

5

Introduction to Path Analysis

This chapter introduces path analysis (PA), the original SEM technique for analyzing structural models with observed variables. Due to its relative seniority in the SEM family and the fact that it lacks a direct way to represent latent variables (other than disturbances), PA is sometimes seen as a less interesting or facile procedure than other varieties of SEM. But there are three reasons why this impression is mistaken. First, sometimes researchers use only a single indicator of some variables they are studying. Whether by design or default, due to things such as resource limitations, a multiple-indicator approach to construct measurement is not always possible. Second, it is important to know the principles of PA because many of these same concepts hold true for more complicated model types. Third, about 25% of roughly 500 applications of SEM published in 16 different psychology research journals from 1993 to 1997 concerned path models (MacCallum & Austin, 2000), so PA is still widely used. Readers who master the fundamentals of PA will thus be better able to understand and critique a wider variety of structural equation models.

5.1 CORRELATION AND CAUSATION

The technique of PA involves the estimation of presumed causal relations among observed variables. However, the basic datum of PA—and of all SEM techniques when means are not analyzed—is the covariance, which includes correlation: $cov_{XY} = r_{XY} SD_X SD_Y$. The reader is probably familiar with the expression “correlation does not imply causation.”

This mantra-like principle is apt because although a substantial correlation could indicate a causal relation, variables can also be associated in ways that have nothing to do with causality—spurious associations are one example. It is also true that an observed correlation of zero does not preclude a true causal relation—the phenomenon of suppression is an example.

The path analytic approach to the study of causality with correlations is as follows. Using the basic building blocks of path models described below, the researcher specifies a model that attempts to explain why X and Y (and other observed variables) are correlated. Part of this explanation may include presumed causal effects (e.g., X causes Y). Other parts of the explanation may reflect presumed noncausal relations, such as a spurious association between X and Y due to common causes. The overall goal of a PA is to estimate causal versus noncausal aspects of observed correlations. Part of the evaluation of a path model involves assessing how well it accounts for the data, that is, the observed correlations or covariances. If the model is not rejected, however, the researcher *cannot* automatically conclude that the hypotheses about causality are correct. This is because the failure to reject a path model (or any type of structural equation model) does *not* prove that it is correct. (This point is elaborated on later.)

The inference of causality requires much more than just acceptable correspondence between the model and the data. For example, to reasonably infer that X is a cause of Y , all of the following conditions must be met: (1) there is time precedence, that is, X precedes Y in time; (2) the direction of the causal relation is correctly specified, that is, X causes Y instead of the reverse or X and Y cause each other in a direct feedback loop; and (3) the association between X and Y does not disappear when external variables such as common causes of both are held constant (i.e., it is not spurious). These conditions are a tall order. For example, the hypothesis that X causes Y would be bolstered if the magnitude of their association is substantial and X is measured before Y (i.e., the design is longitudinal). However, the expected value of r_{XY} may not be zero even if Y causes X and the effect (X) is measured before the cause (Y) (e.g., see Bollen, 1989, pp. 61–65). Even if X actually causes Y , the magnitude of r_{XY} may be low if the interval between their measurement is either too short (e.g., effects on Y take time to materialize) or too long (e.g., the effects are temporary and have dissipated).

though a substantial correlation can also be associated with spurious associations. Observed correlation of zero is the phenomenon of suppression.

of causality with correlating blocks of path models. A model that attempts to correlate variables (e.g., X and Y) are correlated. Part causal effects (e.g., X causes Y) and presumed noncausal relations between X and Y due to common cause are causal versus noncausal evaluation of a path model for the data, that is, the model is not rejected, however, one cannot conclude that the hypotheses are supported by the failure to reject a path model. (It does not prove that it is true.)

more than just acceptable to data. For example, to reasonably follow conditions must be met: (1) X precedes Y in time; (2) causal effects are correctly specified, that is, X and Y do not cause each other in a direct or indirect manner; (3) even if X and Y do not directly cause each other, other causes of both are held constant. These conditions are a tall order. For example, it would be bolstered if the magnitude of the effect of X on Y is measured before Y is measured, the expected value of r_{xy} is greater than zero (X is measured before Y). Even if X actually causes Y, the interval between their measurement on Y take time to materialize (the effect may be temporary and have dissipated).

Despite the problems just described, the assessment of variables at different times at least provides a measurement framework consistent with the specification of directional causal effects. However, longitudinal designs pose potential difficulties, such as subject attrition and the need for additional resources. Probably because of these reasons, most path analytic studies feature concurrent rather than longitudinal measurement. When the variables are concurrently measured, it is not possible to demonstrate time precedence. Therefore, the researcher needs a very clear, substantive rationale for specifying that X causes Y instead of the reverse or that X and Y mutually influence each other when all variables are measured at the same time.

It is only from a solid base of knowledge about theory and research that one can even begin to address these requirements for inferring causation from correlation. Although facility with the statistical details of SEM is essential, it is not a substitute for what could be called wisdom about one's research area. This point is emphasized time and again in this book. It is also why one should adopt the view that just as correlation does not imply causation, statistical causal modeling does not prove causation either. It is why Wilkinson and the Task Force on Statistical Inference (1999) emphasize that use of SEM computer programs "rarely yields any results that have any interpretation as causal effects" (p. 600). What additional evidence is required in order to infer causality after a PA (or any other form of SEM) is considered later.

5.2 SPECIFICATION OF PATH MODELS

The problems of specifying which variables to include in path models and the directionalities of presumed causal effects among them are considered next.

What to Include

This is perhaps the most basic specification issue: given some phenomenon of interest, what are all the variables that affect it? The literature for newer research areas can be limited, so decisions about what to include in the model must sometimes be guided more by the researcher's experience than by published reports. Consulting with

experts in the field about plausible specifications may also help. In more established areas, however, sometimes there is *too much* information. That is, there may be so many potential causal variables mentioned in the literature that it is virtually impossible to include them all. To deal with this situation, the researcher must again rely on his or her judgment about the most crucial variables.

The specification error of omitting causal variables from a path model has the same potential consequence as omitting predictors from a regression equation does: estimates of causal effects of variables included in the model may be inaccurate if there are omitted causal variables that covary with those in the model. The direction of this inaccuracy could be either underestimation of true causal effects or overestimation, depending on the correlations between included and excluded variables. Note that underestimation probably occurs more often than overestimation. However, just as it is unrealistic to expect that all relevant predictors are included in a regression analysis, the same is true about including all causal variables in a path model. Given that most path models may be misspecified in this regard, the best way to minimize potential bias is preventive: make an omitted variable an included one through careful review of extant theory and research.

How to Measure the Hypothetical Construct

The selection of measures is a recurrent research problem. This is especially true in PA because there is only one observed measure of each construct. It is therefore crucial that measures have good psychometric characteristics; that is, their scores should be both reliable and valid. Score reliability is especially critical because one assumption of PA is that the exogenous variables are measured without error. Although this assumption is not required of the endogenous variables, the general consequence of error-prone measures of either exogenous or endogenous variables in PA is that the statistical estimates of presumed causal effects may be inaccurate. However, the nature of this potential inaccuracy is different for exogenous and endogenous variables. The consequences of measurement error in PA are discussed later after prerequisite concepts are introduced. Recall that disattenuating correlations for measurement error is one way to take reliability into account (see Equation 3.5), but it is not a standard part of PA.

ifications may also help. In cases there is *too much* information causal variables mentioned it is impossible to include them in the model. The researcher must again rely on his or her abilities.

causal variables from a path model as omitting predictors from the causal effects of variables in the model. The direction of this distinction of true causal effects or correlations between included and omitted variables probably occurs more often than not. As it is unrealistic to expect to find a regression analysis, the tables in a path model. Given this in mind, the best way to make an omitted variable an important theory and research.

Construct

search problem. This is especially true if the observed measure of each construct does not have good psychometric properties. It is both reliable and valid. One assumption of PA is that the error is without error. Although this is a reasonable assumption for exogenous variables, the general assumption of either exogenous or endogenous variables is that the estimates of presumed causal relations are accurate. The structure of this potential inaccuracy is discussed later after prerequisites for attenuating correlations for reliability into account (see Chapter 10 of PA).

Building Blocks

The basic elements of path models were introduced earlier but are briefly reviewed here. Let X refer to observed exogenous variables, Y to observed endogenous variables, D to disturbances (i.e., unobserved exogenous variables), \curvearrowleft and \curvearrowright to, respectively, the variances and covariances (i.e., unanalyzed associations) of exogenous variables, \rightarrow to presumed direct causal effects, and \rightleftharpoons to reciprocal causal effects. An account of why observed variables are correlated with a path model can reflect two kinds of causal relations or two kinds of noncausal associations. The first kind of causal relation reflects presumed unidirectional influence and includes direct effects of one variable on another (e.g., $X \rightarrow Y$) or indirect effects through at least one mediating variable (e.g., $X \rightarrow Y_1 \rightarrow Y_2$). The second kind concerns feedback loops, either direct or indirect: direct feedback involves only two variables in a reciprocal relation (e.g., $Y_1 \rightleftharpoons Y_2$); indirect feedback involves three or more variables (e.g., $Y_1 \rightarrow Y_2 \rightarrow Y_3 \rightarrow Y_1$). Feedback loops represent mutual influence among variables that are concurrently measured. As an example of a possible feedback relation, consider vocabulary breadth and reading: a bigger vocabulary may facilitate reading, but reading may result in an even larger vocabulary.

The two kinds of noncausal relations are unanalyzed associations and spurious associations. Most path models with more than one observed exogenous variable assume unanalyzed associations between them (e.g., $X_1 \curvearrowright X_2$). An unanalyzed association means just that: the two variables are assumed to covary, but the reasons why they covary—do they affect each other? do they have common causes?—are unknown. It is also possible to represent in path models unanalyzed associations between disturbances (e.g., $D_1 \curvearrowright D_2$). Recall that a disturbance represents all omitted causes of the corresponding endogenous variable. An unanalyzed association between a pair of disturbances is called a disturbance correlation (for standardized variables) or a disturbance covariance (for unstandardized variables); the former term is used herein from this point regardless of whether the variables are standardized or not. A disturbance correlation reflects the assumption that the corresponding endogenous variables share at least one common omitted cause. Accordingly, the absence of the symbol for an unanalyzed association between two disturbances reflects the presumption that the two disturbances are independent.

tion of independence of unmeasured causes. It also represents the hypothesis that the observed correlation between that pair of endogenous variables can be entirely explained by other observed variables in the model. Unlike unanalyzed associations between measured exogenous variables, which are routinely represented in path models, the inclusion of disturbance correlations in a structural model is not so simple. The reasons why this is so are elaborated later.

One other point about disturbances and unanalyzed associations: if measured exogenous variables have unanalyzed associations and disturbances can have them, too, can a disturbance have an unanalyzed association with a measured exogenous variable (e.g., $X \uparrow\downarrow D$)? Theoretically, it can, and such a correlation would imply the presence of an omitted variable (e.g., W) that causes both X and the endogenous variable Y . That is, part of the covariance of X with Y is spurious due to a common cause. However, it is usually assumed in PA that the measured exogenous variables and disturbances are unrelated. (For an example of the estimation of $X \uparrow\downarrow D$ and conditions required to do so, see Kenny, 1979, pp. 93–95.) Part of the reason for this assumption is statistical. Specifically, when multiple regression is used to estimate recursive path models (defined later), it must be assumed that the predictors (observed exogenous variables) and residuals (disturbances) are uncorrelated. Another reason is more conceptual: assuming the independence of disturbances and observed exogenous variables permits the estimation of direct effects of the latter (e.g., $X \rightarrow Y$), with omitted causal variables held constant. Recall that the derivation of regression coefficients accomplishes the same thing but among observed variables. The assumption of unrelated observed exogenous variables and disturbances provides what Bollen (1989) calls *pseudoisolation* of the former from all other unmeasured causes of the corresponding endogenous variable. This is a strong assumption, one that is probably violated in most applications of SEM. As noted earlier, the seriousness of violating this assumption increases with the magnitudes of the correlations between excluded and included variables, which once again highlights the importance of accurate specification.

Spurious associations in path models are represented by specifying common causes. For example, if X is specified as a direct cause of both Y_1 and Y_2 , then at least part of the observed correlation between these two endogenous variables is presumed to be spurious. If the model

ses. It also represents the
tween that pair of endoge-
other observed variables in
between measured exoge-
ented in path models, the
structural model is not so
orated later.

d unanalyzed associations:
lyzed associations and dis-
bance have an unanalyzed
able (e.g., $X \curvearrowright D$)? Theo-
d imply the presence of an
and the endogenous vari-
with Y is spurious due to a
ed in PA that the measured
nrelated. (For an example
as required to do so, see
for this assumption is sta-
is used to estimate recur-
ssumed that the predictors
duals (disturbances) are
ptual: assuming the inde-
ogenous variables permits
(e.g., $X \rightarrow Y$), with omitted
e derivation of regression
ut among observed vari-
l exogenous variables and
lls pseudoisolation of the
he corresponding endoge-
one that is probably vio-
earlier, the seriousness of
magnitudes of the correla-
es, which once again high-
on.

represented by specifying
d as a direct cause of both
correlation between these
be spurious. If the model

contains no other direct or indirect causal effects between these variables (e.g., $Y_1 \rightarrow Y_2$), the entire association between them is presumed to be spurious. Spurious associations can also involve multiple common causes.

Directionality

The specification of the directionalities of presumed causal effects is a crucial part of a PA. The measurement of some variables before others provides one means to specify directionality. When variables are concurrently measured, however, the specification of directionality requires a clear rationale. For example, Lynam, Moffitt, and Stouthamer-Loeber (1993) estimated a path model where poor verbal ability was specified as a direct cause of delinquency, but both variables were concurrently measured in their sample, which raises the question: why this particular direction? Is it not also plausible that certain behaviors associated with delinquency, such as drug use or withdrawal from school, could impair verbal ability? The arguments offered by Lynam et al. for their original specification included the following: their participants were relatively young (about 12 years old), which may rule out delinquent careers long enough to affect verbal skills; the authors also cited results of prospective research that suggest that low verbal ability precedes antisocial acts. Although these particular arguments have been criticized (Block, 1995), they at least exemplify the types of arguments that researchers should provide for their specifications of directionality. Unfortunately, few path analytic articles even consider these issues, which is a serious problem.

What if a researcher is fundamentally uncertain about the directionalities of effects? There are basically three options: (1) forgo PA in favor of techniques that require fewer *a priori* assumptions, such as multiple regression; (2) specify and test alternative path models each with different directionalities; or (3) include reciprocal effects in the model as a way to cover both possibilities (e.g., $Y_1 \leftrightarrow Y_2$). The last two alternatives may seem appealing to someone who is keen to conduct a PA, but neither is a magical solution. In fact, the first option may be preferred. As already mentioned, it is possible in PA (or any other SEM technique) that different models may fit the same data equally well (i.e., equivalent models, defined later). When (not if!) this occurs, there is no statistical basis for choosing one model over another. Also,

the inclusion of reciprocal effects is not a simple matter. The addition of even one reciprocal effect to a model makes it nonrecursive (defined later), which makes the model more difficult to analyze. Thus, there are potential costs to the specification of reciprocal causal effects as a hedge against uncertainty about directionality.

Model Complexity

Given the directionality problems just discussed, why not make life easier for yourself and just specify a path model wherein everything causes everything else? There is a problem with doing so because there is a limit on how many parameters can be represented in a structural equation model. A parameter is a characteristic of a population, and it is estimated with a sample statistic. The limit on the number of model parameters that can be estimated is determined by the number of observations, which is the number of variances and covariances among the observed variables. If v is the number of observed variables, the number of observations equals $v(v + 1)/2$. Suppose that there are four observed variables in a path model. The number of observations here is $4(5)/2$, or 10; thus, there are 10 observations altogether. Note that this number includes only the *unique* entries in a covariance matrix of four variables. That is, only entries above or below the diagonal (i.e., the covariances) are counted and added to the number of entries in the diagonal, which are the variances. The number of observations remains the same regardless of the sample size. If four variables are measured for 100 or 1,000 cases, the number of observations is still 10. Thus, adding cases does not increase the number of observations; only adding variables can do so.

The number of observations for a path model sets the upper limit for the number of parameters that can be estimated. Specifically, the *number of such model parameters cannot exceed the number of observations*. A model may have fewer parameters than observations but no more. The difference between the number of observations and the number of its parameters is the **model degrees of freedom**, df_M . The requirement that there be at least as many observations as parameters to be estimated can be expressed as the requirement that $df_M \geq 0$. The general rule for counting the number of parameters in a path model is as follows:

nple matter. The addition is it nonrecursive (defined to analyze. Thus, there are reciprocal causal effects as a ty.

ussed, why not make life model wherein everything ith doing so because there epresented in a structural stic of a population, and it t on the number of model mined by the number of es and covariances among of observed variables, the uppose that there are four ber of observations here is ; altogether. Note that this covariance matrix of four row the diagonal (i.e., the number of entries in the er of observations remains ur variables are measured rations is still 10. Thus, 10 observations; only adding

model sets the upper limit estimated. Specifically, the eed the number of observa than observations but no of observations and the rees of freedom, df_M . The obsevations as parameters uirement that $df_M \geq 0$. The ameters in a path model is

The total number of variances and covariances (i.e., unanalyzed associations) of exogenous variables that are either observed or unobserved (i.e., disturbances) and direct effects on endogenous variables from other observed variables equals the number of parameters.

Note that the variances and covariances of the endogenous variables are not considered model parameters. This is because endogenous variables are represented in path models as the effects of other measured variables in the model and their own disturbances. Thus, endogenous variables are not free to vary or covary. Instead, the path model as a whole attempts to account for variances of the endogenous variables and their covariances with other variables.

Although the reasons are outlined later, a model with more parameters than observations ($df_M < 0$) is not amenable to empirical analysis. Such models are said to be not identified, which means that it is mathematically impossible to derive unique estimates of each parameter. The most likely symptom of this problem is that an SEM computer program may terminate its run prematurely with error messages. Analogous situations can occur in multiple regression when there are more predictors than cases: the computer cannot carry out the analysis because the model is too complex to be estimated with the available information.

Constraints

The status of each model parameter can be free, fixed, or constrained depending on the researcher's specifications. A **free parameter** is to be estimated by the computer with sample data. In contrast, a **fixed parameter** is specified to equal a constant; that is, the computer "accepts" this constant as the estimate of the parameter regardless of the data. For the present example, the hypothesis that the direct effect of X on Y is zero corresponds to the specification that the coefficient for this path is fixed to zero. The computer will then estimate all other model parameters while holding this path coefficient constant (i.e., equal to 0). It is common in SEM to test hypotheses by specifying that a previously fixed-to-zero parameter becomes a free parameter or vice versa. Results of such analyses may indicate whether to respecify a model by making it more complex (an effect is added) or more parsimonious (an effect is dropped).

A constrained parameter is estimated by the computer within some restriction, but it is not fixed to equal a constant. The restriction typically concerns the relative values of other constrained parameters. An **equality constraint** means that the estimates of two or more parameters are forced to be equal. Suppose that an equality constraint is imposed on the two direct effects that make up a direct feedback loop (e.g., $Y_1 \leftrightarrow Y_2$). This constraint simplifies the analysis because only one path coefficient is needed rather than two. In a multiple-sample SEM analysis, a **cross-group equality constraint** forces the computer to derive equal estimates of that parameter across all groups. This specification corresponds to the null hypothesis that the parameter is equal in all populations from which the samples were drawn (Chapter 11).

Other kinds of constraints are not seen nearly as often in the literature. A **proportionality constraint** forces one parameter estimate to be some proportion of the other. For example, the path coefficient for one direct effect in a reciprocal relation may be forced to be three times the value of the other coefficient. An **inequality constraint** forces the value of a parameter estimate to be less than or greater than a specified value. The specification that the value of an unstandardized path coefficient must be at least 5.00 is an example of an inequality constraint. The imposition of proportionality or inequality constraints generally requires knowledge about the relative magnitudes of effects, but such knowledge is relatively rare in the behavioral sciences. A **nonlinear constraint** imposes a nonlinear relation between two parameter estimates. For example, the value of one estimate may be forced to equal the square of another. A method to correctly analyze a correlation matrix instead of a covariance matrix referred to as constrained estimation uses nonlinear constraints (Chapter 7). So does the Kenny–Judd method, one of two different methods for estimating nonlinear effects of latent variables described in Chapter 13.

5.3 TYPES OF PATH MODELS

There are two basic kinds of path models. **Recursive models** are the most straightforward and have two basic features: their disturbances are uncorrelated, and all causal effects are unidirectional. **Nonrecursive models** have feedback loops or may have correlated disturbances. Con-

l by the computer within a constant. The restriction of constrained parameters. states of two or more parameters an equality constraint is set up a direct feedback loop in the analysis because only two. In a multiple-sample constraint forces the computer to cross all groups. This specification that the parameter is equal in more drawn (Chapter 11). nearly as often in the literature parameter estimate to be the path coefficient for one forced to be three times the constraint forces the value greater than a specified value. standardized path coefficient inequality constraint. The equality constraints generally magnitudes of effects, but such in oral sciences. A nonlinear estimate between two parameter estimate may be forced to equal directly analyze a correlation related to as constrained estimation. So does the Kenny-Judd estimating nonlinear effects 3.

5. Recursive models are the features: their disturbances unidirectional. Nonrecursive correlated disturbances. Con-

sider the path models in Figure 5.1. The model of Figure 5.1(a) is recursive because its disturbances are independent and no variable is both a cause and an effect of another variable, directly or indirectly. For example, X_1 , X_2 , and Y_1 are specified as direct or indirect causes of Y_2 , but Y_2 has no effect back onto one of its presumed causes. In contrast, the model of Figure 5.1(b) is nonrecursive because it has a direct feedback loop in which Y_1 and Y_2 are specified as both causes and effects of each other ($Y_1 \leftrightarrow Y_2$). The model of Figure 5.1(b) also has a disturbance correlation. Note that models with indirect feedback loops, such as $Y_1 \rightarrow Y_2 \rightarrow Y_3 \rightarrow Y_1$, are also nonrecursive.

There is another type of path model, one that has directional effects and correlated disturbances, two examples of which are pre-

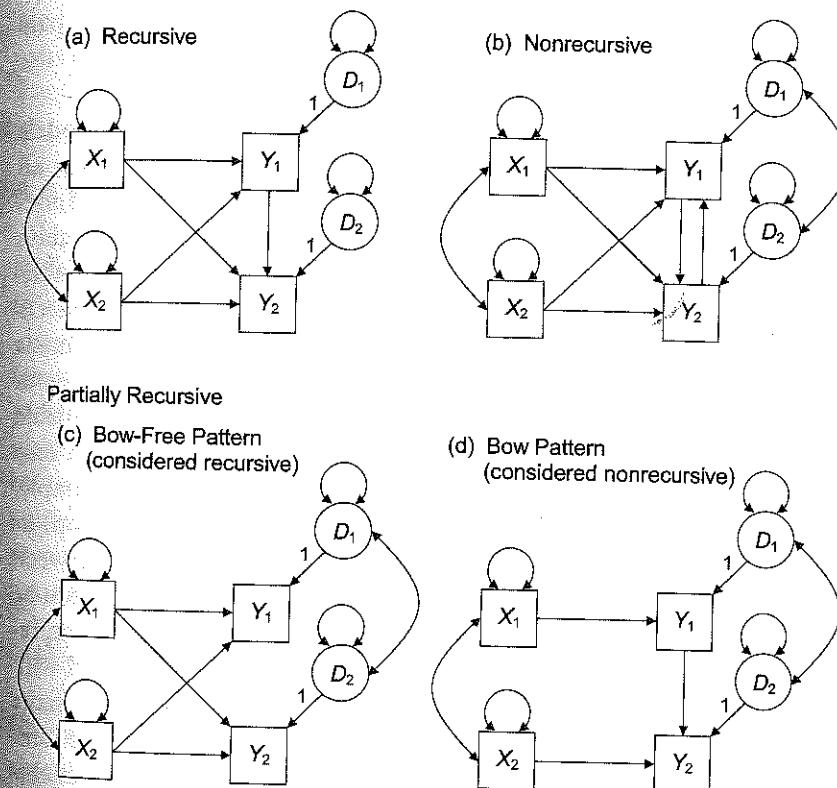


FIGURE 5.1. Examples of recursive, nonrecursive, and partially recursive path models.

sented in Figures 5.1(c) and 5.1(d). Unfortunately, the classification of such models in the SEM literature is not consistent. Some authors call these models nonrecursive, whereas others use the term partially recursive. But more important than the label for these models is the distinction made in the figure. Although the reasons are explained later, partially recursive models with a bow-free pattern of disturbance correlations can be treated in the analysis just like recursive models. A bow-free pattern means that correlated disturbances are restricted to pairs of endogenous variables *without* direct effects between them (see Figure 5.1(c)). In contrast, partially recursive models with a bow pattern of disturbance correlations must be treated in the analysis as nonrecursive models. A bow pattern means that a disturbance correlation occurs *with* a direct effect between the endogenous variables (see Figure 5.1(d); Brito & Pearl, 2003; Kenny, 1979). All ensuing references to recursive and nonrecursive models include, respectively, partially recursive models without and with direct effects among the endogenous variables.

The distinction between recursive and nonrecursive path models has several implications. The assumptions of recursive models that all causal effects are unidirectional and that the disturbances are independent when there are direct effects among the endogenous variables simplify the statistical demands for their analysis. For example, multiple regression can be used to estimate path coefficients for direct effects and disturbance variances in recursive models. Many social scientists are familiar with multiple regression, and computer programs with regression modules are widely available, some even freely available over the Internet. The same assumptions of recursive models that ease the analytical burden are also very restrictive, however. For example, causal effects that are not unidirectional (e.g., as in a feedback loop) or disturbances that are correlated in a model with direct effects between the corresponding endogenous variables (e.g., Figure 5.1(d)) cannot be represented in a recursive model. Although these effects can be represented in nonrecursive models, such models cannot be analyzed with standard multiple regression. Nonrecursive models require more sophisticated methods and may also require additional assumptions. The likelihood of a problem in the analysis of a nonrecursive model is also greater than for a recursive model. One of these problems is that of identification, which is discussed in the next section. Perhaps due to

imately, the classification of inconsistent. Some authors call these models partially recursive. The reason for this is that the reasons are explained by a free pattern of disturbance terms like recursive models. A disturbance is restricted to effects between them (see models with a bow path treated in the analysis as that a disturbance correlates endogenous variables (see 1979). All ensuing references include, respectively, partial direct effects among the

nonrecursive path models of recursive models that all the disturbances are independent of the endogenous variables analysis. For example, multi-coefficients for direct effects levels. Many social scientists computer programs with some even freely available recursive models that ease this, however. For example, as in a feedback loop) or with direct effects between (e.g., Figure 5.1(d)) cannot be represented. These effects can be represented but cannot be analyzed with recursive models require more additional assumptions. One of the problems is that of the next section. Perhaps due to

these difficulties, one sees relatively few nonrecursive path models in the behavioral science literature. But in some disciplines, especially economics, they are much more common, which suggests that the challenges of nonrecursive models can be overcome. The analysis of nonrecursive structural models is discussed in Chapter 9.

Before we continue, let us apply the rules for counting observations and parameters to the four path models of Figure 5.1. Because there are four observed variables in every model, the number of observations for each is $4(5)/2 = 10$. It is assumed that the constants (1) in the figure are fixed parameters that scale the disturbances (Chapter 4; this point is also elaborated later). Summarized in Table 5.1 are the numbers and types of free parameters for each model. Note there that all four models meet the requirement for at least as many observations as parameters. Also note that a reciprocal effect (e.g., $Y_1 \leftrightarrow Y_2$ in the model of Figure 5.1(b)) counts as two direct effects when no equality constraint is imposed. Readers are encouraged to match the parameters listed in Table 5.1 against the corresponding path models of Figure 5.1.

5.4 PRINCIPLES OF IDENTIFICATION

If life were fair, the researcher could proceed directly from specification to collection of the data to estimation. Unfortunately, just as life is sometimes unfair, the analysis of a path model is not always so straightforward. The problem that potentially complicates the evaluation of a path model (or any other kind of structural equation model) is that of identification. A model is said to be identified if it is *theoretically* possible to derive a unique estimate of each parameter. If not, the model is not identified. The word "theoretically" emphasizes identification as a property of the model and not of the data. For example, if a model is not identified, then it remains so regardless of the sample size (100, 1,000, etc.). Therefore, models that are not identified should be respecified; otherwise, attempts to analyze them may be fruitless.

There are two basic requirements for the identification of any kind of structural equation model: (1) there must be at least as many observations as free model parameters ($df_M \geq 0$), and (2) every unobserved (latent) variable must be assigned a scale (metric). Models that violate the first requirement are not identified; specifically, they are

TABLE 5.1. Number and Types of Parameters for Path Models in Figure 5.1

Model	Variances	Covariances	Direct effects on endogenous variables		Total
(a) Recursive	X_1, X_2 D_1, D_2	$X_1 \curvearrowup X_2$	$X_1 \rightarrow Y_1$ $X_1 \rightarrow Y_2$	$X_2 \rightarrow Y_1$ $X_2 \rightarrow Y_2$ $Y_1 \rightarrow Y_2$	10
(b) Nonrecursive	X_1, X_2 D_1, D_2	$X_1 \curvearrowup X_2$ $D_1 \curvearrowup D_2$	$X_1 \rightarrow Y_1$ $Y_1 \rightarrow Y_2$	$X_2 \rightarrow Y_2$ $Y_2 \rightarrow Y_1$	10
(c) Partially recursive (considered recursive)	X_1, X_2 D_1, D_2	$X_1 \curvearrowup X_2$ $D_1 \curvearrowup D_2$	$X_1 \rightarrow Y_1$ $X_1 \rightarrow Y_2$	$X_2 \rightarrow Y_1$ $X_2 \rightarrow Y_2$	10
(d) Partially recursive (considered nonrecursive)	X_1, X_2 D_1, D_2	$X_1 \curvearrowup X_2$ $D_1 \curvearrowup D_2$	$X_1 \rightarrow Y_1$	$X_2 \rightarrow Y_2$ $Y_1 \rightarrow Y_2$	9

underidentified. As an example of how a deficit of observations leads to nonidentification, consider the following equation:

$$a + b = 6 \quad (5.1)$$

Look at this expression as a model, the "6" as an observation (i.e., a datum), and a and b as parameters. Because Equation 5.1 has more parameters (2) than observations (1), it is impossible to find unique estimates for its parameters. In fact, there are an infinite number of solutions, including $(a = 4, b = 2)$, $(a = 8, b = -2)$, and so on, all of which satisfy Equation 5.1. A similar thing happens when a computer tries to derive unique estimates for an underidentified structural equation model: it is impossible to do so, and thus the attempt fails.

Assuming that all disturbances are assigned a scale (discussed next), it is impossible to specify a recursive path model with more parameters than observations. Look again at the path model of Figure 5.1(a). This recursive model already has as many parameters as observations (10). However, it is not possible to add another parameter to this model and still have it be recursive. For example, the addition to

path Models in Figure 5.1

Direct effects on endogenous variables		Total
→Y ₁	X ₂ → Y ₁	10
→Y ₂	X ₂ → Y ₂	
	Y ₁ → Y ₂	
→Y ₁	X ₂ → Y ₂	10
→Y ₂	Y ₂ → Y ₁	
→Y ₁	X ₂ → Y ₁	10
→Y ₂	X ₂ → Y ₂	
→Y ₁	X ₂ → Y ₂	9
	Y ₁ → Y ₂	

deficit of observations leads to equation:

(5.1)

as an observation (i.e., a model). Equation 5.1 has more than impossible to find unique solutions. There are an infinite number of solutions (e.g., $b = -2$), and so on, all of which happens when a computer fails to identify structural equations in the attempt.

signed a scale (discussed in the previous section). The path model with more than the path model of Figure 5.1(b) has as many parameters as observations. If we add another parameter to the model, for example, the addition to

this model of a disturbance correlation or a direct effect from Y₂ to Y₁ (which would result in a direct feedback loop) would make it nonrecursive.

It can be shown that due to their particular characteristics, recursive path models are always identified (e.g., Bollen, 1989, pp. 95–98). Thus, although it is theoretically possible (that word again) to derive unique estimates of the parameters of recursive path models, their analysis can still be foiled by other types of problems. (Remember, life isn't fair!) Data-related problems are one such difficulty. For example, multicollinearity can result in what Kenny (1979) referred to as empirical underidentification: if two observed variables are very highly correlated (e.g., $r_{XY} = .90$), then, practically speaking, they are the same variable, which reduces the effective number of observations below the value of $v(v + 1)/2$. The analysis of a recursive path model can also be foiled if the user of an SEM computer program specifies inaccurate initial estimates of model parameters, also called start values. This problem is discussed later. The good news here is that neither of these potential problems is due to inherent features of a recursive model. The first (multicollinearity) can be addressed through data screening; the second (bad start values), through additional computer runs with better initial estimates.

The situation for nonrecursive path models is more complicated. First, it is possible to inadvertently specify a nonrecursive model with more parameters than observations. Look again at the path model of Figure 5.1(b), which is nonrecursive. Although it appears that this model has "room" for two more direct effects (e.g., X₁ → Y₂), it already has as many parameters as observations (10). Second, particular configurations of paths in a nonrecursive model can make it nonidentified even if there are as many observations as parameters. As a simple example of how a model with equal numbers of parameters and observations can nevertheless fail to have a unique solution, consider the following set of formulas:

$$\begin{aligned} a + b &= 6 \\ 3a + 3b &= 18 \end{aligned} \quad (5.2)$$

Again, look at both expressions as a model, the total scores as observations, and a and b as parameters. Although this model has two observa-

tions and two parameters, it does not have a unique solution. In fact, there are an infinite number of solutions that satisfy Equation 5.2, such as $(a = 4, b = 2)$, $(a = 8, b = -2)$, and so on. This happens due to an inherent characteristic of the model: the second formula in Equation 5.2 ($3a + 3b = 18$) is not unique. Instead, it is simply three times the first formula ($a + b = 6$), which means that it cannot narrow the range of solutions that satisfy the first formula. These two formulas can also be described as linearly dependent.

Although the example just presented oversimplifies the identification issue for nonrecursive models, it points out one of the difficulties in their analysis: a nonrecursive model with at least as many observations as parameters is not necessarily identified. Fortunately, there are some ways that a researcher can determine whether some (but not all) types of nonrecursive models are identified. These procedures are described in Chapter 9, but it is worthwhile to make the following point now: adding exogenous variables is one way to remedy an identification problem of a nonrecursive structural model, but this typically can only be done *before* the data are collected. Thus, it is crucial to evaluate whether a nonrecursive model is identified right after it is specified and before the study is conducted.

Two other terms require definition: just-identification and overidentification. A just-identified model has equal numbers of parameters and observations and is identified. As discussed, a nonrecursive model with equal numbers of parameters and observations is not necessarily identified, so the term "just-identified" does not automatically apply to it. It does, however, apply to recursive models with the same property. As a demonstration of just-identification, consider the following formulas as a model with two observations (6, 10) and two parameters (a, b). Note that the second formula is not linearly dependent on the first:

$$\begin{aligned} a + b &= 6 \\ 2a + b &= 10 \end{aligned} \tag{5.3}$$

This two-observation, two-parameter model has a single solution $(a = 4, b = 2)$; therefore, it is just-identified. Note something else about Equation 5.3: given estimates of its parameters, it can perfectly reproduce the observations (i.e., 6, 10). The same thing is generally true for

a unique solution. In fact, satisfy Equation 5.2, such This happens due to an cond formula in Equation is simply three times the t cannot narrow the range ese two formulas can also

ersimplifies the identifica- out one of the difficulties at least as many observa- ied. Fortunately, there are whether some (but not all) d. These procedures are le to make the following e way to remedy an identi- l model, but this typically l. Thus, it is crucial to eval- d right after it is specified

just-identification and l has equal numbers of d. As discussed, a nonre- ieters and observations is identified" does not auto- to recursive models with st-identification, consider observations (6, 10) and d formula is not linearly

(5.3)

el has a single solution Note something else about ers, it can perfectly repro- thing is generally true for

just-identified path models: not only do they theoretically have unique solutions, but given such they will also perfectly fit the data. The implication of this characteristic for hypothesis testing in PA is discussed later.

A path model can also have fewer parameters than observations. If such models are also identified (true for a recursive model; perhaps not for a nonrecursive one), they are called *overidentified*. As an example of parameter estimation for an overidentified model, consider the following set of formulas with three observations (6, 10, 12) and two parameters (a, b):

$$\begin{aligned} a + b &= 6 \\ 2a + b &= 10 \\ 3a + b &= 12 \end{aligned} \quad (5.4)$$

Try as you might, you will be unable to find values of a and b that satisfy all three formulas. For example, the solution ($a = 4, b = 2$) works only for the first two formulas in Equation 5.4. At first, the absence of a solution appears paradoxical, but there is a way to solve this problem: the imposition of a statistical criterion leads to unique estimates for the parameters of an overidentified model. An example of such a criterion for Equation 5.4 could be the following: find values of a and b that are positive and yield totals such that the sum of the squared differences between the observations (6, 10, 12) and these totals is as small as possible. Applying this criterion to the estimation of a and b in Equation 5.4 yields a solution that not only gives the smallest total squared difference (.67) but that is also unique (using only one decimal place, we obtain $a = 3.0$ and $b = 3.3$). Note that this solution does not perfectly reproduce the observations in Equation 5.4.

Thus, although it is possible to find a unique solution for overidentified models, it may not perfectly reproduce the observations. By the same token, an overidentified path model may not perfectly fit the data. Although this characteristic may at first glance seem like a drawback, it actually has an important role in model testing, one that is explored later.

The second general requirement for identification is that every latent variable must be assigned a scale. This is necessary in order for the computer to be able to calculate estimates of effects that involve

latent variables. Disturbances are the only kind of latent variable in path models. In PA, scales are usually assigned to disturbances through a **unit loading identification (ULI) constraint**. This means that the path coefficient for the direct effect of a disturbance—also called a **residual path coefficient**—is fixed to equal 1.0.¹ This specification has the consequence of assigning to a disturbance a scale that is related to that of the unexplained variance of the corresponding endogenous variable. This scale is not typically identical to that of the endogenous variable because it reflects the predictive power of the presumed causes of the latter. For example, the disturbance D_1 in Figure 5.1 is assigned a scale related to that of the unexplained variance of Y_1 through the imposition of a ULI constraint (i.e., $D_1 \rightarrow Y_1 = 1.0$). Because the residual path coefficient is here a fixed parameter, the computer needs only to estimate the disturbance variance. There is another method to scale latent variables that is discussed in Chapter 7, but it is rarely applied to disturbances in path models. Moreover, most SEM computer programs make it easier to specify a ULI constraint for disturbances.

5.5 SAMPLE SIZE

As already mentioned, the number of cases has no bearing on whether a path model is identified. What, then, is the role of sample size in PA—and all other SEM techniques? Basically the same as for other kinds of statistical methods: results derived within larger samples have less sampling error than within smaller samples. The researcher must then answer the next logical question: how large a sample is required in order for the results to be reasonably stable? Some guidelines about absolute sample size in estimation methods were offered earlier (small, $N < 100$; medium, N between 100 and 200; large, $N > 200$). Another consideration is model complexity. That is, more complex models—those with more parameters—require larger samples than more parsimonious models in order for the estimates to be comparably stable. Thus, a sample size of 200 or even much larger may be necessary for a very complicated path model. Although there are no absolute standards

¹The specification of any constant, such as 2.0 or 17.3, for the residual path coefficient would identify the disturbance, but it is much more common for this constant to equal 1.0.

nd of latent variable in to disturbances through it. This means that the turbance—also called a¹ This specification has a scale that is related to responding endogenous that of the endogenous of the presumed causes Figure 5.1 is assigned a ance of Y_1 through the 1.0). Because the reside computer needs only another method to scale out it is rarely applied to EM computer programs disturbances.

; no bearing on whether ; role of sample size in the same as for other thin larger samples have es. The researcher must e a sample is required in Some guidelines about re offered earlier (small, urge, $N > 200$). Another more complex models—implies than more parsion may be comparably stable. t may be necessary for a re no absolute standards

r the residual path coefficient non for this constant to equal

in the literature about the relation between sample size and path model complexity, the following recommendations are offered: a desirable goal is to have the ratio of the number of cases to the number of free parameters be 20:1; a 10:1 ratio, however, may be a more realistic target. Thus, a path model with 20 parameters should have a minimum sample size of 200 cases. If the cases/parameter ratio is less than 5:1, the statistical precision of the results may be doubtful. Other, more precise ways to estimate minimum sample sizes, such as power analysis, are considered in the next chapter.

5.6 OVERVIEW OF ESTIMATION OPTIONS

There are basically two options for the analysis of recursive path models: (1) multiple regression or (2) estimation with an SEM computer program. The latter typically offers users the choice of different procedures, the most widely used of which is maximum likelihood (ML) estimation. In fact, ML estimation is the default method in most SEM computer programs. For just-identified recursive path models, multiple regression and ML estimation yield identical estimates of direct effects (path coefficients); estimates of disturbance variances may vary slightly because the two procedures use somewhat different denominators in these terms. Values of path coefficients for overidentified recursive path models may be slightly different, but the two procedures generally yield similar results within large samples (i.e., their estimates are asymptotic).

Notwithstanding the similarity of results yielded by multiple regression and ML estimation for recursive path models, there are three reasons why it is well worth the effort to learn how to use an SEM computer program. First, there are numerous statistical indexes of the overall fit of the model to the data that are available in the output of SEM computer programs that are not generated by regression programs. These fit indexes, introduced in Chapter 6, are very useful for testing certain types of hypotheses, especially those that involve the comparison of different models evaluated with the same data. Second, there are several types of results that are automatically calculated by SEM computer programs that must be derived by hand when one uses a regression program. For simple models, these hand calculations are not too

burdensome a chore. For complex models, though, it's a real advantage to let the computer do the work. In contrast, nonrecursive path models and other kinds of structural equation models, such as CFA measurement models or structural regression (SR) models, can be estimated with ML but not multiple regression. Indeed, it would be no exaggeration to describe ML estimation as the motor of SEM. (You are the driver.)

5.7 MAXIMUM LIKELIHOOD ESTIMATION

This section reviews basic characteristics of ML estimation for any kind of structural equation model. Other estimation options, including multiple regression for recursive path models only, are described in the next chapter.

Description

The term **maximum likelihood** describes the statistical principle that underlies the derivation of parameter estimates: the estimates are the ones that maximize the likelihood (the continuous generalization) that the data (the observed covariances) were drawn from this population. That is, ML estimators are those that maximize the likelihood of a sample that is actually observed (Winer, Brown, & Michels, 1991). It is a **normal theory** method because ML estimation assumes that the population distribution for the endogenous variables is multivariate normal. Other methods are based on different parameter estimation theories, but they are not currently used as often. In fact, the use of an estimation method other than ML requires explicit justification (Hoyle, 2000).

Most forms of ML estimation in SEM are simultaneous, which means that estimates of model parameters are calculated all at once. For this reason, ML estimation is described in the statistical literature as a **full-information method**. (There are some partial-information forms of ML estimation, but they are not described in this book.) In contrast, techniques that analyze the equation for only one endogenous variable at a time are known as **partial-information** or **limited-information methods**. Multiple regression used to estimate a recursive path model is

ough, it's a real advantage nonrecursive path models el, such as CFA measure- models, can be estimated l, it would be no exaggerator of SEM. (You are the

ATION

ML estimation for any kind on options, including mul- only, are described in the

ne statistical principle that states: the estimates are the inuous generalization) that drawn from this population. size the likelihood of a sam- (, & Michels, 1991). It is a on assumes that the popu- ples is multivariate normal. meter estimation theories, fact, the use of an estima- plicit justification (Hoyle,

simultaneous, which means ultated all at once. For this atistical literature as a full- rial-information forms of in this book.) In contrast, ly one endogenous variable on or limited-information te a recursive path model is

an example of a partial-information method. So is two-stage least squares, a special type of regression method described in Chapter 9 that can analyze nonrecursive path models. Implications of the difference between full-information versus partial-information estimation when there is specification error are considered in Chapter 6.

The statistical criterion minimized in ML estimation—also known as the fitting function or discrepancy function—is related to discrepancies between the observed covariances and those predicted by the researcher's model. (How model-implied covariances are calculated is discussed in Chapter 6 after prerequisite concepts are introduced here.) In this sense, the ML fitting function is analogous (but not equivalent) to the least squares criterion of multiple regression. The mathematics of how ML estimation actually goes about generating a set of parameter estimates that minimize its fitting function are complex, and it is beyond the scope of this section to describe them in detail (e.g., Eliason, 1993; Kaplan, 2000, pp. 25–28; a less technical presentation is available in Nunnally & Bernstein, 1994, pp. 147–155). The similarities between ML estimates and those derived with multiple regression for recursive path models is reassuring for nonstatisticians, however.

The method of ML estimation is usually so complicated that it is often iterative, which means that the computer derives an initial solution and then attempts to improve these estimates through subsequent cycles of calculations. "Improvement" means that the overall fit of the model to the data generally becomes better from step to step. For most just-identified structural equation models, the fit will eventually be perfect. For overidentified models, the fit of the model to the data may be imperfect, but iterative estimation will continue until the increments of the improvement in model fit fall below a predefined minimum value. Iterative estimation may converge to a solution quicker if the procedure is given reasonably accurate start values, which are initial estimates of a model's parameters. If these initial estimates are grossly inaccurate—for instance, the start value for a path coefficient is positive when the actual direct effect is negative—then iterative estimation may fail to converge, which means that a stable solution has not been reached. Iterative estimation can also fail if the relative variances among the observed variables are very different; that is, the covariance matrix is ill scaled.

Computer programs typically issue a warning message if iterative

estimation is unsuccessful. When this occurs, whatever final set of estimates was derived by the computer may warrant little confidence. Also, some SEM computer programs automatically generate their own start values. It is important to understand, however, that computer-derived start values do not always lead to converged solutions. Although the computer's "guesses" about start values are usually pretty good, sometimes it is necessary for the researcher to provide better start values for the solution to converge. The guidelines for calculating start values for path models presented in Appendix 5.A may be helpful. Another tactic is to increase the program's default limit on the number of iterations to a higher value (e.g., from 30 to 100). Allowing the computer more "tries" may lead to a converged solution.

Although usually not a problem with recursive path models, it can happen in ML and other iterative methods that a converged solution is inadmissible. This is evident by a parameter estimate with an illogical value such as Heywood cases, which include negative variance estimates or estimated correlations between a factor and an indicator with an absolute value greater than 1.0. Heywood cases can be caused by specification errors, nonidentification of the model, the presence of outlier cases that distort the solution, a combination of small sample sizes (e.g., $N < 100$) and only two indicators per factor in a measurement model, bad start values, or extremely high or low population correlations that result in empirical underidentification (Chen, Bollen, Paxton, Curran, & Kirby, 2001). An analogy may help give a context for Heywood cases: ML estimation (and related methods) is like a religious fanatic in that it so believes the model's specification that it will do anything, no matter how implausible, to force the model on the data (e.g., estimated correlations > 1.0). Note that some SEM computer programs do not permit certain Heywood cases to appear in the solution. For example, EQS does not allow the estimate of an error variance to be less than zero; that is, it sets a lower bound of zero (i.e., an inequality constraint) that prevents a negative variance estimate. However, solutions in which one or more estimates have been constrained by the computer to prevent an illogical value may indicate a problem (i.e., they should not be trusted). Researchers should also attempt to determine the source of the problem instead of constraining an error variance to be positive in a computer program for SEM and then rerunning the analysis (Chen et al., 2001).

rs, whatever final set of esti-
rrant little confidence. Also,
lly generate their own start
er, that computer-derived start
olutions. Although the com-
mally pretty good, sometimes
le better start values for the
calculating start values for
ay be helpful. Another tactic
a the number of iterations to
llowing the computer more

recursive path models, it can
s that a converged solution is
ster estimate with an illogical
clude negative variance esti-
factor and an indicator with
ood cases can be caused by
the model, the presence of
combination of small sample
tors per factor in a measure-
y high or low population cor-
identification (Chen, Bollen,
logy may help give a context
related methods) is like a reli-
odel's specification that it will
to force the model on the data
that some SEM computer pro-
esses to appear in the solution
mate of an error variance to be
ind of zero (i.e., an inequality
ance estimate. However, solu-
have been constrained by the
may indicate a problem (i.e.,
; should also attempt to deter-
of constraining an error vari-
m for SEM and then rerunning

The ML method is generally both scale free and scale invariant. The former means that if a variable's scale is linearly transformed, a parameter estimated for the transformed variable can be algebraically converted back to the original metric. The latter means the value of the ML fitting function in a particular sample remains the same regardless of the scale of the observed variables (Kaplan, 2000). However, ML estimation may lose these properties if a correlation matrix is analyzed instead of a covariance matrix. Some special methods to correctly analyze a correlation matrix are discussed in Chapter 7.

When a raw data file is analyzed, standard ML estimation assumes there are no missing values. A special form of ML estimation available for raw data files where some observations are missing at random was described earlier (section 3.2). The statistical assumptions of ML estimation include independence of the observations, multivariate normality of the endogenous variables, independence of the exogenous variables and disturbances, and correct specification of the model. The requirement for correct specification is critical because full-information methods, such as ML estimation, tend to propagate errors throughout the whole model. That is, specification error in one parameter can affect results for other parameters elsewhere in the model. It is difficult to predict the specific direction and magnitude of this error because it depends in part upon the relation between the incorrect parameters and other parameters, but the more serious the specification error, the more serious may be the resulting bias. An additional assumption when a path model is analyzed is that the exogenous variables are measured without error, but this requirement is not specific to ML estimation. See Kaplan (2000, chap. 5) for more information about the statistical assumptions underlying SEM.

Reviewed in Chapter 7 are results of some computer simulation studies of the robustness of ML estimation against violation of the assumption of multivariate normality. Briefly summarized, these results suggest that although the values of parameter estimates generated by ML are relatively robust against nonnormality, results of statistical tests may be positively biased, which means that they lead to the rejection of the null hypothesis too often. Ways to deal with this problem are discussed in Chapter 7, but it can be said here that corrective measures should be taken if distributions of endogenous variables are severely nonnormal.

Interpretation of Parameter Estimates

However complex the mathematics of ML estimation, the interpretation of ML estimates for path models is relatively straightforward. Path coefficients are interpreted as regression coefficients in multiple regression, which means that they control for correlations among multiple presumed causes. In the unstandardized solution, disturbance variances are estimated in the metric of the unexplained variance of the corresponding endogenous variable. Suppose that the observed variance of Y_1 in a recursive path model is 25.00 and the estimated variance of its disturbance is 15.00. From these results it can be concluded that $15.00/25.00 = .60$, or 60%, of the variability in Y_1 is not explained by the observed variables presumed to directly affect it. Accordingly, the term $1 - .60 = .40$, or 40%, is the proportion of total variance in Y_1 explained by the model; it also equals the squared multiple correlation (R_{smc}^2) for this endogenous variable. A technical problem with R_{smc}^2 as a measure of the proportion of explained variance for an endogenous variable in a feedback loop (i.e., the path model is nonrecursive) and a corrected measure are discussed in Chapter 9.

In the standardized solution, the variances of all variables (including the disturbances) equal 1.0. However, some SEM computer programs report standardized estimates for disturbances that are proportions of unexplained variance; that is, they equal $1 - R_{smc}^2$ for each endogenous variable. The square root of $1 - R_{smc}^2$ equals the standardized residual path coefficient for the direct effect of the disturbance on the corresponding endogenous variable. For example, if the proportion of explained variance is $R_{smc}^2 = .40$, then the proportion of unexplained variance is $1 - .40 = .60$ and the standardized residual path coefficient equals $.60^{1/2} = .775$. This example shows that when residual path coefficients are fixed to 1.0 in the unstandardized solution to scale the disturbance (i.e., by imposing a ULI constraint), their values are typically not also 1.0 in the standardized solution.

In standard ML estimation, standard errors are calculated only for the unstandardized solution. This means that results of statistical tests (i.e., ratios of parameter estimates over their standard errors) are available only for the unstandardized solution. Users of SEM computer programs often assume that results of statistical tests of unstandardized estimates apply to the corresponding standardized estimates. For sam-

stimation, the interpretation is relatively straightforward. Path coefficients in multiple regression relations among multiple solution, disturbance variance explained variance of the error that the observed variance and the estimated variance is it can be concluded that variance in Y_1 is not explained by the model affect it. Accordingly, the proportion of total variance in Y_1 unexplained multiple correlation problem with R^2_{smc} as a variance for an endogenous model is nonrecursive) and a 9.

ances of all variables (including some SEM computer programs) disturbances that are proportionally equal $1 - R^2_{smc}$ for each $- R^2_{smc}$ equals the standardized effect of the disturbance on example, if the proportion of unexplained residual path coefficient at when residual path coefficient solution to scale the disturbance), their values are typically

errors are calculated only for that results of statistical tests (standard errors) are available users of SEM computer programs. Statistical tests of unstandardized standardized estimates. For sam-

ples that are large and representative, this assumption may not be problematic. Readers should realize, however, that the level of statistical significance of an unstandardized estimate does not automatically apply to its standardized counterpart. Furthermore, standard ML estimation may derive incorrect standard errors when the variables are standardized. The reasons for this potential inaccuracy are technical (e.g., Bollen, 1989, pp. 123–126; Cudeck, 1989), but this is another argument against the analysis of a correlation matrix instead of a covariance matrix with standard ML estimation. However, there are basically two ways to obtain more accurate estimated standard errors for the standardized solution. Some SEM computer programs, such as Amos, use bootstrapping to generate standard errors for standardized estimates. Another method is constrained estimation, which is available as a user-specified option in some SEM computer programs including SEPATH and RAMONA (Chapter 7).

A common question about path coefficients is this: what indicates a "large" direct effect?—a "small" one? Statistical tests of path coefficients do not really provide an answer to this question. Results of statistical tests reflect not only the absolute magnitudes of path coefficients but also other factors, such as the sample size and intercorrelations among the variables. Because SEM is generally a large-sample technique, it may happen more often than in other kinds of analyses that truly small effects are statistically significant. The interpretation of standardized path coefficients is an alternative, but there are few guidelines about what is a "large" versus a "small" effect. Some suggestions for the interpretation of effect size magnitude for standardized path coefficients are offered in Appendix 5.B.

5.8 OTHER ISSUES

Other conceptual issues about PA are addressed below, but many generalize to other SEM methods.

What to Do after a Path Analysis

General requirements for estimating causality from covariances were discussed earlier. The statistical methods used in PA, such as ML esti-

mation, usually require additional assumptions about the population distributions of either the observed endogenous variables (i.e., multivariate normal) or the disturbances (i.e., normal and homoscedastic). Given a single study, it would be almost impossible to believe that all of these logical and statistical requirements were met. Thus, the interpretation that direct effects in a path model correspond to causal relations in the real world is typically unwarranted. It is only with the accumulation of the following types of evidence that the results of a PA may indicate causality: (1) replication of the model across independent samples; (2) elimination of plausible equivalent models; (3) corroborating evidence from experimental studies of variables in the model that are manipulable; and (4) the accurate prediction of the effects of interventions; see Mulaik (2000) for more information.

Consequences of Measurement Error

An assumption of PA is that the exogenous variables are measured without error. The consequences of minor violations of this requirement are not very critical. More serious violations result in biased estimates of direct effects. This bias affects not only the path coefficients of exogenous variables measured with error but also those of other exogenous variables in the model. However, it is difficult to anticipate the direction of this bias. Depending on the intercorrelations among the exogenous variables, some path coefficients may be biased upward (i.e., they are too high) but others may be biased in the other direction. For example, a direct effect that is actually zero may have a nonzero path coefficient, a true negative direct effect may have a positive path coefficient, or the absolute value of the path coefficient may be less than the true value (i.e., attenuation), all the result of measurement error. In an actual analysis, though, the direction of bias for individual path coefficients is difficult to predict. The amount of bias may be less if intercorrelations among measures of exogenous variables are low or if the magnitudes of true direct effects on the endogenous variables are low.

Although there is no assumption that the endogenous variables are measured without error, the use of psychometrically deficient measures of them has somewhat different consequences than for exogenous variables. The effect of measurement error on endogenous variables is

ns about the population
ogenous variables (i.e.,
., normal and homosced-
ost impossible to believe
ents were met. Thus, the
odel correspond to causal
anted. It is only with the
ice that the results of a PA
model across independent
latent models; (3) corrobora-
tion of variables in the model
rediction of the effects of
information.

s variables are measured
violations of this require-
ments result in biased esti-
mately the path coefficients of
also those of other exoge-
nous variables. It is difficult to anticipate the
tercorrelations among the
s may be biased upward
used in the other direction.
zero may have a nonzero
may have a positive path
th coefficient may be less
the result of measurement
ction of bias for individual
amount of bias may be less
enous variables are low or
e endogenous variables are

e endogenous variables are
metrically deficient measures
es than for exogenous vari-
1 endogenous variables is

manifested through their disturbances in addition to the effects of omitted causes. This leads to a potential interpretive confound, but one that affects *standardized* estimates of direct effects on the endogenous variables but not *unstandardized* ones. Standardized path coefficients tend to be too small when endogenous variables are measured with error. Values of unstandardized path coefficients for the same direct effects, on the other hand, are not biased by measurement error. This pattern of bias due to measurement error in the endogenous variables assumes that the exogenous variables have no error. If the exogenous variables are measured with error, these effects for the endogenous variables could be amplified, diminished, or even canceled out. It is best not to hope for the latter, however.

The use of more than a single measure of a construct is the most common way to deal with the problem of measurement error; this topic is introduced in Chapter 7. Scores from multiple indicators tend to be more reliable and valid than those from a single indicator. The technique of PA does not readily accommodate a multiple-indicator approach to measurement, but other SEM techniques do. In particular, the evaluation of an SR model with both measurement and structural components can be viewed as a type of latent variable PA that allows the use of multiple indicators. This topic is covered in Chapter 8, as is a method to correct for measurement error in single indicators. Readers who wish to learn more about measurement error in SEM are referred to Bollen (1989, chap. 5), Kenny (1979, chap. 5), and Maruyama (1998, chap. 5).

Suppression in Path Models

Suppression can also occur among variables in path models. A general definition is that suppression is indicated when the path coefficient for a predictor has a different sign or is greater in absolute value than that predictor's bivariate correlation with the endogenous variable. Because suppression is the result of correction for correlated causes, at least two causal variables must be involved in suppression. Maasen and Bakker (2001) describe various types of suppression effects that can be observed in path models, and generally they are counterparts of forms of suppression in regression analyses, including classical, reciprocal, and negative suppression. Maasen and Bakker also note that measurement error can mask the presence of a suppression effect in PA.

5.9 SUMMARY

This chapter has introduced the basic logic of specifying structural models with observed variables (path models). It is the same logic that underlies the specification of structural models with latent variables that represent hypothetical constructs. Specification is primarily a rational exercise, that is, it should be based on theory and results of prior empirical studies. Also introduced in this chapter was maximum likelihood estimation, the most widely used method for analyzing most kinds of structural equation models. It is a normal-theory, full-information method that simultaneously analyzes all model equations in an iterative algorithm. General statistical assumptions include independence of observations, independence of exogenous variables and disturbances, multivariate normality, and correct specification of the model. With these concepts in mind, we are ready to delve into the details of path analysis.

5.10 RECOMMENDED READINGS

- Kenny, D. A. (1979). *Correlation and causality*. New York: Wiley. (Chaps. 3–5)
- Klem, L. (1995). Path analysis. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 65–98). Washington, DC: American Psychological Association.

Appendix 5.A

Recommendations for Start Values

These recommendations for the generation of start values for path coefficients and disturbance variances are based on the guidelines about effect size magnitude outlined in Appendix 5.B. If the researcher is using different numerical interpretations of these effect size magnitudes, they can be substituted in the following equations. Suppose that a researcher believes that the direct effect of X on Y is positive and of "medium" magnitude. A reasonable start value for the standardized coefficient for the path $X \rightarrow Y$ would be .30; for the unstandardized coefficient, the start value would be $.30 (SD_Y/SD_X)$. If the expected magnitude were "small" or "large," then, respectively, .10 or .50 would be substituted for .30. Also suppose that a researcher believes that the magnitude of the predictive power of all variables with direct effects on Y (including X) is "large." A component of the start value for the disturbance variance for Y in standardized terms could be .75, which corresponds to 25% explained variance and 75% unexplained. (The former is the square of a "large" correlation, $.50^2$.) The start value for the unstandardized disturbance variance would be $.75 (s_Y^2)$. If instead the researcher believed that the magnitude of the predictive power is "small" or "medium," then .99 (i.e., 1% explained variance, or $.10^2$) or .91 (i.e., 9% explained, or $.30^2$), respectively, would be substituted for .75.

Appendix 5.B

Effect Size Interpretation of Standardized Path Coefficients

It is difficult to suggest interpretive guidelines that will prove useful across different research areas. This is because what may be considered a "large" effect in one area may be seen as modest in another. Also, some types of

endogenous variables are more difficult to predict than others. For example, the predictive validity of IQ scores against scholastic achievement is quite strong, but it is less so against other variables such as occupational success.

Given these limitations, some suggestions about the interpretation of the absolute magnitudes of path coefficients are offered. They are based on recommendations by J. Cohen (1988) about effect size interpretations of correlations in the social sciences. Standardized path coefficients with absolute values less than .10 may indicate a "small" effect; values around .30 a "typical" or "medium" effect; and "large" effects may be indicated by coefficients with absolute values $\geq .50$. These guidelines are intended for cases where the researcher has little theoretical or empirical basis to differentiate between smaller versus larger effects, which is most likely to happen in new research areas. In more established areas, one should look to the research literature (e.g., meta-analytic studies) to gauge whether an observed effect size is smaller or larger. The above guidelines should not be rigidly interpreted. For example, one standardized path coefficient of .49 and another of .51 should not be considered as qualitatively different because the former indicates a "medium" effect and the latter a "large" effect. See Bollen (1989, pp. 137-138) for other recommendations about effect size interpretations of standardized path coefficients.

dict than others. For example, scholastic achievement is influenced by variables such as occupational

about the interpretation of effect sizes offered. They are based on effect size interpretations of standardized path coefficients with "small" effect; values around .20 or greater may be indicated by guidelines intended for use in empirical basis to differentiate which is most likely to happen in different areas, one should look to guidelines (e.g., to gauge whether above guidelines should not be used) to determine whether a standardized path coefficient of .49 is qualitatively different from the former and the latter a "large" effect. Other recommendations about standardized path coefficients.

6

Details of Path Analysis

This chapter deals with the details of conducting a path analysis (PA). The discussion begins with the analysis of the path model of illness factors introduced earlier (section 4.2). How to interpret parameter estimates, calculate indirect and total effects, and compare observed correlations or covariances with their predicted counterparts are discussed in the context of this example. Also introduced here are statistical indexes of overall model fit and strategies for testing hypotheses about path models. Most of these analysis details generalize to the other types of structural equation models considered in later chapters, so they warrant careful review.

6.1 DETAILED ANALYSIS OF A RECURSIVE MODEL OF ILLNESS FACTORS

Briefly reviewed here is the work of D. L. Roth, Wiebe, Fillingim, and Shay (1989), who administered measures of exercise, hardiness, fitness, stress, and illness to 373 university students. Reported in Table 6.1 is a matrix summary of these data. The difference between the smallest and largest observed variances (i.e., $3.80^2 = 14.44$ and $624.80^2 = 390,375.04$ for, respectively, hardiness and illness) is so great that the covariance matrix for the original data is ill scaled, which can cause iterative estimation to fail. To avoid this potential problem, the original variables were multiplied by the constants listed in Table 6.1 in order to make their variances more homogeneous.

Presented in Figure 6.1 is the recursive path model specified by

TABLE 6.1. Input Data (Correlations, Standard Deviations) for Analysis of a Path Model of Illness Factors

Variable	1	2	3	4	5
1. Exercise	1.00				
2. Hardiness	-.03	1.00			
3. Fitness	.39	.07	1.00		
4. Stress	-.05	-.23	-.13	1.00	
5. Illness	-.08	-.16	-.29	.34	1.00
M	40.90	0.00	67.10	4.80	716.70
Original SD	66.50	3.80	18.40	6.70	624.80
Constant	1.00	10.00	4.00	10.00	.10
New SD	66.50	38.00	36.80	67.00	62.48

Note. These data are from D. L. Roth et al. (1989); $N = 373$. Means are reported but not analyzed. Note that low scores on the hardiness measure used by these authors indicate greater hardiness. In order to avoid confusion due to negative correlations, the signs of the correlations that involve the hardiness measure were reversed before they were recorded in this table.

D. L. Roth et al. (1989). The direct effects in the figure depicted with dashed lines were predicted by the authors to be zero. These predictions represent the hypotheses that (1) the effect of exercise on illness is indirect and mediated only by fitness, (2) the effect of hardiness on illness is indirect and mediated only by stress, and (3) the direct effect of fitness on stress is zero. These paths are included in the model analyzed next in order to eventually test these hypotheses. This path model is just-identified: with 5 observed variables, there are $5(6)/2 = 15$ observations; the number of free model parameters is also 15, including the variances of 5 exogenous variables (2 observed and 3 unobserved, that is, the disturbances), 1 covariance between the observed exogenous variables, and a total of 9 direct effects. Thus, the model degrees of freedom are zero ($df_M = 0$).

Parameter Estimates

The path model of Figure 6.1 was fitted to a covariance matrix constructed from the correlations and standard deviations of Table 6.1

Deviations) for Analysis of a Path

	3	4	5
1.00			
-13	1.00		
-29	.34	1.00	
7.10	4.80	716.70	
.840	6.70	624.80	
4.00	10.00	.10	
.680	67.00	62.48	

$N = 373$. Means are reported but not asure used by these authors indicate negative correlations, the signs of the reversed before they were recorded in

s in the figure depicted with ors to be zero. These predicate effect of exercise on illness (2) the effect of hardiness on stress, and (3) the direct effect e included in the model analysis hypotheses. This path variables, there are $5(6)/2 = 15$ parameters is also 15, includes (2 observed and 3 unobtrusive between the observed ect effects. Thus, the model

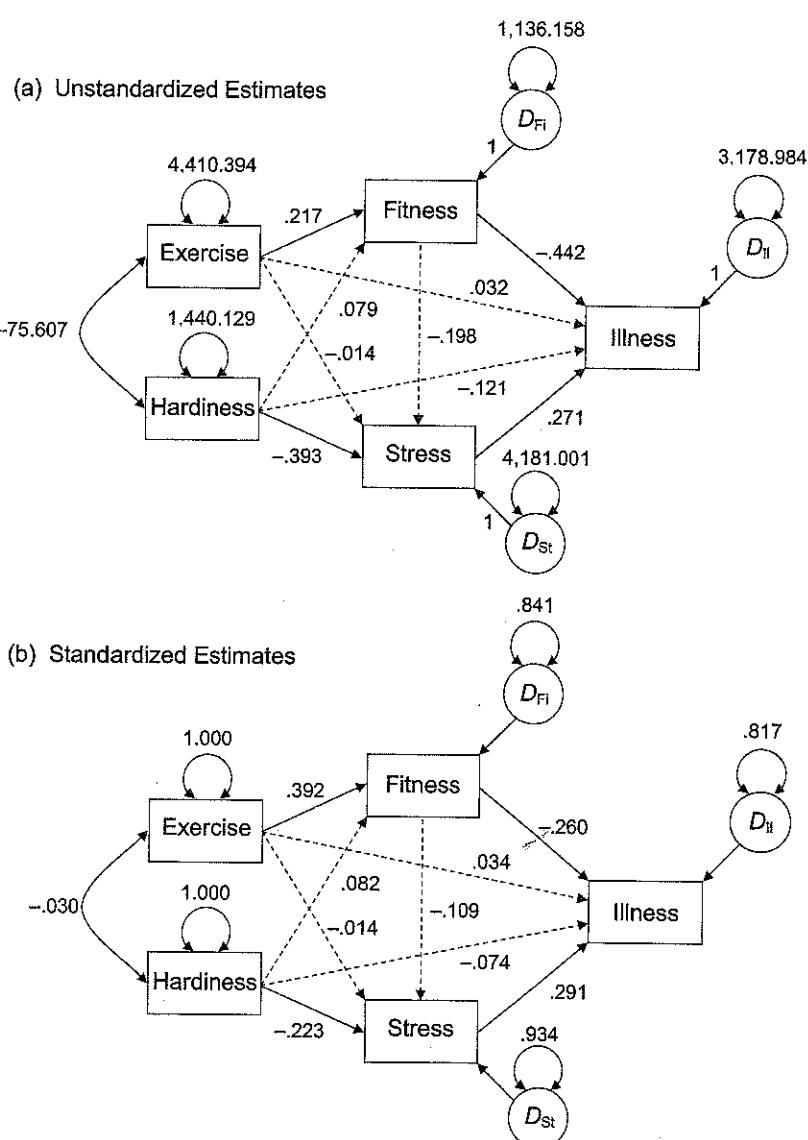


FIGURE 6.1. A recursive path model of illness factors. Paths represented with dashed lines were presumed zero by D. L. Roth et al. (1989). Standardized estimates for the disturbances are proportions of unexplained variance.

to a covariance matrix con lard deviations of Table 6.1

with Amos 5 (Arbuckle, 2003). The analysis converged to an admissible solution. Reported in Table 6.2 are the maximum likelihood (ML) estimates of model parameters. Let us consider first the unstandardized solution, which is also reported in Figure 6.1(a). For example, the unstandardized estimate of the direct effect of exercise on fitness is .217. This means that a 1-point increase on the exercise variable predicts a .217-point increase on the fitness variable, when we control for hardiness. The estimated standard error for this direct effect is .026 (Table 6.2), so $z = .217/.026 = 8.35$, which exceeds the critical value of 2.58 for two-tailed statistical significance at the .01 level. The unstandardized path coefficient for the direct effect of hardiness on fitness is .079; thus, a 1-point increase on the hardiness variable predicts a .079-point increase on the fitness variable, when we control for exercise. The estimated standard error is .046, so $z = .079/.046 = 1.72$. This

TABLE 6.2. Maximum Likelihood Parameter Estimates for a Just-Identified Path Model of Illness Factors

Parameter	Unstandardized	SE	Standardized
Direct effects			
Exercise → Fitness	.217**	.026	.392
Hardiness → Fitness	.079	.046	.082
Exercise → Stress	-.014	.055	-.014
Hardiness → Stress	-.393**	.089	-.223
Fitness → Stress	-.198*	.099	-.109
Exer → Illness	.032	.048	.034
Hardiness → Illness	-.121	.079	-.074
Fitness → Illness	-.442**	.087	-.260
Stress → Illness	.271**	.045	.291
Variances and covariances			
Exercise	4,410.394**	323.386	1.000
Hardiness	1,440.129**	105.595	1.000
Exercise \uparrow Hardiness	-75.607	130.726	-.030
D_{Fl}	1,136.158**	83.307	.841
D_{St}	4,181.001**	306.566	.934
D_{Il}	3,178.984**	233.094	.817

Note. Standardized estimates for disturbance variances are proportions of unexplained variance.

* $p < .05$; ** $p < .01$.

alysis converged to an admissible maximum likelihood (ML) value. Consider first the unstandardized path coefficients in Figure 6.1(a). For example, the effect of exercise on fitness is .35, which exceeds the critical value at the .01 level. The direct effect of hardiness on fitness is .082, which is statistically significant at the .05 level, .079/.046 = 1.72. This

Estimates for a Just-Identified Path

SE	Standardized
.026	.392
.046	.082
.055	-.014
.089	-.223
.099	-.109
.048	.034
.079	-.074
.087	-.260
.045	.291
<hr/>	
<hr/>	
323.386	1.000
105.595	1.000
130.726	-.030
83.307	.841
306.566	.934
233.094	.817

Proportions of unexplained variance are proportions of unexplained

value is less than that required for statistical significance at the .05 level, 1.96. Other unstandardized path coefficients in Table 6.2 and Figure 6.1(a) are interpreted in similar ways. Overall, the pattern of direct effects with statistically significant unstandardized path coefficients is consistent with the predictions of D. L. Roth et al. (1989). The only exception is for the path Fitness → Stress, for which the unstandardized coefficient (-.198) is statistically significant at the .05 level (Table 6.2).

Estimates of the unstandardized variances and covariances of the observed exogenous variables, exercise and hardiness, are just the observed values (Table 6.2, Figure 6.1(a)). The Amos program calculates variances of observed variables as $S^2 = SS/N$ instead of $s^2 = SS/df$, but note that $df/N \times s^2 = S^2$. The estimated disturbance variances reflect unexplained variability for each endogenous variable. For example, the estimated disturbance variance for fitness is 1,136.158. The observed variance of this variable is $s^2 = 36.80^2$ (Table 6.1), so $S^2 = (372/373)1354.24 = 1,350.609$. The ratio of the disturbance variance over the observed variance is $1,136.158/1,350.609 = .841$. That is, the proportion of total variance in fitness not explained by its presumed direct causes, exercise and hardiness, is .841. The proportion of explained variance for fitness thus equals $R^2_{smc} = 1 - .841 = .159$. The estimated disturbance variances for the two other endogenous variables, stress and illness, are interpreted in similar ways. Values of R^2_{smc} for these variables are, respectively, .066 and .183. (Readers should verify these results.)

Because not all variables have the same scale, the unstandardized estimates for different variables cannot be directly compared. However, this is not a problem for the standardized solution, which is reported in Table 6.2 and Figure 6.1(b). Note that there are no standard errors for the standardized estimates, which is typical for the default ML method in most SEM computer programs. (Amos can use bootstrapping to generate standard errors for standardized estimates, but these results are not reported here.) For example, the estimated standardized path coefficients for the direct effects of exercise and hardiness on fitness are, respectively, .392 and .082. That is, a level of exercise one full standard deviation above the mean predicts a fitness level almost .40 standard deviations above the mean, controlling for hardiness. Likewise, a level of hardiness one full standard deviation above the mean is associated with a

fitness level about .10 standard deviations above the mean, when we control for exercise. The magnitude of the standardized direct effect of exercise on fitness is thus about four times greater than that of hardness. Results for the seven other standardized direct effects are interpreted in similar ways. Overall, the pattern of the standardized direct effects is consistent with D. L. Roth and colleagues' hypotheses (1989); specifically, paths predicted to be zero generally have standardized coefficients less than .10 in absolute value (see Figure 6.1(b)). The only exception is the path Fitness → Stress (standardized coefficient = -.109).

Indirect and Total Effects

Indirect effects are estimated statistically as the product of direct effects, either standardized or unstandardized, that comprise them. They are also interpreted just as path coefficients. For example, the standardized indirect effect of fitness on illness through stress is estimated as the product of the standardized coefficients for the paths Fitness → Stress and Stress → Illness, or $-.109 (.291) = -.032$ (see Figure 6.1(b)). The rationale for this derivation is as follows: fitness has a certain direct effect on stress (-.109), but only part of this effect, .291 of it, is transmitted to illness. The result -.032 says that illness level is expected to decrease by about .03 standard deviations for every increase in fitness of one full standard deviation via its prior effect on stress.

The unstandardized indirect effect of fitness on illness through stress is estimated as the product of the unstandardized coefficients for the same two paths, or $-.198 (.271) = -.054$ (see Figure 6.1(a)). That is, the score on the illness variable in its original metric is expected to decrease by about .05 points for every 1-point increase on the fitness variable in its original metric via its prior effect on stress. A full standard deviation on the fitness variable is $S = (372/373)^{1/2} 36.80 = 36.751$ (Table 6.1). Therefore, an increase of one full standard deviation on the fitness variable predicts a decrease of 36.751 (-.054), or -1.985 points on the illness variable through the mediator of fitness. The standard deviation of the illness variable is $S = (372/373)^{1/2} 62.48 = 62.396$. A raw score change of -1.985 points thus corresponds to $-1.985/62.396 = -.032$ standard deviations, which matches the standardized estimate of this indirect effect calculated earlier.

bove the mean, when we condardized direct effect of exer-eater than that of hardiness. rect effects are interpreted in standardizes direct effects is conpotheses (1989); specifically, standardized coefficients less β). The only exception is the cient = $-.109$.

y as the product of direct dized, that comprise them. efficients. For example, the llness through stress is esti-d coefficients for the paths r $-.109 (.291) = -.032$ (see ivation is as follows: fitness , but only part of this effect, esult $-.032$ says that illness standard deviations for every iation via its prior effect on

fitness on illness through standardizes coefficients for $.54$ (see Figure 6.1(a)). That riginal metric is expected to joint increase on the fitness effect on stress. A full stan-($372/373)^{1/2} 36.80 = 36.751$ ill standard deviation on the $.1 (-.054)$, or -1.985 points tor of fitness. The standard $(2/373)^{1/2} 62.48 = 62.396$. A esponds to $-1.985/62.396 =$ he standardized estimate of

Total effects are the sum of all direct and indirect effects of one variable on another. For example, the standardized total effect of fitness on illness is the sum of its direct effect, $-.260$ (see Figure 6.1(b)), and indirect effect through stress, $-.032$, or $-.292$ altogether. Total standardizes effects are also interpreted as path coefficients, and the value of $-.292$ means that increasing fitness by one standard deviation reduces illness by this amount via all presumed direct and indirect causal links between these variables. Unstandardized estimates of total effects are calculated the same way but with unstandardized coefficients. For example, the unstandardized total effect of fitness on illness is the sum of its direct effect, $-.442$, and indirect effect through fitness, $-.198 (.271) = -.054$ (see Figure 6.1(a)), or $-.496$ altogether. That is, for every 1-point increase on the fitness variable in its original metric, we expect about a half-point decrease on the illness variable in its original metric via all presumed causal pathways between these variables.

Some SEM computer programs optionally generate an effects decomposition, a tabular summary of estimated direct, indirect, and total effects. This is fortunate because it can be tedious and error prone to calculate these effects by hand. The Amos program can print both total effects and total indirect effects. The latter is the sum of all indirect effects of a causally prior variable on a subsequent one. For example, fitness is specified to have a single indirect effect on illness (through stress; see Figure 6.1). However, the exercise and hardiness variables each have three possible indirect effects on illness, including one though fitness, another through stress, and the third through both variables (e.g., Exercise \rightarrow Fitness \rightarrow Stress \rightarrow Illness).

In standard ML estimation, Amos does not report standard errors for total effects or total indirect effects. However, Amos can use the method of bootstrapping to estimate the standard error for unstandardized or standardized total effects and total indirect effects. Reported in Table 6.3 is the effects decomposition printed by Amos with standard errors for the unstandardized results only estimated across 1,000 random samples generated by the computer from the observed covariance matrix (Table 6.1). The bootstrapped standard errors for the unstandardized direct effects in Table 6.3 are very similar to the corresponding ones in Table 6.2 from the standard ML analysis. Note in Table 6.3 that stress has only a direct effect on illness, which is also its total effect. The variable fitness has just one indirect effect on illness

TABLE 6.3. Effects Decomposition for a Just-Identified Path Model of Illness Factors

Causal variable	Endogenous variables								
	Fitness			Stress			Illness		
	Unst.	SE	St.	Unst.	SE	St.	Unst.	SE	St.
Exercise									
Direct effect	.217**	.026	.392	-.014	.055	-.014	.032	.047	.034
Total indirect effects	—	—	—	-.043*	.022	-.043	-.112**	.026	-.119
Total effect	.217**	.026	.392	-.057	.050	-.057	-.080	.047	-.085
Hardiness									
Direct effect	.079	.046	.082	-.393**	.091	-.223	-.121	.076	-.074
Total indirect effects	—	—	—	-.016	.013	-.009	-.146**	.039	-.089
Total effect	.079	.046	.082	-.409**	.091	-.232	-.267**	.082	-.163
Fitness									
Direct effect	—	—	—	-.198*	.097	-.109	-.442**	.085	-.260
Total indirect effects	—	—	—	—	—	—	-.054	.028	-.032
Total effect	—	—	—	-.198*	.097	-.109	-.496**	.090	-.292
Stress									
Direct effect	—	—	—	—	—	—	.271**	.045	.291
Total indirect effects	—	—	—	—	—	—	—	—	—
Total effect	—	—	—	—	—	—	.271**	.045	.291

Note. Unst., unstandardized; St., standardized.

* $p < .05$; ** $p < .01$.

(through stress), but only the unstandardized direct and total effects of this predictor are statistically significant. The unstandardized direct effects of exercise and hardiness on illness are also not statistically significant, but the total indirect effects of both variables are so. This pattern of results—statistically significant indirect effects but not direct effects—represents the strongest demonstration for a mediator effect, assuming correct directionality specifications. This pattern of results is also consistent with predictions by D. L. Roth and colleagues (1989;

Path Model of Illness Factors

nous variables

Stress	Illness				
	SE	St.	Unst.	SE	St.
.055	-.014	.032	.047	.034	
.022	-.043	-.112**	.026	-.119	
.050	-.057	-.080	.047	-.085	
.091	-.223	-.121	.076	-.074	
.013	-.009	-.146**	.039	-.089	
.091	-.232	-.267**	.082	-.163	
.097	-.109	-.442**	.085	-.260	
—	—	-.054	.028	-.032	
.097	-.109	-.496**	.090	-.292	
—	—	.271**	.045	.291	
—	—	—	—	—	
—	—	.271**	.045	.291	

direct and total effects of
The unstandardized direct
are also not statistically signifi-
h variables are so. This pat-
direct effects but not direct
ation for a mediator effect,
ns. This pattern of results is
Roth and colleagues (1989;

see Figure 6.1). For more information about the estimation of indirect effects in SEM, see Shrout and Bolger (2002) and Cole and Maxwell (2003).

The Amos program does not calculate standard errors for individual indirect effects of variables with multiple indirect effects on other variables, such as exercise and hardiness in the present example. (This is a characteristic of some other SEM computer programs that report effects decompositions.) However, Baron and Kenny (1986) describe some hand-calculable statistical tests for unstandardized indirect effects in recursive path models that involve only one mediator (e.g., Exercise → Fitness → Illness). The best known of these tests is based on an estimated standard error developed by Sobel (1986), which is described in Appendix 6.A.

Model-Implied or Predicted Covariances and Correlations

The standardized total effect of one variable on another approximates the part of their observed correlation due to presumed causal relations. The sum of the standardized total effects and all other noncausal (e.g., spurious) associations represented in the path model equal model-implied (predicted) correlations that can be compared against the observed correlations. Model-implied covariances have the same general meaning, but they concern the unstandardized solution. All SEM computer programs that calculate model-implied correlations or covariances use matrix algebra methods (e.g., Loehlin, 1998, pp. 43–47). There is an older method for recursive path models amenable to hand calculation known as the tracing rule. It is worthwhile to know about the tracing rule more for its underlying principles than for its now limited utility. The tracing rule is as follows: *a model-implied correlation is the sum of all the causal effects and noncausal associations from all valid tracings between two variables in a recursive path model.* A “valid” tracing means that a variable is not (1) entered through an arrowhead and exited by the same arrowhead nor (2) entered twice in the same tracing.

As an example of the application of the tracing rule to calculate model-implied correlations with the standardized solution, look again

at Figure 6.1(b) and find the exercise and stress variables. There are four valid tracings between them. Two of these tracings make up the standardized total effect (Exercise → Stress, Exercise → Fitness → Stress), which equals $-.057$ (see Table 6.3). The other two tracings involve the unanalyzed association of exercise with another variable, hardness, that has direct or indirect effects on stress. These two tracings are Exercise \curvearrowright Hardiness → Stress and Exercise \curvearrowright Hardiness → Fitness → Stress. Estimates for both tracings are calculated in the same way as indirect effects; that is, as the products of the relevant path coefficients and correlations. The estimate for the Exercise \curvearrowright Hardiness → Stress tracing is calculated as $-.030 (-.223) = .007$, and the estimate for the Exercise \curvearrowright Hardiness → Fitness → Stress tracing is calculated as $-.030 (.082) (-.109) = .0003$. The implied correlation between exercise and stress equals the total effect plus both of the tracings that involve unanalyzed associations between the exogenous variables, or $-.057 + .007 + .0003 = -.0497$. Note that this value rounded to two decimal places ($-.05$) equals the observed correlation between exercise and stress (see Table 6.1). This result is just what we expect because this path model is just-identified ($df_M = 0$).

The difference between a model-implied and an observed correlation is known as a correlation residual. Correlation residuals are standardized covariance residuals—also known as fitted residuals—which are differences between observed and predicted covariances.¹ For the just-identified path model in Figure 6.1, all correlation residuals are zero. For an overidentified path model ($df_M > 0$), however, not all correlation residuals may be zero. There is a rule of thumb in the SEM literature that correlation residuals with absolute values greater than $.10$ suggest that the model does not explain the corresponding observed correlation very well. Although it is difficult to say how many absolute correlation residuals greater than $.10$ are “too many,” the more there are, the worse the explanatory power of the model for specific observed associations. There is no comparable rule of thumb about values of covariance residuals that suggest a poor expla-

¹In EQS output, correlation residuals are labeled as standardized residuals, but in LISREL the same label designates the ratio of a covariance residual over its standard error, which is interpreted in large samples as a z test of whether the population covariance residual is zero.

d stress variables. There are these tracings make up the less, Exercise → Fitness → .3). The other two tracings exercise with another variable, ts on stress. These two tracings and Exercise ↑ Hardiness → ts are calculated in the same cts of the relevant path coefficient the Exercise ↑ Hardiness .23) = .007, and the estimate → Stress tracing is calculated ed correlation between exercis both of the tracings that the exogenous variables, or is value rounded to two decorrrelation between exercise ist what we expect because

ed and an observed correla correlation residuals are stan as fitted residuals—which predicted covariances.¹ For .1, all correlation residuals ($df_M > 0$), however, not all is a rule of thumb in the ith absolute values greater explain the corresponding it is difficult to say how r than .10 are “too many,” atory power of the model is no comparable rule of that suggest a poor expla

standardized residuals, but in riance residual over its standard of whether the population covari

nation because covariances are affected by the scales of the original variables.

Use of the tracing rule is error prone even for relatively simple recursive models because it can be difficult to spot all of the valid tracings. For example, there are a total of 10 different valid tracings between the stress and illness variables of the model in Figure 6.1. All but one of these tracings are for either spurious associations or ones that involve unanalyzed associations between the exogenous variables. This is even another reason to appreciate that many SEM computer programs automatically calculate model-implied correlations as a user-specified option. Note that the tracing rule does not apply to nonrecursive path models. Estimation of indirect and total effects for such models is discussed in Chapter 9.

6.2 ASSESSING MODEL FIT

Overidentified path models with more observations than parameters usually do not perfectly fit the data. There is thus a need to measure the degree of fit of such models. Discussed next are various statistical indexes of overall model fit.

Overview

There are dozens of model fit indexes described in the SEM literature, and new indexes are being developed all the time. Evaluation of the statistical properties of various fit indexes in computer simulation studies is also an active research topic (e.g., Kenny & McCoach, 2003; Marsh, Balla, & Hau, 1996), so the state of knowledge in this area is continuously changing. It is also true that many SEM computer programs print in their output the values of more fit indexes than are typically reported for the analysis.

The availability of so many different fit indexes presents a few problems. One is that different fit indexes are reported in different articles, and another is that different reviewers of the same manuscript may request indexes that they know about or prefer (Maruyama, 1998). This means that it can be difficult for a researcher to decide as to which partic-

ular indexes and which values to report. There is also the possibility for selective reporting of values of fit indexes. For example, a researcher keen to demonstrate good model fit may report only those fit indexes with favorable values. Another problem is that one can become so preoccupied with overall model fit that other crucial information, such as whether the parameter estimates make sense, is overlooked.

Described next is a minimal set of fit indexes that should be reported and interpreted when reporting the results of SEM analyses. This particular set reflects the current state of practice and recommendations about what to report in written summaries of the analysis (e.g., Boomsma, 2000; McDonald & Ho, 2002). These statistics include (1) the model chi-square, (2) the Steiger-Lind root mean square error of approximation (RMSEA; Steiger, 1990) with its 90% confidence interval, (3) the Bentler comparative fit index (CFI; Bentler, 1990), and (4) the standardized root mean square residual (SRMR). Before any of these statistics are described, it is useful to keep in mind the following limitations of basically all fit indexes in SEM:

1. Values of fit indexes indicate only the average or overall fit of a model. It is thus possible that some parts of the model may poorly fit the data even if the value of a particular index seems favorable. See Tomarken and Waller (2003) for discussion of potential problems with models that seem to fit the data well based on values of fit statistics.
2. Because a single index reflects only a particular aspect of model fit, a favorable value of that index does not by itself indicate good fit. This is also why model fit is usually assessed based in part on the values of more than one index. That is, there is no single "magic index" that provides a gold standard for all models.
3. Fit indexes do not indicate whether the results are theoretically meaningful. For example, the signs of some path coefficients may be unexpectedly in the opposite direction. Even if values of fit indexes appear to be favorable, results so anomalous require explanation.
4. Values of fit indexes that suggest adequate fit do not indicate that the predictive power of the model is also high. For example, disturbances of models with even perfect fit to the data can still be large, which means that the model accurately reflects the relative lack of predictive validity among the variables.

re is also the possibility for example, a researcher keen only those fit indexes with he can become so preoccupied information, such as , is overlooked.

t indexes that should be e results of SEM analyses. of practice and recommendaries of the analysis (e.g., These statistics include (1) root mean square error of its 90% confidence inter- FI; Bentler, 1990), and (4) al (SRMR). Before any of keep in mind the following M:

e average or overall fit of a e parts of the model may of a particular index seems : (2003) for discussion of t seem to fit the data well

particular aspect of model s not by itself indicate good ly assessed based in part on That is, there is no single andard for all models.

the results are theoretically of some path coefficients direction. Even if values of sults so anomalous require

lequate fit do not indicate el is also high. For exam- n perfect fit to the data can odel accurately reflects the ong the variables.

5. The sampling distributions of many fit indexes used in SEM are unknown (the RMSEA may be an exception), and the interpretive guidelines suggested later for individual indexes concerning good fit are just that.

Model Chi-Square

The most basic fit statistic is the product $(N - 1) F_{ML}$, where $N - 1$ are the overall degrees of freedom in the sample and F_{ML} is the value of the statistical criterion minimized in ML estimation. In large samples and assuming multivariate normality, $(N - 1) F_{ML}$ is distributed as a Pearson chi-square statistic with degrees of freedom equal to that of the researcher's model, df_M . This statistic is referred to here as the model chi-square, χ^2_M ; it is also known as the likelihood ratio chi-square or generalized likelihood ratio. The value of χ^2_M for a just-identified model generally equals zero and has no degrees of freedom. If $\chi^2_M = 0$, the model perfectly fits the data (i.e., the predicted correlations and covariances equal their observed counterparts). As the value of χ^2_M increases, the fit of an overidentified model becomes increasingly worse. Thus, χ^2_M is actually a "badness-of-fit" index because the higher its value, the worse the model's correspondence to the data.

We continue to assume large samples and multivariate normality. Under the null hypothesis that the researcher's model has perfect fit in the population, χ^2_M approximates a central chi-square distribution. The only parameter of a central chi-square distribution is its degrees of freedom. Other applications of chi-square as a test statistic perhaps more familiar to readers, such as the test of association for two-way contingency tables, also approximate central chi-square distributions as N increases. Also, tables of critical values of chi-square in most statistics textbooks are based on central chi-square distributions.

For an overidentified model, χ^2_M tests the null hypothesis that the model is correct (i.e., it has perfect fit in the population). Suppose that $\chi^2_M = 10.00$ for a model where $df_M = 5$. The exact level of statistical significance associated with this statistic is $p = .075$.² Given this result, the researcher would not reject the null hypothesis at the .05

²This result was obtained from a central chi-square probability calculator freely available over the Internet; see <http://members.aol.com/johnp71/javastat.html>

level. However, if $\chi^2_M = 12.00$ for the same model degrees of freedom, then $p = .035$ and the hypothesis that the model is correct would be rejected at the .05 level (but not at the .01 level). Therefore, it is the *failure* to reject the null hypothesis that supports the researcher's model. This logic is "backward" from the usual *reject-support* context for statistical tests where it is the rejection of the null hypothesis that supports the researcher's theory. However, it is perfectly consistent with an *accept-support* context where the null hypothesis represents the researcher's belief (Steiger & Fouladi, 1997).

Another way of looking at χ^2_M is that it tests the difference in fit between a given overidentified model and a just-identified version of it. Suppose for an overidentified path model that $\chi^2_M > 0$ and $df_M = 5$. Adding five more paths to this model would make it just-identified and reduce both χ^2_M and df_M to zero. This is because almost all just-identified models perfectly explain the data. However, models that are just as complex as the data (i.e., $df_M = 0$) are not very interesting. Raykov and Marcoulides (2000) described each degree of freedom for χ^2_M as a dimension along which the model can potentially be rejected. This is also why, given two different plausible models of the same phenomenon, the one with greater degrees of freedom (i.e., fewest parameters) is generally preferred because it has withstood a greater potential to be rejected. The same idea underlies the *parsimony principle*: given two different models with similar explanatory power for the same data, the simpler model is to be preferred.

There are some problems with relying solely on χ^2_M as a fit statistic. The hypothesis tested by χ^2_M is likely to be implausible. That is, it may be unrealistic to expect a model to have perfect population fit. It is sensitive to the size of the correlations: bigger correlations generally lead to higher values of χ^2_M . This happens in part because larger correlations tend to allow for the possibility of greater differences between observed and model-implied correlations. The model chi-square is also affected by sample size. Specifically, if the sample size is large, which is required in order to interpret the index as a test statistic, the value of χ^2_M may lead to rejection of the model even though differences between observed and predicted covariances are slight. Indeed, rejection of basically any overidentified model based on χ^2_M requires only a sufficiently large number of cases. To reduce the sensitivity of χ^2_M to sample size,

e model degrees of freedom, e model is correct would be 01 level). Therefore, it is the at supports the researcher's usual *reject-support* context n of the null hypothesis that er, it is perfectly consistent e null hypothesis represents , 1997).

it tests the difference in fit a just-identified version of it. el that $\chi^2_M > 0$ and $df_M = 5$. ld make it just-identified and because almost all just-identi-never, models that are just as very interesting. Raykov and ee of freedom for χ^2_M as a ntentially be rejected. This is odels of the same phenomenm (i.e., fewest parameters) is ood a greater potential to be simony principle: given two power for the same data, the

solely on χ^2_M as a fit statistic. implausible. That is, it may rfect population fit. It is sen- r correlations generally lead rt because larger correlations differences between observed el chi-square is also affected ze is large, which is required tatistic, the value of χ^2_M may hough differences between ht. Indeed, rejection of basi- χ^2_M requires only a sufficiently itivity of χ^2_M to sample size,

some researchers divide its value by the degrees of freedom (χ^2_M / df_M), which generally results in a lower value called the **normed chi-square** (NC). However, there is no clear-cut guideline about what value of the NC is minimally acceptable. Bollen (1989) notes that values of the NC of 2.0, 3.0, or even as high as 5.0 have been recommended as indicating reasonable fit and that the NC does not completely correct for the influence of sample size. Other, more sophisticated fit indexes described next are less affected by sample size and have interpretive norms.

Despite problems with χ^2_M as a fit index, it is reported in virtually all reports of SEM analyses. One reason is that the formulas of most if not all other indexes include χ^2_M , so it is a key ingredient. The model chi-square is also useful in the comparison of hierarchical models evaluated with the same data, a topic considered in the next section. Problems with χ^2_M have also motivated the development of numerous supplemental fit statistics, some described next. Basically all of these indexes—and χ^2_M on which they are based—assume multivariate normality for the endogenous variables. If the distributions are severely nonnormal, the value of χ^2_M tends to be too high. This means that (1) true models will be rejected too often when χ^2_M is interpreted as a test statistic and (2) values of supplemental fit indexes based on χ^2_M may be distorted. Corrective measures should be taken if the distributions are severely nonnormal (Chapter 7).

Root Mean Square Error of Approximation

The RMSEA has attracted much interest of late due to a relatively unique combination of properties. It is a parsimony-adjusted index in that its formula includes a built-in correction for model complexity. This means that given two models with similar overall explanatory power for the same data, the simpler model will be favored. It does not approximate a central chi-square distribution. The RMSEA instead approximates a **noncentral chi-square distribution**, which does not require a true null hypothesis. In this case it means that fit of the researcher's model in the population is *not* assumed to be perfect. A noncentral chi-square distribution has an additional parameter known as the **noncentrality parameter**, designated here as δ . This

parameter measures the degree of falseness of the null hypothesis. Specifically, if the null hypothesis is true, then $\delta = 0$ and a central chi-square distribution is indicated. That is, a central chi-square distribution is just a special case of a noncentral chi-square distribution. The value of δ increases as the null hypothesis becomes more and more false. It also serves to shift the noncentral chi-square distribution to the right compared with the central chi-square distribution with the same degrees of freedom (e.g., MacCallum, Browne, & Sugawara, 1996, p. 136.)

In SEM, the parameter δ can be seen as reflecting the degree of misspecification of the researcher's model. It is often estimated as the difference between χ^2_M and df_M or zero, whichever is greater (i.e., the estimate cannot be negative). This estimator is designated here as $\hat{\delta}_M$, which is expressed as follows:

$$\hat{\delta}_M = \max(\chi^2_M - df_M, 0) \quad (6.1)$$

where "max" refers to the maximum value of either of the two expressions in parentheses. Fit indexes, such as the RMSEA, based in part on $\hat{\delta}_M$ reflect the view that the researcher's model is an approximation of reality, not an exact copy of it (Raykov & Marcoulides, 2000).

Along the same lines, Browne and Cudeck (1993) distinguished between two different sources of lack of model fit. One is the error of approximation, which concerns the lack of fit of the researcher's model to the population covariance matrix. The RMSEA measures the error of approximation, and for this reason it is sometimes referred to as a population-based index. The other source of poor model fit concerns the error of estimation, the difference between the fit of the model to the sample covariance matrix and to the population covariance matrix. The error of estimation is affected by sample size—there is greater error in smaller samples—but the error of approximation is not. These two different kinds of errors contribute to overall error, the difference between the population covariance matrix and the model-implied covariance matrix estimated with the sample data. Fit indexes, such as χ^2_M and others that approximate central chi-square distributions, measure overall error and thus are described as sample-based indexes.

The RMSEA is a "badness-of-fit" index in that a value of zero indicates the best fit and higher values indicate worse fit. The formula is

of the null hypothesis, then $\delta = 0$ and a central chi-square distribution. As the null hypothesis becomes more and more noncentral, the chi-square distribution becomes more and more noncentral. MacCallum, Browne, &

s reflecting the degree of fit is often estimated as the difference between the observed and the expected chi-square statistic, designated here as $\hat{\delta}_M$.

) (6.1)

either of the two expressions for RMSEA, based in part on the fact that δ_M is an approximation of $\hat{\delta}_M$. MacCallum et al. (2000).

Cudeck (1993) distinguished between two types of model fit. One is the error of fit of the researcher's model, which measures the error of fit of the model to the covariance matrix. The other type of model fit concerns the fit of the model to the population covariance matrix. The former is referred to as a population model fit, while the latter is referred to as a sample-based model fit. These two types of model fit are related by the formula $\text{RMSEA} = \sqrt{\frac{\delta_M}{df_M(N-1)}}$, where δ_M is the difference between the observed and the expected chi-square statistic, df_M is the degrees of freedom of the model, and N is the sample size.

It is important to note that a value of zero indicates a perfect fit, while a value of .05 indicates a close approximate fit, values between .05 and .08 suggest reasonable error of approximation, and RMSEA $\geq .10$ suggests poor fit (Browne & Cudeck, 1993).

$$\text{RMSEA} = \sqrt{\frac{\hat{\delta}_M}{df_M(N-1)}} \quad (6.2)$$

This equation shows that the RMSEA estimates the amount of error of approximation per model degree of freedom and takes sample size into account. Note that the result $\text{RMSEA} = 0$ says only that $\chi^2_M < df_M$, not that $\chi^2_M = 0$ (i.e., the fit is perfect; see Equation 6.1). A rule of thumb is that $\text{RMSEA} \leq .05$ indicates close approximate fit, values between .05 and .08 suggest reasonable error of approximation, and $\text{RMSEA} \geq .10$ suggests poor fit (Browne & Cudeck, 1993).

Computer programs for SEM that calculate the RMSEA typically also print a 90% confidence interval for the population parameter estimated by the RMSEA, designated here as ε . Confidence intervals for ε are based on the estimated noncentrality parameter $\hat{\delta}_M$, and the lower and upper bounds of such a confidence interval may not be symmetrical around the sample value of the RMSEA (see Steiger & Fouladi, 1997). This interval reflects the degree of uncertainty associated with RMSEA as a point estimate at the 90% level of statistical confidence. It also explicitly acknowledges that the RMSEA (and all other model fit indexes) are sample statistics subject to sampling error.

If the lower bound of a particular 90% confidence interval for ε is less than .05, we would not reject the directional null hypothesis $H_0: \varepsilon_0 \leq .05$, which says that the researcher's model has close approximate fit in the population. (Ideally, zero is the lower bound of a 90% confidence interval for ε .) If it is also true that the upper bound of the same confidence interval does not exceed whatever cutoff value is selected as indicating poor fit (e.g., .10), we can reject the null hypothesis that the fit of the model in the population is just as bad or even worse (e.g., $H_0: \varepsilon_0 \geq .10$). Suppose that $\text{RMSEA} = .045$ with the 90% confidence interval $.045 \pm .150$. Because the lower bound of this interval (.045) is less than .05, the null hypothesis of close approximate fit is not rejected. However, the upper bound of the same confidence interval (.150) exceeds .10, so we cannot reject the hypothesis of poor approximate fit. Thus, the result $\text{RMSEA} = .045$ for this example is subject to a fair amount of sampling error because it is just as consistent with the hypothesis of good approximate fit as it is with the hypothesis of poor approximate fit. This type of "mixed" outcome is more likely to happen in smaller samples.

samples. A larger sample may be required in order to obtain more precise results.

Comparative Fit Index

The CFI was originally associated with EQS but is now printed by other SEM computer programs. It is one of a class of fit statistics known as incremental or comparative fit indexes, which are among the most widely used in SEM. All these indexes assess the relative improvement in fit of the researcher's model compared with a baseline model. The latter is typically the independence model—also called the null model—which assumes zero population covariances among the observed variables. When means are not analyzed, the only parameters of the independence model are the population variances of these variables. Because the independence model assumes unrelated variables, the value of its model chi-square, χ^2_B , is often quite large compared with that of the researcher's model, χ^2_M . To the extent that χ^2_M is less than χ^2_B , the researcher's model is an improvement over the independence model; otherwise, there is no improvement and thus no reason to prefer the researcher's model.

Unlike some older incremental fit indexes described later, the CFI does *not* assume zero error of approximation (i.e., perfect population fit of the researcher's model). The formula is

$$CFI = 1 - \hat{\delta}_M / \hat{\delta}_B \quad (6.3)$$

where $\hat{\delta}_M$ and $\hat{\delta}_B$ estimate the noncentrality parameter of a noncentral chi-square distribution for, respectively, the researcher's model and the baseline model. The estimator $\hat{\delta}_B$ has the same general form as that of $\hat{\delta}_M$ except that the former concerns the difference between χ^2_B and df_B (see Equation 6.1). A rule of thumb for the CFI and other incremental indexes is that values greater than roughly .90 may indicate reasonably good fit of the researcher's model (Hu & Bentler, 1999). However, $CFI = 1.0$ means only that $\chi^2_M < df_M$, not that the model has perfect fit.

All incremental fit indexes have been criticized when the baseline model is the independence (null) model (which is almost always true). This is because the assumption of zero covariances is scientifically

r to obtain more pre-

it is now printed by class of fit statistics es, which are among es assess the relative pared with a baseline odel—also called the variances among the , the only parameters riances of these vari-; unrelated variables, quite large compared extent that χ^2_M is less it over the independ-nd thus no reason to

scribed later, the CFI perfect population fit

(6.3)

neter of a noncentral earcher's model and ame general form as e difference between or the CFI and other roughly .90 may indi-Hu & Bentler, 1999). ot that the model has

ed when the baseline almost always true). nces is scientifically

implausible in many (probably most) applications of SEM. Therefore, finding that the researcher's model has better relative fit than the corresponding independence model may not be very impressive. Although it is possible to specify a different, more plausible baseline model and compute by hand the value of an incremental fit index with its equation, this is rarely done in practice. Widaman and J. Thompson (2003) describe how to specify more plausible baseline models for different kinds of SEM analyses. The Amos 5 program (Arbuckle, 2003) allows the specification of baseline models where covariances among the observed variables are required to be equal instead of zero, which is more realistic. See Marsh et al. (1996) for more information about statistical properties of incremental fit indexes.

Standardized Root Mean Square Residual

The indexes described next are based on covariance residuals, differences between observed and predicted covariances. Ideally, all these residuals should be about zero for good model fit. A statistic called the root mean square residual (RMR) was originally associated with LISREL but is now calculated by other SEM computer programs. It is a measure of the mean absolute value of the covariance residuals. Perfect model fit is indicated by $RMR = 0$, and increasingly higher values indicate worse fit (i.e., it is a badness-of-fit index). One problem with the RMR is that because it is computed with unstandardized variables, its range depends upon the scales of the observed variables. If these scales are all different, it can be difficult to interpret a given value of the RMR. The standardized root mean square residual (SRMR), on the other hand, is based on transforming both the sample covariance matrix and the predicted covariance matrix into correlation matrices. The SRMR is thus a measure of the mean absolute correlation residual, the overall difference between the observed and predicted correlations. Values of the SRMR less than .10 are generally considered favorable.

In addition to the statistics just described, it is also informative to view visual summaries of distributions of the residuals. Specifically, frequency distributions of the correlation residuals or covariance residuals should be generally normal in shape, and a Q-plot of the standardized residuals ordered by their size against their position in the distribution represented by normal deviates (z scores) should follow a diagonal

line. Obvious departures from these patterns may indicate serious misspecification or violation of multivariate normality for the endogenous variables.

Other Fit Indexes

A class of statistics known as predictive fit indexes assess model fit in *hypothetical* replication samples of the same size and randomly drawn from the same population as the researcher's original sample. For this reason, these indexes may be seen as population based rather than sample based. Perhaps the best known predictive fit index under ML estimation is the Akaike information criterion (AIC). It is based on an information theory approach to data analysis that combines estimation and model selection under a single conceptual framework (e.g., D. R. Anderson, Burnham, & W. Thompson, 2000). It is also a parsimony-adjusted index because it favors simpler models. Confusingly, two different formulas for the AIC are reported in the SEM literature. The first is

$$AIC_1 = \chi^2_M + 2q \quad (6.4)$$

where q is the number of free model parameters. Equation 6.4 thus increases the chi-square for the researcher's model by a factor of 2 times the number of free parameters. The second formula is

$$AIC_2 = \chi^2_M - 2df_M \quad (6.5)$$

which decreases the model chi-square by a factor of 2 times the degrees of freedom. Though the two formulas are different, the key is that the relative change in the AIC is the same in both versions, and this change is a function of model complexity.

The AIC and other predictive fit indexes are generally used in SEM to select among competing nonhierarchical models estimated with the same data. (The difference between hierarchical vs. nonhierarchical models is defined in the next section.) Specifically, the model with the smallest AIC is chosen as the one most likely to replicate. This is the model with relatively better fit and fewer parameters compared with competing models. In contrast, more complex models with comparable

patterns may indicate serious
iate normality for the endoge-

fit indexes assess model fit in
same size and randomly drawn
cher's original sample. For this
ulation based rather than sam-
lictive fit index under ML esti-
tion (AIC). It is based on an
alysis that combines estimation
ceptual framework (e.g., D. R.
2000). It is also a parsimony-
models. Confusingly, two dif-
in the SEM literature. The first

$$2q \quad (6.4)$$

parameters. Equation 6.4 thus
er's model by a factor of 2 times
ond formula is

$$2df_M \quad (6.5)$$

a factor of 2 times the degrees
re different, the key is that the
both versions, and this change

exes are generally used in SEM
ical models estimated with the
ierarchical vs. nonhierarchical
specifically, the model with the
likely to replicate. This is the
er parameters compared with
plex models with comparable

overall fit may be less likely to replicate due to greater capitalization on
chance.

Table 6.4 summarizes the characteristics of some miscellaneous fit
indexes that readers may see reported in written summaries of SEM
analyses. Some of these indexes are older than ones described earlier or
have special problems that argue against their use. None are recom-
mended over the indexes described earlier. Listed in the top part of the
table are incremental fit indexes, including the Bentler-Bonett normed
fit index (NFI) and the non-normed fit index³ (NNFI; Bentler &
Bonett, 1980) and the parsimony-adjusted NFI (PNFI; James, Mulaik,
& Brett, 1982). The latter two statistics correct for model complexity
(i.e., they favor simpler models), but note that the PNFI is quite sensi-
tive to model size. This means that the penalty for complexity tends to
be greater when the total number of observed variables in the model is
relatively small, such as 10 or fewer. In small samples, it is possible for
values of the NNFI to be much lower than those of other fit indexes.
The relative noncentrality index (RNI; McDonald & Marsh, 1990) is
similar to the CFI except that its values can be negative. In this regard
the CFI is a better index because its range is 0–1.0.

Described in the middle part of Table 6.4 are miscellaneous predic-
tive fit indexes. Like the AIC, they can be used to select among compet-
ing nonhierarchical models fitted to the same data; specifically, the
model with the lowest values of these indexes is preferred. The
expected cross-validation index (ECVI; Browne & Cudeck, 1993)
results in the same rank ordering of competing models as the AIC, and
some SEM computer programs print a 90% confidence based on the
ECVI. The consistent Akaike information criterion (CAIC; Bozdogan,
1987) takes direct account of sample size, unlike the AIC. Another
index known as the Bayes information criterion (BIC; Raftery, 1993),
penalizes complexity more than either the AIC or the CAIC.

Listed in the bottom part of Table 6.4 are absolute fit indexes that
estimate the proportion of variability in the sample covariance matrix
explained by the model (i.e., the predicted covariance matrix). The
goodness-of-fit index (GFI) was the very first standardized fit index
(Jöreskog & Sörbom, 1981). Originally associated with LISREL, the
GFI is now calculated by some other SEM computer programs. It is

³Also known as the Tucker-Lewis index.

TABLE 6.4. Formulas and Descriptions of Miscellaneous Fit Indexes

Index	Formula	Comment
Incremental fit indexes		
NFI	$1 - \chi_M^2/\chi_B^2$	Sample-based
NNFI	$1 - NC_M/NC_B$	Sample-based, parsimony-adjusted, value can fall outside of range 0–1.0
PNFI	$(df_M/df_B)NFI$	Sample-based, parsimony-adjusted, sensitive to model size
RNI	$1 - (\chi_M^2/df_M)/(\chi_B^2/df_B)$	Population-based, value can be negative
PCFI	$(df_M/df_B)CFI$	Population-based, parsimony-adjusted, sensitive to model size
Predictive fit indexes		
ECVI	$(\chi_M^2 - 2q)/(N - 1)$	Population-based, parsimony-adjusted
CAIC	$\chi_M^2 + q(\ln N + 1)$	Population-based, parsimony-adjusted, takes account of sample size
BIC	$\chi_M^2 + q(\ln N)v$	Population-based, parsimony-adjusted, takes account of sample size, greater penalty for complexity
Absolute fit indexes		
GFI	$1 - V_{res}/V_{tot}$	Sample-based, analogous to R^2 , value can fall outside of range 0–1.0
AGFI	$1 - (1 - GFI)[v(v + 1)/2df_M]$	Sample-based, parsimony-adjusted, analogous to corrected R^2 , value can fall outside of range 0–1.0
PGFI	$\{df_M/[v(v + 1)/2]\}GFI$	Sample-based, parsimony-adjusted, analogous to corrected R^2 , value can fall outside of range 0–1.0, sensitive to model size

Note. NC, normed chi-square; q , number of free parameters; v , number of observed variables; V_{res} , unexplained variability in sample covariance matrix; V_{tot} , total variability in sample covariance matrix; NFI, normed fit index; NNFI, non-normed fit index; PNFI, parsimony NFI; RNI, relative noncentrality index; PCFI, parsimony comparative fit index; ECVI, expected cross-validation index; CAIC, consistent Akaike information criterion; BIC, Bayes information criterion; GFI, goodness-of-fit index; AGFI, adjusted GFI; PGFI, parsimony GFI.

Miscellaneous Fit Indexes

nt

based

based, parsimony-adjusted, value can fall outside of range 0–1.0

based, parsimony-adjusted, sensitive to model size

ion-based, value can be negative

ion-based, parsimony-adjusted, sensitive to model size

ion-based, parsimony-adjusted

ion-based, parsimony-adjusted, takes account of sample size

ion-based, parsimony-adjusted, takes account of sample size, greater penalty for complexity

based, analogous to R^2 , value can fall outside of range 0–1.0based, parsimony-adjusted, analogous to R^2 , value can fall outside of rangebased, parsimony-adjusted, analogous to R^2 , value can fall outside of range, sensitive to model size

parameters; v , number of observed variance matrix; V_{tot} , total variability in NNFI, non-normed fit index; PNFI, parsimony comparative fit index; ECVI, Akaike information criterion; BIC, Bayes AGFI, adjusted GFI; PGFI, parsimony

analogous to a squared multiple correlation (R^2) except that the GFI is a kind of matrix proportion of explained variance. Thus, $GFI = 1.0$ indicates perfect model fit, $GFI > .90$ may indicate good fit, and values close to zero indicate very poor fit. However, values of the GFI can fall outside the range 0–1.0. Values greater than 1.0 can be found with just-identified models or with overidentified models with almost perfect fit; negative values are most likely to happen when the sample size is small or when model fit is extremely poor. Another index originally associated with LISREL is the adjusted goodness-of-fit index (AGFI; Jöreskog & Sörbom, 1981). It corrects downward the value of the GFI based on model complexity; that is, there is a greater reduction for more complex models. However, the AGFI has not performed well in some computer simulation studies, which may explain why it is seen less often in the literature nowadays. The parsimony goodness-of-fit index (PGFI; Mulaik et al., 1989) corrects the value of the GFI by a factor that reflects model complexity, but it is sensitive to model size.

6.3 TESTING HIERARCHICAL MODELS

This section concerns ways to test hypotheses about hierarchical path models with the same data, but the general principles apply to all kinds of hierarchical structural equation models. Two path models are hierarchical—also known as nested—if one is a subset of the other. For example, if a path is dropped from model A to form model B, the two models are hierarchically related (i.e., model B is nested under model A).

Model Trimming and Building

There are two contexts in which hierarchical path models are usually compared. In **model trimming**, the researcher typically begins the analysis with a just-identified model and simplifies it by eliminating paths. This is done by specifying that at least one path previously freely estimated is now constrained to equal zero. The starting point for **model building** is usually a bare-bones, overidentified model to which paths are added. Typically, at least one previously fixed-to-zero path is specified as a free parameter. As a model is trimmed, its overall fit to the data

typically becomes worse (e.g., χ^2_M increases). Likewise, model fit generally improves as paths are added (e.g., χ^2_M decreases). However, the goal of both trimming and building is to find a parsimonious model that still fits the data reasonably well.

Models can be trimmed or built according to one of two different standards, theoretical or empirical. The first represents tests of specific, *a priori* hypotheses. Suppose that a path model contains a direct effect of X on Y_2 and an indirect effect through Y_1 . If the researcher believed that the relation of X to Y_2 is entirely mediated by Y_1 , then he or she could test this hypothesis by constraining the coefficient for the path $X \rightarrow Y_2$ to zero (i.e., the path is trimmed). If the overall fit of this constrained model is not appreciably worse than the one with $X \rightarrow Y_2$ as a free parameter, the hypothesis about a mediated relation of X to Y_2 is supported. The main point, though, is that respecification of a model to test hierarchical versions of it is guided by the researcher's hypotheses. This is not the case, however, for empirically based respecification, in which paths are deleted or added according to statistical criteria. For example, if the sole basis for trimming paths is that their coefficients are not statistically significant, then model respecification is guided by purely empirical considerations. The distinction between theoretically or empirically based respecification has implications for the interpretation of the results of model trimming or building, which are considered after a model comparison test statistic is introduced.

The chi-square difference statistic, χ^2_D , can be used to test the statistical significance of the decrement in overall fit as paths are eliminated (trimming) or the improvement in fit as paths are added (building). As its name suggests, χ^2_D is simply the difference between the χ^2_M values of two hierarchical models estimated with the same data; its degrees of freedom, df_D , equal the difference between the two respective values of df_M . The χ^2_D statistic tests the null hypothesis of identical fit of the two hierarchical models in the population. Specifically, smaller values of χ^2_D lead to the failure to reject the equal-fit hypothesis, but larger values lead to the rejection of this hypothesis. In model trimming, rejection of the equal-fit hypothesis suggests that the model has been oversimplified; the same result in model building, however, supports retention of the path that was just added. Ideally, the more complex of the two models compared with χ^2_D should fit the data reasonably well. Otherwise, it makes little

s). Likewise, model fit generally decreases). However, the goal is to find a parsimonious model that

rding to one of two different
 st represents tests of specific,
 model contains a direct effect
 Y_1 . If the researcher believed
 mediated by Y_1 , then he or she
 g the coefficient for the path
 $X \rightarrow Y_1$. If the overall fit of this con-
 tain the one with $X \rightarrow Y_2$ as a
 ediated relation of X to Y_2 is
 t respecification of a model to
 y the researcher's hypotheses.
 ally based respecification, in-
 ing to statistical criteria. For
 paths is that their coefficients
 l respecification is guided by
 inction between theoretically
 applications for the interpreta-
 uilding, which are considered
 introduced.

χ^2_D , can be used to test the overall fit as paths are added. The difference between models estimated with the same sample size tests the null hypothesis of equal proportions of paths in the population. Specifically, failure to reject the equal-fit hypothesis suggests that the same result in model building is obtained by adding a path that was just added. Two models compared with one another.

sense to compare the relative fit of two hierarchical models neither of which adequately explains the data.

Suppose for an overidentified path model that $\chi^2_M(5) = 18.30$. A single path is added to the model (which reduces the degrees of freedom by 1), and the result is $\chi^2_D(4) = 13.20$. Given both results, $\chi^2_D(1) = 18.30 - 13.20 = 5.10$, which has the exact level of statistical significance $p = .024$. Thus, the overall fit of the new model with an additional path is statistically better than that of the original model at the .05 level (but not at the .01 level). In this example, the chi-square difference test is a univariate one because it concerned a single path (i.e., $df_D = 1$). When two hierarchical models that differ by two or more paths are compared (i.e., $df_D \geq 2$), the chi-square difference test is essentially a multivariate test of all the added (or deleted) paths together. If $p < .05$ for χ^2_D in this case, at least one of the paths is non-zero. The paths may be statistically significant at the .05 level if tested individually, but this is not guaranteed.

Empirical versus Theoretical Respecification

The interpretation of χ^2_D as a test statistic depends in part upon whether the new model is derived empirically or theoretically. For example, if individual paths that are not statistically significant are dropped from the model, it is likely that χ^2_D will not be statistically significant. But if the deleted path is also predicted in advance to be zero, χ^2_D is of utmost theoretical interest. Also, if model specification is entirely driven by empirical criteria such as statistical significance, the researcher should worry about capitalization on chance. That is, a path may be statistically significant due only to chance variation, and its inclusion in the model would be akin to a Type I error. Likewise, a path that corresponds to a true nonzero causal effect may not be statistically significant in a particular sample, and its exclusion from the model would be essentially a Type II error. A sort of buffer against the problem of sample-specific results, though, is a greater role for theory in model respecification.

The issue of capitalization on chance is especially relevant when the researcher uses an "automatic modification" option available in some SEM computer programs such as LISREL. Such purely exploratory options drop or add paths according to empirical criteria such as statisti-

cal significance at the .05 level of a modification index, which is calculated for every path that is fixed to zero. A modification index is actually a univariate version of something known as a Lagrange Multiplier (LM), which in this case is expressed as a χ^2 statistic with a single degree of freedom. The value of an LM in the form of a modification index estimates the amount by which the overall model chi-square statistic, χ^2_M , would decrease if a particular fixed-to-zero path were freely estimated. That is, a modification index estimates $\chi^2_D(1)$ for adding that path. Thus, the greater the value of a modification index, the better the predicted improvement in overall fit if that path were added to the model. Likewise, a multivariate LM estimates the effect of allowing a set of constrained-to-zero paths to be freely estimated. For example, an LM with 5 degrees of freedom approximates the value of $\chi^2_D(5)$ for adding the same 5 paths. Some SEM computer programs, such as Amos and EQS, allow the user to generate modification indexes for specific parameters, which lends a more a priori sense to the use of this statistic.

The Wald W statistic is a related index but one used for model trimming. A univariate Wald W statistic approximates the amount by which the overall χ^2_M statistic would increase if a particular freely estimated path were fixed to zero. That is, a univariate Wald W statistic estimates $\chi^2_D(1)$ for dropping the same path. A value of univariate Wald W that is not statistically significant at, say, the .05 level predicts a decrement in overall model fit that is not statistically significant at the same level. Model trimming that is entirely empirically based would thus delete paths with Wald W statistics that are not statistically significant. A multivariate Wald W statistic approximates the value of χ^2_D for trimming a set of two or more paths from the model. Loehlin (1998) gives us some good advice: A researcher should *not* feel compelled to drop every path that is not statistically significant from the model, especially when the sample size is not large or power is low. Removing such paths might also affect the solution in an important way. It might be better to leave them in the model until replication indicates that the corresponding direct effect is of negligible magnitude.

All of the test statistics just described are sensitive to sample size; thus, even a trivial change in overall model fit due to adding or dropping a path could be statistically significant within a large sample. In addition to noting the statistical significance of a modification index, the researcher should also consider the absolute magnitude of the

fication index, which is calculated as a Lagrange Multiplier (LM), a statistic with a single degree of freedom. A modification index estimates the model chi-square statistic, χ^2_M , if all paths were freely estimated, (1) for adding that path. Thus, the higher the modification index, the better the predicted values added to the model. Likewise, following a set of constrained-to-one rule, an LM with 5 degrees of freedom for adding the same 5 paths. AMOS and EQS, allow the user to specify parameters, which lends a statistic.

A modification index but one used for model fit approximates the amount by which a particular freely estimated path deviates from a univariate Wald W statistic. A value of univariate Wald W at the .05 level predicts a deviation statistically significant at the .05 level. An empirically based modification index would be at least as large as those that are not statistically significant. It approximates the value of χ^2_D for the model. Loehlin (1998) suggests that one should not feel compelled to remove a path from the model, even if power is low. Removing a path in an important way. It might be replication indicates that the path is not statistically significant.

Modification indices are sensitive to sample size; they can be misleading due to adding or dropping paths within a large sample. In the case of a modification index, the absolute magnitude of the

change in the coefficient for the path if it is allowed to be freely estimated or constrained to equal zero. If the magnitude of the change is small, the statistical significance of the modification index may reflect the sample size more than it does the magnitude of the direct effect (see Kaplan, 2000, pp. 121–122).

Specification Searches

MacCallum (1986) and Silvia and MacCallum (1988) conducted computer simulation studies of specification searches. They took known structural equation models, imposed different types of specification errors on them (e.g., a nonzero path is omitted), and evaluated the erroneous models using data generated from populations in which the known models were true. In MacCallum's study (1986), models were modified using empirically based methods (e.g., modification indexes). Most of the time the changes suggested by empirically based respecification were incorrect, which means that they typically did not recover the true model. This pattern was even more apparent for small samples (e.g., $N = 100$). Silvia and MacCallum (1988) followed a similar procedure except that the application of automatic modification was guided by theoretical knowledge, which improved the chances of discovering the true model. The implication of these studies is clear: learn from your data, but your data should not be your teacher.

A relatively new research area in SEM concerns the development of automated yet "intelligent" specification searches based on heuristics that attempt to optimize respecification compared with "dumb" specification searches (e.g., automatic model modification). These algorithms are generally based on principles of machine learning or data mining in computer science. For example, Marcoulides and Drezner (2001) describe an adaptive search algorithm based on principles of genetics and natural selection that evaluates models through successive "generations" from parent to child models. Marcoulides and Drezner (2003) describe a search algorithm that mimics the behavior of an ant colony as it collectively tries to achieve a certain goal, in this case model optimization. Intelligent specification search algorithms are not yet widely implemented in SEM computer programs, but this may change. It remains to be seen whether such algorithms are as intelligent as they are claimed to be.

Example of Model Trimming

Model trimming guided by both a priori and empirical considerations is demonstrated for the just-identified path model of illness analyzed earlier. Recall that D. L. Roth et al. (1989) believed that five direct effects in this model were actually zero (see Figure 6.1). To test this hypothesis, the model with these five paths deleted was fitted to the covariance matrix based on the data in Table 6.1 with Amos 5 (Arbuckle, 2003). Amos Basic syntax for this analysis is presented in Appendix 6.B. Values of selected fit indexes are $\chi^2_M(5) = 11.078$, $p = .050$, $NC = 11.078/5 = 2.216$, $CFI = .961$, $SRMR = .051$, and $RMSEA = .057$, with the 90% confidence interval .001–.103. These results present a mixed picture. For example, the null hypothesis of perfect model fit in the population is technically *not* rejected at the .05 level because the probability of the model chi-square (.050) is not below .05, but just barely so. The value of the RMSEA (.057) is just higher than the ideal value of .05, but the upper bound of its 90% confidence interval (.103) is consistent with the hypothesis of poor approximate model fit in the population ($H_0: \epsilon_0 \geq .10$). On the other hand, values of the CFI (.961) and SRMR (.051) are both favorable.

Only one correlation residual had an absolute value greater than .10. This residual, -.13, is between the fitness and stress variables. Thus, the model with five deleted paths, including Fitness → Stress (see Figure 6.1), underpredicts this correlation by .13. The only statistically significant modification index, $\chi^2(1) = 5.096$, $p = .024$, is for the same path. Based on these results, the path Fitness → Stress was added back to the model and the respecified model with only four deleted paths was estimated with the same data. Values of selected fit indexes from this second analysis are $\chi^2_M(4) = 9.921$, $p = .205$, $NC = 1.480$, $CFI = .988$, $SRMR = .034$, and $RMSEA = .036$ with the 90% confidence interval 0–.092. The results of the chi-square difference test are

$$\chi^2_D(1) = \chi^2_M(5) - \chi^2_M(4) = 11.078 - 9.921 = 5.157, p = .023$$

which means that the improvement in overall fit due to adding back the path Fitness → Stress is statistically significant at the .05 level. [Note that values of $\chi^2_D(1)$ and of the modification index $\chi^2(1)$ are

and empirical considerations. A path model of illness analyzed by Bentler and Satorra (1989) believed that five direct paths deleted was fitted to the data in Table 6.1 with Amos 5. The results of this analysis are presented in Table 6.1. These results present a hypothesis of perfect model fit at the .05 level because the χ^2 value (5) = 11.078, $p = .051$, SRMR = .051, and RMSEA = .103–.103. These results present a hypothesis of perfect model fit at the .05 level because the χ^2 value (5) = 11.078, $p = .051$, SRMR = .051, and RMSEA = .103–.103. These results present a hypothesis of perfect model fit at the .05 level because the χ^2 value (5) = 11.078, $p = .051$, SRMR = .051, and RMSEA = .103–.103.

The absolute value greater than .05 for the χ^2 statistic indicates that the model does not fit the data well. The path coefficients for the stress variables are significant, including Fitness → Stress (.13). The only statistically significant path coefficient is the path Fitness → Stress (.13). The only statistically significant path coefficient is the path Fitness → Stress (.13). The only statistically significant path coefficient is the path Fitness → Stress (.13).

$$\chi^2(1) = 5.096, p = .024,$$

The overall fit due to adding back the significant path is significant at the .05 level. The values of selected fit indexes for the respecified model with only one path deleted are $\chi^2(4) = 9.921$, $p = .034$, $R = .034$, and RMSEA = .036–.036. The results of the chi-square test are $\chi^2(1) = 5.096, p = .024$.

The results of the chi-square test are $\chi^2(1) = 5.096, p = .024$. The values of selected fit indexes for the respecified model with only one path deleted are $\chi^2(4) = 9.921, p = .034, R = .034$, and RMSEA = .036–.036. The results of the chi-square test are $\chi^2(1) = 5.096, p = .024$. The values of selected fit indexes for the respecified model with only one path deleted are $\chi^2(4) = 9.921, p = .034, R = .034$, and RMSEA = .036–.036.

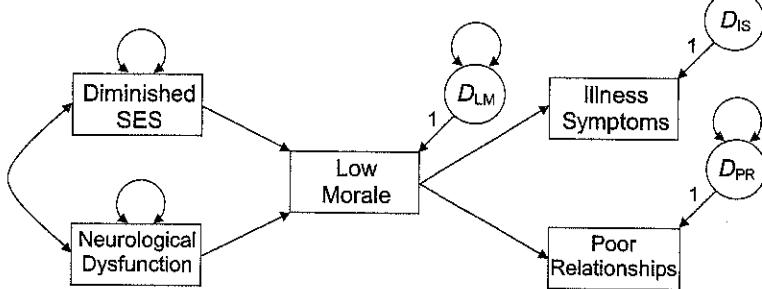
6.4 COMPARING NONHIERARCHICAL MODELS

Sometimes researchers specify alternative models that are not hierarchically related. Although the values of χ^2 from two different nonhierarchical models can still be compared, the difference between them cannot be interpreted as a test statistic. This is where the AIC and other predictive fit indexes come in handy: within a set of competing nonhierarchical models, the one with the lowest value is preferred. An example follows. Presented in Figure 6.2 are two different path models of recovery after cardiac surgery evaluated by Romney, Jenkins, and Bynner (1992). The *psychosomatic model* of Figure 6.2(a) represents the hypothesis that morale mediates the effects of neurological dysfunction and diminished socioeconomic status (SES) on physical symptoms and social relationships. The *conventional medical model* of Figure 6.2(b) depicts different assumptions about causal relations among the same variables.

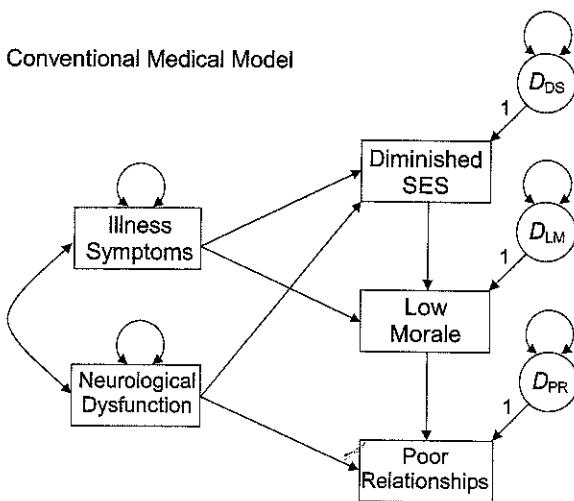
Reported in Table 6.5 are the correlations among the observed variables reported by Romney et al. (1992) for a sample of 469 patients. Unfortunately, Romney et al. did not report variable means or standard deviations, and the analysis of a correlation matrix with standard ML estimation is not generally recommended. To deal with this problem, the correlation matrix in Table 6.5 was analyzed with the SEPATH module of STATISTICA 6 (StatSoft Inc., 2003). This module offers a program option that, if selected, results in the automatic use of the method of constrained estimation (described in Chapter 7), which can correctly analyze a correlation matrix. Each model of Figure 6.2 was fitted to the correlation matrix in Table 6.1 with the method of constrained ML estimation in SEPATH. Both analyses converged to admissible solutions. Values of selected fit indexes for the psychosomatic model (Figure 6.2(a)) are

$$\chi^2(5) = 40.402, p < .001, CFI = .918, SRMR = .065, RMSEA = .120 (.086–.156)$$

(a) Psychosomatic Model



(b) Conventional Medical Model

**FIGURE 6.2.** Alternative nonhierarchical path models of adjustment after cardiac surgery.

and corresponding values for the conventional medical model (Figure 6.2(b)) are

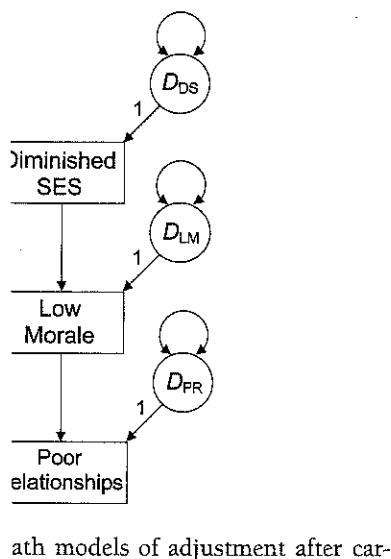
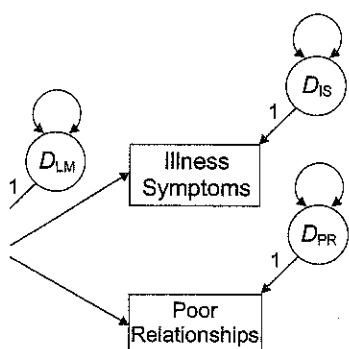
$$\chi^2_M(3) = 3.238, p = .283, \text{CFI} = .999, \\ \text{SRMR} = .016, \text{RMSEA} = .013 (0-.080)$$

It is not surprising that the overall fit of the more complex conventional medical model ($df_M = 3$) is better than that of the more parsimonious psychosomatic model ($df_M = 5$). However, the fit advantage of the

TABLE 6.5. Input Data (Correlations) for Analysis of Nonhierarchical Path Models of Recovery after Cardiac Surgery

Variable	1	2	3	4	5
1. Low Morale	1.00				
2. Illness Symptoms	.53	1.00			
3. Neurological Dysfunction	.15	.18	1.00		
4. Poor Relationships	.52	.29	-.05	1.00	
5. Diminished SES	.30	.34	.23	.09	1.00

Note. These data are from Romney et al. (1992); N = 469.



path models of adjustment after cardiac surgery.

conventional medical model (Figure

33, CFI = .999,
RMSEA = .013 (.0–.080)

of the more complex conventional medical model than that of the more parsimonious psychosomatic model; however, the fit advantage of the

more complex model is enough to offset the penalty for having more parameters imposed by the AIC as defined by Equation 6.4. For the conventional medical model, AIC = 27.238, but for the psychosomatic model, AIC = 60.402. Because the former model has the lowest AIC value, it is preferred over the latter model.

6.5 EQUIVALENT MODELS

After a final path model is selected from among hierarchical or nonhierarchical alternatives, then equivalent models should be considered. Equivalent models yield the same predicted correlations or covariances but with a different configuration of paths among the same observed variables. Equivalent models also have equal goodness-of-fit indexes, including χ^2_M (and df_M) and all other fit statistics described earlier. For a given path model—or any structural equation model—there may be many equivalent variations; thus, it behooves the researcher to explain why his or her final model should be preferred over mathematically identical ones.

Readers already know that just-identified path models perfectly fit the data. By default, any variation of a just-identified path model exactly matches the data, too, and thus is an equivalent model. Equivalent versions of overidentified path models can be generated using the Lee-Herschberger replacing rules (Herschberger, 1994; Lee & Herschberger, 1990), which are summarized below:

1. Within a block of variables at the beginning of a model that is just-identified and with unidirectional relations to subsequent variables, direct effects, correlated disturbances, and equality-constrained reciprocal effects (i.e., the two unstandardized direct effects are specified to be equal) are interchangeable. For example, $Y_1 \rightarrow Y_2$ may be replaced by $Y_2 \rightarrow Y_1$, $D_1 \uparrow D_2$, or $Y_1 \rightleftharpoons Y_2$. If two variables are specified as exogenous, then an unanalyzed association may be substituted, too.
2. At subsequent places in the model where two endogenous variables have the same causes and their relations are unidirectional, all of the following may be substituted for one another: $Y_1 \rightarrow Y_2$, $Y_2 \rightarrow Y_1$, $D_1 \uparrow D_2$, and the equality-constrained reciprocal effect $Y_1 \rightleftharpoons Y_2$.

Note that substituting reciprocal direct effects (\rightleftharpoons) for other types of paths would make the model nonrecursive, but it is assumed that the new model is identified. How to tell whether a nonrecursive structural model is identified is considered in Chapter 9. Also, some equivalent versions of a structural model may be implausible due to the nature of the variables or the time of measurement. For example, a model that contains a direct effect from an acculturation variable to gender would be illogical; also, the assessment of Y_1 before Y_2 is inconsistent with the specification $Y_2 \rightarrow Y_1$. When an equivalent model cannot be disregarded, however, it is up to the researcher to provide a rationale for preferring one over the other.

Relatively simple structural models may have few equivalent versions, but more complicated ones may have hundreds or even thousands (e.g., MacCallum et al., 1993). We learn in Chapter 7 that measurement models can have infinitely many equivalent versions. Thus, it is unrealistic that researchers consider all possible equivalent models. As a compromise, researchers should generate at least a few substantively meaningful equivalent models. Unfortunately, even this limited step is often neglected. MacCallum et al. (1993) reviewed 53 articles published in behavioral science research journals in which SEM results were reported. Although nonhierarchical models were considered in about a dozen works, none of the authors explicitly acknowledged the existence of equivalent models. In their review of about 500 applications of SEM published in 16 different research journals, MacCallum

he beginning of a model that is
ectional relations to subsequent
ted disturbances, and equality-
(i.e., the two unstandardized
be equal) are interchangeable.
placed by $Y_2 \rightarrow Y_1, D_1 \leftrightarrow D_2$, or
specified as exogenous, then an
substituted, too.

del where two endogenous vari-
id their relations are unidirec-
be substituted for one another:
1 the equality-constrained recip-

irect effects (\rightleftarrows) for other types
ursive, but it is assumed that the
hether a nonrecursive structural
apter 9. Also, some equivalent
implausible due to the nature of
ent. For example, a model that
ration variable to gender would
before Y_2 is inconsistent with the
ivalent model cannot be disre-
er to provide a rationale for pre-

s may have few equivalent ver-
y have hundreds or even thou-
Ve learn in Chapter 7 that mea-
any equivalent versions. Thus, it
all possible equivalent models.
generate at least a few substan-
Unfortunately, even this limited
al. (1993) reviewed 53 articles
h journals in which SEM results
cal models were considered in
ors explicitly acknowledged the
ir review of about 500 applica-
t research journals, MacCallum

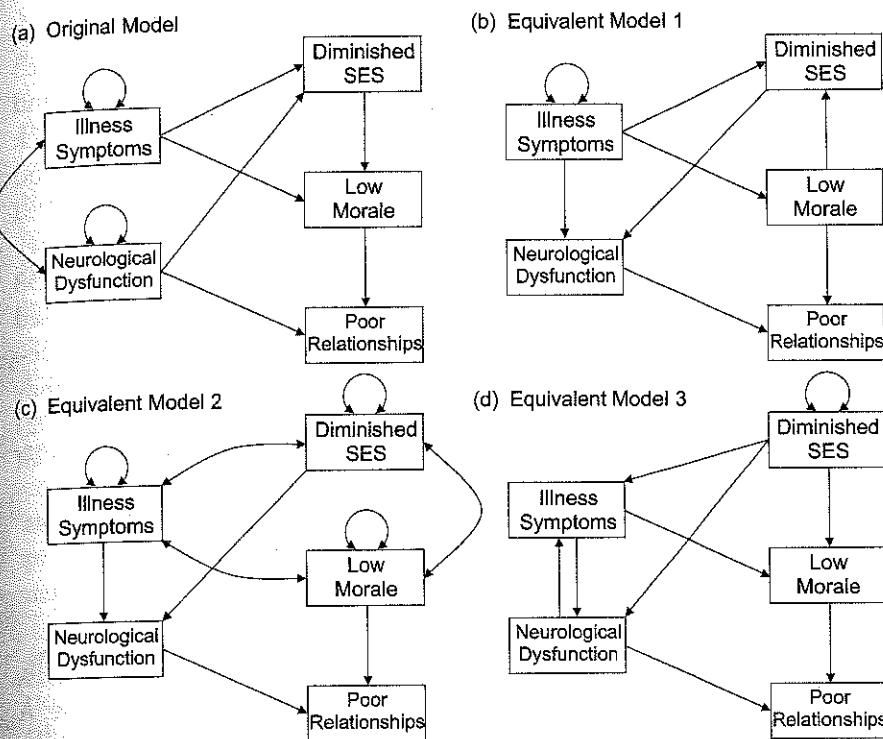


FIGURE 6.3. Four equivalent path models of adjustment after cardiac surgery. Disturbances are omitted.

and Austin (2000) found that researchers are still mainly unaware of the phenomenon of equivalent models or choose to ignore it. This particular kind of confirmation bias is one of the most serious limitations of today's practice of SEM.

Presented in Figure 6.3(a) is Romney and associates' original *conventional medical model* (1992) shown without disturbances to save space. The other three models in Figure 6.3 are generated from the original using the Lee-Herschberger replacing rules and thus are equivalent to it. For example, the equivalent model of Figure 6.3(b) substitutes a direct effect for a covariance between the illness symptoms and neurological dysfunction variables; it also reverses the direct effects between diminished SES and low morale and between diminished SES and neurological dysfunction. The equivalent model of Figure 6.3(c)

replaces two of three direct effects that involve diminished SES with unanalyzed associations. The equivalent model of Figure 6.3(d) replaces the unanalyzed association between illness symptoms and neurological dysfunction with an equality-constrained direct feedback loop. It also reverses the direct effects between illness symptoms and diminished SES and between neurological dysfunction and diminished SES. Because all four path models of Figure 6.3 are equivalent, they explain the data in Table 6.5 equally well. For example, $\chi^2_M(3) = 3.238$ for all four models. Therefore, the choice among them must be based on theoretical rather than statistical grounds. Applying the Lee-Herschberger replacing rules to any of the models in Figure 6.3 may generate even more equivalent models. See Meehl and Walker (2002) and Reichardt (2002) for additional discussions about the testing of alternative directionality specifications for structural models.

6.6 POWER ANALYSIS

Power analysis can determine the probability that the results of a statistical test will lead to rejection of the null hypothesis when it is false. A power analysis in SEM can be conducted at the level of individual paths or for the whole model. One way to estimate the power of the test for a particular unstandardized path coefficient in a recursive path model is to use a method for the technique of multiple regression (J. Cohen, 1988, chap. 9). This method is amenable to hand calculation and estimates the power of tests of individual unstandardized regression coefficients. It requires the specification of the population proportion of unique variance explained by the direct effect of interest. This is the amount by which the proportion of total explained variance in the endogenous variable is expected to increase due to adding that direct effect to the model. (It also equals the squared part correlation.) The researcher also specifies in this method the level of α (usually .05 or .01), calculates the noncentrality parameter for the appropriate F distribution, λ , and then uses special tables in J. Cohen (1988) to estimate power. A variation is to specify a desired level of power (e.g., .80) and then estimate the minimum sample size needed to obtain it.

A method by Saris and Satorra (1993) for estimating power to detect an added path is more specific to SEM, and it can be used with

t involve diminished SES with
lent model of Figure 6.3(d)
between illness symptoms and
ity-constrained direct feedback
between illness symptoms and
cal dysfunction and diminished
Figure 6.3 are equivalent, they
ll. For example, $\chi^2_M(3) = 3.238$
ce among them must be based
grounds. Applying the Lee-
the models in Figure 6.3 may
See Meehl and Walker (2002)
scussions about the testing of
for structural models.

bility that the results of a statis-
ll hypothesis when it is false. A
at the level of individual paths
mate the power of the test for a
nt in a recursive path model is
multiple regression (J. Cohen,
e to hand calculation and esti-
standardized regression coeffi-
the population proportion of
t effect of interest. This is the
otal explained variance in the
ase due to adding that direct
squared part correlation.) The
the level of α (usually .05 or
eter for the appropriate F dis-
in J. Cohen (1988) to estimate
l level of power (e.g., .80) and
needed to obtain it.
993) for estimating power to
SEM, and it can be used with

basically any kind of structural equation model. Suppose that a researcher believes that the population unstandardized direct effect of X on Y is 5.0 (i.e., a 1-point increase in X leads to an increase on Y of 5 points when all other causal variables are held constant). Using this and the other *a priori* values of the parameters of the model of interest, the researcher's next step is to generate a predicted covariance matrix by employing the tracing rule (recursive path models only) or methods based on matrix algebra (e.g., Loehlin, 1998, pp. 43–47). This model-implied covariance matrix is then specified as the input data to an SEM computer program. The model analyzed does *not* include the path $X \rightarrow Y$ (i.e., it is fixed to zero), and the sample size specified is a planned value for the study (e.g., $N = 200$). The χ^2_M statistic from this analysis approximates a noncentral chi-square statistic. Next, the researcher consults a special table for noncentral chi-square for estimating power as a function of degrees of freedom and level of α (e.g., Loehlin, 1998, p. 263). The researcher uses $df = 1$ in these tables to obtain the estimated probability of detecting the added path when testing for it.

An approach to power analysis at the model level by MacCallum et al. (1996) is based on the RMSEA and noncentral chi-square distributions for the three different kinds of null hypotheses: (1) $H_0: \varepsilon_0 \leq .05$ (the model has *close fit* in the population); (2) $H_0: \varepsilon_0 \geq .05$ (there is *not close fit*); and (3) $H_0: \varepsilon_0 = 0$ (there is *exact fit*). The exact fit hypothesis is tested by χ^2_M , but because this hypothesis may be typically implausible, only the close-fit and not-close-fit hypotheses are considered here. Higher power for a test of the close-fit hypothesis means that we are better able to reject an incorrect model. In contrast, higher power for a test of the not-close-fit hypothesis implies better ability to detect a reasonably correct model. A power analysis for either null hypothesis (or both) would be conducted by specifying N , α , df_M , and a suitable value of the parameter estimated by the RMSEA under the alternative hypothesis, ε_1 . For example, ε_1 could be specified for the close-fit null hypothesis as .08, the recommended upper limit for reasonable approximation error, or ε_1 , could be specified for the not-close-fit null hypothesis as .01, which represents the case of a very good model. A variation is to determine the minimum sample size needed to reach a target level of power (e.g., .80) given α , ε_0 , and ε_1 .

Estimated power or minimum sample sizes can be obtained by consulting special tables in MacCallum et al. (1996) or Hancock and

Freeman (2001) or by using a computer program. MacCallum et al. (1996) give syntax for the SAS/STAT program (SAS Institute Inc., 2000) for a power analysis based on the method just described. Another option is to use the Power Analysis module in STATISTICA 6 (StatSoft Inc., 2003), which can estimate power for a structural equation model over ranges of ϵ_1 (with ϵ_0 fixed to .05), α , df_M , and N . The ability to inspect power curves as functions of sample size is very useful for planning a study.

The Power Analysis module of STATISTICA was used to estimate the power to reject the close-fit hypothesis, $H_0: \epsilon_0 \leq .05$, assuming $\epsilon_1 = .08$ and $\alpha = .05$ for the Romney et al. *conventional medical model* (1992), where $df_M = 3$ (see Figure 6.2). The sample size in this analysis was specified as $N = 469$, the size of the original sample. The estimated power calculated by the Power Analysis program is .276. That is, if this model actually does *not* have close fit in the population, then the estimated probability that we can reject this incorrect model is just somewhat greater than 25% for a sample size of 469 cases, given the other assumptions for this analysis. These results reflect a general trend that power at the model level may be low when there are few model degrees of freedom even for reasonably large samples. For models with only one or two degrees of freedom, sample sizes in the thousands may be required in order for model-level power to be greater than .80 (e.g., Loehlin, 1998, p. 264). Sample size requirements for the same level of power drop to some 300–400 cases for models when df_M is about 10. Even smaller samples may be needed for 80% power if $df_M > 20$, but the sample size should not be less than 100 in any event. As noted by Loehlin (1998), the results of power analysis in SEM can be sobering. Specifically, if an analysis has a low probability of rejecting a false model, this fact should temper the researcher's enthusiasm for his or her preferred model.

6.7 OTHER ESTIMATION OPTIONS

Some other estimation procedures for data with multivariate normal distributions available in many SEM computer programs are briefly described here. Two iterative methods include generalized least squares (GLS) and unweighted least squares (ULS). Both are based on

er program. MacCallum et al. gram (SAS Institute Inc., 2000) thod just described. Another le in STATISTICA 6 (StatSoft or a structural equation model α , df_M , and N . The ability to ple size is very useful for plan-

STATISTICA was used to estimate iesis, $H_0: \varepsilon_0 \leq .05$, assuming al. conventional medical model he sample size in this analysis riginal sample. The estimated rogram is .276. That is, if this the population, then the esti-incorrect model is just some- of 469 cases, given the other ts reflect a general trend that n there are few model degrees nples. For models with only zes in the thousands may be to be greater than .80 (e.g., ements for the same level of oodels when df_M is about 10. 0% power if $df_M > 20$, but the) in any event. As noted by ysis in SEM can be sobering. bability of rejecting a false cher's enthusiasm for his or

ta with multivariate normal nputer programs are briefly include generalized least res (ULS). Both are based on

the least squares criterion, but unlike multiple regression (i.e., ordinary least squares) they are full-information methods that estimate all parameters at once. A drawback of the ULS method is that it requires all observed variables to have the same scale; that is, this method is neither scale free nor scale invariant. It is also generally less efficient than ML estimation, which means that standard errors tend to be somewhat larger in ULS estimation. The GLS method is a member of a larger family of estimation methods known as weighted least squares (WLS), and some other WLS methods can be used for severely non-normal data (Chapter 7). One potential advantage of ULS and GLS over ML estimation is that they may require less computation time and presumably less computer memory. However, this potential advantage is not as meaningful nowadays given fast processors and abundant memory available in relatively inexpensive personal computers.

Two partial-information methods that analyze only one equation at a time are described next. Neither method is iterative, so no start values are required. Neither method assumes multivariate normality, and both are available in many software programs for general statistical analyses. The first is standard multiple regression, which was the "classical" estimation method for PA before SEM computer programs such as LISREL were introduced in the early 1970s. Recall that one requirement of multiple regression is that the residuals are uncorrelated with the predictors. This aspect of least squares estimation requires in PA the assumption that the disturbances are uncorrelated with all presumed causes of the corresponding endogenous variable. This assumption explains why standard multiple regression cannot be used to estimate nonrecursive path models. It also cannot be used to analyze models with latent variables that represent hypothetical constructs. How to estimate a recursive path model with multiple regression is described in Appendix 6.C. Although there are ways to estimate the overall fit of overidentified recursive path models to the data with multiple regression (e.g., Kline, 1998, pp. 152-154; Schumacker & Lomax, 1996, pp. 44-45), they require tedious and error-prone hand calculations. It is generally better to use an SEM computer program.

Another partial-information, noniterative method is two-stage least squares (2SLS). Unlike standard multiple regression, 2SLS can be used to estimate nonrecursive path models. There are also forms of 2SLS for models with latent variables, such as measurement models.

For example, LISREL uses a type of 2SLS estimation to calculate start values that are then passed to iterative, full-information methods such as ML estimation. The method of 2SLS is also widely used in other areas, such as econometrics, and its range of potential application for model estimation is much wider than that of standard multiple regression. The 2SLS method is described in Chapter 9.

Because partial-information methods such as 2SLS estimate only one equation at a time, they may be more robust against specification error propagation than full-information methods. However, results of a study by Kaplan and Wenger (1993) indicate that it is difficult to predict exactly how specification error in one part of the model will affect estimates for other parts of the model. This is because error propagation is determined first by the initial model specification and then by how the model is subsequently respecified. Bollen (2001) found that 2SLS estimates were generally less biased by specification error propagation than ML estimates, but only under the restrictive assumptions that specification error elsewhere in the model does not affect the equation being analyzed and that this equation is correctly specified. Thus, it is difficult at present to recommend partial-information over full-information estimation methods such as ML.

The availability of so many different estimation methods can sometimes seem overwhelming for newcomers to SEM. Loehlin (1998, p. 54) cites the following proverb that may describe this experience: a person with one watch always knows what time it is; a person with two never does. Actually, the situation is not so bewildering because ML estimation works just fine for most types of structural equation models if the data have been properly screened and the distributions of the endogenous variables are reasonably multivariate normal. Also, these different estimation procedures tend to yield similar solutions for such "well-behaved" data.

6.8 SUMMARY

The details of conducting a path analysis and testing hypotheses were the subject of this chapter. Researchers often begin the analysis with a just-identified model that is trimmed; that is, the model is simplified by eliminating paths. As models are trimmed, their fit to the data usually

estimation to calculate start LL-information methods such as also widely used in other of potential application for of standard multiple regression Chapter 9.

such as 2SLS estimate only robust against specification methods. However, results of a ate that it is difficult to pre-part of the model will affect is is because error propagation specification and then by 1. Bollen (2001) found that y specification error propa-the restrictive assumptions del does not affect the equa-is correctly specified. Thus, rtial-information over full-LL.

imation methods can some-s to SEM. Loehlin (1998, describe this experience: a time it is; a person with two o bewildering because ML structural equation models ad the distributions of the variate normal. Also, these 1 similar solutions for such

becomes progressively worse. Researchers can also test models in the opposite direction: an initial, overidentified model can be made more complex by adding paths. At each step, the fit of each respecified model can be compared against that of the original model with the chi-square difference test. The goal of both model trimming and building is to a find in a series of hierarchical models one that is parsimonious and yet fits the data reasonably well. Model testing should be guided as much as possible by theoretical rather than empirical considerations (e.g., paths are added solely according to their statistical significance levels). After the researcher selects a final path model, it is critical to generate and evaluate some plausible equivalent models that explain the data just as well but with a different configuration of paths.

Given an understanding of path analysis, one knows essentially half the rationale of SEM—that of structural models. The other half concerns measurement models, in which observed variables are represented as indicators of underlying latent variables, the subject of Chapter 7. Once the reader understands the principles of measurement models, then he or she will have acquired the conceptual framework necessary to understand the whole of SEM for the analysis of covariance structures as variations on these two major themes, structural analysis and measurement.

6.9 RECOMMENDED READINGS

- MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, 114, 185–199.
- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, 7, 422–445.
- Tomarken, A. J., & Waller, N. G. (2003). Potential problems with “well-fitting” models. *Journal of Abnormal Psychology*, 112, 578–598.

Appendix 6.A

Statistical Tests for Indirect Effects in Recursive Path Models

Suppose that a is the unstandardized coefficient for the path $X \rightarrow Y_1$ and that SE_a is its standard error; let b and SE_b , respectively, represent the same things for the path $Y_1 \rightarrow Y_2$. The product ab estimates the unstandardized indirect effect of X on Y_2 through Y_1 . Sobel's (1986) estimated standard error of ab is

$$SE_{ab} = \sqrt{b^2 SE_a^2 + a^2 SE_b^2} \quad (6.6)$$

In a large sample, the ratio ab/SE_{ab} is interpreted as a z test of the unstandardized indirect effect and is called the *Sobel test*. There is also a Web page that automatically calculates the Sobel test after the required information is entered in graphical dialogs.⁴ Results of the Sobel test for the individual unstandardized indirect effects of exercise on illness for the recursive path model of Figure 6.1 based on the results reported in Table 6.2 are summarized in the accompanying table.

Indirect effect	a	SE_a	b	SE_b	ab	SE_{ab}	z
Exer \rightarrow Fit \rightarrow Ill	.217	.026	-.442	.087	-.096	.022	-4.34**
Exer \rightarrow Stress \rightarrow Ill	-.014	.055	.271	.045	-.004	.015	.25
Hardi \rightarrow Fit \rightarrow Ill	.079	.046	-.442	.087	-.035	.021	-1.63
Hardi \rightarrow Stress \rightarrow Ill	-.393	.089	.271	.045	-.107	.030	-3.56**

** $p < .01$.

The author is unaware of a hand-calculable test of the statistical significance of indirect effects through two or more mediators, but a rule of thumb by J. Cohen and P. Cohen (1983) seems reasonable: if all of its component unstandardized path coefficients are statistically significant at the same level of α , then the whole indirect effect can be taken as statistically significant at that level of α , too. All three of the component unstandard-

⁴www.unc.edu/~preacher/sobel/sobel.htm

Effects in Recursive

ized coefficients of the path Exercise → Fitness → Stress → Illness meet this requirement at the .05 level (see Table 6.2), so the whole indirect effect can be considered statistically significant at the same level. Note that the only other indirect effect through two mediators, Hardiness → Fitness → Stress → Illness, fails to meet this standard because the unstandardized path coefficient of its first component direct effect is not statistically significant at the .05 level (Table 6.2). See Shrout and Bolger (2002) for more information about the analysis of mediator effects.

efficient for the path $X \rightarrow Y_1$ and $X \rightarrow Y_2$, respectively, represent the same ab estimates the unstandardized Sobel's (1986) estimated standard

$$\chi^2 SE_b^2 \quad (6.6)$$

interpreted as a z test of the unstandardized Sobel test. There is also a Web test after the required information of the Sobel test for the indirect exercise on illness for the recursive results reported in Table 6.2 are

	SE_b	ab	SE_{ab}	z
1	.087	-.096	.022	-4.34**
2	.045	-.004	.015	.25
3	.087	-.035	.021	-1.63
4	.045	-.107	.030	-3.56**

utable test of the statistical significance of more mediators, but a rule of thumb is reasonable: if all of its components are statistically significant at the .05 level, the indirect effect can be taken as statistically significant. Note that the effect of the component unstandardized

Appendix 6.B

Amos Basic Syntax

```

Sub Main()
    Rem Figure 6.1
    Dim SEM As New AmosEngine
    SEM.TextOutput
    SEM.Standardized
    SEM.Smc
    SEM.SampleMoments
    SEM.BeginGroup "Roth.csv"
        SEM.Structure "Fitness = Exercise + (p1) Hardiness" _
            & "+ D_Fi (1)"
        SEM.Structure "Stress = (p2) Exercise + Hardiness" _
            & "+ (p3) Fitness + D_St (1)"
        SEM.Structure "Illness = (p4) Exercise + (p5) " _
            Hardiness" & "+ Fitness + Stress + D_Il (1)"
        SEM.Model "Trim 5", "p1 = p2 = p3 = p4 = p5 = 0"
        SEM.Model "Trim 4", "p1 = p2 = p4 = p5 = 0"
    End Sub

```

Appendix 6.C

Estimation of Recursive Path Models with Multiple Regression

Follow these steps:

1. *Variances and covariances (i.e., unanalyzed associations) of observed exogenous variables.* These are simply the observed values. For example, the unstandardized estimate of an unanalyzed association is the covariance, and the standardized estimate is the Pearson correlation.
2. *Direct effects on endogenous variables from other observed variables.* For each endogenous variable, enter all observed variables specified to directly affect it as predictors in a regression equation. The unstandardized regression coefficients are the unstandardized path coefficients, and the beta weights are the standardized path coefficients.
3. *Disturbance variances.* Record R^2 from each analysis in the previous step. The term $1 - R^2$ is the proportion of unexplained variance. The product $(1 - R^2)s_y^2$, where s_y^2 is the observed variance of the corresponding endogenous variable, equals the unstandardized disturbance variance. The term $(1 - R^2)^{1/2}$ is the standardized residual path coefficient.
4. *Disturbance covariances (i.e., unanalyzed associations).* This step concerns only partially recursive models with disturbance correlations restricted to pairs of endogenous variables without direct effects between them (see Figure 5.1). The unstandardized estimate is the partial covariance between the endogenous variables when we control for their common causes, and the standardized estimate is the corresponding partial correlation (e.g., Kenny, 1979, pp. 52–61).