Introduction

It is both a pleasure and honor to introduce the fifth edition of this book. Like the previous editions, structural equation modeling (SEM) is presented in an accessible way for readers without strong quantitative backgrounds. Included in this edition are many new examples of SEM applications in disciplines that include health, political science, international studies, cognitive neuroscience, developmental psychology, sport and exercise, and psychology, among others. Some examples were selected due to technical problems in the analysis, but such examples provide a context for discussing how to deal with challenges that can and do occur in SEM, especially in samples that are not large. So not all applications of SEM described in this book are trouble free, but neither are actual research problems.

WHAT'S NEW

The many changes in this edition are intended to enhance the pedagogical presentation and cover recent developments. The biggest changes are summarized next:

1. The fourth edition of this book was one of the first introductory works to incorporate Judea Pearl's nonparametric approach to SEM, also called the structural causal model (SCM), into the larger SEM family that dates to the development of path analysis by Sewall Wright in the 1920s–1930s and to the publication of LISREL III in 1976 as the first widely available computer program for covariance structure analysis, also called covariance-based SEM. In the same tradition, this fifth edition includes composite SEM, also referred to as partial least squares path modeling or variance-based SEM, as the third full member of the SEM family. Composite SEM has developed from a set of methods seen in the 1980s–1990s as more suitable for exploratory research that emphasized prediction over explanation to a suite of full-fledged modeling techniques for exploratory or confirmatory analyses, including theory testing. Both the SCM and composite SEM offer unique perspectives on causal modeling that can benefit researchers more familiar with traditional, covariance-based SEM. This means that researchers acquainted with all three members of the SEM family can test a wider range of hypotheses about measurement and causation. I try to make good on this promise throughout the fifth edition.

2. Traditional SEM and composite SEM are described within Edward Rigdon's concept proxy framework that links data with theoretical concepts through proxies, which approximate concepts based on

correspondence rules—also called auxiliary theory—about presumed causal directionality between concepts and data. This point refers to the distinction between reflective measurement, where proxies for latent variables are common factors, and formative measurement, where proxies for emergent variables are composites of observed variables. The choice between the two measurement models just mentioned should be based on theory, not by default due to the researcher's lack of awareness about SEM techniques for analyzing composites.

3. There are additional new chapters on SEM analyses in small samples and recent developments in mediation analysis. Surveyed works about mediation analysis concern research designs and definitions of mediated effects, including natural direct and indirect effects and interventional direct and indirect effects estimated in clinical trials, among other topics. There is also coverage of new reporting standards for SEM studies by the American Psychological Association (APA) and the technique of piecewise SEM, which is based on concepts from Pearl's SCM. There are also extended tutorials on modern techniques for dealing with missing data, including multiple imputation and full information maximum likelihood (FIML), and also about instrumental variable methods as a way to deal with the confounding of target causal effects.

4. The topics of specification and identification versus analysis were described in separate chapters in the fourth edition. They are now combined into individual chapters for each technique described in the fifth edition. I believe this more closely integrated presentation helps readers to more quickly and easily develop a sense of mastery for a particular kind of SEM technique.

5. There is greater emphasis on freely available software for SEM analyses in this new edition. For example, the R package `lavaan` package was used in most analyses described in this book. It is a full-featured computer program for both basic and advanced SEM analyses. It has the capability to analyze both common factors and composites as proxies for theoretical concepts. The syntax in `lavaan` is both straightforward and used in some other R packages, including cSEM for composite SEM, to specify structural equation models, so it has application beyond `lavaan`.

Other R packages used for detailed examples in the fifth edition include `semTools`, piecewiseSEM, MBESS, MIIVsem, psych, WebPower, systemfit, `sem`, `bmem`, CauseAndCorrelation, dagitty, and ggm. Together with the `lavaan` package, a wide variety of analyses for nonparametric, parametric, and composite models in SEM is demonstrated, all with no-cost software. Commercial software for SEM is still described, including Mplus, which can feature state-of-the-art analyses before they appear in other computer tools, but free SEM software is now nearly as capable as commercial products. Also, I would guess that free software could be used in the large majority of published SEM studies.

6. Extended presentations on regression fundamentals, significance testing, and measurement and psychometrics beloved by readers of the fourth edition are freely available in updated form as primers on the book's website. This change was necessary to include the new material in the fifth edition. The topics just mentioned are still covered in the new edition but in a more concise way. New to the fifth edition in the main text is a self-test of knowledge about background concepts in statistics and measurement. There is a scoring key, too, so readers can check their understanding of fundamentals. Readers with higher scores could directly proceed to substantive chapters on SEM analyses, and readers with lower scores can consult any of the primers on the website for more information and exercises.

BOOK WEBSITE

The address for the book's website is *https://www.guilford.com/kline-materials*. From the site, you can freely access the computer files—data, syntax, and output files—for all detailed examples in this book. The website promotes a learning-by-doing approach. The availability of both syntax and data files means that readers can reproduce the analyses in this book by using the corresponding R packages. Even without doing so, readers can still open the output file on their own computers for a particular analysis and view the results. This is because all computer files are simple text files that can be opened with any basic text editor, such as Notepad (Windows), Emacs (Linux/UNIX), or TextEdit (macOS), among others. Syntax files are annotated with extensive comments. Even if readers use a different computer tool, such as LISREL, it is still worthwhile to review the files on the website generated in the R environment. This is because it can be helpful to view the same analysis from somewhat different perspectives. Some of the

exercises for this book involve extensions of the analyses for these examples, so there are plenty of opportunities for practice with real data sets.

PEDAGOGICAL APPROACH

Something that has not changed in the fifth edition is pedagogical style: I still speak to readers (through my author's voice) as one researcher to another, not as statistician to the quantitatively naïve. For example, the instructional language of statisticians is matrix algebra, which conveys a lot of information in a short amount of space, but readers must already be versed in linear algebra to understand the message. There are other, more advanced works about SEM that emphasize matrix presentations (Bollen, 1989; Kaplan, 2009; Mulaik, 2009b), and these works can be consulted when you are ready. Instead, fundamental concepts about SEM are presented here in the language of applied researchers: words, tabular summaries, and data graphics, not matrix equations. I will not shelter you from some of the more technical aspects of SEM, but I aim to cover fundamental concepts in accessible ways that promote continued learning.

PRINCIPLES > SOFTWARE

You may be relieved to know that you are not at a disadvantage at present if you have no experience using an SEM computer tool. This is because the coverage of topics in this book is not based on the symbolism, syntax, or user interface associated with a particular software package. In contrast, there are many books linked to specific SEM computer programs. They can be invaluable for users of a particular program, but perhaps less so for others. Instead, key principles of SEM that users of *any* computer tool must understand are emphasized here. In this way, this book is more like a guide to writing style than a handbook about how to use a particular word processor. Besides, becoming proficient with a particular software package is just a matter of practice. But without strong conceptual knowledge, the output from a computer tool for statistical analyses—including SEM—may be meaningless or, even worse, misleading.

SYMBOLS AND NOTATION

Advanced works on SEM often rely on the symbols and notation associated with the original matrix-based syntax for LISREL, which features a profusion of doubly subscripted lowercase Greek letters for individual model parameters, uppercase Greek letters for matrices of parameters for the whole model, and two-letter acronyms in syntax for matrices. For example, the symbols

$$\lambda_{12}^{(x)},\ \mathbf{\Lambda}_x,\ \text{and LX}$$

refer in LISREL notation to, respectively, a specific loading on an exogenous (explanatory) factor, the parameter matrix of loadings for all such factors, and LISREL syntax that designates the matrix (Lambda-X). Although I use here and there some symbols from LISREL notation, I do not oblige readers to memorize LISREL notation to get something out of the book. This is appropriate because LISREL symbolism can be confusing unless one has learned the whole system by rote.

ENJOY THE RIDE

Learning a new set of statistical techniques is not everyone's idea of fun. (If doing so is fun for you, that's okay, I understand and agree.) But I hope the combination of accessible language that respects your intelligence, examples of SEM analyses in various disciplines, free access to background tutorials (i.e., the primers) and computer files for detailed examples, and the occasional bit of sage advice offered in this book will help to make the experience a little easier, perhaps even enjoyable. It might also help to think of this book as a kind of travel guide about language and customs, what to know and pitfalls to avoid, and what lies just over the horizon in SEM land.

PLAN OF THE BOOK

Part I introduces fundamental concepts, reporting standards, preparation of the data, and computer tools. Chapter 1 lays out both the promise of SEM and widespread problems in its application. Concepts in regression, significance testing, and psychometrics that are especially relevant for SEM are reviewed in Chapter 2, which also include the self-test in these areas. Basic

steps in SEM and reporting standards are introduced in Chapter 3 along with an example from a recent empirical study. How to prepare the data for analysis in SEM and options for dealing with common problems, including missing data, are covered in Chapter 4, and computer tools for SEM, both commercial and free, are described in Chapter 5.

Part II deals with the fundamentals of hypothesis testing in SEM for classical path models, which in the analysis phase feature a single observed measure for each theoretical variable, also called single-indicator measurement. It begins in Chapter 6, which introduces nonparametric SEM as described by Judea Pearl (i.e., the SCM). The SCM is graphical in nature; specifically, causal hypotheses are represented as directed graphs where theoretical variables are depicted with no commitment to any distributional assumptions or specific operational definitions for any variable. Graphs in nonparametric SEM can be analyzed by special computer tools *without* data. This capability allows researchers to test their *ideas* before collecting the data. For example, the analysis of a directed graph may indicate that a particular causal effect cannot be estimated unless additional variables are measured. After the data are collected, it is a parametric model that is typically analyzed, and such models and their assumptions are described in Chapter 7. The technique of piecewise SEM, which connects the two perspectives, nonparametric and parametric, through novel techniques for analyzing path models, is covered in Chapter 8.

Chapters 9–12 are perhaps the most important ones in the book. They concern how to test hypotheses and evaluate models in complete and transparent ways that respect both reporting standards for SEM and best practices. These presentations are intended as counterexamples to widespread dubious practices that plague many, if not most, published SEM studies. That is, the state of SEM practice is generally poor, and one of my goals is to help readers distinguish their work above this din of mediocrity. Accordingly, Chapter 9 outlines methods for simultaneous estimation of parameters in structural equations models and explains how to analyze means along with covariances. Chapter 10 deals with the critical issue of how to properly assess model fit after estimates of its parameters are in hand. A critical point is that model fit should be routinely adjudged from at least two perspectives: global or overall fit, and local fit at the level of residuals, which in SEM concerns differences between sample and predicted associations for each pair of measured variables. Chapters 11–12 extend these ideas to, respectively, the comparison of alternative models all fit to the same data and the simultaneous analysis of a model over data from multiple groups, also called multiple-group SEM.

Part III deals with the analysis of models where at least some theoretical concepts are approximated with multiple observed variables, or multiple-indicator measurement. Such models are often referred to as "latent variable models," but for reasons explained in Chapter 13, our models include only proxies for latent variables, not latent variables themselves. These proxies are of two general types: common factors based on reflective measurement models and composites based on formative measurement models. The analysis of pure reflective measurement models in the technique of confirmatory factor analysis (CFA) is described in Chapter 14, and Chapter 15 deals with the analysis of structural regression (SR) models—also called latent variable path models—where causal effects between observed variables or common factors are estimated. Chapter 16 is about composite SEM, which analyzes causal models with multiple-indicator measurement based on formative, not reflective, measurement and where proxies for conceptual variables are composites, not common factors. Application of the technique of confirmatory composite analysis (CCA), the composite analog to CFA, is demonstrated.

Part IV is about advanced techniques. How to deal with SEM analyses in small samples is addressed in Chapter 17, and Chapter 18 concerns the analysis of categorical data in CFA. Chapter 19 explains how to analyze nonrecursive models with causal loops that involve two or more endogenous (outcome) variables assumed to influence each other, and Chapter 20 surveys recent developments that enhance, improve, and extend ways to assess hypotheses of causal mediation, or indirect causal effects that involve at least one intervening variable. The state of mediation analysis in the literature is problematic, but some of the newer approaches and methods described in this chapter seem promising. The analysis of latent growth models for longitudinal data is the subject of Chapter 21, and the application of multiple-group CFA to test hypotheses of measurement invariance is dealt with in Chapter 22. The capstone of the book is the summary of best practices in SEM in Chapter 23. Also mentioned in this chapter are common mistakes with the aim of helping you to avoid them.

8

Local Estimation and Piecewise SEM

There are two broad families of estimation methods in SEM, local and global. In **local estimation**—also called **limited-information methods**, **partial-information methods,** or **single-equation methods**—equations for endogenous variables are analyzed one at a time. Conceptually, (1) the full model is decomposed into a series of submodels, one for each outcome, and (2) presumed causal effects for each outcome are estimated in separate regression analyses. In **global estimation**, the whole model is analyzed all at once; that is, equations for all outcomes and their presumed causes are simultaneously estimated. Until recently, (1) local estimation in SEM was mainly restricted to manifest-variable path analysis models, and (2) global estimation was the sole practical choice for models with common factors as proxies for latent variables. But the availability of relatively new computer tools for local estimation of path models has expanded analysis options for both types of models just mentioned. This chapter covers local estimation for manifest-variable path models and the related method of piecewise SEM. Options for local estimation or global estimation of models with common factors are covered in later chapters.

RATIONALE OF LOCAL ESTIMATION

In local estimation, the researcher conducts a series of regression analyses, one for each outcome variable in the model. A suitable regression technique should be used for a particular outcome. This means that

1. The **link function**, which associates a linear combination of predictors to a parameter for the distribution of the outcome, is appropriate, given the data type of the outcome (e.g., count, ordinal, or binary data).
2. Distributional assumptions of the technique, if any, should be plausible.
3. Functional forms of statistical associations between predictors and outcomes are properly specified in parametric methods, or nonparametric methods are used that do not assume a particular functional form (Appendix 7.1).

For example, ordinary least squares (OLS), or standard multiple regression (identity link), is for continuous outcomes with linear relations to all predictors. Curvilinear relations can also be estimated if the computer is instructed to include in the equation polynomial terms, such as X^2 for quadratic effects of X, along with variable X itself (Appendix 7.1). Normal distributions for observed scores are not required in the OLS method, but distributions of residuals for cases should be normal. Dichotomous outcomes could be analyzed in logistic regression (logit link) or probit regression (probit link), among other options for binary regression, and outcomes that are count variables could be analyzed in Poisson regression (log link), and so on. The point is that there should a be a good match between the distributional assumptions and types of functional relations estimated in a particular regression technique and the outcome analyzed with its presumed causes.

Potential advantages of local estimation are listed next (Bollen, 2019; Lefcheck, 2016; Shipley, 2000):

1. No specialized software is needed. This is because local estimation can be conducted using software for general statistical analyses, such as SPSS, SAS/STAT, or native regression procedures in R, such as function "`lm( )`" (linear regression models) for continuous outcomes.

2. Local estimation accommodates a wide range of variable types and distributions, and it also allows the use of parametric or nonparametric regression methods. Thus, no universal requirements or assumptions, such as multivariate normality, necessarily apply over the analyses for all outcome variables.

3. Local estimation may be less susceptible to **propagation of specification error** compared with global estimation. Because each outcome is separately analyzed in local estimation, effects of specification error for one outcome may not spread to different outcomes with correctly specified equations.

4. Global estimation methods are typically asymptotic; that is, they require large samples for precise estimation. They also require statistically identified models; otherwise, estimation may fail. In contrast, local estimation is generally less demanding about sample size, and it may be possible in local estimation to generate estimates for individual outcomes even though the whole model is not identified.

5. The availability of significance tests of overall model fit was once the near-exclusive domain of global estimation, but there are now computer tools that conduct global fit testing in the context of local estimation, too. For observed-variable path models, these global fit tests are generally based on the concept of d-separation and simultaneously test all model-implied conditional independencies for a union basis set (Chapter 6). Such tests can be conducted without estimating a single model parameter; that is, d-separation tests can be conducted prior to local estimation. Doing so is part of the rationale for piecewise SEM, which is described next.

PIECEWISE SEM

Shipley (2000) described the basic logic of **piecewise SEM**, also called **confirmatory path analysis**, for recursive path models with no causal loops and no correlated errors. The method was later expanded to include path models with correlated errors, multilevel analysis of path models, comparison of alternative models fitted to the same data, evaluation of path models over multiple samples, and models with proxies for latent variables (Shipley, 2003, 2009; Shipley & Douma, 2020, 2021). Lefcheck (2016) described the freely available piecewiseSEM package for the R computing environment.

The basic steps of piecewise SEM for recursive path models with continuous outcomes with linear relations (Shipley, 2000) are summarized next:

1. The path model is expressed as a directed acyclic graph (DAG).
2. The union basis set of implied conditional independencies is derived. Recall that the union basis set controls for all parents of each nonadjacent pair of variables, or those not directly connected by a path in the graph. It consists of the smallest number of nonoverlapping conditional independence claims that generate all such hypotheses encoded by the DAG.
3. In the data, calculate the value of the Pearson correlation or partial correlation that corresponds to each implied conditional independency in the union basis set. Each of these coefficients is also a correlation residual, or the difference between the observed (sample) correlation and the predicted value, which is zero (i.e., conditional independence). Correlation residuals are measures of local fit because they involve a single pair of variables, not all variables in the model considered at once.
4. Next, for each observed correlation test the null hypothesis that the corresponding parameter equals zero against a nondirectional alternative hypothesis. For example, if $r_{XY \bullet W}$ is the sample coefficient for the implied conditional independence $X \perp Y \mid W$, then the null and alternative hypotheses are, respectively,

$$H_0: \rho_{XY \bullet W} = 0 \quad \text{and} \quad H_1: \rho_{XY \bullet W} \neq 0$$

   Depending on the computer tool, the test statistic could be t $(N - 2 - c)$, where c is the number of variables for which we are controlling ($c = 1$ in this example), or it could be the normal deviate z based on the Fisher transformation for correlation coefficients.
5. Conduct the **d-separation (d-sep) test**, which is a

multivariate significance test of all implied conditional independencies in the union basis set. The test statistic is *C* (Fisher, 1954), and its formula is

$$C = -2\sum_{i=1}^{k} \ln(p_i) \quad (8.1)$$

where ln is the natural log transformation to base *e* (approximately 2.7182), and p_i is the *p* value from each of the individual significance tests described in Step 4. The *C* statistic is distributed over random samples as central chi-square with $df = 2k$, where *k* is the number of independence claims in the union basis set. The null hypothesis tested by *C* is

$$H_0\text{: } \mathbf{p}_{k\times 1} = \mathbf{0}_{k\times 1}$$

where $\mathbf{p}_{k\times 1}$ is the population vector of *p* values from the tests of all implied conditional independences and $\mathbf{0}_{k\times 1}$ is the zero vector of the same dimension where all elements equal zero.

6. If the model *fails* the d-sep test (e.g., *C* is statistically significant at $p < .05$), then the researcher may decide to respecify it. Model respecification is considered in more detail in a later chapter, but for now we will treat a failed d-sep test as indicating a *potential* problem with the original model. The d-sep test should also be conducted for any respecified model.
7. If the original model or any respecified version is eventually retained, the last step is to locally estimate the equation for each outcome. Path coefficients for presumed causal effects are generally identified through the specification of adjustment (conditioning) sets of covariates in the OLS method or through the specification of instruments in instrumental variables regression, such as the two-stage least squares (2SLS) method.

Two elaborations are needed. First, the question of what is the minimally acceptable absolute correlation residual before concluding that an independence claim is deficient is not clearly specified in the works on piecewise SEM cited to this point. Statistical significance as the sole decision criterion (e.g., reject the model if $p < .05$ for the *C* statistic) is problematic because it ignores effect size and power. For instance, in a large sample, the test for a sample partial correlation that is close to zero, such as $r = .002$, could be statistically significant, but this degree of departure from zero may be seen as negligible.[1] A d-sep test based on partial correlations that all differ trivially from zero could likewise be significant in a large sample. But in a small sample due to low power, the d-sep test may fail to be significant even though some sample partial correlations are much larger, such as $r = .20$, which indicates a 100-fold greater departure from zero than $r = .002$.

Suggested next is a rule of thumb for interpreting correlation residuals based more on an effect size perspective than on outcomes of significance testing: Absolute discrepancies between observed and predicted correlations that exceed .10 may signal appreciable model–data disagreement. This standard has been suggested for exploratory factor analysis (Pett et al., 2003; Tabachnick & Fidell, 2013) and, in my judgment, it seems like a reasonable guideline when continuous variables are analyzed in SEM, too. Although it is difficult to say how many absolute correlation residuals > .10 is "too many," the more there are, the worse the correspondence between model and data concerning implied conditional independencies. An example of considering effect size when conducting the d-sep test follows.

Suppose in a large sample that $p = .001$ for *C* statistic, so the model "fails" the d-sep test at a conventional level of statistical significance. The researcher inspects the absolute values for the whole set of partial correlations and finds that none exceeds .01. If these degrees of departure from zero are all considered unimportant, then the researcher could decide to ignore the failed d-sep test for the model. That is, the model is not rejected, given the relatively low magnitudes of partial correlations even though the global significance test was failed. Now suppose in a small sample that a model "passes" the d-sep test (i.e., *C* is not significant) even though some, and perhaps most, absolute partial correlations exceed .10. Low power of the d-sep test could explain this pattern of results. Accordingly, the researcher could decide in this case to ignore the passed d-sep test (i.e., the model is not retained), given the magnitudes of the correlation residuals.

[1] An alternative is to test correlation residuals for significance against nonzero values, such as .05 in absolute value or some other reasonably small value—see Thoemmes and Rosseel (2018) for more information and examples of R code that implement this type of test.

A second issue in piecewise SEM concerns local estimation. The version of piecewiseSEM that I used in the upcoming detailed example permitted the specification of a single regression equation for each outcome, generally the one that includes all parents for a specific outcome (Lefcheck, 2020). There may be other equations for the same outcome that feature different covariates or the inclusion of instruments for particular causal variables, but multiple equations for the same outcome cannot be analyzed in a single execution (run) of the function "`psem( )`," which is used in the piecewiseSEM package to specify the equations and fit them to the data. But it is not problematic to specify and analyze additional equations for the same outcome in regression analyses conducted outside of the piecewiseSEM package.

DETAILED EXAMPLE

Let's recap the ongoing example to this point: In Chapter 6, we specified as a directed acyclic graph (DAG) the nonparametric version of the Roth et al. (1989) recursive path model of illness in Figure 6.7. We analyzed the graph with a computer tool (analysis 1, Table 6.3) that generated the union basis set, or the smallest number of conditional independencies (5 in total) located by the d-separation criterion that (1) are mutually independent; (2) imply all other conditional independencies; and (3) include the parents of both variables in the conditioning (adjustment) set (Rules 6.1–6.2). The union basis set for the Roth et al. (1989) model is listed in Table 6.4. For example, the graph predicts that the fitness and stress outcomes are independent after controlling for both of their parents, exercise and hardy.

For the same DAG in Chapter 6, we applied graphical identification criteria (analysis 2, Table 6.3) to generate (1) minimally sufficient adjustment sets of covariates to estimate causal effects in ordinary least squares (OLS) regression (Rules 6.3–6.4) or in two-stage least squares (2SLS) regression with instruments or partial instruments (Rules 6.5–6.6). The results of these analyses—see Table 6.5—provide a "roadmap" or analysis plan for local estimation in this chapter.

In Chapter 7, we specified the parametric version of the Roth et al. (1989) path model depicted using full McArdle–McDonald RAM graphical symbolism in Figure 7.5. We determined that the model is identified (Rules 7.3–7.4) and that the degrees of freedom are $df_M = 5$ (Rules 7.1–7.2), which exactly equals the size of the union basis set for this model (5). For continuous variables in linear recursive models, each element of the union basis set corresponds to a vanishing partial correlation that can be compared with a sample partial correlation coefficient. If the model is correctly specified, then the two values—predicted (zero) and observed—should be similar with the bounds of sampling error or effect size (i.e., any discrepancy is considered trivial); otherwise, the model should not be retained.

Listed in Table 8.1 are the analyses, annotated script files, and R packages used in the piecewise SEM analysis of the Roth et al. (1989) parametric path model of illness in Figure 7.5. All files can be freely downloaded from this book's website. The version of the piecewiseSEM package used in this example (Lefcheck, 2020) could not read summary data (i.e., raw data input is required). So, I generated in analysis 1 a raw data file in comma separated values (.csv) format based on the summary statistics in Table 4.3 for the Roth et al. (1989) data set in a sample of $N = 373$. Specifically, I used the "`kd( )`" function in `semTools` (Jorgensen et al., 2022) for the **Kaiser-Dickman algorithm** (Kaiser & Dickman, 1962) to create raw scores for 373 cases where variable descriptive statistics (covariances, means) exactly match those in Table 4.3 for the actual data. These generated raw scores were specified as the input data for the analyses 2–4 in Table 8.1.[2]

Partial Correlations and the d-Separation Test

For analysis 2 in Table 8.1, I used the psych package (Revelle, 2022) and the piecewiseSEM package to calculate the sample partial correlation and p value for each of the five conditional independence claims in the union basis set and also to conduct the multivariate d-sep test for the whole model. Recall that sample (observed) partial correlations in these analyses are also correlation residuals because they all correspond to predicted values that are zero. The results of analysis 2 are summarized in Table 8.2. There is one absolute correlation that is just .10 or more. This result, –.103 (shown in boldface in the table), is for the pair fitness and stress. The model implies that fitness and stress are independent, given exercise and hardy, but

[2]Note that "`kd( )`" generates a different set of raw scores each time it is run, but score descriptive statistics always exactly match those of target covariances and means. Thus, all analysis results described in this chapter are identical in any raw data so generated for this example.

TABLE 8.1. Analyses, Script Files, and Packages in R for Piecewise SEM Analyses of a Recursive Path Model of Illness

Analysis	Script files	R packages
1. Generate unstandardized scores that exactly match sample covariances, means	`roth-generate-scores.r`	`semTools` `lavaan`
2. Estimate and test implied conditional independencies	`roth-d-sep-test.r`	piecewiseSEM psych
3. Local estimation of causal effects		
a. Covariate adjustment (OLS)	`roth-effects-ols.r`	piecewiseSEM
b. Instruments (2SLS)	`roth-effects-2sls.r`	systemfit
4. Bootstrapped standard errors and confidence intervals for indirect effects	`roth-bootstrap-ci.r`	`bmem` `sem`

Note. The external raw data file created in analysis 1 is `roth.csv`. Output files have the same names except the extension is "`.out`." Packages `semTools` and `lavaan` are also used in analyses 2–4. OLS, ordinary least squares; 2SLS, two-stage least squares.

their observed residual differs by what I would consider to be worrisome. In Figure 7.5, there is a single back-door path between fitness and stress:

$$\text{Fitness} \leftarrow \text{Exercise} \curvearrowleft \text{Hardy} \rightarrow \text{Stress}$$

A possible specification error is that fitness and stress are related through paths omitted from the original model. For instance, perhaps fitness affects stress (Fitness → Stress), or vice versa (Stress → Fitness). We will deal with respecification later, but we have already detected a local fit problem.

The value of the *C* statistic for the d-sep test calculated for these data in the piecewiseSEM package is 19.521 (see the output file for analysis 2, Table 8.1). With a total of 5 conditional independence claims in the union basis set, the degrees of freedom are 5 × 2, or 10. For χ^2 (10) = 19.521, p = .034, so the model *fails* the d-sep test at the .05 level. Thus, there is covariance evidence against the model from the perspective of significance testing. The sample size here is not large (N = 373), one absolute correlation exceeds .10 (for fitness and stress), and other absolute correlations are nearly as large (e.g., .089 for hardy and fitness)—see Table 8.2—so local fit problems are indicated from an effect size perspective, too. Exercise 1 asks you to calculate *C* for this analysis, given the *p* values for the partial correlations in Table 8.2.

Given the results to this point, I would *reject* the model as inconsistent with the data and thus begin the

TABLE 8.2. Sample Partial Correlations and *p* Values for a Union Basis Set of Implied Conditional Independencies for a Recursive Path Model of Illness

Conditional independence	Adjustment set	Partial correlation	*p*
Exercise ⊥ Stress	Hardy	–.058	.260
Exercise ⊥ Illness	Fitness, Stress	.039	.455
Hardy ⊥ Fitness	Exercise	.089	.087
Hardy ⊥ Illness	Fitness, Stress	–.081	.118
Fitness ⊥ Stress	Exercise, Hardy	**–.103**	.048

Note. The *p* values are for two-tailed tests that the population correlation is zero.

respecification phase. But in this pedagogical example, we continue next to local estimation using two different methods, covariate selection with estimation in OLS regression and estimation with instruments in 2SLS regression. Doing so gives us the opportunity to appreciate that multiple estimators for the same causal effect may be available in local estimation of path models.

Estimates of Direct Causal Effects

I used the piecewiseSEM package for analysis 3a in Table 8.1 to generate the OLS estimators of unstandardized direct effects in the Roth et al. (1989) path model that are listed in the second and third columns of Table 8.3 and shown in boldface. These results control for the parents of each outcome.. Because there are no causes of fitness other than exercise and also no backdoor paths between these two variables—see Figure 7.5—the adjustment set is empty (i.e., no covariates). This means that the bivariate regression of fitness on exercise is the sole OLS estimator for this effect. The unstandardized coefficient is .108 (see the table), which indicates that fitness is expected to increase by .108 points in its raw score metric, given a 1-point increase in the raw score metric of exercise. Exercise 2 asks you to interpret the unstandardized OLS coefficient for the direct effect of hardy on stress.

There are a total of three OLS estimators for the unstandardized direct effect of fitness on illness, each with a different adjustment set (see Table 8.3). Their values are generally consistent and range from –1.036 when controlling for exercise to –.849 when controlling for stress, the other parent of illness in the original model (Figure 7.5). The result just mentioned says that for every 1-point increase in fitness, there is an expected decline in illness of .849 points while holding stress constant. There are also a total of three estimators for the unstandardized direct effect of stress on illness, each with just one of the variables fitness, exercise, or hardy as the adjustment set (Table 8.3). Values of these alternative estimators are all positive and generally consistent. Exercise 3 asks you to interpret the OLS coefficient for the direct effect of stress on illness while controlling for fitness, the other parent of illness.

Reported in the fourth and fifth columns in Table 8.3 are the 2SLS estimates of unstandardized direct effects in the Roth et al. (1989) path model. These results are inconsistent or plainly anomalous for some effects, and thus problematic. For example, the estimate for the direct effect of exercise on fitness is *negative*, or –.646 (i.e., more exercise, less fitness) when the instrument is hardy, but the coefficient for the same direct effect is *positive*, or .719 (i.e., more exercise, more fitness) when the instrument is stress. There is a similar inconsistent

TABLE 8.3. Unstandardized Local Estimates for Direct Effects in a Recursive Path Model of Illness

	OLS		2SLS	
Effect	Estimate	Adjustment set	Estimate	Instrument
Exercise → Fitness	**.108 (.013)**	—	***–.646 (1.377)***	Hardy
			.719 (.687)	Stress
Hardy → Stress	**–.203 (.045)**	—	***1.469 (3.252)***	Exercise
			–1.637 (1.240)	Fitness
Fitness → Illness	**–.849 (.162)**	Stress	–.558 (.443)	Exercise | Stress
	–1.036 (.183)	Exercise	–6.927 (8.533)	Hardy | Stress
	–.951 (.168)	Hardy		
Stress → Illness	**.574 (.089)**	Fitness	***88.191 (5,980.901)***	Exercise |
	.628 (.091)	Exercise	1.180 (.431)	Fitness
	.597 (.093)	Hardy		Hardy | Fitness

Note. OLS, ordinary least squares; 2SLS, two-stage least squares. Adjustment sets are minimally sufficient. Standard errors are reported in parentheses. Values in boldface for OLS estimates control for parents of each outcome, and values in italic boldface for 2SLS estimates are contradictory in sign for the same effect or out-of-bounds (invalid).

pattern of 2SLS estimates for the direct effect of hardy on stress depending on the instrument, exercise or fitness (see the table). Both 2SLS estimators for the direct effect of fitness on illness are negative, but the magnitude of the result when the conditional instrument is

Hardy | Stress

exceeds by more than 10-fold the magnitude of the estimate when the conditional instrument is

Exercise | Stress

or, respectively, –6.927 versus –.558 (Table 8.3). Finally, the standard error for the 2SLS estimate of the direct effect of stress on illness, or 5,980.901, is so large compared with the observed standard deviation of the illness variable (62.48; Table 4.3) that no meaningful interpretation seems possible (i.e., it is out-of-bounds, and thus invalid).

There are features of the Roth et al. (1989) data set that handicap estimation with instrumental variables. For instance, the sample correlation between exercise and hardy is practically zero ($r = –.03$; Table 4.3), so these variables would be weak instruments for one another. Another example is that the conditional instrument

Exercise | Stress

is essentially unrelated to stress ($r = –.001$) and, thus, it is a weak instrument when estimating the coefficient for the direct effect of stress on illness (the estimate was invalid; see Table 8.3). Given these problems, estimation with the 2SLS method is not pursued further in this example.

Disturbance Variances

The second column of Table 8.4 lists the observed variances (s^2) for the outcome variables fitness, stress, and illness, and the third column lists the values of R^2 where the predictor variables are the parents of each outcome. Proportions of explained variation range from .053 for stress to .177 for illness. The standardized disturbance variances are calculated as $1 – R^2$, or the proportion of variance *not* explained, for each outcome. For example, $R^2 = .152$ for fitness, $1 – .152 = .848$, so exercise does *not* explain .848 of the total variation in fitness.[3] Unstandardized disturbance variances are calculated as $(1 – R^2)\ s^2$. For fitness, the unstandardized disturbance variance is calculated as .848 (338.56), or 287.009. Exercise 4 asks you to interpret the results in Table 8.4 for illness.

Parametric Model Diagram with Estimates

Now we have unstandardized OLS estimates for all direct effects and disturbance variances in the Roth et al. (1989) path model. They are shown in their proper places in Figure 8.1(a) depicted in full McArdle–McDonald RAM symbolism. Estimates for direct effects on illness control for both of its parents, stress and illness (Table 8.3). Because the variables exercise and hardy are exogenous, their variances and covariances are also model parameters, but these values are just the corresponding descriptive statistics (Table 4.3).

Because not all measured variables in Figure 8.1(a) have the same raw score metric, values of unstandard-

[3] Values of R^2 adjusted for shrinkage could be substituted for unadjusted R^2 in these calculations.

TABLE 8.4. Unstandardized Ordinary Least Squares Estimates of Disturbance Variances for a Recursive Path Model of Illness

Outcome	s^2	R^2	$1 – R^2$	Unstandardized estimate
Fitness	338.56	.152	.848	287.099
Stress	1,122.25	.053	.947	1,062.771
Illness	3,903.75	.177	.823	3,212.786

Note. The parent(s) of fitness, stress, and illness are, respectively, exercise, hardy, and both fitness and stress. The $1 – R^2$ values are the standardized estimates of disturbance variances.

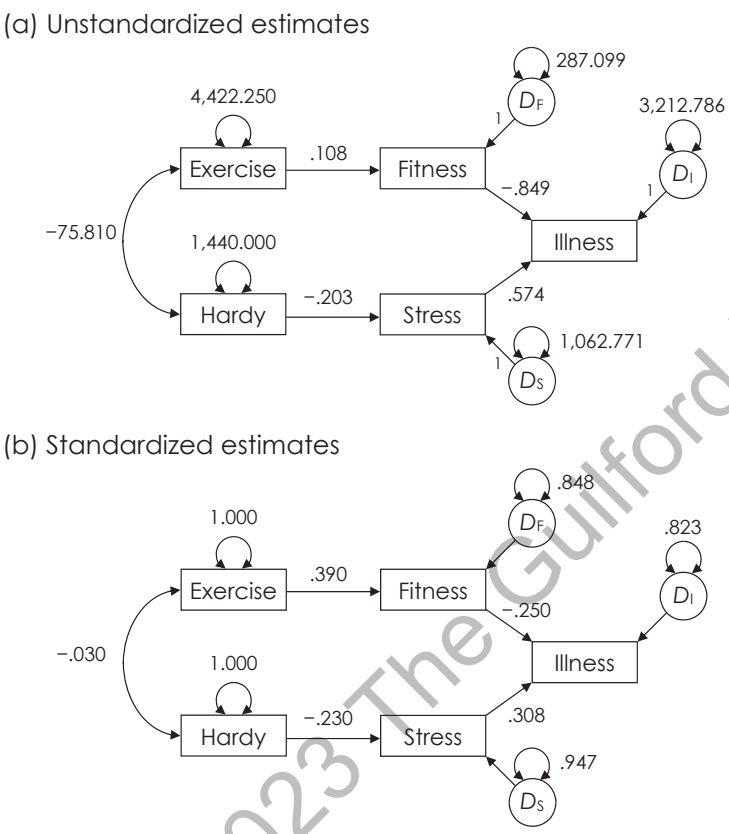

FIGURE 8.1. A recursive path model of illness with ordinary least squares estimates. Results for illness are based on both fitness and stress as parents.

ized path coefficients, such as for direct effects on illness from fitness (–.849) and stress (.574), cannot be directly compared. This is not a problem for the standardized coefficients, which are presented in Figure 8.1(b). For example, the standardized direct effect of fitness on illness is –.250, which says that for every increase in fitness of 1 standard deviation, the level of illness is expected to decline by .25 standard deviations while controlling for stress. The standardized direct effect of stress on illness is .308, so the level of illness is predicted to increase by about .30 standard deviations for every increase in stress of 1 standard deviation while controlling for fitness. Because both results just mentioned are expressed in a common metric (standard deviation units), they can be directly compared: The absolute magnitude of the standardized direct effect of stress on illness exceeds that of fitness by about 23% (.308/.250 = 1.23). Exercise 5 asks you to interpret the standardized direct effects of exercise and hardy on their respective outcomes.

Figure 8.1(a) for the unstandardized estimates does not include regression intercepts, or values of predicted scores when scores on all predictors equal zero, for fitness, stress, and illness, the outcome variables. (All intercepts are zero in the standardized solution.) Intercepts are usually reported in output from regression computer procedures—see the output files for analysis 3a for this example (Table 8.1)—and their values could be reported for each outcome along with those for unstandardized regression coefficients. In contrast,

some SEM computer tools do not generate or print intercepts unless specifically instructed to do so, but we will consider this issue in the next chapter.

Indirect Effects

The fact that indirect effects do not automatically warrant interpretation as "mediation" in cross-sectional designs with no temporal precedence or a clear conceptual time-ordering of cause, mediator, and outcome was discussed in Topic Boxes 6.1 and 7.1. For historical completeness, **the four steps** by Baron and Kenny (1986) for testing mediation are described in Topic Box 8.1. Note that simply following the four steps does not by itself "prove" mediation. *That is, analysis is insufficient to establish mediation without strong theory.*

The direct effects of exercise and hardy on illness in the Roth et al. (1989) path model are both fixed to zero. Each presumed causal variable just mentioned has a single indirect effect on illness, exercise through fitness, and hardy through stress (Figure 8.1). Both of these indirect effects are also total effects, so they can be estimated in two different ways: (1) as products of coefficients from the direct effects that comprise each part of the whole indirect pathway, and (2) through covariate adjustment. Both types of estimates just mentioned are described next.

The second column of Table 8.5 gives the values of the product estimators for both indirect effects in the Roth et al. (1989) path model. The unstandardized estimate for the effect of exercise on illness through fitness is –.092, which equals the product of the two unstandardized coefficients for the two direct effects that make up the indirect pathway, or .108 (–.849) (see Figure 8.1(a)). Note that the second term of the product for the effect of fitness on illness, –.849, controls for stress, the other parent of illness. In words, the unstandardized estimate for the indirect effect means that .092 is the expected *decrease* in illness in its raw score metric while holding exercise constant and increasing fitness to whatever value it would attain under a one-point increase in the raw score metric of exercise (Pearl, 2009, pp. 355–358). This definition is actually counterfactual because it expresses what *could* happen (a decrease in illness), if a previous condition had been different (increasing fitness to the level it would be after an increase in exercise).

The standardized product estimator for the indirect effect of exercise on illness through fitness is –.099 (Table 8.5). It is calculated as .390 (–.250), which is the product of the standardized coefficients for the direct effects that compose the indirect pathway (Figure 8.1(b)). Thus, illness is predicted to decrease by .099 standard deviations while keeping exercise constant and increasing fitness to the level it would be under an increase in exercise of a full standard deviation. Exercise 6 asks you to reproduce the calculations for product estimators of the unstandardized and standardized indirect effect of hardy on illness through stress reported in Table 8.5 and interpret both path coefficients. Note that indirect effects in the table are based on the direct effect of fitness on illness controlling for stress or on the direct effect of stress on illness controlling for fitness. There are other potential product estimators for direct effects of fitness and stress on illness that control for different variables—see the OLS estimates in Table 8.3. Thus, multiple product estimators for each indirect effect through fitness and stress are available in this example.

Because product estimators of indirect effects have complex distributions over random samples, it can be challenging to estimate their standard errors. The best-known example of a method amenable to hand calculation for unstandardized indirect effects that involve just three variables is the **Sobel approximate standard error** (Sobel, 1982). Suppose that a is the unstandardized coefficient for the direct effect $X \rightarrow W$ and that SE_a is its standard error. Let b and SE_b, respectively, stand for the same things for the direct effect $W \rightarrow Y$. The product ab estimates the unstandardized indirect effect of X on Y through W, and its standard error is approximated as

$$SE_{ab} = \sqrt{b^2 SE_a^2 + a^2 SE_b^2} \qquad (8.2)$$

Values of the Sobel standard errors for both unstandardized indirect effects in the Roth et al. (1989) path model are reported in Table 8.5. Exercise 7 asks you to reproduce the calculations for the standard error of the unstandardized indirect effect of exercise on illness through fitness.

In large samples, the ratio $z = ab/SE_{ab}$ is the **Sobel test** for the unstandardized indirect effect. A web page by K. Preacher automatically calculates the Sobel test.[4] The same calculator also gives results for the **Aroian test** and the **Goodman test**, each of which is based on somewhat different approximations of the standard error compared with the Sobel test. Specifically, the value of the Sobel standard error is *smaller* than the

[4] *http://www.quantpsy.org/sobel/sobel.htm*

TOPIC BOX 8.1

Mediation: The Four Steps and Assumptions

Baron and Kenny (1986) described the application of multiple regression over four steps to estimate indirect effects among continuous variables. The steps were originally phrased in terms of statistical significance, but that language was later changed to refer to zero versus nonzero coefficients. This is because coefficients that are trivially small can be significant in large samples while very large coefficients can fail to be significant in small samples (Kenny, 2021). *This means that statistical significance is not a decision criterion when estimating indirect effects* (or any other kinds of effects). The four steps listed next refer to Figure 8.2, where *X*, *M*, and *Y* designate, respectively, the hypothesized cause, mediator, and outcome, and where *a*, *b*, *c*′, and *c* represent coefficients for direct effects between these variables:

1. The cause affects the outcome ignoring the mediator, that is, coefficient *c* in Figure 8.2(a) is not zero.
2. The cause affects the mediator, that is, coefficient *a* in Figure 8.2(b) is not zero.
3. The mediator affects the outcome controlling for the cause, that is, coefficient *b* in Figure 8.2(a) is not zero.
4. To claim that the mediator is completely responsible for the relation between cause and outcome, coefficient *c*′ should be zero.

In Figure 8.2(b), the product *ab* estimates the indirect effect of *X* on *Y* through *M*. The quantity *ab* + *c*′ estimates the total effect of *X* on *Y*, or the sum of the direct and indirect effects of *X*. It also equals coefficient *c* in Figure 8.2(a) where *X* is the sole cause of *Y*. For continuous variables,

$$c - c' = ab \tag{8.3}$$

That is, the difference between the total effect of *X* in Figure 8.2(a) ignoring *M* and the direct effect of *X* in Figure 8.2(b) controlling for *M* equals the product estimator for the indirect effect. Equation 8.3 does not hold when *Y* is a binary outcome variable analyzed using logistic regression or probit regression. This is

(continued)

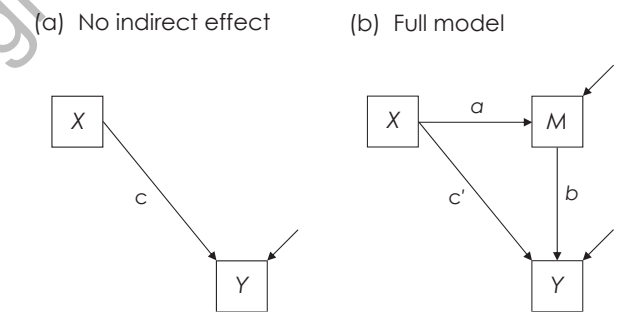

FIGURE 8.2. Models for putative cause, mediator, and outcome variables, respectively, *X*, *M*, and *Y*. Partial model with no indirect effect (a). Full model with direct and indirect effects (b).

because the variance of the outcome variable is not fixed across the models in Figures 8.2(a) and 8.2(b) analyzed in logistic or probit regression. In contrast, the scale in standard regression analyses for continuous outcomes is constant over equations—see MacKinnon (2008, chap. 11) for more information and examples.

The requirement in Step 1 just listed that coefficient c should not be zero is problematic because it does not allow for **inconsistent mediation**—also called **competitive mediation**—where the signs of ab and c' in Figure 8.2(b) are different. In this case, the total effect c in Figure 8.2(a) could be zero even though the size of the product estimator ab is appreciable. Suppose that

$$a = -.50,\ b = .30, \text{ and } c' = .15$$

for Figure 8.2(b). The indirect effect of X on Y through M is $ab = -.50(.30) = -.15$, which exactly cancels out the direct effect of X on Y, or .15, when the total effect is computed, or

$$c = ab + c' = -.15 + .15 = 0$$

The situation where the coefficients for the two constituent direct effects of an indirect effect for three variables (e.g., a and b in Figure 8.2(b) have the same sign is called **consistent mediation** or **complementary mediation** (Zhao et al., 2010).

James and Brett (1984) argued that Step 3 just listed should be modified by not controlling for the causal variable X in Figure 8.2(b) when estimating coefficient b for the direct effect of the mediator, if the hypothesis involves **complete mediation**. This means that the cause is unrelated to the outcome when the mediator is held constant. If so, then including the cause adds nothing to the prediction of the outcome over what is already explained by the mediator (i.e., $c' = 0$). Step 4 refers to the expectation for complete mediation. In contrast, **partial mediation** is indicated when $c' \neq 0$; that is, the mediator is not solely responsible for the observed association between cause and outcome. Because complete mediation is not always expected, though, there may be little harm in routinely controlling for the cause in Step 3 (Tate, 2015).

In consistent mediation, controlling for the mediator *weakens* the association between the cause and outcome variables, or $c' < c$ in absolute value for Figure 8.2. Suppression can be described as special cases of inconsistent mediation where controlling for the mediator *strengthens* the association between cause and outcome, or $c' > c$ in absolute value. That is, the relation between cause and outcome is strengthened by the mediator's omission. In general, suppression is indicated when the indirect effect and the total effect have opposite signs (Rucker et. al., 2011). Another indication of suppression is when the magnitudes of the direct and indirect effects exceed that of the total effect (Lachowicz et al., 2018)—see MacKinnon et al. (2000) and Zhao et al. (2010) for more information.

Here is an example of suppression in an actual mediation study of demoralization in breast cancer patients after primary therapy (Peng et al., 2021): The variables are X = stress, M = demoralization, Y = psychological well-being estimated as common factors and

$$a = .48,\ b = -.85, \text{ and } c' = .37$$

The indirect effect is $ab = .48(-.85) = -.41$, the total effect is $-.41 + .37 = -.04$, which is also the model-implied correlation between X and Y. Thus, the relation between stress and well-being increases from –.04

(continued)

ignoring demoralization to .37 after controlling for it. Without including demoralization in the analysis, a researcher could falsely conclude that stress and well-being are basically unrelated among these patients.

In too many mediation studies, researchers have uncritically followed the four-step method while applying statistical significance as basically the sole criterion for interpretation. The "logic" works like this: If the product estimator *ab* is "significant," then variable *M* mediates at least part of cause *X* on outcome *Y*. But without also addressing assumptions, effect size, research design, and theory, the conclusion just stated is unwarranted; Zhao et al. (2010) described additional myths in mediation analysis. New developments in mediation analysis, outlined in Chapter 20, are making the four-step method ever more obsolete.

Aroian standard error but *larger* than the Goodman standard error (e.g., MacKinnon et al., 2002, p. 85). Thus, it can happen that the same indirect effect is "significant" in the Goodman test but "not significant" in the Sobel test or in the Aorian test. It can be difficult to know which outcome is correct in this case because all results are approximate. Thus, *p* values from the Sobel and related tests should not be overinterpreted (i.e., avoid dichotomania).

The Sobel test assumes normality, but distributions of product estimators are not generally normal; instead, such distributions are often asymmetrical with high kurtosis (MacKinnon et al., 2002). The test requires large samples, and *p* values in small samples can be very inaccurate. The test is restricted to unstandardized indirect effects composed of just three variables. An alternative method is nonparametric bootstrapping, which does not assume normality. Nonparametric bootstrapping can be applied to direct or indirect effects, and indirect effects can be composed of ≥ 3 variables. The method generates a bootstrapped confidence interval for a particular effect. If the value of zero is *not* included within the bounds of a bootstrapped 95% confidence interval, then the corresponding effect could be considered as "significant" at the .05 level for a nondirectional test. But if the confidence interval includes zero, then the effect could be considered as "not significant."

Of course, there is no requirement to interpret a confidence interval as a significance test. This is because from the perspective of interval estimation, all values within a confidence interval are considered as basically equivalent within the limits of sampling error at a particular level of confidence (i.e., $1 - \alpha$). For example, if zero falls within the bounds of a confidence, it has no more special status than any numerical value contained by the interval.

Preacher and Hayes (2004) described macros for SPSS and SAS/STAT that generate bootstrapped confidence intervals for unstandardized indirect effects

TABLE 8.5. Unstandardized and Standardized Estimates of Indirect Effects in a Recursive Path Model of Illness

Effect	Product estimator	Estimated as total effect	Adjustment set
Exercise → Fitness → Illness	–.092[a] (.021) –.099	–.080 (.048) –.085 –.059 (.046) –.063	Hardy Stress
Hardy → Stress → Illness	–.116[b] (.031) –.071	–.267 (.084) –.163 –.231 (.081) –.140	Exercise Fitness

Note. The estimator is ordinary least squares for all results. Adjustment set is minimally sufficient when estimating each indirect effect as a total effect. Standard errors for product estimators are Sobel standard errors. Estimates are reported as unstandardized (standard error) standardized.

[a]Bootstrapped 95% confidence interval is [–.131, –.053].

[b]Bootstrapped 95% confidence interval is [–.195, –.065].

that involve just three variables. Preacher and Hayes (2008) described revised SPSS and SAS/STAT macros and also syntax for Mplus and LISREL that extend nonparametric bootstrapping methods to models with multiple intervening variables or indirect pathways composed of ≥ 3 variables. Hayes (2022) described PROCESS, a macro for R, SPSS, and SAS/STAT for analyzing a wide range of models with indirect effects based on nonparametric bootstrapping. There are also R packages that can generate bootstrapped confidence intervals for direct or indirect effects in path models. Examples include MBESS (Kelley, 2022), and `bmem` (Zhang & Wang, 2022), which also has extensive capabilities for handling missing data in mediation studies (Zhang & Wang, 2013).

Confidence intervals or significance tests based on nonparametric bootstrapping are not magic. For example, bootstrapped estimates can be severely biased in small samples, especially if sample distributions do not reflect population distributions. There are various corrections for small sample bias (Dwivedi et al., 2017), but whether corrected estimates in a particular analysis are trustworthy is generally unknown. Also, values of the lower and upper bounds for a bootstrapped confidence interval are potentially not unique unless the researcher specifies a **seed**, or the initial value of the random number generator used by the computer to select cases. Suppose for a particular seed that the value zero falls just *inside* the bounds of a 95% bootstrapped confidence interval, so the corresponding effect is "not significant" at the .05 level. The analysis is rerun except for a different seed, and the value zero now falls just *outside* the bounds of the second confidence interval. Now the same effect is "significant," again at the .05 level. This "disagreement" is not surprising because statistical results based on simulated random sampling are typically indeterminate (i.e., not unique).

Presented in the third and fourth columns of Table 8.5 are results for the indirect effects of exercise and hardy on illness estimated as total causal effects through covariate adjustment (see also Table 6.5). For example, two different minimally sufficient sets identify the total effect of exercise on illness. The unstandardized and standardized estimators with hardy as the covariate are, respectively, –.080 and –.085, and the corresponding estimators derived with stress as the covariate are, respectively, –.059 and –.063. Results across the three different estimators (including the product estimators) of the same indirect effect are generally similar: The unstandardized coefficients for the indirect effect of exercise on illness range from –.092 to –.059, and the standardized coefficients range from –.099 to –.063 (see Table). Exercise 8 asks you to verify that outcomes of significance testing are *not* consistent over different estimators for the same indirect effect.

For analysis 4 in Table 8.1, I used the `bmem` package to generate bootstrapped 95% confidence intervals for estimates of *all* model parameters, but next we consider only results for the indirect effects. An advantage of `bmem` in this analysis is that it allowed the specification that direct effects of both exercise and hardy on illness are zero; that is, effects of these causal variables are solely indirect—see Figure 8.1. The method specified was the **bias-corrected bootstrap**, which adjusts for possible asymmetry in the empirical sampling distribution by determining the proportion of bootstrapped estimates that fall below the observed result (Efron, 1987). The default total of 1,000 generated samples was not changed in these analyses. The path model was specified for analysis in `bmem` using syntax from the `sem` package (Fox et al., 2022).

The bootstrapped estimate of the standard error for indirect effect of exercise on illness through fitness is .020, or very similar to the Sobel standard error for this effect (.021; see Table 8.5). The bootstrapped 95% confidence interval is [–.131, –.053], and the corresponding point estimate is –.092—see Table 8.5. For the indirect effect of hardy on illness through stress, the bootstrapped standard error of .032 is just slightly larger than the Sobel standard error of .031 for this effect, and the bootstrapped 95% confidence interval is [–.195, –.065] for the point estimate of –.116. Neither bootstrapped confidence interval includes zero, so both point estimates are statistically significant at the .05 level, but keep in mind the limitations of significance tests based on nonparametric bootstrapping considered earlier.

Whether *any* effect, indirect or otherwise, is significant or not significant at some arbitrary level of α may be irrelevant, especially if the researcher emphasizes interval estimation and also considers whether observed effect sizes are large and precise enough to matter in a particular research area. In Chapter 20, which covers enhanced mediation analysis, you will learn about additional ways to describe the magnitudes of indirect effects. The next chapter concerns global estimation, and we will consider analysis results for the same model and data as covered in this detailed example of local estimation.

SUMMARY

In local estimation, the equation for just one outcome at a time is analyzed using a suitable regression method. It may be less susceptible to the propagation of specification error than global estimation. This means that if the equation for a particular outcome is wrong, then the error need not inevitably contaminate estimates for other outcomes. Local estimation is also the last step in the method of piecewise SEM. The initial steps are to (1) express a path model as a DAG; (2) derive the union basis set of conditional independencies implied by graph; (3) test each of these hypothesized independence relations against sample data; and (4) conduct the multivariate d-separation test over all implied conditional independencies. A failed (i.e., significant) d-separation test signals covariance evidence against the model, but effect size, or the empirical magnitudes of departures from implied conditional independence, should be considered, too. If either the original model or a respecified version is retained, its parameters are locally estimated. There may be multiple estimators for some parameters, but their results should generally be consistent; otherwise, a problem is indicated. Global estimation is introduced in the next chapter.

LEARN MORE

Shipley (2000) and Lefcheck (2016) outline the logic of piecewise SEM, and Zhao et al. (2010) classify types of mediated effects and caution against blindly applying the classical four-step method for testing mediation.

Lefcheck, J. S. (2016). piecewiseSEM: Piecewise structural equation modelling in R for ecology, evolution, and systematics. *Methods in Ecology and Evolution, 7*(5), 573–579.

Shipley, B. (2000). A new inferential test for path models based on directed acyclic graphs. *Structural Equation Modeling, 7*(2), 206–218.

Zhao, X., Lynch, J. G., Jr., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research, 37*(2), 197–206.

EXERCISES

1. Calculate C for the d-sep test based on the results in Table 8.2.

2. Interpret the OLS result in Table 8.3 for the direct effect of hardy on stress.

3. Interpret the OLS result in Table 8.3 for the direct effect of stress on illness, controlling for fitness.

4. Interpret the results in Table 8.4 for illness.

5. Interpret the standardized coefficients in Figure 8.1(b) for the direct effects of exercise and hardy on their respective outcomes.

6. Calculate and interpret the product estimators for the unstandardized and standardized indirect effects of hardy on illness through stress, given the coefficients in Figure 8.1.

7. In Table 8.5, the Sobel approximate standard error for the unstandardized indirect effect of exercise on illness through fitness is .021. Calculate this result based on the information in Table 8.3.

8. Verify in Table 8.5 that outcomes of significance tests for estimators of the unstandardized indirect effect of exercise on illness through fitness are not consistent.