composite score, zero indicates 0% correct on the Wechsler unit weighted subtests. Negative scores in the early occasions reflect the fact that items from the Stanford-Binet are relatively easy compared with those on the Wechsler. Records for a random 25% subsample are shown in Figure 1. It can be seen in the figure that there is an initial rapid growth phase until the early teenage years, at which point performance levels off to become almost linear. In late adulthood most of the participants have declining scores, whereas in a few notable cases participants actually continue to improve. Although the overall pattern is similar for the

Table 1
The Measurement Design for Bradway and McArdle's Longitudinal Study (Left) and Ages of 4 Individuals Who Had Incomplete Data (Right)

	Measurement design			Age			
Occasion	Year	Median age	*N*	Individual 1	Individual 2	Individual 3	Individual 4
1	1931	4.3	74	2.0	2.5	4.3	4.5
2	1941	14.1	74	11.1	12.6	14.2	14.5
3	1956	29.9	74	26.5	28.4	29.9	29.9
4	1969	42.9	56	—	41.6	43.3	—
5	1984	57.7	53	54.2	55.8	—	—
6	1992	66.2	54	—	—	—	67.1

Note. Dashes indicate missing values.

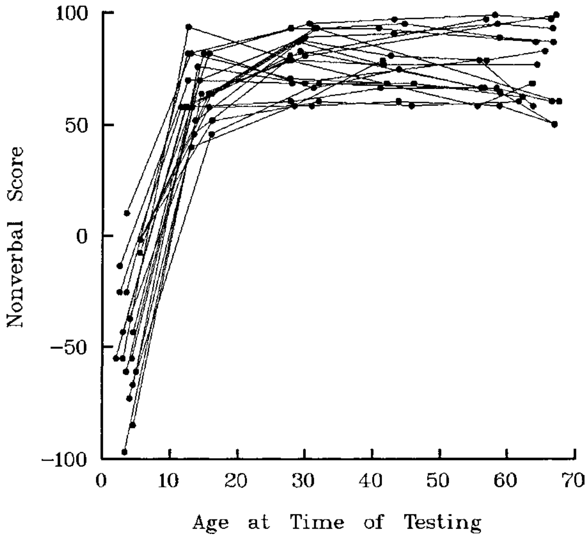

Figure 1. Data from Bradway (Bradway & Thompson, 1962) and McArdle's (McArdle & Hamagami, 1996) study showing growth over the life span on a composite measure of nonverbal intelligence. $N = 74$ for the entire sample (25% random subsample shown).

sample as a whole, there are clear individual differences.

Example 2: Response to Glucose Challenge

A glucose challenge is administered to individuals who may have diabetes to assess their ability to process glucose. Figure 2 shows data that were obtained in an experiment with 20 obese and 13 control patients who took a standard glucose tolerance test (from Zerbe, 1979, p. 219). The measurements are plasma inorganic phosphate, a metabolite whose concentration varies inversely with glucose level. In the early stage of the reaction, phosphate level rapidly declines, whereas in the later phase it increases gradually. The elapsed time at which the change in pattern occurs differs between the groups. It is approximately 2 hr after the beginning of the experiment for obese patients but about 1.5 hr for control patients. A quadratic model could give an adequate summary of these data. On the other hand, the pattern of rapid decline and quick rebound is so sharp that the overall reaction appears like two distinct linear phases.

Example 3: Skill Acquisition on a Rotary Pursuit Task

The *rotary pursuit task* is a classic experimental procedure designed to assess learning in a domain that is unaffected by verbal, spatial, or quantitative abilities. In the rotary pursuit task, participants learn to track a point on a rotating disk with a stylus. Time on target over trials of the experiment is the dependent variable. The left side of Figure 3 shows performance on the rotary pursuit task for a 10% subsample of the 242 individuals who participated in an experiment during sessions that were conducted on 3 different days (from Fox, Hershberger, & Bouchard, 1996). Seventy-five trials were summarized into 15 blocks of 5 trials each. The right side of Figure 3 shows the mean performance over the experiment for all 242 individuals. The most salient feature of the figure is the jump in performance between Days 1 and 2 and between Days 2 and 3. This kind of spontaneous improvement is called *reminiscence* (Payne, 1987). It is

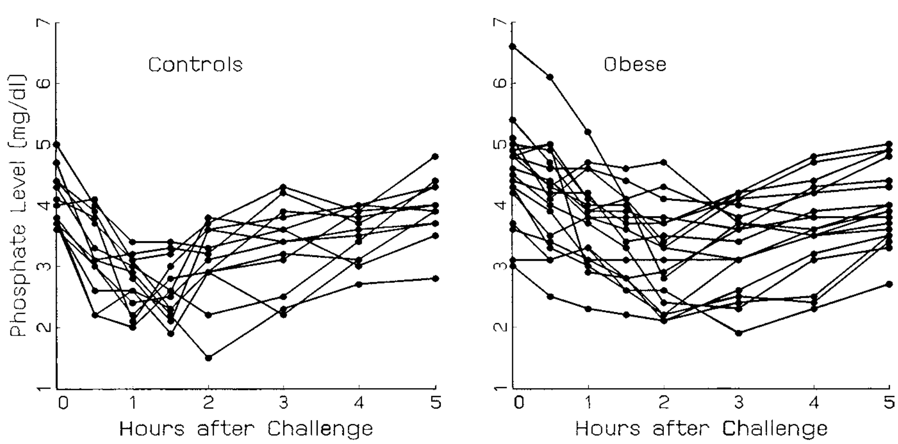

Figure 2. Plasma phosphate levels of 20 control participants (left) and 13 patients with obesity (right) collected during a glucose challenge.

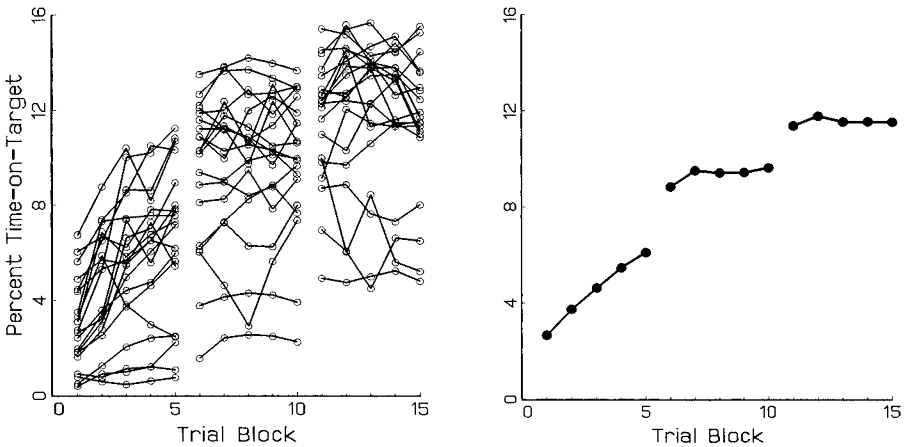

Figure 3. Fifteen trial blocks over 3 days on the rotary pursuit task. Experiment is based on 242 participants. A 10% sample of the individual records is shown on the left; means from the total sample are shown on the right.

interesting because participants appear to learn even when they are not practicing. The reminiscence effect highlights the phaselike appearance of the acquisition. Because the behavior is discontinuous between days, no single simple function would work well here.

Multiphase Models

The principal attraction of multiphase models is their wide applicability. A consequence of this flexibility is that there are many choices regarding the form of the model and the restrictions that may be imposed. The main issues are (a) the number of segments, (b) the algebraic form of the submodels, and (c) the degree of smoothness at the change points. Although we have studied problems involving several phases, most examples require only two or three segments. Consequently, we focus on the basic two- and three-phase structures, assuming that extensions to four or more segments are apparent. The focus in this section is the description of data from 1 individual. The following section reviews the way that data from many subjects are combined in the mixed-effects model.

Let y_{ij} denote the measurement of the ith subject on the jth occasion, with $j = 1, \ldots, n$. The occasion of measurement or the elapsed time from the beginning of the study to the jth assessment is x_j. The collection for an individual is $\mathbf{x} = (x_1, \ldots, x_n)'$. The model we consider is

$$y_{ij} = f_j(\mathbf{x}, \boldsymbol{\beta}) + e_{ij} \quad j = 1, \ldots, n \tag{1}$$

in which the fitted component, $f_j(\mathbf{x}, \boldsymbol{\beta})$, has sections $g_k(\mathbf{x}, \boldsymbol{\beta})$, $k = 1, \ldots, K + 1$. Specifically, for $K + 1 = 3$ segments the systematic part of the model is

$$f_j(\mathbf{x}, \boldsymbol{\beta}) = \begin{cases} g_1(\mathbf{x}, \boldsymbol{\beta}) & x_j < \tau_1 \\ g_2(\mathbf{x}, \boldsymbol{\beta}) & \tau_1 \le x_j < \tau_2 \\ g_3(\mathbf{x}, \boldsymbol{\beta}) & \tau_2 \le x_j \end{cases} \tag{2}$$

The residuals are $\mathbf{e}_i = (e_{i1}, \ldots, e_{in})'$ and $\boldsymbol{\beta}$ $(p \times 1)$ are the coefficients specified for all of the segments. The submodels of each phase can be any nonlinear model, although in practice quite simple functions are generally adequate. The values τ_k, $k = 1, \ldots, K$, are the change points of the function.

To give some idea of the flexibility of the approach, Figure 4 shows examples using two phases. Curve A exhibits linear growth followed by linear decline. Within each phase, the rate of change is constant. It is the combination of both pieces that makes the process nonlinear. Curve B has linear increase followed by extremely rapid exponential increase. Development improves gradually up to a point, after which the process changes appreciably, so much so that it appears to be different in kind. Curve C combines cubic and quadratic sections, producing a complicated pattern with quite different rates of change at several occasions plus two periods where the increase essentially stops.

For the sake of concreteness, consider the linear–linear model with change at τ. Suppose measurements are taken at the beginning of the study and at Months 3, 6, 9, and 12. Then $\mathbf{x} = (0,3,6,9,12)'$. Suppose from

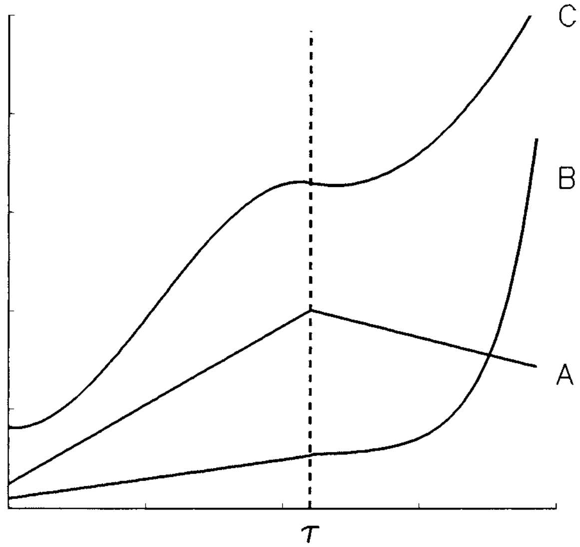

Figure 4. Examples of two-piece models with change point at τ. Models have phases composed of sections that are (A) linear–linear, (B) linear–exponential, and (C) cubic–quadratic.

previous research it is known that the change point occurs at month $\tau = 8$. Then, in the first three occasions, the process uses $\alpha_0 + \alpha_1 x_j$; after that time, the model is $\gamma_0 + \gamma_1 x_j$. Combining parameters gives the vector $\boldsymbol{\beta} = (\alpha_0, \alpha_1, \gamma_0, \gamma_1)'$. The overall setup is

$$f_j(\mathbf{x}, \boldsymbol{\beta}) = \begin{cases} \alpha_0 + \alpha_1 x_j & x_j \leq \tau \\ \gamma_0 + \gamma_2 x_j & \tau < x_j \end{cases} \tag{3}$$

where $g_1(\mathbf{x}, \boldsymbol{\beta}) = \alpha_0 + \alpha_1 x_j$ and $g_2(\mathbf{x}, \boldsymbol{\beta}) = \gamma_0 + \gamma_1 x_j$. In many experiments, the change point is unknown and must be estimated. When this occurs, τ is included in the list of parameters, $\boldsymbol{\beta} = (\alpha_0, \alpha_1, \gamma_0, \gamma_1, \tau)'$. The estimation problem becomes more complicated if τ is unknown, but the conceptual idea otherwise is identical to the situation where τ is specified beforehand.

Behavior at the Change Point

The model switches from one segment to the next at the change points. With some problems, the transition is gradual so that even though there is a shift in function, the changeover to the new section is steady and incremental. In other problems, the change is abrupt, producing a pronounced jump at τ_k. Examples 1 and 3 above (in the *Nonverbal Performance Over the Life Span* and *Skill Acquisition on a Rotary Pursuit Task* sections) demonstrate these possibilities. To handle such different situations, one must specify conditions regarding behavior of the function at the transition. The three conditions are illustrated by the functions in Figure 5.

In standard linear or even nonlinear regression models, there is nothing exactly comparable to these conditions. In deciding how to handle each issue, a researcher tailors the model so that it is most appropriate for the situation under investigation. These options cannot be processed in a mechanical way, and it is not possible to state how one should treat them in general. The conditions reflect important features of the developmental process. How they are handled requires knowledge about, or makes assumptions regarding, the behavior in question.

1. Segments are discontinuous at τ_k: In some situations the shift between segments of the function is accompanied by a distinct increase or decrease in the response. At the change point $x_j = \tau_k$, the difference in predicted values is given by

$$\delta_k = g_k(\tau_k, \boldsymbol{\beta}) = g_{k+1}(\tau_k, \boldsymbol{\beta}) \tag{4}$$

This is illustrated in Function A of Figure 5. The discontinuity may be interesting substantively as a measure of the reaction that accompanies the change in function. A clear example is the reminiscence effect in Figure 3 where the discontinuity indicates the amount of learning that has occurred between practice

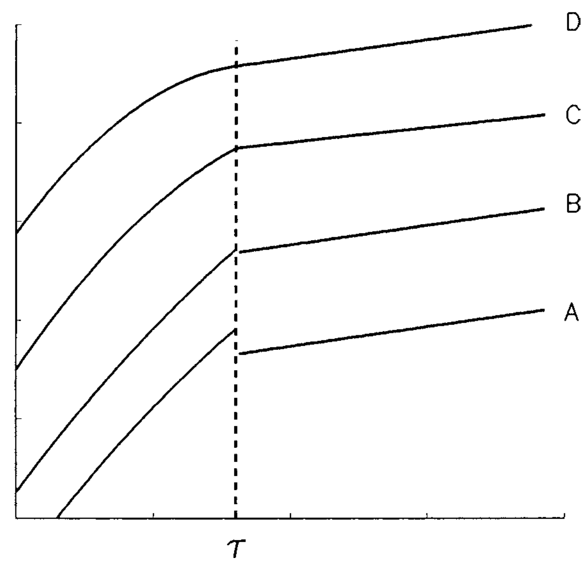

Figure 5. Behavior at the change point of various two-phase models: (A) linear–linear model with discontinuous segments; (B) linear–linear model with continuity between segments; (C) quadratic–linear model with continuity between segments; (D) quadratic–linear model in which segments are continuous and first derivatives are equal.

sessions. Consequently, it can be useful to examine the discontinuity formally. This can be done directly by testing the hypothesis H_0: $\delta_k = 0$. Alternatively, one may evaluate the fit of the full model in which the segments are unconstrained at τ to the fit of a reduced model (reviewed next) in which the segments are restricted to join at τ.

2. Segments are continuous at τ_k: If the transition from one phase to another is continuous, then at τ_k the sections join:

$$g_k(\tau_k, \boldsymbol{\beta}) = g_{k+1}(\tau_k, \boldsymbol{\beta}) \tag{5}$$

Function B in Figure 5 is an illustration based on two linear phases as in model (3). If the segments are continuous at τ, then (5) requires that

$$\alpha_0 + \alpha_1\tau = \gamma_0 + \gamma_1\tau$$

When this condition is true, one of the parameters is redundant and can be solved in terms of the others. There are usually several ways to do this. An investigator can determine which specific approach to take based on the problem at hand. For example, in the linear–linear model the second intercept can be computed from the others as $\gamma_0 = \alpha_0 + \alpha_1\tau - \gamma_1\tau$. Then the original five-parameter model of (3) becomes a four-parameter system

$$f_j(\mathbf{x}, \boldsymbol{\beta}) = \begin{cases} \alpha_0 + \alpha_1 x_j & x_j \leq \tau \\ \alpha_0 + \alpha_1\tau + \gamma_1(x_j - \tau) & \tau < x_j \end{cases} \tag{6}$$

with a reduced set of coefficients $\boldsymbol{\beta} = (\alpha_0, \alpha_1, \gamma_1, \tau)'$.

3. Segments are continuous, and first derivatives are equal at τ_k: Even if the function is continuous at τ_k, the transition from one segment to another need not be smooth. This is illustrated in the example of Figure 5C where at τ the derivatives of the two segments are different. The segments join, but the connection is like a corner. To make the change smooth, both the function values and the rates of change must be equal at τ_k. The new requirement is that at τ_k the derivatives be equal:

$$\frac{dg_k(\tau_k, \boldsymbol{\beta})}{d\tau_k} = \frac{dg_{k+1}(\tau_k, \boldsymbol{\beta})}{d\tau_k} \tag{7}$$

When both of the constraints (5) and (7) apply, then two of the parameters are redundant. The function shown in Figure 5D illustrates the effect of these conditions on a model that is composed of quadratic and linear sections. The original six-parameter function, again assuming the change point is a coefficient to estimate so that $\boldsymbol{\beta} = (\alpha_0, \alpha_1, \alpha_2, \gamma_0, \gamma_1, \tau)'$, is

$$f_j(\mathbf{x}, \boldsymbol{\beta}) = \begin{cases} \alpha_0 + \alpha_1 x_j + \alpha_2 x_j^2 & x_j \leq \tau \\ \gamma_0 + \gamma_1 x_j & \tau < x_j \end{cases} \tag{8}$$

From (5), $\alpha_0 + \alpha_1\tau + \alpha_2\tau^2 = \gamma_0 + \gamma_1\tau$. One can solve for the first intercept in terms of the other coefficients as

$$\alpha_0 = \gamma_0 + \gamma_1\tau - \alpha_1\tau - \alpha_2\tau^2 \tag{9}$$

From (7), $\alpha_1 + 2\alpha_2\tau = \gamma_1$. One can solve for the linear coefficient of the first segment in terms of the others:

$$\alpha_1 = \gamma_1 - 2\alpha_2\tau \tag{10}$$

Then a version of (8) that is continuous at τ and also has a smooth transition between phases is the following four-parameter function:

$$f_j(\mathbf{x}, \boldsymbol{\beta}) = \begin{cases} \gamma_0 + \gamma_1 \mathrm{x_j} + \alpha_2(\mathrm{x_j} - \tau)^2 & \mathrm{x_j} \leq \tau \\ \gamma_0 + \gamma_1 \mathrm{x_j} & \tau < \mathrm{x_j} \end{cases} \tag{11}$$

The reduced set of parameters consists of $\boldsymbol{\beta} = (\alpha_2, \gamma_0, \gamma_1, \tau)'$.

An Alternative Representation

A multiphase model is a composite of segments; (2) is composed of three parts but is really a single function. For theoretical and computational purposes, it is sometimes convenient to rewrite the model as one expression. This is straightforward to do when the phases are polynomials, so that the basic arrangement is a kind of regression spline (Durrleman & Simon, 1989; Seber & Wild, 1989; P. L. Smith, 1979). The sections of (2) can be combined by means of an operator called the "+" function. The value of this approach is that it allows the function to be expressed in several different but equivalent ways. As in the previous section, there is no single form that is always preferable. One form may be especially useful because it permits an effect to be tested directly or has parameters that are interpretable. For example, the most general version of the two-piece linear–linear model is (3), where a discontinuity between segments is allowed. In some situations it can be valuable to use an equivalent form of the linear–linear model in which the amount of discontinuity appears as a parameter. One could then compare individuals on the magnitude of the discontinuity or summarize a population by the mean discontinuity. The version of the model in (3) does not contain this information explicitly, but it can be attractive to have in some situations. To represent alternative versions of the model, we may use a helpful operator.

Suppose the range of a function u is defined for both positive and negative values. The "+" function is a truncation operation:

$$u_+ = \begin{cases} 0 & u \le 0 \\ u & u > 0 \end{cases}$$

A special rule is adopted for the situation where u is raised to the power zero. The convention is

$$u_+^0 = 0 \text{ if } u \le 0 \quad u_+^0 = 1 \text{ if } u > 0$$

In splines the "+" function is used with differences. Here are some examples:

$$\begin{aligned} &(4-5)_+ = 0 \quad (9-2)_+ = 7 \quad (6-3)_+^2 = 9 \\ &(3-1)_+^0 = 1 \quad (1-3)_+^0 = 0 \end{aligned}$$

As an illustration of the operator, consider the general linear–linear process with discontinuity of (3). It can be combined into a single function of five parameters $\boldsymbol{\beta} = (\alpha_0, \alpha_1, \delta_0, \delta_1, \tau)'$ as

$$f_j(\mathbf{x}, \boldsymbol{\beta}) = \alpha_0 + \alpha_1 x_j + \delta_0 (x_j - \tau)_+^0 + \delta_1 (x_j - \tau)_+ \tag{12}$$

To see the correspondence with (3) note that for $x_j \le \tau$ the last two terms in (12) are zero, and the result is obviously the first segment of (3). When $x_j > \tau$, (12) is

$$\begin{aligned} f_j &= \alpha_0 + \alpha_1 x_j + \delta_0 + \delta_1 x_j - \delta_1 \tau \\ &= (\alpha_0 + \delta_0 - \delta_1 \tau) + (\alpha_1 + \delta_1) x_j \\ &= \gamma_0 + \gamma_1 x_j \end{aligned}$$

where $\gamma_0 = \alpha_0 + \delta_0 - \delta_1 \tau$ and $\gamma_1 = \alpha_1 + \delta_1$. This means that δ_1 is the difference in slopes between the segments,

$$\delta_1 = \gamma_1 - \alpha_1$$

and δ_0, as defined in (4), is the difference in function values at τ:

$$\begin{aligned} \delta_0 &= \gamma_0 - \alpha_0 + (\gamma_1 - \alpha_1)\tau \\ &= \gamma_0 - \alpha_0 + (\gamma_1 - \alpha_1)\tau \\ &= (\gamma_0 + \gamma_1 \tau) - (\alpha_0 + \alpha_1 \tau) \end{aligned}$$

Consequently, $\delta_0 = 0$ implies that the segments are continuous at τ. To test formally for continuity, one can construct a confidence interval for δ_0 and delete the term if the interval includes zero. To impose continuity, set $\delta_0 = 0$ by using (12) with the third term excluded. This results in an equivalent form of model (6) with parameters $\boldsymbol{\beta} = (\alpha_0, \alpha_1, \delta_1, \tau)'$.

In general, if there are $K + 1$ segments with K change points, each segment a polynomial of degree D, and no continuity restrictions at τ_k, the multiphase model can be written (from P. L. Smith, 1979) as

$$f_j = \sum_{m=0}^{D} \gamma_{0m} x_j^m \sum_{k=1}^{K} \sum_{m=0}^{D} \gamma_{km} (x_j - \tau_k)_+^m$$

If terms $\gamma_{k0}\,(x_j - \tau_k)_+^0$ are excluded, then the segments join at τ_k. And if terms $\gamma_{k1}(x_j - \tau_k)_+$ are also excluded, then the derivatives of the segments adjoining τ_k are equal as well. For example, the $K + 1 = 3$ phase quadratic that is continuous at the $K = 2$ change points with continuous first derivatives is

$$\begin{aligned} f_j(\mathbf{x}, \boldsymbol{\beta}) = {} & \gamma_{00} x_j^0 + \gamma_{01} x_j^1 + \gamma_{02} x_j^2 \\ & + \gamma_{12}(x_j - \tau_1)_+^2 + \gamma_{22}(x_j - \tau_2)_+^2 \end{aligned}$$

with $\boldsymbol{\beta} = (\gamma_{00}, \ldots, \gamma_{22}, \tau_1, \tau_2)'$. This is a complex seven-parameter model. The merit of the function is that it would be appropriate for an equally complex behavioral process that develops in three distinctive phases.

This development means that it is usually possible to write a given multiphase model in several ways. For example, the two-phase quadratic–linear model with first- and second-order continuity can be written as (8) with the constraints on α_0 and α_1 given by (9) and (10). A second form uses (11), where the constraints are substituted into the first segment. Third, note by means of the "+" function that (11) is compactly

$$f_j(\mathbf{x}, \boldsymbol{\beta}) = \gamma_0 + \gamma_1 x_j + \alpha_2 (\tau - x_j)_+^2 \tag{13}$$

In this particular instance there is no compelling reason to pick one form rather than the others. The version based on (8) with (9) and (10) makes both the two-phase nature of the model as well as the constraints explicit; the version in (13) is compact and emphasizes that the piecewise collection of segments is one unified model. In other situations, in particular when there is interest in evaluating the magnitude of discontinuity at a change point, one specific form may be preferable to other versions.

In summary, the basic multiphase model for repeated measures from one individual is defined in (2) as a collection of several simpler submodels. Each component of the composite applies to a section of the complete interval. The submodels are most frequently low-order polynomials such as the quadratic, cubic, or linear but many other functions can be used. Depending on the characteristics of the experiment, the model can be further tailored by allowing a discontinuity at the change point or by restricting the function to be continuous at the change point with possible continuous first derivatives. Equivalent forms of the model can be specified, and an interesting version is obtained by means of the "+" function. This flexibility allows the model to be adapted directly for the problem at hand. A number of different repeated measures processes can be accommodated with this framework.

Example With Data From One Individual

Although most studies that gather repeated measures are based on a sample of cases, there are instances when a single individual is investigated, and the analysis proceeds exactly as outlined above. One could also use the model in the form presented here when the data set consists of a single response from each of several individuals. This is the case discussed in detail by Seber and Wild (1989). To illustrate, we briefly examine data obtained in a series of experiments conducted by Chaiken (1994). The response is reaction time required to judge whether a geometrical object presented on a display device can be identified subsequently among several other stimuli—a procedure known as the *visual search task.* There were 24 trials. As participants learn this task, there often is a rapid decrease in reaction time followed by a period of more gradual and nearly linear improvement. Consequently the model of (11), a two-phase quadratic–linear structure with first- and second-order continuity, seems appropriate. Appendix A shows the data and setup for estimating the model for one case by least squares with the SAS nonlinear regression program PROC NLIN. To make the constraints of this model as clear as possible, the form used in the computer code is (8) with (9) and (10). Estimates and corresponding standard errors are

$$\hat{\alpha}_2 = 0.048(0.01) \quad \hat{\gamma}_0 = 1.34(0.11)$$
$$\hat{\gamma}_1 = -0.018(0.007) \quad \hat{\tau} = 6.54(0.74)$$

The change between the two phases is estimated to occur midway between Trials 6 and 7 at $\hat{\tau} = 6.54$. The rate in the second phase shows gradual improvement with a decrease in reaction time of $\hat{\gamma}_1 = -0.018$ per trial. The data and fitted function are plotted in Figure 6.

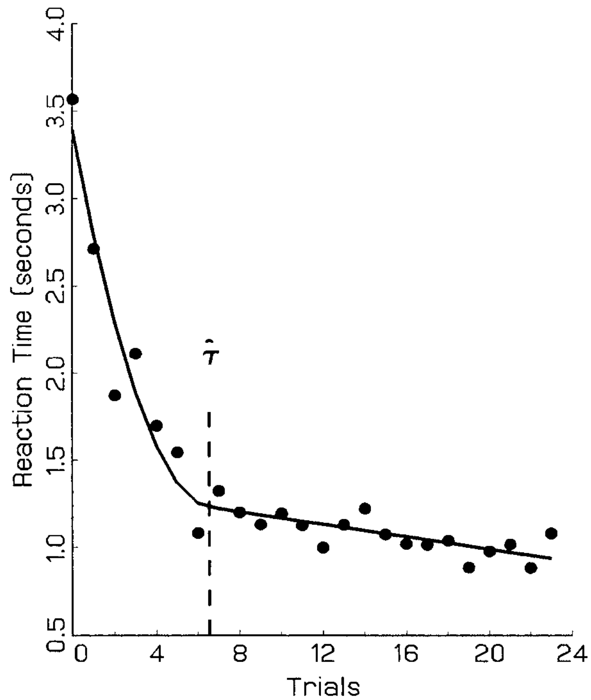

Figure 6. Data from one individual on the visual search task. Fitted model is the two-phase, quadratic–linear structure with first- and second-order continuity. Estimated change point is at $\hat{\tau} = 6.5$.

Extension to a Mixed-Effects Model

The multiphase model introduced to this point specifies a change process that would apply to repeated measures obtained from a single individual or to the case where one measurement is taken on many members of a sample. It can be generalized in several useful ways to handle the complexity that is inherent in repeated measure studies when members of a population are under investigation. The main ideas that are part of the mixed-effects model system are, first of all, regression coefficients that vary over individuals. The second element is the summarization of the varying coefficients into an average set of coefficients and their covariance matrix, and also of the regression residuals. This describes the between- and within-subject variability. The third feature is a flexible measurement system that accommodates missing data, unbalanced designs, and random assessment.

Random Coefficients

A mixed-effects model for repeated measures is based on a function describing the change process over time. The function employs a set of parameters and includes one or more covariates. Consider model (6) as a simple example with parameters $\boldsymbol{\beta} = (\alpha_0, \alpha_1, \gamma_1, \tau)$ and covariate $\mathbf{x}$. The key idea of the mixed-effects model is that each individual in the population has a unique set of regression coefficients that are optimal for predicting their repeated measures utilizing this common function. Parameters are quantities of statistical distributions that pertain to a population. When an individual is the focus of attention and the goal is to summarize their particular data by the common function, the coefficients of the function are known as *stochastic parameters*, or *random coefficients*. They are unknown quantities that can be "estimated," but because they apply to one particular individual they are not population parameters in the more familiar sense. In the analysis of the visual search data of the previous section, the values "estimated" from the data are the stochastic parameters for that person. Again with model (6) as the example, the individual coefficients are denoted

$$\boldsymbol{\beta}_i = (\alpha_{0i}, \alpha_{1i}, \gamma_{1i}, \tau_i)'$$

Then (1) is written as a process exclusively for the *i*th individual as

$$\mathbf{y}_i = \mathbf{f}(\mathbf{x}, \boldsymbol{\beta}_i) + \mathbf{e}_i \tag{14}$$

where the order of $\mathbf{y}_i$, $\mathbf{f}(\mathbf{x}, \boldsymbol{\beta}_i)$, and $\mathbf{e}_i$ is $(n \times 1)$, while $\boldsymbol{\beta}_i$ is a $p \times 1$ vector.

Because $\boldsymbol{\beta}_i$ varies from person to person, the systematic part of the model, $\mathbf{f}(\mathbf{x}, \boldsymbol{\beta}_i)$, varies as well. In fact, because the change point is random, the shift from one segment to another can occur at quite different times for different persons. Suppose $n = 5$ measurements are taken at occasions $\mathbf{x} = (0,3,6,9,12)'$, and the segments are continuous at the change point. If $\tau_i = 8$, then the assessments y_{i1} to y_{i3} follow the first segment whereas y_{i4} and y_{i5} follow the second. The arrangement is

$$\begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \\ y_{i4} \\ y_{i5} \end{pmatrix} = \begin{pmatrix} \alpha_{i0} \\ \alpha_{i0} + 3\alpha_{i1} \\ \alpha_{i0} + 6\alpha_{i1} \\ \alpha_{i0} + 9\alpha_{i1} + \gamma_{i1}(9 - \tau_i) \\ \alpha_{i0} + 12\alpha_{i1} + \gamma_{i1}(12 - \tau_i) \end{pmatrix} + \begin{pmatrix} e_{i1} \\ e_{i2} \\ e_{i3} \\ e_{i4} \\ e_{i5} \end{pmatrix}$$
$$\mathbf{y}_i = \mathbf{f}(\mathbf{x}, \boldsymbol{\beta}_i) + \mathbf{e}_i$$

Of course, the fitted values $\mathbf{f}(\mathbf{x}, \boldsymbol{\beta}_i)$ might vary appreciably for another person who had different elements in $\boldsymbol{\beta}_i$. This is particularly evident as τ_i varies.

Between- and Within-Subject Variability

The random coefficients and residuals, $\boldsymbol{\beta}_i$ and $\mathbf{e}_i$, are individual-level variables that vary from person to person. As with any set of statistical variates, an important part of the model is their distribution, mean vector, and covariance matrix. For the sake of concreteness, one can imagine a table with coefficients for observations arranged by rows as in the upper section of Table 2. The most common and convenient assumption about the distribution of the random coefficients is that they are multivariate normal. Represented in the bottom of the table are the associated mean vector and covariance matrix. They are parameter matrices of order $p \times 1$ and $p \times p$, respectively, defined as

$$\boldsymbol{\beta} = E(\boldsymbol{\beta}_i) \qquad \boldsymbol{\Phi} = cov(\boldsymbol{\beta}_i)$$

Elements of $\boldsymbol{\beta}$ are interpreted as the coefficients of the response curve for the "average individual" (cf. Zeger, Liang, & Albert, 1988). Diagonal elements of $\boldsymbol{\Phi}$ describe the variability of each of the random coefficients, whereas off-diagonal elements describe the

Table 2
Random Coefficients for Each Individual (Stochastic Parameters) With the Mean Vector and Covariance Matrix (Population Parameters)

Individual	α_{i0}	α_{i1}	γ_{i1}	τ_i
1	α_{10}	α_{11}	γ_{11}	τ_1
2	α_{20}	α_{21}	γ_{21}	τ_2
3	α_{30}	α_{31}	γ_{31}	τ_3
⋮	⋮	⋮	⋮	⋮
Mean vector ($\boldsymbol{\beta}$)	$E(\alpha_{i0})$	$E(\alpha_{i1})$	$E(\gamma_{i1})$	$E(\tau_i)$
Covariance matrix ($\boldsymbol{\Phi}$)	$\text{var}(\alpha_{i0})$			
	$\text{cov}(\alpha_{i1}, \alpha_{i0})$	$\text{var}(\alpha_{i1})$		
	$\text{cov}(\gamma_{i1}, \alpha_{i0})$	$\text{cov}(\gamma_{i1}, \alpha_{i1})$	$\text{var}(\gamma_{i1})$	
	$\text{cov}(\tau_i, \alpha_{i0})$	$\text{cov}(\tau_i, \alpha_{i1}$	$\text{cov}(\tau_i, \gamma_{i1})$	$\text{var}(\tau_i)$

Note. For the sake of concreteness, the coefficients of model (6) have been assumed.

extent to which pairs of random coefficients covary. The shorthand way to summarize these results is by writing $\boldsymbol{\beta}_i \sim N(\boldsymbol{\beta}, \boldsymbol{\Phi})$.

The residuals, $\mathbf{e}_i$, also vary from person to person. They could be represented schematically as in Table 2 with columns marked for each term $e_{i1}, \ldots, e_{in}$. As with $\boldsymbol{\beta}_i$, the most common assumption is that $\mathbf{e}_i$ has a normal distribution with mean vector of zero and covariance matrix $\boldsymbol{\Lambda}$:

$$\mathbf{e}_i \sim N(\mathbf{0}, \boldsymbol{\Lambda})$$

Many different structures have been suggested to describe the pattern of variances and correlations among the residuals. When the number of repeated measures is large, there are many terms in $\mathbf{e}_i$ and the number of elements in the $n \times n$ matrix $\boldsymbol{\Lambda}$ is large. In practice one attempts to summarize $\boldsymbol{\Lambda}$ in a way that is reasonable for the problem at hand, but to do so parsimoniously. Some common forms are listed in Table 3. Other alternatives are reviewed in, for example, Davidian and Giltinan (1993), Verbeke and Molenberghs (2000, chapter 10), and Vonesh and Chinchilli (1997, chapter 7).

In the literature on experimental design and multilevel models (e.g., Kreft & de Leeuw, 1998; McCulloch & Searle, 2001), the distribution of residuals is known as within-subject, or Level 1 variability. This is because this distribution describes the fluctuations of an individual's repeated measures around their particular response function. The distribution of the random coefficients is known as between-subjects or Level 2 variability.

Missing Data, Unbalanced and Random Measurements

One reason for the popularity of mixed-effects models is that they can produce accurate estimates of the change process even in the presence of missing data. In longitudinal studies where a sample may be followed for many years, the likelihood is high that some participants will not be observed at all planned occasions. A statistical method that produces unbiased estimates in the presence of missing data is obviously valuable.

This feature can be achieved if the process that generates the missing data is not systematically related to the unobserved values had they actually been recorded (Little & Rubin, 1987). An example where there is a systematic relationship between the data that are missing and the process that produces the missingness arises in studies of health functioning in older people (Mirowsky & Reynolds, 2000). Individuals who have chronic health problems tend to miss scheduled appointments because of general discomfort or the effort required to travel to the research site. Because of self-selection and differential dropout, the data that are recorded may tend to be from those who are healthy enough to participate, whereas others who are less healthy are increasingly underrepresented over time. Information that is collected is biased in the direction of greater health than is true of the population. In contrast, if missing data were missing only because of logistical problems on the part of the experimenters, which would affect the more and less healthy equally often, then the resulting sample data would be unbiased with respect to the population.

When missing data are not systematically related to the unobserved scores, a mixed-effects model can utilize available information completely and do so without bias in estimates. In this situation the missing data mechanism is said to be *ignorable* for the mixed-effects model. When there is a systematic relationship between missing data and unobserved scores, then the missing data problem is *nonignorable.* The statistical literature regarding missing data and possible remedies for the latter situation is extensive and is an order of magnitude more complex than when missing data are ignorable (cf. Hedeker & Gibbons, 1997). The presentation by Verbeke and Molenberghs (2000, chapters 14–21) is a convenient summary on many of these issues. In the remainder of this article, we assume ignorable missing data for the examples where information is incomplete.

Missing data occur when measurements that were originally planned are not actually collected. Suppose a design calls for assessments to be made at 0, 6, 12, and 24 months after a study begins. The temporal spacing between assessments is unequal; however, if all individuals are observed at each of the four occasions, then the design is said to be *balanced* (e.g., McCulloch & Searle, 2001, especially chapter 7). If one or more planned measurements are not collected for any participant, then the missing data cause the design to be *unbalanced.*

In addition to missing data situations, unbalanced designs arise in another context. It often happens that some individuals are measured earlier or later than stipulated by a protocol because of scheduling conflicts or other difficulties in implementation. If the effect of a treatment is time dependent, then it is important to use the actual date at which the outcome variable is measured so that the relation between be-

Table 3
Representative Covariance Structures for the Residuals, $\mathbf{e}_i$

Name	Covariance structure	No. of parameters required
Sphericity	$\begin{pmatrix} \varphi^2 & & & \\ 0 & \varphi^2 & & \\ 0 & 0 & \varphi^2 & \\ 0 & 0 & 0 & \varphi^2 \end{pmatrix}$	1
Diagonal	$\begin{pmatrix} \varphi_1^2 & & & \\ 0 & \varphi_2^2 & & \\ 0 & 0 & \varphi_3^2 & \\ 0 & 0 & 0 & \varphi_4^2 \end{pmatrix}$	n
Compound symmetry	$\begin{pmatrix} \varphi_1^2 & & & \\ \varphi_2 & \varphi_1^2 & & \\ \varphi_2 & \varphi_2 & \varphi_1^2 & \\ \varphi_2 & \varphi_2 & \varphi_2 & \varphi_1^2 \end{pmatrix}$	2
Equicorrelation	$\begin{pmatrix} \varphi_1^2 & & & \\ \rho\varphi_2\varphi_1 & \varphi_2^2 & & \\ \rho\varphi_3\varphi_1 & \rho\varphi_3\varphi_2 & \varphi_3^2 & \\ \rho\varphi_4\varphi_1 & \rho\varphi_4\varphi_2 & \rho\varphi_4\varphi_3 & \varphi_4^2 \end{pmatrix}$	$n + 1$
First-order auto-regressive	$\begin{pmatrix} \varphi^2 & & & \\ \rho\varphi^2 & \varphi^2 & & \\ \rho^2\varphi^2 & \rho\varphi^2 & \varphi^2 & \\ \rho^3\varphi^2 & \rho^2\varphi^2 & \rho\varphi^2 & \varphi^2 \end{pmatrix}$	2
Toeplitz	$\begin{pmatrix} \varphi^2 & & & \\ \rho_1\varphi^2 & \varphi^2 & & \\ \rho_2\varphi^2 & \rho_1\varphi^2 & \varphi^2 & \\ \rho_3\varphi^2 & \rho_2\varphi^2 & \rho_1\varphi^2 & \varphi^2 \end{pmatrix}$	n

Note. All forms are appropriate for balanced designs with equal intervals between measurements. Sphericity and compound symmetry are also valid for unequally spaced data and random measurement designs.

havior and time is accurately represented. When the developmental process is closely tied to chronological time, it is essential to conduct the analysis using elapsed time study rather than the less precise occasion of measurement. Continuing with the above example, an individual might have complete data with four valid scores, but with measurements taken at decimal months 0.25, 6.0, 11.75, and 24.50. Another, also with complete data, may be assessed at months 0.50, 6.33, 12.25, and 22.0. If the independent variable is potentially different for any participant, then even when there are no missing data, the design is unbalanced. In this case the measurements are sampled at random occasions. Of course, missing data and unbalanced designs can occur together as when a participant misses one of his or her own uniquely scheduled appointments.

In some interesting random measurement designs there may be unequal numbers of assessments across participants and yet no missing data. For example, a probation officer may check up on adolescents in the caseload not only at distinct times during the year but also on differing numbers of occasions. Five assessments might be available for one case but seven for another with completely unique contact dates for each. Again, there are no missing data in this example in the sense of planned measurements that were not obtained, as long as contacts do not reflect any systematic tendency to exclude or include particular cases. The salient feature of this design is not that an unequal number of data points are collected, but rather that the behavior is sampled randomly. Random measurement schemes are a natural data collection design in some research domains. It is noteworthy, and of considerable scientific value, that the same facility of mixed-effects models that handles ignorable missing data also applies to more general unbalanced designs and random measurement strategies.

These various data collection plans fit into a more comprehensive framework by modifying model (14) and its components. Let the number of measurements recorded for the ith individual be n_i, and let $\mathbf{x}_i = (x_{i1}, \ldots, x_{n_i})'$ be the elapsed times since the beginning of the study for the jth response of individual i. Both n_i and the set of values in $\mathbf{x}_i$ may vary from person to person. As before, the random coefficients are stochastic and the number of them is always p. The model is

$$\mathbf{y}_i = \mathbf{f}_i(\mathbf{x}_i, \boldsymbol{\beta}_i) + \mathbf{e}_i \tag{15}$$

where the order of $\mathbf{f}_i(\mathbf{x}_i, \boldsymbol{\beta}_i)$ and $\mathbf{e}_i$ is also $(n_i \times 1)$.

Because the number of elements in $\mathbf{e}_i$ may vary across persons, the order of the covariance matrix of residuals varies as well. All of the covariance structures in Table 3 are appropriate for designs that are balanced and have equal intervals between measurements but can be modified to handle ignorable missing data problems (Davidian & Giltinan, 1993). The approach that is used is to define a structure for individuals with complete data, then specialize the structure according to the pattern of missing data. For example, suppose an analysis is based on $n = 4$ measurements and specifies a model using the diagonal structure. Most cases have complete data, but one is missing measurement y_{i3} and a second is missing measurements y_{i1} and y_{i4}. The pattern of the covariance matrix for participants with complete data is exactly as in the second example (the diagonal covariance structure) in Table 3. The patterns for the two special cases are as follows:

$$\underbrace{\begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i4} \end{pmatrix}}_{\mathbf{y}_i} \underbrace{\begin{pmatrix} \varphi_1^2 & & \\ 0 & \varphi_2^2 & \\ 0 & 0 & \varphi_4^2 \end{pmatrix}}_{\boldsymbol{\Lambda}_i} \qquad \underbrace{\begin{pmatrix} y_{i2} \\ y_{i3} \end{pmatrix}}_{\mathbf{y}_i} \underbrace{\begin{pmatrix} \varphi_2^2 & \\ 0 & \varphi_3^2 \end{pmatrix}}_{\boldsymbol{\Lambda}_i}$$

The notation used in customizing the covariance structure to fit the available data is $\boldsymbol{\Lambda}_i = \mathrm{cov}(\mathbf{e}_i)$. The subscript on $\boldsymbol{\Lambda}$ indicates that the order of the matrix and the elements it contains depend on the pattern of available data and so varies between persons.

In summary, the general form of the multiphase model we use is given by (15) with the response for an individual similar to (2). Within- and between-subjects covariance structures for the residuals and random coefficients, respectively, are written as

$$\mathbf{e}_i \sim N(\mathbf{0}, \boldsymbol{\Lambda}_i) \qquad \text{Level 1 variability} \tag{16a}$$

$$\boldsymbol{\beta}_i \sim N(\boldsymbol{\beta}, \boldsymbol{\Phi}) \qquad \text{Level 2 variability} \tag{16b}$$

Estimating the Model

The multiphase model is a special case of the more general nonlinear mixed-effects model. Because of the flexibility of the nonlinear mixed-effects framework, the estimation demands for the multiphase model can be computationally intensive. This is because the model combines features of nonlinear regression, which is a difficult estimation problem, with mixed-effects models, which is also a difficult estimation problem. The actual amount of calculation depends on the particular function under investigation. For example, if a model specifies random change points and is estimated by maximum likelihood, then the computational overhead is substantial (Cudeck & du Toit, in press; Davidian & Giltinan, 1995, chapters 4–6; Pinheiro & Bates, 1995; Roe, 1997). On the other hand, if the change points are fixed parameters without random effects, then the model fits into the rubric of partially nonlinear mixed-effects regression, and estimation by maximum likelihood is rather straightforward (Blozis & Cudeck, 1999; du Toit & Cudeck, 2001; Vonesh & Chinchilli, 1997). If the change points are assumed to be known values that are specified by the investigator, then the problem is simply a linear mixed-effects model, and estimates can be obtained without great difficulty. Several other approaches that do not rely on maximum likelihood have been taken to the estimation problem. These work well in many circumstances (e.g., Beal &

Sheiner, 1982; Davidian & Giltinan, 1995, chapters 5–8; Lindstrom & Bates, 1990; Vonesh & Chinchilli, 1997, chapter 7). Currently at least four popular programs—LISREL (Jöreskog, Sörbom, du Toit, & du Toit, 1999), SAS NLMIXED (Wolfinger, 1999), S-PLUS (Pinheiro & Bates, 2000), and Mx (Neale, 1999)—can produce maximum-likelihood estimates for versions of the model that specify unknown change points. If the model has fixed change points, these same programs, plus several other packages that are available for estimating the linear mixed-effects model (Kreft, de Leeuw, & van der Leeden, 1994), can be used. Many special cases of the general multiphase model are readily estimated using them. Appendix B shows details for one version of the second example that was estimated using SAS NLMIXED. All the other examples reviewed below can be handled similarly.

Analysis of Nonverbal Performance Data

We illustrate the model with data from Bradway and McArdle's investigation of intellectual performance over the life span. It is apparent in Figure 1 that the dramatic early growth in nonverbal skill in childhood and adolescence ceases almost entirely in adulthood, at which point the process changes to a pattern that is essentially linear in form. One interesting problem is to estimate the age when the rapid early phase changes over to the more stable pattern of the second period. Another problem is to determine the typical rate of decline in late old age.

An effective representation of these data is possible with the quadratic–linear model of (11) and (13), where again x_{ij} is age at the jth occasion of measurement for the ith individual. In addition to the unknown change point, τ, the slope of the linear phase, γ_1, is a valuable feature of the model. It can be used to test whether there is overall improvement or decline across the late adult years. With first- and second-order continuity, there are a total of four coefficients in this function.

Proceeding with the version in (13) and adding random effects to each term allows for individual differences on both the change point and the slope of the second phase as well as the other parameters

$$f_{ij}(\mathbf{x}_i, \boldsymbol{\beta}_i) = \gamma_{0i} + \gamma_{1i}x_{ij} + \alpha_{2i}(\tau_i - x_{ij})^2_+$$

with the set of random coefficients defined as $\boldsymbol{\beta}_i = (\gamma_{0i}, \gamma_{1i}, \alpha_{2i}, \tau_i)'$. The covariance matrix of the random effects in (16a) is of order four. The Level 1 covariance matrix in (16b) is assumed to be homogeneous, $\boldsymbol{\Lambda}_i = \sigma_e^2\mathbf{I}_{n_i}$, with residual variance σ_e^2.

Maximum-likelihood estimates of the parameters are as follows:

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\gamma}_0 \\ \hat{\gamma}_1 \\ \hat{\alpha}_2 \\ \hat{\tau} \end{pmatrix} = \begin{pmatrix} 80.4 \\ -.129 \\ -.592 \\ 18.5 \end{pmatrix}$$

$$\hat{\boldsymbol{\Phi}} = \begin{pmatrix} 49.1 & & & \\ .473 & .018 & & \\ -.205 & .012 & .031 & \\ -2.44 & .137 & .404 & 9.25 \end{pmatrix} \quad \hat{\sigma}_e^2 = 82.5$$

The most interesting feature of the regression estimates are the last two terms. The estimated age at which the transition from the first phase to the second occurs is $\hat{\beta}_4 = \hat{\tau} = 18.5$ years, with $SE(\hat{\tau}) = 0.59$. There is appreciable variability in this change point, however: $[\hat{\boldsymbol{\Phi}}]_{44} = 9.25$. This indicates that some participants switch to the linear, adult pattern earlier or later than the average. The slope of the second phase is negative, $\hat{\beta}_2 = \hat{\gamma}_1 = -0.129$, $SE(\hat{\gamma}_1) = 0.04$, and a small, but not zero, estimated variance: $[\hat{\boldsymbol{\Phi}}]_{33} = .018$. The mean function that corresponds to these estimates is shown in Figure 7, where the change point is marked as the dashed vertical line.

Although the trend in the second phase is decreasing overall, a few individuals show little decline in cognitive functioning in old age and others actually

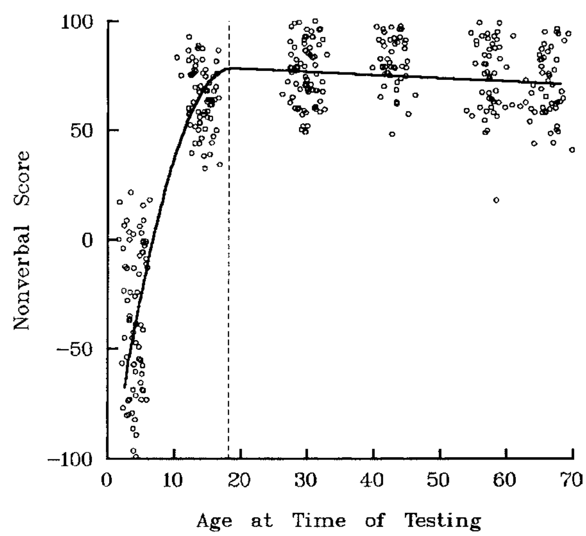

Figure 7. Data and fitted mean function for the quadratic–linear model. The change point is indicated by the vertical dashed line.

continue to improve. Plots in Figure 8 display fitted functions for a few selected cases. The 2 individuals in Figure 8A had large differences in Phase 2 intercept (70.8 vs. 91.9). Those in Figure 8B had the largest difference in Phase 2 slope (−0.32 vs. 0.04). Those in Figure 8C had the largest differences in transition age (14.1 vs. 23.6).

Altogether, this analysis seems satisfactory for the data and gives results that are reasonable. This does not mean the model is the last word for this interesting problem, but rather that this description is plausible. For the sake of comparison, a few other models were also examined to obtain more information about the performance of alternative structures. Table 4 summarizes these models with fit statistics for each. The number of parameters in the kth model is q_k. In this context, the Akaike information criterion has the form (cf. Vonesh & Chinchilli, 1997, chapter 6.4) $AIC_k = -2LnL_k + 2q_k$, where LnL_k is the value of the log-likelihood function at the solution. For AIC_k, lower values indicate better model performance.

Model 2 is a two-phase, linear–linear model. Segments are constrained to be continuous at the change point as in (6). This model has the same number of parameters as the quadratic–linear model but fits slightly worse. Although the model captures the phases that are evident in the data, it does not seem as appropriate scientifically as the quadratic–linear version. The specification in Model 2 of a constant rate throughout childhood with no deceleration at all until the change point may not be plausible. Model 2 also implies that individuals suddenly switch to an adult pattern with a different rate. In spite of these reservations, the adult linear pattern in the second phase is similar to the second phase under Model 1.

Model 3 is the nonlinear exponential function that is often used for data that show increase to an asymptote

$$f_{ij} = \theta_F + (\theta_0 - \theta_F)\exp(-\alpha x_{ij})$$

where θ_0 is initial performance, θ_F is performance at an advanced stage of learning, and α is rate of improvement. This function is nonlinear and increasing throughout adolescence. However, the specification that the function approaches an upper bound of θ_F but does not decline is an undesirable aspect. Rate during the older years is a feature to estimate. Consequently, it would be expected to fit these data less well than Model 1 or 2 because many cases show a decrease in knowledge in later adulthood. In fact, the measures of fit for Model 3 are poorer than for the other two.

Models 4 and 5 are the quadratic and cubic structures:

$$f_{i,j} = \alpha_0 + \alpha_1 x_j + \alpha_2 x_{ij}^2 \qquad f_{ij} = \alpha_0 + \alpha_1 x_{ij} + \alpha_2 x_{ij}^2 + \alpha_3 x_{ij}^3$$

These are perhaps the two most commonly applied nonlinear functions and might be thought to be appropriate here as a first approximation. The long, virtually linear response in adulthood does not match these models well on a priori grounds, however. It was found in each case that random effects are needed for only the first two coefficients: α_2 in Model 4 and α_2 and α_3 in Model 5 are fixed coefficients. These models fit the data much more poorly than the others of this group.

A Model for the Glucose Challenge Data

The linear–linear model was not effective in the previous example. In other settings, it may perform quite well. A case in point is the glucose challenge data from obese and control participants introduced in Figure 2. The purpose of the investigation was to assess the pattern of phosphate elimination after introduction. Phosphate level is expected to decrease immediately after administration but to recover subsequently. It can be seen in Figure 2 that the timing of this reaction is variable and appears to be delayed as much as an hour for patients in the group with obesity.

Table 4
Measures of Fit of Alternative Models

Model	Description	LnL_k	qk	AIC_k
1	Quadratic–linear	−1538.2	15	3106.4
2	Linear–linear	−1542.5	15	3115.0
3	Exponential	−1551.2	10	3122.4
4	Quadratic with two random effects	−1752.4	7	3518.8
5	Cubic with two random effects	−1671.9	8	3359.8

Note. LnL_k is the value of the log-likelihood function at the solution; q_k is the number of parameters in the kth model; AIC_k is the Akaike information criterion.

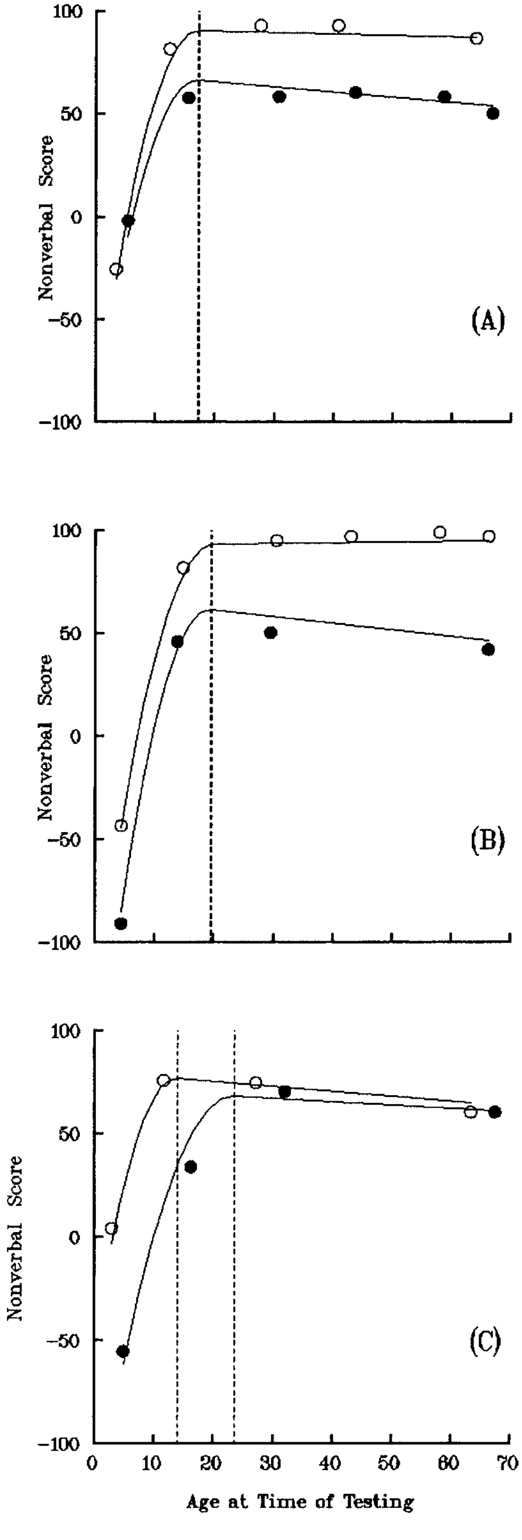

Figure 8. Individual data and individual functions illustrating (A) differences in intercept in Phase 2, (B) differences in Phase 2 slope, and (C) differences in change point. The change point is indicated by the vertical dashed line.

Estimating the rebound is an important and interesting part of the problem.

It is assumed that the phases are continuous at the change point, so that after simplifying (12) because $\delta_0 = 0$, the basic model is

$$f_{ij}(\mathbf{x}_i,\boldsymbol{\beta}_i) = \alpha_{0i} + \alpha_{1i}x_{ij} + \delta_{1i}(x_{ij} - \tau_i)_+$$

where the random coefficients are $\boldsymbol{\beta}_i = (\alpha_{0i}, \alpha_{1i}, \delta_{1i}, \tau_i)'$. As was reviewed earlier, α_{0i} and α_{1i} are the intercept and slope of the first segment, and $\delta_{1i} = \gamma_{1i} - \alpha_{1i}$ is the difference in slopes between the two segments. To distinguish between populations, we represent the fixed parameters for those participants who were in the control group with subscript C, and those for patients with obesity with subscript E (participants who were not obese were controls and participants with obesity were the experimental group in this study). The eight parameters, four for each population, are

$$\boldsymbol{\beta}_{\mathrm{C}} = (\alpha_{\mathrm{C}0}, \alpha_{\mathrm{C}1}, \delta_{\mathrm{C}1}, \tau_{\mathrm{C}})' \qquad \boldsymbol{\beta}_{\mathrm{E}} = (\alpha_{\mathrm{E}0}, \alpha_{\mathrm{E}1}, \delta_{\mathrm{E}1}, \tau_{\mathrm{E}})'$$

We specify that the participants in the control population have random coefficients with mean vector $\boldsymbol{\beta}_{\mathrm{C}}$ and covariance matrix $\boldsymbol{\Phi}$: $\boldsymbol{\beta}_i \sim N(\boldsymbol{\beta}_{\mathrm{C}}, \boldsymbol{\Phi})$. Those in the experimental population have mean vector $\boldsymbol{\beta}_{\mathrm{E}}$ with covariance matrix $\boldsymbol{\Phi}$: $\boldsymbol{\beta}_i \sim N(\boldsymbol{\beta}_{\mathrm{E}}, \boldsymbol{\Phi})$. Thus, the fixed effects differ in the two populations, but the covariance matrix of the random coefficients is the same.

In this example, estimates were obtained from SAS NLMIXED using the first-order linear approximation method of Beal and Sheiner (1982). The SAS code for the final version of this analysis is included in Appendix B. In the first use of the model, the estimated covariances between τ_i and the other random coefficients were essentially zero. In the interest of parsimony we fixed these three covariances to zero. There was no appreciable loss of fit with this simplification. Estimates of the modified model are as follows, with standard errors in parentheses:

$$\hat{\boldsymbol{\beta}}_{\mathrm{C}} = [3.98(.20), -1.30(.14), 1.63(.15), 1.06(.11)]$$

$$\hat{\boldsymbol{\beta}}_{\mathrm{E}} = [4.52(.15), -70(.07), 1.02(.09), 2.08(.13)]$$

$$\hat{\boldsymbol{\Phi}} = \begin{pmatrix} .41 & & & \\ -.09 & .06 & & \\ .07 & -.07 & .08 & \\ 0^* & 0^* & 0^* & .09 \end{pmatrix} \qquad \sigma_e^2 = .12$$

The values in the last row of $\hat{\boldsymbol{\Phi}}$ marked by an asterisk are fixed zeros. Estimated standard errors of the diagonal elements of $\hat{\boldsymbol{\Phi}}$ are $\mathrm{SE}[\hat{\boldsymbol{\Phi}}]_{11} = .11$, $\mathrm{SE}[\hat{\boldsymbol{\Phi}}]_{22} =$

.01, SE$[\hat{\boldsymbol{\Phi}}]_{33}$ = .01, SE$[\hat{\boldsymbol{\Phi}}]_{44}$ = .04. Of course, all the standard errors are approximate because of the small sample. The estimates of initial status differ by more than half a unit in the two populations, $\hat{\alpha}_{C0}$ = 3.98 (mg/dl) versus $\hat{\alpha}_{E0}$ = 4.52 (mg/dl), and the slope in the control population is steeper than for patients with obesity, $\hat{\alpha}_{C1} = -1.30$ versus $\hat{\alpha}_{E1} = -0.70$. The change point marks when the decline in phosphate levels gives way to recovery in the second phase. The time at which rebound occurs is a full hour earlier for control participants than for the group with obesity, $\hat{\tau}_C$ = 1.06 versus $\hat{\tau}_E$ = 2.08. The variability in the initial status coefficients, α_{0i}, is especially large, $[\hat{\boldsymbol{\Phi}}]_{11}$ = .41, whereas the variability in the other random effects is smaller but evidently reliably different from zero.

Figure 9 shows the sample means and fitted functions for the two populations. Figure 10 contains data and individual functions for some of the participants with control participants in the top two rows and participants with obesity in the last two rows. As always, individual differences in all components of the function are substantial. Not surprisingly, data from some individuals, for example, Participant 6 in the control group, are not well described by the model. For many others, the fit appears to be excellent. As illustrated in the four plots in the bottom row of Figure 10, the differences in individual change points is occasionally striking.

One other comparison is informative. An obvious model for data such as these is the simple quadratic model. The version used in this analysis specified three parameters for each population. The distribution of the random coefficients was common in both populations. This leads to a setup with six coefficients:

$$
\begin{aligned}
f_{ij} &= \alpha_{0i} + \alpha_{1i}x_j + \alpha_{2i}x_{ij}^2 \qquad \text{(control)} \\
&= \gamma_{0i} + \gamma_{1i}x_j + \gamma_{2i}x_{ij}^2 \qquad \text{(experimental)}
\end{aligned}
$$

An additional seven variances and covariances are specified for the matrices of (16). On the basis of parameter estimates and graphical analysis, the model seems plausible. It has a fit measure of AIC = 422. On the other hand, the two-phase linear model requires 16 parameters and produces a value of AIC = 347. Thus, the two-phase model uses more parameters but fits much better than the quadratic model. The problem with the quadratic model in this problem is that the response before and after the change point is not symmetric. The two-phase model accommodates the difference in patterns easily.

Summary

A multiphase model may be just the right structure for behavior that exhibits distinct regimes. The versions examined here were made up of two simple pieces, either linear or quadratic. Combined into one model, they produce a function that is flexible and effective. As was illustrated in the examples, a two-phase model may perform better than other simple candidates such as a low-order polynomial or one of the common nonlinear functions. Because each piece is itself interpretable, the overall model is easy to understand and can provide a very valuable summary.

Multiphase models have been applied extensively in nonlinear regression (e.g., Seber & Wild, 1989,

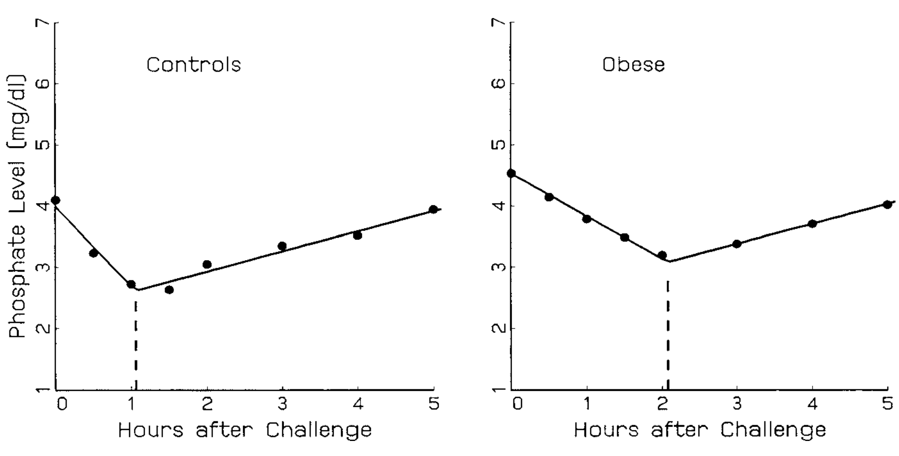

Figure 9. Sample means and fitted mean functions of linear–linear model for phosphate levels, with estimated change points indicated by the dashed vertical lines.

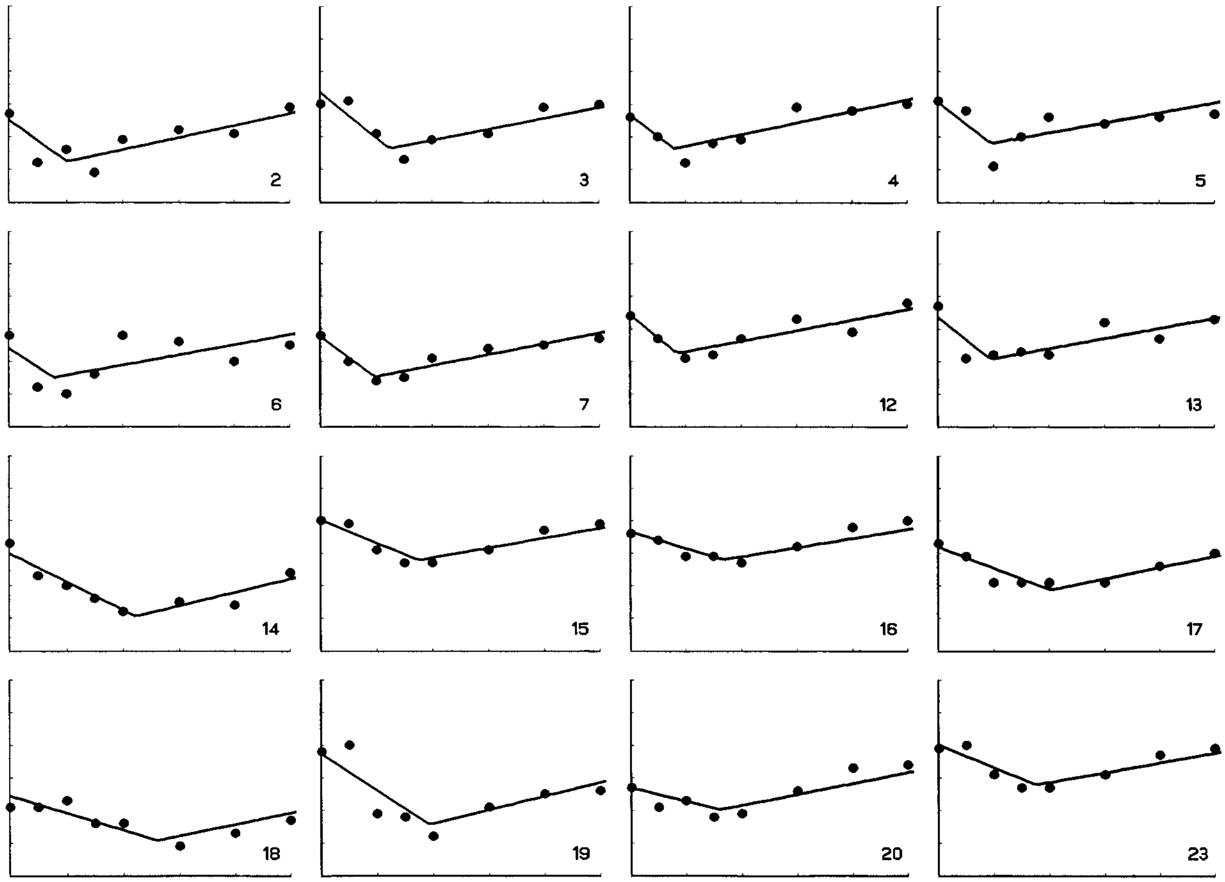

Figure 10. Data and individual functions for selected individuals. Plots in the upper two rows are those of control participants, and plots in the lower two rows are those of patients with obesity.

chapter 9) as a dependable way to describe a process that is possibly quite complex. They are even more appealing in the mixed-model context because the random effects on various combinations of coefficients are especially satisfactory. The examples reviewed here illustrate these possibilities. Related work on self-modeling nonlinear regression (Lawton, Sylvestre, & Maggio, 1972; Lindstrom, 1995), an interesting semiparametric approach to repeated measures data, shares many of these same attributes.

One of the most attractive elements of a multiphase model is the change point, the age or time when behavior switches from one process to another. For many problems this parameter is important in its own right as a marker of qualitative change in behavior. The change point can be set at a known value, can be specified as a fixed parameter that characterizes the entire population but without random effects, or can be a random coefficient on which individual differences are expressed. This latter possibility gives a multiphase model considerable generality. Different persons may transition between phases at different times, which of course implies that the amount of time one spends in any regime is variable as well.

In addition to their use as a descriptive tool for a complex pattern of change, several practical applications have been suggested for this model. Pauler and Laird (2000), for example, took the point of transition between segments as an indication of noncompliance in a clinical trial, where noncompliance is operationalized as an appreciable change in rate compared with a baseline period. Slate and Turnbull (2000) showed how each new measurement obtained for an individual during an ongoing study may be used to assess the probability that the change point has occurred. When this probability is high, then the evidence indicates that the participant has moved from the first phase to the second. If the first phrase represents standard or normal progress and the second represents an abnormal pattern, then knowledge that the change point has occurred could be used for early detection or remediation.

Another practical feature of the general model is the facility to specify varying degrees of smoothness between phases at the transition. In some cases, a switch in phase produces not only a change in form but also a discontinuity or jump in the response at the change point. The degree of discontinuity may be informative substantively as is illustrated in Figure 3 with the reminiscence effect of the rotary pursuit data of Fox et al. (1996). If either or both phases are nonlinear, then smoothness can be specified to include derivatives as well as the functions themselves. P. L. Smith (1979) suggested several formal tests of smoothness to investigate this matter directly.

Examples reviewed here are based on two pieces that are either linear or quadratic. If the change point has no random effects, the model fits into the linear mixed-effects family and is easy to estimate. If instead the change point has random effects, then the structure is a nonlinear mixed-effects model and is more difficult to estimate. Many forms of the general model can be fit using existing software.

References

Bates, D. M., & Watts, D. G. (1988). *Applied nonlinear regression.* New York: Wiley.

Beal, S. L., & Sheiner, L. B. (1982). Estimating population kinetics. *CRC Critical Reviews in Biomedical Engineering, 8,* 195–222.

Blozis, S. A., & Cudeck, R. (1999). Conditionally linear mixed-effects models with latent variable covariates. *Journal of Educational and Behavioral Statistics, 24,* 245–270.

Bock, R. D. (1991). Prediction of growth. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 126–136). Washington, DC: American Psychological Association.

Bradway, K., & Thompson, C. (1962). Intelligence at adulthood: A twenty-five year followup. *Journal of Educational Psychology, 53,* 1–14.

Brown, H., & Prescott, R. (1999). *Applied mixed models in medicine.* New York: Wiley.

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

Burchinal, M., & Appelbaum, M. I. (1991). Estimating individual developmental functions: Methods and their assumptions. *Child Development, 62,* 23–43.

Chaiken, S. R. (1994). The inspection time not studied: Processing speed ability unrelated to psychometric intelligence. *Intelligence, 19,* 295–316.

Crowder, M. J., & Hand, D. J. (1990). *Analysis of repeated measures.* London: Chapman & Hall.

Cudeck, R. (1996). Mixed-effects models in the study of individual differences with repeated measures data. *Multivariate Behavioral Research, 31,* 371–403.

Cudeck, R., & du Toit, S. H. C. (in press). Nonlinear multilevel models for repeated measures data. In N. Duan & S. P. Reise (Eds.), *Multilevel modeling: Methodological advances, issues and applications.* Mahwah, NJ: Erlbaum.

Davidian, M., & Giltinan, D. M. (1993). Some simple methods for estimating intraindividual variability in nonlinear mixed effects models. *Biometrics, 49,* 59–73.

Davidian, M., & Giltinan, D. M. (1995). *Nonlinear models for repeated measurement data.* London: Chapman & Hall.

Diggle, P. J., Liang, K.-Y., & Zeger, S. L. (1994). *Analysis of longitudinal data.* Oxford, England: Oxford University Press.

Durrleman, S., & Simon, R. (1989). Flexible regression models with cubic splines. *Statistics in Medicine, 8,* 551–561.

du Toit, S. H. C. (1993). *Analysis of multilevel models.* Pretoria, South Africa: Human Sciences Research Council.

du Toit, S. H. C., & Cudeck, R. (2001). The analysis of nonlinear random coefficient regression models with LISREL using constraints. In R. Cudeck, S. H. C. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Current perspectives* (pp. 259–278). Chicago: Scientific Software International.

Fox, P. W., Hershberger, S. L., & Bouchard, T. J. (1996). Genetic and environmental contributions to the acquisition of a motor skill. *Nature, 384,* 356–359.

Gallant, A. R. (1987). *Nonlinear statistical models.* New York: Wiley.

Gallant, A. R., & Fuller, W. A. (1973). Fitting segmented polynomial regression models whose join points have to be estimated. *Journal of the American Statistical Association, 68,* 144–147.

Goldstein, H. I. (1995). *Multilevel statistical models.* London: Arnold.

Hedeker, D., & Gibbons, R. D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods, 2,* 64–78.

Jones, R. H. (1993). *Longitudinal data with serial correlation: A state–space approach.* London: Chapman & Hall.

Jöreskog, K. G., Sörbom, D., du Toit, S. H. C., & du Toit, M. (1999). *LISREL 8: New statistical features.* Chicago: Scientific Software International.

Kreft, I., & de Leeuw, J. (1998). *Introducing multilevel modeling.* Thousand Oaks, CA: Sage.

Kreft, I. G. G., de Leeuw, J., & van der Leeden, R. (1994). Review of five multilevel analysis programs: BMDP-5V, GENMOD, HLM, ML3, VARCL. *American Statistician, 48,* 324–335.

Lawton, W. H., Sylvestre, E. A., & Maggio, M. S. (1972). Self-modeling nonlinear regression. *Technometrics, 14,* 381–392.

Lewis, D. (1966). *Quantitative methods in psychology.* Iowa City: University of Iowa Press.

Leyland, A. H., & Goldstein, H. (2001). *Multilevel modelling of health statistics.* New York: Wiley.

Lindsey, J. K. (1999). *Models for reported measurements* (2nd ed.). Oxford, England: Oxford University Press.

Lindstrom, M. J. (1995). Self-modeling with random shift and scale parameters and a free knot spline shape function. *Statistics in Medicine, 14,* 2009–2021.

Lindstrom, M. J., & Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures. *Biometrics, 46,* 673–687.

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

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

McArdle, J. J., & Hamagami, F. (1996). Multilevel models from a multiple group structural equation perspective. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling* (pp. 89–124). Mahwah, NJ: Erlbaum.

McCulloch, C. E., & Searle, S. R. (2001). *Generalized, linear, and mixed models.* New York: Wiley.

Mirowsky, J., & Reynolds, J. R. (2000). Age, depression, and attrition in the National Survey of Families and Households. *Sociological Methods and Research, 28,* 476–504.

Morrell, C. H., Pearson, J. D., Carter, H. B., & Brant, L. J. (1995). Estimating unknown transition times using a piecewise nonlinear mixed-effects model in men with prostate cancer. *Journal of the American Statistical Association, 90,* 45–53.

Neale, M. C. (1999). *Mx: Statistical modeling* (5th ed.). Richmond, VA: Virginia Commonwealth University, Medical College of Virginia, Department of Psychiatry.

Pauler, D. K., & Laird, N. M. (2000). A mixture model for longitudinal data with application to assessment of noncompliance. *Biometrics, 56,* 464–472.

Payne, D. O. (1987). Hypermnesia and reminiscence in recall: A historical and empirical review. *Psychological Bulletin, 101,* 5–27.

Pinheiro, J. C., & Bates, D. M. (1995). Approximations to the loglikelihood function in the nonlinear mixed effects model. *Journal of Computational and Graphical Statistics, 4,* 12–35.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS.* New York: Springer-Verlag.

Roe, D. J. (1997). Comparison of population pharmacokinetics modeling methods using simulated data: Results from the population modeling workgroup (with discussion). *Statistics in Medicine, 16,* 1241–1262.

Seber, G. A. F., & Wild, C. J. (1989). *Nonlinear regression.* New York: Wiley.

Skates, S. J., Pauler, D. K., & Jacobs, I. J. (2001). Screening based on the risk of cancer calculation from Bayesian hierarchical changepoint and mixture models of longitudinal markers. *Journal of the American Statistical Association, 96,* 429–439.

Slate, E. H., & Turnbull, B. W. (2000). Statistical models for longitudinal biomarkers of disease onset. *Statistics in Medicine, 19,* 617–637.

Smith, A. F. M., & Cook, D. G. (1980). Straight lines with a change-point: A Bayesian analysis of some renal transplant data. *Applied Statistics, 29,* 180–189.

Smith, P. L. (1979). Splines as a useful and convenient statistical tool. *American Statistician, 33,* 57–62.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* Thousand Oaks, CA: Sage.

Terman, L. M., & Merrill, M. A. (1937). *Measuring intelligence.* Boston: Houghton Mifflin.

Verbeke, G., & Molenberghs, G. (1997). *Linear mixed models in practice.* New York: Springer.

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data.* New York: Springer.

Vonesh, E. F., & Chinchilli, V. M. (1997). *Linear and nonlinear models for the analysis of repeated measurements.* New York: Marcel Dekker.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4th ed.). Baltimore, MD: Williams & Wilkins.

Wolfinger, R. D. (1999). *Fitting nonlinear mixed models with the new NLMIXED procedure* (Tech. Paper No. 287). Cary, NC: SAS Institute.

Zeger, S. L., Liang, K. Y., & Albert, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics, 44,* 1049–1060.

Zerbe, G. O. (1979). Randomization analysis of the completely randomized design extended to growth and response curves. *Journal of the American Statistical Association, 74,* 215–221.

Appendix A

Fitting a Two-Phase Model to Data From 1 Individual With SAS PROC NLIN

```
/*
**   One individual's data from Scott Chaiken's (1994) visual search experiment
**   x: Trials of the experiment (coded 0 to 23)
**   y: Reaction time (recorded in milliseconds, transformed to seconds)
**
**   Two phase model, quadratic-linear, with unknown change point, plus
**   first- and second-order continuity at change point
**
**                       y = f(x, beta) + e
**
**   where
**
**                      | a0 + a1 * x + a2 * x^2        x <= tau
**   f(x, beta) = |
**                      | g0 + g1 * x                   x > tau
**
**   There are 6 parameters in the above model (a0, a1, a2, g0, g1, tau).
**   First- and second-order continuity at tau implies that two parameters
**   are dependent on the others. This example uses the restrictions
**
**            a1 = g1 − 2 * a2 * tau
**
**            a0 = g0 + (g1 − a1) * tau − a2 * tau^2
**
**   Consequently, the actual number of parameters to estimate are the four
**   coefficients beta = (a2, g0, g1, tau)
*/

title1 'Two phase model with estimated change point for a single individual';

data RT;
   input x y @@;
      y = y / 1000;
         datalines;
            0    3566     1    2712     2    1872     3    2111     4    1697
            5    1544     6    1082     7    1322     8    1198     9    1130
           10    1194    11    1126    12     999    13    1128    14    1222
           15    1074    16    1020    17    1014    18    1037    19     885
           20     978    21    1017    22     883    23    1080
                   ;

/*   Nonlinear regression for least squares estimation of the model    */
proc nlin;
   parms a2 = .1 g0 = 1 g1 = −.1 tau = 6;
   a1 = g1 − 2 * a2 * tau;
   a0 = g0 + (g1 − a1) * tau − a2 * tau** 2;
   if x <= tau then
      model y = a0 + a1 * x + a2 * x**2;
   else
      model y = g0 + g1 * x;
   output out=res predicted=yhat;
run;

   /*   Setup and plot fitted values and response    */
legend1 frame cfram=ligr label=none cborder=black
            position=center value=(justify=center);
axis1 label=(angle=90 rotate=0) minor=none;
```

(*Appendixes continue*)

```
axis2 minor=none;
proc gplot;
   plot y*x yhat*x/frame cframe=ligr legend=legend1
   vaxis=axis1 haxis=axis2 overlay;
run;
```

Appendix B

Fitting the Two-Phase, Linear–Linear Mixed-Effects Model

```
/*
**   Use SAS PROC NLMIXED to fit a two-phase, linear-linear model with random
**   effects. Data from Zerbe (1979) J. Amer. Stat. Assn., 74, 215–221.
**
**   Two samples: N1 = 13 normal (grp = 1), N2 = 20 obese (grp = 2)
**
**   Four parameters for each population
**
**            betc = (ac0, ac1, dc1, tc)       bete = (ae0, ae1, de1, te)
**
**   Random coefficients for individuals in the two populations have distribution
**
**            bet_i ~ N(betc, phi) (normals)       bet_i ~ N(bete, phi) (obese)
**
**   For phosphate level (y_ij) of individual i at the jth measurement, with
**   xi = (x_1, . . ., x_8) equal hours from start of study), the model is
**
**                              y_ij = f_ij(xi,bet_i) + e_ij
**   where
**
**         f_ij = bet_i1 + bet_i2 * x_j + bet_i3 * max (0, x_j-bet_14)
**
**   Level 1 variability:
**         Residuals are normal, with mean vector zero, variances of sig^2
**
**   Level 2 variability:
**         Random effects are normal, with mean vector zero, covariance matrix phi.
**
**   Estimation based on first-order Taylor expansion of Beal & Shiner (1982).
*/

title1 'Linear-linear mixed effects model, estimated change points for each sample';

data gfile;
   filename d_source 'c:/zerbe.dat';
   infile d_source;
      input indiv grp hour phosphat;

/* --------- Setup for linear-linear model ---------------------------- */

proc nlmixed method=firo;

   /*   3 covariances in 4th row of random effects matrix are fixed at zero   */
     parms  ac0 = 4.0          ac1 = −1.3          dc1 = 1.6          tauc = 1.1
            ae0 = 4.5          ae1 =  −.70         de1 = 1.0          taue = 2.1

            v11 =   .41
            c21 = −.08         v22 =   .08
            c31 =   .07        c32 = −.07          v33 = .06
                                                                      v44 = .15

          var_e = .10;
```

```
/*   Define individual coefficients, with parameters that differ by group   */
     if (grp = 1) then bet_i1 = ac0 + u1;
     if (grp = 1) then bet_i2 = ac1 + u2;
     if (grp = 1) then bet_i3 = dc1 + u3;
     if (grp = 1) then bet_i4 = tauc + u4;

     if (grp = 2) then bet_i1 = ae0 + u1;
     if (grp = 2) then bet_i2 = ae1 + u2;
     if (grp = 2) then bet_i3 = de1 + u3;
     if (grp = 2) then bet_i4 = taue + u4;

/*
**   Specify the linear-linear model with random change point.
**   Function "max()" works like the "+" operator.
*/

   fn = bet_i1 + bet_i2 * hour + bet_i3 * max (0, hour-bet_i4);

model
   phospat ~normal (fn, var_e);

random
   u1 u2 u3 u4 ~ normal ([0,0,0,0], [11,c21,v22,c31,c32,v33,0,0,0,v44])
   subject = indiv
   out = ran_effect;
run;

/*   The dataset "ran_effect" has predicted random effects for each subject   */

proc print data = ran_effect;
run;
```

The groups are coded as "grp = 1" and "grp = 2", for normal and obese, respectively. Phosphate measurements were taken at hours 0, 0.5, 1, 1.5, 2, 3, 4, and 5. The eight recordings for the first two cases in the normal group as they appear in the input file "zerbe.dat" are listed below.

1	1	0.0	4.3
1	1	0.5	3.3
1	1	1.0	3.0
1	1	1.5	2.6
1	1	2.0	2.2
1	1	3.0	2.5
1	1	4.0	3.4
1	1	5.0	4.4
2	1	0.0	3.7
2	1	0.5	2.2
2	1	1.0	2.6
2	1	1.5	1.9
2	1	2.0	2.9
2	1	3.0	3.2
2	1	4.0	3.1
2	1	5.0	3.9

Received July 11, 2000
Revision received October 25, 2001
Accepted October 25, 2001 ■