CHAPTER 2

ON INTERPRETABLE REPARAMETERIZATIONS OF LINEAR AND NONLINEAR LATENT GROWTH CURVE MODELS

Kristopher J. Preacher
Vanderbilt University

Gregory R. Hancock
University of Maryland

One of the primary goals of longitudinal modeling is to estimate and interpret free model parameters that reflect meaningful aspects of change over time in a parsimonious manner. Perhaps the simplest example is to estimate a linear slope in regression analysis when the predictor is time (t) and the criterion (y) is measured repeatedly at several occasions; this slope may be interpreted straightforwardly as the expected change in y given a unit change in t in the population, holding constant all other predictors. Of course, the nature of longitudinal change may be far from this simplest

Advances in Longitudinal Methods in the Social and Behavioral Sciences, pages 25–58
Copyright © 2012 by Information Age Publishing
All rights of reproduction in any form reserved.

scenario, and as such researchers are increasingly seeking theoretically appropriate ways to model more complex nonlinear systems as well (see, e.g., Grimm & Ram, 2009; Ram & Grimm, 2007).

In many circumstances, a given linear or nonlinear model may be sensible from a theoretical perspective and may fit the data well, yet may have parameters that are difficult to interpret in a meaningful way. Such models often may be reparameterized. *Reparameterization* is the reexpression of a target model so that parameters of the reexpressed model more closely align with questions of substantive interest to the researcher. In this chapter we discuss various ways in which reparameterization may be used in the context of latent growth curve modeling (LGM), a powerful and flexible application of structural equation modeling (SEM) often used to model trends in longitudinal data and individual differences in those trends.

As is seen in this chapter, reparameterization can be extremely useful to researchers for several reasons. First and foremost, it is often more convenient to directly estimate some quantity of interest as a model parameter than it is to compute it post hoc as a function of other estimated parameters. Direct estimation of these parameters permits researchers to obtain both point and interval estimates of the quantity of interest, allowing the researcher to test hypotheses and determine precision of estimation. Reparameterization also allows researchers to determine whether an aspect of change or some other important quantity is predicted or moderated by other variables, and generally allows the flexibility of treating a parameter as a fixed known value, as an unknown value to be estimated, or even (in some cases) as a random coefficient reflecting individual differences in some aspect of the function.

To provide a framework for our discussion, we describe three illustrative (but by no means exhaustive) classes of reparameterization, and present exemplars from each class: (1) quantifying homogeneity or heterogeneity of individuals, (2) estimating and predicting aspects of change, and (3) estimating and predicting time-specific individual differences. To illustrate the quantification of homogeneity or heterogeneity, we use the context of adolescents' delinquent peer associations to discuss how a linear growth curve model may be reparameterized so that the aperture (the point in time associated with the least variable individual differences) is directly estimated as a model parameter. As an example of estimation and prediction of aspects of change, we demonstrate the estimation of individually varying surge points and surge slopes (Choi, Harring, & Hancock, 2009) in elementary school children's mathematics scores, and show how gender can be used as a person-level predictor of these random effects. To illustrate the estimation and prediction of time-specific individual differences, we show how to reparameterize a nonlinear function of infant growth so that individual differences can be operationalized as a random effect and predicted by maternal breast-

feeding behavior. Clearly, individual differences are central to each of these exemplars, as they often are in models of change over time. In each case, the target of reparameterization will become an estimated parameter or a random effect reflecting individual differences in the target aspect.

This chapter thus serves as an introduction to the general concept of reparameterization. It is our intention and hope that readers will gain an understanding of reparameterization that will allow them to generalize the procedure to new contexts. In the next section, we discuss the concept of reparameterization as a general tool that can be used in the context of modeling longitudinal data. Then we elaborate on our empirical examples as exemplars reflecting the three broad classes of reparameterization. At the end, we suggest other potential reparameterization classes as well.

REPARAMETERIZATION

As indicated earlier, reparameterization is the reexpression of a target model so that parameters of the reexpressed model more closely align with questions of substantive interest to the researcher. The use of reparameterization assumes that the researcher has first identified an appropriate target function to represent growth or change over time, and that the researcher wishes to quantify some aspect of the function (e.g., a point in time, an aspect of change, or a prediction coefficient) not already represented in the standard parameterization of the function. A reparameterized model should have the same number of estimated (free) parameters, and ideally will be statistically equivalent to the original model (although sometimes reparameterization will involve approximation).

In the methodological literature there is a history of reparameterizing conventional models to aid in addressing specific substantive questions. For example, using SEM, Choi et al. (2009) reparameterized a logistic model of change to estimate lower and upper asymptotes, surge points, and jerk points. The latter two parameters represent points in time corresponding to key points of change in the logistic trend. Cudeck and du Toit (2002) reparameterized the common and familiar quadratic curve to estimate the intercept, the time at which the curve attains its maximum/minimum, and the predicted value at that point in time, using single- and multilevel regression modeling. Using multilevel modeling, Rausch (2004, 2008) reparameterized the negative exponential curve to estimate a "half-life" parameter, which represents the amount of time that must elapse for the mean trend to reach a point halfway between the current point and the upper asymptote. Finally, Harring, Cudeck, and du Toit (2006) reparameterized two-segment linear spline models so that the knot (or transition point) is estimated directly as a parameter. In each of these examples, the objective of reparameterization is to recast some aspect of change as either a model parameter or as a random effect that varies and covaries across sampled cases.

Despite the evident potential and usefulness of reparameterization, it is rarely applied outside the methodological literature. We can speculate about why this strategy has failed to catch on among social scientists. First, reparameterization typically has been demonstrated in the context of a single functional form in isolation. It may not be clear to potential users how essentially the same procedure could be adapted to a variety of functional forms. Second, when reparameterization has appeared in prior literature, it typically has not been the primary focus of the study, but rather a tool used to achieve a specific end (e.g., enhancing the probability of successful convergence or isolating a scientifically interesting aspect of the functional form). Third, there have been too few linkages with applied topics to motivate substantive researchers to make the extra effort to use unconventional model specifications.

Clearly, there is a need for a general explanatory framework for reparameterizing models, the goals and outcome of which are closely tied to substantive questions. Below we describe a general approach for obtaining interpretable reparameterizations using LGM. In broad strokes, this general approach involves the following four steps:

1. Reparameterizing the target function to contain substantively important parameters or random coefficients.
2. Linearizing the target function to render it specifiable in SEM software.
3. Specifying the model using the structured latent curve approach.
4. Estimating model parameters (point and interval estimates).

We begin by describing this framework conceptually, and then in the next section illustrate the details in the context of the exemplars of our three classes of reparameterization. Throughout, we highlight the generality of the approach and the new substantively relevant information that can be obtained by using it.

1. Reparameterizing the Target Function

Reparameterization begins with a model expression. The researcher needs to decide what aspect of that model could benefit from explicit quantification. If the aspect of interest is already represented in the model (e.g., slope mean, intercept variance), then there is no need to reparameterize the model. Assuming the desired aspect is not already in the model as a parameter or random effect, the researcher must determine how it could be expressed in terms of existing model parameters. An expression

is derived, often using simple calculus, for that aspect of change in terms of existing parameters; the result is then solved back in terms of an existing parameter and substituted into the original model expression. The result is the reparameterized model.

2. Linearizing the Target Function

In many cases, reparameterization will result in an intrinsically nonlinear function, in the sense that no transformation will render a linear function. For example, some parameters may enter the model embedded within reciprocals, radicals, trigonometric terms, exponents, or logarithms. This intrinsic nonlinearity poses a practical problem in latent growth curve modeling because SEM is a fundamentally linear framework. Hence, the researcher may need to "linearize" the function to enable fitting the model in SEM software.

To linearize the target function, we approximate it with a first-order Taylor series expansion, which can be described simply as consisting of the target function itself (evaluated at the parameter estimates) plus the linear sum of coefficients (e.g., intercepts, slopes, asymptotes) in the target function times partial first derivatives of the target function taken with respect to each coefficient. Taylor series linearization has long been used in fitting nonlinear regression and nonlinear mixed models (e.g., Beal & Sheiner, 1982; Davidian & Giltinan, 1995; Hand & Crowder, 1996).

3. Specifying the Model

As a third step, we employ the principles of *structured latent curve modeling* (SLCM) (Browne, 1993; Browne & du Toit, 1991) to rearrange the linearized function in a way that makes it possible to specify the model using SEM. SLCM involves treating the partial derivatives obtained in Step 2 as factor loadings. New parameters that conceptually reflect person-level (level-2) characteristics can be treated as fixed values or estimated parameters. New within-person (level-1) aspects of a reparameterized model may be treated as fixed values, estimated parameters, or random coefficients that vary from person to person. In fact, nearly any aspect of growth can be treated as a random coefficient in this framework.

4. Estimating Model Parameters

Once the model is specified, it can be fit using SEM software capable of imposing nonlinear constraints.[1] It is important to note that the framework described here accommodates missing data that are missing at random, and can be modified to accommodate individually varying occasions of measurement. Both capabilities are made possible by the use of full-information maximum likelihood (FIML) estimation.

EXEMPLARY APPLICATIONS

Thus far, we have covered in conceptual terms four steps that can be taken to proceed from a conventionally parameterized latent growth model to a reparameterized model with parameters serving specific interpretive purposes. Now that the groundwork has been laid, we present concrete details in the context of the exemplars in our three illustrative classes of reparameterization.

Quantifying Homogeneity or Heterogeneity of Individuals

Our first exemplar is of the broader class of reparameterizations of longitudinal models that quantify homogeneity or heterogeneity of individuals. Clinicians and education researchers are often interested in tracking the degree of children's affiliation with delinquent peers. In particular, researchers may wish to locate the point in time when children are the most similar to each other in their degree of affiliation with delinquent peers, before their trajectories begin to diverge from one another. Locating this point may help clinicians properly time interventions to delay or prevent negative behaviors that tend to spread through peer associations, such as drug use, truancy, and juvenile crime.

More generally, the point in time at which individuals demonstrate the greatest similarity is termed the *aperture* (Hancock & Choi, 2006). The aperture has previously been discussed by Rogosa and Willett (1985) (who called it the *centering point*) and by Mehta and West (2000). The aperture may be directly estimated in a number of ways. We use a method based on the four steps described in the preceding section. To illustrate the method, we make use of peer affiliation data reported by Stoolmiller (1994) and subsequently analyzed by Hancock and Choi (2006).

First, we parameterize the linear latent growth model (with random intercepts and slopes) such that the point in time corresponding to the minimum model-implied variance is represented as a parameter in the model. Note that there is only one such point in time in a linear trajectory model, and it characterizes the entire sample, so it cannot be treated as a random effect that varies across people. We begin with a model expression for an

unconditional latent growth model, which may be expressed in an equation for the outcome y at time i for individual j:

$$y_{ij} = \eta_{1j} + \eta_{2j}\left(t_{ij} - t^*\right) + \varepsilon_{ij}. \tag{2.1}$$

In equation 2.1, y_{ij} is delinquent peer association (a measure of aggregated child, parent, and teacher responses); t_{ij} is the value of time (in this example, grade in school); t^* is the time chosen as the origin (zero point) of the time variable for all subjects, η_{1j} and η_{2j} are the latent intercept and slope, respectively, for subject j; and ε_{ij} is an occasion-specific disturbance term, here assumed to have homoscedastic[2] variance over time and to be uncorrelated across occasions, with $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$. The latent growth factors are assumed to be jointly normally distributed:

$$\begin{bmatrix} \eta_{1j} \\ \eta_{2j} \end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \psi_{11} & \\ \psi_{21} & \psi_{22} \end{bmatrix}\right). \tag{2.2}$$

Thus, there are six parameters (two factor means, three level-2 (co)variances, and a level-1 disturbance variance). The model is depicted graphically in Figure 2.1.

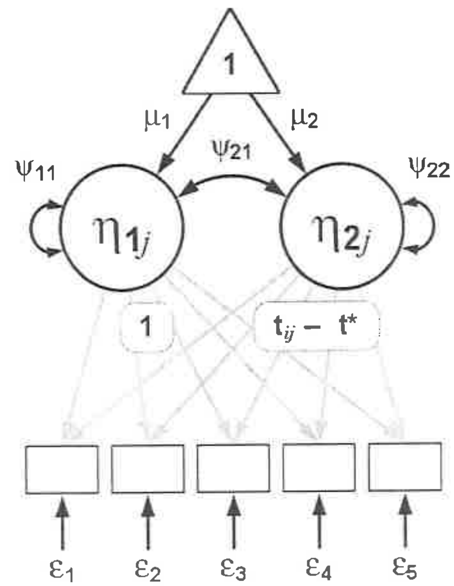

Figure 2.1 A linear latent growth curve model for five repeated measures.

In this example we would like to estimate the aperture—the value of t_{ij} at which children are the most similar to one another in terms of affiliation with delinquent peers—as a model parameter. The aperture occurs where the model-implied variance of y_{ij} is the smallest, a characteristic upon which we can capitalize. The first step, then, is to express the model-implied variance of y_{ij} in terms of existing model parameters. That variance is:

$$\begin{aligned}\sigma_y^2 &= \mathrm{var}\left[\eta_{1j} + \eta_{2j}\left(t_{ij} - t^*\right) + \varepsilon_{ij}\right] \\ &= \mathrm{var}\left(\eta_{1j}\right) + 2\left(t_{ij} - t^*\right)\mathrm{cov}\left(\eta_{1j}, \eta_{2j}\right) + \left(t_{ij} - t^*\right)^2 \mathrm{var}\left(\eta_{2j}\right) + \mathrm{var}\left(\varepsilon_{ij}\right) \\ &= \psi_{11} + 2\left(t_{ij} - t^*\right)\psi_{21} + \left(t_{ij} - t^*\right)^2 \psi_{22} + \sigma_\varepsilon^2.\end{aligned} \tag{2.3}$$

To find the temporal reference point of t_{ij} at which the time variable would have to be centered in order to achieve minimum variance (i.e., t^*), we apply elementary calculus, setting the first partial derivative of equation 2.3 with respect to centered time ($t_{ij} - t^*$) equal to zero:

$$\begin{aligned}\frac{\partial \sigma_y^2}{\partial\left(t_{ij} - t^*\right)} &= 2\psi_{21} + 2\left(t_{ij} - t^*\right)\psi_{22} \\ 0 &= \psi_{21} + \left(t_{ij} - t^*\right)\psi_{22} \\ \left(t_{ij} - \eta_a\right) &= -\frac{\psi_{21}}{\psi_{22}}.\end{aligned} \tag{2.4}$$

The quantity η_a is the aperture. Notice also that if we center at η_a, $\psi_{21} = 0$. This permits us to reexpress the model equivalently as:

$$y_{ij} = \eta_{1j} + \eta_{2j}\left(t_{ij} - \eta_a\right) + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N\left(0, \sigma_\varepsilon^2\right) \tag{2.5}$$

where

$$\begin{bmatrix}\eta_{1j} \\ \eta_{2j} \\ \eta_a\end{bmatrix} \sim \mathrm{MVN}\left(\begin{bmatrix}\mu_1 \\ \mu_2 \\ \mu_a\end{bmatrix}, \begin{bmatrix}\psi_{11} & & \\ 0 & \psi_{22} & \\ 0 & 0 & 0\end{bmatrix}\right). \tag{2.6}$$

The reparameterized model in equations 2.5 and 2.6 still has six parameters; we have simply sacrificed the estimation of ψ_{21} (which we know to be zero at the aperture; see equation 2.4) for the ability to estimate the aperture, parameterized as the mean of a latent variable with zero variance.

The second step is to linearize the model in equations 2.5 and 2.6 in order to make it possible to specify the model in SEM software. The model is already in linear form, so no linearization is technically necessary here, but we apply linearization anyway to illustrate the method in a simple context. A first-order Taylor series approximation is given by the function:

$$\tilde{y} \approx y\big|_{\mu} + \sum_p \frac{\partial y}{\partial \eta_{pj}}\bigg|_{\mu_p}\left(\eta_{pj} - \mu_p\right) \tag{2.7}$$

where $y|_{\mu}$ is the target function evaluated at the parameters,

$$\frac{\partial y}{\partial \eta_{pj}}\bigg|_{\mu_p}$$

is the first partial derivative of the reparameterized model with respect to the pth coefficient evaluated at the coefficient means, and $(\eta_{pj} - \mu_p)$ is the pth mean-centered growth coefficient. In the reparameterized model there are three coefficients (η_{1j}, η_{2j}, and η_a), and the required partial derivatives evaluated at the coefficient means are, respectively, 1, $(t_{ij} - \mu_a)$, and $-\mu_2$. Although it may not be immediately obvious, this linearized form adheres to the standard expression of the confirmatory factor model, where $y|_{\mu} = \Lambda\mu$,

$$\sum_p \frac{\partial y}{\partial \eta_{pj}}\bigg|_{\mu_p}\left(\eta_{pj} - \mu_p\right) = \Lambda\eta_j^*,$$

and η_j^* is a vector of *mean-centered* latent variables. The loadings in Λ are the derivatives evaluated at each occasion of measurement:

$$\Lambda = \left[\begin{array}{c:c:c}1 & (0 - \mu_a) & -\mu_2 \\ 1 & (1 - \mu_a) & -\mu_2 \\ \cdots & \cdots & \cdots \\ 1 & (T - \mu_a) & -\mu_2\end{array}\right]. \tag{2.8}$$

The reparameterized model is depicted in Figure 2.2 in a somewhat simplified form.[3]

Three aspects of change (intercept, linear slope, and aperture) are represented as factors in the reparameterized model. The intercept and slope are represented as random coefficients that do not covary, whereas the ap-

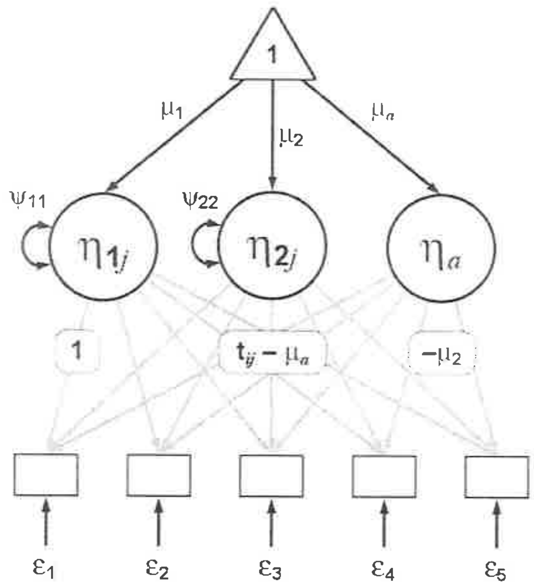

Figure 2.2 A reparameterized linear latent growth curve model for five repeated measures, with the aperture parameterized as μ_a.

erture is represented as a fixed coefficient (a latent variable with zero variance) because it cannot vary across individuals.

Applying this reparameterized model to Stoolmiller's (1994) data, we find that the model fits well ($\chi^2_8 = 1.725$, $p = .988$; RMSEA = .000, 90% CI = [.000,.000], NNFI = 1.277)[4] and yields the following parameter estimates (and standard errors):

$$\begin{bmatrix} \eta_{1j} \\ \eta_{2j} \\ \eta_a \end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix} 4.048 & (.489) \\ 0.229 & (.067) \\ 3.668 & (.926) \end{bmatrix}, \begin{bmatrix} 5.819\ (1.168) & & \\ 0 & .290\ (.098) & \\ 0 & 0 & 0 \end{bmatrix} \right) \quad (2.9)$$

and $\hat{\sigma}^2_\varepsilon = 5.117(.365)$. The aperture is estimated to occur at $t_a = 3.668$; that is, according to the model, children are most similar in their affiliation with delinquent peers approximately ⅔ of the way through the third grade. Therefore, third grade might be a good opportunity to implement interventions designed to inoculate children against peer influence and prevent them from engaging in negative behaviors that are spread through peer associations. LISREL syntax is provided at http://quantpsy.org/pubs/preacher_hancock_2012_code.pdf.

This may seem like a rather involved procedure for estimating just one additional parameter. We agree, and certainly there are easier ways to estimate the aperture parameter. We merely used the aperture to illustrate a general procedure that can be used in much more complex settings, as we illustrate next. One advantage of using the specific approach to reparameterization we described in steps 1–4 above in the case of apertures is that the procedure not only yields a point estimate and an interval estimate, but also treats the aperture as a variable in the model (η_a). Representing the aperture as a variable allows the researcher to test moderation hypotheses; for example, we can test whether the timing of greatest similarity differs between boys and girls, or as a function of social network size or socioeconomic status.

Estimation and Prediction of Aspects of Change

Our second example is also drawn from the field of education. We make use of the Early Childhood Longitudinal Study, Kindergarten Class of 1998–1999 (ECLS-K) data (Tourangeau, Nord, Lê, Sorongon, & Najarian, 2009). The ECLS-K data span kindergarten through the eighth grade. The limited dataset we use includes data on math, reading, and gender. Here we focus on modeling the nonlinear trend in math skills assessed in the fall and spring of kindergarten, fall and spring of first grade, and spring of third, fifth, and eighth grades. A random subset of 250 children's scores are depicted in Figure 2.3.

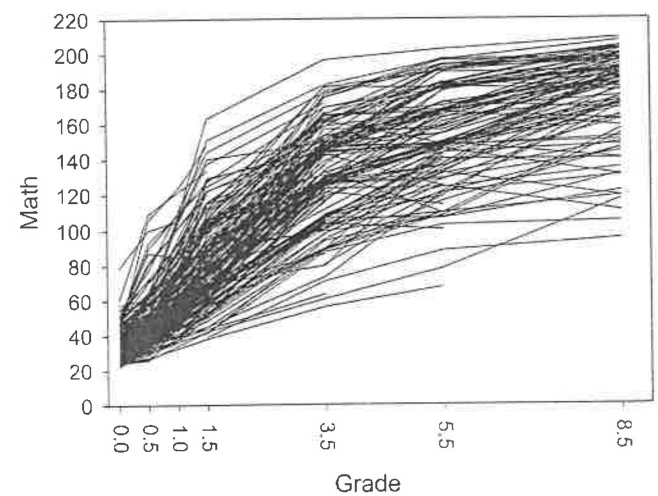

Figure 2.3 A random sample of 250 children's observed mathematics score trajectories from the ECLS-K data.

An appropriate model for learning data such as those in ECLS-K might be the sigmoidal Gompertz curve, which increases slowly from a lower asymptote to a region of relatively faster change, followed by a gradual approach to an upper asymptote. Unlike logistic functions, Gompertz curves need not be symmetric about their points of inflection—a potentially realistic reflection of actual learning rates. There are several versions of the Gompertz curve, some with three parameters (e.g., Browne, 1993; Gompertz, 1825; Sit & Poulin-Costello, 1994; Winsor, 1932) and some with four (e.g., Grimm, Ram, & Estabrook, 2010). We begin with a parameterization of the Gompertz curve presented by Sit and Poulin-Costello (1994), with all coefficients random:

$$y_{ij} = \eta_{1j}\exp\left(-\exp\left(\eta_{2j} - \eta_{3j}t_{ij}\right)\right) + \varepsilon_{ij}, \tag{2.10}$$

where η_{1j} is the upper asymptote, η_{2j} controls the shape of the curve, and η_{3j} shifts the curve horizontally.

Of the three parameters in the expression in equation 2.10, only η_{1j} may be of interest to education researchers. However, the curve is governed by two other parameters that are rather more difficult to treat substantively. Thus, we wish to reparameterize the function to replace η_{2j} and η_{3j} with more interesting parameters: the *surge point* (the point in time at which the maximum rate of change occurs; Choi et al., 2009) and the *surge slope* (the slope, or instantaneous rate of change, at that point). These parameters may be of greater interest to researchers; treating the surge point as a random coefficient would enable educators to determine what environmental factors moderate the surge point, and what individual differences predict children's different rates of learning at that point.

We begin with the target function in equation 2.10. In this example, we allow heteroscedastic error variances over time. The first step is to reparameterize the target function to contain the three desired aspects of change (upper asymptote, surge point, and surge slope) as either estimated parameters or random coefficients. For full generality, we choose to treat all three as random coefficients. The target function and its first and second derivatives are:

$$y_{ij} = \eta_{1j}\exp\left(-\exp\left(\eta_{2j} - \eta_{3j}t_{ij}\right)\right) \tag{2.11}$$

$$\frac{\partial y_{ij}}{\partial t_{ij}} = \eta_{1j}\eta_{3j}\exp\left(\eta_{2j} - \eta_{3j}t_{ij}\right)\exp\left(-\exp\left(\eta_{2j} - \eta_{3j}t_{ij}\right)\right)$$

$$\frac{\partial^2 y_{ij}}{\partial t_{ij}^2} = -\eta_{1j}\eta_{3j}^2\exp\left(\eta_{2j} - \eta_{3j}t_{ij}\right)\exp\left(-\exp\left(\eta_{2j} - \eta_{3j}t_{ij}\right)\right)$$

$$+\eta_{1j}\eta_{3j}^2\left(\exp\left(\eta_{2j} - \eta_{3j}t_{ij}\right)\right)^2\exp\left(-\exp\left(\eta_{2j} - \eta_{3j}t_{ij}\right)\right)$$

The surge point is defined to occur where

$$\frac{\partial^2 y_{ij}}{\partial t_{ij}^2} = 0 .$$

Setting the second derivative to zero and solving for t_{ij} yields the new surge point random coefficient t_{0j}:

$$t_{0j} = \frac{\eta_{2j}}{\eta_{3j}}. \tag{2.12}$$

At this point it is possible to express either η_{2j} or η_{3j} in terms of t_{0j}, yielding:

$$\begin{aligned}\eta_{2j} &= t_{0j}\eta_{3j}\\ \eta_{3j} &= \eta_{2j}/t_{0j}.\end{aligned} \tag{2.13}$$

We could replace either η_{2j} or η_{3j}, leading to

$$\begin{aligned}y_{ij} &= \eta_{1j}\exp\left(-\exp\left(t_{0j}\eta_{3j} - \eta_{3j}t_{ij}\right)\right)\\ &= \eta_{1j}\exp\left(-\exp\left(\eta_{3j}\left(t_{0j} - t_{ij}\right)\right)\right)\end{aligned} \tag{2.14}$$

or

$$\begin{aligned}y_{ij} &= \eta_{1j}\exp\left(-\exp\left(\eta_{2j} - \frac{\eta_{2j}}{t_{0j}}t_{ij}\right)\right)\\ &= \eta_{1j}\exp\left(-\exp\left(\eta_{2j}\left(1 - \frac{t_{ij}}{t_{0j}}\right)\right)\right).\end{aligned} \tag{2.15}$$

We chose equation 2.14 because its parameters are potentially of greater interest and are more easily interpretable; it has the additional benefit that it does not involve division by t_{0j}, thus avoiding potential problems of dividing by zero or by very small numbers. The reparameterized function is:

$$y_{ij} = \eta_{1j}\exp\left(-\exp\left(\eta_{3j}\left(t_{0j} - t_{ij}\right)\right)\right), \tag{2.16}$$

where η_{1j} and η_{3j} are defined as before, and t_{0j} is the surge point.

We also want the surge slope as a coefficient in the model. This slope is the first derivative (local linear slope) at the surge point (Choi et al., 2009). We obtain the first derivative using the new parameterization to avoid reintroducing η_{2j} into the model. The target function and its first derivative are:

$$y_{ij} = \eta_{1j} \exp\left(-\exp\left(\eta_{3j}\left(t_{0j} - t_{ij}\right)\right)\right) \tag{2.17}$$

$$\frac{\partial y_{ij}}{\partial t_{ij}} = \eta_{1j}\eta_{3j} \exp\left(\eta_{3j}\left(t_{0j} - t_{ij}\right)\right)\exp\left(-\exp\left(\eta_{3j}\left(t_{0j} - t_{ij}\right)\right)\right)$$

At the surge point t_{0j}, $(t_{0j} - t_{ij}) = 0$, yielding the new surge slope random coefficient s_{0j}:

$$s_{0j} = \eta_{1j}\eta_{3j} \exp(-1). \tag{2.18}$$

Reexpressing η_{3j} in terms of the surge slope, we obtain:

$$s_{0j} = \eta_{1j}\eta_{3j} \exp(-1) \tag{2.19}$$

$$\eta_{3j} = \frac{s_{0j}\mathrm{e}}{\eta_{1j}}$$

where $\mathrm{e} = \exp(1)$, or Euler's number. Replacing η_{3j} in equation 2.17 yields:

$$y_{ij} = \eta_{1j} \exp\left(-\exp\left(\frac{s_{0j}\mathrm{e}}{\eta_{1j}}\left(t_{0j} - t_{ij}\right)\right)\right) \tag{2.20}$$

where η_{1j} and t_{0j} are defined as before, and s_{0j} is the surge slope.

The second step is to linearize the reparameterized model to express it in a form more palatable to SEM software. The first partial derivatives of equation 2.20 with respect to $\eta = [\eta_{1j}\ t_{0j}\ s_{0j}]'$ and evaluated at the coefficient means $\mu = [\eta_1\ t_0\ s_0]'$ are:

$$\frac{\partial y}{\partial \eta} = \begin{bmatrix} \exp\left(-\exp\left(\dfrac{s_0\mathrm{e}\left(t_0 - t_{ij}\right)}{\eta_1}\right)\right) \\ + \dfrac{s_0\mathrm{e}\left(t_0 - t_{ij}\right)\exp\left(\dfrac{s_0\mathrm{e}\left(t_0 - t_{ij}\right)}{\eta_1}\right)\exp\left(-\exp\left(\dfrac{s_0\mathrm{e}\left(t_0 - t_{ij}\right)}{\eta_1}\right)\right)}{\eta_1} \\ -s_0\mathrm{e}\exp\left(\dfrac{s_0\mathrm{e}\left(t_0 - t_{ij}\right)}{\eta_1}\right)\exp\left(-\exp\left(\dfrac{s_0\mathrm{e}\left(t_0 - t_{ij}\right)}{\eta_1}\right)\right) \\ -\mathrm{e}\left(t_0 - t_{ij}\right)\exp\left(s_0\mathrm{e}\left(t_0 - t_{ij}\right)\eta_1^{-1}\right)\exp\left(-\exp\left(s_0\mathrm{e}\left(t_0 - t_{ij}\right)\eta_1^{-1}\right)\right) \end{bmatrix}. \tag{2.21}$$

We use these derivatives as factor loadings, substituting values of t_{ij} where appropriate. The linearized model can then be expressed in matrix form as:

$$y_j = \Lambda\mu + \Lambda\eta_j^* + \varepsilon_j, \tag{2.22}$$

where $\Lambda\mu$, the loading matrix multiplied by the factor means, represents the target model evaluated at the parameter estimates, and $\Lambda\eta_j^*$ represents the deviation of individual j's trajectory from the mean implied by $\Lambda\mu$. The vector ε_j contains occasion-specific residuals for individual j.

In the third step, we make use of SLCM to specify the model in equation 2.22. This involves treating the derivatives in equation 2.21 as factor loadings using software capable of imposing nonlinear constraints. In SLCM, parameters that enter the reparameterized target function linearly have corresponding factor means estimated in μ, whereas those that enter the function nonlinearly have corresponding factor means constrained to zero. As the ambitious reader may derive, η_{1j} and s_{0j} enter the reparameterized target function linearly, whereas t_{0j} enters nonlinearly because the first partial derivative of the function with respect to t_{0j} contains t_{0j}; thus, $\mu = [\eta_1\ 0\ s_0]'$. This step is taken so that $\Lambda\mu$ will equal the desired mean trajectory at the parameter estimates.[5] The model is depicted graphically in Figure 2.4, Panel A. Symbols for elements of the random coefficient covariance matrix are omitted from the figure for simplicity, but can be represented as:

$$\Psi = \begin{bmatrix} \psi_{\eta_1} & & \\ \psi_{t_0,\eta_1} & \psi_{t_0} & \\ \psi_{s_0,\eta_1} & \psi_{s_0,t_0} & \psi_{s_0} \end{bmatrix}. \tag{2.23}$$

This *unconditional* model does not consider gender as a covariate (we include gender next). Fitting the model to data yields:

$$\begin{bmatrix} \eta_{1j} \\ t_{0j} \\ s_{0j} \end{bmatrix} \sim \mathrm{MVN}\left(\begin{bmatrix} 145.775(.232) \\ .000(\mathrm{N/A}) \\ 22.601(.056) \end{bmatrix}, \begin{bmatrix} 407.249\ (9.018) & & \\ -3.069\ (.147) & .250\ (.006) & \\ 93.129\ (1.772) & -2.007\ (.047) & 37.868\ (.563) \end{bmatrix}\right) \tag{2.24}$$

with the variance of ε_{ij} increasing from 12.165 to a maximum of 87.257 in the spring of first grade, and decreasing to 61.129 by the spring of eighth grade. The estimated mean surge point was $\hat{t}_0 = 1.303\ (.005)$, with corresponding surge slope $\hat{s}_0 = 22.601\ (.056)$.

The goal of this example was not only to show how parameters such as the surge point and surge slope may be treated as random coefficients within

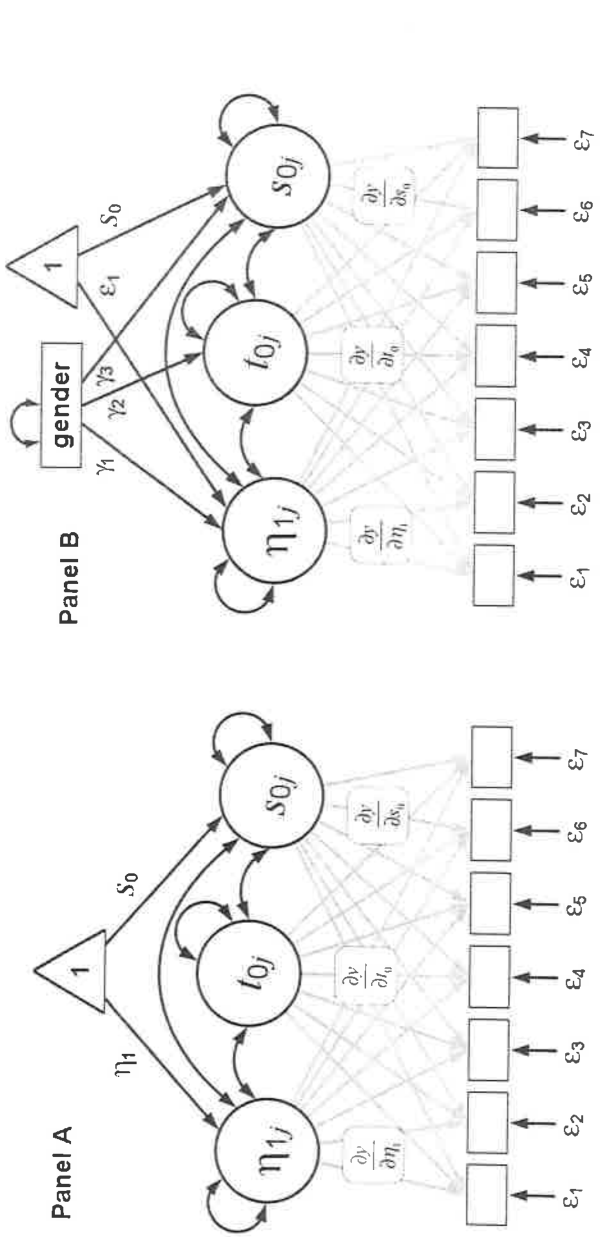

Figure 2.4 Panel A: An unconditional Gompertz growth curve model for mathematics learning among grade school children. For full loading expressions, see equation 2.21. Panel B: The same model is represented with gender as a level-2 predictor of growth coefficients. Because gender is coded (boys = 0, girls = 1), the γ coefficients represent the expected difference for girls with respect to boys.

the SEM/LGM framework, but also to predict individual differences in these coefficients using gender as a level-2 predictor. The structural equation for the random coefficients $\boldsymbol{\eta}_j$ in the *conditional* model may be written as:

$$\boldsymbol{\eta}_j = \boldsymbol{\mu} + \boldsymbol{\Gamma}\mathbf{x}_j + \boldsymbol{\zeta}_j, \tag{2.25}$$

where $\boldsymbol{\Gamma}$ contains structural coefficients linking growth factors to exogenous measured variables in $\mathbf{x}_j$. Regressing the growth factors on gender (boy = 0, girl = 1) yields the following results:

$$\begin{bmatrix} \eta_{0j} \\ t_{0j} \\ s_{0j} \end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix} 146.447(.308) \\ .000(\text{N/A}) \\ 23.238(.079) \end{bmatrix}, \begin{bmatrix} 413.179\ (8.886) & & \\ -3.052\ (.137) & .228\ (.006) & \\ 96.052\ (1.806) & -1.884\ (.047) & 37.838\ (.575) \end{bmatrix} \right) \tag{2.26}$$

$$\hat{\boldsymbol{\Gamma}} = \begin{bmatrix} -1.454\,(.408) \\ .043\,(.008) \\ -1.339\,(.106) \end{bmatrix}. \tag{2.27}$$

Because boys were coded 0 and girls were coded 1, the means in equation 2.26 reflect model-implied values of the growth coefficients for boys. Thus, the mean surge point for boys was $\hat{t}_0 = 1.282(.007)$, with corresponding surge slope $\hat{s}_0 = 23.238(.079)$. Based on equation 2.27, being a girl is associated with a slightly lower value of the upper asymptote (i.e., 146.447 – 1.454 = 144.993), a significantly later surge point (i.e., 1.282 + .043 = 1.325), and a significantly lower surge slope (i.e., 23.238 – 1.339 = 21.899) than for boys. Mplus syntax for both the unconditional and conditional models is provided at http://quantpsy.org/pubs/preacher_hancock_2012_code.pdf.

Using these reparameterized models, we were able to predict individual differences in the timing of key growth events: the point in time at which children are learning math the fastest. The advantages of specifying the moderated Gompertz curve with random coefficients in SEM are clear. Unlike in multilevel modeling, we could additionally consider latent covariates if the situation demanded it.

Estimation and Prediction of Time-Specific Individual Differences

Our third example is drawn from the field of public health. Public health researchers often are concerned with tracking infant development in countries that are susceptible to child malnutrition. Recent studies sponsored by UNICEF (2003, 2008a, 2008b) found that African and South Asian children are particularly likely to suffer from stunted growth due to malnutrition, and stunting can begin *in utero*. The aim of these studies is to identify determinants of optimal and suboptimal infant growth so that interventions can be implemented. In order to identify these determinants, it is necessary to have a model that accurately reflects growth in infant weight. However, we may also want not only to describe individual differences in growth trends, but also to predict these individual differences—at any desired age—with environmental variables such as breastfeeding versus bottle-feeding, rural versus urban status, and others. In this section, we show how to reparameterize a nonlinear function of infant growth so that individual differences can be operationalized as a random effect that is potentially predicted by breastfeeding behavior.

The Cebu Longitudinal Health and Nutrition Survey (Adair & Popkin, 1996; Adair et al., 2011) includes weight data for Filipino infants every 2 months from ages 0 to 24 months (n = 2632).[6] The aims of the survey study included tracking individual differences in weight gain at various ages, as well as discovering the extent to which environmental factors (including maternal breastfeeding behavior) impact weight gain. The first step in our approach to modeling these data was to choose a plausible functional form to describe individual and mean trajectories of change. We selected the Jenss-Bayley model (Jenss & Bayley, 1937) because it was designed specifically to model human growth in the first 6 years of life. The Jenss-Bayley function is well suited for infant growth data because early biological growth often follows an exponential process where growth acts to limit further growth, but the asymptote is not horizontal during the early years as in an ordinary exponential function. One common expression of the Jenss-Bayley model (with random coefficients) is:

$$\begin{aligned} y_{ij} &= \eta_{1j} + \eta_{2j}t_{ij} - \exp(\eta_{3j}^{*} + \eta_{4j}t_{ij}) + \varepsilon_{ij} \\ &= \eta_{1j} + \eta_{2j}t_{ij} - \eta_{3j}\exp\left(\eta_{4j}t_{ij}\right) + \varepsilon_{ij} \end{aligned}, \qquad (2.28)$$

where η_{1j} and η_{2j} are the intercept and slope coefficients for the line defining the asymptote of the function, $\eta_{3j} = \exp(\eta_{3j}^{*})$ is the vertical distance between the intercept of the Jenss-Bayley function and the linear asymptote's intercept, and $\exp(\eta_{4j})$ is the ratio of the acceleration of growth at age t to

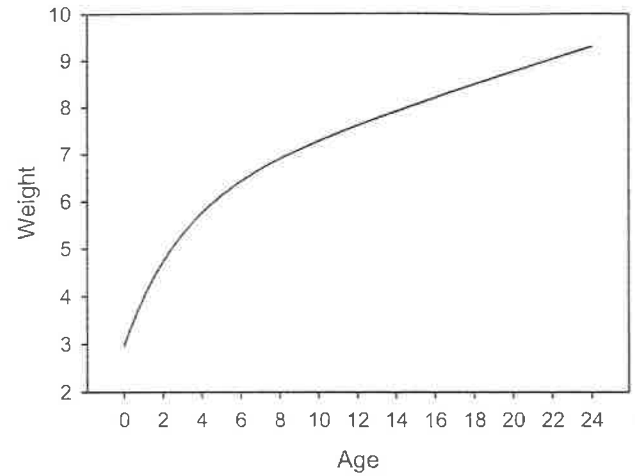

Figure 2.5 The Jenss-Bayley function defined by η_1 = 6, η_2 = .137, η_3 = 1.11, and η_4 = –.334.

that at age $t - 1$. Thus, the Jenss-Bayley function combines exponential and linear growth. A generic Jenss-Bayley function is depicted in Figure 2.5.

For this example we are interested in reparameterizing the Jenss-Bayley function to permit the prediction of individual differences in model-implied weight at any given point in time between 0 and 24 months of age. In addition, the reparameterization needs to be such that this model-implied weight is treated as a random effect in the model. In a linear growth curve model this would be a straightforward task—simply centering age at any desired value would render the intercept factor interpretable as model-implied weight at that age, and the intercept factor then could serve as the dependent variable in a latent structural regression. However, because the Jenss-Bayley model is intrinsically nonlinear we do not have this luxury. Centering age elsewhere would result in a different functional form. With intrinsically nonlinear functional forms like the Jenss-Bayley function, the task is not so simple.

To reparameterize the Jenss-Bayley function, we express model-implied weight at an arbitrary age t^* of the investigator's choosing. The model-implied weight for the jth infant, η_j^*, can then be expressed as:

$$\eta_j^* = \eta_{1j} + \eta_{2j}t^* - \eta_{3j}\exp(\eta_{4j}t^*). \qquad (2.29)$$

This expression, in turn, can be solved for an existing coefficient (we chose η_{1j}) and the result substituted back into equation 2.28, yielding:

$$y_{ij} = \eta_j^* + \eta_{2j}(t_{ij} - t^*) + \eta_{3j}\exp(\eta_{4j}t^*) - \eta_{3j}\exp(\eta_{4j}t_{ij}) + \varepsilon_{ij}. \qquad (2.30)$$

In equation 2.30, we have given up the option of estimating parameters associated with the linear intercept η_{1j}, a random coefficient that may not be particularly interesting, for the ability instead to estimate parameters associated with η_j^*, the random coefficient of central interest.

The model in equation 2.30 is still intrinsically nonlinear, so in order to specify it using SEM we must first obtain a linear approximation. The partial derivatives of equation 2.30 with respect to each coefficient (i.e., η_j^*, η_{2j}, η_{3j}, and η_{4j}), evaluated at the growth factor means, are:

$$\frac{\partial y}{\partial \eta} = \left[\begin{array}{c|c|c|c} 1 & (t_{ij} - t^*) & e^{\mu_3}(e^{\mu_4 t^*} - e^{\mu_4 t_{ij}}) & e^{\mu_3}\left[(t^*)e^{\mu_4 t^*} - t_{ij}e^{\mu_4 t_{ij}}\right] \end{array}\right]'. \qquad (2.31)$$

We use these derivatives as factor loadings:

$$\Lambda = \left[\begin{array}{c|c|c|c} 1 & (0 - t^*) & e^{\mu_3}(e^{\mu_4 t^*} - 1) & e^{\mu_3}(t^*)e^{\mu_4 t^*} \\ 1 & (2 - t^*) & e^{\mu_3}(e^{\mu_4 t^*} - e^{\mu_4(2)}) & e^{\mu_3}\left[(t^*)e^{\mu_4 t^*} - (2)e^{\mu_4(2)}\right] \\ \cdots & \cdots & \cdots & \cdots \\ 1 & (T - t^*) & e^{\mu_3}(e^{\mu_4 t^*} - e^{\mu_4(T)}) & e^{\mu_3}\left[(t^*)e^{\mu_4 t^*} - (T)e^{\mu_4(T)}\right] \end{array}\right] \qquad (2.32)$$

where t^* is a constant time point chosen by the researcher, corresponding to the age at which individual differences in weight are desired.

Specifying the linearized model in SEM proceeds as before. That is, we use the general confirmatory factor analysis model in equation 2.22 with the loading matrix in equation 2.32 and the latent mean vector $\mu = [\mu_1\ \mu_2\ \mu_3\ 0]'$. The mean of the latent variable corresponding to the exponential rate coefficient is set equal to zero so that $\Lambda\mu$ will equal the desired mean trajectory at the parameter estimates (η_{4j} enters the model nonlinearly). The model is depicted graphically in Figure 2.6, Panel A. As with the previous example, elements of the random coefficient covariance matrix are omitted from the figure, but can be represented as:

$$\Psi = \begin{bmatrix} \psi_{\eta^*} & & & \\ \psi_{\eta_2,\eta^*} & \psi_{\eta_2} & & \\ \psi_{\eta_3,\eta^*} & \psi_{\eta_3,\eta_2} & \psi_{\eta_3} & \\ \psi_{\eta_4,\eta^*} & \psi_{\eta_4,\eta_2} & \psi_{\eta_4,\eta_3} & \psi_{\eta_4} \end{bmatrix}. \qquad (2.33)$$

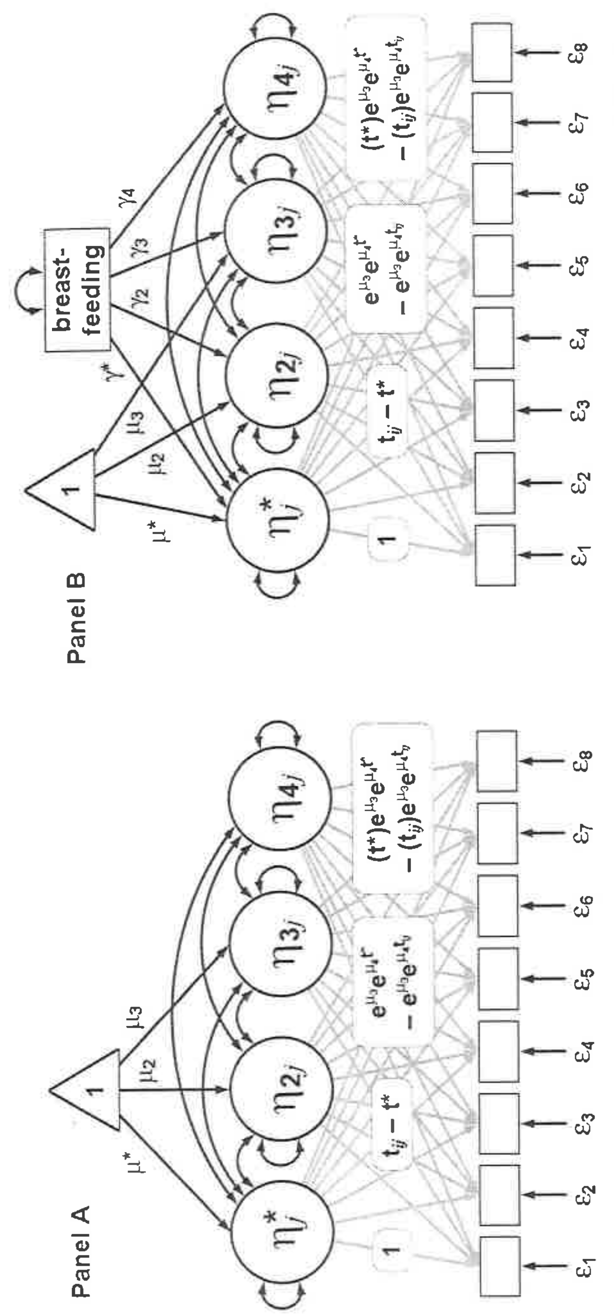

Figure 2.6 Panel A: The reparameterized and linearized Jenss-Bayley model specified as a structured latent curve model. The Cebu data had 13 repeated measurements; only 8 are depicted here for simplicity. The mean of model-implied infant weight is represented by μ^*, whereas the variance of the η_j^* factor represents interindividual variability in this model-implied weight. Panel B: The same model is represented with cumulative breastfeeding as a level-2 predictor of growth coefficients. Because cumulative breastfeeding is the average of a bimonthly binary breastfeeding indicator coded (0 = did not breastfeed the previous day, 1 = breastfed the previous day), γ^* represents the weight difference between hypothetical infants who always received breastfeeding versus those who never received breastfeeding (on days prior to measurement).

The unconditional model does not include breastfeeding as a predictor. As one example of the kind of output generated by this model, consider what happens when we set $t^* = 6$, that is, we wish to estimate the mean and variance of model-implied individual differences in infant weight at age = 6 months. We also permit residuals at adjacent occasions to covary with a constant covariance. Fitting the unconditional model yields adequate fit ($\chi^2_{88} = 1274.024$, $p < .0001$; RMSEA = .072, 90% CI = [.068, .075], NNFI = .982) and parameter estimates (and standard errors) as follows:

$$\begin{bmatrix} \eta^*_i \\ \eta_{2i} \\ \eta_{3i} \\ \eta_{4i} \end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix} 6.778\,(.017) \\ .136\,(.001) \\ 1.841\,(.008) \\ .000\,(\text{N/A}) \end{bmatrix}, \begin{bmatrix} .899\;(.035) & & & \\ -.008\;(.002) & .002\;(<.001) & & \\ .515\;(.026) & -.014\;(.001) & .380\;(.021) & \\ .019\;(.002) & -.001\;(<.001) & .016\;(.001) & .002\;(<.001) \end{bmatrix} \right) \tag{2.34}$$

with

$$\begin{bmatrix} \varepsilon_{ij[t]} \\ \varepsilon_{ij[t-1]} \end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} .102\;(.003) & \\ .028\;(.001) & .102\;(.003) \end{bmatrix} \right), \tag{2.35}$$

and $\hat{\mu}_4 = -.334(.004)$. The model-implied mean weight at age 6 months is 6.778 kg, with a standard deviation of $(.899)^{1/2} = .948$ kg. This procedure can be repeated any number of times, substituting a new t^* each time, yielding a series of models that are equivalent in terms of fit. The model-implied mean weight is plotted at 13 different ages in Figure 2.7, along with a 95% interval based on the estimated time-specific variance and the observed means (i.e., $\hat{\mu}^* \pm 1.96\sqrt{\hat{\psi}_{\eta^*}}$).

The preceding represents an informative, convenient way to model and illustrate the model-implied mean and variance of infant weight at any desired age, even if the desired age falls somewhere between two observations. However, our primary interest—and the reason we reparameterized the model in the first place—lies in predicting individual differences in weight at any given age by cumulative breastfeeding *up to that point*. At every occasion of measurement, mothers were asked whether they had breastfed the previous day (0,1). We defined *cumulative breastfeeding* as the average of all breastfeeding responses up to a given point in time. It serves as an indicator of how much a given infant has been breastfed since birth, relative to other infants of the same age. We introduced cumulative breastfeeding as an

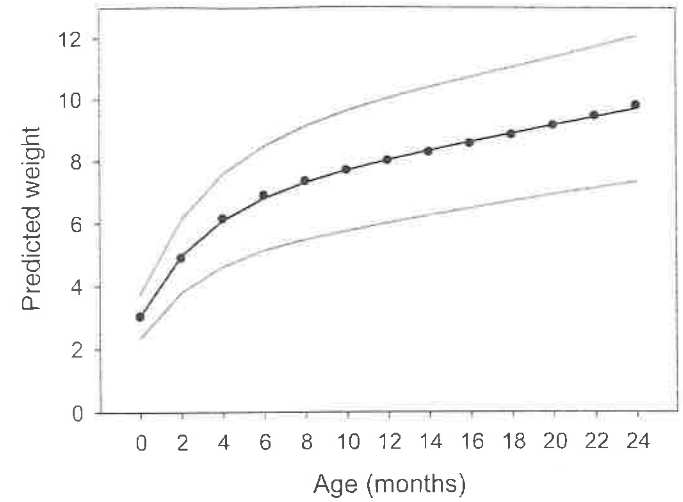

Figure 2.7 Observed and model-implied infant weight at 13 different ages, plotted with 95% interval based on the estimated time-specific variance (i.e., $\hat{\mu}^* \pm 1.96\sqrt{\hat{\psi}_{\eta^*}}$). Solid dots represent observed means.

infant-level predictor of all four random coefficients. Our main interest was in the effect of cumulative breastfeeding on infant weight at particular ages.

As with the second example, the structural equation for the random coefficients η_i in the conditional model may be written as equation 2.25. Regressing the model-implied weight factor (along with the other growth factors) on cumulative breastfeeding at each point in time yields different parameter estimates and model fit at each age because cumulative breastfeeding is a different variable at each occasion of measurement. All of the models fit well by standard criteria (minimum RMSEA = .051 at ages 4–8 months; maximum RMSEA = .062 at birth). The pattern of effects in Figure 2.8 tells us something useful: The effect of cumulative breastfeeding on infant weight is positive in the early months, but negative at 10 months and beyond. A steep drop occurs in the effect of breastfeeding on weight by about the 6-month mark. This corresponds to the age at which the American Academy of Pediatrics, the World Health Organization, and UNICEF recommend supplementing breast milk with solid food. Investigating the reasons for the change from a positive effect to a negative effect by 10 months is beyond the scope of this example, but we may conjecture that infants who still rely heavily on breastfeeding at or beyond 10 months of age are probably experiencing other dietary deficiencies that explain their relatively lower weight. Perhaps mothers need to continue breastfeeding because of other comorbid factors, like poverty, which are linked to these other deficiencies. Mplus syntax for both the

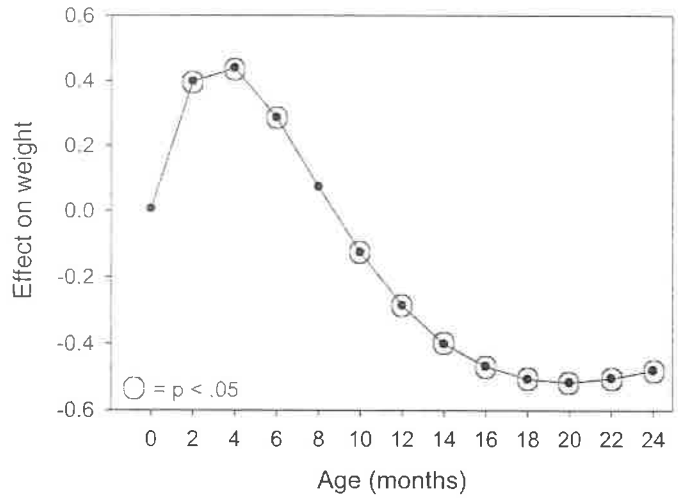

Figure 2.8 The effect of cumulative breastfeeding on model-implied infant weight at 13 ages. Circled slope estimates are statistically significant at $\alpha = .05$.

unconditional and conditional models is provided at http://quantpsy.org/pubs/preacher_hancock_2012_code.pdf.

SUMMARY AND EXTENSIONS

In this chapter we have provided a framework for reparameterizing linear and nonlinear LGMs to yield new parameters and latent variables to answer important substantive questions. Often a model can be reparameterized to yield a point estimate and interval estimate for a new parameter that is of central interest to the researcher because it represents some salient aspect of change that was initially inaccessible. In other models, reparameterization permits treating the new parameter as a random coefficient that varies across individuals.

We presented four steps for using SEM (specifically, the SLCM approach) to model trends using reparameterized models. Again, they are:

1. Reparameterizing the target function to contain substantively important parameters or random coefficients.
2. Linearizing the target function to render it specifiable in SEM software.
3. Specifying the model using the structured latent curve approach.
4. Estimating model parameters (point and interval estimates).

We described this approach conceptually, then gave more detailed explanations in the context of exemplars representing different classes of the kinds of reparameterizations in which researchers might reasonably be interested. Specifically, we illustrated the steps using a clinical/educational psychology example on affiliation with delinquent peers, a mathematics learning example in which aspects of learning could be predicted by gender, and a public health example involving prediction of infant growth.

Reparameterization has the potential to make the already versatile SEM framework even more flexible. Particularly in longitudinal settings, reparameterization permits the researcher to treat virtually any aspect of change as a fixed and known value, an estimated parameter, or in many cases a random coefficient that varies across individuals. The latter two options provide a way to investigate whether these aspects of change are predicted or moderated by level-2 predictors (e.g., mathematics learning or cumulative breastfeeding). These examples illustrated three kinds of reparameterization researchers might be interested in using. The first example focused on estimating parameters associated with homogeneity or heterogeneity; the second example treated three aspects of change as random coefficients; and the third example focused on the estimation and prediction of time-specific individual differences in an outcome. We feel that reparameterization has great potential for helping researchers create models that more closely align with theoretical predictions, and should be emphasized to a greater extent in graduate training.

We now close this chapter by emphasizing that the general approach to reparameterization described in our steps is by no means limited to the classes of reparameterization illustrated in this chapter. Other examples of potential use are readily devised, as follows:

1. Hipp, Bauer, Curran, and Bollen (2004) fit a partially nonlinear growth curve model to seasonal crime trend data in which intrinsically nonlinear parameters were treated as fixed coefficients. Using the reparameterization approach described in this chapter, similar models could be fit to cyclic data treating, for example, instantaneous rate of change as an individual-differences variable, amenable to prediction by person-level characteristics. The variance of this random coefficient could serve as a proxy for cycle synchrony versus asynchrony.
2. The field of metabolism biochemistry is full of examples in which it is of interest to assess the rate at which a drug is absorbed and metabolized. Often the functions used to model these dynamics are complex nonlinear trajectories (see Davidian & Giltinan, 1995). Reparameterization could be used in tandem with SLCM to treat

salient characteristics of growth as random effects, and to embed the trajectory function into a larger causal network.

3. In education research it is often of interest to gauge the rate of learning from year to year when academic years are separated by summer breaks (Entwistle & Alexander, 1992; Hancock & Koran, 2005). When multiphase segmented spline models are used, it is possible through reparameterization to treat the angle separating the slopes of adjacent linear segments (in radians or degrees) as a random coefficient.

We are confident that many more uses for reparameterization will become apparent to researchers, and we look forward to the expanded scope of research questions that these methods will help to answer.

APPENDIX
LISREL and Mplus Code to Accompany Examples

Three Ways to Estimate an Aperture Parameter

Method 1: Lambda shift method

```
LATENT GROWTH CURVE OF HANCOCK & CHOI DATA, EX. 2
DA NI=4 NO=198 MA=CM
CM
11.000
5.860 13.000
6.205 8.094 14.000
6.103 8.798 10.177 16.000
ME
3.3 3.7 4.0 4.2
MO NY=4 NE=2 LY=FU,FI BE=FU,FI TY=FI AL=FR PS=SY,FI TE=DI,FR AP=1
LE
INT SLP
FR PS 1 1 PS 2 2 !PS 2 1
VA 1 LY 1 1 LY 2 1 LY 3 1 LY 4 1
CO LY 1 2 = 4-PA 1
CO LY 2 2 = 6-PA 1
CO LY 3 2 = 7-PA 1
CO LY 4 2 = 8-PA 1
PD
OU IT=5000 AD=OFF ND=4
```

Method 2: Phantom variable approach

```
LATENT GROWTH CURVE OF HANCOCK & CHOI DATA, EX. 2
DA NI=4 NO=198 MA=CM
CM
```



```
11.000
5.860 13.000
6.205 8.094 14.000
6.103 8.798 10.177 16.000
ME
3.3 3.7 4.0 4.2
MO NY=4 NE=3 LY=FU,FI BE=FU,FI TY=FI AL=FI PS=SY,FI TE=DI,FR
LE
INT SLP PHANTOM
FR AL 1 AL 2
PA PS
 1
 0 1
 0 0 0
MA PS
 .5
  0 .5
  0 0 0
PA LY
0 0 0
0 0 0
0 0 0
0 0 0
MA LY
1 4 -1
1 6 -1
1 7 -1
1 8 -1
PA BE
 0 0 0
 0 0 0
 0 1 0
MA BE
 0 0 0
 0 0 0
 0 -4.3 0
PD
OU IT=5000 AD=OFF ND=4
```

Method 3: Structured latent curve approach

```
LATENT GROWTH CURVE OF HANCOCK & CHOI DATA, EX. 2
DA NI=4 NO=198 MA=CM
CM
11.000
5.860 13.000
6.205 8.094 14.000
6.103 8.798 10.177 16.000
ME
3.3 3.7 4.0 4.2
MO NY=4 NE=3 LY=FU,FI TY=FI AL=FR PS=SY,FI TE=DI,FR AP=1
LE
```

```
INT SLP APERTURE
PA PS
 1
 0 1
 0 0 0
MA PS
 .5
  0 .5
  0  0  0
PA LY
 0 0 0
 0 0 0
 0 0 0
 0 0 0
MA LY
 1 0 0
 1 0 0
 1 0 0
 1 0 0
CO LY 1 2 = 4-PA 1
CO LY 2 2 = 6-PA 1
CO LY 3 2 = 7-PA 1
CO LY 4 2 = 8-PA 1
CO LY 1 3 = -1*AL(2)
CO LY 2 3 = -1*AL(2)
CO LY 3 3 = -1*AL(2)
CO LY 4 3 = -1*AL(2)
CO AL 3 = PA 1
PD
OU IT=5000 AD=OFF ND=4
```

A Reparameterized Gompertz Structured Latent Growth Curve Model with Random Coefficients

Example application to ECLS-K mathematics data (kindergarten through eighth grade). Random coefficients represent the upper asymptote, surge point, and surge slope (see Choi et al., 2009) for definitions of these terms.

```
TITLE: ECLS-K math, reparameterized Gompertz curve;
DATA: FILE = eclsk.dat;
VARIABLE: NAMES = id
!gender !0=M/1=F; uncomment gender for conditional model
m_fk m_sk m_f1 m_s1 m_s3 m_s5 m_s8;
USEVARIABLES ARE gender m_fk-m_s8;
MISSING = .;
ANALYSIS:
ALGORITHM = INTEGRATION;
INTEGRATION = MONTECARLO;
```



```
MODEL:

!Factor Loadings
g0 BY m_fk*(Lg01)
m_sk-m_s8(Lg02-Lg07);
t0 BY m_fk*(Lt01)
m_sk-m_s8(Lt02-Lt07);
g3 BY m_fk*(Lg31)
m_sk-m_s8(Lg32-Lg37);

!Means
[m_fk-m_s8@0]; [g0*147](mu_g0); [t0@0]; [g3*22](mu_g3);

!Variances and covariances
m_fk-m_s8; g0*495 t0*.24 g3; g0 WITH t0 g3; t0 WITH g3;

!Regressions
!g0 t0 g3 ON gender; !uncomment line for conditional model

MODEL CONSTRAINT:
NEW(mu_t0*.7); !Introduce mean of surge point

!Asymptote loadings
Lg01 = exp(-1*exp(((mu_g3*exp(1)*(mu_t0-0.0))/mu_g0)))
+((mu_g3*exp(1)*(mu_t0-0.0)*exp(mu_g3*exp(1)*(mu_t0-0.0)/mu_g0)
*exp(-1*exp(mu_g3*exp(1)*(mu_t0-0.0)/mu_g0)))/mu_g0);
Lg02 = exp(-1*exp(((mu_g3*exp(1)*(mu_t0-0.5))/mu_g0)))
+((mu_g3*exp(1)*(mu_t0-0.5)*exp(mu_g3*exp(1)*(mu_t0-0.5)/mu_g0)
*exp(-1*exp(mu_g3*exp(1)*(mu_t0-0.5)/mu_g0)))/mu_g0);
Lg03 = exp(-1*exp(((mu_g3*exp(1)*(mu_t0-1.0))/mu_g0)))
+((mu_g3*exp(1)*(mu_t0-1.0)*exp(mu_g3*exp(1)*(mu_t0-1.0)/mu_g0)
*exp(-1*exp(mu_g3*exp(1)*(mu_t0-1.0)/mu_g0)))/mu_g0);
Lg04 = exp(-1*exp(((mu_g3*exp(1)*(mu_t0-1.5))/mu_g0)))
+((mu_g3*exp(1)*(mu_t0-1.5)*exp(mu_g3*exp(1)*(mu_t0-1.5)/mu_g0)
*exp(-1*exp(mu_g3*exp(1)*(mu_t0-1.5)/mu_g0)))/mu_g0);
Lg05 = exp(-1*exp(((mu_g3*exp(1)*(mu_t0-3.5))/mu_g0)))
+((mu_g3*exp(1)*(mu_t0-3.5)*exp(mu_g3*exp(1)*(mu_t0-3.5)/mu_g0)
*exp(-1*exp(mu_g3*exp(1)*(mu_t0-3.5)/mu_g0)))/mu_g0);
Lg06 = exp(-1*exp(((mu_g3*exp(1)*(mu_t0-5.5))/mu_g0)))
+((mu_g3*exp(1)*(mu_t0-5.5)*exp(mu_g3*exp(1)*(mu_t0-5.5)/mu_g0)
*exp(-1*exp(mu_g3*exp(1)*(mu_t0-5.5)/mu_g0)))/mu_g0);
Lg07 = exp(-1*exp(((mu_g3*exp(1)*(mu_t0-8.5))/mu_g0)))
+((mu_g3*exp(1)*(mu_t0-8.5)*exp(mu_g3*exp(1)*(mu_t0-8.5)/mu_g0)
*exp(-1*exp(mu_g3*exp(1)*(mu_t0-8.5)/mu_g0)))/mu_g0);

!Surge point loadings
Lt01 = -1*mu_g3*exp(1)*exp(mu_g3*exp(1)*(mu_t0-0.0)/mu_g0)
*exp(-1*exp(mu_g3*exp(1)*(mu_t0-0.0)/mu_g0));
Lt02 = -1*mu_g3*exp(1)*exp(mu_g3*exp(1)*(mu_t0-0.5)/mu_g0)
*exp(-1*exp(mu_g3*exp(1)*(mu_t0-0.5)/mu_g0));
Lt03 = -1*mu_g3*exp(1)*exp(mu_g3*exp(1)*(mu_t0-1.0)/mu_g0)
*exp(-1*exp(mu_g3*exp(1)*(mu_t0-1.0)/mu_g0));
```

```
Lt04 = -1*mu_g3*exp(1)*exp(mu_g3*exp(1)*(mu_t0-1.5)/mu_g0)
*exp(-1*exp(mu_g3*exp(1)*(mu_t0-1.5)/mu_g0));
Lt05 = -1*mu_g3*exp(1)*exp(mu_g3*exp(1)*(mu_t0-3.5)/mu_g0)
*exp(-1*exp(mu_g3*exp(1)*(mu_t0-3.5)/mu_g0));
Lt06 = -1*mu_g3*exp(1)*exp(mu_g3*exp(1)*(mu_t0-5.5)/mu_g0)
*exp(-1*exp(mu_g3*exp(1)*(mu_t0-5.5)/mu_g0));
Lt07 = -1*mu_g3*exp(1)*exp(mu_g3*exp(1)*(mu_t0-8.5)/mu_g0)
*exp(-1*exp(mu_g3*exp(1)*(mu_t0-8.5)/mu_g0));

!Surge slope loadings
Lg31 = -1*exp(1)*(mu_t0-0.0)*exp(mu_g3*exp(1)*(mu_t0-0.0)/mu_g0)
*exp(-1*exp(mu_g3*exp(1)*(mu_t0-0.0)/mu_g0));
Lg32 = -1*exp(1)*(mu_t0-0.5)*exp(mu_g3*exp(1)*(mu_t0-0.5)/mu_g0)
*exp(-1*exp(mu_g3*exp(1)*(mu_t0-0.5)/mu_g0));
Lg33 = -1*exp(1)*(mu_t0-1.0)*exp(mu_g3*exp(1)*(mu_t0-1.0)/mu_g0)
*exp(-1*exp(mu_g3*exp(1)*(mu_t0-1.0)/mu_g0));
Lg34 = -1*exp(1)*(mu_t0-1.5)*exp(mu_g3*exp(1)*(mu_t0-1.5)/mu_g0)
*exp(-1*exp(mu_g3*exp(1)*(mu_t0-1.5)/mu_g0));
Lg35 = -1*exp(1)*(mu_t0-3.5)*exp(mu_g3*exp(1)*(mu_t0-3.5)/mu_g0)
*exp(-1*exp(mu_g3*exp(1)*(mu_t0-3.5)/mu_g0));
Lg36 = -1*exp(1)*(mu_t0-5.5)*exp(mu_g3*exp(1)*(mu_t0-5.5)/mu_g0)
*exp(-1*exp(mu_g3*exp(1)*(mu_t0-5.5)/mu_g0));
Lg37 = -1*exp(1)*(mu_t0-8.5)*exp(mu_g3*exp(1)*(mu_t0-8.5)/mu_g0)
*exp(-1*exp(mu_g3*exp(1)*(mu_t0-8.5)/mu_g0));
```

The Jenss-Bayley Model Reparameterized to Estimate the Effect of Cumulative Breastfeeding on Infant Weight at Any Desired Age

We did not have permission to post the Cebu infant data, but we provide Mplus code below to show how to estimate the model.

```
TITLE: cebu growth data (jenss-bayley) with mobile intercept;
DATA: FILE IS cebu_wide_more.dat;
VARIABLE: NAMES ARE id momht rural male age0-age12 br0-br12 h0-h12
w0-w12 b0 b2 b4 b6 b8 b10 b12 b14 b16 b18 b20 b22 b24 xb6 xb8 xb10
xb12 xb14 xb16 xb18 xb20 xb22 xb24;
USEVARIABLES ARE cbf !comment out 'cbf' for unconditional model
x0-x12; !b0-b24 cumul BF means from age 0m; xb6-xb24 from age 6m
MISSING ARE ALL (-999);
USEOBSERVATIONS ARE id NE 1600044; !omit outlier
DEFINE: !below, choose to model height or weight
!x0=h0;x1=h1;x2=h2;x3=h3;x4=h4;x5=h5;x6=h6;
!x7=h7;x8=h8;x9=h9;x10=h10;x11=h11;x12=h12;
x0=w0;x1=w1;x2=w2;x3=w3;x4=w4;x5=w5;x6=w6;
x7=w7;x8=w8;x9=w9;x10=w10;x11=w11;x12=w12;
!comment out next line for unconditional model
cbf=b12; ! <- SET AGE FOR CUMULATIVE BREASTFEEDING (mos.);
ANALYSIS: ESTIMATOR IS ML; ITERATIONS ARE 10000; !BOOTSTRAP IS 300;
MODEL: [x0-x12@0]; x0-x12*.1(v1); x0-x11 PWITH x1-x12*.027(v2);
```



```
fa*1.2; fb*.005; fc*.288; fd*.02; fa WITH fb fc fd;
fb WITH fc fd; fc WITH fd;
[fa*9](mfa); [fb*.13](mfb); [fc*1.2](mfc); [fd@0];
! fa, fb, & fc are linear
fa BY x0-x12@1;
fb BY x0*(b0); fb BY x1-x12(b1-b12);
fc BY x0*(c0); fc BY x1-x12(c1-c12);
fd BY x0*(d0); fd BY x1-x12(d1-d12);
!comment out next line for unconditional model
fa ON cbf*-.53; fb ON cbf*.05; fc ON cbf*-.55; fd ON cbf*-.23;

!jenss-bayley with a mobile intercept
MODEL CONSTRAINT: NEW(mfd*-.334 t0);
t0=12; ! <- SET INTERCEPT LOC. (mos.)

b0=0-t0; c0=exp(mfc+mfd*t0)-exp(mfc+mfd*0);
b1=2-t0; c1=exp(mfc+mfd*t0)-exp(mfc+mfd*2);
b2=4-t0; c2=exp(mfc+mfd*t0)-exp(mfc+mfd*4);
b3=6-t0; c3=exp(mfc+mfd*t0)-exp(mfc+mfd*6);
b4=8-t0; c4=exp(mfc+mfd*t0)-exp(mfc+mfd*8);
b5=10-t0; c5=exp(mfc+mfd*t0)-exp(mfc+mfd*10);
b6=12-t0; c6=exp(mfc+mfd*t0)-exp(mfc+mfd*12);
b7=14-t0; c7=exp(mfc+mfd*t0)-exp(mfc+mfd*14);
b8=16-t0; c8=exp(mfc+mfd*t0)-exp(mfc+mfd*16);
b9=18-t0; c9=exp(mfc+mfd*t0)-exp(mfc+mfd*18);
b10=20-t0; c10=exp(mfc+mfd*t0)-exp(mfc+mfd*20);
b11=22-t0; c11=exp(mfc+mfd*t0)-exp(mfc+mfd*22);
b12=24-t0; c12=exp(mfc+mfd*t0)-exp(mfc+mfd*24);

d0=t0*exp(mfc+mfd*t0)-0*exp(mfc+mfd*0);
d1=t0*exp(mfc+mfd*t0)-2*exp(mfc+mfd*2);
d2=t0*exp(mfc+mfd*t0)-4*exp(mfc+mfd*4);
d3=t0*exp(mfc+mfd*t0)-6*exp(mfc+mfd*6);
d4=t0*exp(mfc+mfd*t0)-8*exp(mfc+mfd*8);
d5=t0*exp(mfc+mfd*t0)-10*exp(mfc+mfd*10);
d6=t0*exp(mfc+mfd*t0)-12*exp(mfc+mfd*12);
d7=t0*exp(mfc+mfd*t0)-14*exp(mfc+mfd*14);
d8=t0*exp(mfc+mfd*t0)-16*exp(mfc+mfd*16);
d9=t0*exp(mfc+mfd*t0)-18*exp(mfc+mfd*18);
d10=t0*exp(mfc+mfd*t0)-20*exp(mfc+mfd*20);
d11=t0*exp(mfc+mfd*t0)-22*exp(mfc+mfd*22);
d12=t0*exp(mfc+mfd*t0)-24*exp(mfc+mfd*24);

OUTPUT: TECH1 TECH3 STDYX; !CINTERVAL(BCBOOTSTRAP);
```

ACKNOWLEDGMENT

We thank Kevin Grimm and Jeff Harring for valuable help with the Gompertz curve example.

NOTES

1. SEM software packages capable of imposing nonlinear constraints currently include LISREL, Mplus, SAS PROC CALIS and PROC TCALIS, Mx, and OpenMx.
2. We assume homoscedasticity for parsimony, not because the model requires it.
3. The Taylor series approximation uses mean-centered latent variables, but the diagrams and model specifications used in this chapter simplify matters by giving some of the latent variables mean parameters.
4. Computation of NNFI required a nonstandard null model; see Widaman and Thompson (2003).
5. If the parameters entering the function nonlinearly are fixed coefficients rather than random coefficients, the model is termed *conditionally linear* (Blozis & Cudeck, 1999).
6. We omitted from our analyses one infant who grew to be nearly twice the weight of his peers, yielding $n = 2{,}631$.

REFERENCES

Adair, L. S., & Popkin, B. M. (1996). Low birth weight reduces the likelihood of breast-feeding among Filipino infants. *Journal of Nutrition, 126*, 103–112.

Adair, L. S., Popkin, B. M., Akin, J. S., Guilkey, D. K., Gultiano, S., Borja, J., et al. (2011). Cohort profile: The Cebu longitudinal health and nutrition survey. *International Journal of Epidemiology, 40*, 619–625.

Beal, S. L., & Sheiner, L. B. (1982), Estimating population kinetics. *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.

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, The Netherlands: Elsevier-North Holland.

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*. Washington, DC: American Psychological Association.

Choi, J., Harring, J. R., & Hancock, G. R. (2009). Latent growth modeling for logistic response functions. *Multivariate Behavioral Research, 44*, 620–645.

Cudeck, R., & du Toit, S. H. C. (2002). A version of quadratic regression with interpretable parameters. *Multivariate Behavioral Research, 37*, 501–519.

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

Entwistle, D. R., & Alexander, K. L. (1992). Summer setback: Race, poverty, school composition, and mathematics achievement in the first two years of school. *American Sociological Review, 57*, 72–84.

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London, 115*, 513–585.

Grimm, K. J., & Ram, N. (2009). Nonlinear growth models in M*plus* and SAS. *Structural Equation Modeling: A Multidisciplinary Journal, 16*, 676–701.

Grimm, K. J., Ram, N., & Estabrook, R. (2010). Nonlinear structured growth mixture models in Mplus and OpenMx. *Multivariate Behavioral Research, 45*, 887–909.

Hancock, G. R., & Choi, J. (2006). A vernacular for linear latent growth models. *Structural Equation Modeling: A Multidisciplinary Journal, 13*, 352–377.

Hancock, G. R., & Koran, J. M. (2005, April). *A latent growth modeling strategy for interrupted longitudinal processes*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada.

Hand, D., & Crowder, M. (1996). *Practical longitudinal data analysis*. Boca Raton, FL: Chapman & Hall/CRC.

Harring, J. R., Cudeck, R., & du Toit, S. H. C. (2006). Fitting partially nonlinear random coefficient models as SEMs. *Multivariate Behavioral Research, 41*, 579–596.

Hipp, J., Bauer, D. J., Curran, P. J., & Bollen, K. A. (2004). Crimes of opportunity or crimes of emotion: Testing two explanations of seasonal change in crime. *Social Forces, 82*, 1333–1372.

Jenss, R. M., & Bayley, N. (1937). A mathematical method for studying the growth of a child. *Human Biology, 9*, 556–563.

Mehta, P. D., & West, S. G. (2000). Putting the individual back in individual growth curves. *Psychological Methods, 5*, 23–43.

Ram, N., & Grimm, K. (2007). Using simple and complex growth models to articulate developmental change: Matching theory to method. *International Journal of Behavioral Development, 31*, 303–316.

Rausch, J. R. (2004). *Designing longitudinal studies of negative exponential growth according to the reliabilities of growth parameter estimators*. Unpublished thesis, University of Notre Dame, Notre Dame, IN.

Rausch, J. R. (2008). *Parametrizations for the negative exponential growth model*. Paper presented at the international meeting of the Psychometric Society, Durham, NH.

Rogosa, D. R., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50*, 203–228.

Sit, V., & Poulin-Costello, M. (1994). *Catalog of curves for curve fitting*. Biometrics Information, Handbook no. 4, B.C. Ministry of Forests, Victoria, BC, Canada.

Stoolmiller, M. (1994). Antisocial behavior, delinquent peer association and unsupervised wandering for boys: Growth and change from childhood to early adolescence. *Multivariate Behavioral Research, 29*, 263–288.

Tourangeau, K., Nord, C., Lê, T., Sorongon, A. G., & Najarian, M. (2009). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K), Combined User's Manual for the ECLS-K Eighth-Grade and K–8 Full Sample Data Files and Electronic Codebooks (NCES 2009–004)*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

UNICEF. (2003). *Drought, HIV/AIDS and child malnutrition in southern Africa: Part 1, Preliminary analysis of nutritional data on the humanitarian crisis*. Retrieved from www.tulane.edu/~internut/NIPSA%201.pdf.

UNICEF. (2008a). *The state of Africa's children 2008: Child survival.* Retrieved from www.unicef.org/sowc08/docs/SOAC_2008_EN_A4.pdf.

UNICEF. (2008b). *The state of Asia-Pacific's children 2008: Child survival.* Retrieved from www.unicef.org/publications/files/SOAPC_2008_080408.pdf.

Widaman, K. F., & Thompson, J. S. (2003). On specifying the null model for incremental fit indices in structural equation modeling. *Psychological Methods, 8*, 16–37.

Winsor, C. P. (1932). The Gompertz curve as a growth curve. *Proceedings of the National Academy of Sciences, 18*, 1–8.

Advances in Longitudinal Methods in the Social and Behavioral Sciences

edited by

Jeffrey R. Harring
University of Maryland

Gregory R. Hancock
University of Maryland

INFORMATION AGE PUBLISHING, INC.
Charlotte, NC • www.infoagepub.com