A First Course *in* Structural Equation Modeling

Second Edition

Tenko Raykov
Michigan State University

and

George A. Marcoulides
California State University, Fullerton

LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
2006 Mahwah, New Jersey London

Contents

CHAPTER ONE

Fundamentals of Structural Equation Modeling

WHAT IS STRUCTURAL EQUATION MODELING?

Structural equation modeling (SEM) is a statistical methodology used by social, behavioral, and educational scientists as well as biologists, economists, marketing, and medical researchers. One reason for its pervasive use in many scientific fields is that SEM provides researchers with a comprehensive method for the quantification and testing of substantive theories. Other major characteristics of structural equation models are that they explicitly take into account measurement error that is ubiquitous in most disciplines, and typically contain latent variables.

Latent variables are theoretical or hypothetical constructs of major importance in many sciences, or alternatively can be viewed as variables that do not have observed realizations in a sample from a focused population. Hence, latent are such variables for which there are no available observations in a given study. Typically, there is no direct operational method for measuring a latent variable or a precise method for its evaluation. Nevertheless, manifestations of a latent construct can be observed by recording or measuring specific features of the behavior of studied subjects in a particular environment and/or situation. Measurement of behavior is usually carried out using pertinent instrumentation, for example tests, scales, self-reports, inventories, or questionnaires. Once studied constructs have been assessed, SEM can be used to quantify and test plausibility of hypothetical assertions about potential interrelationships among the constructs as well as their relationships to measures assessing them. Due to the mathematical complexities of estimating and testing these relationships and assertions,

computer software is a must in applications of SEM. To date, numerous programs are available for conducting SEM analyses. Software such as AMOS (Arbuckle & Wothke, 1999), EQS (Bentler, 2004), LISREL (Jöreskog & Sörbom, 1993a, 1993b, 1993c, 1999), M*plus* (Muthén & Muthén, 2004), SAS PROC CALIS (SAS Institute, 1989), SEPATH (Statistica, 1998), and RAMONA (Browne & Mels, 2005) are likely to contribute in the coming years to yet a further increase in applications of this methodology. Although these programs have somewhat similar capabilities, LISREL and EQS seem to have historically dominated the field for a number of years (Marsh, Balla, & Hau, 1996); in addition, more recently M*plus* has substantially gained in popularity among social, behavioral, and educational researchers. For this reason, and because it would be impossible to cover every program in reasonable detail in an introductory text, examples in this book are illustrated using only the LISREL, EQS, and M*plus* software.

The term *structural equation modeling* is used throughout this text as a generic notion referring to various types of commonly encountered models. The following are some characteristics of structural equation models.

1. The models are usually conceived in terms of not directly measurable, and possibly not (very) well-defined, theoretical or hypothetical constructs. For example, anxiety, attitudes, goals, intelligence, motivation, personality, reading and writing abilities, aggression, and socioeconomic status can be considered representative of such constructs.
2. The models usually take into account potential errors of measurement in all observed variables, in particular in the independent (predictor, explanatory) variables. This is achieved by including an error term for each fallible measure, whether it is an explanatory or predicted variable. The variances of the error terms are, in general, parameters that are estimated when a model is fit to data. Tests of hypotheses about them can also be carried out when they represent substantively meaningful assertions about error variables or their relationships to other parameters.
3. The models are usually fit to matrices of interrelationship indices—that is, covariance or correlation matrices—between all pairs of observed variables, and sometimes also to variable means.[1]

[1]It can be shown that the fit function minimized with the maximum likelihood (ML) method used in a large part of current applications of SEM, is based on the likelihood function of the raw data (e.g., Bollen, 1989; see also section "Rules for Determining Model Parameters"). Hence, with multinormality, a structural equation model can be considered indirectly fitted to the raw data as well, similarly to models within the general linear modeling framework. Since this is an introductory book, however, we emphasize here the more direct process of fitting a model to the analyzed matrix of variable interrelationship indices, which can be viewed as the underlying idea of the most general asymptotically distribution-free method of model fitting and testing in SEM. The maximization of the likelihood function for the raw data is equivalent to the minimization of the fit function with the ML method, F_{ML}, which quantifies

This list of characteristics can be used to differentiate structural equation models from what we would like to refer to in this book as classical linear modeling approaches. These classical approaches encompass regression analysis, analysis of variance, analysis of covariance, and a large part of multivariate statistical methods (e.g., Johnson & Wichern, 2002; Marcoulides & Hershberger, 1997). In the classical approaches, typically models are fit to raw data and no error of measurement in the independent variables is assumed.

Despite these differences, an important feature that many of the classical approaches share with SEM is that they are based on linear models. Therefore, a frequent assumption made when using the SEM methodology is that the relationships among observed and/or latent variables are linear (although modeling nonlinear relationships is increasingly gaining popularity in SEM; see Schumacker & Marcoulides, 1998; Muthén & Muthén, 2004; Skrondal & Rabe-Hesketh, 2004). Another shared property between classical approaches and SEM is model comparison. For example, the well-known F test for comparing a less restricted model to a more restricted model is used in regression analysis when a researcher is interested in testing whether to drop from a considered model (prediction equation) one or more independent variables. As discussed later, the counterpart of this test in SEM is the difference in chi-square values test, or its asymptotic equivalents in the form of Lagrange multiplier or Wald tests (e.g., Bentler, 2004). More generally, the chi-square difference test is used in SEM to examine the plausibility of model parameter restrictions, for example equality of factor loadings, factor or error variances, or factor variances and covariances across groups.

Types of Structural Equation Models

The following types of commonly used structural equation models are considered in this book.

1. *Path analysis models.* Path analysis models are usually conceived of only in terms of observed variables. For this reason, some researchers do not consider them typical SEM models. We believe that path analysis models are worthy of discussion within the general SEM framework because, although they only focus on observed variables, they are an important part of the historical development of SEM and in particular use the same underlying idea of model fitting and testing as other SEM models. Figure 1 presents an example of a path analysis model examining the effects of several explanatory variables on the number of hours spent

---

the distance between that matrix and the one reproduced by the model (see section "Rules for Determining Model Parameters" and the Appendix to this chapter).

watching television (see section "Path Diagrams" for a complete list and discussion of the symbols that are commonly used to graphically represent structural equation models).

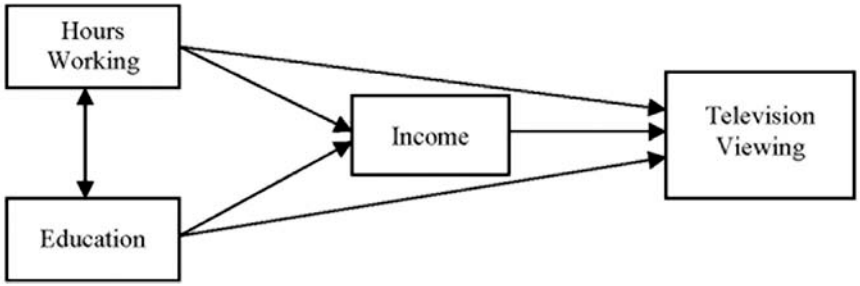

FIG. 1. Path analysis model examining the effects of some variables on television viewing. Hours Working = Average weekly working hours; Education = Number of completed school years; Income = Yearly gross income in dollars; Television Viewing = Average daily number of hours spent watching television.

2. *Confirmatory factor analysis models.* Confirmatory factor analysis models are frequently employed to examine patterns of interrelationships among several latent constructs. Each construct included in the model is usually measured by a set of observed indicators. Hence, in a confirmatory factor analysis model no specific directional relationships are assumed between the constructs, only that they are potentially correlated with one another. Figure 2 presents an example of a confirmatory factor analysis model with two interrelated self-concept constructs (Marcoulides & Hershberger, 1997).

3. *Structural regression models.* Structural regression models resemble confirmatory factor analysis models, except that they also postulate particular explanatory relationships among constructs (latent regressions) rather than these latent variables being only interrelated among themselves. The models can be used to test or disconfirm theories about explanatory relationships among various latent variables under investigation. Figure 3 presents an example of a structural regression model of variables assumed to influence returns of promotion for faculty in higher education (Heck & Johnsrud, 1994).

4. *Latent change models.* Latent change models, often also called latent growth curve models or latent curve analysis models (e.g., Bollen & Curran, 2006; Meredith & Tisak, 1990), represent a means of studying change over time. The models focus primarily on patterns of growth, decline, or both in longitudinal data (e.g., on such aspects of temporal change as initial status and rates of growth or decline), and enable re-

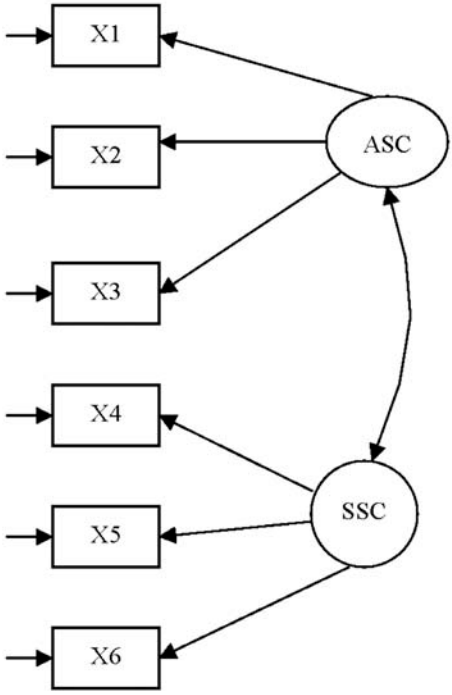

FIG. 2. Confirmatory factor analysis model with two self-concept constructs. ASC = Academic self-concept; SSC = Social self-concept.

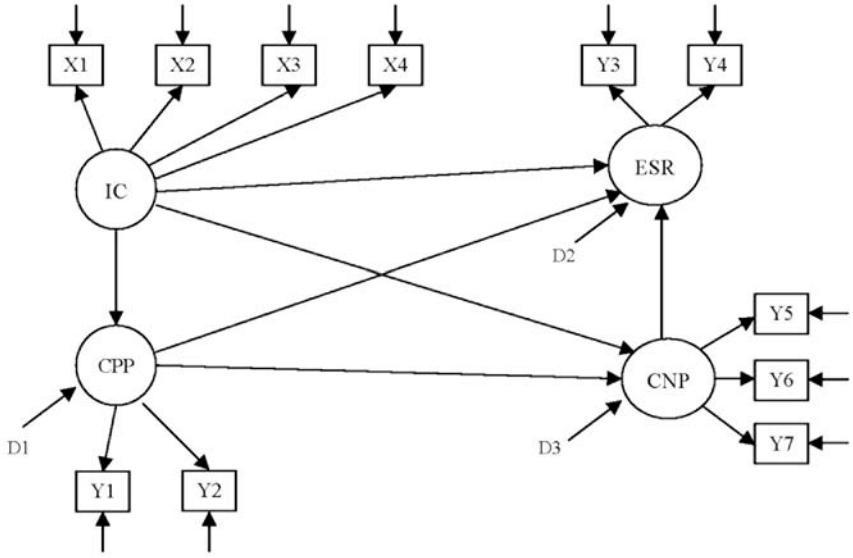

FIG. 3. Structural regression model of variables influencing return to promotion. IC = Individual characteristics; CPP = Characteristics of prior positions; ESR = Economic and social returns to promotion; CNP = Characteristics of new positions.

searchers to examine both intraindividual temporal development and interindividual similarities and differences in its patterns. The models can also be used to evaluate the relationships between patterns of change and other personal characteristics. Figure 4 presents the idea of a simple example of a two-factor growth model for two time points, although typical applications of these models occur in studies with more than two repeated assessments as discussed in more detail in Chapter 6.

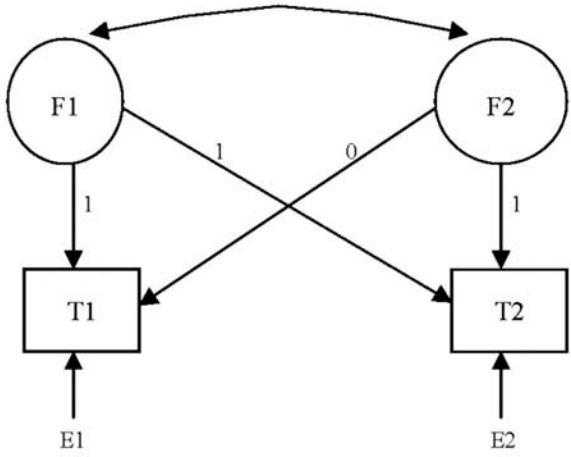

FIG. 4. A simple latent change model.

When and How Are Structural Equation Models Used?

Structural equation models can be utilized to represent knowledge or hypotheses about phenomena studied in substantive domains. The models are usually, and should best be, based on existing or proposed theories that describe and explain phenomena under investigation. With their unique feature of explicitly modeling measurement error, structural equation models provide an attractive means for examining such phenomena. Once a theory has been developed about a phenomenon of interest, the theory can be tested against empirical data using SEM. This process of testing is often called *confirmatory mode* of SEM applications.

A related utilization of structural models is *construct validation*. In these applications, researchers are interested mainly in evaluating the extent to which particular instruments actually measure a latent variable they are supposed to assess. This type of SEM use is most frequently employed when studying the psychometric properties of a given measurement device (e.g., Raykov, 2004).

Structural equation models are also used for theory development purposes. In *theory development*, repeated applications of SEM are carried

out, often on the same data set, in order to explore potential relationships between variables of interest. In contrast to the confirmatory mode of SEM applications, theory development assumes that no prior theory exists—or that one is available only in a rudimentary form—about a phenomenon under investigation. Since this utilization of SEM contributes both to the clarification and development of theories, it is commonly referred to as *exploratory mode* of SEM applications. Due to this development frequently occurring based on a single data set (single sample from a studied population), results from such exploratory applications of SEM need to be interpreted with great caution (e.g., MacCallum, 1986). Only when the findings are replicated across other samples from the same population, can they be considered more trustworthy. The reason for this concern stems mainly from the fact that results obtained by repeated SEM applications on a given sample may be capitalizing on chance factors having lead to obtaining the particular data set, which limits generalizability of results beyond that sample.

Why Are Structural Equation Models Used?

A main reason that structural equation models are widely employed in many scientific fields is that they provide a mechanism for explicitly taking into account measurement error in the observed variables (both dependent and independent) in a given model. In contrast, traditional regression analysis effectively ignores potential measurement error in the explanatory (predictor, independent) variables. As a consequence, regression results can be incorrect and possibly entail misleading substantive conclusions.

In addition to handling measurement error, SEM also enables researchers to readily develop, estimate, and test complex multivariable models, as well as to study both direct and indirect effects of variables involved in a given model. *Direct effects* are the effects that go directly from one variable to another variable. *Indirect effects* are the effects between two variables that are mediated by one or more intervening variables that are often referred to as a mediating variable(s) or mediator(s). The combination of direct and indirect effects makes up the *total effect* of an explanatory variable on a dependent variable. Hence, if an indirect effect does not receive proper attention, the relationship between two variables of concern may not be fully considered. Although regression analysis can also be used to estimate indirect effects—for example by regressing the mediating on the explanatory variable, then the effect variable on the mediator, and finally multiplying pertinent regression weights—this is strictly appropriate only when there are no measurement errors in the involved predictor variables. Such an assumption, however, is in general unrealistic in empirical research in the social and behavioral sciences. In addition, standard errors for relevant estimates are difficult to compute using this sequential ap-

plication of regression analysis, but are quite straightforwardly obtained in SEM applications for purposes of studying indirect effects.

What Are the Key Elements of Structural Equation Models?

The key elements of essentially all structural equation models are their parameters (often referred to as model parameters or unknown parameters). Model parameters reflect those aspects of a model that are typically unknown to the researcher, at least at the beginning of the analyses, yet are of potential interest to him or her. *Parameter* is a generic term referring to a characteristic of a population, such as mean or variance on a given variable, which is of relevance in a particular study. Although this characteristic is difficult to obtain, its inclusion into one's modeling considerations can be viewed as essential in order to facilitate understanding of the phenomenon under investigation. Appropriate sample statistics are used to estimate parameter(s). In SEM, the parameters are unknown aspects of a phenomenon under investigation, which are related to the distribution of the variables in an entertained model. The parameters are estimated, most frequently from the sample covariance matrix and possibly observed variable means, using specialized software.

The presence of parameters in structural equation models should not pose any difficulties to a newcomer to the SEM field. The well-known regression analysis models are also built upon multiple parameters. For example, the partial regression coefficients (or slope), intercept, and standard error of estimate are parameters in a multiple (or simple) regression model. Similarly, in a factorial analysis of variance the main effects and interaction(s) are model parameters. In general, parameters are essential elements of statistical models used in empirical research. The parameters reflect unknown aspects of a studied phenomenon and are estimated by fitting the model to sampled data using particular optimality criteria, numeric routines, and specific software. The topic of structural equation model parameters, along with a complete description of the rules that can be used to determine them, are discussed extensively in the following section "Parameter Estimation."

PATH DIAGRAMS

One of the easiest ways to communicate a structural equation model is to draw a diagram of it, referred to as *path diagram*, using special graphical notation. A path diagram is a form of graphical representation of a model under consideration. Such a diagram is equivalent to a set of equations defining a model (in addition to distributional and related assumptions), and is typically used as an alternative way of presenting a model pictorially. Path diagrams not only enhance the understanding of structural equation models and their communication among researchers with various backgrounds,

but also substantially contribute to the creation of correct command files to fit and test models with specialized programs. Figure 5 displays the most commonly used graphical notation for depicting SEM models, which is described in detail next.

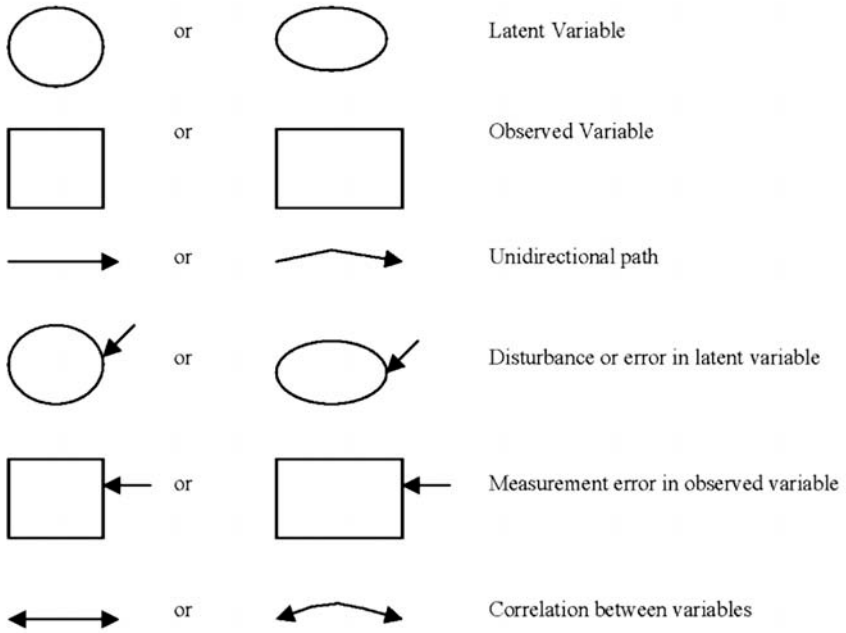

FIG. 5. Commonly used symbols for SEM models in path diagrams.

Latent and Observed Variables

One of the most important initial issues to resolve when using SEM is the distinction between observed variables and latent variables. *Observed variables* are the variables that are actually measured or recorded on a sample of subjects, such as manifested performance on a particular test or the answers to items or questions in an inventory or questionnaire. The term manifest variables is also often used for observed variables, to stress the fact that these are the variables that have actually been measured by the researcher in the process of data collection. In contrast, *latent variables* are typically hypothetically existing constructs of interest in a study. For example, intelligence, anxiety, locus of control, organizational culture, motivation, depression, social support, math ability, and socioeconomic status can all be considered latent variables. The main characteristic of latent vari-

ables is that they cannot be measured directly, because they are not directly observable . Hence, only proxies for them can be obtained using specifically developed measuring instruments, such as tests, inventories, self-reports, testlets, scales, questionnaires, or subscales. These proxies are the indicators of the latent constructs or, in simple terms, their measured aspects. For example, socioeconomic status may be considered to be measured in terms of income level, years of education, bank savings, type of occupation. Similarly, intelligence may be viewed as manifested (indicated) by subject performance on reasoning, figural relations, and culture-fair tests. Further, mathematical ability may be considered indicated by how well students do on algebra, geometry, and trigonometry tasks. Obviously, it is quite common for manifest variables to be fallible and unreliable indicators of the unobservable latent constructs of actual interest to a social or behavioral researcher. If a single observed variable is used as an indicator of a latent variable, it is most likely that the manifest variable will generally contain quite unreliable information about that construct. This information can be considered to be one-sided because it reflects only one aspect of the measured construct, the side captured by the observed variable used for its measurement. It is therefore generally recommended that researchers employ multiple indicators (preferably more than two) for each latent variable in order to obtain a much more complete and reliable picture of it than that provided by a single indicator. There are, however, instances in which a single observed variable may be a fairly good indicator of a latent variable, e.g., the total score on the Stanford-Binet Intelligence Test as a measure of the construct of intelligence.

The discussed meaning of latent variable could be referred to as a traditional, classical, or 'psychometric' conceptualization. This treatment of latent variable reflects a widespread understanding of unobservable constructs across the social and behavioral disciplines as reflecting proper subject characteristics that cannot be directly measured but (a) could be meaningfully assumed to exist separately from their measures without contradicting observed data, and (b) allow the development and testing of potentially far-reaching substantive theories that contribute significantly to knowledge accumulation in these sciences. This conceptualization of latent variable can be traced back perhaps to the pioneering work of the English psychologist Charles Spearman in the area of factor analysis around the turn of the 20th century (e.g., Spearman, 1904), and has enjoyed wide acceptance in the social and behavioral sciences over the past century. During the last 20 years or so, however, developments primarily in applied statistics have suggested the possibility of extending this traditional meaning of the concept of latent variable (e.g., Muthén, 2002; Skrondal & Rabe-Hesketh, 2004). According to what could be referred to as its modern conceptualization, any variable without observed realizations in a studied sample from a population of interest can be considered

a latent variable. In this way, as we will discuss in more detail in Chapter 6, patterns of intraindividual change (individual growth or decline trajectories) such as initial true status or true change across the period of a repeated measure study, can also be considered and in fact profitably used as latent variables. As another, perhaps more trivial example, the error term in a simple or multiple regression equation or in any statistical model containing a residual, can also be viewed as a latent variable. A common characteristic of these examples is that individual realizations (values) of the pertinent latent variables are conceptualized in a given study or modeling approach—e.g., the individual initial true status and overall change, or error score—which realizations however are not observed (see also Appendix to this chapter).[2]

This extended conceptualization of the notion of latent variable obviously includes as a special case the traditional, 'psychometric' understanding of latent constructs, which would be sufficient to use in most chapters of this introductory book. The benefit of adopting the more general, modern understanding of latent variable will be seen in the last chapter of the book. This benefit stems from the fact that the modern view provides the opportunity to capitalize on highly enriching developments in applied statistics and numerical analysis that have occurred over the past couple of decades, which allow one to consider the above modeling approaches, including SEM, as examples of a more general, latent variable modeling methodology (e.g., Muthén, 2002).

Squares and Rectangles, Circles and Ellipses

Observed and latent variables are represented in path diagrams by two distinct graphical symbols. Squares or rectangles are used for observed variables, and circles or ellipses are employed for latent variables. Observed variables are usually labeled sequentially (e.g., X_1, X_2, X_3), with the label centered in each square or rectangle. Latent variables can be abbreviated according to the construct they present (e.g., *SES* for socioeconomic status) or just labeled sequentially (e.g., F_1, F_2; F standing for "factor") with the name or label centered in each circle or ellipse.

Paths and Two-Way Arrows

Latent and observed variables are connected in a structural equation model in order to reflect a set of propositions about a studied phenomenon,

[2]Further, individual class or cluster membership in a latent class model or cluster analysis can be seen as a latent variable. Membership to a constituent in a finite mixture distribution may also be viewed as a latent variable. Similarly, the individual values of random effects in multi-level (hierarchical) or simpler variance component models can be considered scores on latent dimensions as well.

which a researcher is interested in examining (testing) using SEM. Typically, the interrelationships among the latent as well as the latent and observed variables are the main focus of study. These relationships are represented graphically in a path diagram by one-way and two-way arrows. The one-way arrows, also frequently called paths, signal that a variable at the end of the arrow is explained in the model by the variable at the beginning of the arrow. One-way arrows, or paths, are usually represented by straight lines, with arrowheads at the end of the lines. Such paths are often interpreted as symbolizing causal relationships—the variable at the end of the arrow is assumed according to the model to be the effect and the one at the beginning to be the cause. We believe that such inferences should not be made from path diagrams without a strong rationale for doing so. For instance, latent variables are oftentimes considered to be causes for their indicators; that is, the measured or recorded performance is viewed to be the effect of the presence of a corresponding latent variable. We generally abstain from making causal interpretations from structural equation models except possibly when the variable considered temporally precedes another one, in which case the former could be interpreted as the cause of the one occurring later (e.g., Babbie, 1992, chap. 1; Bollen, 1989, chap. 3). Bollen (1989) lists three conditions that should be used to establish a causal relation between variables—isolation, association, and direction of causality. While association may be easier to examine, it is quite difficult to ensure that a cause and effect have been isolated from all other influences. For this reason, most researchers consider SEM models and the causal relations within them only as approximations to reality that perhaps can never really be proved, but rather only disproved or disconfirmed.

In a path diagram, two-way arrows (sometimes referred to as two-way paths) are used to represent covariation between two variables, and signal that there is an association between the connected variables that is not assumed in the model to be directional. Usually two-way arrows are graphically represented as curved lines with an arrowhead at each end. A straight line with arrowheads at each end is also sometimes used to symbolize a correlation between variables. Lack of space may also force researchers to even represent a one-way arrow by a curved rather than a straight line, with an arrowhead attached to the appropriate end (e.g., Fig. 5). Therefore, when first looking at a path diagram of a structural equation model it is essential to determine which of the straight or curved lines have two arrowheads and which only one.

Dependent and Independent Variables

In order to properly conceptualize a proposed model, there is another distinction between variables that is of great importance—the differentiation

between dependent and independent variables. Dependent variables are those that receive at least one path (one-way arrow) from another variable in the model. Hence, when an entertained model is represented as a set of equations (with pertinent distributional and related assumptions), each dependent variable will appear in the left-hand side of an equation. Independent variables are variables that emanate paths (one-way arrows), but never receive a path; that is, no independent variable will appear in the left-hand side of an equation, in that system of model equations. Independent variables can be correlated among one another, i.e., connected in the path diagram by two-way arrows. We note that a dependent variable may act as an independent variable with respect to another variable, but this does not change its dependent-variable status. As long as there is at least one path (one-way arrow) ending at the variable, it is a dependent variable no matter how many other variables in the model are explained by it.

In the econometric literature, the terms exogenous variables and endogenous variables are also frequently used for independent and dependent variables, respectively. (These terms are derived from the Greek words exo and endos, for being correspondingly of external origin to the system of variables under consideration, and of internal origin to it.) Regardless of the terms one uses, an important implication of the distinction between dependent and independent variables is that there are no two-way arrows connecting any two dependent variables, or a dependent with an independent variable, in a model path diagram. For reasons that will become much clearer later, the variances and covariances (and correlations) between dependent variables, as well as covariances between dependent and independent variables, are explained in a structural equation model in terms of its unknown parameters.

An Example Path Diagram of a Structural Equation Model

To clarify further the discussion of path diagrams, consider the factor analysis model displayed in Fig. 6. This model represents assumed relationships among Parental dominance, Child intelligence, and Achievement motivation as well as their indicators.

As can be seen by examining Fig. 6, there are nine observed variables in the model. The observed variables represent nine scale scores that were obtained from a sample of 245 elementary school students. The variables are denoted by the labels V_1 through V_9 (using V for 'observed Variable'). The latent variables (or factors) are Parental dominance, Child intelligence, and Achievement motivation. As latent variables (factors), they are denoted F_1, F_2, and F_3, respectively. The factors are each measured by three indicators, with each path in Fig. 6 symbolizing the factor loading of the observed variable on its pertinent latent variable.

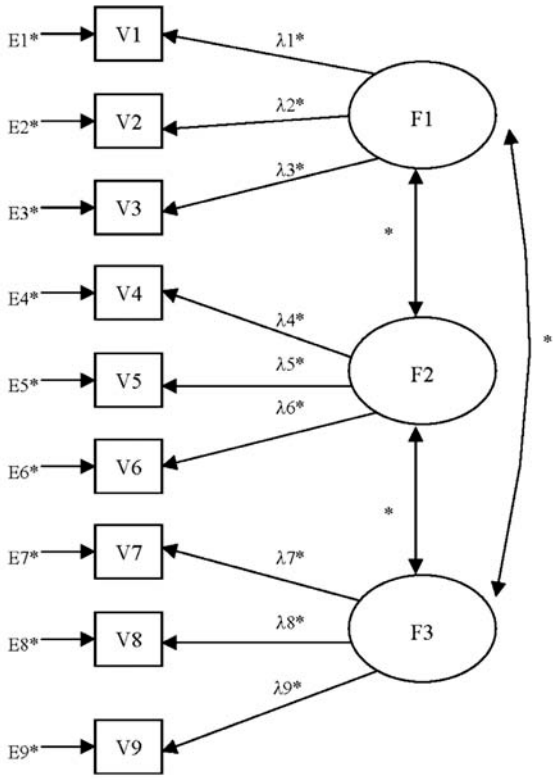

FIG. 6. Example factor analysis model. F_1 = Parental dominance; F_2 = Child intelligence; F_3 = Achievement motivation.

The two-way arrows in Fig. 6 designate the correlations between the latent variables (i.e., the factor correlations) in the model. There is also a residual term attached to each manifest variable. The residuals are denoted by *E* (for Error), followed by the index of the variable to which they are attached. Each residual represents the amount of variation in the manifest variable that is due to measurement error or remains unexplained by variation in the corresponding latent factor that variable loads on. The unexplained variance is the amount of indicator variance unshared with the other measures of the particular common factor. In this text, for the sake of convenience, we will frequently refer to residuals as errors or error terms.

As indicated previously, it is instrumental for an SEM application to determine the dependent and the independent variables of a model under consideration. As can be seen in Fig. 6, and using the definition of error, there are a total of 12 independent variables in this model—these are the

three latent variables and nine residual terms. Indeed, if one were to write out the 9 model definition equations (see below), none of these 12 variables will ever appear in the left-hand side of an equation. Note also that there are no one-way paths going into any independent variable, but there are paths leaving each one of them. In addition, there are three two-way arrows that connect the latent variables—they represent the three factor correlations. The dependent variables are the nine observed variables labeled V_1 through V_9. Each of them receives two paths—(i) the path from the latent variable it loads on, which represents its factor loading; and (ii) the one from its residual term, which represents the error term effect.

First let us write down the *model definition equations*. These are the relationships between observed and unobserved variables that formally define the proposed model. Following Fig. 6, these equations are obtained by writing an equation for each observed variable in terms of how it is explained in the model, i.e., in terms of the latent variable(s) it loads on and corresponding residual term. The following system of nine equations is obtained in this way (one equation per dependent variable):

$$\begin{aligned}
V_1 &= \lambda_1 F_1 + E_1, \\
V_2 &= \lambda_2 F_1 + E_2, \\
V_3 &= \lambda_3 F_1 + E_3, \\
V_4 &= \lambda_4 F_2 + E_4, \\
V_5 &= \lambda_5 F_2 + E_5, \\
V_6 &= \lambda_6 F_2 + E_6, \\
V_7 &= \lambda_7 F_3 + E_7, \\
V_8 &= \lambda_8 F_3 + E_8, \\
V_9 &= \lambda_9 F_3 + E_9,
\end{aligned} \tag{1}$$

where λ_1 to λ_9 (Greek letter *lambda*) denote the nine factor loadings. In addition, we make the usual assumptions of uncorrelated residuals among themselves and with the three factors, while the factors are allowed to be interrelated, and that the nine observed variables are normally distributed, like the three factors and the nine residuals that possess zero means. We note the similarity of these distributional assumptions with those typically made in the multiple regression model (general linear model), specifically the normality of its error term, having zero mean and being uncorrelated with the predictors (e.g., Tabachnick & Fidell, 2001).

According to the factor analysis model under consideration, each of the nine Equations in (1) represents the corresponding observed variable as the sum of the product of that variable's factor loading with its pertinent factor, and a residual term. Note that on the left-hand side of each equation there is only one variable, the dependent variable, rather than a combination of variables, and also that no independent variable appears there.

Model Parameters and Asterisks

Another important feature of path diagrams, as used in this text, are the asterisks associated with one-way and two-way arrows and independent variables (e.g., Fig. 6). These asterisks are symbols of the unknown parameters and are very useful for understanding the parametric features of an entertained model as well as properly controlling its fitting and estimation process with most SEM programs. In our view, a satisfactory understanding of a given model can only then be accomplished when a researcher is able to locate the unknown model parameters. If this is done incorrectly or arbitrarily, there is a danger of ending up with a model that is unduly restrictive or has parameters that cannot be uniquely estimated. The latter problematic parameter estimation feature is characteristic of models that are unidentified—a notion discussed in greater detail in a later section—which are in general useless means of description and explanation of studied phenomena. The requirement of explicit understanding of all model parameters is quite unique to a SEM analysis but essential for meaningful utilization of pertinent software as well as subsequent model modification that is frequently needed in empirical research.

It is instructive to note that in difference to SEM, in regression analysis one does not really need to explicitly present the parameters of a fitted model, in particular when conducting this analysis with popular software. Indeed, suppose a researcher were interested in the following regression model aiming at predicting depression among college students:

$$\begin{aligned}\text{Depression} = a + b_1\text{Social-Support} + \\ b_2\text{Intelligence} + b_3\text{Age} + \text{Error},\end{aligned} \tag{2}$$

where a is the intercept and b_1, b_2, and b_3 are the partial regression weights (slopes), with the usual assumption of normal and homoscedastic error with zero mean, which is uncorrelated with the predictors. When this model is to be fitted with a major statistical package (e.g., SAS or SPSS), the researcher is not required to specifically define a, b_1, b_2 and b_3, as well as the standard error of estimate, as the model parameters. This is due to the fact that unlike SEM, a regression analysis is routinely conducted in only one possible way with regard to the set of unknown parameters. Specifically, when a regression analysis is carried out, a researcher usually only needs to provide information about which measures are to be used as explanatory variables and which as the dependent variables; the utilized software automatically determines then the model parameters, typically one slope per predictor (partial regression weight) plus an intercept for the fitted regression equation and the standard error of estimate.

This automatic or default determination of model parameters does not generally work well in SEM applications and in our view should not be encouraged when the aim is a meaningful utilization of SEM. We find it particularly important in SEM to explicitly keep track of the unknown parameters in order to understand and correctly set up the model one is interested in fitting as well as subsequently appropriately modify it if needed. Therefore, we strongly recommend that researchers always first determine (locate) the parameters of a structural equation model they consider. Using default settings in SEM programs will not absolve a scientist from having to think carefully about this type of details for a particular model being examined. It is the researcher who must decide exactly how the model is defined, not the default features of a computer program used. For example, if a factor analytic model similar to the one presented in Fig. 6 is being considered in a study, one researcher may be interested in having all factor loadings as model parameters, whereas others may have reasons to view only a few of them as unknown. Furthermore, in a modeling session one is likely to be interested in several versions of an entertained model, which differ from one another only in the number and location of their parameters (see chapters 3 through 6 that also deal with such models). Hence, unlike routine applications of regression analysis, there is no single way of assuming unknown parameters without first considering a proposed structural equation model in the necessary detail that would allow one to determine its parameters. Since determination of unknown parameters is in our opinion particularly important in setting up structural equation models, we discuss it in detail next.

RULES FOR DETERMINING MODEL PARAMETERS

In order to correctly determine the parameters that can be uniquely estimated in a considered structural equation model, six rules can be used (cf. Bentler, 2004). The specific rationale behind them will be discussed in the next section of this chapter, which deals with parameter estimation. When the rules are applied in practice, for convenience no distinction needs to be made between the covariance and correlation of two independent variables (as they can be viewed equivalent for purposes of reflecting the degree of linear interrelationship between pairs of variables). For a given structural equation model, these rules are as follows.

Rule 1. All variances of independent variables are model parameters. For example, in the model depicted in Fig. 6 most of the variances of independent variables are symbolized by asterisks that are associated with each error term (residual). Error terms in a path diagram are generally attached to each dependent variable. For a latent dependent variable, an associated error term symbolizes the structural regression

disturbance that represents the variability in the latent variable unexplained by the variables it is regressed upon in the model. For example, the residual terms displayed in Fig. 3, D_1 to D_3, encompass the part of the corresponding dependent variable variance that is not accounted for by the influence of variables explicitly present in the model and impacting that dependent variable. Similarly, for an observed dependent variable the residual represents that part of the variance of the former, which is not explained in terms of other variables that dependent variable is regressed upon in the model. We stress that all residual terms, whether attached to observed or latent variables, are (a) unobserved entities because they cannot be measured and (b) independent variables because they are not affected by any other variable in the model. Thus, by the present rule, the variances of all residuals are, in general, model parameters. However, we emphasize that this rule identifies as a parameter the variance of any independent variable, not only of residuals. Further, if there were a theory or hypothesis to be tested with a model, which indicated that some variances of independent variables (e.g., residual terms) were 0 or equal to a pre-specified number(s), then Rule 1 would not apply and the corresponding independent variable variance will be set equal to that number.

Rule 2. All covariances between independent variables are model parameters (unless there is a theory or hypothesis being tested with the model that states some of them as being equal to 0 or equal to a given constant(s)). In Fig. 6, the covariances between independent variables are the factor correlations symbolized by the two-way arrows connecting the three constructs. Note that this model does not hypothesize any correlation between observed variable residuals—there are no two-way arrows connecting any of the error terms—but other models may have one or more such correlations (e.g., see models in Chap. 5).

Rule 3. All factor loadings connecting the latent variables with their indicators are model parameters (unless there is a theory or hypothesis tested with the model that states some of them as equal to 0 or to a given constant(s)). In Fig. 6, these are the parameters denoted by the asterisks attached to the paths connecting each latent variable to its indicators.

Rule 4. All regression coefficients between observed or latent variables are model parameters (unless there is a theory or hypothesis tested with the model that states that some of them should be equal to 0 or to a given constant(s)). For example, in Fig. 3 the regression coefficients are represented by the paths going from some latent variables and ending at other latent variables. We note that Rule 3 can be considered a special case of Rule 4, after observing that a factor loading can be conceived of as a regression coefficient (slope) of the observed variable when regressed on the pertinent factor. However, performing this regression is typically impossi-

ble in practice because the factors are not observed variables to begin with and, hence, no individual measurements of them are available.

Rule 5. The variances of, and covariances between, dependent variables as well as the covariances between dependent and independent variables are never model parameters. This is due to the fact that these variances and covariances are themselves explained in terms of model parameters. As can be seen in Fig. 6, there are no two-way arrows connecting dependent variables in the model or connecting dependent and independent variables.

Rule 6. For each latent variable included in a model, the metric of its latent scale needs to be set. The reason is that, unlike an observed variable there is no natural metric underlying any latent variable. In fact, unless its metric is defined, the scale of the latent variable will remain indeterminate. Subsequently, this will lead to model-estimation problems and unidentified parameters and models (discussed later in this chapter). For any independent latent variable included in a given model, the metric can be fixed in one of two ways that are equivalent for this purpose. Either its variance is set equal to a constant, usually 1, or a path going out of the latent variable is set to a constant (typically 1). For dependent latent variables, this metric fixing is achieved by setting a path going out of the latent variable to equal a constant, typically 1. (Some SEM programs, e.g., LISREL and M*plus*, offer the option of fixing the scales for both dependent and independent latent variable).

The reason that Rule 6 is needed stems from the fact that an application of Rule 1 on independent latent variables can produce a few redundant and not uniquely estimable model parameters. For example, the pair consisting of a path emanating from a given latent independent variable and this variable's variance, contains a redundant parameter. This means that one cannot distinguish between these two parameters given data on the observed variables; that is, based on all available observations one cannot come up with unique values for this path and latent variance, even if the entire population of interest were examined. As a result, SEM software is not able to estimate uniquely redundant parameters in a given model. Consequently, one of them will be associated with an arbitrarily determined estimate that is therefore useless. This is because both parameters reflect the same aspect of the model, although in a different form, and cannot be uniquely estimated from the sample data, i.e., are not identifiable. Hence, an infinite number of values can be associated with a redundant parameter, and all of these values will be equally consistent with the available data. Although the notion of identification is discussed in more detail later in the book, we note here that unidentified parameters can be made identified if one of

them is set equal to a constant, usually 1, or involved in a relationship with other parameters. This fixing to a constant is the essence of Rule 6.

A Summary of Model Parameters in Fig. 6

Using these six rules, one can easily summarize the parameters of the model depicted in Fig. 6. Following Rule 1, there are nine error term parameters, viz. the variances of E_1 to E_9, as well as three factor variances (but they will be set to 1 shortly, to follow Rule 6). Based on Rule 2, there are three factor covariance parameters. According to Rule 3, the nine factor loadings are model parameters as well. Rule 4 cannot be applied in this model because no regression-type relationships are assumed between latent or between observed variables. Rule 5 states that the relationships between the observed variables, which are the dependent variables of the model, are not parameters because they are supposed to be explained in terms of the actual model parameters. Similarly, the relationships between dependent and independent variables are not model parameters.

Rule 6 now implies that in order to fix the metric of the three latent variables one can set their variances to unity or fix to 1 a path going out of each one of them. If a particularly good, that is, quite reliable, indicator of a latent variable is available, it may be better to fix the scale of that latent variable by setting to 1 the path leading from it to that indicator. Otherwise, it may be better to fix the scale of the latent variables by setting their variances to 1. We note that the paths leading from the nine error terms to their corresponding observed variables are not considered to be parameters, but instead are assumed to be equal to 1, which in fact complies with Rule 6 (fixing to 1 a loading on a latent variable, which an error term formally is, as mentioned above). For the latent variables in Fig. 6, one simply sets their variances equal to 1, because all their loadings on the pertinent observed variables are already assumed to be model parameters. This setting latent variances equal to 1 means that these variances are no more model parameters, and overrides the asterisks that would otherwise be attached to each latent variable circle in Fig. 6 to enhance pictorially the graphical representation of the model.

Therefore, applying all six rules, the model in Fig. 6 has altogether 21 parameters to be estimated—these are its nine error variances, nine factor loadings, and three factor covariances. We emphasize that testing any specific hypotheses in a model, e.g., whether all indicator loadings on the Child intelligence factor have the same value, places additional parameter restrictions and inevitably decreases the number of parameters to be estimated, as discussed further in the next section. For example, if one assumes that the three loadings on the Child intelligence factor in Fig. 6 are equal to one another, it follows that they can be represented by a single model parameter.

In that case, imposing this restriction decreases by two the number of unknown parameters to 19, because the three factor loadings involved in the constraint are not represented by three separate parameters anymore but only by a single one.

Free, Fixed, and Constrained Parameters

There are three types of model parameters that are important in conducting SEM analyses—free, fixed, and constrained. All parameters that are determined based on the above six rules are commonly referred to as *free parameters* (unless a researcher imposes additional constraints on some of them; see below), and must be estimated when fitting the model to data. For example, in Fig. 6 asterisks were used to denote the free model parameters in that factor analysis model. *Fixed parameters* have their value set equal to a given constant; such parameters are called fixed because they do not change value during the process of fitting the model, unlike the free parameters. For example, in Fig. 6 the covariances (correlations) among error terms of the observed variables V_1 to V_9 are fixed parameters since they are all set equal to 0; this is the reason why there are no two-way arrows connecting any pair of residuals in Fig. 6. Moreover, following Rule 6 one may decide to set a factor loading or alternatively a latent variance equal to 1. In this case, the loading or variance in question also becomes a fixed parameter. Alternatively, a researcher may decide to fix other parameters that were initially conceived of as free parameters, which might represent substantively interesting hypotheses to be tested with a given model. Conversely, a researcher may elect to free some initially fixed parameters, rendering them free parameters, after making sure of course that the model remains identified (see below).

The third type of parameters are called constrained parameters, also sometimes referred to as restricted or restrained parameters. *Constrained parameters* are those that are postulated to be equal to one another—but their value is not specified in advance as is that of fixed parameters—or involved in a more complex relationship among themselves. Constrained parameters are typically included in a model if their restriction is derived from existing theory or represents a substantively interesting hypothesis to be tested with the model. Hence, in a sense, constrained parameters can be viewed as having a status between that of free and of fixed parameters. This is because constrained parameters are not completely free, being set to follow some imposed restriction, yet their value can be anything as long as the restriction is preserved, rather than locked at a particular constant as is the case with a fixed parameter. It is for this reason that both free and constrained parameters are frequently referred to as model parameters. Oftentimes in the literature, all free parameters plus a representative(s) for the parameters involved in each

restriction in a considered model, are called independent model parameters. Therefore, whenever we refer to number of model parameters in the remainder, we will mean the number of independent model parameters (unless explicitly mentioned otherwise).

For example, imagine a situation in which a researcher hypothesized that the factor loadings of the Parental dominance construct associated with the measures V_1, V_2, and V_3 in Fig. 6 were all equal; such indicators are usually referred to in the psychometric literature as tau-equivalent measures (e.g., Jöreskog, 1971). This hypothesis amounts to the assumption that these three indicators measure the same latent variable in the same unit of measurement. Hence, by using constrained parameters, a researcher can test the plausibility of this hypothesis. If constrained parameters are included in a model, however, their restriction should be derived from existing theory or formulated as a substantively meaningful hypothesis to be tested. Further discussion concerning the process of testing parameter restrictions is provided in a later section of the book.

PARAMETER ESTIMATION

In any structural equation model, the unknown parameters are estimated in such a way that the model becomes capable of "emulating" the analyzed sample covariance or correlation matrix, and in some circumstances sample means (e.g., Chap. 6). In order to clarify this feature of the estimation process, let us look again at the path diagram in Fig. 6 and the associated model definition Equations 1 in the previous section. As indicated in earlier discussions the model represented by this path diagram, or system of equations, makes certain assumptions about the relationships between the involved variables. Hence, the model has specific implications for their variances and covariances. These implications can be worked out using a few simple relations that govern the variances and covariances of linear combinations of variables. For convenience, in this book these relations are referred to as the four laws of variances and covariances; they follow straightforwardly from the formal definition of variance and covariance (e.g., Hays, 1994).

The Four Laws for Variances and Covariances

Denote variance of a variable under consideration by 'Var' and covariance between two variables by 'Cov.' For a random variable X (e.g., an intelligence test score), the first law is stated as follows:

Law 1:

$$\text{Cov}(X,X) = \text{Var}(X).$$

Law 1 simply says that the covariance of a variable with itself is that variable's variance. This is an intuitively very clear result that is a direct consequence of the definition of variance and covariance. (This law can also be readily seen in action by looking at the formula for estimation of variance and observing that it results from the formula for estimating covariance when the two variables involved coincide; e.g., Hays, 1994.)

The second law allows one to find the covariance of two linear combinations of variables. Assume that X, Y, Z, and U are four random variables—for example those denoting the scores on tests of depression, social support, intelligence, and a person's age (see Equation 2 in the section "Rules for Determining Model Parameters"). Suppose that a, b, c, and d are four constants. Then the following relationship holds:

Law 2:

$$\text{Cov}(aX + bY, cZ + dU) = ac\ \text{Cov}(X,Z) + ad\ \text{Cov}(X,U) + bc\ \text{Cov}(Y,Z) + bd\ \text{Cov}(Y,U).$$

This law is quite similar to the rule of disclosing brackets used in elementary algebra. Indeed, to apply Law 2 all one needs to do is simply determine each resulting product of constants and attach the covariance of their pertinent variables. Note that the right-hand side of the equation of this law simplifies markedly if some of the variables are uncorrelated, that is, one or more of the involved covariances is equal to 0. Law 2 is extended readily to the case of covarying linear combinations of any number of initial variables, by including in its right-hand side all pairwise covariances pre-multiplied with products of pertinent weights.[3]

Using Laws 1 and 2, and the fact that $\text{Cov}(X,Y) = \text{Cov}(Y,X)$ (since the covariance does not depend on variable order), one obtains the next equation, which, due to its importance for the remainder of the book, is formulated as a separate law:

[3]Law 2 reveals the rationale behind the rules for determining the parameters for any model once the definition equations are written down (see section "Rules for Determining Model Parameters" and Appendix to this chapter). Specifically, Law 2 states that the covariance of any pair of observed measures is a function of (i) the covariances or variances of the variables involved and (ii) the weights by which these variables are multipled and then summed up in the equations for these measures, as given in the model definition equations. The variables mentioned in (i) are the pertinent independent variables of the model (their analogs in Law 2 are X, Y, Z, and U); the weights mentioned in (ii) are the respective factor loadings or regression coefficients in the model (their analogs in Law 2 are the constants a, b, c, and d). Therefore, the parameters of any SEM model are (a) the variances and covariances of the independent variables, and (b) the factor loadings or regression coefficients (unless there is a theory or hypothesis tested within the model that states that some of them are equal to constants, in which case the parameters are the remaining of the quantities envisaged in (a) and (b)).

Law 3:

$$\begin{aligned}\mathrm{Var}(aX + bY) &= \mathrm{Cov}(aX + bY, aX + bY)\\ &= a^2\,\mathrm{Cov}(X,X) + b^2\,\mathrm{Cov}(Y,Y) + ab\,\mathrm{Cov}(X,Y) + ab\,\mathrm{Cov}(X,Y),\end{aligned}$$

or simply

$$\mathrm{Var}(aX + bY) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y) + 2ab\,\mathrm{Cov}(X,Y).$$

A special case of Law 3 that is used often in this book involves uncorrelated variables X and Y (i.e., $\mathrm{Cov}(X,Y) = 0$), and for this reason is formulated as another law:

Law 4: If X and Y are uncorrelated, then

$$\mathrm{Var}(aX + bY) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y).$$

We also stress that there are no restrictions in Laws 2, 3, and 4 on the values of the constants a, b, c, and d—in particular, they could take on the values 0 or 1, for example. In addition, we emphasize that these laws generalize straightforwardly to the case of linear combinations of more than two variables.

Model Implications and Reproduced Covariance Matrix

As mentioned earlier in this section, any considered model has certain implications for the variances and covariances (and means, if included in the analysis) of the involved observed variables. In order to see these implications, the four laws for variances and covariances can be used. For example, consider the first two manifest variables V_1 and V_2 presented in Equations 1 (see the section "Rules for Determining Model Parameters" and Fig. 6). Because both variables load on the same latent factor F_1, we obtain the following equality directly from Law 2 (see also the first two of Equations (1)):

$$\begin{aligned}\mathrm{Cov}(V_1,V_2) &= \mathrm{Cov}(\lambda_1 F_1 + E_1, \lambda_2 F_1 + E_2)\\ &= \lambda_1\lambda_2\,\mathrm{Cov}(F_1,F_1) + \lambda_1\,\mathrm{Cov}(F_1,E_2) + \lambda_2\,\mathrm{Cov}(E_1,F_1) + \mathrm{Cov}(E_1,E_2)\\ &= \lambda_1\,\lambda_2\,\mathrm{Cov}(F_1,F_1)\\ &= \lambda_1\lambda_2\,\mathrm{Var}(F_1)\\ &= \lambda_1\lambda_2. \qquad (3)\end{aligned}$$

To obtain Equation 3, the following two facts regarding the model in Fig. 6 are also used. First, the covariance of the residuals E_1 and E_2, and the covariance of each of them with the factor F_1, are equal to 0 according to our earlier assumptions when defining the model (note that in Fig. 6 there are no two-headed arrows connecting the residuals or any of them with F_1); second, the variance of F_1 has been set equal to 1 according to Rule 6 (i.e., $\mathrm{Var}(F_1) = 1$).

Similarly, using Law 2, the covariance between the observed variables V_1 and V_4 say (each loading on a different factor) is determined as follows:

$$\begin{aligned}\text{Cov}(V_1,V_4) &= \text{Cov}(\lambda_1 F_1 + E_1, \lambda_4 F_2 + E_4)\\ &= \lambda_1\lambda_4\,\text{Cov}(F_1,F_2) + \lambda_1\,\text{Cov}(F_1,E_4) + \lambda_4\,\text{Cov}(E_1,F_2) + \text{Cov}(E_1,E_4)\\ &= \lambda_1\lambda_4\phi_{21}, \qquad (4)\end{aligned}$$

where ϕ_{21} (Greek letter *phi*) denotes the covariance between the factors F_1 and F_2.

Finally, the variance of the observed variable V_1, say, is determined using Law 4 and the previously stated facts, as:

$$\begin{aligned}\text{Var}(V_1) &= \text{Cov}(\lambda_1 F_1 + E_1, \lambda_1 F_1 + E_1)\\ &= \lambda_1^2\,\text{Cov}(F_1,F_1) + \lambda_1\,\text{Cov}(F_1,E_1) + \lambda_1\,\text{Cov}(E_1,F_1) + \text{Cov}(E_1,E_1)\\ &= \lambda_1^2\,\text{Var}(F_1) + \text{Var}(E_1)\\ &= \lambda_1^2 + \theta_1, \qquad (5)\end{aligned}$$

where θ_1 (Greek letter *theta*) symbolizes the variance of the residual E_1.

If this process were continued for every combination of say p observed variables in a given model (i.e., V_1 to V_9 for the model in Fig. 6), one would obtain every element of a variance-covariance matrix. This matrix will be denoted by $\Sigma(\gamma)$ (the Greek letter *sigma*), where γ denotes the set or vector of all model parameters (see, e.g., Appendix to this chapter). The matrix $\Sigma(\gamma)$ is referred to as the reproduced, or model-implied, covariance matrix. Since $\Sigma(\gamma)$ is symmetric, being a covariance matrix, it has altogether $p(p + 1)/2$ nonredundant elements; that is, it has 45 elements for the model in Fig. 6. This number of nonredundant elements will also be used later in this chapter to determine the degrees of freedom of a model under consideration, so we make a note of it here.

Hence, using Laws 1 through 4 for the model in Fig. 6, the following reproduced covariance matrix $\Sigma(\gamma)$ is obtained (displaying only its nonredundant elements, i.e., its diagonal entries and those below the main diagonal and placing this matrix within brackets):

$$\Sigma(\gamma) = \begin{bmatrix}
\lambda_1^2+\theta_1 & & & & & & & & \\
\lambda_1\lambda_2 & \lambda_2^2+\theta_2 & & & & & & & \\
\lambda_1\lambda_3 & \lambda_2\lambda_3 & \lambda_3^2+\theta_3 & & & & & & \\
\lambda_1\lambda_4\phi_{21} & \lambda_2\lambda_4\phi_{21} & \lambda_3\lambda_4\phi_{21} & \lambda_4^2+\theta_4 & & & & & \\
\lambda_1\lambda_5\phi_{21} & \lambda_2\lambda_5\phi_{21} & \lambda_3\lambda_5\phi_{21} & \lambda_4\lambda_5 & \lambda_5^2+\theta_5 & & & & \\
\lambda_1\lambda_6\phi_{21} & \lambda_2\lambda_6\phi_{21} & \lambda_3\lambda_6\phi_{21} & \lambda_4\lambda_6 & \lambda_5\lambda_6 & \lambda_6^2+\theta_6 & & & \\
\lambda_1\lambda_7\phi_{31} & \lambda_2\lambda_7\phi_{31} & \lambda_3\lambda_7\phi_{31} & \lambda_4\lambda_7\phi_{32} & \lambda_5\lambda_7\phi_{32} & \lambda_6\lambda_7\phi_{32} & \lambda_7^2+\theta_7 & & \\
\lambda_1\lambda_8\phi_{31} & \lambda_2\lambda_8\phi_{31} & \lambda_3\lambda_8\phi_{31} & \lambda_4\lambda_8\phi_{32} & \lambda_5\lambda_8\phi_{32} & \lambda_6\lambda_8\phi_{32} & \lambda_7\lambda_8 & \lambda_8^2+\theta_8 & \\
\lambda_1\lambda_9\phi_{31} & \lambda_2\lambda_9\phi_{31} & \lambda_3\lambda_9\phi_{31} & \lambda_4\lambda_9\phi_{32} & \lambda_5\lambda_9\phi_{32} & \lambda_6\lambda_9\phi_{32} & \lambda_7\lambda_9 & \lambda_8\lambda_9 & \lambda_9^2+\theta_9
\end{bmatrix}.$$

We stress that the elements of $\Sigma(\gamma)$ are all nonlinear functions of model parameters. In addition, each element of $\Sigma(\gamma)$ has as a counterpart a corresponding numerical element (entry) in the observed empirical covariance matrix that is obtained from the sample at hand for the nine observed variables under consideration here. Assuming that this observed sample covariance matrix, denoted by S, is as follows:

$$S = \begin{bmatrix} 1.01 & & & & & & & & \\ 0.32 & 1.50 & & & & & & & \\ 0.43 & 0.40 & 1.22 & & & & & & \\ 0.38 & 0.25 & 0.33 & 1.13 & & & & & \\ 0.30 & 0.20 & 0.30 & 0.70 & 1.06 & & & & \\ 0.33 & 0.22 & 0.38 & 0.72 & 0.69 & 1.12 & & & \\ 0.20 & 0.08 & 0.07 & 0.20 & 0.27 & 0.20 & 1.30 & & \\ 0.33 & 0.19 & 0.22 & 0.09 & 0.22 & 0.12 & 0.69 & 1.07 & \\ 0.52 & 0.27 & 0.36 & 0.33 & 0.37 & 0.29 & 0.50 & 0.62 & 1.16 \end{bmatrix},$$

then the top element value of S (i.e., 1.01) corresponds to $\lambda_1^2 + \theta_1$ in the reproduced matrix $\Sigma(\gamma)$. Similarly, the counterpart of the element .72 in the sixth row and fourth column of S, is $\lambda_4\lambda_6$ in the $\Sigma(\gamma)$ matrix; conversely, the counterpart of the element in the 3rd row and 1st column in $\Sigma(\gamma)$, viz. $\lambda_1\lambda_3$, is .43 in S, and so on.

Now imagine setting the pairs of counterpart elements in S and $\Sigma(\gamma)$ equal to one another, from the top-left corner of S to its bottom-right corner; that is, for the model displayed in Fig. 6, set $1.01 = \lambda_1^2 + \theta_1$, then $0.32 = \lambda_1\lambda_2$, and so on until for the last elements, $1.16 = \lambda_9^2 + \theta_9$ is set. As a result of this equality setting, for the model in Fig. 6 a system of 45 equations (viz., the number of nonredundant elements in its matrix $\Sigma(\gamma)$ or S) is generated with as many unknowns as there are model parameters—that is, 21, as there are 21 asterisks in Fig. 6. Hence, one can conceive of the process of fitting a structural equation model as solving a system of possibly nonlinear equations. For each equation, its left-hand side is a subsequent numerical entry of the sample covariance matrix S, whereas its right-hand side is its counterpart element of the matrix $\Sigma(\gamma)$, i.e., the corresponding expression of model parameters at the same position in the model reproduced covariance matrix. Therefore, fitting a structural equation model is conceptually equivalent to solving this system of equations obtained according to the consequences of the model, whereby this solution is sought in an optimal way that is discussed in the next section.

The preceding discussion in this section also demonstrates that the model presented in Fig. 6 implies, as does any structural equation model, a specific structuring of the elements of the covariance matrix (and sometimes means; e.g., Chapter 6) that is reproduced by that model in terms of

particular expressions—in general, nonlinear functions—of unknown model parameters. Hence, if certain values for the parameters were entered into these expressions or functions, one would obtain a covariance matrix that has numbers as elements. In fact the process of fitting a model to data with SEM programs can be thought of as a repeated insertion of appropriate values for the parameters in the matrix $\Sigma(\gamma)$ until a certain optimality criterion, in terms of its proximity to the sample covariance matrix S, is satisfied (see below).

Every available SEM software has built into its memory the exact way in which these functions of model parameters in $\Sigma(\gamma)$ can be obtained (e.g., see Appendix to this chapter). For ease of computation, most programs make use of matrix algebra, with the software in effect determining each of the parametric expressions involved in these $p(p+1)/2$ equations. This occurs quite automatically once a researcher has communicated to the program the model with its parameters (and a few other related details discussed in the next chapter).

How Good Is a Proposed Model?

The previous section illustrated how a given structural equation model leads to a reproduced covariance matrix $\Sigma(\gamma)$ that is fit to the observed sample covariance matrix S through appropriate choice of values for the model parameters. Now it would seem that the next logical question is, "How can one measure or evaluate the extent to which the matrices S and $\Sigma(\gamma)$ differ?" This is a particularly important question in SEM because it permits one to assess the goodness of fit of the model. Indeed, if the difference between S and $\Sigma(\gamma)$ is small for a particular (optimal) set of values of the unknown parameters, then one can conclude that the model represents the observed data reasonably well. On the other hand, if this difference is large, one can conclude that the model is not consistent with the observed data. There are at least two reasons for such inconsistencies: (a) the proposed model may be deficient, in the sense that it is not capable of emulating well enough the analyzed matrix of variable interrelationships even with most favorable parameter values, or (b) the data may not be good, i.e., are deficient in some way, for example not measuring well the aspects of the studied phenomenon that are reflected in the model. Hence, in order to proceed with model fit assessment, a method is needed for evaluating the degree to which the reproduced matrix $\Sigma(\gamma)$ differs from the sample covariance matrix S.

In order to clarify this method, a new concept needs to be introduced, that of distance between matrices. Obviously, if the values to be compared were scalars, i.e., single numbers, a simple subtraction of one from the other (and possibly taking absolute value of the resulting difference) would suffice to evaluate the distance between them. However, this cannot be

done directly with the two matrices S and $\Sigma(\gamma)$. Subtracting the matrix S from the matrix $\Sigma(\gamma)$ does not result in a single number; rather, according to the rules of matrix subtraction (e.g., Johnson & Wichern, 2002) a matrix consisting of counterpart element differences is obtained.

There are some meaningful ways to evaluate the distance between two matrices, however, and the resulting distance measure ends up being a single number that is easier to interpret. Perhaps the simplest way to obtain this number involves taking the sum of the squares of the differences between the corresponding elements of the two matrices. Other more complicated ways involve the multiplication of these squares with some appropriately chosen weights and then taking their sum (discussed later). In either case, the single number obtained in the end represents a generalized measure of the distance between two matrices considered, in particular S and $\Sigma(\gamma)$. The larger this number, the more different these matrices are; the smaller the number, the more similar they are. Since in SEM this number typically results after comparing the elements of S with those of the model-implied covariance matrix $\Sigma(\gamma)$, this generalized distance is a function of the model parameters as well as the observed variances and covariances. Therefore, it is customary to refer to the relationship between the matrix distance, on the one hand, and the model parameters and S, on the other hand, as a *fit function*. Being defined as the distance between two matrices, the fit function value is always positive or 0. Whenever its value is 0, and only then, the two matrices involved are identical.

It turns out that depending on how the matrix distance is defined, several fit functions result. These fit functions, along with their corresponding methods of parameter estimation, are discussed next.

Methods of Parameter Estimation

There are four main estimation methods and types of fit functions in SEM: unweighted least squares, maximum likelihood, generalized least squares, and asymptotically distribution free (often called weighted least squares). The application of each estimation approach is based on the minimization of a corresponding fit function.

The unweighted least squares (ULS) method uses as a fit function, denoted F_{ULS}, the simple sum of squared differences between the corresponding elements of S and the model reproduced covariance matrix $\Sigma(\gamma)$. Accordingly, the estimates for the model parameters are those values for which F_{ULS} attains its smallest value. The ULS estimation approach can typically be used when the same or similar scales of measurement underlie the analyzed variables.

The other three estimation methods are based on the same sum of squares as the ULS approach, but after specific weights have been used to multiply

each of the squared element-wise differences between S and $\Sigma(\gamma)$, resulting in corresponding fit functions. These functions are designated F_{GLS}, F_{ML} and F_{ADF}, for the generalized least squares (GLS), maximum likelihood (ML), and asymptotically distribution free (ADF) method, respectively (see Appendix to this chapter for their definition). The ML and GLS methods can be used when the observed data are multivariate normally distributed. This assumption is quite frequently made in multivariate analyses and can be examined using general-purpose statistical packages (e.g., SAS or SPSS); a number of consequences of it can also be addressed with SEM software. As discussed in more detail in alternative sources (e.g., Tabachnick & Fidell, 2001; Khattree & Naik, 1999), examining multinormality involves several steps. The simplest way to assess univariate normality, an implication of multivariate normality, is to consider skewness and kurtosis, and statistical tests are available for this purpose. Skewness is an index that reflects the lack of symmetry of a univariate distribution. Kurtosis has to do with the shape of the distribution in terms of its peakedness relative to a corresponding normal distribution. Under normality, the univariate skewness and kurtosis coefficients are 0; if they are markedly distinct from 0, the univariate and hence multivariate normality assumption is violated. (For statistical tests of their equality to 0, see Tabachnick & Fidell, 2001.) There is also a measure of multivariate kurtosis called Mardia's multivariate kurtosis coefficient, and its normalized estimate is of particular relevance in empirical research (e.g., Bentler, 2004). This coefficient measures the extent to which the multivariate distribution of all observed variables has tails that differ from the ones characteristic of the normal distribution, with the same component means, variances and covariances. If the distribution deviates only slightly from the normal, Mardia's coefficient will be close to 0; then its normalized estimate, which can be considered a standard normal variable under normality, will probably be nonsignificant.

Although it may happen that multivariate normality holds when all observed variables are individually normally distributed, it is desirable to also examine bivariate normality that is generally not a consequence of univariate normality. In fact, if the observations are from a multivariate normal distribution, each bivariate distribution should also be normal, like each univariate distribution. (We stress that the converse does not hold— univariate and/or bivariate normality does not imply multivariate normality.) A graphical method for examining bivariate normality involves looking at the scatter plots between all pairs of analyzed variables to ensure that they have (at least approximately) cigar-shaped forms (e.g., Tabachnik & Fidell, 2001). A formal method for judging bivariate normality is based on a plot of the chi-square percentiles and the mean distance measure of individual observations (e.g., Johnson & Wichern, 2002; Khattree & Naik, 1999; Marcoulides & Hershberger, 1997). If the distribution is normal, the plot of appropriately chosen chi-square percentiles and the individual score distance to mean

should resemble a straight line. (For further details on assessing multivariate normality, see also Marcoulides & Hershberger, 1997, pp. 48–52.)

In recent years, research has shown that the ML method can also be employed with minor deviations from normality (e.g., Bollen, 1989; Jöreskog & Sörbom, 1993b; see also Raykov & Widaman, 1995), especially when one is primarily interested in parameter estimates. In this text for a first course in SEM that is chiefly concerned with the fundamentals of SEM, we consider only the ML method for parameter estimation and model testing purposes with continuous latent and manifest variables, and refer to its robustness features in cases with slight deviations from observed variable normality (e.g., Bentler, 2004). In more general terms, the ML method aims at finding estimates for the model parameters that maximize the likelihood of observing the available data if one were to collect data from the same population again. (For a nontechnical introduction to ML, in particular in the context of missing data analysis via SEM, see, e.g., Raykov, 2005). This maximization is achieved by selecting, using a numerical search procedure across the space of all possible parameter values, numerical values for all model parameters in such a way that they minimize the corresponding fit function, F_{ML}.

With more serious deviations from normality, the asymptotically distribution free (or weighted least squares) method can be used as long as the analyzed sample is fairly large. Sample size plays an important role in almost every statistical technique applied in empirical research. Although there is universal agreement among researchers that the larger the sample relative to the population the more stable the parameter estimates, there is no agreement as to what constitutes large, due to the exceeding complexity of this matter. This topic has received a considerable amount of attention in the literature, but no easily applicable and clear-cut general rules of thumb have been proposed. To give only an idea of the issue involved, a cautious and simplified attempt at a rule of thumb might suggest that sample size would desirably be more than 10 times the number of free model parameters (cf. Bentler, 1995; Hu, Bentler, & Kano, 1992). Nevertheless, it is important to emphasize that no rule of thumb can be applied indiscriminately to all situations. This is because the appropriate size of a sample depends on many factors, including the psychometric properties of the variables, the strength of the relationships among the variables considered and size of the model, and the distributional characteristics of the variables (as well as, in general, the amount of missing data). When all these above mentioned issues are considered, samples of varying magnitude may be needed to obtain reasonable parameter estimates. If the observed variable distribution is not quite normal and does not demonstrate piling of cases at one end of the range of values for manifest variables, researchers are encouraged to use the Satorra-Bentler robust ML method of parameter estimation (e.g., Bentler, 2004; Muthén & Muthén, 2004; du Toit & du Toit, 2001). This promising approach is based on corrected statistics

obtainable with the ML method, and for this reason is frequently referred to as robust ML method. It is available in all three programs used later in this book, EQS, LISREL, and M*plus* (see, e.g., Raykov, 2004, for evaluation of model differences with this method) and provides overall model fit test statistics and parameter standard errors (see below) that are all robust to mild deviations from normality.

Another alternative to dealing with nonnormality is to make the data appear more normal by introducing some normalizing transformation on the raw data. Once the data have been transformed and closeness to normality achieved, normal theory analysis can be carried out. In general, transformations are simply reexpressions of the data in different units of measurement. Numerous transformations have been proposed in the literature. The most popular are (a) power transformations, such as squaring each data point, taking the square root of it, or reciprocal transformations; and (b) logarithmic transformations. Last but not least, one may also want to consider alternative measures of the constructs involved in a proposed model, if such are readily available.

When data stem from designs with only a few possible response categories, the asymptotically distribution free method can be used with polychoric or polyserial correlations, or a somewhat less restrictive categorical data analysis approach can be utilized that is available within a general, latent variable modeling framework (e.g., Muthén, 2002; Muthén & Muthén, 2004). For example, suppose a questionnaire included the item, "How satisfied are you with your recent car purchase?", with response categories labeled, "Very satisfied", "Somewhat satisfied", and "Not satisfied". A considerable amount of research has shown that ignoring the categorical attributes of data obtained from items like these can lead to biased SEM results obtained with standard methods, such as that based on minimization of the ordinary ML fit function. For this reason, it has been suggested that use of the polychoric-correlation coefficient (for assessing the degree of association between ordinal variables) and the polyserial-correlation coefficient (for assessing the degree of association between an ordinal variable and a continuous variable) can be made, or alternatively the above mentioned latent variable modeling approach to categorical data analysis may be utilized. Some research has also demonstrated that when there are five or more response categories, and the distribution of data could be viewed as resembling normal, the problems from disregarding the categorical nature of responses are likely to be relatively minimal (e.g., Rigdon, 1998), especially if one uses the Satorra-Bentler robust ML approach. Hence, once again, examining the data distribution becomes essential.

From a statistical perspective, all four mentioned parameter estimation methods lead to consistent estimates. *Consistency* is a desirable feature insuring that with increasing sample size the estimates converge to the un-

known, true population parameter values of interest. Hence, consistency can be considered a minimal requirement for parameter estimates obtained with a given estimation method in order for the latter to be recommendable. With large samples, the estimates obtained using the ML and GLS approaches under multinormality, or the ADF method, possess the additional important property of being normally distributed around their population counterparts. Moreover, with large samples these three methods yield under their assumptions efficient estimates, which are associated with the smallest possible variances across the set of consistent estimates using the same data information and therefore allow one to evaluate most precisely the model parameters.

Iterative Estimation of Model Parameters

The final question now is, "Using any of the above estimation methods, how does one actually evaluate the parameters of a given model in order to render the empirical covariance matrix S and the model reproduced covariance matrix $\Sigma(\gamma)$ as close as possible?" In order to answer this question, one must resort to special numerical routines. Their goal is to minimize the fit function corresponding to the chosen method of estimation. These numerical routines proceed in a consecutive, or iterative, manner by selecting values for model parameters according to the following principle. At each step, the method specific distance—that is, fit function value—between S and $\Sigma(\gamma)$ with the new parameter values, should be smaller than this distance with the parameter values available at the preceding step. This principle is followed until no further improvement in the fit function can be achieved. At that moment, there is no additional decrease possible in the generalized distance between the empirical covariance matrix S and the model reproduced covariance matrix $\Sigma(\gamma)$, as defined by the used estimation method (e.g., Appendix to this chapter).

This iterative process starts with initial estimates of the parameters, i.e., start values for all parameters. Quite often, these values can be automatically calculated by the SEM software used, although researchers can provide their own initial values if they so choose with some complicated models. The iteration process terminates (i.e., converges) if at some step the fit function does not change by more than a very small amount (typically .000001 or a very close number, yet this value may be changed by the researcher if they have a strong reason). That is, the iterative process of parameter estimation converges at that step where practically no further improvement is possible in the distance between sample covariance matrix S and model reproduced covariance matrix $\Sigma(\gamma)$. The numerical values for the parameters obtained at that final iteration step represent the required estimates of the model parameters. We emphasize that in order for a set of

parameter estimates to be meaningful, it is necessary that the iterative process converges, i.e., terminates and thus yields a final solution. If convergence does not occur, a warning sign is issued by the software utilized, which is easily spotted in the output, and the parameter estimates at the last iteration step are in general meaningless, except perhaps being useful for tracking down the problem of lack of convergence.

All converged solutions also provide a measure of the sampling variability for each obtained parameter estimate, called *standard error*. The magnitude of the standard error indicates how stable the pertinent parameter estimate is if repeated sampling, at the same size as the analyzed sample, were carried out from the studied population. With plausible models (see following section on fit indices), the standard errors are used to compute t values that provide information about statistical significance of the associated unknown parameter. The t values are computed by the software as the ratio of parameter estimate to its standard error. If for a free parameter, for example, its t value is greater than $+2$ or less than -2, the parameter is referred to as significant at the used significance level (typically .05) and can be considered distinct from 0 in the population. Conversely, if its t value lies between $+2$ and -2, the parameter is nonsignificant and may be considered 0 in the population. Furthermore, based on the large-sample normality feature of parameter estimates, adding 1.96 times the standard error to and subtracting 1.96 times the standard error from the parameter estimate yields a confidence interval (at the 95% confidence level) for that parameter. This confidence interval represents a range of plausible values for the unknown parameter in the population, and can be conventionally used to test hypotheses about a prespecified value of that parameter there. In particular, if this interval covers the preselected value, the hypothesis that the parameter equals it in the population can be retained (at significance level .05); otherwise, this hypothesis can be rejected. Moreover, the width of the interval permits one to assess the precision with which the parameter has been estimated. Wider confidence intervals are associated with lower precision (and larger standard errors), and narrower confidence intervals go together with higher precision of estimation (and smaller standard errors). In fact, these features of the standard errors as measures of sampling variability of parameter estimates make them quite useful—along with a number of model goodness-of-fit indices discussed in the next section—for purposes of assessing goodness of fit of a considered model (see below).

We note that a reason for the numerical procedure of fit function minimization not to converge could be that the proposed model may simply be misspecified. Misspecified models are inadequate for the analyzed data, that is, they contain unnecessary parameters or omit important ones (and/or such variables); in terms of path diagrams, misspecified models contain wrong paths or two-way arrows, and/or omit important one-way paths or covariances. Another reason for lack of convergence is lack of

model identification. A model is not identified, frequently also referred to as unidentified, if it possesses one or more unidentified parameters. These are parameters that cannot be uniquely estimated—unlike identified parameters—even if one gathers data on the analyzed variables from the whole population and calculates their population covariance matrix to which the model is subsequently fitted. It is therefore of utmost importance to deal only with identifiable parameters, and thus make sure one is working with an identified model. Due to its special relevance for SEM analyses, the issue of parameter and model identification is discussed next.

PARAMETER AND MODEL IDENTIFICATION

As indicated earlier in this chapter, a model parameter is unidentified if there is not enough empirical information to allow its unique estimation. In that case, any estimate of an unidentified parameter computed by a SEM program is arbitrary and should not be relied on. A model containing at least one unidentified parameter cannot generally be relied on either, even though some parts of it may represent a useful approximation to the studied phenomenon. Since an unidentified model is generally useless in empirical research (although it may be useful in some theoretical discussions), one must ensure the positive identification status of a model, which can be achieved by following some general guidelines.

What Does It Mean to Have an Unidentified Parameter?

In simple terms, having an unidentified parameter implies that it is impossible to compute a defendable estimate of it. For example, suppose one considers the equation $a + b = 10$, and is faced with the task of finding unique values for the two unknown constants a and b. One solution obviously is $a = 5, b = 5$, yet another is also $a = 1, b = 9$. Evidently, there is no way in which one can determine unique values for a and b satisfying the above equation, because one is given a single equation with two unknowns. For this reason, there is an infinite number of solutions to the equation since there are more unknown values (viz. two parameters) than known values (one equation, i.e., one known number—namely 10—related to the two parameters). Hence, the 'model' represented by this equation is underidentified, or unidentified, and any pair of estimates that could be obtained for a and b—like the above two pairs of values for them—are just as legitimate as any other pair satisfying the equation and hence there is no reason to prefer any of its infinitely many solutions. As will become clear later, the most desirable condition to encounter in SEM is to have more equations than are needed to obtain unique solutions for the parameters; this condition is called overidentification.

A similar identification problem occurs when the only information available is that the product of two unknown parameters λ and ϕ, say, is equal to 55. Knowing only that $\lambda\phi = 55$ does not provide sufficient information to come up with unique estimates of either λ or ϕ. Obviously, one could choose a value of λ in a completely arbitrary manner (except, of course, 0 for this example) and then take $\phi = 55/\lambda$ to provide one solution. Since there are an infinite number of solutions for λ and ϕ, neither of these two unknowns is identified until some further information is provided, for example a given value for one of them.

The last product example is in fact quite relevant in the context of our earlier consideration of Rule 6 for determining model parameters. Indeed, if neither the variance ϕ of a latent variable nor a path λ going out of it are fixed to a constant, then a situation similar to the one just discussed is created in any routinely used structural equation model, with the end result that the variance of the latent variable and the factor loading for a given indicator of it become entangled in their product and hence unidentified.

Although the two numerical examples in this section were deliberately simple, they nonetheless illustrate the nature of similar problems that can occur in the context of structural equation models. Recall from earlier sections that SEM can be thought of as an approach to solving, in an optimal way, a system of equations—those relating the elements of the sample covariance matrix S with their counterparts in the model reproduced covariance matrix $\Sigma(\gamma)$. It is possible then, depending on the model, that for some of its parameters the system may have infinitely many solutions. Clearly, such a situation is not desirable given the fact that SEM models attempt to estimate what the parameter values would be in the population. Only identified models, and in particular estimates of identified parameters, can provide this type of information; models and parameters that are unidentified cannot supply such information. A straightforward way to determine if a model is unidentified is presented next.

A Necessary Condition for Model Identification

The above introduced parallel between SEM and solving a system of equations is also useful for understanding a simple and necessary condition for model identification. Specifically, if the system of equations relating the elements of the sample covariance matrix S with their counterparts in the model-implied covariance matrix $\Sigma(\gamma)$ contains more parameters than equations, then the model will be unidentified. This is because that system has more unknowns than could possibly be uniquely solved for, and hence for at least some of them there are infinitely many solutions. Although this condition is easy to check, it should be noted that it is not a suf-

ficient condition. That is, having less unknown parameters than equations in that system does not guarantee that a model is identified. However, if a model is identified then this condition must hold, i.e., there must be as many or fewer parameters in that system of equations than nonredundant elements of the empirical covariance matrix S.

To check the necessary condition for model identification, one simply counts the number of (independent) parameters in the model and subtracts it from the number of nonredundant elements in the sample covariance matrix S, i.e., from $p(p+1)/2$ where p is the number of observed variables to which the model is fitted. The resulting difference,

$$p(p + 1)/2 - (\text{Number of model parameters}), \tag{6}$$

is referred to as *degrees of freedom* of the considered model, and is usually denoted *df*. If this difference is positive or zero, the model degrees of freedom are nonnegative. As indicated above, having nonnegative degrees of freedom represents a necessary condition for model identification.

If the difference in Equation 6 is 0, however, the degrees of freedom are 0 and the model is called saturated. *Saturated models* have as many parameters as there are nonredundant elements in the sample covariance matrix. In such cases, there is no way that one can test or disconfirm the model. This is because a saturated model will always fit the data perfectly, since it has just as many parameters as there are nonredundant elements in the empirical covariance matrix S. The number of parameters then equals that of nonredundant equations in the above mentioned system obtained when one sets equal the elements of S to their counterpart entries in the model reproduced covariance matrix $\Sigma(\gamma)$, which system therefore has a unique solution.

This lack of testability for saturated models is not unique to the SEM context, and in fact is analogous to the lack of testability of any other statistical hypothesis when pertinent degrees of freedom equal zero. For example, in an analysis of variance design with at least two factors and a single observation per cell, the hypothesis of factorial interaction is not testable because its associated degrees of freedom equal 0. In this situation, the interaction term for the underlying analysis of variance model (decomposition) is confounded with the error term and cannot be disentangled from it. For this reason, the hypothesis of interaction cannot be tested, as a result of lack of empirical information bearing upon the interaction term. The same sort of lack of empirical information renders a saturated structural equation model untestable as well.

Conversely, if the difference in Equation 6 is negative, the model degrees of freedom are negative. In such cases, the model is unidentified since it violates the necessary condition for identification. Then the system of equa-

tions relating the elements of S to their counterpart entries of the implied covariance matrix $\Sigma(\gamma)$ contains more unknown than equations. Such systems of equations, as is well known from elementary algebra, do not possess unique solutions, and hence the corresponding structural equation model is not associated with a unique set of values for its parameters that could be obtained from the sample data, no matter how large the sample size is. This is the typical deficiency of unidentified structural equation models, which renders them in general useless in practice.

From this discussion it follows that since one of the primary reasons for conducting a SEM analysis is to test the fit of a proposed model, the latter must have positive degrees of freedom. That is, there must be more nonredundant elements in the data covariance matrix than unknown model parameters. This insures that there are over-identifying restrictions placed on the model parameters, which are obtained when the elements of S are set equal to the corresponding entries of $\Sigma(\gamma)$. In such cases, it becomes of interest to evaluate to what extend these restrictions are consistent with the data. This evaluation is the task of the process of model fit evaluation.

Having said that, we stress that as mentioned earlier the condition of non-negative degrees of freedom is only a necessary but not a sufficient condition for model identification. As a matter of fact, there are many situations in which the degrees of freedom for a model are positive and yet some of its parameters are unidentified. Hence, passing the check for non-negative degrees of freedom, which sometimes is also referred to as the *t*-rule for model identification, does not guarantee identification and that a model could be a useful means of description and explanation of a studied phenomenon. Model identification is in general a rather complex issue that requires careful consideration and handling, and is further discussed next.

How to Deal with Unidentified Parameters in Empirical Research?

If a model under consideration is carefully conceptualized, the likelihood of unidentified parameters will usually be minimized. In particular, using Rules 1 to 6 will most likely ensure that the proposed model is identified. However, if a model is found to be unidentified, a first step toward identification is to see if all its parameters have been correctly determined or whether all the latent variables have their scales fixed. In many instances, a SEM program will signal an identification problem with an error message and even correctly point to the unidentified parameter, but in some cases the software may point to the wrong parameter or even may miss an unidentified parameter and model. Hence, the best strategy is for the researcher to examine the issue of model identification and locate the unidentified parameter(s) in an unidentified model, rather than rely com-

pletely on a SEM program. This could be accomplished on a case-to-case basis, by trying to solve the above system of equations obtained when elements of the empirical covariance matrix S are set equal to their counterpart entries in the model implied covariance matrix $\Sigma(\gamma)$. If for at least one model parameter there are infinitely many solutions possible from the system, that parameter is not identified and the situation needs to be rectified (see below). If none of the parameters is unidentified, all of them are identified and the model is identified as well.

Once located, as a first step one should see if the unidentified parameter is a latent variance or factor loading, since omitting to fix the scale of a latent variable will lead to lack of identification of at least one of these two parameters pertaining to that variable. When none of the unidentified parameters results because of such an omission, a possible way of dealing with these parameters is to impose appropriate, substantively plausible constraints on them or functions of them that potentially involve also other parameters. This way of attempting model identification may not always work, in part because of lack of such constraints. In those cases, either a completely new model may have to be contemplated—one that is perhaps simpler and does not contain unidentified parameters—or a new study and data-collection process may have to be designed.

MODEL-TESTING AND -FIT EVALUATION

The SEM methodology offers researchers a method for quantification and testing of theories. Substantive theories are often representable as models that describe and explain phenomena under investigation. As discussed previously, an essential requirement for all such models is that they be identified. Another requirement, one of no lesser importance, is that researchers consider for further study only those models that are meaningful from a substantive viewpoint and present plausible means of data description and explanation.

SEM provides a number of inferential and descriptive indices that reflect the extent to which a model can be considered an acceptable means of data representation. Using them together with substantive considerations allows one to make a decision whether a given model should reasonably be rejected as a means of data explanation or could be tentatively relied on (to some extent). The topic of structural equation *model fit evaluation*, i.e., the process of assessing the extent to which a model fits an analyzed data set, is very complex and in some aspects not necessarily uncontroversial. Due to the introductory nature of this book, this section develops a relatively brief discussion of the underlying issues, which can be considered a minimalist scheme for carrying out model evaluation. For further elaboration on these issues, we refer the reader to Bentler (2004), Bollen (1989), Byrne (1998), Jöreskog and Sörbom (1993a, 1993b), Muthén & Muthén (2004).

Substantive Considerations

A major aspect of fit evaluation involves the substantive interpretations of results obtained with a proposed model. All models considered in research should first be conceptualized according to the latest knowledge about the phenomenon under consideration. This knowledge is usually obtained via extensive study of the pertinent literature. Fitted models should try to embody in appropriate ways the findings available from previous studies. If a researcher wishes to critically examine aspects of available theories, alternative models with various new restrictions or relationships between involved variables can also be tested. However, the initial conceptualization of any proposed model can only come after an informed study of the phenomenon under consideration that includes also a careful study of past research and accumulated knowledge in the respective subject-matter domain.

Regardless of the specifics of a model in this regard, the advantages of the SEM methodology can only be used with variables that have been validly and reliably assessed. Even the most intricate and sophisticated models are of no use if the variables included in the model are poorly assessed. A model cannot do more than what is contained in the data themselves. If the data are poor, in the sense of reflecting substantial unreliability in assessing aspects of a studied phenomenon, the results will be poor, regardless of the particulars of used models.

Providing an extensive discussion of the various ways of ensuring satisfactory measurement properties of variables included in structural equation models is beyond the scope of this introductory book. These issues are usually addressed at length in books dealing specifically with psychometrics and measurement theory (e.g., Allen & Yen, 1979; Crocker & Algina, 1986; Suen, 1990). The present text instead assumes that the researcher has sufficient knowledge of how to organize a reliable and valid assessment process for the variables included in a given model.

Model Evaluation and the True Model

Before particular indexes of model fit are discussed, a word of warning is in order. Even if all fit indexes point to an acceptable model, one cannot claim in empirical research to have found the true model that has generated the analyzed data. (The cases in which data are simulated according to a known model are excluded from this consideration.) This fact is related to another specificity of SEM that is different from classical modeling approaches. Whereas classical methodology is typically interested in rejecting null hypotheses because the substantive conjecture is usually reflected in the alternative rather than null hypotheses (e.g., alternative hypotheses of difference or change), SEM is pragmatically concerned with finding a model that does not contradict the data. That is, in an empirical SEM session, one is typically

interested in retaining a proposed model whose validity is the essence of a pertinent null hypothesis. In other words, statistically speaking, when using SEM one is usually 'interested' in not rejecting the null hypothesis.

However, recall from introductory statistics that not rejecting a null hypothesis does not mean that it is true. Similarly, because model testing in SEM involves testing the null hypothesis that the model is capable of perfectly reproducing with certain values of its unknown parameters the population matrix of observed variable interrelationship indices, not rejecting a fitted model does not imply that it is the true model. In fact, it may well be that the model is not correctly specified (i.e., wrong), yet due to sampling error it appears plausible. Similarly, just because a model fits a data set well does not mean that it is the only model that fits the data well or nearly as well. There can be a plethora of other models that fit the data equally well, better, or only marginally worse. In fact, there can be a number (possibly very many; e.g., Raykov & Marcoulides, 2001; Raykov & Penev, 1999) of equivalent models that fit the data just as well as a model under consideration. Unfortunately, at present there is no statistical means for discriminating among these equivalent models—especially when the issue is choosing one (or more) of them for further consideration or interpretation. Which one of these models is better and which one is to be ruled out, can only be decided on the basis of a sound body of substantive knowledge about the studied phenomenon. This is partly the reason why substantive considerations are so important in model-fit evaluation. In addition, one can also evaluate the validity of a proposed model by conducting replication studies. The value of a given model is greatly enhanced if it can be replicated in new samples from the same studied population.

Parameter Estimate Signs, Magnitude, and Standard Errors

It is worth reiterating at this point that one cannot in general meaningfully interpret a model solution provided by SEM software if the underlying numerical minimization routine has not converged, that is has not ended after a finite number of iterations. If this routine does not terminate, one cannot trust the program output for purposes of solution interpretation (although the solution may provide information that is useful for tracking down the reasons for lack of convergence).

For a model to be considered further for fit evaluation, the parameter estimates in the final solution of the minimization procedure should have the right sign and magnitude as predicted or expected by available theory and past research. In addition, the standard errors associated with each of the parameter estimates should not be excessively large. If a standard error of a parameter estimate is very large, especially when compared to other parameter estimate standard errors, the model does not provide reliable information with regard to that parameter and should be interpreted with great

caution; moreover, the reasons for this finding should be clarified before further work with the model is undertaken.

Goodness-of-Fit Indices

The Chi-Square Value Evaluation of model fit is typically carried out on the basis of an inferential goodness-of-fit index as well as a number of other descriptive or alternative indices. This inferential index is the so-called chi-square value. The index represents a test statistic of the goodness of fit of the model, and is used when testing the null hypothesis that the model fits the corresponding population covariance matrix perfectly. This test statistic is defined as

$$T = (N - 1)\, F_{\min}\,, \tag{7}$$

where N is the sample size and $F_{\min}$ denotes the minimal value of the fit function for the parameter estimation method used (e.g., ML, GLS, ADF).

The name chi-square value derives from the fact that with large samples the distribution of T approaches a chi-square distribution if the model is correct and fitted to the covariance matrix S. This large-sample behavior of T follows from what is referred to as likelihood ratio theory in statistics (e.g., Johnson & Wichern, 2002). The test statistic in (7) can be obtained in the context of comparing a proposed model with a saturated model, which as discussed earlier fits perfectly the data, using the so-called likelihood ratio. After multiplication with –2, the distribution of the logarithm of this ratio approaches with increasing sample size a chi-square distribution under the null hypothesis (e.g., Bollen, 1989), a result that for our purposes is equivalent to a chi-square distribution of the right-hand side of Equation (7).

The degrees of freedom of this limiting chi-square distribution are equal to those of the model. As mentioned previously, they are determined by using the formula $df = (p(p+1)/2) - q$, where p is the number of observed variables involved in the model and q is the number of model parameters (see Equation 6 in section "Parameters and Model Identification").

When a considered model is fit to data using SEM software, the program will judge the obtained chi-square value T in relation to the model's degrees of freedom, and output its associated p-value. This p-value can be examined and compared with a preset significance level (often .05) in order to test the null hypothesis that the model is capable of exactly reproducing the population matrix of observed variable relationship indices. Hence, following the statistical null hypothesis testing tradition, one may consider rejection of the model when this p-value is smaller than a preset significance value (e.g., .05), and alternatively retention of the model if this p-value is higher than that significance level.

Although this way of looking at statistical inference in SEM may appear to be the reverse of the one used within the framework of traditional hypothesis testing, as exemplified with the framework of the general linear model, it turns out that at least from a philosophy-of-science perspective the two are compatible. Indeed, following Popperian logic (Popper, 1962), one's interest lies in rejecting models rather than confirming them. This is because there is in general no scientific way of proving the validity of a given model. That is, in empirical research no structural equation model can be proved to be the true model (see discussion in previous section).

In this context, it is also important to note that in general there is a preference for dealing with models that have a large number of degrees of freedom. This is because an intuitive meaning of the notion of degree of freedom is as a dimension along which a model can be disconfirmed. Hence, the more degrees of freedom a model has, the more dimensions there are along which one can reject the model, and hence the higher the likelihood of disconfirming it when it is tested against data. This is a desirable feature of the testing process because, according to Popperian logic, empirical science can only disconfirm and not confirm models. Therefore, if one has two models that are plausible descriptions of a studied phenomenon, the one with more degrees of freedom is a stronger candidate for consideration as a means of data description and explanation. The reason is that the model with more degrees of freedom has withstood a greater chance of being rejected; if the model was not rejected then, the results obtained with it may be viewed as more trustworthy. This reasoning is essentially the conceptual basis of the parsimony principle widely discussed in the SEM literature (e.g., Raykov & Marcoulides, 1999, and references therein). Hence, Popperian logic, which maintains that a goal of empirical science is to formulate theories that are falsifiable, is facilitated by an application of the parsimony principle. If a more parsimonious model is found to be acceptable, then one may also place more trust in it because it has withstood a higher chance of rejection than a less parsimonious model. However, researchers are cautioned that rigid and routine applications of the parsimony principle can lead to conclusions favoring an incorrect model and implications that are incompatible with those of the correct model. (For a further discussion, see Raykov & Marcoulides, 1999.)

The chi-square value T has received a lengthy discussion in this section for two reasons. First, historically and traditionally, it has been the index that has attracted a great deal of attention over the past 40 years or so and especially in the 1970s through 1990s. In fact, most of the fit indexes devised over the past several decades in the SEM literature are functions of the chi-square value. Second, the chi-square value has the important feature of being an inferential fit index. That is to say, by using it one is in a position to make a generalization about the fit of the model in a studied population.

This is due to the fact that the large-sample distribution of T is known—namely central chi-square, when a correct model is fitted to the covariance matrix—and a p-value can be attached to each particular sample's value of T. This feature is not shared by most of the other goodness of fit indexes.

However, it may not be always wise to strictly follow this statistical evaluation of plausibility of a model using the chi-square value T, due to the fact that with very large samples T cannot be really relied on. The reason is readily seen from its definition in Equation 7. Since the value of T is obtained by multiplying $N - 1$ (sample size less 1) by the attained minimum of the fit function, increasing the sample size typically leads to an increase in T as well. Yet the model's degrees of freedom remain the same because the model has not changed, and hence so does the reference chi-square distribution against which T is judged for significance. Consequently, with very large samples there is a spurious tendency to obtain large values of T, which tend to be associated with small p-values. Therefore, if one were to use only the chi-square's p-value as an index of model fit, there will be an artificial tendency with very large samples to reject models even if they were only marginally inconsistent with the analyzed data.

Alternatively, there is another spurious tendency with small samples for the test statistic T to remain small, which is also explained by looking at the above Equation (7) and noting that the multiplier $N - 1$ is then small. Hence, with small samples there is a tendency for the chi-square fit index to be associated with large p-values, suggesting a considered model as a plausible data-description means. Thus, the chi-square index and its p-value alone cannot be fully trusted in general as means for model evaluation. Other fit indices must also be examined in order to obtain a better picture of model fit.

Descriptive Fit Indices The above limitations of the chi-square value indicate the importance of the availability of other fit indices to aid the process of model evaluation. A number of descriptive-fit indices have been proposed mostly in the 1970s and 80s that provide a family of fit measures useful in the process of assessing model fit.

The first developed descriptive-fit index is the goodness-of-fit index (GFI). It can be loosely considered a measure of the proportion of variance and covariance that a given model is able to explain. The GFI may be viewed as an analog in the SEM field of the widely used R^2 index in regression analysis. If the number of parameters is also taken into account in computing the GFI, the resulting index is called adjusted goodness-of-fit index (AGFI). Its underlying logic is similar to that of the adjusted R^2 index also used in regression analysis. The GFI and AGFI indexes range between 0 and 1, and are usually fairly close to 1 for well-fitting models. Unfortunately, as with many other descriptive indices, there are no strict norms for the GFI and AGFI below which a model cannot be considered a plausible description of the

analyzed data and above which one could rest assured that the model approximates the data reasonably well. As a rough guide, it may be suggested that models with a GFI and AGFI in the mid-.90s or above may represent a reasonably good approximation of the data (Hu & Bentler, 1999).

There are two other descriptive indices that are also very useful for model-fit evaluation purposes. These are the normed fit index (NFI) and the non-normed fit index (NNFI) (Bentler & Bonnet, 1980). The NFI and NNFI are based on the idea of comparing the proposed model to a model in which no interrelationships at all are assumed among any of the variables. The latter model is referred to as the independence model or the null model, and in some sense may be seen as the least attractive, or "worst", model that could be considered as a means of explanation and description of one's data. The name independence or null model derives from the fact that this model assumes the variables only have variances but that there are no relationships at all among them, that is, all their covariances are zero. Thus, the null model represents the extreme case of no relationships among the studied variables, and interest lies in comparing a proposed model to the corresponding null model. When the chi-square value of the null model is compared to that of a model under consideration, one gets an idea of how much better the model of concern fits the data relative to how bad a means of data description and explanation that model could possibly be. This is the basic idea that underlies the NFI and NNFI descriptive-fit indices.

The NFI is computed by relating the difference of the chi-square value for a proposed model to the chi-square value for the independence or null model. The NNFI is a variant of the NFI that takes into account also the degrees of freedom of the proposed model. This is done in order to account for model complexity, as reflected in the degrees of freedom. The reason this inclusion is meaningful, is that for a given data set more complex models have more parameters and hence fewer degrees of freedom, whereas less complex models have less parameters and thus more degrees of freedom. Therefore, one can consider degrees of freedom as an indicator of complexity of a model (given a set of observed variables to which it is fitted).

Similar to the GFI and AGFI, models with NFI and NNFI close to 1 are considered to be more plausible means of describing the data than models for which these indices are further away from 1. Unfortunately, once again, there are no strict norms above which one can consider the indices as supporting model plausibility and below which one can safely reject the model. As a rough guide, models with NNFI and NFI in the mid-.90s or higher are viewed likely to represent reasonably good approximations to analyzed data (Hu & Bentler, 1999).

In addition to the GFI, AGFI, NNFI, and NFI, there are more than a dozen other descriptive-fit indices that have been proposed in the SEM literature

over the past 30 years or so. Despite this plethora of descriptive-fit indices, most of them are directly related to the chi-square value T and represent reexpressions of it or its relationships to other models' chi-square values and related quantities. The interested reader may refer to more advanced SEM books that provide mathematical definitions of each of these indexes (e.g., Bollen, 1989) as well as the program manuals for EQS, LISREL, and M*plus* (Bentler, 2004; Jöreskog & Sörbom, 1993a; Muthén & Muthén, 2004).

Alternative Fit Indices A family of alternative fit indices are based on an altogether different conceptual approach to the process of hypothesis testing in SEM, which can be referred to as an alternative approach to model assessment. These indices have been developed over the past 25 years and largely originate from an insightful paper by Steiger and Lind (1980). The basis for alternative-fit indices is the noncentrality parameter (NCP), denoted δ. The NCP basically reflects the extent to which a model does not fit the data. For example, if a model is correct and the sample large, the test statistic T presented in Equation 7 follows a (central) chi-square distribution, but if the model is not quite correct, i.e., is misspecified to a small degree, then T follows a noncentral chi-square distribution. As an approximation, a noncentral chi-square distribution can roughly be thought of as resulting when the central chi-square distribution is shifted to the right by δ units (and its variance correspondingly enlarged). In this way, the NCP can be viewed as an index reflecting the degree to which a model under consideration fails to fit the data. Thus, the larger the NCP, the worse the model; and the smaller the NCP, the better the model. It can be shown that with not-too-misspecified models, normality, and large samples, δ approximately equals $(N-1)F_{\mathrm{ML},0}$, where $F_{\mathrm{ML},0}$ is the value of the maximum likelihood fit function when the model is fit to the population covariance matrix. The NCP is estimated in a given sample by $\hat{\delta} = T - d$ if $T \geq d$, and by 0 if $T < d$, where for simplicity d denotes the model degrees of freedom.

Within the alternative approach to model testing, the conventional null hypothesis that a proposed model perfectly fits the population covariance matrix is relaxed. This is explained by the observation that in practice every model is wrong even before it is fitted to data. Indeed, the reason why a model is used when studying a phenomenon of interest is that the model should represent a useful simplification and approximation of reality rather then be a precise replica of it. That is, by its very nature, a model cannot be correct because then it would have to be an exact copy of reality and therefore useless. Hence, in the alternative approach to model testing the conventional null hypothesis of perfect model fit that has been traditionally tested in SEM by examining the chi-square index and its p-value, is really of

no interest. Instead, one is primarily concerned with evaluating the extent to which the model fails to fit the data. Consequently, for the reasonableness of a model as a means of data description and explanation, one should impose weaker requirements for degree of fit.

This is the logic of model testing that is followed by the so-called root mean square error of approximation (RMSEA) index that has recently become quite a popular index of model fit. In a given sample, the RMSEA is evaluated as

$$\pi = \sqrt{(\mathrm{T} - \mathrm{d}) / (\mathrm{dn})} \tag{8}$$

when $T \geq d$, or as 0 if $T < d$, where $n = N - 1$ is sample size less 1. The RMSEA, similar to other fit indices, also takes into account model complexity, as reflected in the degrees of freedom. It has been suggested that a value of the RMSEA of less than .05 is indicative of the model being a reasonable approximation to the analyzed data (Browne & Cudeck, 1993). Some research has found that the RMSEA is among the fit indices least affected by sample size; this feature sets the RMSEA apart from many other fit indices that are sample-dependent or have characteristics of their distribution, such as the mean, depending on sample size (Marsh et al., 1996; Bollen, 1989).

The RMSEA is not the only index that can be obtained as a direct function of the noncentrality parameter. The comparative-fit index (CFI) also follows the logic of comparing a proposed model with the null model assuming no relationships between the observed measures (Bentler, 1990). The CFI is defined as the ratio of improvement in noncentrality when moving from the null to a considered model, to the noncentrality of the null model. Typically, the null model has considerably higher noncentrality than a proposed model because the former could be expected to fit the data poorly. Hence, values of CFI close to 1 are considered likely to be indicative of a reasonably well-fitting model. Again, there are no norms about how high the CFI should be in order to safely retain or reject a given model. CFI's in the mid-.90s or above are usually associated with models that are plausible approximations of the data.

The expected cross-validation index (ECVI) was also introduced as a function of the noncentrality parameter (Browne & Cudeck, 1993). The ECVI represents a measure of the degree to which one would expect a given model to replicate in another sample from the same population. In a set of several proposed models for the same studied phenomenon, a model is preferred if it minimizes the value of ECVI relative to the other models. The ECVI was developed partly as a reaction to the fact that because the RMSEA is only weakly related to sample size, it cannot account for the fact that with small samples it would be unwise to fit a very complex

model (i.e., one with many parameters). The ECVI accounts for this possibility, and when the maximum likelihood method of estimation is used it will be identical up to a multiplicative constant to the Akaike information criterion (AIC). The AIC is a special type of fit index that takes into account both the measure of fit and model complexity (Akaike, 1987), and resembles the so-called Bayesian information criterion (BIC). The two indices, AIC and BIC, are widely used in applied statistics for purposes of model comparison. Generally, models with lower values of ECVI, AIC, and BIC are more likely to be better means of data description than models with higher such indexes. The ECVI, AIC, and BIC have become quite popular in SEM and latent variable modeling applications, particularly for the purpose of examining competing models, i.e., when a researcher is considering several models and wishes to select from them the one with best fit. According to these indices, models with smaller values on them are preferred to models with higher values.

Another important feature of this alternative approach to model assessment involves the routine use of confidence intervals, and specifically for the noncentrality parameter and RMSEA. Recall from basic statistics that a confidence interval provides a range of plausible values for the population parameter being estimated, at a given confidence level. The width of the interval is also indicative of the precision of estimation of the parameter using the data at hand. Of special interest to the alternative approach of model testing is the left endpoint of the 90% confidence interval of the RMSEA index for an entertained model. In particular, if this endpoint is considerably smaller than .05 and the interval not too wide (e.g., the right endpoint not higher than .08), it can be argued that the model is a plausible means of describing the analyzed data. Hence, if the RMSEA is smaller than .05 or the left endpoint of its confidence interval markedly smaller than .05, with this interval being not excessively wide, the pertinent model could be considered a reasonable approximation of the analyzed data.

In conclusion of this section on model testing and fit evaluation, we would like to emphasize that no decision on goodness of fit should be based on a single index, no matter how favorable for the model that index may appear. As indicated earlier, every index represents a certain aspect of the fit of a proposed model, and in this sense is a source of limited information as to how good the model is or how well it can be expected to perform in the future (e.g., on another sample from the same population). Therefore, a decision to reject or retain a model should always be based on multiple goodness-of-fit indices (and if possible on the results of replication studies). In addition, as indicated in the next section, important insights regarding model fit can be sometimes obtained by also conducting an analysis of residuals.

Analysis of Residuals

All fit indices discussed in the previous section should be viewed as overall measures of model fit. In other words, they are summary measures of fit and none of them provides information about the fit of individual parts of the model. As a consequence, it is possible for a given model to be seriously misspecified in some parts (i.e., incorrect with regard to some of the variables and their relationships) but very well fitting in others, so that an evaluation of the previously discussed fit criteria suggests that the model may be judged plausible.

For example, consider a model that is substantially off the mark with respect to an important relationship between two particular observed variables (e.g., the model omits this relationship). In such a case, the difference between the sample covariance and the covariance reproduced by the model at the final solution—called residual for that pair of variables—may be substantial. This result would suggest that the model cannot be considered a plausible means of data description. However, at the same time the model may do an excellent job of explaining all of the remaining covariances and variances in the sample covariance matrix S, and overall result in a nonsignificant chi-square value and favorable descriptive as well as alternative fit indices. Such an apparent paradox may emerge because the chi-square value T, or any of the other fit indices discussed above, is a measure of overall fit. Hence, all that is provided by overall measures of model fit is a summary picture of how well a model fits the entire analyzed matrix, but no information is contained in them about how well the model reproduces the individual elements of that matrix.

To counteract this possibility, the so-called covariance residuals—often also referred to as model residuals—can be examined. There are as many generic residuals of this kind as there are nonredundant elements of the sample covariance matrix of the variables to which a model is fitted. They result from an element-wise comparison of each sample variance and covariance to the value of its counterpart element in the implied covariance matrix obtained with the parameter estimates when the model is fitted to data. In fact, there are two types of model residuals that can be examined in most SEM models and are provided by the used software. The unstandardized residuals index the amount of unexplained variable covariance in terms of the original metric of the raw data. However, if this metric is quite different across measured variables, it is impossible to examine meaningfully these residuals and determine which are large and which are small. A standardization of the residuals to a common metric, as reflected in the standardized residuals, makes this comparison much easier.

A standardized residual above 2 generally indicates that the model considerably underexplains a particular relationship between two variables.

Conversely, a standardized residual below –2 generally indicates that the model markedly overexplains the relationship between the two variables. Using this residual information, a researcher may decide to either add or remove some substantively meaningful paths or covariance parameters, which could contribute to a smaller residual associated with the involved two variables and hence a better-fitting model with regard to their relationship (see Appendix to this chapter).

Overall, good-fitting models will typically exhibit a steam-and-leaf plot of standardized residuals that closely resembles a symmetric distribution. In addition, examining the so-called *Q* plot of the standardized residuals is a useful means of checking the plausibility of a proposed model. The *Q* plot graphs the standardized residuals against their expectations if the model were a good means of data description. With well-fitting models, a line drawn through the marks of the residuals on that plot will be close to the dotted, equidistant line provided on it by the software (Jöreskog & Sörbom, 1993c). Marked departures from a straight line indicate serious model misspecifications or possibly violations of the normality assumption (e.g., nonlinear trends in the relationships between some observed variables).

An important current limitation in SEM applications is the lack of evaluation of estimated individual-case residuals. Individual-case residuals are routinely used in applications of regression analysis because they help researchers with model evaluation and modification. In regression analysis, residuals are defined as the differences between individual raw data and their model-based predictions. Unfortunately, SEM developers have only recently begun to investigate more formally ways in which individual-case residuals can be defined within this framework (e.g., Bollen & Arminger, 1991; Raykov & Penev, 2001). The development of means for defining individual-case residuals is also hampered by the fact that most structural equation models are based on latent variables, which cannot be directly observed or precisely measured. Therefore, very important pieces of information that are needed in order to arrive at individual-case residuals similar to those used in regression analysis are typically missing.

Modification Indices

A researcher usually conducts a SEM analysis by fitting a proposed model to available data. If the model does not fit, one may accept this fact and leave it at that (which is not really commonly recommended), or alternatively may consider answering the question, "How could the model be altered in order to improve its fit?" In the SEM literature, the modification of a specified model with the aim of improving fit has been termed a specification search (Long, 1983; MacCallum, 1986). Accordingly, a specification search is conducted with the intent to detect and correct specification error in a pro-

posed model, that is, its deviation from the true model characterizing a studied population and relationships among analyzed variables.

Although in theory researchers should fully specify and deductively hypothesize a model prior to data collection and model testing, in practice this often may not be possible, either because a theory is poorly formulated or because it is altogether nonexistent. As a consequence, specification searches have become nearly a common practice in many SEM applications. In fact, most currently available SEM programs provide researchers with options to conduct specification searches to improve model fit, and some new search procedures (e.g., using genetic algorithms, ant colony optimization, and Tabu search) have also been developed to automate this process (see Marcoulides, Drezner, & Schumacker, 1998; Marcoulides & Drezner, 2001, 2002; Scheines, Spirtes, Glymour, Meek & Richardson, 1998).

Specification searches are clearly helpful for improving a model that is not fundamentally misspecified but is incorrect only to the extent that it has some missing paths or some of its parameters are involved in unnecessarily restrictive constraints. With such models, it can be hypothesized that their unsatisfactory fit stems from overly strong restriction(s) on its parameters that are either fixed to 0 or set equal to other parameter(s), or included in a more complex relationship. The application of any means of model improvement is only appropriate when the model modification suggested is theoretically sound and does not contradict the results of previous research in a particular substantive domain. Alternatively, the results of any specification search that do not agree with past research should be subjected to further analysis based on new data before any real validity can be claimed.

The indexes that can be used as diagnostic statistics about which parameters could be changed in a model are called modification indices (a term used in the LISREL and M*plus* programs) or Lagrange multiplier test statistics (a term used in the EQS program). The value of a *modification index* (term is used generically) indicates approximately how much a proposed model's chi-square would decrease if a particular parameter were freely estimated or freed from a constraint it was involved in the immediately preceding modeling session. There is also another modification index, called the Wald index, which takes an alternative approach to the problem. The value of the *Wald index* indicates how much a proposed model's chi-square would increase if a particular parameter were fixed to 0 (i.e., if the parameter were dropped from a model under consideration).

The modification indexes address the question of how to improve an initially specified model that does not fit satisfactorily the data. Although no strict rules-of-thumb exist concerning how large these indexes must be to warrant a meaningful model modification, based on purely statistical considerations one might simply consider making changes to parameters asso-

ciated with the highest modification indices (see below for a guideline regarding their magnitude). If there are several parameters with high modification indices, one may consider freeing them one at a time, beginning with the largest, because like in the general linear modeling framework a single change in a structural equation model can affect other parts of the solution (Jöreskog & Sörbom, 1990; Marcoulides et al., 1998). When LISREL or M*plus* are used, modification indices larger than 5 generally merit close consideration. Similarly, when EQS is used, parameters associated with significant Lagrange-multiplier or Wald-index statistics also deserve close consideration.

It must be emphasized, however, that any model modification must first be justified on theoretical grounds and be consistent with already available theories or results from previous research in the substantive domain under consideration, and only second must be in agreement with statistical optimality criteria such as those mentioned. Blind use of modification indices can turn out to be a road to models that lead researchers astray from their original substantive goals. It is therefore imperative to consider changing only those parameters that have a clear substantive interpretation. Additional statistics, in the form of the estimated change for each parameter, can also be taken into account before one reaches a final decision regarding model modification.

In conclusion, we emphasize that results obtained from any model-improvement specification search may be unique to the particular data set, and that capitalization on chance can occur during the search (e.g., MacCallum, 1986). Consequently, once a specification search is conducted a researcher is entering a more exploratory phase of analysis. This has also purely statistical implications in terms of not keeping the overall significance level at the initially prescribed nominal value (the preset significance level, usually .05). Hence, the possibility exists of arriving at such statistically significant results regarding aspects of the model due only to chance fluctuations. Thus, the likelihood of falsely declaring at least one of the conducted statistical tests of the model or any of its parameters to be significant, is increased rather than being the same as that of any single test. For this reason, any models that result from specification searches must be cross-validated before real validity can be claimed for any of its findings.

APPENDIX TO CHAPTER 1

The structural equation models considered in this book are special cases of the so-called *general LInear Structural RELationships* (*LISREL*) *model*. To define it formally, denote by Y the vector of p observed variables in a considered study ($p > 1$), by η that of q latent factors assumed in it ($q > 0$), and by ε the vector of p pertinent residuals (error terms). Designating by Λ the $p \times q$ matrix of factor loadings, by ν the $p \times 1$ vector of observed variable mean intercepts, by α the $q \times 1$ vector of latent variable intercepts, and by ζ the $q \times 1$ vector of corresponding latent disturbance terms, the general LISREL model is represented by the following pair of equations:

$$Y = \nu + \Lambda\eta + \varepsilon \text{ and} \tag{A1.1}$$

$$\eta = \alpha + B\eta + \zeta\,, \tag{A1.2}$$

where in addition the matrix $I - B$ is assumed invertible. In this introductory text, for simplicity the additional assumption of normality of the variables in Y, η, ε, and ζ is made, which are continuous, as well as that of uncorrelatedness of ε and ζ, ε and η, and ζ and η. Components of each of the vectors ε, ζ, and η may be correlated among themselves, however, assuming overall model identification. Further, the model classes considered in this book also assume $\nu = \alpha = 0$, which does not pose any special restriction of generality in many social and behavioral studies when the units and origins of measurement are not meaningful (or the analyzed variables may be considered mean-centered without loss of relevant information for a particular investigation.) The validity of Equations (A1.1) and (A1.2) is assumed for each individual in a sample, but for simplicity of notation we suppress the subject subindex in this Appendix. We stress that the LISREL model defined by Equations (A1.1) and (A1.2) also includes the case where manifest variables influence observed and/or latent variables, which is seen by noting that observed predictors can be formally represented by error-free latent variables (i.e., η's with corresponding ε's being 0 and unitary pertinent elements of Λ) in the right-hand sides of each of these two equations.

Under these assumptions, via simple algebra Equations (A1.1) and (A1.2) entail that the implied, observed variable covariance matrix Σ has the following form:

$$\Sigma = \Lambda(I–B)^{-1}\Psi(I–B')^{-1}\Lambda' + \Theta\,, \tag{A1.3}$$

where priming denotes transposition, $\Psi = \text{Cov}(\xi)$ is the latent disturbance terms covariance matrix, and $\Theta = \text{Cov}(\varepsilon)$ is that of the residuals.

A simple inspection of the right-hand side of Equation (A1.3) demonstrates that the parameters of the general LISREL model are among: (a) the

factor loadings (elements of Λ); (b) the latent regression coefficients (elements of B); and (c) the variances and covariances of the latent disturbance terms and residuals (the elements of Ψ and Θ) that collectively represent the independent variables of the model. (In case a considered model has no latent dependent variable, as in confirmatory factor analysis models, set $\eta = \zeta$ and $B = 0$ in (A1.2) and use the immediately preceding sentence in this paragraph to obtain its model parameters.) This observation presents the rationale behind the rules for determining model parameters discussed in this chapter. Also, denoting by γ the vector of all model parameters (i.e., all variances of and covariances between independent variables as well as all regression coefficients and factor loadings), Equation (A1.3) states that $S = \Sigma(\gamma)$. That is, an implication of any considered structural equation model is the structuring of all elements of the pertinent population covariance matrix in terms of fewer in number, more fundamental parameters, viz. those in γ. This is the essence of the model parameterization that is invoked as a consequence of adopting a particular structural equation model.

The fit function for the model fitting and estimation approach used throughout this book, that of the maximum likelihood method, is defined as

$$F_{ML} = \text{“distance”}(S, \Sigma(\gamma)) = -1n|S\,\Sigma(\gamma)^{-1}| + \text{tr}(S\,\Sigma(\gamma)^{-1}) - p\,, \qquad \text{(A1.4)}$$

where $|.|$ denotes matrix determinant and tr(.) trace (sum of main diagonal elements). Using special numerical optimization algorithms, this fit function is minimized across the parameter space, i.e., the set of all admissible values of all model parameters (in particular, typically positive variances). When the model is correct and fitted to the covariance matrix for a large sample, $T = n\,F_{ML,\,min}$ follows a central chi-square distribution with degrees of freedom being those of the fitted model, where $F_{ML,\,min}$ is the minimum of (A1.4), $n = N - 1$ and N is sample size. This fit function is derivable in the context of the likelihood ratio test theory, when comparing a given model to a saturated model fitted to the same set of observed variables (e.g., Bollen, 1989). The minimizer of (A1.4), $\hat{\gamma}$, consists of the point estimates of all model parameters; their standard errors are obtained as the corresponding elements on the main diagonal of the inverted observed information matrix.

The unweighted least squares (ULS) method is based on minimizing, across the set of all possible values for γ, the fit function

$$F_{ULS} = .5\ \text{tr}[(S - \Sigma(\gamma))^2]\,, \qquad \text{(A1.5)}$$

while the generalized least squares (GLS) approach minimizes the fit function

$$F_{GLS} = .5\ \text{tr}[(I - S^{-1}\Sigma(\gamma))^2]\,. \qquad \text{(A1.6)}$$

The asymptotically distribution free (ADF) method (also called weighted least squares method) minimizes the fit function

$$F_{ADF} = (s - \sigma(\gamma))' W^{-1} (s - \sigma(\gamma)) \,, \tag{A1.7}$$

where s denotes the strung-out vector of nonredundant elements of S, $\sigma(\gamma)$ the similar vector of their counterparts in $\Sigma(\gamma)$, and W is a weight matrix that represents a consistent estimate of the large sample covariance matrix of the elements of S (considered as random variables themselves). As shown in the literature (e.g., Bollen, 1989), with large samples the ULS, ML, and GLS estimation methods can be considered special cases of the ADF method, obtained with appropriate choices of the matrix W. Corrections of the chi-square value and parameter standard errors, which are functions of the matrix W, yield from the ML test statistic and standard errors robust ML test statistics and robust standard errors, respectively (e.g., Bentler, 2004).

The matrix of covariance residuals, $S - \Sigma(\hat{\gamma})$, which is also often referred to as matrix of model residuals, contains information about local goodness of fit of the model. These residuals, i.e., the elements of $S - \Sigma(\hat{\gamma})$, are expressed in the original metrics of manifest variables that may be quite dissimilar across variables, and thus are in general hard to interpret. Their standardized counterparts, called standardized residuals, are expressed in a uniform metric across all variables and can therefore be used to locate pairs of variables whose interrelationship indices are markedly misfit. Like the individual case residuals in regression analysis, in SEM the model residuals are not unrelated to one another. However, those of them whose standardized versions are in absolute value higher than 2 generally indicate parts of the model that are considerably inconsistent with the data. A positive residual means underprediction by the model of the covariance for the two variables involved, and may be made smaller by introducing a parameter additionally contributing to their interrelationship index (as reflected in its counterpart element of $\Sigma(\hat{\gamma})$). Conversely, a negative residual means overprediction by the model of that covariance, and may be rendered smaller in magnitude (i.e., absolute value) by deleting a parameter contributing to this interrelationship index (as reflected in its counterpart in $\Sigma(\hat{\gamma})$). An examination of model residuals, in particular standardized residuals, is therefore recommendable as an essential step in the process of model fit evaluation.