CHAPTER 2

*Modeling Longitudinal Data from Research on Adolescence**

TODD D. LITTLE, NOEL A. CARD, KRISTOPHER J. PREACHER, AND ELIZABETH MCCONNELL

The study of adolescent development generally relies on two types of study, cross-sectional and longitudinal. Cross-sectional studies are quite common and involve comparing two or more cohorts of youth who are assessed at a single, concurrent measurement occasion. Longitudinal studies, in contrast, involve collecting data at two or more occasions, with the interval between occasions allowing for some sort of meaningful change to occur. As we describe next, there are several advantages associated with longitudinal designs and the analysis of longitudinal data.

ADVANTAGES OF LONGITUDINAL DATA

Longitudinal data offer several advantages over cross-sectional data in the study of adolescent development. The first advantage is that longitudinal data allow us to draw more valid conclusions regarding developmental changes in levels (i.e., means) and processes (i.e., associations) of phenomena than can be drawn with cross-sectional data. Researchers may try to infer such changes from concurrent data collected from participants over a developmental range (e.g., across age from early, middle, and late adolescence). However, in cross-sectional designs, age effects are confounded with cohort effects, leading to ambiguity in interpretation (Baltes, 1968; Schaie, 1965). For instance, if concurrent data reveal that a phenomenon is more prevalent or frequent among older than middle adolescents and among middle than early adolescents, these differences *may* be due to developmental differences. However, they also may be due to cohort differences, such that those born earlier (i.e., the late adolescents) experience different sociohistorical inputs and exhibit higher levels of the phenomenon than those born later (i.e., the early adolescents). With longitudinal data, we can hold constant the cohort effect so that changes can be attributed to developmental differences (see Baltes, 1968; Schaie, 1965). Some longitudinal designs (i.e., accelerated longitudinal designs, described later) can be used to identify and separate both developmental and cohort effects.

A second advantage of a longitudinal study is that it allows for inferences regarding various estimates of the cross-time relations among

*Correspondence regarding this chapter should be addressed to Todd D. Little (yhat@ku.edu), Department of Psychology, University of Kansas, 1415 Jayhawk Blvd., Lawrence, KS 66045-7555; Noel A. Card (ncard@email.arizona.edu), Division of Family Studies and Human Development, University of Arizona, Tucson, AZ 85721-0033; or Kristopher J. Preacher (preacher@ku.edu), Department of Psychology, University of Kansas, 1415 Jayhawk Blvd., Lawrence, KS 66045-7555.

This work was supported in part by grants from the NIH to the University of Kansas Intellectual and Developmental Disabilities Research Center (5 P30 HD002528). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

a set of variables. These cross-time relations have many forms, including stability, stationarity, and equilibrium of the individual differences associations, as well as mean changes and intraindividual differences. Generally speaking, *stability* refers to the strength of the relation between the relative standing of a person on the same construct measured at two or more measurement occasions. These associations are often referred to as autoregressive paths, or interindividual stability. *Stationarity* refers to whether the autoregressive paths are equal in magnitude over multiple time intervals (e.g., two or more intervals among three or more measurement occasions, assuming equal length intervals; Kenny, 1979). For example, if the stability coefficient is of the same magnitude between Times 1 and 2 and Times 2 and 3, stationarity holds. A third cross-time relation of interest is the stability in the pattern of associations among two or more constructs across two or more measurement occasions. Termed *equilibrium*, this cross-time relation refers to the equality or homogeneity of the within-time covariances (Cole & Maxwell, 2003; Dwyer, 1983; Kessler & Greenberg, 1981).

Each of these three terms (stability, stationarity, and equilibrium) refers to across-time associations among constructs; however, they do not address the across-time similarity or change in mean levels of a construct (i.e., is the typical developmental trend one of increase or decrease in a construct across adolescence?). To address this issue, we need to consider within-person, or intraindividual, stability or change across time. Fully understanding these different conceptualizations of stability can be confusing, so we discuss these further in the context of panel models and growth curve models below. Regardless of the form considered, knowing the degree of stability is important because it helps us understand whether certain phenomena are transient (less stable) or persistent (more stable) across adolescent development. The key point here is that addressing any conceptualization of stability versus change requires longitudinal data.

The third advantage of a longitudinal study is that it allows us to make qualified inferences regarding the cause–effect relations among constructs. Although causality can be inferred only within properly conducted experimental designs, it is less often recognized that such designs are necessarily longitudinal in that some time must elapse between the experimental manipulation and measured outcome. As Gollob and Reichardt (1987) stated simply: “Causes take time to exert their effects” (p. 81). For many aspects of adolescent development, experimental manipulation is difficult or not ethically possible, and in these situations longitudinal naturalistic research provides our best basis for inference regarding directions of influence. These issues will be further discussed later in this chapter (in the section on panel designs), but the point here is that directions of causal influence from nonexperimental data can be evaluated legitimately only within longitudinal designs.

A fourth advantage of a longitudinal study is the ability to model the processes through which effects are expressed over time. With multiple observation occasions, both direct and indirect pathways of influence can be modeled and, in multivariate growth curve models, dynamic associations can be modeled. Often, the patterns of indirect influence over time provide the most meaningful information for theoretical consideration. For example, Cole and Maxwell (2003) describe how indirect processes of associations over time, when equilibrium of within time associations is achieved, allows one to make strong inferences of mediation (we discuss this model in more detail in the section on mediation and moderation below). Similarly, Greenwood and Little (2008) describe how the regression discontinuity design for intervention evaluations, when coupled with longitudinal designs, allows inferences of intervention effectiveness that are as valid and strong as those obtained with the gold standard of the randomized clinical trial (which also incorporates longitudinal data in its inferential arsenal; see Shadish, Cook, & Campbell, 2002).

Given the value of longitudinal data, it is not surprising that they are frequently used in developmental research. Card and Little (2007a) examined the prevalence of longitudinal designs in six premier developmental journals during a one-year publication period, finding that only 41% of studies published in these journals analyzed longitudinal data. Many developmental journals give preference to longitudinal studies, and at least one of the leading journals of adolescent development, *Journal of Research on Adolescence*, explicitly requests that submitted manuscripts include analyses of longitudinal data. Although such policies may dismiss the value of many cross-sectional studies, and we certainly do not want to deemphasize the value of a well-conceived and executed concurrent study, the advantages of longitudinal analyses highlight the merits of collecting and analyzing longitudinal data. However, the advantages and benefits of longitudinal data can be realized only if the theoretical rationale, measurement operations, and statistical model are coordinated.

Rationale for Longitudinal Data

Notwithstanding the many advantages of longitudinal data, we emphasize that researchers should not collect longitudinal data just to get longitudinal data. Too often, longitudinal data are collected with little thought devoted to the rationale, and lack of forethought can lead to poor design decisions and wastes precious resources on collecting data that are unable to deliver the sought-for answers. In studying adolescence, longitudinal data are very useful for understanding mechanisms of change and processes of influence as well as the interplay of the adolescent and the context (see Card, Little, & Bovaird, 2007; Little, Bovaird, & Card, 2007). However, studies that collect large batteries of questionnaire protocols annually or semiannually often lack correspondence with either a well-considered theoretical model or the needs of the underlying statistical model. In fact, we will consistently echo Collins's (2006) call for integrating the theoretical model, the temporal design, and the statistical model (see also Ram & Grimm, 2007).

The kinds of answers that state-of-the-science methods can glean from longitudinal data are only as good as the theoretical rationale and design of the study itself. A number of considerations must go into planning for longitudinal data. In the next section, we highlight a number of considerations that are particularly relevant to understanding developmental mechanisms and processes in adolescence.

DESIGN AND DATA CONSIDERATIONS

In this section, we will discuss a number of important design and data issues that need to be carefully considered before a statistical model is adopted and specified.

Design Considerations

The first critical issue, of course, is the theoretical model driving the research. As found throughout this *Handbook*, many theoretical models exist that make explicit the mechanisms and processes of change. However, many theoretical models are not sufficiently developed to provide clear guidance. Longitudinal data are best suited for testing hypotheses derived from well articulated models of change. In this regard, theory is an analyst's best friend. Key theoretical considerations that need to be addressed include: what changes, what drives change, what is the functional form of change, what mediating and/or moderating mechanism are impacting change, how quickly change occurs, and whether the available measures are sufficiently calibrated and sensitive to capture all of these features. Explicitly addressing these and related considerations in the planning stages of a study provides the optimal basis for designing a useful and informative longitudinal investigation.

Well-articulated models of change and the theoretical expectations derived from them will drive a number of key design considerations. As we describe next, some key considerations include the interval of measurement, the functional form of change, how we might represent time, several issues of measurement, and whether we will rely on manifest or latent variable analyses.

Interval of Measurement

The first issue is the interval of measurement. The vast majority of studies on adolescent development conduct annual or semiannual assessments with little thought to whether the change process under study will be captured adequately. For example, many studies of the dynamics of adolescent social development hypothesize fluid and relatively rapid transmission of influence. Measurements that do not occur at the pace of (or faster than) the change process cannot yield much meaningful information about it. Figure 2.1 illustrates the problem. Measurement intervals that do not keep pace with the change process provide, at best, a static replication of the cross-sectional information identified at a first measurement occasion. Any associations with change from one time point to the next do not reflect true change but rather are likely to reflect common contextual features associated with the timing of the measurement. For example, an annual school-based study that assesses indicators of affect at each measurement occasion (fall semester of a school year), may find modest stability of affect and may find that increases or decreases in affect are reliably predicted across the one-year interval. Attributing causal processes to such findings is likely unwarranted because the observed associations are confounded with the context of the fall semester of a school year (i.e., beginning of new school year with new classes, teachers, and peers). The true processes of change associated with affect likely operate over much shorter intervals and the observed associations are therefore spurious.

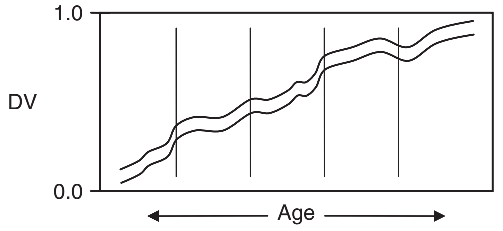


FIGURE 2.1 When measurement interval does not capture rate of change. *Note:* At a given age range, the overall nonlinear pattern can be reasonably approximated with a linear trend. If one measured across the entire age range, then a nonlinear model could be fit to the data.

Functional Form of Change

Another consideration is the functional form of the change process. Many theoretical models propose a functional form that is generally nonlinear in nature (e.g., initially rapid growth dampened by a deceleration with age). Globally speaking, across the life span, nonlinear change functions are likely and perhaps common. However, depending on the adequacy of the study design, the functional form of the process under study may not be captured sufficiently to model it with a nonlinear statistical model. Locally speaking, an adequate statistical model may in fact fit quite adequately as a linear function. Figure 2.2 illustrates this issue. Here, the appropriate statistical model may very well be a simple linear approximation even though the underlying theoretical model is explicitly nonlinear. In order to statistically capture the expected nonlinear nature of the growth pattern, measurements need to be taken well before and well after the expected bend of a curve. These extra measurements provide stability for estimating the "tails" accurately so that the bend can be modeled. Extra

FIGURE 2.2 Comparison of local linear change and global nonlinear change.

measurements at the bend of the expected curve can also provide more stable estimates of the functional form.

However, researchers should not be dismayed if their theory suggests nonlinear growth. If they are not able to collect data at sufficient numbers of time points, their data may well be reasonably approximated with a linear trajectory. Local linear approximations can be quite powerful in capturing the information found in a narrow band of otherwise nonlinear change. However, we encourage researchers to design data collection intervals that allow one to adequately test a theoretical model that hypothesizes nonlinear growth (see e.g., Blozis, Conger, & Harring, 2007).

Representing Time

A third issue involves the unit of time used to index the developmental process. Age in years is a common index for representing developmental processes. However, age alone may not be the optimal index. Rather than automatically thinking of time in terms of age in years, we encourage adolescence researchers to consider what the true index of time is over which change occurs. Next, we offer just a few possibilities to highlight the various ways that time can be represented in longitudinal models—these are just some of a few interesting and innovative ways in which time can be conceptualized and utilized in longitudinal studies (see Wohlwill, 1973).

One alternative index of development is experiential time. This type of index of time is common in studies of childhood and adolescence in which researchers use grade in school to reflect the "time" axis. Here, the developmental changes of participants are assumed to be mostly determined by grade-related experiences and not the underlying maturational age of the adolescent. For some phenomena, this assumption is probably reasonable, particularly for academic-related outcomes. For other outcomes, grade in school may not be an optimal index of development. Other examples of experiential time include amount of experience in extracurricular activities (e.g., first versus fourth year in club) or number of relationships (e.g., first versus later boyfriend/girlfriend), to suggest just a few.

Another alternative index of development is time to or since a key developmental episode. Puberty is a good example of a meaningful episode that has significant implications for understanding development. Here, the index of development would be time leading up to the episode (pubertal onset) and time after the episode. So, we might consider "time" as being two years before puberty, one year before, year of pubertal transition, one year after, and so on. Other relevant episodes during adolescence and emergent adulthood on which we might index time include school transitions (e.g., transition to high school, graduation from high school), moving out of the home, or marriage.

In these alternative conceptualizations of developmental processes, the chronological age of individuals can, and probably should, also be included as a context covariate. For example, research on pubertal onset has shown that early versus late onset, relative to adolescents' chronological age, has implications for adjustment (e.g., Zehr, Culbert, Sisk, & Klump, 2007). However, models of the change process prior to and following puberty that give rise to the adjustment difficulties are not possible if chronological age is used as the primary index of time. The data can be reorganized to center on the episode (e.g., pubertal onset) and how the repeated measurements before and after the episode can be used to model the dynamics of change both before and after the event. Note that such rearrangements of the data lead to a number of unobserved measurement occasions for each individual. Later, we discuss why the unobserved occasions can be effectively treated as a simple missing data problem whereby the missing information can be accurately and validly imputed.

Measurement Across Time

A fourth issue involves measurement. In this context, a number of features of measurement

TABLE 2.1 A Simple Taxonomy of the Nature of the Measured Variables Crossed with the Nature of the Latent Variables Found in General SEM Models

		Nature of the Unobserved, Latent Variables	
		Categorical	**Metrical**
Nature of the Observed, Manifest Variables	**Categorical**	Latent Transition Analysis	Latent Trait Analysis; IRT
	Metrical	Latent Class; Mixture Modeling	Latent Variable Analysis; CFA

Note: *Categorical* refers to variables that reflect nominal or polytomous categories. *Metrical* refers to variables that reflect ordinal, interval, or ratio-level properties. Many models for longitudinal data can contain more than one of these kinds of variables.

need to be considered. The first feature is the nature of the observed variable (our measures) and the nature of the latent variable (the underlying construct about which we wish to draw conclusions). Table 2.1 shows a general taxonomy of the kinds of analyses that are appropriate given the nature of the observed and latent variables. Quite often, a given project will contain a mix of nominal (or polytomous) and metrical (e.g., interval, ratio) manifest and latent variables. Mapping these measurements onto the appropriate statistical tool is important for ensuring valid conclusions. Commonly, a given study will have a combination of types of variables at both the manifest and latent levels. Most modern statistical programs can readily estimate models that contain such combinations (e.g., Muthén & Muthén, 2007), but often do so by either correcting for the restrictions of the noncontinuous measures or imposing assumptions in the estimation phase that may or may not be reasonable. Our advice is to consider carefully the nature of the variables to be collected and spend both thought and time to develop better and more precise measures. In our view, many researchers do not prepare thoroughly enough before embarking on a study. Examining a battery of measures and attempting to improve on their fidelity of measurement, for example, would enhance the overall ability of a longitudinal study to detect change. For a list of resources on innovations in measurement, visit www.Quant.KU.edu.

A second measurement feature is the developmental appropriateness of the measures. Embretson (2007), for example, outlines some serious measurement problems that can occur when trying to model developmental change processes. When measures are not appropriately calibrated for age-related differences in the phenomenon under study, they can provide dramatically biased estimates of the true growth function. Embretson's simulations are based on the logic of item response theory (IRT). Instead of an item's being appropriate for a person of low ability or high ability, items are appropriate for persons who are more or less developed in the skill, attitude, belief, trait, or behavior that the item is intended to measure. In this regard, measurement instruments need to provide sufficient sensitivity across the low to high levels of the qualities that individuals in a study may possess. The sensitivity and appropriateness of a measure are dependent on both the developmental age of the person and the person's individual characteristics. The broader implication here is that IRT methodology should be applied much more regularly across the developmental, social, and behavioral sciences in order to evaluate the breadth of item coverage of a measure.

A related measurement issue is the homotypic versus heterotypic expression of a construct across time (Sears, 1947). In many developmental studies, the expression of the underlying construct changes over time; this changing expression of a common underlying construct is referred to as heterotypic continuity (in contrast to homotypic continuity, or the same expression of the construct across time). Many of the behaviors that indicate aggression, for example, change as the child matures through adolescence into adulthood. To account for this heterotypic continuity, it is necessary to change the measures used to assess the construct across time. However, this change should not be an all-or-nothing decision; too often, researchers implement a wholesale switch from one measurement battery to another when the participants reach an age at which the former instrument is no longer deemed developmentally appropriate. Such decisions in the design of measurement protocols can have disastrous effects on the ability to model change over time. The key problem here is the lack of a linking function to allow one to map the meaning of a score on the earlier measure to a score on the later measure. A simple remedy to this problem is to phase in the new instrument and to phase out the old. Having one or more measurement occasions in which all items (or a key subset of items) from both instruments are assessed allows one to statistically calibrate the scores across the two instruments, which then allows one to model growth trends across the measures used in a different span of the study. Here, again, IRT methodology is ideally suited to provide the calibration that would allow one to model growth in circumstances when the item pool for measuring a construct must change.

Another related issue has to do with the cross-time factorial invariance (i.e., measurement equivalence) of the constructs. In all studies that do not use latent variable structural equation modeling (SEM) procedures, factorial invariance is simply assumed. Factorial invariance refers to the idea that a construct's indicators retain their relative relation with the construct, in terms of their pattern of intercepts and factor loadings, across time (and across groups; Card & Little, 2006; Little, Card, Slegers, & Ledford, 2007; Meredith, 1993; Selig, Card, & Little, 2008; Vandenberg & Lance, 2000; Widaman & Reise, 1997). Fortunately, with SEM procedures, the assumption that the construct's relative measurement characteristics have not changed over time (i.e., factorial invariance) can be readily assessed. Demonstrating the measurement equivalence of a set of indicators over time is critically important for the quality of generalizations that one can draw. If invariance of the loadings and the intercepts holds, then one can attribute all observed changes to changes in the latent constructs; that is, the potential confound of changes in the relative measurement properties of the items has been ruled out. If one does not establish measurement invariance, then this potential confound of changes in the measurement properties remains a very real threat to validity (Shadish et al., 2002).

Partial measurement invariance refers to the idea that one or more loadings or intercepts are not proportionally invariant over time. Although rules of thumb vary, if a reasonable number of items or indicators show invariance while a minority of items do not, one can compare and discuss changes that reflect true changes in the underlying construct (Reise, Widaman, & Pugh, 1993; Byrne, Shavelson, & Muthén, 1989). Finding noninvariance of a few of the indicators of a construct reflects an important outcome of a study that should be examined and discussed in some detail, because the lack of invariance likely was caused by some developmental process or external factor. Here, the sensitivity of the indicator may have changed or the composition of the construct may have changed.

Latent versus Manifest Variables

Longitudinal data have their advantages, and so, too, do particular methods of data analysis. Although we will discuss some manifest (observed) variable techniques for the analysis of longitudinal data, we will emphasize the latent variable approaches encompassed the class of techniques termed *structural equation modeling* (SEM; Bollen, 1989), which encompasses confirmatory factor analysis (CFA; latent variable analysis with unstructured bivariate associations among latent variables; Brown, 2006) and mean and covariance structures (MACS; latent variable analysis including information about variable means; Little, 1997). Latent variable analyses are advantageous for a number of important reasons. First, when multiple indicators of a construct are employed, the common variance among the indicators provides information about the construct that is (in theory) free of measurement error (Little, Lindenberger, & Nesselroade, 1999). Measurement error is a nefarious problem in manifest variable approaches because it contributes to the unreliability of a measure and contaminates the true score information that one seeks to measure. The assumptions and accuracy of inferences based on classical manifest variable analyses (e.g., regression, analysis of variance [ANOVA]) are undermined when variables are measured with some degree of error (i.e., classical techniques assume that variables are measured without error). In the behavioral and social sciences, measurement is replete with different degrees of unreliability. Unreliability leads to systematic under- or overestimation of the true relations or associations among the constructs analyzed, which can lead to various errors of inference and generalizability (for a detailed discussion of these issues, see Cole & Maxwell, 2003; and Widaman, Little, Preacher, & Sawalani, in press).

In addition to estimating and correcting for unreliability, latent variable SEM approaches provide a wealth of important validity information. For example, the degree to which the multiple indicators converge on the measurement of a common construct provides information about the content validity of the indicators. When multiple constructs, each with multiple indicators, are included in an SEM analysis, the convergent and discriminant patterns of validity in both the indicators and the constructs are fully exposed. Well-fitting models signify, for example, that indicators converge on the constructs they are intended to measure and the pattern of latent correlations among the constructs informs the degree of discrimination among the focal constructs as well as the criterion-related validity of each construct.

Another major assumption of classical techniques is that a measure provides equivalent measurement across time and across subgroups of individuals. As mentioned, this assumption of factorial invariance is one that can be easily specified and tested in the context of latent variable SEM approaches.

We raise the values of latent variable analysis in this section because these techniques require proper planning at the design stage. Specifically, researchers need to plan to collect multiple indicators of the constructs of interest. At a minimum, these multiple indicators might be multiple items from one scale. Ideally, these multiple indicators might come from different scales or be based on different information sources, so as to reduce the specific yet undesired shared components of these multiple indicators (e.g., shared reporter variance). In order to produce what is known as a just-identified measurement structure, we recommend that researchers collect at least three indicators of a construct (see Little, Slegers, & Card, 2006). If more than three indicators are obtained, researchers can use parceling techniques to include all of this information within this just-identified measurement structure (see Little, Cunningham, Shahar, & Widaman, 2002). The key points here are that latent variable analyses have definite advantages over manifest variable approaches, but researchers need to plan to have multiple indicators of constructs in their designs in order to conduct such analyses.

MISSING DATA

It is often said that the best way to handle missing data is to not have missing data. There is some truth to this adage, and researchers should make every effort to collect all data from all participants (e.g., repeatedly visiting schools to obtain data from students absent on testing days). In reality, however, data will be missing no matter how Herculean the effort to collect complete data. As we describe next, ample evidence points to the conclusion that full information maximum likelihood (FIML) estimation and iterative imputation algorithms (e.g., expectation maximization, EM; Markov chain Monte Carlo, MCMC) are the best approaches to handling missing data (Enders, in press; Little & Rubin, 2002; Schafer & Graham, 2002).

In longitudinal research, the reasons why data may be missing are numerous. Each of these reasons would reflect a missing data mechanism. One critical missing data mechanism in longitudinal research is attrition. Participants drop out from a study for various reasons such as family mobility, lack of interest in further participation, or unfavorable circumstances that coincide with a measurement occasion. Clearly, all reasonable effort should be taken to retain participants in a study. Using appropriate incentives, staying in contact, and providing "something" in return are each useful in maximizing retention. When efforts such as these fail, however, the missing observations are not to be viewed as a complete loss of information from the study, as they would be treated using classical methods of pairwise or listwise deletion. Instead, one can utilize all available information from each participant in a study to inform the estimation of parameters of the statistical model employed (Little, Lindenberger, & Maier, 2000).

Before describing the ways to manage missing data, it is useful to consider the three common terms used to describe patterns of missingness. As outlined in Table 2.2, data can be missing completely at random (MCAR), missing functionally at random (MAR), or not missing at random (NMAR). As the name implies, MCAR data are missing due to a completely random process. Examples of how this might arise are if some questionnaires were accidentally destroyed or a pesky gremlin randomly deleted cells within our database (and we were unable to replace these scores). The unrealistic nature of these examples speaks to the likely unreality of the MCAR pattern. The second pattern, MAR, is somewhat misleading in its label. Data that are MAR are not technically missing at random. Instead, there is a relation between the missing data mechanism and information contained in the dataset (variables that have been measured), but *no relation* between the missing data mechanism and information not contained in the dataset (variables that have not been measured). For example, if missing data are highly predicted by absenteeism (students who frequently miss school are less likely to complete questionnaires at that time point) and we have a measure of attendance (e.g., access to school records), then missingness would be considered MAR. Finally, NMAR implies that the missingness is associated with some variable

TABLE 2.2 Types of missing data patterns

	No Association with *any* observed variable(s)	**An association with *Analyzed* variables**	**An association with *Unanalyzed* variables**
	MCAR	**MAR**	**MAR**
No Association with unobserved/unmeasured variables	Fully recoverable Fully unbiased	Mostly recoverable Mostly unbiased	Mostly recoverable Mostly unbiased
	NMAR	**NMAR + MAR**	**NMAR + MAR**
An Association with unobserved/unmeasured variables	Not recoverable As biased as not imputing	Partly recoverable Less biased than not imputing	Partly recoverable Less biased than not imputing

Note: 'Recoverable' refers to recovering the missing data processes and 'bias' refers to the accuracy of conclusions relative to analyzing complete case data only. In all instances, power will be maximized by estimating missing data. The 'association' here refers to the reliable relation between the measured or unmeasured variables and the missing data process. In most cases, this association is assumed to be linear. The distinction between analyzed vs. unanalyzed variables refers to the variables selected for a given analysis vs. the variables on the dataset that are not selected for a given analysis.

that is *not* contained in the dataset. From the previous example, if the researcher did not have a measure of attendance, then the pattern would be considered NMAR.

Data can be missing for many reasons and at many levels and, for the most part, each of these possible mechanisms (MCAR, MAR, and NMAR) will contribute to the missing data found in a longitudinal study. In each situation, it is better to use FIML procedures or impute missing data than to rely only on complete case analysis (i.e., listwise or pairwise deletion). When data are MCAR, any approach (imputation or deletion) will yield unbiased parameter estimates (i.e., means, standard deviations, and associations); however, as we describe later, imputation (or FIML) will provide the maximum statistical power. In the more common cases of MAR or NMAR, imputation (or FIML) is also preferable to deletion. Inferences become compromised when the missing data process results from, or is associated with, variables that are unmeasured (NMAR). Here, the reasons for the missingness, because they are unmeasured, cannot be used to help inform the estimation algorithms, and the missing data process will not be recovered. However, the inferences to the population based on the sample will be just as biased as those based on complete case analyses (i.e., using listwise or pairwise deletion). When there is a relation between information contained in the dataset (variables that have been measured) and the missing data mechanism (MAR), the best-practice techniques for handling missing data do a very good job of recreating the missing data process so that inferences are less biased than otherwise.

Because missing data are almost always missing due to processes that arise from all three of these mechanisms, utilizing all the information in the dataset in the estimation or imputation process will lead to two significant advantages: (1) the full power of the original sample will be retained, and (2) inferences will be as generalizable as possible. Here, an important distinction needs to be made between analyses that use FIML estimation of a reduced set of the possible variables contained in the dataset versus imputation procedures that utilize all variables in the dataset (analyzed and unanalyzed). If a variable that is associated with the missing data process is not included in an analysis, then it cannot inform the estimation process, and the missing data process associated with the unanalyzed variable will not be recovered. In such situations, the bias that could have been corrected by including the unanalyzed variable will not be corrected. Because longitudinal datasets often contain a broad array of variables in the data set (including many that are not relevant to a particular analysis or research question), we recommend that researchers use all available variables in the data set and use a multiple imputation procedure such as Norm (see Schafer, 1997), SAS PROC MI (www.SAS.com), or the R module Amelia II (Honaker, King, & Blackwell, 2008) (for information on these and other imputation software, visit www.Quant.KU.edu).

In longitudinal data, a number of aspects of imputation must be considered. First, all imputation algorithms assume that the association between the observed data and the missing data process are linear. Potential nonlinearities can be included by creating additional variables such as interaction terms and powered polynomials and using these informative variables in the missing data imputation process. Even if these informative variables are not to be utilized to address any theoretical questions, they can provide more refined information to better condition the imputation of missing data.

Collecting and analyzing data from an initial wave of data collection is common practice when embarking on a longitudinal study. Any missing data at this initial stage would be imputed, and analyses based on these data would be subsequently reported. A question emerges, however, when the second wave of data is collected. Should the missing data at Wave 1 be reestimated using the information from Wave 2? In our view, the answer is yes. The logic for this answer is that information about how an individual changes at Wave 2 provides additional information that increases the likelihood of accurately recovering the missing data process. Keep in mind that modern missing data imputation is an agnostic affair when it comes to causality or time-ordered relations (Enders, in press). All that a missing data procedure attempts to accomplish is to optimize a variance–covariance matrix and mean vector that looks as much like the population as possible given all the available information provided by the sample at hand. To do so, the procedures impute likely values where data are missing and then estimates sufficient statistics; it then uses these sufficient statistics to reevaluate the likely values imputed into the missing data cells, reestimates these likely values, and then reestimates the sufficient statistics. This process continues until the change from one iteration to the next is trivially small. Investigators should not worry about whether a variable in the data set is an outcome, predictor, or a diagnostic category.

The agnostic process of estimation means that one can estimate missing information for any type of variable. Gender, for example, can be estimated in this manner. If an estimated value for gender comes out to be 0.7 and gender is coded 0 and 1, then the most likely gender of this individual is the sex coded as 1 (male). For categorization purposes (e.g., in multiple-group analyses), this case can be treated as if gender were male, but for analytic purposes the imputed value can be left as 0.7 in order to minimize any bias (Enders, in press).

Because dropout and attrition often lead to a significant percentage of missing data, longitudinal studies should routinely utilize multiple imputations. As described above, a single imputation is an iterative process. This iterative process is somewhat dependent on the initial guesses plugged in for the missing data. Multiple imputations allow the uncertainty and potential influence of the starting point on the standard errors of estimation to be minimized. Early work on multiple imputation suggested that a small number of imputations would be sufficient to capture this uncertainty. More recently, however, work in this area indicates that a larger number of imputations (>20) are needed in order to accurately capture the true information and inherent variability. With the increase in computational capabilities and the availability of software routines to easily combine and summarize the results from the multiple analyses of the multiple imputations, we recommend that researchers err on the side of more rather then fewer imputations (software routines and links to routines that summarize imputation results can be found at www.Quant.KU.edu). Graham, Olchowski, and Gilreath (2007) provide sound guidance for determining the appropriate number of imputations given the fraction of missing data present in a given sample (see also Enders, in press).

One key implication of this discussion of missing data is that one should carefully attend to, and plan to assess, important potential predictors of the missing data mechanism. Measuring those variables that consistently associate with missingness helps inform the imputation process such that the missing data are more likely to be MAR (recoverable) than NMAR (unrecoverable). A comprehensive review of the literature on adolescent development would provide a solid list of variables that show consistent associations with missing data processes and reasons for attrition. These variables should be routinely measured for their use in the imputation stages of a project (a partial list of recommended variables can be found at www.Quant.KU.edu).

One consequence of the ability to recover missing data using modern techniques has to do with intentionally missing data collection designs (see Graham, Taylor, Olchowski, & Cumsille, 2006). Such designs involve randomly assigning participants to different patterns of data collection occasions and/or variables in the protocol. As long as each possible pairwise association has sufficient coverage to accurately estimate the population covariance, the random nature of the missing data design will yield a missing data process that is fully recoverable.

ANALYSIS TECHNIQUES

Given that one has carefully considered the design needs for answering the questions of interest, the theoretical model should be matched with the appropriate statistical model. Many different considerations are involved in choosing the appropriate statistical model. Rosel and Plewis (2008) provide a taxonomy of statistical models that all fall under the broad umbrella of SEM. A first consideration is whether the model should be univariate or multivariate. Given that models of a single measure are relatively simple and specific to circumscribed questions, we will focus on multivariate models (models that include two or more constructs). A second consideration is whether the constructs should be modeled as observed, manifest variables or as unobserved, latent variables. We outlined the various advantages of latent variable approaches above, and focus more of our discussion below on latent variable applications (though we will give some consideration to manifest variable approaches). Other dimensions that can be considered include the presence or absence of covariates, interactions, multiple levels, and nonlinear effects (Rosel & Plewis, 2008). In addition to these considerations, models can be applied to categorical or metrical variables (Collins, 2006).

In the next three sections, we describe the basic features and issues in specifying panel models, growth curve models, and intensive time series models. We also address, when applicable, inclusion of covariates, handling multiple levels, and modeling nonlinear effects. We devote the Conclusions section to a discussion of mediation and moderation in the context of the three general classes of technique for modeling longitudinal data (for thorough treatments of these techniques, see, e.g., Bijleveld & van der Kamp, 1998; Cairns, Bergman, & Kagan, 1998; Collins & Horn, 1991; Collins & Sayer, 2001; Hedeker & Gibbons, 2006; Little, Schnabel, & Baumert, 2000; Magnusson, Bergman, Rudinger, & Törestad, 1991; Menard, 1991; Moskowitz & Hershberger, 2002; Nesselroade & Baltes, 1979; Saldaña, 2003).

PANEL MODELS

The first type of longitudinal model that we address is termed the *panel model.* Panel models are also referred to as autoregressive models. We prefer panel models to emphasize that data for these models usually reflect multiple measures at a few assessment occasions. In this regard, panel models consist, at a minimum, of data collected at two time points in which the presumed antecedent(s) is measured at Time 1 and the presumed consequence(s) is measured at Time 2. We emphasize that this is a minimum condition, and there are substantial advantages to measuring both antecedents and consequences at both times, measuring alternative predictor variables at Time 1 (and preferably also at Time 2), and obtaining data from more than two time points. We elaborate on each of these considerations in the next subsection, describing the logic of panel models and the kinds of questions that can be answered. We also note that it is possible to analyze panel data using either manifest (i.e., multiple regression) or latent variable analysis, as described in the second and third subsections (respectively) below.

Logic of and Questions Answered by Panel Models

As mentioned, panel models involve measuring the presumed antecedents and consequences at two (or more) time points. The focus of panel models is on interindividual (i.e., between person) differences. Specifically, analyses of panel models answer the question of whether (and to what extent) interindividual differences in the presumed antecedent at Time 1 are predictive of later (Time 2) interindividual differences in the presumed consequent. In other words, does a person's relative standing on variable X at one time point relate to that person's relative standing on variable Y at a later time point?

The key analytic consideration of panel models is the presence of associations among variables across time. Three aspects of these associations (e.g., correlations) merit attention. First, as alluded to earlier, associations involve the covariation of interindividual differences among variables, in this case among variables over time. These covariances or correlations tell us nothing about whether the mean level of a variable or set of variables is increasing, decreasing, or staying the same. To address these questions, growth curve models are preferred (see next section). Second, there are several ways to assess correlations, including a manifest variable (i.e., regression) framework as well as a latent variable (i.e., structural equation) framework. Although there are several advantages of the latter approach, which we highlight throughout, we want to emphasize that there is not a single right way to analyze panel models. As with all statistical models, the important issue is whether the analytic model answers the questions the researcher wants to ask (see preceding discussion). The third consideration of our initial statement is that there is intentional breadth in terms of variables included. Examining the across-time correlation of a single variable provides important information regarding the interindividual (but not intraindividual, or within-person) stability of that variable. More commonly, researchers think of panel models involving two variables, in which associations between X and Y across time are of interest. However, panel models are not restricted to just two variables, and it is often valuable to compare the relative predictive strength of multiple variables across time, or whether these longitudinal associations are mediated or moderated (see the section, *Mediation and Moderation in Longitudinal Data*).

Imagine that a researcher believes that X causes Y over time. More explicitly, imagine that this researcher believes that an adolescent's level of X relative to peers (i.e., interindividual differences in X) leads to an adolescent's being high or low (relative to peers) on Y at some specific later time. The first task of the researcher is to explicate the amount of time over which the presumed influence is expected to occur (see Gollob & Reichardt, 1987). Then, the researcher collects data measuring (at a minimum) X at one time, waits the expected amount of time over which the influence occurs, and then collects data measuring (at a minimum) Y at this second time point. The researcher then analyzes these data to determine the existence and magnitude of the association between Time 1 X and Time 2 Y.

Imagine that the researcher does indeed find an association between Time 1 X (denoted X_1) and Time 2 Y (denoted Y_2), as shown in Figure 2.3, situation A. Based on this association, can it be concluded that X causes Y? Although many researchers might be tempted to jump to this conclusion, such a conclusion would *not* be appropriate. A failure to find this association would refute the researcher's hypothesis that X causes Y over the specified time period (to the extent that there was adequate statistical power to detect an effect of a certain size). A finding that X_1 is associated with Y_2 *might* mean that X causes Y, but there are at least two alternative explanations.

The first alternative explanation is that, instead of X causing Y, Y actually causes X. This possibility is illustrated in Figure 2.3, situation B, where we have shown earlier values of Y (denoted Y_0) causing X (at Time 1, X_1). If Y is also stable across time (in terms of interindividual differences), then an association between Y_0 and Y_2 is also likely. Based on these two associations, the detected association between X_1 and Y_2 can be considered spurious, in that it is due *not* to X causing Y, but rather Y causing X and Y being stable across time. One way to account for this alternative explanation is to measure X and Y at both occasions, and evaluate the extent to which X_1 predicts later Y_2 (controlling for Y_1) *and* the extent to which Y_1 predicts later X_2 (controlling for X_1). This evaluation is depicted in Figure 2.3, situation C. If the researcher finds that X_1 predicts later

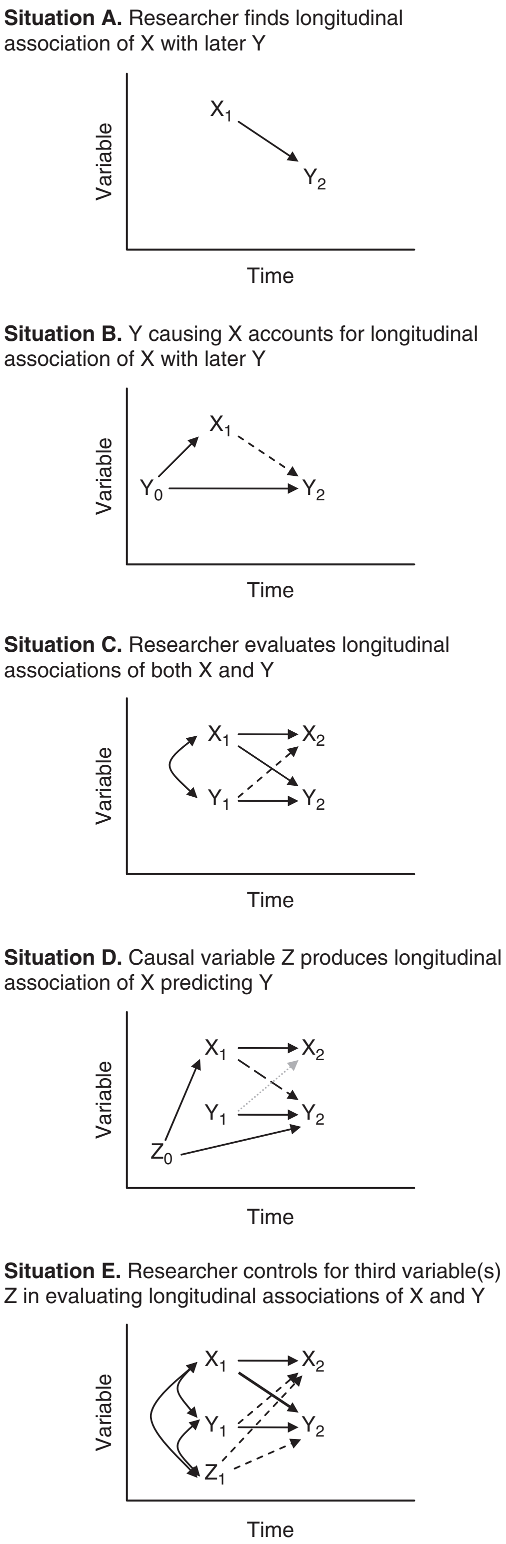


FIGURE 2.3 An illustration of different panel models in longitudinal research.

Y_2 but Y_1 does not predict later X_2 (which is why this path is indicated with a dashed line in Figure 2.3, situation C), then it is plausible to rule out this alternative explanation if two assumptions can be made.

The first assumption is that the time span of influence between measurements of the two variables is equal. In this hypothetical example, we emphasized that the researcher carefully considered the time span over which X was expected to influence Y (i.e., the researcher has measured the variables at the "correct" interval to detect this association). There is no guarantee, however, that this time span is also that over which Y might influence X. Analysis of data in which an inappropriate time span for the influence of Y on X would not provide an appropriate test of this explanation; therefore, the failure to find a predictive relation of Y_1 toward X_2 among these data would not provide a convincing case for ruling out this direction of influence. Researchers who wish to evaluate competing directions of influence between X and Y must carefully consider the time span over which *both* processes occur. If it can be convincingly (and not just conveniently) argued that the two processes occur across the same period of time, then the analyses depicted in Figure 2.3, situation C can compare these two directions of influence. If this argument cannot be made, then more than two occasions of measurement are needed, as described below.

The second assumption that must be made is even more difficult to support. This assumption is that no third variable causes both X and Y. This threat to inferences of causality is pervasive and can never fully be ruled out in a naturalistic (i.e., nonexperimental) study; it is known as the "third variable problem." This problem is illustrated in Figure 2.3, situation D, in which variable Z represents a third variable, potentially causing both X and Y. The threat is that the observed relation between X and Y may not be due to X causing Y, but rather to the mutual reliance of X and Y on the same third variable Z. Specifically, the threat is that some third variable Z causes X_1 as well as Y_2, and this common cause is what accounts for the spurious longitudinal association between X and Y. The third variable problem is widely invoked by critics of research attempting to draw causal inferences from longitudinal naturalistic data. We propose two rebuttals to these sorts of critique.

The first rebuttal is to ask critics to explicate the time span of influence of the third variable Z. Given the existence of the pattern of associations shown in Figure 2.3, situation C, the critique of a potential third variable as the true cause should specify one of two conditions. First, a critic invoking the third variable threat should be pressed to specify that the influence of Z on X occurs more quickly than the influence of Z on Y. Alternatively, one invoking the third variable critique might specify that the third variable Z causes Y but not X, and that X is merely associated with Z. Too often, the third variable critique is raised in a nonspecific way that does not consider the specifics of influence. Although it is the researcher's responsibility to rule out competing explanations, we believe that those raising this critique should offer a reasonable explication of this differential time span of influence.

The second—and more authoritative—rebuttal to the third variable problem is to actually measure and include the third variable in one's model. Although there is a potentially infinite number of third variables (i.e., multiple Zs) that could be invoked, the researcher should plan to address the most theoretically defensible possibilities by including measures of these third variables (at a minimum, at Time 1 to evaluate the prediction of later X and Y from earlier Zs). If X_1 is found to predict later Y_2 even after controlling for relevant Z_1s, this offers evidence against the threat of the third variable as the true causal mechanism (this scenario is depicted in Figure 2.3, situation E). Of course, the effectiveness of controlling for Zs in ruling out the third variable threat depends on the expected influence of Z occurring over the same time span as the $X \rightarrow Y$ relation.

Despite these methods of evaluating competing $X \rightarrow Y$ versus $Y \rightarrow X$ explanations and potential rebuttals to third variable problem critiques, analysis of panel models cannot conclusively demonstrate causality. There is always the possibility that Y causes X over time spans other than those studied, which limits our conclusions regarding the direction of influence between X and Y to the particular time span investigated. The third variable problem can never be fully ruled out, as there are potentially infinite Zs that we have not considered that could account for the observed longitudinal associations. At the same time, analysis of panel models with adequate consideration of these potential problems allows us to build a strong case for causality (especially if we control for most theoretically viable third variable causes), and certainly the strongest case that can be made without experimental manipulation. Like all statistical models, panel models allow us to evaluate only whether the data are consistent (or not) with our theoretical predictions.

Manifest Variable Analysis of Panel Data

Manifest variable analysis of panel models is typically performed using a series of multiple regressions. We will describe these analyses in a way paralleling the increasing sophistication of analyses described in the previous subsection (and Figure 2.3).

Consider first the simplest, but least optimal, case in which the researcher measures X at Time 1 and Y at Time 2. The regression model is straightforward:

$$Y_2 = B_0 + B_1 X_1 + e \tag{1}$$

Here, the estimate of interest is B_1, the regression coefficient of X_1. Specifically, one is interested in the statistical significance of this coefficient (does X predict later Y?) as well as its sign (does X predict higher or lower Y?) and magnitude (how strongly does X predict Y?). Magnitude can be evaluated using the standardized regression coefficient.

When researchers have measures of both X and Y at both time points, the regression analyses yield much more convincing information. Consider first the evaluation that X predicts Y from the following equation (for alternative possibilities, see Duncan, 1969):

$$Y_2 = B_0 + B_1Y_1 + B_2X_1 + e \quad (2)$$

Here, Y_2 is predicted by two variables. The first variable is the earlier value of the variable itself (Y_1). Inclusion of this predictor is important in two ways. First, the regression coefficient of this predictor (B_1) indicates the magnitude of interindividual stability in Y (usually evaluated in terms of significance and the magnitude of the standardized regression coefficient). Second, including the initial level of the dependent variable (often termed the autoregressive component) means that B_1 is interpreted as the extent to which X_1 predicts Y_2 above and beyond the stable variability of Y. In other words, the regression coefficient of X_1 (B_2) represents the extent to which X_1 predicts instability, or interindividual change, in Y_2.

A key value of including both X and Y at both time points is the ability to evaluate both directions of prediction, that is, X → Y and Y → X. Therefore, it is common to evaluate a second regression analysis in addition to that described earlier (Equation 2), assuming such an effect is theoretically tenable:

$$X_2 = B_0 + B_1X_1 + B_2Y_1 + e \quad (3)$$

The coefficients of this regression are interpreted in a way parallel to those just described. Namely, B_1 is interpreted as the interindividual stability in X, and B_2 is interpreted as the prediction of change in X from Y. We stated earlier that direction of influence between X and Y is indicated by finding that one variable predicts the other (e.g., X predicts change in Y) but that the other variable does not predict the first (e.g., Y fails to predict change in X). This interpretation suggests one of the critical limitations to using manifest variable regression analyses to analyze panel models—substantive decisions are based on inferential conclusions that are not easily comparable. An illustration should clarify this point. Imagine that the researcher found that X significantly predicts change in Y, with $B = 0.30$ and $p = 0.049$. In a separate regression predicting X, the researcher found that Y fails to predict change in X, with $B = 0.29$ and $p = 0.051$. Although letter-of-the-law hypothesis testing would lead to conclusions that X predicts Y but not vice versa, this example shows clearly that X's ability to predict Y is not substantially stronger than Y's ability to predict X. Statistically, methods of comparing these two predictions within a multiple regression framework are difficult (see Widaman, 2000), and are much more easily accomplished within an SEM framework as described below.

Consider next the situation in which the researcher wishes to control for one or more additional variables in order to rule out alternative causal explanations. Following from previous equations (for alternatives, see Duncan, 1969), the researcher would fit two regression equations, one with each Time 2 X and Y serving as dependent variables:

$$\begin{aligned} Y_2 &= B_0 + B_1Y_1 + B_2X_1 + B_3Z_{a1} \\ &\quad + B_4Z_{b1} + \cdots + e \end{aligned} \quad (4)$$

$$\begin{aligned} X_2 &= B_0 + B_1X_1 + B_2Y_1 + B_3Z_{a1} \\ &\quad + B_4Z_{b1} + \cdots + e \end{aligned} \quad (5)$$

These equations make clear that the researcher can control for as many potential alternative causal variables as desired.

Latent Variable Analysis of Panel Data

Analyzing panel models in a latent variable framework has several advantages over analysis in a multiple regression framework (e.g., Anderson, 1987; Little, Preacher, Selig & Card, 2007). In this subsection, we will first describe these advantages. We will then briefly describe the practice of fitting latent variable path models. However, space constraints preclude a full description of these practices, and we refer interested readers to Little, Preacher, et al. (2007; see Figure 2.4 for an example of a latent variable panel model) for further description.

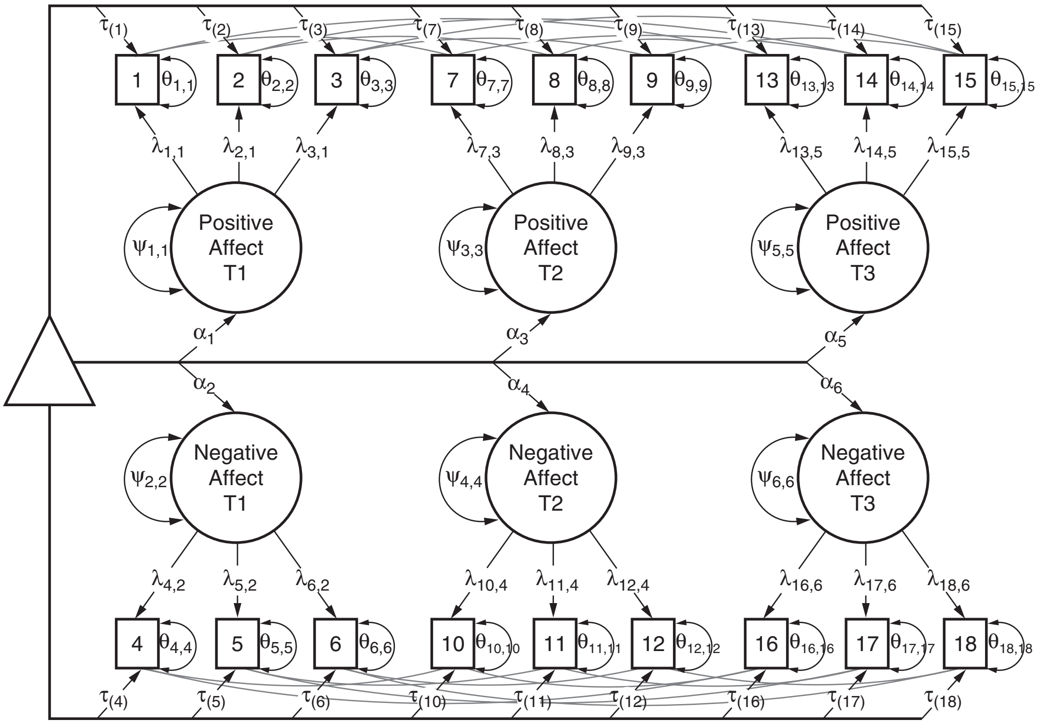


FIGURE 2.4 A detailed model of a three-wave panel model of two constructs, including all relevant para-meter estimates.

The first advantage of latent variable panel analysis over manifest variable regression techniques is simply the correction for unreliability. It is well known that unreliability attenuates (i.e., diminishes the estimated values of) correlations among variables; what is less often recognized is the unpredictable biases this causes in manifest variable regression analyses, such as the analysis of panel models. If we consider the longitudinal association between X_1 and Y_2, we know that unreliability of either of these variables will attenuate the longitudinal correlation between these variables. However, if we also consider unreliability of Y_1, the picture becomes more complex: now we have an attenuated stability estimate of Y and an attenuated estimate of the concurrent correlation between X_1 and Y_1. The net result of these two attenuations will be to inflate the unique association of X_1 predicting Y_2. We have no way of knowing if this regression coefficient is underestimated (due to attenuated correlation between X_1 and Y_2), overestimated (due to attenuated control of Y_1), or the competing attenuations balance each other to produce an unbiased estimate (the possibility that most researchers using regression seem to hope for, but that is probably least likely). If we also add additional third variables (i.e., Zs) to the equation, the problem is even further compounded. In short, attenuation due to unreliability biases parameter estimates in manifest variable regression in ways that are difficult to predict. The extent to which this problem has led to inappropriate substantive conclusions in regression analyses of panel models is frightful to consider. In contrast, latent variable representations of panel models correct for this unreliability to produce unbiased parameter estimates.

A second advantage of latent variable panel models is that they allow for evaluation of measurement equivalence across time. This is a considerable advantage when we consider the impact of changing measurement properties across time. Using manifest variable regression, one might regress Y_2 onto Y_1 and X_1. If we conclude that X predicts change in Y, we are really concluding that there is a change in the *construct* of Y that is predictable from earlier levels of X. An alternative possibility, if we do not establish equivalence of measurement across time, is that change in the *measurement* of Y is predictable from earlier levels of X. One possible cause of this alternative possibility is that some aspect of Y is becoming increasingly salient to the measurement of Y across time, and X is simply associated (not necessarily in a causal manner) with that particular aspect.

The remaining advantages are aspects of the general modeling tools of SEM, but also apply to manifest variable path analysis. However, if one were to go to the trouble of using these modeling programs, it would be advisable to conduct latent variable analyses given the advantages just described (assuming that one has multiple measured indicators of the constructs).

A third advantage of latent variable panel analysis is that paths within the model can be easily directly compared. We described earlier the challenges of comparing competing prediction paths across separate regression equations. This problem is solved within a latent variable framework. To compare whether an $X \rightarrow Y$ path is stronger than a $Y \rightarrow X$ path, one can statistically compare a model in which these two paths are freely estimated to a model in which these are constrained equally. If this constraint significantly worsens model fit (as evaluated by a significant increase in χ^2), then we can conclude that these two paths are significantly different from one another. This formal statistical comparison allows much greater clarity in decisions regarding direction of influence. Such a comparison, however, would have to be performed estimating the paths as standardized path coefficients if X and Y are measured on different scales; this is achieved through the use of phantom variables (Little, 1997; Rindskopf, 1984).

Latent variable panel models also allow for the evaluation of complex data. One complexity is the addition of more than two time points. This addition not only provides a replication of two time-point results (i.e., do Time 1 → Time 2 relations also emerge at Time 2 → Time 3?), but also allows for greater flexibility in evaluating various time spans of influence. This flexibility might allow researchers to detect, for example, not only whether X predicts change in Y over a shorter time frame (e.g., between Times 1 and 2) but also whether Y predicts X over a longer time frame (e.g., between Times 1 and 3). Increasing the number of time points also allows for evaluation of more complex longitudinal processes, such as longitudinal mediation (e.g., $X_1 \rightarrow M_2 \rightarrow Y_3$; see Cole & Maxwell, 2003; Gollob & Reichardt, 1991; Little, Card, Bovaird, Preacher, & Crandall, 2007; and discussion of mediation below).

Finally, multiple-group latent variable panel models allow several opportunities that are difficult or impossible in manifest variable regression. One opportunity is to evaluate whether the longitudinal predictions are moderated across different groups (e.g., differences by sex or ethnicity). This method of assessing moderation is accomplished by fitting the longitudinal model in multiple groups simultaneously, and then evaluating whether parameters (e.g., $X \rightarrow Y$) differ across groups. Although such interactions can be evaluated in multiple regression with the use of product terms, inclusion of these product terms quickly becomes cumbersome and the statistical power to detect these interactions with unreliable variables is often inadequate. Moreover, analysis of moderation via multiple-group SEM avoids the often unrealistic assumption of multiple regression of homoskedasticity (that the residual variances are equal across moderator groups; multiple-group SEM allows these residual variances to be freely estimated). A second opportunity afforded in multiple-group latent variable panel models is in the analysis of accelerated longitudinal designs (see Bell, 1953; Little, Card et al., 2007; Schaie, 1965). These designs involve collecting longitudinal data from several cohorts so that the cohort groups partially overlap in age. Treating these cohorts as groups in a multiple-group analysis allows one to evaluate whether the longitudinal predictions are similar or different across a wide range of age and cohort (for details, see Little, Card et al., 2007).

In addition to these basic considerations, users of SEM panel models might consider several recent advances not described in detail here. One of these issues involves the method of scaling the latent variable. In longitudinal studies in which the measures have a meaningful scale (e.g., frequency ratings), a recently proposed effects coding approach might be a useful method of scale setting (see Little, Slegers, & Card, 2006; Marsh, Wen, Hau, Little, Bovaird, & Widaman, 2007). In addition, Widaman and Thompson (2003) have described the necessity of specifying alternative null models for the computation of relative fit indices (e.g., comparative fit index [CFI], Tucker-Lewis index [TLI]) for cases in which the usual null model of independence is not nested within the theoretical model. It is appropriate to use these alternative null models in repeated-measures SEM applications, such as the latent variable panel model.

GROWTH CURVE MODELS

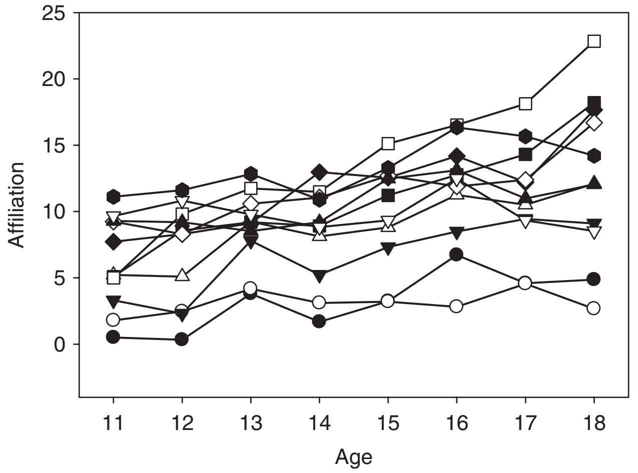
As mentioned, panel models focus on interindividual differences in changes over time. In the next section, we will describe techniques for modeling within-person, or intraindividual, changes over time.

The Logic of Growth Curve Models

Many modeling paradigms have a tendency to lose sight of the individual in an attempt to accurately model nomothetic laws guiding the relations among variables. Panel models are useful for modeling the temporal stability of (inter-)individual differences in constructs and the longitudinal causal linkages among different constructs. However, other models are required for modeling within-person (i.e., intraindividual) change over time, as well as interindividual differences in this intraindividual change. For example, throughout the adolescent years, individuals tend to steadily increase in affiliation with peers (time spent with friends rather than family; e.g., Rubin, Bukowski, & Parker, 2006). However, some adolescents likely increase in affiliation more quickly than others, and some individuals begin adolescence with lower or higher levels of affiliation relative to their peers. For example, the data may appear as in Figure 2.5, where each line represents the measurements of a single adolescent from ages 11 to 18.

For situations in which both the mean trajectory and variability in individual trajectories are of interest, *growth curve modeling* has become very popular. Unlike panel models, growth curve models focus on trajectories of change within the individual (*intraindividual change*) and individual differences in trajectories of change (*interindividual differences in interindividual change*). Depending on what is being measured, the appropriate unit of time may be minutes, weeks, months, or even years. A single study may be characterized by anywhere from 2 measurements to 50 or more. Measurements may occur simultaneously or at different times, and spacing among occasions may differ from occasion to occasion and from case to case. The researcher may be interested not only in modeling change, but in modeling the antecedents (predictors) and sequelae (outcomes) of change in several variables measured simultaneously. For example, a researcher may be interested in investigating potential predictors of change in teen affiliation and alcohol use in the same sample of high school students, as well as the potential for aspects of these trajectories to predict SAT scores. Such data characteristics would be difficult or impossible to address using older analyses such as analysis of covariance (ANCOVA), but growth curve modeling can be used to model change under all of these circumstances and more.

We briefly discuss two multivariate approaches to modeling change over time: (a) the multilevel model for repeated measures and (b) the latent growth curve model. Subsequently,



Note: Affiliation is measured as the average number of meaningful social encounters per day over the course of one 2-week period at roughly the same time every year.

FIGURE 2.5 Individual observed trajectories of affiliation for 11 fictitious adolescents, followed from age 11 to age 18.

we discuss the pros and cons of a recent extension of the latent growth curve model that is intended to identify latent classes characterized by unique trajectories—the growth mixture model.

The Multilevel Modeling Approach

Multilevel modeling (MLM), or hierarchical linear modeling (HLM), is an extension of regression analysis applied to hierarchically organized data that would otherwise violate the independence assumption of ordinary regression. For example, students "nested within" the same classroom likely are more similar on key educational outcomes than are students drawn from different classrooms; a shared environment can be expected to lead to increased similarity. MLM is very useful for analyzing repeated measures data, because repeated measures can be seen as nested within individuals in the same way that students are nested within classrooms.

The basic *level 1* equation for repeated measures data is:

$$y_{ij} = \beta_{0j} + \beta_{1j} time_{ij} + \varepsilon_{ij} \tag{6}$$

where y_{ij} is the measurement on some outcome y for individual j at time i, β_{0j} and β_{1j} are, respectively, the intercept and slope, and ε_{ij} is a residual term. The residuals are assumed to be normally distributed with mean 0 and variance σ_{ε}^2. Unlike intercepts and slopes in ordinary regression, the intercepts and slopes in multilevel regression can be treated as *random effects* with their own *level 2* equations:

$$\beta_{0j} = \gamma_{00} + u_{0j} \tag{7}$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \tag{8}$$

where γ_{00} and γ_{10} are the means of the distributions of β_{0j} and β_{1j}, u_{0j} and u_{1j} are level 2 residuals representing the deviations of individuals from those means, and u_{0j} and u_{1j} are assumed to be multivariate normally distributed with means 0, variances τ_{00} and τ_{11}, and covariance τ_{10}. That is, each individual is implicitly allowed to have his or her own intercept and slope. These individual coefficients are not actually estimated; rather, their means, variances, and covariances are estimated.

We could expand this basic linear model in many ways. For example, we could add predictors to the level 1 equation that are assessed at each occasion, in which case they are termed *time-varying covariates*. Returning to our affiliation example:

$$Affil_{ij} = \beta_{0j} + \beta_{1j} time_{ij} + \beta_{2j} mood_{ij} + \varepsilon_{ij} \quad (9)$$

We could also, or instead, add student-level predictors of intercepts and slopes (*time-invariant covariates*) to the level 2 equations, for example:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} IQ_j + u_{0j} \quad (10)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} IQ_j + u_{1j} \quad (11)$$

Substitution of Equations 10 and 11 into Equation 9 demonstrates that models with level 2 predictors of slopes lead to cross-level interaction effects, which are often of great interest to researchers:

$$\begin{aligned} y_{ij} &= \gamma_{00} + \gamma_{01} IQ_j + \gamma_{10} time_{ij} \\ &+ \gamma_{11} \underbrace{IQ_j \times time_{ij}}_{\substack{\text{cross-level} \\ \text{interaction term}}} + u_{0j} + u_{1j} time_{ij} + \varepsilon_{ij} \end{aligned} \quad (12)$$

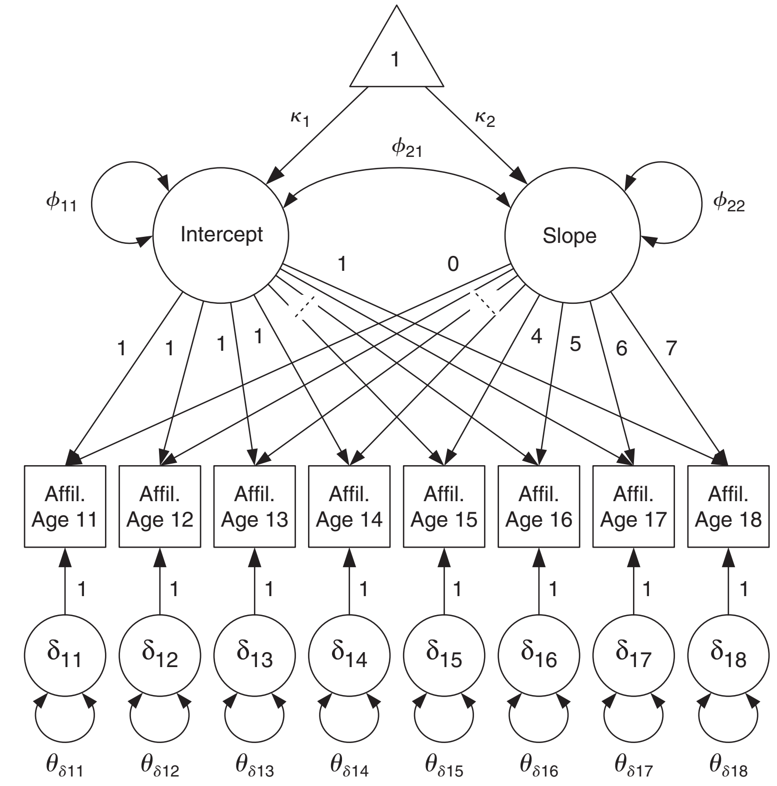
Multilevel models are advantageous for many reasons. Foremost, they avoid the parameter bias that occurs from mistakenly assuming independence. They also grant the researcher the ability to model regression weights as dependent variables in their own right, and to partition variance in meaningful ways to shed light on developmental phenomena. Software for fitting multilevel models is plentiful and can now be found in all major general statistics packages, including SPSS, SAS (PROC MIXED), R, Stata, and others, and in more specialized applications such as HLM, MLwiN, LISREL, and Mplus. Good introductions to MLM can be found in Hofmann (1997), Luke (2004), and Singer and Willett (2003). Many book-length treatments are available, including Bickel (2007), Hox (2002), Snijders and Bosker (1999), Kreft and de Leeuw (1998), and Raudenbush and Bryk (2002).

The Latent Growth Curve Approach

Latent growth curve modeling (LGM) is the application of SEM to the study of trajectories. LGM has its roots in factor analysis (Rao, 1958; Tucker, 1966), in which the aim was to recover polynomial trends in repeated measures data as factors. The main development leading to modern LGM occurred with the application of confirmatory factor analysis to repeated measures data (Meredith & Tisak, 1990). This development allowed scientists to test theory-based models of change in a confirmatory mode.

Because individuals are hypothesized to vary in terms of individual intercepts and slopes, and these aspects of change are not directly observable, it is natural to model these random effects as latent variables, or factors. Each of these *aspects of change* (e.g., intercept, linear slope, quadratic slope, and so on) is modeled as a latent variable, with the growth trend reflected in the pattern of loadings. The latent variables are often permitted to vary, indicating that individuals are permitted by the model to differ in their trajectories.

Consider the case of a simple linear model where individuals are hypothesized to differ in terms of both intercept and slope. A simple linear growth curve is depicted in the path diagram in Figure 2.6. Using standard SEM notation, yearly assessments of affiliation from ages 11 to 18 are arrayed as squares from left to right in chronological order. Circles represent latent variables and residual terms, and the triangle represents a constant. Affiliation at time j is modeled as a function of all the variables from which it receives arrows, weighted by the value of the associated path coefficient. For example, "Affil. Age 14" receives arrows from the intercept and slope factors, and from the random disturbance term δ_{14}. Its equation is therefore:



Note: Greek symbols represent parameters to be estimated, including κ_1 (mean intercept); κ_2 (mean slope); ϕ_{11}, ϕ_{22}, and ϕ_{21} (intercept and slope variances and covariance); and $\theta_{\delta 11} - \theta_{\delta 18}$ (occasion-specific disturbance variances).

FIGURE 2.6 A latent growth curve model specifying linear change in affiliation.

$$Affil_{14} = Intercept + Slope \times 3 + \delta_{14} \qquad (13)$$

More generally, affiliation at time j can be modeled as a function of time as:

$$Affil_j = Intercept + Slope \times time_j + \delta_j \qquad (14)$$

where $time_j$ is simply the value assigned to the factor loading associated with slope. In matrix notation, all of the equations can be summarized as:

$$\mathbf{Y} = \mathbf{\Lambda}\boldsymbol{\zeta} + \boldsymbol{\delta} \qquad (15)$$

where **Y** is a vector containing the repeated measures of affiliation, **Λ** is a matrix of factor loadings containing a column for each aspect of change (1s for intercept and linear increasing integers for slope), ζ is a vector containing the intercept and slope factors, and **δ** is a vector of disturbance residuals. The elements of **Λ** should be chosen with care, because any model parameters associated with the Intercept are conditional on the point at which the time metric equals "0." See our earlier discussion of how to choose an appropriate metric of time.

This model is termed an *unconditional* growth model because no predictors of intercepts or slopes are included. So the model for the observed means is very simple:

$$\boldsymbol{\mu}_y = \mathbf{\Lambda}\boldsymbol{\kappa} \qquad (16)$$

where $\boldsymbol{\mu}_y$ is a vector of modeled means and **κ** is a vector of latent variable means containing the mean intercept and slope. The associated covariance structure explaining the variances and covariances among the repeated measures is:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda\Phi\Lambda}' + \boldsymbol{\Theta}_\delta \tag{17}$$

where $\boldsymbol{\Phi}$ contains the variances and covariance of intercept and slope and $\boldsymbol{\Theta}_\delta$ is a diagonal matrix of disturbance variances. These models can be extended to include predictors of intercepts and slopes, but our purpose here is to provide an introduction to these models.

The primary benefit of LGM is its flexibility. For example, researchers are not limited to modeling equally spaced occasions—the factor loadings in $\boldsymbol{\Lambda}$ can be adapted to reflect any well-defined functional trend. Some software applications (Mplus and Mx) can even fit models to data in which each individual is assessed using his or her own unique measurement schedule. Missing data pose no obstacle in LGM (unless they are NMAR). Change in multiple variables may be assessed simultaneously, which permits estimation of covariances of intercepts and slopes across variables. LGM permits the assessment of change in latent variables with multiple indicators. Because aspects of change are treated as variables in LGM, the entire model in Figure 2.6 can be included in a larger structural model in a modular fashion. Furthermore, because LGM is a special application of SEM, researchers have access to an array of fit indices to aid in assessing model fit.

Numerous introductions to growth curve modeling are now available. Some of the more accessible ones include Byrne and Crombie (2003), Chan (1998), Curran (2000), Curran and Hussong (2003), Hancock and Lawrence (2006), and Willett and Sayer (1994). For more comprehensive treatments of LGM, we refer the reader to Bollen and Curran (2006); Preacher, Wichman, MacCallum, and Briggs (2008), and Duncan, Duncan, and Strycker (2006).

Interestingly, the basic form of both the multilevel model of change and the latent growth curve model are identical for a broad class of applications. Random effects in repeated measures of MLM are equivalent to latent aspects of change in LGM, and for many models the parameter estimates are identical across these two general methods (e.g., γ_{00} in Equation 7 is equivalent to κ_1 in Figure 2.6; see Bauer, 2003; Chou, Bentler, & Pentz, 1998; Curran, 2003; MacCallum, Kim, Malarkey, & Kiecolt-Glaser, 1997; Preacher et al., 2008; Rovine & Molenaar, 2000; Willett & Sayer, 1994).

Growth Mixture Modeling: Identifying Subgroups of Growth

One increasingly popular model for treating longitudinal data in developmental research is the *latent growth mixture model* (LGMM; see Muthén, 2008). Mixture modeling involves identifying subgroups or classes of individuals characterized by different patterns of model parameters. Rather than manually dividing the sample into groups, the researcher simply specifies the number of groups, and separate sets of model parameters are estimated for each group, along with probabilities of class membership. An LGMM, then, is mixture modeling applied to longitudinal data in an attempt to identify *K* subgroups (or latent classes) characterized by different latent growth trajectories (e.g., intercepts, slopes; see e.g., Barker, Tremblay, Nagin, Vitaro, & Lacourse, 2006; Muthén & Muthén, 2000).

Approaches to LGMM include those popularized by Nagin and colleagues (1999, 2005; Nagin & Tremblay, 1999) in which no heterogeneity is permitted in growth trajectories within classes, and those by Muthén and colleagues (Muthén, 2001; Muthén & Shedden, 1999) in which variability both within and between classes is permitted. Currently, the most flexible and popular software packages for fitting LGMMs are Mplus (Muthén & Muthén, 2007) and Proc TRAJ (Jones, Nagin, & Roeder, 2001; Nagin, 1999).

The rapid increase in the use of LGMM in developmental research has arguably outpaced careful consideration of the pros and cons of this method. Next, we describe both (1) the arguments against LGMMs, which should caution researchers against the uninformed use of these models; and (2) suggestions for using LGMMs, in which we describe some basic principles to follow if one feels theoretically justified in using LGMMs.

Arguments Against LGMM

Despite the rapidly increasing popularity of growth mixture modeling, Bauer (2007) offers a sobering litany of both methodological and theoretical concerns that, when carefully considered, should make developmental researchers very reluctant to apply such models without ample justification. First, the data to which LGMMs are applied are often nonnormal. Yet, because of the assumption of within-class normality, it is often easy for LGMMs to recover several artifactual classes, within each of which the assumption of normality is satisfied. In other words, nonnormal data can easily lead to the retention of spurious latent classes (Bauer, 2007; Bauer & Curran, 2003a, 2003b; Tofighi & Enders, 2007). Second, the specific growth model specifications within classes can give rise to spurious compensatory retention of too many classes as the model attempts to recover the observed variability in the data (Bauer & Curran, 2004). Class retention depends on the proper specification of a within-class model of growth. For example, if the effects of exogenous predictors of growth are actually nonlinear but are modeled as linear, spurious classes may result (Bauer, 2007). Third, if missing data are not missing at random (NMAR; see earlier discussion of missing data types) but are assumed to be MAR or MCAR, the number of classes can be underestimated (if multiple classes do exist). Fourth, in complex (e.g., stratified) samples, sampling probabilities need to be considered during parameter estimation, or else bias will result. It follows that if the researcher does not properly consider sample weights (the rule rather than the exception), then the number of classes, class proportions, and within-class parameter estimates can be considerably off the mark (for details of these issues, see Bauer, 2007).

Beyond purely methodological concerns, there are also theoretical and practical concerns that limit the current usefulness of LGMM. For example, LGMM is often used in cases where there is no reasonable basis for suspecting the existence of latent classes. Through random slopes, latent growth curve models already allow for the possibility of variability in trajectories. To justify LGMM, theory would have to predict the existence of discrete classes, each characterized by a different growth trajectory. Justifying the existence of latent taxa can be very difficult when continua provide a more parsimonious and (usually) more realistic alternative (Bauer, 2007; MacCallum, Zhang, Preacher, & Rucker, 2002).

Practically speaking, LGMMs are difficult to understand and use (though software advances are ameliorating this problem), improper solutions frequently occur, spurious classes are routinely extracted even in homogeneous populations, parameters tend to be very sensitive to starting values, and very large samples are required for accurate estimation. Furthermore, the question of how to properly apply fit statistics and model selection criteria to LGMMs is not entirely settled. We believe the nearly immediate popularity of LGMMs so soon after their introduction has done more harm than good. With time, many of the issues described above may be addressed and resolved by methodologists. Until then, however, we recommend that researchers employ extreme caution and conservatism when using LGMMs. To use LGMMs the researcher must appropriately address nonnormality (either by transforming variables or by properly modeling nonnormality) and have very strong theoretical reasons to suspect the existence of discrete classes. However, Bauer and Shanahan (2007) provide a sound example of how these techniques can be fruitfully employed that does not rely on a class interpretation, but instead they emphasize their use to approximate an unknown nonlinear function.

Suggestions If One Uses LGMM

Given the methodological difficulties and potential pitfalls of LGMM, researchers should exercise extreme restraint and caution if using LGMMs. If one feels amply justified on the basis of strong theory to employ LGMMs, one needs to follow the principles of replicability, interpretability, and predictability. These principles are general ones that cannot replace very strong theory that clearly states (1) why qualitatively distinct classes should exist, (2) how many classes should exist, and (3) what the functional form of the growth trajectories within each class should be. These principles also cannot replace quality measurement to provide data that meet the estimation needs of LGMMs. Nevertheless, these principles can strengthen the validity of conclusions drawn on the basis of LGMM.

Because of the susceptibility and tendency to find groups where groups may not exist, replicate the subgroup compositions using either cross-validation techniques (if sample sizes are very large; Browne & Cudeck, 1989) or independent samples (or subsamples). One form of replication that may be fruitful in this context is to employ an $N-k$ resampling technique (i.e., sample size, N, minus a sample fraction, k; called group jackknife; see Efron, 1979; Efron &Tibshirani, 1993) to derive subgroups and to compare the profiles and probabilities of group memberships across the repeated subsamples. This criterion, however, is the weakest supporting evidence because if the data that one analyzes have problematic characteristics that give rise to spurious classes, then these spurious classes are likely to reemerge. The best use of this criterion is to collect new data with improved measures that clearly satisfy the estimation needs of LGMMs.

Interpretability refers to the theoretical basis for the expectation of subgroups that emerge from LGMMs. Here, specifying a priori, and grounding in sound theory, the number and expected nature of all groups that should emerge would provide a basis to feel encouraged that the derived groups are meaningful. Too often researchers succumb to the tendency to "Label After Results Are Known" (LARKing; c.f. HARKing, Kerr, 1998). Post hoc labeling of subgroups encourages reification of the potentially random subgroups that emerge. The more that theory can be brought to bear in LGMMs, the more the technique has a confirmatory nature to it and the more likely it is that the identified groups are meaningful.

The third key to supporting the validity of the LGMM results is predictability. Again, on the basis of *very strong predictions grounded in theory*, specifying a set of variables that can predict and reliably differentiate the subgroup characteristics would add the strength of criterion validity information to the work. Moreover, specifying a set of criterion variables with which the subgroups show reliable associations would further strengthen the validity arguments in favor of meaningful subgroups. These criterion associations also need to be grounded in theory and should provide clear statements of both the direction and magnitude of associations.

These basic principles can provide a broad set of methodologically grounded arguments to support the results of LGMMs. For the most part, however, we strongly encourage researchers to resist the temptation to apply these models simply because of their intuitive appeal. Unlike other statistical methods in which violations of assumptions do not completely undermine the utility of the procedures, LGMMs are quite sensitive to assumption violations.

Combining Panel Models With Latent Growth Curve Models

Bollen and Curran (2004, 2006) and Curran and Bollen (2001) emphasize that panel models and latent growth curve models need not be treated as mutually exclusive choices for modeling longitudinal data. They propose an *autoregressive latent trajectory* (ALT) model (Figure 2.7) that incorporates aspects of both kinds of models. The ALT model includes factors representing aspects of change (intercept, slope, etc., as in Figure 2.6) but also

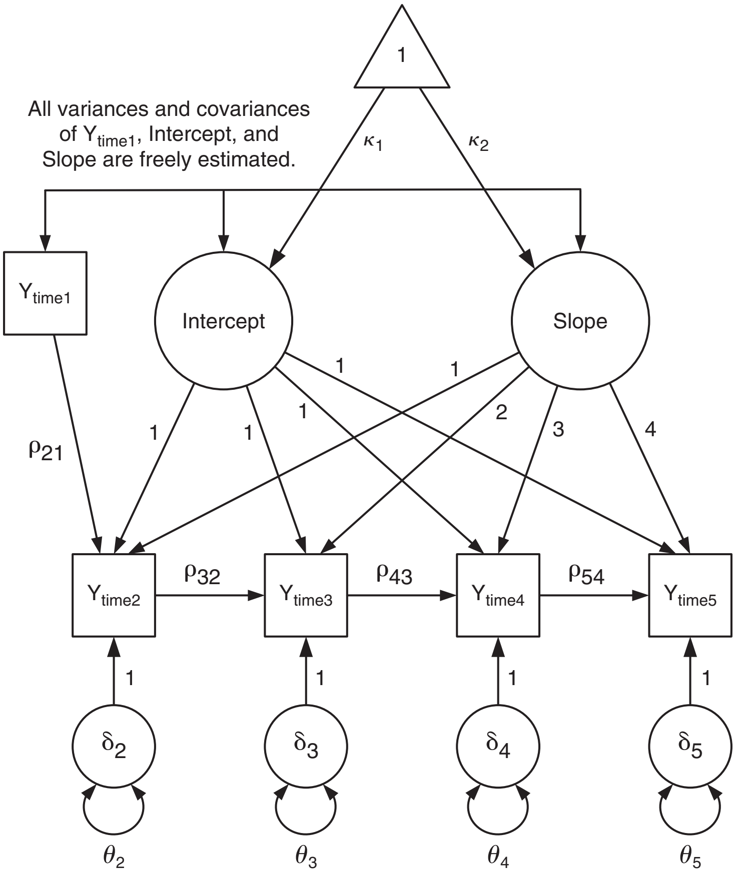


FIGURE 2.7 An autoregressive latent trajectory model with five repeated measurements.

directional paths linking adjacent measurements, as in Figure 2.3. An example of this ALT model is displayed in Figure 2.7.

In the ALT model, y at time t is modeled as a function of (1) an underlying latent trajectory that determines growth over time *and* (2) y at time $t - 1$. Thus, one can think of the ALT model as a special case of LGM with time-varying covariates, where the covariate at each occasion after the first is simply the repeated measure at the previous occasion. One could also think of the panel model and the standard latent growth curve model as nested within the ALT model as special cases. The ρ coefficients are autoregressive parameters that reflect the degree of association between adjacent measurements after controlling for the growth curve. The ρ coefficients are sometimes constrained to equality. Because the panel model treats the first measurement as exogenous to the system, in ALT models the initial measure is also treated as exogenous and permitted to covary with the latent curve factors.

TIME SERIES AND RELATED MODELS

The methods for analyzing longitudinal data discussed thus far are generally implemented in data collection designs involving a large sample of participants and a small but adequate number of repeated observations. Often, research questions involve longitudinal designs where a small number of individuals (perhaps even one) are measured repeatedly over a large number of intervals. In time series analysis, numerous (often 50 or more) observations provide the basis for estimating key parameters depicting the associations of the measures over time (Box & Pierce, 1970). Designs such as this are the ideal longitudinal design for many questions in the social and behavioral sciences (Velicer & Fava, 2003). Tracking key variables at intensive intervals is relatively easy using handheld devices or daily diary data, providing a long series of measurements (Walls & Schafer, 2006). The goal of time series (and related) designs and analyses are to identify time-related patterns in the sequence of numbers where the patterns are correlated, but offset in time.

Manifest Variable Time Series Analysis

The need for specific techniques of time series analysis arises when considering the problems of classical methods, such as multiple regression, in analyzing these intensive repeated data. The primary rationale for use of time series procedures rather than multiple regression analysis is the inherent dependency that results from making repeated observations of the same participant or group of participants, referred to as autocorrelation. Use of multiple regression in the presence of autocorrelation is an explicit violation of the assumption of independence of errors. As a result, Type I error rates would be substantially increased. In addition, false patterns may either obscure or spuriously enhance effects of predictors (e.g., onset of a contextual change) unless the autocorrelation is accounted for in the model.

There are several classes of time series analysis (see Hershberger, Molenaar, & Corneal, 1996). The most common time series application in behavioral research (Velicer & Fava, 2003) is the autoregressive integrated moving average (ARIMA; Box, Jenkins, & Reinsel, 1994) class of models in which the pattern of change in a dependent variable is assessed over time (see also Boker, Deboeck, Edler, & Keel; 2008; Hershberger et al., 1996; van Buuren, 1997). The basic properties of an ARIMA model are characterized by three parameters. The key elements of an ARIMA (p, d, q) time series analysis are the lingering effects of preceding scores called *autoregressive* elements (p), trends in the data called *integrated* elements (d), and the lingering effects of preceding shocks called the *moving average* element (q). All ARIMA models also have random process error terms called *shocks*, but differ in the *order* (how many preceding observations must be considered when determining the dependency) of the p, d, and q parameters.

A time series analysis consists of three steps. The first step is to identify which mathematical model best represents the data, focusing on the autocorrelation function, potential cyclic patterns, autoregressive components (p), and moving-average components (q). Here, the task is to specify values of the p, d, and q parameters that adequately fit the pattern of change over time. One typically begins by identifying d, the integrative element, representing the average pattern of no change in mean ($d = 0$), linear increase or decrease ($d = 1$), or linear and quadratic change ($d = 2$) (or potentially higher order change processes, although these are rare) across the *entire* time span. Once this overall pattern is identified, it is removed (in time series terminology, the series is made *stationary*), and the next two parameters are estimated. These parameters specify whether scores (p) or moving averages (q) are unrelated to previous scores (e.g., $p = 0$), related only to the previous score (i.e., simple autocorrelation, $p = 1$), or related to even earlier scores or shocks (p or $q > 1$). The second step, *estimation*, is to reconfigure the dependent observed variable into a serially independent variable through a transformation appropriate for the identified model. The third step, *diagnosis*, is to estimate the model parameters through generalized least squares and examine the residuals for unaccounted patterns.

Time series analysis can be used to answer several types of research questions. First, time series analysis can be used to evaluate patterns of overall trends and autocorrelation, such as whether there are linear increases or decreases over time, or whether previous scores/shocks impact later scores. Time series analyses can also be used to evaluate potential cycles and trends, therefore identifying potential seasonal or other periodic patterns (e.g., patterns of substance use across a week that are high near weekends and lower during the middle weekdays), and whether longer trends are distinct from (or similar to) "local" patterns. Time series analyses are sometimes used for forecasting, or attempting to predict value of observations in the future; this use is more common in the field of economics, but might be considered in adolescent research (e.g., predicting patterns of school violence across the school year).

Time series analyses can also consider covariates of change (see McDowell, McCleary, Meidinger, & Hay, 1980). When considering presumed causes of change, the field of time series analysis often uses the term *interventions* to reference these predictors of change. Following this tradition, we use here the term *intervention* broadly to refer to any event (or shock) expected to impact the adolescent across time, whether intentionally introduced by the researcher (e.g., social skills training) or naturally occurring (e.g., an adolescent joins a school club). The impact of these interventions or events is evaluated after accounting for other change patterns (e.g., cyclical waves). Similarly, we can evaluate whether this impact is abrupt and permanent (e.g., disasters, parental divorce that permanently affects the adolescent's family environment) or abrupt but temporary (e.g., parental separation that resolves to a stable family environment). It is also possible to evaluate the impact of multiple interventions, assuming there is some separation of these interventions in time (e.g., parental divorce and later transition to a different school). Similarly, one can evaluate the impact of introduction, then removal, of an intervention (e.g., an adolescent joins and then leaves a school club).

Although time series analyses often use data from a single person, they can also incorporate data from multiple participants or compare across individuals or groups to assess the degree of similarity in the patterns for different populations. Several options for combining time series data across participants exist. In *pooled* time series analysis (Hsiao, 1986), all observations for all participants are included in a single vector, and a patterned transformation matrix is utilized to convert the serially dependent variable into a serially independent variable. Another alternative parallels meta-analysis. Here, individual participant time series are combined rather than individual studies. However, the meta-analytic approach is difficult in that there is a lack of statistical time series information in the published literature (many reports still rely on visual analysis) and an appropriate definition of effect size is needed for time series data. Multilevel or mixed modeling, as described in the previous sections, can be considered as a means of utilizing data from multiple participants, where the elements of the time series are nested within individuals, resulting in a two-level hierarchy. A multilevel approach to time series analysis is easily conducted using traditional multilevel software (for example, HLM or SAS PROC MIXED). In SEM packages, specifying a time series model can be more challenging because of the large number of observation occasions (e.g., typically > 50) and the need to model such data in "wide" format.

Multivariate time series models involve measuring multiple variables at each time point for the same individual. A basic approach can be to determine the cross-lagged correlational structure between the multiple variables, where *lag* refers to the time relation between two variables. If one variable can be conceptualized as a dependent variable and the remaining variables can be considered covariates, then a *concomitant variable time series analysis* (Glass, Willson, & Gottman, 1975) can be conducted as a direct analog to the ANCOVA. We discuss the possibilities of multivariate time series models in the next section on latent variable analyses.

Latent Variable Time Series Analysis

Most techniques described as "time series" models are applied to manifest variable relations and often to a single variable. In the early developments of factor analysis, however, p-technique factor analysis was introduced to provide a basis for modeling multivariate time series data as a reduced set of latent factors (Cattell, 1952, 1963, 1988). Simply stated, p-technique factor analysis utilizes factor extraction methods for a set of variables, but rather than factor across multiple individuals, the cases used to generate the analyzed covariance matrix are intensively repeated observations of one or a few individuals. In this regard, the "sample size" used in the analysis is the number of repeated observations, and the actual number of participants may be only a single individual. P-technique factor analysis, as originally conceived, has some known limitations. For example, if there is any serial dependency inherent in the time series, it can result in underestimated factor loadings (i.e., due to positive autocorrelations; Wood & Brown, 1994). When there is little or no dependency in the time series data, p-technique can provide useful factors that can inform important research questions (see Hawley & Little, 2003, for example).

The basic idea of p-technique SEM analysis is that indicators of a construct will ebb and flow over time in a consistent manner such that they will reflect an underlying latent construct that is defined on the basis of the change patterns among the indicators (Nesselroade, McArdle, Aggen, & Meyers, 2001). The relations among multiple constructs can then be assessed and compared on the basis of their cross-time changes with one another. For example, asking a person to rate his or her mood by providing ratings on three indicators of positive affect (e.g., happy, glad, up) and three indicators of negative affect (e.g., sad, down, blue) every day for, say, 100 days would yield a covariance matrix among the indicators that would allow one to test a model in which two constructs, each with three indicators, underlie the data. Such a model would be a simple CFA specifying two constructs, and the appropriateness of the model would be evaluated using standard model fit criteria.

A time-ordered data matrix such as this easily captures contemporaneous relations among the indicator processes, but it does not capture lagged dynamic influences. To address this limitation, dynamic p-technique SEM utilizes the inherent time-ordered information of such a data matrix to analyze a lagged covariance matrix wherein the effects of the constructs at occasion o can be evaluated for their influence on the latent constructs at occasion $o + 1$ (Hawley & Little, 2003). In other words, the data set can be lagged such that concurrent covariation and lagged covariation among a set of variables can be modeled. Because SEM is a covariance structure modeling technique, covariance matrices of this nature can be analyzed in much the same way that a covariance matrix derived from multiple individuals can be analyzed. Both panel-like and growth curve–like models are perfectly suited to model dynamic covariance matrices. Because the dynamic version of p-technique SEM analysis explicitly models any serial dependency in the data (Hershberger, 1998), the dependency issue no longer presents a potentially biasing influence on model parameters.

Figure 2.8 depicts a basic block Toeplitz covariance matrix that is modeled in a dynamic p-technique SEM (see Hawley & Little, 2003; Little, Bovaird, & Slegers, 2006). A lagged covariance matrix contains three distinct structural features. The first feature is the simultaneous or synchronous relations among the variables. In this example, these simultaneous relations among the variables are represented twice, within Lag 0 and again within Lag 1, in the triangles directly below the major diagonal. The variances of the variables are located along the major diagonals and covariances are located off the diagonals. For the most part, the corresponding elements between these two sections would be nearly or exactly identical (see Hawley & Little,

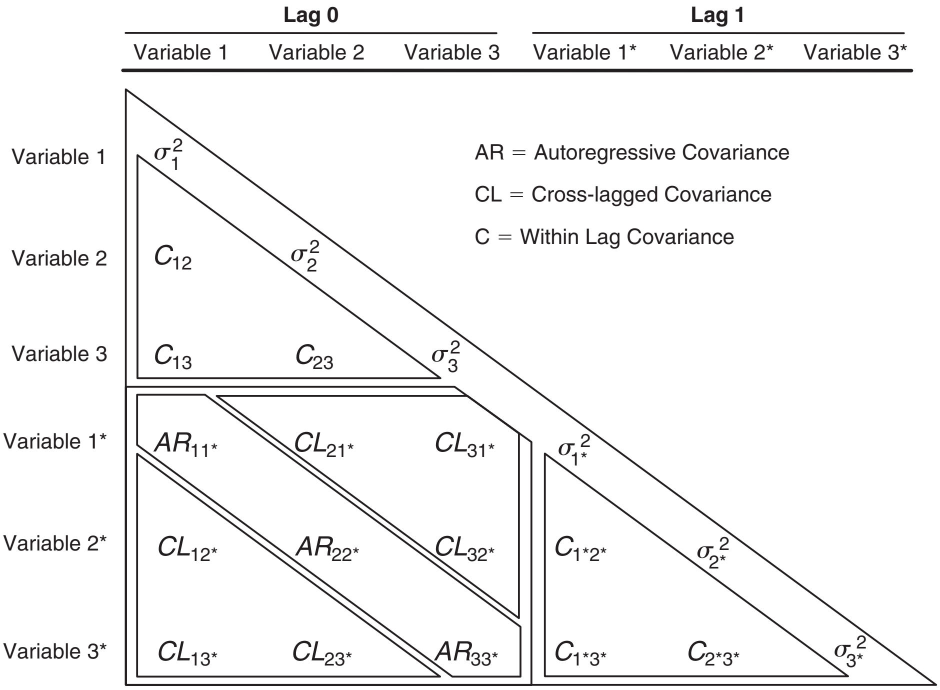


FIGURE 2.8 Toeplitz matrix.

2003). The lower quadrant of a lagged covariance matrix contains the lagged information among the variables, which reflects two sources of lag information. The autoregressive lagged relations between each pair of variables are located on the subdiagonal of this lower quadrant (denoted $AR_{1,1*}$, $AR_{2,2*}$, and $AR_{3,3*}$; see Figure 2.8). This information reflects a variable's correlation with itself between lag 0 and lag 1. Cross-lagged relations among the variables are represented in the upper and lower triangles of this quadrant (e.g., $CL_{1,2*}$, $CL_{1,3*}$, and $CL_{2,3*}$). $CL_{1,2*}$, for example, represents the covariation between variable 1 at lag 0 (V_1) and variable 2 at lag 1 (V_{2*}) (see e.g., Molenaar, 1985, 1994; Molenaar, De Gooijer, & Schmitz, 1992; Wood & Brown, 1994). With dynamic p-technique models the structural relations specified for the variables within lag 0 would be identical for lag 1 because these two parts of the block Toeplitz dynamic matrix are essentially identical. The key feature of dynamic p-technique SEM models therefore is the unique parameters that link the constructs estimated within lag 0 with their lagged counterparts within lag 1. These parameters reflect the dynamic relations among the constructs.

Dynamic p-technique SEM is a very powerful technique to model developmental phenomena (Little, Bovaird & Slegers, 2006). Dynamic p-technique factor analysis has been applied in the domains of mood, personality, and locus of control (see Jones & Nesselroade, 1990, for a review). Unfortunately, the technique has yet to be fully utilized as a potentially powerful research tool in adolescent developmental research. Given that the intensive study of a single individual allows one to model the truly dynamic interplay among a set of constructs (Nesselroade & Molenaar, 1999), this lack of utilization is disappointing. Dynamic p-technique SEM is particularly well suited to examine questions regarding topics such as the person–situation debate and the social–personality nexus (see Fleeson & Jolley, 2006).

Dynamic p-technique SEM has the additional advantages of traditional applications of SEM such as the ability to model change relations as error-free constructs. Moreover, the lagged covariance matrix depicted in Figure 2.8 can be extended to include multiple potential lags.

Although dynamic p-technique SEM models can be fit to a single individual, the broader usefulness of this approach emerges when one compares the resulting dynamic models of change across a sample of individuals. With this approach, the key sample size issue is ensuring sufficient data points for each individual to establish a well-conditioned model for each individual. In such situations, the models can be compared across a relatively small number of individuals to draw conclusions about similarities and differences in the modeled dynamic processes. In this regard, the number of persons needed to make reasonably sound nomothetic generalizations is relatively small. In situations in which the number of observations for a given subject is not sufficiently large to allow estimation of a well-conditioned covariance matrix, data from a number of participants can be chained. In particular, Nesselroade and Molenaar (1999) proposed combining pooled time series data with dynamic factor analysis to overcome the limitation of the number of observations in the time series needed for stable estimation of the population covariance matrices.

In short, dynamic p-technique SEM is particularly useful for intensive repeated measures designs that are geared to understand dynamic change processes. Advantages of the dynamic p-technique include accounting for the autocorrelation among indicators, allowing cross-lagged influences, and correcting for measurement error (when multiple indicators of latent variables are employed). Dynamic p-technique SEM enjoys nearly limitless expandability to be able to incorporate static covariates, time-varying effects, and static outcomes (Little, Bovaird, & Slegers, 2006). With multiple-group capabilities of SEM programs, comparing models across groups of individuals allows nomothetic assessments of the similarities and differences in the dynamic patterns among individuals.

MEDIATION AND MODERATION IN LONGITUDINAL DATA

Questions about mediation and moderation abound, but these terms are often misused or misunderstood. *Mediation* is said to occur when part of the effect of X on Y occurs indirectly through some intermediate variable M. That is, X causes Y because X causes M, which in turn causes Y, or M is one mechanism through which X exerts its effect on Y. A question about mediation, therefore, generally takes the form of "by what means does variable X exert influence on variable Y?" In other words, is there a potential causal chain that links X to Y via some mediating influence M? Here, M is the "delivery agent" or "carrier" that transmits the influence of X to Y.

Moderation, however, occurs when the magnitude or direction of the effect of X on Y depends on some third variable W. A question about moderation generally takes the form of "is the relation between variable X and variable Y impacted by some moderating influence, W?" Here, W is the "changer" of a relation between two (or more) variables; in other words, a question about moderation typically is answered with an "it depends on"–type statement: The relation between X and Y depends on the level of variable W.

Mediation and moderation are often confused, despite repeated attempts to educate researchers on the difference (e.g., Baron & Kenny, 1986; Frazier, Tix, & Barron, 2004). In this section we assume the reader is already familiar with the basic concepts of—and distinctions between—mediation and moderation. A good understanding of the basic concepts involved in mediation can be gained from Frazier et al. (2004); MacKinnon (2008); MacKinnon, Fairchild, and Fritz (2007); MacKinnon, Lockwood, Hoffman, West, and Sheets (2002); Preacher and Hayes (2008b); and Shrout and Bolger (2002). Basic material on modeling moderation (interaction) effects is provided by Aiken and West (1991); Cohen, Cohen, West, and Aiken (2003); Jaccard, Turrisi, and Wan (1990); and Aguinis (2004). Here, we discuss some ways in which mediation and moderation effects may be incorporated into models for longitudinal data. In what follows, we assume that the *indirect effect* (the product of the X → M and M → Y paths) has been adopted to represent mediation (e.g., Dearing & Hamilton, 2006).

Mediation in Longitudinal Settings

The first questions to ask regarding mediation are "What effect is hypothesized to be mediated?" and "By what?" The object of a mediation analysis is to determine by what intermediate steps an effect unfolds (or by what mechanism the effect occurs). Once potential mediators (mechanisms) are identified, attention can be devoted to modeling the effect. Traditional cross-sectional models provide a weak basis for causal inference (and hence inferences of mediation effects) because they allow no time for effects to unfold. Thus, panel models are ideally suited for investigating mediation effects. Gollob and Reichardt (1991) and Cole and Maxwell (2003) strongly advocate using models like that in Figure 2.9 to address hypotheses of mediation. Such models have many advantages relative to other models used to assess mediation. First, the temporal separation necessary for establishing causal effects is considered. Second, because repeated measures of each variable are included, more precise estimates of key path coefficients are obtained. Third, panel models can easily incorporate latent variables to correct for measurement error. Fourth, by controlling for previous measurement of M and Y, only the portions of the variances of M and Y that do not remain stable over time contribute to the estimation of the mediation effect. This serves to reduce bias and paint a more realistic picture of the indirect process by which X effects change in Y via M.

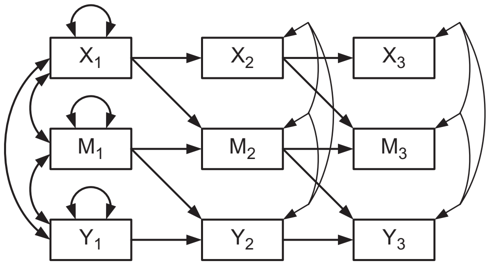


FIGURE 2.9 A longitudinal mediation panel model.

Gollob and Reichardt (1991) argue that there is no single mediation effect characterizing a set of variables X, M, and Y. Rather, a potentially different mediation effect exists for every choice of lag separating the assessments of X and M, and M and Y (see also Cole & Maxwell, 2003; Maxwell & Cole, 2007). Thus, serious attention should be devoted to choosing the optimal lags to separate measurement of key variables involved in the mediation effect. It is important to note that this optimal lag may not correspond to the lag associated with the largest effect. The best lag to use instead may be dictated by theory or context. For example, if the researcher is interested in gauging how (and by what means) an intervention initiated at the beginning of the school year affects grades at the end of the school year, the beginning and end points of the study are predetermined. It remains only to choose the appropriate occasion (or preferably occasions) to measure the mediator. Detailed instructions for how to estimate and test mediation effects in panel models, as well as other important design and modeling issues to consider, can be found in Cole and Maxwell (2003), Maxwell and Cole (2007), and Little et al. (2007).

Moderation in Longitudinal Settings

As with mediation, the first questions to ask with regard to moderation are "What effect is hypothesized to be moderated?" and "By what?" Once these questions are answered, the appropriate modeling strategy follows naturally. In the context of panel models, for example, the lagged effect of X_1 on Y_2 may depend on W_1 (in order for W to influence this lagged effect, W necessarily must be measured before Y_2, hence the "1" subscript), in which case the researcher may consider using W_1 and the product of X_1 and W_1 as additional predictors of Y_2. There are numerous ways in which moderation can be examined with longitudinal data. We focus on two examples to give readers an idea of the many possibilities.

Moderated Autoregressive Effects

Consider first a (univariate) time series consisting of four equally spaced repeated measurements of affiliation. The simplex model depicted in Figure 2.10 could be applied to these data. Coefficients a_1, b_1, and c_1 may or may not be constrained to equality, depending on how stationary the researcher believes the process to be.

One way to incorporate a moderation effect, assuming it is justified by theory, would be to hypothesize that the autoregressive weights a_1, b_1, and c_1 are moderated by self-monitoring (SM), which might be assumed to be a trait characteristic and so measured only once, at Time 1. Moderation by SM may be incorporated by computing the product SM × affiliation at Time $j - 1$ and including this interaction term as a predictor at Time $j - 1$, as in Figure 2.11. The significance of the a_3, b_3, and c_3 coefficients can be used as a basis for deciding whether the autoregressive effect of affiliation at Time $j - 1$ on affiliation at Time j is moderated by SM.

As with many interaction effects, if the interaction is found to be significant, the researcher

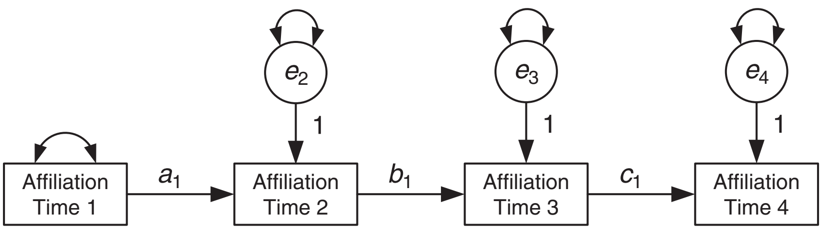


FIGURE 2.10 A simplex model for four repeated measurements.

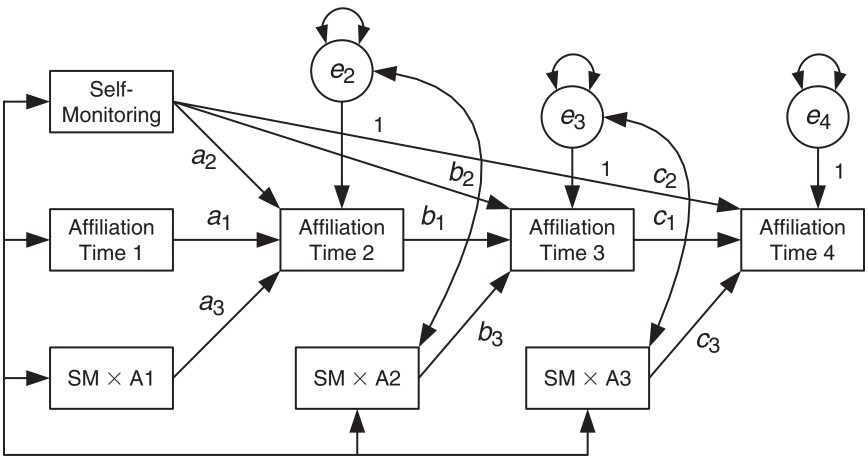


FIGURE 2.11 A moderated simplex model in which the autoregressive effect depends on the moderator self-monitoring.

may wish to explore the effect further by (1) plotting the interaction effect for various interesting conditional values of SM (say, ±1 SD and the mean) and (2) testing the *simple slope* of affiliation at Time j regressed on affiliation at Time $j-1$ (i.e., the *simple autoregression*) for significance at conditional values of SM. Methods for accomplishing this are exactly analogous to well-known methods for plotting and probing significant interaction effects in ordinary regresssion (see Aiken & West, 1991; Preacher & Rucker, 2005; Preacher, Curran, & Bauer, 2006). For example, if we wished to test the significance of the autoregression of affiliation at Time 3 on affiliation at Time 2 when SM = 3, the simple autoregression is:

$$\hat{\omega} = \hat{b}_1 + \hat{b}_3(3) \tag{18}$$

and the standard error for this estimate is:

$$se_{\hat{\omega}} = \sqrt{\hat{s}_1^2 + 2(3)\hat{s}_{31} + (3^2)\hat{s}_3^2} \tag{19}$$

where $\hat{s}_1^2$, $\hat{s}_3^2$, and $\hat{s}_{31}$ are the asymptotic variances and covariance of the parameter estimates $\hat{b}_1$ and $\hat{b}_3$. The critical ratio $\hat{\omega}/se_{\hat{\omega}}$ can be compared to the standard normal distribution (assuming a large sample size) to determine whether the autoregression is significant when SM = 3. All of the requisite coefficient estimates can be found in standard SEM output, and the asymptotic (co)variances are provided by special request in most SEM programs.

Predictors of Slopes in MLM and LGM

Moderation effects can also be modeled as part of the growth curve models discussed earlier, although these effects may not be immediately recognizable as moderation. Recall that any variable that affects the relation between two other variables can be considered a moderator. In LGM, the slope factor represents the relation between time and the outcome variable (this latent variable itself may or may not vary across people). Thus, including predictors of the slope factor amounts to including moderators of the effect of time. See Curran, Bauer, and Willoughby (2004) and Preacher et al. (2006) for more detailed discussions and worked examples. By the same token, including level 2 predictors of slopes in MLM qualifies as moderation (Bauer & Curran, 2005). In both the LGM and MLM contexts, if predictors of slopes are included, the same variables must also be included as predictors of intercepts.

Extensions

Clearly, the concepts of mediation and moderation have many potential applications in models for longitudinal data. We illustrated a few here, but the precise way in which mediators or moderators are included in specific applications will depend on the research context. Some studies may require multiple mediating variables (MacKinnon, 2000; Preacher & Hayes, 2008a). Multiple-group analysis can be construed as a kind of moderation analysis, in which some model parameters can vary according to the moderating group variable. Polynomial growth functions (e.g., quadratic trajectories) can be considered moderation effects, in which one variable (e.g., time) moderates its own linear influence on the outcome. Interactions among three or more variables may be entertained.

Mediation and moderation may even be combined into one model. For example, a given mediation effect may be hypothesized to vary across levels of a moderator (*moderated mediation*; see Bauer, Preacher, & Gil, 2006; Edwards & Lambert, 2007; Preacher, Rucker, & Hayes, 2007). Or a moderation effect may be hypothesized to be mediated (*mediated moderation*). Latent moderation effects, in which either the independent variable or the moderator is latent (or both) are beyond the scope of this chapter, but interested readers can consult Schumacker and Marcoulides (1998) and Little, Bovaird, and Widaman (2006) to learn more.

CONCLUSIONS

In this chapter, we have provided a broad overview of issues and techniques in modeling longitudinal data during adolescence. As we have indicated throughout, a number of detailed resources are available to delve more deeply into these issues and techniques. We close by discussing some opportunities and future directions for longitudinal data analysis.

Opportunities

As should be clear, longitudinal data can provide extremely rich information to inform our understanding of developmental mechanisms and processes. Of course, issues of causality are more challenging in the context of nonexperimental methodology, but when coupled with a broad-based program of research that includes experimentation and rigorous inferential designs, the basis for causal inference can be significantly strengthened. In fact, each of the techniques and issues that we have highlighted herein can be appropriately coupled with rigorous experimental designs such as randomized intervention trials and regression discontinuity designs (Greenwood & Little, 2008). In situations in which it is not possible or ethical to manipulate the presumed cause, longitudinal studies without experimental manipulation often represent our best approaches to evaluating presumed causal relations (or at least the temporal primacy of constructs; Card & Little, 2007).

We emphasize again that the opportunities afforded by longitudinal research will not be realized if researchers do not attend carefully to the critical design and measurement issues we outlined above. Moreover, the statistical model that one chooses to employ must match the theoretical model in order to capitalize on the strong inferential capabilities of modern analytic procedures. Quite often, researchers will need to utilize more than one statistical approach to fully exploit the available information contained in longitudinal data. Panel analyses, for example, can inform the direct and indirect pathways by which antecedent variables influence consequent variables, while multivariate growth curve models can reveal different information such as the strength of the dynamic relations in the changes over time.

Future Directions

In terms of future directions, we highlight some that are particularly promising for use in the study of adolescent development. First, Bayesian estimation methods have emerged as useful tools for estimating a variety of models, including complex growth curve models that can be difficult or impossible to estimate using current maximum likelihood estimation (MLE)-based software. Bayesian methods for analyzing longitudinal data in the study of adolescent development are particularly useful because of their ability to incorporate prior information in estimating both simple and complex models. In addition to being an alternative to the MLE method, Bayesian methods also have unique strengths, such as systematically incorporating prior information from previous studies. Bayesian methods are also particularly well-suited to analyze small sample data (Zhang, Hamagami, Wang, Grimm, & Nesselroade, 2007; Wang & McArdle, 2008).

A second important direction in the development of tools for the analysis of longitudinal data is in the use of bootstrap estimation of sampling distributions of parameter estimates (and corresponding confidence intervals). Bootstrap estimation has been around for some time but is only recently receiving the attention that the approach deserves. Although various methods of bootstrapping exist, the basic idea is that one resamples from the original sample a very large number of replicate samples (e.g., 2,000 to 5,000 resamples). These resamples have the same sample size as the original sample, but the key is that the resamples are drawn from the original sample *with replacement.* Because the sampling is done with replacement a given observational record might occur many times in a given resample and not occur at all in another resample.

The inherent variability across the large number of resamples provides an empirical basis to estimate the confidence intervals for each estimated parameter of a given model. The model is run using each of the large number of replicate samples and the estimated parameters are stored. The estimated parameters from the original sample provide the point estimate for a given analysis. The values demarcating the lower and upper 2.5% of the distribution of estimated parameters from the bootstrapped samples provides an empirical calculation of the 95% confidence interval of each model parameter. A nice feature of such empirically derived distributions is that they do not need to be symmetric and the inherent degree of sampling variability is fully captured. When the collected data do not meet the assumptions of the analytic procedure (e.g., normality), the quality of generalizations made from bootstrapped estimation of confidence intervals often are significantly better than those derived from the theoretical distributions used in other estimation methods such as maximum likelihood and least squares estimation.

Final Thoughts

Given the lack of training opportunities found in many universities, students and researchers often must rely on self-guided learning. A chapter such as this cannot substitute for in-depth training, but it can provide a guide to the directions and issues that one should cover. In this chapter we provided a broad overview of issues in, and techniques for, longitudinal data analysis. Because a handbook chapter such as this cannot provide the details or address the nuances of any of the topics we have presented, we have attempted to provide a wealth of readings and resources to help readers find the additional information needed. We have also highlighted a number of web-based resources (e.g., www.Quant.KU.edu) that provide updated guidance and information relevant to the techniques we have described here. Because these techniques continue to develop and become more capable as well as more refined, we encourage all researchers to focus concerted efforts to stay informed with the state-of-the-science in quantitative methodology.

REFERENCES

Aguinis, H. (2004). *Regression analysis for categorical moderators*. New York: Guilford Press.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage.

Anderson, J. G. (1987). Structural equation models in the social and behavioral sciences: Model building. *Child Development*, *58*, 49–64.

Baltes, P. B. (1968). Longitudinal and cross-sectional sequences in the study of age and generation effects. *Human Development*, *11*, 145–171.

Barker, E. D., Tremblay, R. E., Nagin, D. S., Vitaro, F., & Lacourse, E. (2006). Development of male proactive and reactive physical aggression during adolescence. *Journal of Child Psychology and Psychiatry*, *47*, 783–790.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.

Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics*, *28*, 135–167.

Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research*, *42*, 757–786.

Bauer, D. J., & Curran, P. J. (2003a). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, *8*, 338–363.

Bauer, D. J., & Curran, P. J. (2003b). Over-extracting latent trajectory classes: Much ado about nothing? Reply to Rindskopf (2003), Muthén (2003), and Cudeck and Henly (2003). *Psychological Methods*, *8*, 384–393.

Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods*, *9*(1), 3–29.

Bauer, D. J., & Curran, P. J. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral Research*, *40*, 373–400.

Bauer, D. J., & Shanahan, M. J. (2007). Modeling complex interactions: Person-centered and variable-centered approaches. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 255–284). Mahwah, NJ: Lawrence Erlbaum.

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods*, *11*, 142–163.

Bell, R. Q. (1953). Convergence: An accelerated longitudinal approach. *Child Development*, *24*, 145–152.

Bickel, R. (2007). *Multilevel analysis for applied research: It's just regression!* New York: Guilford Press.

Bijleveld, C. C. J. H., & van der Kamp, L. J. T. (1998). *Longitudinal data analysis: Designs, models and methods*. Thousand Oaks, CA: Sage.

Blozis, S. A., Conger, K. J., & Harring, J. R. (2007). Nonlinear latent curve models for multivariate longitudinal data. *International Journal of Behavioral Development*, *31*, 340–346.

Boker, S. M., Deboeck, P. R., Edler, C., & Keel, P. K. (2008). Generalized local linear approximation of derivatives from time series. In S. Chow, E. Ferrer, & F. Hsieh (Eds.), *Statistical* *methods for modeling human dynamics: An interdisciplinary dialogue*. Boca Raton, FL: Taylor & Francis.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons.

Bollen, K. A., & Curran, P. J. (2004). Autoregressive latent trajectory (ALT) models: A synthesis of two traditions. *Sociological Methods and Research*, *32*, 336–383.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Hoboken, NJ: John Wiley & Sons.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time series analysis: Forecasting and control*. (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Box, G. E. P., & Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, *65*, 1509–1526.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Browne, M. W., & Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research*, *24*, 445–455.

Byrne, B. M., & Crombie, G. (2003). Modeling and testing change: An introduction to the latent growth curve model. *Understanding Statistics*, *2*, 177–203.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456–466.

Cairns, R. B., Bergman, L. R., & Kagan, J. (Eds.) (1998). *Methods and models for studying the individual*. Thousand Oaks, CA: Sage.

Card, N. A. & Little, T. D. (2007). Longitudinal modeling of developmental processes. *International Journal of Behavioral Development*, *31*, 297–302.

Card, N. A., & Little, T. D. (2006). Analytic considerations in cross-cultural research on peer relations. In X. Chen, D. C. French, & B. Schneider (Eds.), *Peer relations in cultural context* (pp. 75–95). New York: Cambridge University Press.

Card, N. A., Little, T. D., & Bovaird, J. A. (2007). Modeling ecological and contextual effects in longitudinal studies of human development. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 1–11). Mahwah, NJ: Lawrence Erlbaum.

Cattell, R. B. (1952). The three basic factor-analytic research designs—their interrelations and derivatives. *Psychological Bulletin*, *49*, 499–551.

Cattell, R. B. (1963). The structuring of change by P- and incremental R- technique. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 1671–198). Madison: University of Wisconsin.

Cattell, R. B. (1988). The data box. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 69–130). New York: Plenum.

Chan, D. (1998). The conceptualization and analysis of change over time: An integrative approach incorporating longitudinal mean and covariance structures analysis (LMACS) and multiple indicator latent growth modeling (MLGM). *Organizational Research Methods*, *1*, 421–483.

Chou, C.-P., Bentler, P. M., & Pentz, M. A. (1998). Comparisons of two statistical approaches to study growth curves: The multilevel model and the latent curve analysis. *Structural Equation Modeling*, *5*, 247–266.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology*, *112*, 558–577.

Collins, L. M. (2006). Analysis of longitudinal data: The integration of theoretical models, temporal design, and statistical model. *Annual Review of Psychology*, *57*, 505–528.

Collins, L. M., & Horn, J. L. (Eds.) (1991). *Best methods for the analysis of change: Recent advances, unanswered questions, future directions*. Washington, DC: American Psychological Association.

Collins, L. M., & Sayer, A. G. (Eds.) (2001). *New methods for the analysis of change*. Washington, DC: American Psychological Association.

Curran, P. J. (2000). A latent curve framework for the study of developmental trajectories in adolescent substance use. In J. S. Rose, L. Chassin, C. C. Presson, & S. J. Sherman (Eds.), *Multivariate applications in substance use research*. Mahwah, NJ: Lawrence Erlbaum.

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research*, *38*, 529–569.

Curran, P. J., Bauer, D. J., & Willoughby, M. T. (2004). Testing main effects and interactions in latent curve analysis. *Psychological Methods*, *9*, 220–237.

Curran, P. J., & Bollen, K. A. (2001). The best of both worlds: Combining autoregressive and latent curve models. In L. M. Collins, & A. G. Sayer (Eds.), *New methods for the analysis of change*, (pp. 105–136). Washington, DC: American Psychological Association.

Curran, P. J., & Hussong, A. M. (2003). The use of latent trajectory models in psychopathology research. *Journal of Abnormal Psychology*, *112*, 526–544.

Dearing, E., & Hamilton, L. C. (2006). Contemporary advances and classic advice for analyzing mediating and moderating variables. *Monographs of the Society for Research in Child Development*, *71*, 88–104.

Duncan, O. D. (1969). Some linear models for two-wave, two-variable panel analysis. *Psychological Bulletin*, *72*, 177–182.

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2006). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Dwyer, J. H. (1983). *Statistical models for the social and behavioral sciences*. New York: Oxford University Press.

Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: A general analytical framework using moderated path analysis. *Psychological Methods*, *12*, 1–22.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, *7*, 1–26.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

Embretson, S. E. (2007). Impact of measurement scale in modeling development processes and ecological factors. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 63–87). Mahwah, NJ: Lawrence Erlbaum.

Enders, C. K. (in press). *Applied missing data analysis*. New York: Guilford Press.

Fleeson, W., & Jolley, S. (2006). A proposed theory of the adult development of intraindividual variability in trait-manifesting behavior. In D. K. Mroczek & T. D. Little (Eds.), *Handbook of personality development* (pp. 41–60). Mahwah, NJ: Lawrence Erlbaum.

Frazier, P. A., Tix, A. P., & Barron, K. E. (2004). Testing moderator and mediator effects in counseling psychology research, 115–134. *Journal of Counseling Psychology*, *51*, 157.

Glass, G. V., Willson, V. L., & Gottman, J. M. (1975). *Design and analysis of time-series experiments*. Boulder, CO: Colorado Associated University Press.

Gollob, H. F., & Reichardt, C. S. (1987). Taking account of time lags in causal models. *Child Development*, *58*, 80–92.

Gollob, H. F., & Reichardt, C. S. (1991). Interpreting and estimating indirect effects assuming time lags really matter. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions*. Washington, DC: American Psychological Association.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, *8*, 206–213.

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods*, *11*, 323–343.

Greenwood, C. R., & Little, T. D. (2008). *Use of regression discontinuity designs in special education research*. Paper commissioned as one in a series of invited NCSER, IES papers devoted to special education research methodology topics Hyattsville, MD: Optimal Solutions Group.

Hancock, G. R., & Lawrence, F. R. (2006). Using latent growth models to evaluate longitudinal change. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course*. Greenwich, CT: Information Age.

Hawley, P. H., & Little, T. D. (2003). Modeling intraindividual variability and change in biobehavioral developmental processes. In B. Pugesek, A. Tomer, & A. von Eye (Eds.), *Structural equations modeling: Applications in ecological and evolutionary biology research* (pp. 143–170). Cambridge, UK: Cambridge University Press.

Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis*. Hoboken, NJ: John Wiley & Sons.

Hershberger, S. L. (1998). Dynamic factor analysis. In G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 217–249). Mahwah, NJ: Lawrence Erlbaum.

Hershberger, S. L., Molenaar, P. C. M., & Corneal, S. E. (1996). A hierarchy of univariate and multivariate structural time series models. In G. A. Marcoulides & R. E. Schumaker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 159–194). Mawah, NJ: Lawrence Earlbaum.

Hofmann, D. A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management*, *23*, 723–744.

Honaker, J., King, G., & Blackwell, M. (2008). *Amelia II: A program for missing data*. Accessed on 10/25/2008 at http://gking.harvard.edu/amelia/.

Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum.

Hsiao, C. (1986). *Analysis of panel data*. Cambridge University Press.

Jaccard, J., Turrisi, R., & Wan, C. K. (1990). *Interaction effects in multiple regression.* Newberry Park, CA: Sage.

Jones, B., Nagin, D., & Roeder, K. (2001). A SAS procedure based on mixture models for estimating developmental trajectories. *Sociological Methodology Research*, *29*, 374–393.

Jones, J. C., & Nesselroade, J. R. (1990). Multivariate, replicated, single-subject, repeated measures designs and P-technique factor analysis: A review of intraindividual change studies. *Experimental Aging Research*, *16*, 171–183.

Kenny, D. A. (1979). *Correlation and causality*. New York: John Wiley & Sons.

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*, 196–217.

Kessler, R. C., & Greenberg, D. F. (1981). Linear panel analysis: *Models of quantitative change*. New York: Academic Press.

Kreft, I. G. G., & de Leeuw, J. (1998). *Introducing multilevel modeling*. London: Sage.

Little, R. J. A., & Rubin D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York: John Wiley & Sons.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, *32*, 53–76.

Little, T. D., Bovaird, J. A., & Card, N. A. (Eds.) (2007). *Modeling contextual effects in longitudinal studies*. Mahwah, NJ: Lawrence Erlbaum.

Little, T. D., Bovaird, J. A., & Slegers, D. W. (2006). Methods for the analysis of change. In D. K. Mroczek & T. D. Little (Eds.) *Handbook of personality development* (pp. 181–211). Mahwah, NJ: Lawrence Erlbaum.

Little, T. D., Bovaird, J. A., & Widaman, K. F. (2006). On the merits of orthogonalizing powered and product terms: Implications for modeling interactions among latent variables. *Structural Equation Modeling*, *13*, 497–519.

Little, T. D., Card, N. A., Bovaird, J. A., Preacher, K. J., & Crandall, C. S. (2007). Structural equation modeling of mediation and moderation with contextual factors. In T. D. Little, J. A., Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 207–230). Mahwah, NJ: Lawrence Erlbaum.

Little, T. D., Card, N. A., Slegers, D. W. & Ledford, E. C. (2007). Representing contextual effects in multiple-group MACS models. In T. D. Little, J. A., Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 121–147). Mahwah, NJ: Lawrence Erlbaum.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, *9*, 151–173.

Little, T. D., Lindenberger, U., & Maier, H. (2000). Selectivity and generalizability in longitudinal research: On the effects of continuers and dropouts. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 187–200). Mahwah, NJ: Lawrence Erlbaum.

Little, T. D., Lindenberger, U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When "good" indicators are bad and "bad" indicators are good. *Psychological Methods*, *4*, 192–211.

Little, T. D., Preacher, K. J., Selig, J. P., & Card, N. A. (2007). New developments in latent variable panel analysis of longitudinal data. *International Journal of Behavioral Development*, *31*, 357–365.

Little, T. D., Schnabel, K. U., & Baumert, J. (Eds.) (2000). *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples*. Mahwah, NJ: Lawrence Erlbaum.

Little, T. D., Slegers, D. W., & Card, N. A. (2006). A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. *Structural Equation Modeling*, *13*, 59–72.

Luke, D. A. (2004). *Multilevel modeling*. Thousand Oaks, CA: Sage.

MacCallum, R. C., Kim, C., Malarkey, W. B., & Kiecolt-Glaser, J. K. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research*, *32*, 215–253.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*, 19–40.

MacKinnon, D. P. (2000). Contrasts in multiple mediator models. In J. S. Rose, L. Chassm, C. C. Presson, & S. J. Sherman (Eds.), *Multivariate applications in substance use research: New methods for new questions* (pp. 141-160). Mahwah, NJ: Lawrence Erlbaum.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York: Lawrence Erlbaum.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology*, *58*, 593–614.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test the significance of the mediated effect. *Psychological Methods*, *7*, 83–104.

Magnusson, D., Bergman, L. R., Rudinger, G., & Törestad, B. (Eds.) (1991). *Problems and methods in longitudinal research: Stability and change*. New York: Cambridge University Press.

Marsh, H. W., Wen, Z., Hau, K-T., Little, T. D., Bovaird, J. A., & Widaman, K. F. (2007). Unconstrained structural equation models of latent interactions: Contrasting residual- and mean-centered approaches. *Structural Equation Modeling*, *14*, 570–580.

Maxwell, S. E., & Cole, D. A. (2007). Bias in cross sectional analysis of longitudinal mediation. *Psychological Methods*, *12*, 23–44.

McDowell, D., McCleary, R., Meidinger, E. E., & Hay, R. A. (1980). *Interrupted time series analysis*. Thousand Oaks, CA: Sage.

Menard, S. W. (1991). *Longitudinal research*. Newbury Park, CA: Sage.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525–543.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107–122.

Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series data. *Psychometrika*, *57*, 333–349.

Molenaar, P. C. M. (1994). Dynamic latent variable models in developmental psychology. In A.von Eye & C. C.Clogg (Eds.), *Latent variables analysis* (pp.155–180). Thousand Oaks, CA: Sage.

Molenaar, P. C. M., De Gooijer, J.G., & Schmitz, B. (1992). Dynamic factor analysis of nonstationary multivariate time series. *Psychometrika*, *57*, 333–349.

Moskowitz, D. S., & Hershberger, S. L. (Eds.). (2002). *Modeling intraindividual variability with repeated measures data: Methods and applications*. Mahwah, NJ: Lawrence Erlbaum.

Muthén, B. O. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumaker (Eds.), *New developments and techniques in structural equation modeling* (pp. 1–33). Mahwah, NJ: Lawrence Erlbaum.

Muthén, B. O. (2008). Latent variable hybrids: Overview of old and new models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models*. Charlotte, NC: Information Age.

Muthén, B. O., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*, *24*, 882–891.

Muthén, B. O., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, *55*, 463–469.

Muthén, L. K., & Muthén, B. O. (2007). *Mplus user's guide* (version 5). Los Angeles: Muthén & Muthén.

Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods*, *4*, 139–157.

Nagin, D. S. (2005). *Group-based modeling of development*. Cambridge, MA: Harvard University Press.

Nagin, D. S., & Tremblay, R. E. (1999). Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Development*, *70*, 1181–1196.

Nesselroade, J. R., & Baltes, P. B. (Eds.). (1979). *Longitudinal research in the study of behavior and development*. New York: Academic Press.

Nesselroade, J. R., McArdle, J. J., Aggen, S. H., & Meyers, J. (2001). Dynamic factor analysis models for representing process in multivariate time-series. In D. M. Moskowitz & S. L. Hershberger (Ed.), *Modeling intraindividual variability with repeated measures data: Methods and applications* (pp. 235–265). Mahwah, NJ: Lawrence Erlbaum.

Nesselroade, J. R., & Molenaar, P. C. M. (1999). Pooling lagged covariance structures based on short, multivariate time series for dynamic factor analysis. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research*. (pp. 223–250). Thousand Oaks, CA: Sage.

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, *31*, 437–448.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*, 879–891.

Preacher, K. J., & Hayes, A. F. (2008). Contemporary approaches to assessing mediation in communication research. In A. F. Hayes, M. D. Slater, & L. B. Snyder (Eds.), *The Sage sourcebook of advanced data analysis methods for communication research* (pp. 13–54). Thousand Oaks, CA: Sage.

Preacher, K. J., & Rucker, D. D. (2005, January). *Probing significant interaction effects in path analysis*. Poster presentation, the annual meeting of the Society for Personality and Social Psychology, New Orleans, LA.

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*, *42*, 185–227.

Preacher, K. J., Wichman, A. L., MacCallum, R. C., & Briggs, N. E. (2008). *Latent growth curve modeling*. Thousand Oaks, CA: Sage.

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

Rao, C. R. (1958). Some statistical models for comparison of growth curves. *Biometrics*, *14*, 1–17.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, *114*, 552–566.

Rindskopf, F. (1984). Using phantom and imaginary latent variables to parameterize constraints in linear structural models. *Psychometrika*, *49*, 37–47.

Rosel, J., & Plewis, I. (2008). Longitudinal data analysis with structural equations. *Methodology*, *4*, 37–50.

Rovine, M. J., & Molenaar, P. C. M. (2000). A structural modeling approach to a multilevel random coefficients model. *Multivariate Behavioral Research*, *35*, 51–88.

Rubin, K. H., Bukowski, W. M., Parker, & J. G. (2006). Peer interactions, relationships, and groups. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology*, Vol. *3*: *Social, emotional, and personality development* (6th ed.; pp. 571–645). Hoboken, NJ: John Wiley & Sons.

Saldaña, J. (2003). *Longitudinal qualitative research: Analyzing change through time*. Walnut Creek, CA: Alta Mira.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman & Hall.

Schafer, J. L., & Graham, J. W. (2002) Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147–177.

Schaie, K. W. (1965). A general model for the study of developmental problems. *Psychological Bulletin*, *64*, 92–107.

Schumacker, R. E., & Marcoulides, G. A. (1998). *Interaction and nonlinear effects in structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum.

Sears, R. R. (1947). Child psychology. In W. Dennis (Ed.), *Current trends in psychology* (pp. 50–74). Pittsburgh, PA: University of Pittsburgh Press.

Selig, J. P., Card, N. A., & Little, T. D. (2008). Latent variable structural equation modeling in cross-cultural research: Multigroup and multilevel approaches. In F. J. R. van de Vijver, D. A. van Hemert, & Y. H. Poortinga (Eds.), *Individuals and Cultures in Multilevel Analysis*. (pp. 93–119). Mahwah, NJ: Lawrence Earlbaum.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton, Mifflin and Company.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, *7*, 422–445.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.

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

Tofighi, D., & Enders, C. K. (2007). Identifying the correct number of classes in a growth mixture model. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 317–341). Greenwich, CT: Information Age.

Tucker, L. (1966). Learning theory and multivariate experiment: Illustration by determination of parameters of generalized learning curves. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology* (pp. 476–501). Chicago: Rand McNally.

van Buuren, S. (1997). Fitting ARMA time series by structural equation models. *Psychometrika*, *62*, 215–236.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4–70.

Velicer, W. F., & Fava, J. L. (2003). Time series analysis. In W. F. Velicer & J. A. Schinka (Ed.), *Handbook of psychology: Research methods in psychology*, vol. *2*; (pp. 581–606). New York: John Wiley & Sons.

Walls, T. A., & Schafer, J. L. (Eds.) (2006). *Models for intensive longitudinal data*. New York: Oxford University Press.

Wang, L., & McArdle, J. J. (2008). A simulation study comparison of Bayesian estimation with conventional methods for estimating unknown change points. *Structural Equation Modeling*, *15*, 52–74.

Widaman, K. F. (2000). *T*esting cross-group and cross-time constraints on parameters using the general linear model. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 163–186). Mahwah, NJ: Lawrence Erlbaum.

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.

Widaman, K. F., Little, T. D., Preacher, K. J., & Sawalani, G. (in press). Creating and using short forms in archival data. In B. Donnellen & K. Trzesniewski (Eds.), *Archival data analysis*. Washington, DC: American Psychological Association.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC: American Psychological Association.

Willett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin*, *116*, 363–381.

Wohlwill, J. F. (1973). *The study of behavioral development*. New York: Academic Press.

Wood, P., & Brown, D. (1994). The study of intraindividual differences by means of dynamic factor models: Rationale, implementation, and interpretation. *Psychological Bulletin*, *116*, 166–186.

Zehr, J. L., Culbert, K. M., Sisk, C. L., & Klump, K. L. (2007). An association of early puberty with disordered eating and anxiety in a population of undergraduate women and men. *Hormones and Behavior*, *52*, 427–435.

Zhang, Z., Hamagami, F., Wang, L., Grimm, K. J., & Nesselroade, J. R. (2007). Bayesian analysis of longitudinal data using growth curve models. *International Journal of Behavioral Development*, *31*, 347–383.