LATENT GROWTH CURVE MODELING

Kristopher J. Preacher
University of Kansas

Aaron L. Wichman
Ohio State University

Robert C. MacCallum
University of North Carolina at Chapel Hill

Nancy E. Briggs
University of Adelaide

CHAPTER 1. INTRODUCTION

A fundamental observation in the social and behavioral sciences is that people change over time, but not necessarily in the same way or at the same rate. For example, verbal ability increases steadily throughout the elementary school years, but it does not increase at the same rate for all students. There are individual differences in the rate and direction of change in many contexts, and these individual differences in change are often of scientific or practical interest. Change over time can be measured in seconds, as in studies of cardiac reactivity, or in decades, as in life span development studies. Marital arguments, for example, may cause significant endocrinological change over the course of only a few minutes, but these changes can be quite different for husbands than for wives (Kiecolt-Glaser et al., 1997). Flora and Chassin (2005) examined growth in drug use among adolescents as a function of parent alcoholism, following participants over a number of years through young adulthood. Early approaches to investigating change were very limited in that (a) they focused exclusively either on group-level or on individual-level growth and (b) they addressed only two occasions of measurement, resulting in data too impoverished to allow examination of some of the most basic and interesting hypotheses about change over time.

More extensive data and more advanced statistical methods are often needed to enable scientists to discern and understand not only the shape and direction of change (*trajectory*) but also to identify the sources and consequences of change.

Longitudinal designs can yield valuable information about trends in psychological phenomena as well as individual differences in aspects of change over time. The richness of such data increases with the number of waves of data collection. Willett (1989) and Willett and Sayer (1994) discuss several advantages associated with longitudinal data drawn from multiple waves of data collection, relative to two-wave data: (a) the quality of research findings will be enhanced, (b) psychological theories may suggest appropriate functional forms for growth, (c) it is possible to test hypotheses about systematic interindividual differences in growth, (d) it is possible to associate features of growth with background characteristics, and (e) precision and reliability in growth measurement is a rapid monotonic increasing function of the number of waves. In short, longitudinal data not only enhance the statistical power of hypothesis tests but also enable researchers to test hypotheses they would be unable to address with cross-sectional or two-wave data.

Although many techniques have been developed to capitalize on these desirable features of longitudinal data, the focus of this book is on *latent growth curve modeling* (LGM). LGM represents a broad class of statistical methods that permit better hypothesis articulation, provide enhanced statistical power, and allow greater correspondence between the statistical model and the theory under investigation relative to competing methods. LGM permits straightforward examination of *intra*individual (within-person) change over time as well as *inter*individual (between-person) variability in intraindividual change. LGM is appealing not only because of its ability to model change but also because it allows investigation into the antecedents and consequents of change.

A selection of the kinds of questions the LGM framework can enable scientists to articulate and test include the following:

- What is the shape of the mean trend over time?
- Does the initial level predict rate of change?
- Do two or more groups differ in their trajectories?
- Does rate of change or degree of curvature in the mean trend predict key outcomes?
- What variables are systematically associated with change over time?
- Are theoretical hypotheses about the trajectory tenable given observed data?

- Does significant between-person variability exist in the shape of the trajectory?
- Is change over time in one variable related to change over time in another variable?

This list is by no means exhaustive. We now provide an overview of the rest of the book.

Overview of the Book

This book is intended for readers interested in studying change in phenomena over time. However, because LGM is an application of structural equation modeling (SEM), it is recommended that the reader have a basic working knowledge of SEM. It is important that researchers interested in LGM also be modestly familiar with topics such as multiple linear regression, the use of path diagrams to represent models, model identification issues, and the concept of fixed and free parameters. Useful introductory SEM texts include Kline (2004), Maruyama (1997), and Raykov and Marcoulides (2000). A more advanced treatment can be found in Bollen (1989).

We demonstrate the utility and flexibility of the LGM technique by beginning with a very basic model and building on this basic model with common extensions that permit deeper understanding of change processes and better articulated hypothesis tests. With each step, we apply the model to an existing data set to demonstrate how researchers can approach problems in practice. We demonstrate how to use LISREL (Jöreskog & Sörbom, 1996), Mx (Neale, Boker, Xie, & Maes, 2003), and Mplus (L. K. Muthén & Muthén, 1998–2006)—three popular SEM applications—to fit these models to data, although other user-friendly SEM packages could be used (e.g., EQS, AMOS). All results were identical (or nearly so) in LISREL and Mx. Syntax is provided on our Web site,[1] but most of the models we describe can be applied using virtually any SEM software package. In addition, we provide an extensive reference section so that interested readers will know where to go for more information. Thus, this book will bring researchers up to speed in LGM, but it also serves as a gateway to literature that explores these topics in greater depth.

We begin with a summary of research leading up to the development of LGM. We then describe the formal model specification, followed by sections on parameter estimation and model evaluation. Following this introductory material, we describe the data set we use throughout this book and the software used for analyses. The remainder of the book is devoted

to descriptions of specific models likely to be encountered or used in practice. Beginning with a basic (null) model, we explore more complex growth curve models by gradually relaxing constraints on parameters and adding additional variables to the model. We then explore some interesting and common extensions to the basic LGM, including related growth curves (multiple growth processes modeled simultaneously), cohort-sequential designs, the addition of time-varying covariates, and more complicated growth functions. Following these examples, we discuss relationships between LGM and other techniques, including growth mixture modeling, piecewise growth curves, modeling change in latent variables, and the interface between multilevel (random coefficients) modeling and LGM.

Latent Growth Curve Modeling: A Brief History and Overview

Historically, *growth curve models* (e.g., Potthoff & Roy, 1964) have been used to model longitudinal data in which repeated measurements are observed for some outcome variable at a number of occasions. The *latent* growth curve approach is rooted in the *exploratory factor analysis* (EFA) and *principal components analysis* (PCA) literature. Covariances among repeated measures can be modeled with EFA (Baker, 1954; Rao, 1958) or PCA (Tucker, 1958, 1966). Factors or components are then conceptualized as *aspects of change* or *chronometric* (as opposed to psychometric) *factors* (McArdle, 1989; McArdle & Epstein, 1987), and loadings may be interpreted as parameters representing the dependence of the repeated measures on these unobservable aspects of change. These aspects of change could include, for example, linear, quadratic, or S-shaped trends. These approaches have a number of problems for the study of change, however. One primary obstacle to using these approaches in practice is rotational indeterminacy—there is no clear rotation criterion that would select a loading pattern conforming to interpretable aspects of change (e.g., a set of polynomial curves). Although attempts have been made to develop rotation criteria that could be used to identify smooth functions (e.g., Arbuckle & Friendly, 1977; Tucker, 1966), none were completely satisfactory. An additional limitation of these methods was that they approached the problem of modeling change from the standpoint of estimating free loadings representing unknown functional trends (an exploratory approach) rather than testing the feasibility of a *particular* set of loadings (a confirmatory approach). The ability to test specific hypothesized trends is of great interest to substantive researchers.

Meredith and Tisak (1990) described *latent curve analysis* (LCA), an application of *confirmatory factor analysis* (CFA) that neatly sidesteps the rotational indeterminacy problem by allowing researchers to specify loadings reflecting specific hypothesized trends in repeated-measures data. This LCA approach is equivalent to what we call LGM. Because LGM is an application of CFA, which in turn is a special case of SEM, growth curve models can be imbedded in larger theoretical models. For readers interested in more details of the historical development of LGM, Bollen and Curran (2006) provide a thorough overview of its history.

Several advantages are associated with the use of LGM over competing methods, such as ANCOVA and multilevel modeling. LGM permits the investigation of interindividual differences in change over time and allows the researcher to investigate the antecedents and consequences of change. LGM provides group-level statistics such as mean growth rate and mean intercept, can test hypotheses about specific trajectories, and allows the incorporation of both time-varying and time-invariant covariates. LGM possesses all the advantages of SEM, including the ability to evaluate the adequacy of models using model fit indices and model selection criteria, the ability to account for measurement error by using latent repeated measures, and the ability to deal effectively with missing data. It is straightforward to compare growth across multiple groups or populations. LGM is a very flexible modeling strategy and can be easily adapted to new situations with unusual requirements.

Curran and Willoughby (2003) make an important point in stating that growth curve models "might be viewed as residing at an intersection between variable-centered and person-centered analysis" (p. 603). An exclusively variable-centered (nomothetic) perspective of change emphasizes mean trends over time, whereas an exclusively person-centered (idiographic) perspective focuses only on idiosyncratic trends characterizing individuals. Important insights can be gained from each perspective. Rather than focusing on one or the other, LGM capitalizes on both nomothetic aspects of change over time (mean trends) and idiographic aspects (individual departures from the mean trend).

Model Specification and Parameter Interpretation

A latent growth model can be represented as a special case of SEM. SEM is a general modeling framework for specifying and testing hypothesized patterns of relationships among sets of variables, some of which are measured (observed) while others are latent (unobserved). Latent variables often serve as proxies for psychological constructs that are impossible to

measure directly. A typical structural equation model contains a small number of latent variables linked by path coefficients, which are interpreted as regression weights. Latent variables, in turn, are represented by measured indicator variables. The relationship between latent variables and indicators corresponds to the factor analysis model. That is, factor loadings represent effects of latent variables on their indicators.

A special case of this general SEM system yields the basic latent growth curve model. We present this special case here, and in subsequent developments involving more complex LGM, we employ more of the full SEM framework. In LGM, the measured variables are repeated measures of the same variable *y.* The latent variables of primary importance are not psychological constructs; they instead represent patterns, or aspects, of change in *y.* In a basic LGM, often two factors are specified to represent aspects of change. These factors are defined by specifying factor loadings of repeated measures of *y* such that the factor loadings describe trends over time in *y.* The *intercept factor* represents the level of the outcome measure, *y*, at which the time variable equals zero, and the *slope factor* represents the linear rate at which the outcome measure changes. For example, a researcher interested in the rate of linear change in children's externalizing could collect repeated measurements of externalizing behavior, then treat these repeated measurements as indicators of intercept and slope factors (constraining loadings to reflect the expected pattern of change in that variable). As we will illustrate, the flexibility of the LGM framework permits the specification of more sophisticated models as well.

An LGM can be represented in matrix notation in terms of a *data model*, a *covariance structure*, and a *mean structure.* The data model represents the relationship between the factors and the repeated measures of *y*.[2] This model represents the $p \times 1$ vector of observations ($\mathbf{y}$) as a linear function of intercepts ($\boldsymbol{\tau}_y$, $p \times 1$), m latent variables representing aspects of change ($\boldsymbol{\eta}$, $m \times 1$), and disturbance terms ($\boldsymbol{\varepsilon}$, $p \times 1$), treating factor loadings ($\boldsymbol{\Lambda}_y$, $p \times m$) as regression coefficients[3]:

$$\mathbf{y} = \boldsymbol{\tau}_y + \boldsymbol{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\varepsilon}. \tag{1.1}$$

The $\boldsymbol{\tau}_y$ term is typically fixed to zero for model identification reasons. In expanded form (for $m = 2$), this model represents y_{ti}, the score at occasion t for individual i, as a function of two latent variables (η_{1i} and η_{2i}) and an error term (ε_{ti}):

$$y_{ti} = \lambda_{1t}\eta_{1i} + \lambda_{2t}\eta_{2i} + \varepsilon_{ti}. \tag{1.2}$$

The latent variables, in turn, may be expressed as functions of latent means (α_1 and α_2) and individual deviations away from those means:

$$\eta_{1i} = \alpha_1 + \zeta_{1i}, \tag{1.3}$$

$$\eta_{2i} = \alpha_2 + \zeta_{2i}. \tag{1.4}$$

The latent variables η_{1i} and η_{2i} are often referred to as *random coefficients.* The ζ residuals, representing individuals' deviations from the means of η_{1i} and η_{2i}, are sometimes referred to as *random effects.*

From the data model in Equation 1.1, one can derive a covariance structure and a mean structure. The covariance structure represents the population variances and covariances of the repeated measures of *y* as functions of model parameters, and the mean structure represents the population means of those repeated measures as another function of model parameters. The mean and covariance structures differ from the data model in that they do not contain scores for individuals on the factors (e.g., intercept and slope factors). These models are often used for parameter estimation and model evaluation. In the covariance structure, the variances and covariances of observed variables ($\mathbf{\Sigma}$, $p \times p$) are represented as functions of factor loadings ($\mathbf{\Lambda}_y$), factor variances and covariances ($\mathbf{\Psi}$, $m \times m$), and disturbance variances and covariances ($\mathbf{\Theta}_\varepsilon$, $p \times p$) (Bollen, 1989):

$$\mathbf{\Sigma} = \mathbf{\Lambda}_y \mathbf{\Psi} \mathbf{\Lambda}_y' + \mathbf{\Theta}_\varepsilon. \tag{1.5}$$

The mean structure, obtained by taking the expectation of the data model, represents population means of observed variables ($\boldsymbol{\mu}_y$, $p \times 1$) as functions of intercepts ($\boldsymbol{\tau}_y$, $p \times 1$) and latent variable means ($\boldsymbol{\alpha}$, $m \times 1$):

$$\boldsymbol{\mu}_y = \boldsymbol{\tau}_y + \mathbf{\Lambda}_y \boldsymbol{\alpha}. \tag{1.6}$$

In LGM, the elements of $\boldsymbol{\tau}_y$ are typically (but not always) constrained to zero, yielding a simplified data model and mean structure. Thus, the parameters of interest are contained in the matrices $\mathbf{\Lambda}_y$, $\mathbf{\Psi}$, and $\mathbf{\Theta}_\varepsilon$ and the vector $\boldsymbol{\alpha}$. Columns of $\mathbf{\Lambda}_y$ are known as *basis curves* or *latent growth vectors* (Singer & Willett, 2003).

In the model in Figure 1.1, Y1 through Y5 represent equally spaced repeated measures of variable Y. Here, change in Y is modeled as a function of two basis curves, and thus, the loading matrix $\mathbf{\Lambda}_y$ has two columns. Loadings on the intercept factor are fixed to 1.0 to represent the influence of a constant on the repeated measures. Loadings on the slope factor are

fixed to a linear progression to represent linearly increasing growth over time. Although it is traditional to begin the slope loadings at 0 to indicate that the first occasion of measurement indicates the initial response, this is by no means necessary, and indeed is often contraindicated. Additional factors are possible, each representing additional aspects of growth, often, but not necessarily, polynomial (see Model 10 in Chapter 2). In addition, the covariances among these aspects of change can be estimated by specifying covariance paths among factors. The ability to estimate these covariances can be important in situations where, for example, it is of interest to determine whether rate of growth in some variable is related to initial status. The triangle in Figure 1.1 represents the constant 1.0. Thus, the path coefficients linking the triangle to the basis factors are regressions onto a constant and, thus, represent means of the intercept and slope factors.

In the matrix representation of the covariance structure of Y1 through Y5, specifying the loadings of the repeated measures on intercept and slope factors equates to completely specifying the contents of $\mathbf{\Lambda}_y$, the factor loading matrix. The important parameters, including the necessary constraints for a simple linear growth model with homoscedastic and uncorrelated disturbance variances, are

$$\mathbf{\Lambda}_y = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}, \tag{1.7}$$

$$\mathbf{\Psi} = \begin{bmatrix} \psi_{11} & \\ \psi_{21} & \psi_{22} \end{bmatrix}, \tag{1.8}$$

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}, \tag{1.9}$$

$$\mathbf{\Theta}_\varepsilon = \begin{bmatrix} \theta_\varepsilon & & & & \\ 0 & \theta_\varepsilon & & & \\ 0 & 0 & \theta_\varepsilon & & \\ 0 & 0 & 0 & \theta_\varepsilon & \\ 0 & 0 & 0 & 0 & \theta_\varepsilon \end{bmatrix}. \tag{1.10}$$

Interpretation of the parameters in these matrices is straightforward. Elements of $\mathbf{\Lambda}_y$ are fixed by the researcher to represent hypothesized

trajectories, where each column of loadings represents a hypothesized aspect of change. In Equation 1.7, all elements in the first column are constrained to 1.0 to reflect the fact that each individual's intercept

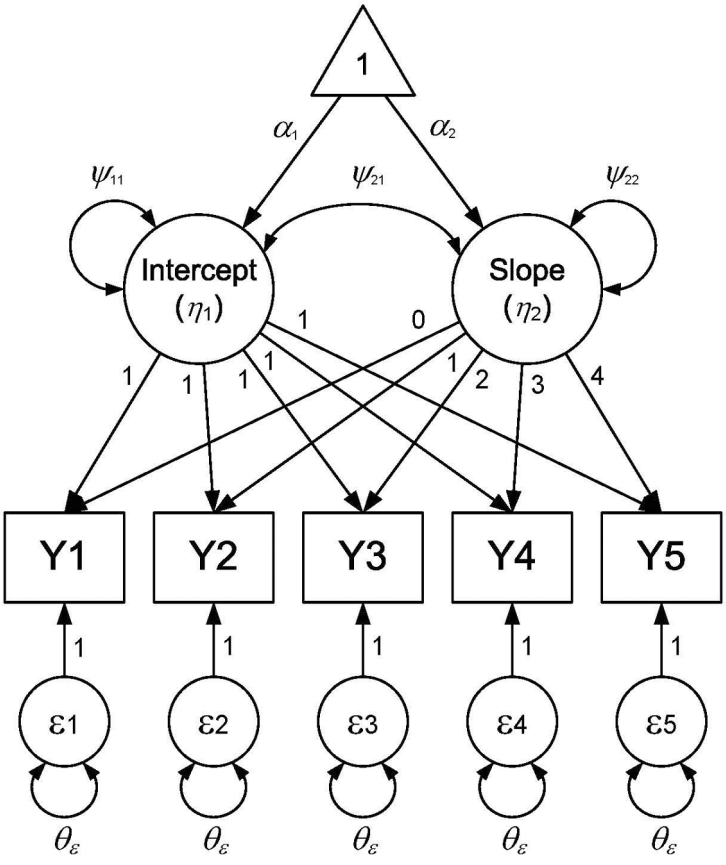

Figure 1.1 A Complete Path Diagram for a Typical Latent Growth Curve Model, Including Random Intercept, Random Linear Slope, and Intercept–Slope Covariance Parameters.

NOTE: By SEM convention, circles represent latent variables, squares represent measured variables (here, Y1 through Y5 are equally spaced repeated measures), triangles represent constants, double-headed arrows represent variances or covariances, and single-headed arrows represent regression weights. Numerical values correspond to fixed parameters, whereas symbols represent free parameters (those to be estimated).

remains constant over the repeated measures. The linear progression in the second column of $\mathbf{\Lambda}_y$ reflects the hypothesis of linear growth with equal time intervals. Elements of $\mathbf{\Psi}$ represent the variances and covariances of these aspects of change. In the case of simple linear growth represented in Figure 1.1, the $\mathbf{\Psi}$ matrix includes the intercept variance (ψ_{11}), the slope variance (ψ_{22}), and the covariance of intercepts and slopes (ψ_{21}). The elements of $\boldsymbol{\alpha}$ are regression coefficients predicting aspects of change from a constant (1.0), and can be interpreted as the mean intercept and slope. In particular, α_2 in Equation 1.9 represents the expected change in the outcome variable associated with a change of one unit in the time metric. Finally, elements of $\mathbf{\Theta}_\varepsilon$ are variances and covariances of disturbance terms, representing the portion of the variance in data not associated with the hypothesized latent curves.[4] If homoscedastic disturbance variance is assumed (as they are in Equation 1.10), the single disturbance variance may be represented by placing an equality constraint on the diagonal elements of $\mathbf{\Theta}_\varepsilon$. The off-diagonal terms in $\mathbf{\Theta}_\varepsilon$ are usually fixed to zero to represent the hypothesis that disturbances are uncorrelated over time, although this assumption is not required. This basic LGM may be extended to incorporate predictors of intercept and/or slope, parallel growth curves for multiple outcomes, and parameter constraints (e.g., the covariance of the intercept and slope may be constrained to zero). These extensions and others are discussed later.

To summarize, there are six parameters estimated in a typical linear application of LGM, regardless of the number of repeated measurements. These include the mean intercept and slope (α_1 and α_2), the intercept and slope variances and covariance (ψ_{11}, ψ_{22}, and ψ_{21}), and a disturbance variance (θ_ε) that remains constant over repeated measurements. The remaining parameters are typically constrained to zero or to values consistent with a particular pattern of change. As you will see, some of these constraints can be changed or relaxed in various ways, depending on the characteristics of particular applications. Additional free parameters are added to the model in more complex models, such as those involving nonlinear growth or predictors of change over time.

The Scaling of Time

In Figure 1.1 and Equation 1.7, time was coded in a way that expressed linear change, placing the origin of the time scale at the first occasion of measurement ($\lambda_{1,2} = 0$). The loading pattern illustrated in Equation 1.7, in which the intercept is interpreted as "initial status," is typical, but other

patterns may be useful as well, depending on the particular research question under consideration. Consider a case in which the last occasion of measurement (e.g., graduation from a drug rehabilitation program) is the occasion of most interest. In such situations, it is more sensible to code the final occasion of measurement as 0, as in the following loading matrix:

$$\boldsymbol{\Lambda}_y = \begin{bmatrix} 1 & -4 \\ 1 & -3 \\ 1 & -2 \\ 1 & -1 \\ 1 & 0 \end{bmatrix}. \tag{1.11}$$

In Equation 1.11, the slope loadings increase linearly, but the intercept is defined to lie at the fifth, and last, occasion. In addition, if it is more sensible to code time in terms of months rather than years, Equation 1.11 could be reparameterized as follows:

$$\boldsymbol{\Lambda}_y = \begin{bmatrix} 1 & 0 \\ 1 & 12 \\ 1 & 24 \\ 1 & 36 \\ 1 & 48 \end{bmatrix}. \tag{1.12}$$

The interpretation of some model parameters related to the slope factor (e.g., the mean slope or regression weights associated with predictors of slope) will be different for loading matrices in Equations 1.11 and 1.12, but the fit of the overall model will not. In fact, any linear transformation of the loading matrix will not alter model fit, although transformation of $\boldsymbol{\Lambda}_y$ does have important consequences for parameter interpretation (Biesanz, Deeb-Sossa, Papadakis, Bollen, & Curran, 2004; Hancock & Lawrence, 2006; Mehta & West, 2000; Rogosa & Willett, 1985; Stoel, 2003; Stoel & van den Wittenboer, 2003; Stoolmiller, 1995; Willett, Singer, & Martin, 1998). It is rarely sensible, for example, to center the time variable at its mean because the central occasion of measurement is not usually of most interest.[5] Rather, time should be coded in a way that facilitates answering substantive questions. The zero point of the time scale should be placed at a meaningful occasion of measurement, in part because this choice determines the point in time at which to interpret the mean intercept, intercept variance, and intercept–slope covariance (Rogosa & Willett, 1985). Stoel (2003) and Stoel and van den Wittenboer (2003) suggest that the intercept is meaningful only

for growth processes with a natural origin; if the time origin is chosen arbitrarily, interpretation of any intercept-related parameters should be avoided. Similarly, the metric of time should be chosen so as to maximize interpretability (Biesanz et al., 2004). In some circumstances, this may involve using two metrics of time simultaneously in the same model (e.g., age and wave; see McArdle & Anderson, 1990).

A common question concerns the optimal number of repeated measures to use. To a large extent this choice will be dictated by practical concerns such as time and cost. The minimum number of repeated measures necessary to achieve a model with at least one degree of freedom (*df*) for *m* polynomial growth factors, regardless of whether or not the disturbance variances are constrained to equality over time, is $m + 1$. This formula will always hold, assuming that the estimated parameters include only factor means, factor (co)variances, and disturbance variances. A model with at least 1 *df* (i.e., an overidentified model) is necessary because, if there are at least as many free parameters as there are sample means and (co)variances, the model will not be identified and will thus be untestable (see Bollen, 1989; Bollen & Curran, 2006, for more details on model identification).

But how many repeated measures should be used, given $m + 1$ as an absolute minimum? Stoolmiller (1995) suggests that four to five measurement occasions are probably sufficient for modeling linear growth. MacCallum, Kim, Malarkey, and Kiecolt-Glaser (1997) note that there is no reliable rule of thumb, but suggest that linear models demand at least four or five repeated measures, whereas more complex models may demand "substantially more than that" (p. 217). To obtain adequate power for testing covariances among slope factors in parallel process latent growth curve models (Model 7 in Chapter 2), Hertzog, Lindenberger, Ghisletta, and von Oertzen (2006) recommend that at least six repeated measures be collected, but this figure may shift up or down with changes in effect size, sample size, or growth curve reliability (the ratio of total variance explained by aspects of growth). As in almost every other aspect of data analysis, more is better—more data yield more information, and that is never a bad thing. However, it is our experience that parsimonious linear models often have trouble adequately fitting more than six repeated measures. This should come as no surprise, as few natural processes are likely to follow precisely linear trajectories. *No model is correct*, so researchers should be prepared to see simple models fit poorly as information accumulates. LGM is best suited for modeling trends measured over a limited number of occasions in large samples. Regardless of the number of measurements, the *range* of measurement should be sufficient to span the entire time frame of theoretical interest.

Finally, we note that occasions of measurement need not be equally spaced. For example, if longitudinal data were collected in 1977, 1979, 1983, and 1984, the intervals between occasions of measurement become 2, 4, and 1:

$$\boldsymbol{\Lambda}_y = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 6 \\ 1 & 7 \end{bmatrix}. \tag{1.13}$$

The loading matrix in Equation 1.13 still represents linear change; the intervals between slope loadings are themselves linear rescalings of the intervals between occasions of measurement. If one takes 1977 as a baseline, subsequent measurement occasions occurred 2, 6, and 7 years after it. In nonlinear models (such as the polynomial latent curve models described in Chapter 2 or the structured latent curve models discussed in Chapter 3), it is often a good idea to not only collect more than five repeated measures but also to space measurements more closely together during periods when change is occurring the most rapidly. This helps avoid estimation problems and to more accurately estimate parameters characterizing change.

Our point is that there is no "one-size-fits-all" design and corresponding growth curve model. The number of measurement waves, times between waves, units of time, and placement of the time origin may vary greatly both within and across studies. More in-depth discussion of the scaling of time may be found in Biesanz et al. (2004), Curran and Willoughby (2003), Hancock and Choi (2006), Schaie (1986), and Stoolmiller (1995).

Asynchronous Measurement

Loadings in $\boldsymbol{\Lambda}_y$ are fundamentally unlike other fixed parameters with which users of SEM may be familiar. Regardless of how time is coded, the contents of $\boldsymbol{\Lambda}_y$ represent *functions of time.* For example, the slope loadings contain values of what might be considered the predictor *time* in other modeling contexts. Many applications of LGM make the oversimplifying assumption that data are collected at the same occasions across all individuals (Mehta & West, 2000). Such data are referred to as *time-structured data* (Bock, 1979). The fact that all subjects share the same occasions of measurement permits these values to be placed in a common $\boldsymbol{\Lambda}_y$ matrix. But this is unrealistic in most applications. In situations where individuals are not measured at the same occasions, or are measured at the

same occasions but at different ages, the basic LGM described earlier will not be sufficient. As we now discuss, there are two main strategies for estimating growth curve models in situations when individuals are measured at different occasions. Both strategies allow for individual differences in factor loadings.

When all individuals are not measured at the same occasions, but there is a limited set of measurement schedules, a multiple-group strategy can be employed in which multiple models, each characterized by a distinct $\mathbf{\Lambda}_y$ matrix, are estimated simultaneously. We discuss this strategy in more detail in the next section, but basically it involves placing individuals with the same measurement schedule into groups and fitting the model simultaneously to all such groups (Allison, 1987; T. E. Duncan, Duncan, Strycker, Li, & Alpert, 1999; McArdle & Bell, 2000; McArdle & Hamagami, 1991; B. Muthén, Kaplan, & Hollis, 1987). Thus, all individuals measured only at times 1, 3, 5, and 6 may belong to Group 1, whereas all individuals measured only at times 1, 2, 4, and 5 may belong to Group 2.

But consider the case in which there are too many distinct measurement schedules to be accommodated by the multiple-groups solution. A more general solution exists. An attractive feature of the Mx and Mplus programs is that they can accommodate individual slope loadings via implementation of *definition variables*, or *individual data vectors*—special parameter vectors containing fixed values for each individual (Hamagami, 1997; Neale et al., 2003). Whereas the traditional approach involves applying a model in which all individuals are assumed to share the same basis curves and thus the same $\mathbf{\Lambda}_y$, the use of definition variables involves creating a set of slope factor loadings unique to each individual. In the special case of longitudinal research involving age, this is referred to as *scaling age across individuals* (Mehta & West, 2000). A sample Mx script demonstrating this technique is included in Appendix A of Mehta and West (2000) and at our Web site (http://www.quantpsy.org/).

Assumptions

LGM with maximum likelihood (ML) estimation invokes certain important assumptions. Most assumptions involve the distributions of latent variables (in LGM, these are, for instance, intercept, slopes, and disturbances). Because these variables are by definition unobservable, it is customary to assume their characteristics. We assume that the means of residuals and disturbance terms in Equations 1.2, 1.3, and 1.4 are zero. At each occasion, this assumption applies to means computed across the population of individuals and across theoretical repeated observations of the same person. In other

words, if it were possible to measure the same individual repeatedly at a given occasion, we assume that the mean of the disturbances across those measurements is zero. In a similar sense, the covariances among all residual terms are assumed to be zero within and between occasions, and all covariances between residuals in Equation 1.2 and random intercepts and slopes are assumed to be zero. To use ML estimation, it is necessary to make the additional assumption that observed variables are derived from population distributions with roughly the same multivariate kurtosis as a multivariate normal distribution. The assumptions underlying LGM are treated extensively by Bollen and Curran (2006).

Byrne and Crombie (2003) discuss three additional assumptions. They require the assumptions that the trajectory be linear, that the disturbances be uncorrelated across occasions, and that the disturbance variances remain equal across occasions. In fact, these are not assumptions of LGM or of ML. Byrne and Crombie's assumption of linearity refers to the linearity of the growth trajectory, not to the loadings. Technically, this is not an assumption; rather, it is the central hypothesis under scrutiny. Two ways to test the hypothesis of linearity are to compare the fit of a linear growth curve model to a baseline of absolute fit or relative to an unspecified trajectory model (see Model 11 in Chapter 2). The other two assumptions discussed by Byrne and Crombie (independence and homoscedasticity of disturbances) may be common aspects of model specification in LGM but are not required; in fact, the ability to estimate different occasion-specific disturbance variances is considered a strength of the LGM approach and is required for approaches that combine LGM with autoregressive strategies (Curran & Bollen, 2001; McArdle, 2001).

Parameter Estimation and Missing Data

Parameter estimation in SEM traditionally is accomplished with ML estimation, the use of which assumes that measured variables are multivariate normally distributed. For models such as latent growth curve models that are designed to explain covariances as well as means of measured variables, the data are typically in the form of a sample covariance matrix, $\mathbf{S}$, and sample mean vector, $\bar{\mathbf{y}}$, computed from complete data (the data are sometimes forced to be complete through listwise deletion or one of several data imputation methods). Matrix $\mathbf{S}$ is of order $p \times p$ and contains the sample variances and covariances of the p repeated measures of y. Vector $\bar{\mathbf{y}}$ contains the sample means of those p repeated measures. According to the covariance and mean structure models in Equations 1.5 and 1.6, the population covariance matrix, $\boldsymbol{\Sigma}$, and mean vector, $\boldsymbol{\mu}$, are functions of model parameters. If we

let all the parameters in Equations 1.5 and 1.6 be organized into a single vector, θ, then the objective in parameter estimation is to find parameter estimates in $\hat{\theta}$ such that the resulting implied $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$ are as similar as possible to $\mathbf{S}$ and $\overline{\mathbf{y}}$, respectively. In ML estimation, this optimality is defined using the multivariate normal likelihood function. That is, ML estimation results in a set of parameter estimates $\hat{\theta}$ that maximize the log of the likelihood function:

$$\ln L = -\frac{1}{2}\sum_{i=1}^{N}\left\{p\ln 2\pi + \ln|\boldsymbol{\Sigma}| + (\mathbf{y}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\right\}. \quad (1.14)$$

Extending developments by Jöreskog (1967), it can be shown that ln$L$ is maximized when the following *discrepancy function* is minimized:

$$F_{ML} = \ln|\boldsymbol{\Sigma}| - \ln|\mathbf{S}| + \mathrm{tr}\left[(\mathbf{S} - \boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right] + (\overline{\mathbf{y}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{y}} - \boldsymbol{\mu}). \quad (1.15)$$

Thus, given $\mathbf{S}$ and $\overline{\mathbf{y}}$, ML estimation seeks a vector of parameter estimates, $\hat{\theta}$, that produce implied $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$ matrices (from Equations 1.5 and 1.6) that minimize F_{ML}. Note that if $\boldsymbol{\Sigma} = \mathbf{S}$ and $\boldsymbol{\mu} = \overline{\mathbf{y}}$—that is, if the model perfectly reproduces the data—then $F_{\mathrm{ML}} = 0$.

The minimization of the F_{ML} discrepancy function assumes that complete sample data are used to obtain $\mathbf{S}$ and $\overline{\mathbf{y}}$. If some data are missing—for example, if individuals are measured at different occasions or if data simply are not obtained for some individuals at some occasions—this strategy will not work because covariance matrices computed using the available data may not be internally consistent. Fortunately, options are available for dealing with missing data. The multiple-group strategy mentioned earlier in the context of multiple measurement schedules can be considered a general model-based approach to addressing missing data (T. E. Duncan & Duncan, 1995; Marini, Olsen, & Rubin, 1979; McArdle & Hamagami, 1992; B. Muthén et al., 1987). For example, if some individuals are measured at occasions 1, 2, 3, and 5 and others are measured at occasions 1, 3, 4, and 5, a two-group model may be specified in which all members within each group share the same occasions of measurement. For multisample analyses, the discrepancy function in Equation 1.15 is generalized to a multisample expression and is then minimized so that optimal fit is obtained to all groups simultaneously.

The multiple-groups approach to dealing with missing data becomes impractical when there are more than a few distinct measurement schedules. Until recently, this limitation presented a real problem, given the prevalence

of missing data in real longitudinal designs. Advances in estimation that allow models to be fit directly to raw data have made it possible to include incomplete cases in the analysis. In situations involving missing (partially complete) data, the *full information maximum likelihood* (FIML) method is often recommended to obtain ML parameter estimates. To allow for incomplete data, the log likelihood in Equation 1.14, which implies the availability of complete data, can be modified as follows:

$$\ln L = -\frac{1}{2}\sum_{i=1}^{N}\left\{p_i \ln 2\pi + \ln|\boldsymbol{\Sigma}_i| + (\mathbf{y}_i - \boldsymbol{\mu}_i)'\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i)'\right\}, \qquad (1.16)$$

where $\mathbf{y}_i$ is the measured portion of the data vector for individual i and $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are the modeled mean vector and covariance matrix, respectively, with rows and columns corresponding to the data present for individual i (Arbuckle, 1996; Wothke, 2000). FIML estimation involves maximizing this function. FIML is more efficient and less biased than methods involving data imputation or deletion of partial data and yields unbiased estimates when data are missing completely at random (MCAR) or *missing at random* (MAR; Neale, 2000; Rubin, 1976).[6] When compared with pairwise and listwise deletion and imputation methods, FIML has been shown to have a lower incidence of convergence failure, higher efficiency, lower bias, and more accurate model rejection rates (Enders & Bandalos, 2001). The characteristics of the two methods are similar. In fact, they are equivalent when no data are missing. For a clear, more in-depth description of the FIML algorithm with comparisons to other methods, see Enders (2001).

FIML is only one of several discrepancy functions that can be minimized to yield parameter estimates in the presence of missing data. Other approaches include the application of generalized least squares, unweighted least squares, the E-M algorithm, and asymptotically distribution free methods to data that have been rendered "complete" through pairwise or listwise deletion or through single or multiple imputation. FIML is often preferable to these methods because (a) it uses all available information to estimate parameters, (b) it does not require extremely large samples, and (c) standard errors may be obtained by inverting the asymptotic covariance matrix of parameter estimates. Pairwise and listwise deletion omit some data from consideration, and pairwise deletion risks the possibility of encountering a covariance matrix that is not positive definite. Furthermore, FIML is now a standard estimation option in most SEM software, including AMOS (Arbuckle & Wothke, 1999), Mplus (L. K. Muthén & Muthén, 1998–2006), Mx (Neale et al., 2003), EQS (Bentler, 1995), and LISREL

(Jöreskog & Sörbom, 1996). In LISREL, for example, FIML is automatically invoked if raw data are used as input. Missing data issues are discussed in more depth by Allison (1987, 2002). The primary drawback to using FIML estimation is that, if some data are missing, the full array of ML fit indices is not available.[7]

Model Evaluation and Selection

Specifying and testing models as representations of theoretical predictions is fundamental to the practice of modern empirical science. In developing a model to be fit to observed data, it is critical that the specified model accurately reflect the predictions or implications of a substantive theory of growth (Collins, 2006; Curran, 2000; Curran & Hussong, 2003; Curran & Willoughby, 2003). In addition, although the focus here is on model evaluation, it is usually preferable not to specify and evaluate models in isolation, but rather to compare competing, theoretically derived models.

Given a theoretically plausible model, hypotheses in LGM can be tested by assessing the statistical and practical significance of model parameters, including the means of the intercept and slope factors and the variances and covariances among aspects of change. An informal test of the significance of a parameter is conducted by dividing the point estimate by its standard error; if the ratio exceeds about 2.00 (1.96 in very large samples), the parameter estimate is said to be significantly different from zero at the .05 level. The determination of practical significance depends heavily on the context.

In SEM, the fit of an entire model also can be assessed. Indeed, good fit by global criteria is usually a prerequisite for interpreting parameter estimates.[8] Under multivariate normality and under the null hypothesis of perfect fit, $\hat{F}_{ML} \times (N - 1)$ is distributed as χ^2 with degrees of freedom $df = [p(p + 3)/2] - q^*$, where p is the number of variables and q^* is the effective number of free model parameters.[9] This χ^2 statistic forms the basis for an array of fit indices that can be used to gauge the match between a model's predictions and observed data. We recommend the *root mean square error of approximation* (RMSEA; Browne & Cudeck, 1993; Steiger & Lind, 1980):

$$\text{RMSEA} = \sqrt{\frac{\max\left\{\left(\hat{F}_{ML} - \frac{df}{N-1}\right), 0\right\}}{df}}, \tag{1.17}$$

available in several SEM programs. The numerator under the radical in Equation 1.17 is an estimate of model misfit (discrepancy) in the population. Thus, the quantity under the radical represents estimated population

model error per degree of freedom and thus smaller values are better. RMSEA is preferred because it is an estimate of misfit in the population rather than simply a measure of misfit in the sample. Importantly, one can obtain confidence intervals for RMSEA, providing a measure of precision of this fit index in addition to a point value.

Because χ^2-based indices are known to suffer from problems associated with near-singular matrices (Browne, MacCallum, Kim, Andersen, & Glaser, 2002), violations of distributional assumptions (Curran, West, & Finch, 1996), and large N (Tucker & Lewis, 1973), it is also advisable to examine simple residuals between elements of $\mathbf{S}$ and $\hat{\mathbf{\Sigma}}$. A common index based only on residuals is the *standardized root mean square residual* (SRMR; Jöreskog & Sörbom, 1996), a summary measure of the magnitude of the residuals. SRMR is the square root of the average squared absolute difference between observed correlations and model-implied correlations (and thus smaller values are better). Like RMSEA, SRMR is included as default output in many SEM programs, although it should be kept in mind that SRMR assesses fit of the covariance structure only and is not sensitive to misfit in the mean structure.

Another family of fit indices reflects the *incremental fit* of the specified model over the fit of an appropriately specified null model (see Model 0 in Chapter 2). One example of this sort of fit index is the *nonnormed fit index* (NNFI; Bentler & Bonett, 1980; Tucker & Lewis, 1973):

$$\text{NNFI} = \min\left\{\left(\frac{\frac{\chi_0^2}{df_0} - \frac{\chi_k^2}{df_k}}{\frac{\chi_0^2}{df_0} - 1}\right), 1\right\}, \tag{1.18}$$

where χ_0^2 and df_0 are computed with respect to the null model, and χ_k^2 and df_k are computed with respect to the model of interest. NNFI has been demonstrated to be relatively robust to violations of distributional assumptions (Lei & Lomax, 2005). We elaborate on the appropriate null model in the next chapter. In this book, we report χ^2, RMSEA, NNFI, and SRMR for all fitted models.

In addition to evaluation of models in isolation, a model selection approach can be used to evaluate the relative fit of nested or nonnested models. One model is said to be nested in another if the estimated parameters of one (Model A) are a subset of those in the latter (Model B). In other words, if some of the parameters in B are constrained to yield A, A is nested within B. When some data are partially missing, some fit indices can no longer be computed (Enders, 2001), for example, GFI and SRMR.

Under the null hypothesis of no difference between models, the difference in χ^2 statistics ($\Delta\chi^2$) for complete data, or between $-2\ln L$ values for incomplete data, is itself distributed as a χ^2 statistic, with *df* equal to the difference in the number of parameters estimated. For models that are not nested, information-based model selection criteria (e.g., Akaike information criterion, Bayesian information criterion) may be used to select models.

Statistical Power

As with most applications of inferential statistics, statistical power is important in the LGM context. Power refers to the probability of correctly rejecting a false null hypothesis. In the LGM context, the null hypothesis is the researcher's latent growth curve model, so power is the probability that one's model of growth will be rejected if it is not correct in the population. LGM models are generally never exactly correct in the population, so a high level of statistical power will tend to ensure rejection of models that are very good, but not perfect. This rejection of good models is, of course, not desirable in practice, but is a well-known limitation of the likelihood ratio test of model fit. In practice, this situation is remedied through the use of various descriptive measures of fit such as those described earlier rather than focusing exclusively or heavily on the likelihood ratio test of model fit.

MacCallum, Browne, and Sugawara (1996) describe a method of computing power (given sample size) or minimum required sample size (given a desired level of power) that involves the RMSEA fit index. The researcher chooses null and alternative hypotheses corresponding to values of RMSEA that reflect, respectively, good fit (ε_0) and poor fit (ε_A). For example, the *test of exact fit* might involve selecting $\varepsilon_0 = 0.00$ (exact fit) and $\varepsilon_A = 0.08$ (mediocre fit). A test of close fit might involve selecting $\varepsilon_0 = 0.05$ (close fit) and $\varepsilon_A = 0.10$ (unacceptable fit). SAS code provided by the authors will supply the minimum *N* necessary for rejecting a poor model at a given level of power, model *df*, and a pair of null and alternative hypotheses defining good and poor fits in terms of RMSEA. Conversely, their code will supply the level of statistical power associated with a pair of null and alternative hypotheses given a particular *N* and model *df.*

In addition to the power to reject a poor model, it is also sensible to consider the power to detect nonzero parameters. This kind of power represents a largely unstudied topic in the LGM context. An exception is a recent study by Hertzog et al. (2006), who found that the power to detect slope covariances in parallel process latent growth curve models (see Model 7 in Chapter 2) depends heavily on effect size, the number of repeated measures, growth curve reliability, and sample size.

Notes

1. http://www.quantpsy.org/

2. For those readers familiar with the LISREL framework for model specification in SEM, the mathematical representation here uses the "all-*y*" model, where all latent variables are considered as endogenous.

3. Vectors are denoted by underscores, and matrices are denoted by boldface.

4. It is important to remember that elements of $\boldsymbol{\Theta}_{\varepsilon}$ do not represent error variance in the usual sense. Rather, the *t*th variance in $\boldsymbol{\Theta}_{\varepsilon}$ reflects the degree to which a linear model does not adequately capture individuals' scores at occasion *t*. Much of this variability may be due to error, but it also may be amenable to prediction by other variables.

5. Stoolmiller (1995), on the other hand, recommends centering the time variable in polynomial models to avoid estimation problems due to linear dependence among the polynomial terms.

6. Missing data (for *x*) are MCAR when "missingness" depends neither on observed nor unobserved responses of *x* or any other variable. Missing *x* data are MAR when missingness depends neither on observed nor unobserved responses of *x* after controlling for other variables in the data set (Allison, 2002).

7. Mplus is the only SEM software, of which we are aware, that will compute fit indices when some data are missing.

8. Perhaps owing to their highly constrained nature, it is common for growth curve models to fit poorly by global fit criteria. However, models may fit poorly in situations even when individual growth curves are approximated well. Coffman and Millsap (2006) recommend that global fit indices be supplemented with individual fit criteria.

9. This formulation of *df* assumes that sample data comprise both means and covariances. Growth curve models are almost always fit to means and covariances. If the mean structure is not modeled, $df = [p(p + 1)/2] - q^*$.