PREFACE

Social and behavioral scientists routinely analyze complex relationships between abstract concepts that defy simple measurement. Concepts such as depression, democracy, or organizational culture are difficult to capture with single measures. Furthermore, theories or conceptual models guiding social and behavioral science analyses often specify complex webs of relationships among the concepts. These relationships may include confounders of focal relationships, chains of effects that reflect mechanisms linking focal concepts, or intermediate processes, among other possibilities.

Structural equation models (SEMs) provide an elegant statistical framework that captures this theoretical complexity. At their core, SEMs offer two powerful capabilities that set them apart from other statistical models. First, SEMs permit the specification of latent (unobserved) variables representing abstract concepts, linked to manifest (observed) variables through flexible measurement models. This approach acknowledges the inherent challenges in measuring social and behavioral constructs and provides rigorous methods to account for measurement error. Second, SEMs accommodate multiple dependent or endogenous variables within a single integrated model. This capability allows researchers to specify a statistical model that reflects the interrelationships present in many conceptual or theoretical models.

This book serves as an introduction to structural equation modeling for social and behavioral scientists. We begin with the general model, providing researchers with the tools to translate complex theoretical frameworks into testable statistical models, to fit models to data, and to evaluate the fit of specified models. In addition, we devote a chapter to an extended consideration of mediation analysis that highlights assumptions underlying these analyses and incorporates contemporary developments in causal inference. We also include a chapter on fitting SEMs with categorical or limited endogenous variables. This foundation prepares readers for more advanced models in the SEM family, including latent growth models for longitudinal processes and latent class models for categorical latent variables.

This book is intended for intermediate to advanced graduate students and researchers in the social and behavioral sciences. Readers should be familiar

with linear regression models. Those needing a refresher may consult excellent primers such as Fox (2016), Gujarati (2018), or Lewis-Beck (2015). An understanding of models for limited and categorical dependent variables will be helpful for Chapter 6. Long's (1997) *Regression Models for Categorical and Limited Dependent Variables* provides an excellent introduction. For readers interested in specific SEM applications, the Quantitative Analysis in the Social Sciences (QASS) series offers volumes on measurement models (Roos & Bauldry, 2021), nonrecursive simultaneous equations models (Paxton et al., 2011), latent growth models (Preacher et al., 2008), and multilevel SEMs (Silva et al., 2019), among others. For a more advanced treatment of the general SEM, Bollen's (1989) classic text *Structural Equations With Latent Variables* remains an essential resource, and Mulaik's (2009) *Linear Causal Modeling With Structural Equations* is also excellent.

CHAPTER 1. INTRODUCTION

Social and behavioral science research frequently focuses on the analysis of complex connections among abstract or theoretical constructs. Concepts such as depression, democracy, or a supportive organizational culture can be difficult to operationalize with a single measure. For example, depression is a psychological condition that can manifest itself with a variety of symptoms, such as feelings of sadness, loss of interest in activities, changes in sleep patterns, and difficulty concentrating. Similarly, democracy is a multifaceted political system that encompasses various aspects, such as free and fair elections, political participation, and the protection of civil liberties. Capturing these intricate concepts typically requires drawing on multiple measures that reflect different facets of the underlying construct.

Theories or conceptual models guiding social and behavioral science analyses often specify complex webs of relationships among the concepts. Such models may include important confounders of a focal relationship that need to be taken into account in any analysis. For example, a conceptual model of the relationship between education and health likely includes background factors, such as parental education and childhood health, that affect both adult education and adult health. Theoretical and conceptual models may also include intermediate variables that transmit the effect of a focal independent variable to an outcome. Continuing with our example of education and health, a theoretical model may specify socioeconomic resources, supportive networks, and health behaviors as a few of the mechanisms that link education and health. These intermediate variables may also have relationships with each other. To give another example, a conceptual model of organizational performance might specify a chain of effects in which leadership style influences employee motivation, which, in turn, affects job satisfaction and ultimately leads to improved organizational outcomes. Such complex relationships require statistical methods that can adequately account for the intricate connections among variables of interest.

Structural equation models (SEMs) provide a statistical framework that captures and reflects the richness of theoretical and conceptual models. SEMs permit the specification of models with latent, or unobserved, variables that represent social and behavioral concepts. Latent variables are typically measured by manifest, or observed, variables that are indicators of concepts. For example, a latent variable that represents depressive symptoms could be measured by a set of survey items that ask respondents about various symptoms they experienced during the past week. With multiple

indicators, SEMs can more accurately capture an underlying concept of interest while accounting for measurement error in individual items.

SEMs allow for a broad range of measurement models linking concepts to the variables that measure them. In addition, SEMs permit the specification of models with multiple dependent (or endogenous) variables, either latent or observed. The ability to handle multiple dependent variables allows researchers to analyze various forms of complex relationships present in many theoretical and conceptual models. For example, a SEM based on a theoretical model of education and health could include various confounders, such as childhood health, and intermediate variables, such as smoking. Following the estimation of model parameters, SEMs provide a means of decomposing a total effect (e.g., the effect of education on health) into indirect (e.g., the effect of education on health that operates through smoking) and direct effects (e.g., the effect of education on health net of smoking). Given a set of assumptions that we discuss in Chapter 5, this decomposition enables researchers to test hypotheses about the processes through which one variable influences another, such as the mediating role of smoking in the relationship between education and health.

The general SEM can be considered to consist of two component models, the measurement model and the structural model. For SEMs that include one or more latent variables, the measurement model specifies the relationships between any latent variables and their observed measures, often referred to as indicators of the latent variables. The structural model, in contrast, specifies the relationships between latent variables or between observed variables that are not typically themselves indicators of latent variables. We now turn to a discussion of each of these component models before bringing them together in the general model.

Empirical Example

To help introduce various features of the general SEM, we will draw on an extended empirical example that we will return to throughout the book. Suppose that we are interested in relationships between marriage, psychological distress, and drinking among young adults. To analyze these relationships, we can draw on data for respondents ages 26 to 29 from the 2023 wave of the National Survey on Drug Use and Health (NSDUH; Center for Behavioral Health Statistics and Quality, 2025).

These data include a measure of marital status that differentiates respondents who are currently married, widowed, divorced, and never married. Given our focus on young adults, we excluded a small number of respondents who reported being widowed or divorced from our analysis

sample and constructed a binary variable for respondents who were currently married at the time of the survey versus having never been married (1 = currently married, 0 = never been married).

NSDUH includes six indicators that comprise the Kessler Psychological Distress Scale (K6; (K6; Kessler et al., 2002)). The indicators begin with the following stem: "During the past 30 days, how often did you feel " and include (1) nervous, (2) restless or fidgety, (3) so sad or depressed that nothing could cheer you up, (4) hopeless, (5) that everything was an effort, and (6) down on yourself, no good, or worthless. The response categories range from 1 = none of the time to 5 = all of the time (note: the response categories are reverse-coded from the original survey items).

As a measure of drinking, we draw on the number of days that respondents reported consuming alcohol in the past month. In addition, we include two demographic covariates, sex (1 = female, 0 = male) and race/ethnicity (1 = Hispanic, 2 = non-Hispanic Black, 3 = non-Hispanic White). We excluded a small number of respondents from other racial/ethnic groups from the analysis sample. Table 1.1 provides descriptive statistics for our focal variables along with the two demographic covariates.

It is important to note that our empirical example involves a number of simplifications for pedagogical purposes. For example, the category of people who have never been married is heterogeneous and may include people in long-term relationships or cohabiting. In addition, many factors beyond the two demographic characteristics relate to marriage, psychological distress, and drinking. Our simple empirical example, nonetheless, allows for an illustration of many features of SEMs.

Latent and Observed Variables

As discussed above, a key feature of SEMs is the ability to specify and analyze latent variables that represent social or behavioral concepts. There are several benefits of working with latent variables. First, as in our example with psychological distress, there can be multiple ways to operationalize or measure some concepts. Specifying a model with a latent variable allows an analyst to avoid having to choose a single measure and instead take advantage of multiple measures. Second, any given observed measure of a latent variable includes some amount of measurement error (i.e., variance in the measure that stems from sources other than the concept). For example, only a portion of the variance in the second indicator concerning how restless respondents felt in the past month is due to respondents' levels of psychological distress. Some of the variance in the indicator reflects idiosyncratic variation in responses across respondents, and perhaps some

Table 1.1 Descriptive statistics for marriage, psychological distress, and drinking e mpirical example; $N = 3{,}153$.

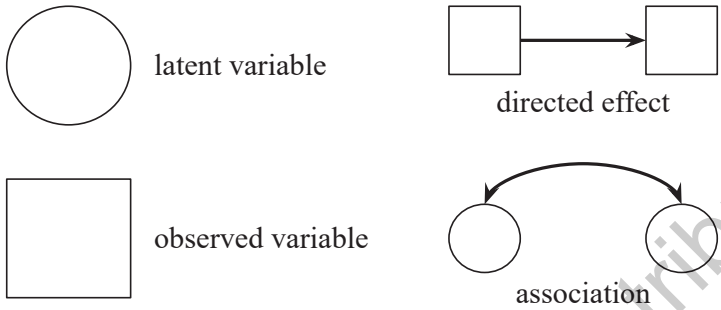
	Mean	*SD*	*Range*
Focal outcome			
drinks	4.24	6.43	0–30
Psychological distress			
nervous	2.37	1.12	1–5
restless	2.13	1.16	1–5
depressed	1.74	1.03	1–5
hopeless	1.76	1.06	1–5
effort	2.19	1.26	1–5
worthless	1.87	1.13	1–5
Focal predictor			
married	0.36	0.48	0,1
Demographic characteristics			
female	0.56	0.50	0,1
Hispanic	0.26	0.44	0,1
non-Hispanic Black	0.12	0.33	0,1

Source: National Survey on Drug Use and Health, 2023 wave, respondents ages 26 to 29. $N = 94$ cases excluded due to missing data for indicators of psychological distress. Descriptive statistics obtained using R.

reflects systematic biases (e.g., method effects such as the ordering of questions on the survey). It is also possible to construct a scale from the six indicators by taking the average. Scales help reduce measurement error, but they do retain some measurement error and, as such, remain imperfect measures of latent variables. Finally, including latent variables in a model allows researchers to evaluate measurement properties of indicators, such as their reliability, as part of an analysis. For example, the measurement component of a SEM that involves psychological distress reveals the relative quality of each indicator.

SEMs can be represented either graphically or as a system of equations. Figure 1.1 illustrates the key features for the graphical representation

Figure 1.1 Graphical syntax for SEMs.

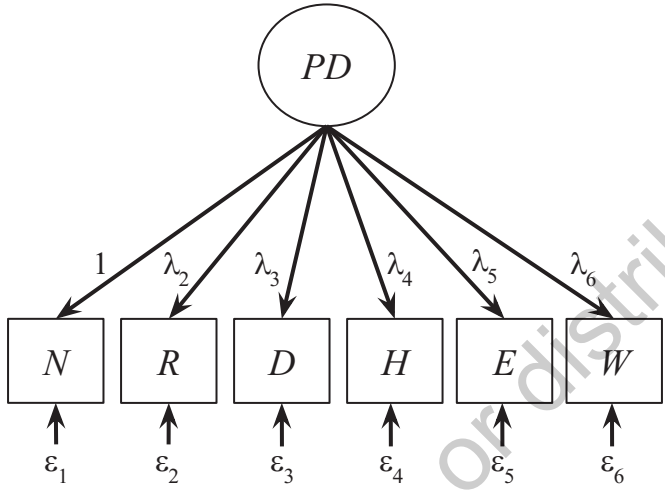


of SEMs. By convention, latent variables are indicated with circles or ovals, while observed variables are indicated with rectangles or squares. Single-headed arrows represent the effect of one variable on another, and double-headed arrows represent an association or correlation between two variables. These simple conventions can be used to represent a very broad array of statistical models.

Returning to our example, we can depict the measurement component of our model with a latent variable for psychological distress and six indicators of the latent variable (see Figure 1.2). In this figure, the latent variable is denoted by an oval labeled "*PD*" for psychological distress, and the six indicators of the latent variable are denoted by squares labeled with initials for each indicator. Direct effects stem from the latent variable and point toward the indicators. This represents a standard approach to measurement that assumes that the latent variable is a cause of each of the measures, which are called effect or reflective indicators. Latent variables can also be measured by causal or formative indicators for which the direction of effects is in the opposite direction (Bollen & Bauldry, 2011). Chapter 2 provides an example of this type of measurement model. In addition to the latent variable representing psychological distress and the observed variables, Figure 1.2 also includes six errors, $\varepsilon_1 - \varepsilon_6$, which have direct effects on the respective indicators. These errors, which capture sources of variation in the indicators that are independent of psychological distress, can also be thought of as latent variables, and they are sometimes represented in diagrams enclosed in circles.

Figure 1.2 includes labels for the direct effects of psychological distress on each indicator. The first label is simply 1, which represents fixing the effect of the latent variable on the first indicator to be 1. This is one approach to providing a scale for the latent variable, an issue we discuss in

Figure 1.2 Measurement model for psychological distress.



detail in Chapter 3. The five remaining labels represent factor loadings and, by convention, are labeled using the Greek letter lambda ($\lambda_2 - \lambda_6$,). Factor loadings give the effect of the latent variable on the respective indicators and are estimated from fitting the model to data.

A diagram representing a SEM is often all an analyst needs to convey the structure of a model, but it can also be useful to represent a SEM as a system of equations. The measurement component of our model for psychological distress can be represented by the following equations:

$$N_i = \alpha_1 + (1)PD_i + \varepsilon_{1i} \tag{1.1}$$

$$R_i = \alpha_2 + \lambda_2 PD_i + \varepsilon_{2i} \tag{1.2}$$

$$D_i = \alpha_3 + \lambda_3 PD_i + \varepsilon_{3i} \tag{1.3}$$

$$H_i = \alpha_4 + \lambda_4 PD_i + \varepsilon_{4i} \tag{1.4}$$

$$E_i = \alpha_5 + \lambda_5 PD_i + \varepsilon_{5i} \tag{1.5}$$

$$W_i = \alpha_6 + \lambda_6 PD_i + \varepsilon_{6i} \tag{1.6}$$

where i indexes cases; *PD* refers to latent psychological distress; *nervous* (*N*), *restless* (*R*), *depressed* (*D*), *hopeless* (*H*), *effort* (*E*), and *worthless* (*W*) are indicators; and $\varepsilon_1 - \varepsilon_6$ are errors. As noted above, the first indicator is

treated as the scaling indicator with a factor loading fixed to 1. The model parameters to be estimated include six intercepts ($\alpha_1 - \alpha_6$), five factor loadings ($\lambda_1 - \lambda_6$), six error variances ($var(\varepsilon_1) - var(\varepsilon_6)$), and the variance of latent psychological distress. In Chapter 3, we cover commonly used estimators for these parameters. In this example, we treat the indicators as continuous measures, which is often the case for ordinal variables with five or more values. However, we could treat the indicators as ordinal measures, a possibility that we cover in Chapter 6 when we discuss SEMs with categorical dependent or endogenous variables.

With respect to interpretation, factor loadings provide estimates of the effect of a latent variable on each indicator (with the exception of the scaling indicator). As in linear regression models, the intercepts provide an estimate of the mean of the indicator when the latent variable (the only predictor) is 0. These are not often interpreted and, in fact, are not even mentioned as parameters in some textbook treatments of measurement models. In addition, similar to linear regression models, the variances of the errors can be used in the calculation of R-squares for each indicator to measure the proportion of variance in the indicator explained by the latent variable. In this context, R-squares can be interpreted as measures of the reliability of each indicator. Indicators for which a latent variable explains a high proportion of variance exhibit higher reliability than indicators for which a latent variable explains a low proportion of variance. In our example with psychological distress, we would ideally like to see factor loadings with magnitudes indicating strong relationships with latent psychological distress and R-squares for all indicators that suggest that a high proportion of variance is accounted for by the latent variable. What constitutes "strong relationships" and "a high proportion of variance" will vary between research contexts. We discuss how to obtain these parameter estimates and additional details about interpretation in Chapter 3.

In addition to parameter estimates, some SEMs permit tests of the overall fit of the model with the data. These tests provide an indication of how well a specified model reproduces patterns among the means, variances, and covariances of the observed variables. Such tests may lead a researcher to decide that a specified model is inadequate and to consider alternatives or help select among competing models. In our example measurement model for psychological distress, we have 9 degrees of freedom and can therefore evaluate how well our proposed model fits the NSDUH data. We cover how to know if an overall model fit test is available for a SEM and where degrees of freedom are from in our discussion of model specification in Chapter 2, and we cover tests of overall model fit in Chapter 4.

Multiple Endogenous Variables

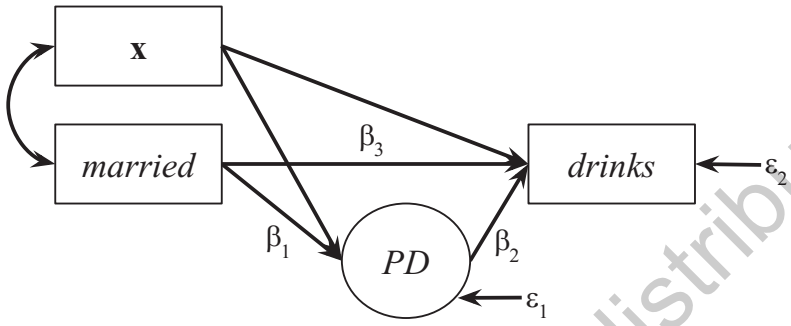
In different areas of statistics, analysts refer to variables with respect to their position in equations with different labels. Perhaps the most common references are to independent and dependent variables, with independent variables on the right-hand side of equations and dependent variables on the left-hand side. With SEMs, we typically refer to exogenous and endogenous variables. Exogenous variables are those that have no directed paths pointing to them or, equivalently, never appear on the left-hand side of an equation. These variables are exogenous in the sense that they are determined outside of the specified model. Endogenous variables are those that have at least one directed path pointing toward them or appear on the right-hand side of at least one equation. It is possible for an endogenous variable to be a dependent variable in one equation and an independent variable in another.

For SEMs that include latent variables, the measurement component of the model relates indicators to the latent variables. In contrast, the structural component of a SEM captures predictive or causal relationships among observed and latent variables (if any). A second key feature of SEMs is the ability to specify models that include more than one endogenous variable. This feature allows researchers to model the interdependencies among all variables or focal variables in an analysis. Returning to our example, we are interested in relationships between marriage, psychological distress, and drinking. In this case, we might specify a SEM that has psychological distress and drinking as endogenous variables predicted by marriage and demographic characteristics. Furthermore, we might also specify psychological distress as a predictor of drinking, in which case it represents an intermediate variable in the overall model.

Figure 1.3 illustrates the structural component of our model for the relationships between marriage, psychological distress, and drinking. Being married has direct effects on psychological distress, depicted as a latent variable, and the number of days drinking in the past month. In addition, psychological distress has a direct effect on drinking. As endogenous variables, both psychological distress and drinking have errors pointing to them. These capture unmeasured factors that influence each variable, just as in a standard linear regression model.

The figure also shows a vector **x** representing demographic characteristics (gender and race/ethnicity) included in the analysis. These demographic characteristics have direct effects on both psychological distress and drinking. In addition, they are correlated with being married. In general, in SEMs, all exogenous variables are assumed to be correlated with each other, which is equivalent to how independent variables are treated in standard regression models.

Figure 1.3 Structural model relating marriage, psychological distress, and drinking.



In our example, we wish to adjust for demographic characteristics that could influence psychological distress and drinking when assessing their relationships with marriage and each other. We might also be interested in how demographic characteristics relate to being married, in which case, we could instead specify direct effects from the demographic characteristics to being married rather than correlations. We discuss decisions along these lines in the next chapter, which examines model specification. Finally, our measure of drinking is a count of the number of days respondents consumed alcohol in the past month. We could treat this measure as continuous; however, SEMs are also capable of handling count endogenous variables, which we cover in Chapter 6.

As with the measurement component of the model, we can express the structural component of a SEM with a system of equations. For our example, we have the following:

$$PD_i = \alpha_1 + \beta_1 married_i + \boldsymbol{\beta}'_{x1}\mathbf{x}_i + \varepsilon_{1i} \tag{1.7}$$

$$drinks_i = \alpha_2 + \beta_2 PD_i + \beta_3 married_i + \boldsymbol{\beta}'_{x2}\mathbf{x}_i + \varepsilon_{2i}\,, \tag{1.8}$$

where *PD* (latent psychological distress) and *drinks* are our two endogenous variables, *married* is a binary variable for being married, **x** is a vector of demographic characteristics, and ε_1 and ε_2 are errors. The parameters in the model include the intercepts (α_1 and α_2) for each endogenous variable, the regression coefficients for each of the predictors (β_1, β_2, β_3, $\boldsymbol{\beta}_{x1}$, $\boldsymbol{\beta}_{x2}$), and the variances for the two errors ($var(\varepsilon_1)$ and $var(\varepsilon_2)$). In addition, we can obtain estimates of the covariances or correlations among the exogenous

variables. The interpretations of the parameter estimates and R-squares are the same as with standard linear regression models. For example, a 1-unit increase in psychological distress is associated with a β_2 difference in the number of days drinking.

The simultaneous estimation of all regression parameters facilitates the decomposition of total effects into direct and indirect effects. The total effect of a focal predictor on a focal outcome is given as the sum of direct and indirect effects. The direct effect is simply the coefficient for the path from the focal predictor to the focal outcome. Indirect effects are calculated by multiplying the coefficients along the paths that connect the focal predictor to the focal outcome. In our example, we have one such path—the path from marriage to psychological distress and then from psychological distress to drinking. As such, the total effect of marriage on drinking is given by the sum of its indirect and direct effects, or $\beta_1 \beta_2 + \beta_3$. This decomposition is frequently used in mediation analysis. A substantively meaningful indirect effect via a particular focal mediator (or set of mediators) is interpreted as evidence that the focal mediator is a mechanism linking the focal predictor to the focal outcome. In Chapter 5, we take a close look at mediation analysis with a particular emphasis on the assumptions supporting the analysis and contemporary thinking around causal analysis.

Although not a feature of our example, the structural component of a SEM can include interactions and nonlinear relationships in the same manner as standard regression models. For example, if we expected that marriage has a different relationship with psychological distress for women and men, we could include that interaction as a predictor in the equation for psychological distress. To give another example, if we were interested in the relationships between marriage, psychological distress, and drinking for a larger age range and included age as a predictor in our model, then we could also allow for nonlinear relationships between age and psychological distress or drinking by including squared age, logged age, or terms for other functional forms depending on our theoretical or substantive expectations.

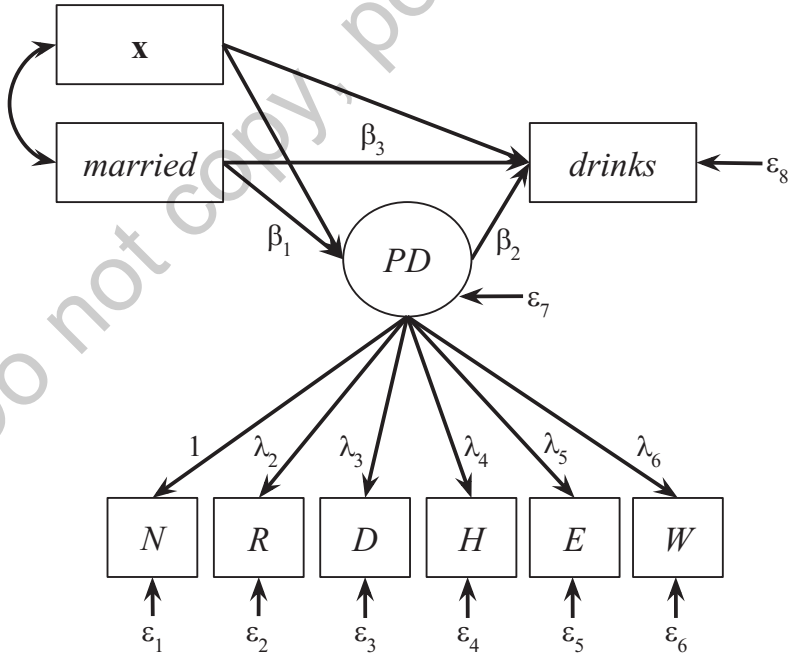
The specification of the structural component of a SEM typically stems from a researcher's understanding of or assumptions about the causal processes linking social and behavioral concepts. However, the fact that model specification is often based on assumptions about causal effects does not mean that the estimates obtained from fitting a SEM to data should be interpreted as causal effects. Whether such an interpretation is merited derives from features of the research design (e.g., the data come from a randomized controlled study) or explicit assumptions about the statistical analysis (e.g., all confounders of a focal relationship have been included in

the model). In this way, SEMs are no different from any other statistical tool, such as linear regression or propensity score matching estimators; a causal interpretation of estimates necessarily rests on specific features of a research design or untestable assumptions. We discuss this issue in more detail in Chapters 2 and 5.

General Structural Equation Model

The general SEM represents the combination of measurement and structural models in a single model. Figure 1.4 illustrates this for our example with marriage, psychological distress, and drinking. We see psychological distress represented as a latent variable with six indicators, as well as structural relationships among marriage, psychological distress, and drinking. The combined model includes all parameters from both the measurement and structural models and maintains the same interpretations we discussed above.

Figure 1. 4 General SEM relating marriage, psychological distress, and drinking.



We can represent the general model as a system of equations by combining the equations for the structural model and measurement model as follows:

$$PD_i = \alpha_7 + \beta_1 married_i + \boldsymbol{\beta}'_{x1}\mathbf{x}_i + \varepsilon_{7i} \quad (1.9)$$

$$drinks_i = \alpha_8 + \beta_2 PD_i + \beta_3 married_i + \boldsymbol{\beta}'_{x2}\mathbf{x}_i + \varepsilon_{8i} \quad (1.10)$$

$$N_i = \alpha_1 + (1)PD_i + \varepsilon_{1i} \quad (1.11)$$

$$R_i = \alpha_2 + \lambda_2 PD_i + \varepsilon_{2i} \quad (1.12)$$

$$D_i = \alpha_3 + \lambda_3 PD_i + \varepsilon_{3i} \quad (1.13)$$

$$H_i = \alpha_4 + \lambda_4 PD_i + \varepsilon_{4i} \quad (1.14)$$

$$E_i = \alpha_5 + \lambda_5 PD_i + \varepsilon_{5i} \quad (1.15)$$

$$W_i = \alpha_6 + \lambda_6 PD_i + \varepsilon_{6i}, \quad (1.16)$$

where the subscripts for the intercepts and the errors for the structural component of the model have been adjusted to avoid duplication with the intercepts and errors of the measurement component of the model. Note that the order in which the equations are written is arbitrary. For some purposes, it can be useful to write the system of equations as a matrix equation. Working with matrices is generally not necessary for an introduction to SEMs, but for those who are interested, the Appendix provides a matrix representation of the general SEM.

The general SEM can be used to represent a very broad range of statistical models (see the Appendix for a few derivations). For example, if there are no latent variables, and thus there is no measurement component to a model, then we are modeling relationships among observed variables with a simultaneous equations model. If, in addition, there is only a single endogenous variable, then we have a standard linear regression model. Alternatively, if a model has one or more latent variables but no direct effects among them, then we have a confirmatory factor analysis model. Furthermore, the general SEM can also be applied to a variety of data structures and lead to other types of models. For example, suppose that we have longitudinal data with repeated measures of a given construct (e.g., reading test scores). If repeated measures are treated as indicators of latent variables representing a trajectory (e.g., growth in reading ability), then we have a latent growth model. The general SEM thus provides a foundation for fitting a wide variety of statistical models and can be flexibly applied to many analysis contexts.

Although the general SEM provides a powerful tool for examining complex theoretical relationships, it is important to be aware of several limitations that we will explore in more detail in the following chapters. First, as

we will see in Chapter 2, SEMs require substantial theoretical grounding to guide model specification, as researchers need to explicitly articulate both how abstract concepts relate to observed measures and how variables interrelate with one another. The potential complexity of SEMs necessitates careful thought to connect research questions to specific parameters. Additionally, although SEMs are frequently discussed using causal language, obtaining estimates that merit a causal interpretation requires particular features of research designs (e.g., random assignment to a treatment) or explicit assumptions about confounding and other threats to inference. The specification of SEMs alone does not produce causal estimates. Finally, as we will discuss in Chapter 4, specifying a SEM that has a good fit with a given source of data does not preclude the possibility of the existence of equivalent or even better-fitting models. These challenges underscore the importance of treating SEMs as tools for evaluating theoretically or substantively grounded models and being appropriately cautious in the interpretation of model fit and parameter estimates.

Statistical Software and Code

Researchers have a number of options for statistical software to fit SEMs. Most general statistical software packages, such as R, Stata, and SAS, have the capability to fit various types of SEMs. In addition, software designed specifically for SEMs, such as Mplus and AMOS, can be used for an even wider range of SEMs. Because the capabilities and syntax of the various statistical software packages continue to evolve, the book does not include snippets of code or output. Instead, code for multiple software packages (R, Stata, and Mplus), along with data for every example in the book, is available on a companion website https://collegepublishing.sagepub.com/products/introduction-to-structural-equation-models-1-295187 to serve as a resource for readers looking to replicate the examples or for templates to use in their own work.

Outline of Book

This introduction to SEMs follows a standard workflow of a SEM analysis through the next three chapters. Chapter 2 covers model specification with a consideration of a variety of measurement models and different forms of structural models. In addition, Chapter 2 includes a discussion of how SEMs relate to directed acyclic graphs (DAGs), an alternative graphical syntax for evaluating causal models. Chapter 3 turns to the issue of model identification—whether it is possible to find unique parameter estimates—and

fitting SEMs to data. Chapter 3 focuses on maximum likelihood (ML) estimators and variants that address deviations from normality and missing data. In addition, Chapter 3 includes an introduction to one of the alternative estimators for SEMs, a model-implied instrumental variable (MIIV) estimator that has some advantages over ML estimators. Chapter 4 discusses tests of overall model fit as well as the relative fit of nested models. Such tests can lead analysts to modify models, a process also discussed in this chapter.

The first four chapters provide a complete introduction to the use of the general SEM. Chapter 5 offers an extended discussion of mediation analysis, one of the most common uses of SEMs. Chapter 6 provides an overview of how to account for categorical or limited endogenous variables in SEMs. Finally, Chapter 7 offers a few general conclusions, notes a range of extensions to the general SEM (e.g., applications to other data structures and alternative estimators), and discusses limitations and common pitfalls with SEMs. In this book, some more advanced sections have been included to give readers a sense of contemporary developments or more technical aspects of SEMs. The advanced sections are denoted with asterisks and can be skipped by readers more interested in the foundational material.

Further Reading

For those interested in a supplementary introduction to SEMs, Kline's (2023) *Principles and Practice of Structural Equation Modeling* is an accessible text. For more advanced and mathematically oriented treatments, *Structural Equations With Latent Variables* by Bollen (1989) offers a precise and technical presentation of the theoretical underpinnings of SEMs that, even after 30 years, remains an authoritative source, and Mulaik's (2009) *Linear Causal Modeling With Structural Equations* provides another excellent mathematical introduction to SEMs. In addition, Roos and Bauldry (2021) provide an overview of measurement models and working with latent variables in a SEM framework.

Beyond alternative textbook coverage of SEMs, readers are encouraged to identify empirical research articles that employ SEMs in their specific fields of interest. Engaging with real-world applications of SEMs, of which there are many, can greatly enhance understanding and provide practical insights into how these models are used to address complex research questions across the social and behavioral sciences. Whether in psychology, sociology, education, health sciences, or any other discipline, studying research using SEMs will enrich your understanding of its capabilities.

CHAPTER 2. MODEL SPECIFICATION

Now that we have a general idea of what SEMs can do, in this chapter, we turn to the issue of model specification. Model specification refers to the process of determining the structure of a SEM in a given research context. Key decisions include (1) deciding whether there are any latent variables and, if so, what observed variables will serve as indicators; (2) selecting which variables, if any, have direct effects on other variables (i.e., which variables are exogenous and which are endogenous); and (3) identifying whether there are correlations between errors of endogenous variables that should be included in the model.

As noted in Chapter 1, it can be useful to develop the measurement component and the structural component of a model in isolation. In our discussion, we maintain this separation and first consider some of the range of possibilities that analysts may encounter with the measurement component of SEMs and then shift to possibilities that analysts may consider with the structural component of SEMs. Before doing so, it is helpful to outline more generally where ideas or decisions regarding model specification come from in the first place.

Specifying a SEM may draw on information from at least four sources: prior studies, substantive expertise, characteristics of the research design or anticipated data for analysis, and a general sense of logic of social and behavioral science patterns. Suppose that a researcher is interested in developing a model to analyze the relationships between higher education, democracy, and economic growth, using countries rather than individuals as the unit of analysis. With this example in mind, we can see some ways in which each of these sources plays a role in the specification of a SEM.

First, prior studies almost always play an essential role in guiding the specification of a SEM. Previous studies can be a source of theoretical or conceptual models that delineate expected direct effects or causal relationships among focal variables. In our example, a theoretical model from previous research might suggest that the expansion of higher education leads to the advancement of democratic institutions, which in turn leads to higher economic growth. This chain of effects could be represented in our SEM. Previous studies can also help analysts identify important confounders of focal relationships, as well as potential intermediate variables that should be specified as such in the SEM. In our example, perhaps urbanization or population size should be included as potential confounders. Finally, past studies can also help researchers identify concepts that could be represented as latent variables and possible indicators of any latent variables.

Democracy is such a concept with several different approaches to measurement, all of which provide researchers with ideas for suitable indicators of democracy.

Second, in some cases, researchers move beyond previous studies and need to rely on their substantive expertise to guide the specification of a SEM. Substantive expertise could be used to propose and develop new measures of a latent variable or to treat a concept as latent in the first place. For example, such a process likely guided political scientists to recognize the value of treating democracy as a concept with multiple indicators and to develop different measures to capture the extent of democracy between countries and over time. In addition, substantive expertise could also lead researchers to consider alternative theoretical models, to account for potential confounders, or to explore possible mediators that have been overlooked in previous studies.

Third, specific aspects of the research design or the expected data for a study can provide important information for the specification of a model. For example, if an analyst expects to use experimental data to analyze the effect of an intervention on feelings of empathy (a latent variable), then consideration of potential confounders is less relevant due to randomization into treatment. Or, suppose that in our example with higher education, democracy, and economic growth, an analyst has access to longitudinal data. Such data could be used, for instance, to help establish temporal order by specifying a model with lags between the focal variables.

Finally, a general sense of logic or knowledge of common patterns in social and behavioral science can help researchers make decisions about model specification. For example, a measurement model that includes latent variables for democracy at two points in time could be based on a set of indicators developed by different sources at each time point. An analyst may decide to include correlations among the errors for indicators from the same source to reflect potential bias due to a given source (e.g., perhaps an organization consistently provides low ratings of democratic institutions from year to year). A general sense of logic can be particularly relevant when considering modifications to a SEM in light of evidence of poor fit with the data, a process that we examine in Chapter 4.

Specifying a SEM typically involves a multifaceted process informed by various sources. Previous studies often provide a foundational background, while substantive experience can aid in selecting the appropriate variables or including specific effects. The characteristics of the data and the research design can sometimes be leveraged in ways that inform the specification of SEMs, as can logic or knowledge of social and behavioral science patterns. A combination of these sources is sufficient in most research contexts to guide the specification of a model.

Measurement Model

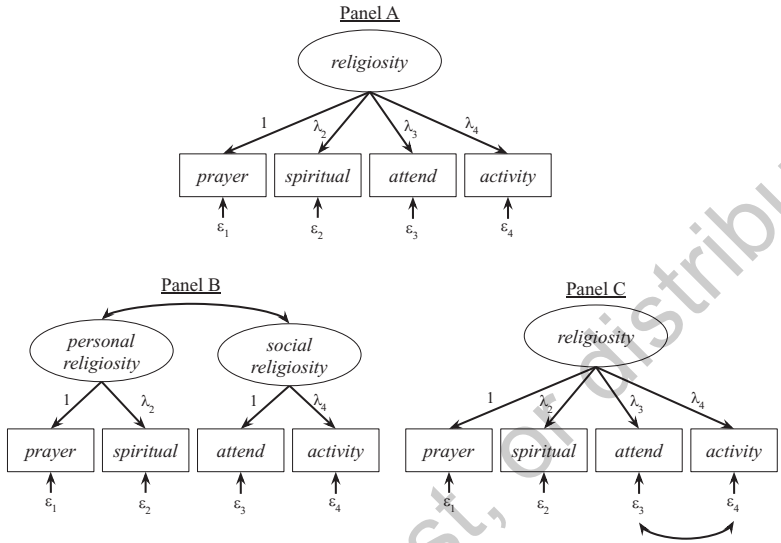
Although specifying a SEM is based on knowledge of the specific research context, it can be useful to have a sense of the structures of common models and the settings in which they arise. If an analyst anticipates that a given SEM will include latent variables, then it is often easiest to begin with the measurement component of the model. In this section, we discuss a few common forms of measurement models, and then in the following section, we turn to common forms of structural models.

In Chapter 1, we encountered an example that involved a latent variable representing psychological distress measured by six indicators. These indicators were designed to measure a one-dimensional concept, psychological distress, although we might question this, as we will see in Chapter 4. In some cases, analysts might find that indicators appear to represent more than one dimension of a concept. For example, suppose that an analyst is interested in developing a measurement model for religiosity and has four indicators—one measuring the importance of prayer, another measuring the importance of spirituality, a third measuring the frequency of attending services, and a fourth measuring the frequency of participating in other religious activities. The analyst may consider religiosity as a one-dimensional concept and specify a model with latent religiosity measured by all four indicators. Alternatively, the analyst may consider a measurement model with two dimensions for religiosity, one representing personal religiosity measured by the first two indicators and one representing social religiosity measured by the second two indicators.

Figure 2.1 illustrates the difference between a model specifying one latent variable (Panel A) and a model specifying two latent variables that reflect different dimensions of religiosity (Panel B). The SEM with latent variables for personal and social religiosity includes a correlation between the two dimensions of religiosity but, in contrast to the SEM in Panel A with only one dimension of religiosity, fixes an additional factor loading to 1. This is needed for model identification, which we discuss in the next chapter.

A measurement model with two dimensions of religiosity permits a more nuanced analysis in that each dimension might have different predictors or different relationships with an outcome of interest. However, it is possible that separating religiosity into two dimensions is inconsistent with a theoretical model or perhaps unnecessary for a focal analysis. In this case, an analyst may consider an alternative measurement model that maintains religiosity as a one-dimensional latent variable while recognizing that the indicators for attendance and activities have additional shared variance due to their social nature. For example, above and beyond the religious

Figure 2.1 Measurement models for religiosity.



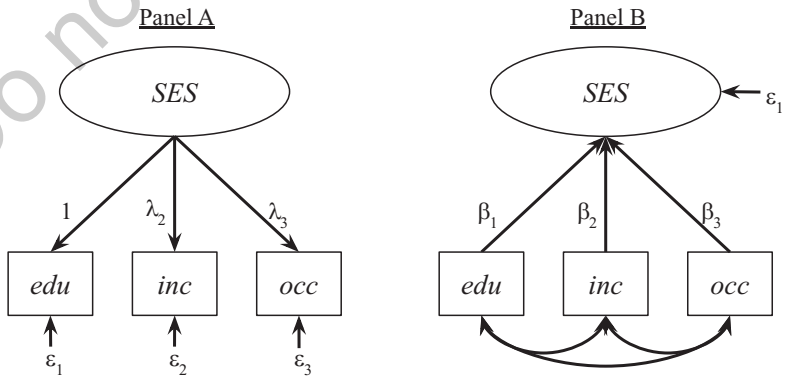
component, perhaps some people attend church and engage in other religious activities as a form of social engagement. In this case, we would expect the errors for the two indicators to contain this shared variance, and an analyst could include a correlation between the two errors (see Figure 2.1, Panel C). Correlated errors that reflect shared variance between indicators that is not accounted for by their common dependence on a latent variable can stem from a variety of sources (e.g., similar wording across survey items or a common data source for observed indicators, such as a firm providing democracy ratings).

Our discussion so far has been based on the perspective of a researcher making choices about model specification based on past work, substantive expertise, features of the research design, or general knowledge. However, in some situations, multiple model specifications may be equally plausible to the researcher. For example, perhaps all of the measurement models for religiosity represented in Panels A through C are reasonable in a specific research context. In these situations, rather than arbitrarily choosing one model specification, a researcher can, in some cases, turn to empirical tests of how well each candidate model fits with their data or whether a given model has a better fit with their data than an alternative model. All candidate measurement models for religiosity have sufficient degrees of freedom to test the overall model fit with a given source of data, which we cover in

detail in Chapter 4. In addition, it is possible to test the relative fit of the measurement model with correlated errors (Panel C) versus the model without the correlated errors (Panel A), and some comparative measures of model fit are available to help adjudicate between the model with two dimensions (Panel B) versus one dimension of religiosity (Panel A), which we also cover in detail in Chapter 4. Notably, it is not possible to empirically adjudicate between the two-dimensional measurement model (Panel B) and the correlated error measurement model (Panel C). These are referred to as *chi-square equivalent* (or *mathematically identical*) models and will have identical fit to a given data source. In these cases, an empirical test cannot differentiate the two models, which is important to keep in mind and underscores the role of theory and substantive knowledge in SEMs.

So far, our examples have involved measurement models in which the latent variable is a cause of the indicators. We would expect a person with higher religiosity to report praying more often, being more spiritual, and attending services and religious activities more frequently than a person with lower religiosity. We refer to indicators that are caused by latent variables as *effect* or *reflective* indicators. It is also possible for indicators to be causes of a latent variable. We refer to these types of indicators as *causal* or *formative* indicators. Consider socioeconomic status as a latent variable with three indicators—educational attainment, income, and occupational prestige. Figure 2.2 shows two potential measurement models for socioeconomic status. In Panel A, the three indicators are treated as effect indicators, each of which is affected by socioeconomic status. In contrast, Panel B shows a measurement model that treats the three indicators as causal

Figure 2.2 Measurement models for socioeconomic status.



indicators, with each being causes of socioeconomic status. There are several important differences between the two measurement models. In the measurement model with causal indicators (Panel B), the indicators are exogenous and therefore are treated as correlated with each other, as would be the case with predictors in a linear regression model. There are no errors associated with each indicator in this measurement model. In addition, the latent variable, socioeconomic status, is endogenous in the causal indicator measurement model and therefore has an error attached to it. Finally, as a reflection of the different relationships causal indicators have with a latent variable, we sometimes denote their coefficients as betas (β) rather than lambdas (λ) for factor loadings.

A thought experiment can be quite helpful in determining whether indicators are best considered effect or causal indicators. Imagine a change in a latent variable and consider whether you would expect a set of indicators to change simultaneously as well. In our religiosity example, this seems likely because a change in religiosity should result in a change in each indicator. This is less clear for socioeconomic status. A change in socioeconomic status does not necessarily simultaneously lead to a change in education, income, and occupational prestige. Rather, it seems much more likely that the opposite occurs. That is, a change in any one of the indicators (e.g., a higher level of education or a promotion leading to a higher level of income) leads to a change in socioeconomic status. Such a thought experiment suggests that socioeconomic status indicators are more likely to be appropriately treated as causal rather than effect indicators. In our examples, it is likely that all the indicators of religiosity are best treated as effect indicators, and all the indicators of socioeconomic status are best treated as causal indicators. However, measures of a latent variable may contain a mix of effect and causal indicators—see, for example, Perreira et al. (2005) for a measurement analysis of the indicators of depressive symptoms that comprise the Center for Epidemiologic Studies Depression, or CESD, scale.

SEMs involving latent variables measured by causal indicators can pose challenges with respect to model identification and interpretation, so analysts often strive to identify suitable effect indicators for latent variables or consider alternative measurement approaches (e.g., constructing an index based on a set of causal indicators). Despite these practical challenges, when developing a measurement model for any latent variables in a SEM, researchers should consider whether indicators are best treated as effect or causal indicators and specify the measurement component of the model accordingly.

In this section, we covered a few of the most common forms of measurement models that analysts may wish to consider when working out the

specification of the measurement component of a SEM. However, our discussion has not been exhaustive. There are many additional forms of measurement models tailored to specific analytic contexts—for example, higher-order models in which latent variables are themselves measured by other latent variables or multitrait, multimethod models in which multiple latent variables are measured by the same indicators (see, e.g., Roos & Bauldry, 2021). Furthermore, some models that are not typically thought of as measurement models fit into this framework (e.g., latent growth models for longitudinal data represent a form of a measurement model). The conclusion to this chapter includes additional resources for readers interested in extended treatments of measurement models.

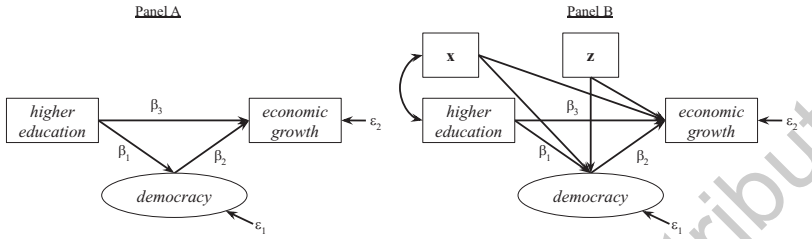
Structural Model

In the previous section, we discussed the choices surrounding the specification of the measurement component of a SEM that involves latent variables. In this section, we turn to consideration of choices related to the specification of the structural component of a SEM. As noted above, in any given research context, model specification derives from a combination of theory, substantive expertise, features of the research design and data, and general knowledge. Our discussion of the structural model is intended to provide a sense of the common considerations and options that researchers may encounter when specifying this component of a SEM.

Contemporary thinking about statistical analyses in general emphasizes connecting research questions to specific estimands before selecting an estimator and obtaining estimates (Lundberg et al., 2021). The same thinking applies to SEMs; in fact, the specification of a structural model brings the connection between research questions and estimands to the forefront. An *estimand* is a description of the quantity or parameter that will be estimated in a given study. It defines what a researcher aims to learn or infer from the data, regardless of the method used to estimate it or the data collected. In addition to a statement of the quantity or parameter, key components of an estimand may include a target population, a focal independent variable, a focal dependent variable, and a sense of whether the estimand is descriptive, predictive, or causal in nature. Some research questions may lead to estimands that include additional focal variables (e.g., moderators or mediators).

To see this in practice, we return to our example of higher education, democracy, and economic growth. An analyst may pose a research question regarding the role of democracy in the link between the expansion of higher education and economic growth. This research question involves mediation

Figure 2.3 Structural models for higher education, democracy, and economic growth.



with higher education as the focal independent variable, democracy as the focal mediator, and economic growth as the focal outcome (see Figure 2.3, Panel A). The primary estimand in this case is the indirect effect of higher education on economic growth through democracy, which is given by $\beta_1 \beta_2$. Further, suppose that the analyst is interested in obtaining a causal estimate of the indirect effect, at least as much as is possible in this setting with no intervention or other exogenous source of variation. As such, the analyst includes potential confounders of the relationships between higher education and democracy, higher education and economic growth, and democracy and economic growth in the SEM (see Figure 2.3, Panel B). To do so, an analyst needs to decide which variables represent different confounders, represented by the vectors **x** and **z** in the figure. Contemporary work on causal mediation analysis from the potential outcomes framework has brought to light a number of subtle issues present in this example, which we examine in Chapter 5.

Two other features of Panel B in Figure 2.3 merit attention. First, direct effects of both **x** and **z** on democracy and economic growth are not labeled with regression coefficients (i.e., βs). Like a regression model, when fitting this SEM to data, we obtain estimates for all direct effects. These are not labeled because they are not focal estimands based on the research question. Of course, it is important to examine these estimates to ensure that the relationships operate as anticipated or to uncover potential issues or unexpected findings. An analyst should be cautious in interpreting these direct effects in the same way as the estimand because the model has not been specified with any of the variables contained in **x** or **z** as focal variables. In the past, some researchers have been inclined to interpret all direct effects in a SEM as causal estimates, a critique that has been levied against regression models more generally (Westreich & Greenland, 2013). This has been a common source of criticism of applied work using SEMs from those advocating a contemporary perspective on causal inference.

The second feature of Panel B in Figure 2.3 that deserves attention is the lack of direct effects of higher education on the vector **z** and direct effects from the vector **x** to the vector **z**. We discuss the specific reasons behind these omissions in Chapter 5 with an extended consideration of mediation. For our current purposes, a more general point is that it is the lack of a direct effect that reflects a causal assumption in the specification of a SEM. All paths, whether direct effects or correlations, in a SEM represent potential relationships. When a SEM is fit to data, any given path may turn out to be within sampling error of 0, but that is not something the analyst decided in advance. In contrast, the absence of a direct effect or correlation between two variables reflects a strong causal assumption that the relationship is, in fact, 0. It is these assumptions that "buy" degrees of freedom to test the overall fit of a SEM with a given dataset. As we will discuss in Chapter 4, an analyst might obtain an overall model fit statistic that suggests poor fit with the data, which then calls into question the causal assumptions underlying the specification of the model. Unfortunately, the opposite conclusion is not valid—a SEM that exhibits strong fit with a given dataset does not provide proof that the causal assumptions underlying the model specification are valid. This is because it is always possible that there are models based on alternative causal assumptions that exhibit equally good or even better fit with a given dataset.

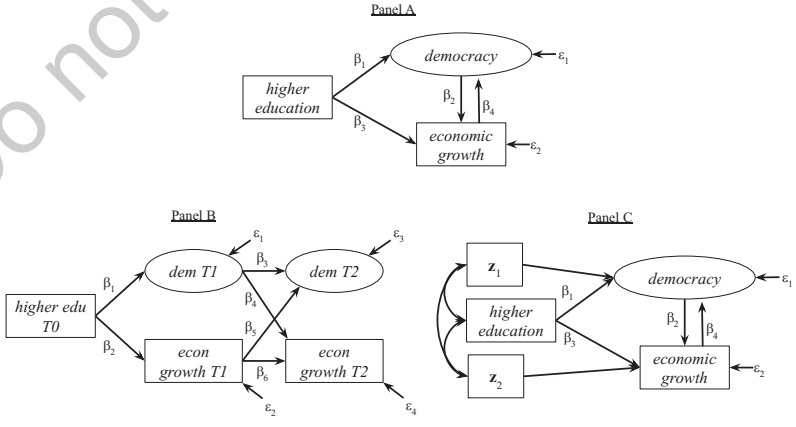
At this point, one might reasonably wonder whether it is possible to specify a SEM without any causal assumptions—that is, a SEM that specifies all possible paths among a set of variables. With the exception of the trivial model that specifies no latent variables and merely correlates all observed variables with each other, the answer is "no." This impossibility stems from a fundamental constraint: Specifying all possible direct effects among variables quickly exhausts the available degrees of freedom, leaving insufficient information to obtain unique parameter estimates. Consider a simple example with three observed variables, x, y, and z, each directly affecting the others. This model would require estimating six regression coefficients, six intercepts, and three error variances—15 parameters in total—but provides only 9 degrees of freedom (we discuss where this calculation comes from in Chapter 3). This mathematical limitation is why causal assumptions, which constrain certain paths to be 0, are essential to SEM specification.

For our last set of examples of structural components of SEMs, we return to the relationships between higher education, democracy, and economic growth. Suppose that a researcher believes that democracy and economic growth have a reciprocal relationship: Improvements in democratic governance might lead to higher economic growth, but higher economic growth might also lead to improvements in democratic governance. Reciprocal relationships or reverse causation are frequently encountered in some areas

of social and behavioral science. Figure 2.4, Panel A illustrates this model. To the extent that this model is an accurate reflection of reality, simply ignoring the effect of economic growth on democracy will lead to a biased estimate of the effect of democracy on economic growth (and potentially other parameter estimates, as we will see in Chapter 3). However, the SEM illustrated in Panel A with reciprocal effects is not identified, so it is not possible to obtain unique parameter estimates.

One option for a path forward is to incorporate temporal information into the model. If a researcher has access to longitudinal data with repeated measures of the indicators of democracy and economic growth, then it is possible to specify a SEM in which democracy at Time 1 (T1) influences economic growth at Time 2 (T2) and vice versa. This model, sometimes referred to as a cross-lagged panel model, is illustrated in Figure 2.4, Panel B. In this model specification, we include direct effects from democracy at T1 to democracy at T2 and similarly for economic growth from T1 to T2. This alters our interpretation of the effect of democracy on economic growth, as it is now net of economic growth at T1. This is sometimes usefully thought of in terms of the effect of democracy on the change in economic growth from T1 to T2. The important thing to keep in mind is how this shift relates to the guiding research question and estimand, which may need to be adjusted to reflect the change in the research design and the analytic strategy (i.e., the use of longitudinal data to help address the reciprocal effects). In addition, this model assumes that there are no contemporaneous effects of democracy on economic growth and vice versa.

Figure 2.4 Alternative models for higher education, democracy, and economic growth.



Two additional observations about the model in Figure 2.4, Panel B are in order. First, the cross-lagged panel model is not the only way to leverage longitudinal data in a SEM framework. Depending on the research questions and the specific analytic concerns, analysts may prefer to use longitudinal data to identify trajectories of democracy and economic growth and how these trajectories relate to each other. Alternatively, analysts could instead specify a fixed-effects model that addresses potential unobserved country-level confounders of the relationship between democracy and economic growth. Second, this general model structure can appear in other contexts that do not necessarily involve longitudinal data. One prominent area is analyses involving dyadic effects, such as between spouses or romantic partners. For example, a similar SEM could be specified to analyze the effects of husbands' perceptions of relationship quality on wives' depression and vice versa. The conclusion of this chapter provides resources for readers interested in specifying SEMs with longitudinal or dyadic data.

Figure 2.4, Panel C presents another option for specifying a model involving a reciprocal relationship. In this case, a researcher maintains the contemporaneous reciprocal effect between democracy and economic growth but brings additional variables, $\mathbf{z}_1$ and $\mathbf{z}_2$, to the model. In particular, $\mathbf{z}_1$ represents a vector of covariates (or possibly a single covariate) that have direct effects on democracy and no effects on economic growth. Similarly, $\mathbf{z}_2$ represents a vector of covariates that have direct effects on economic growth and no effects on democracy. The absence of direct effects on democracy and economic growth, respectively, represent causal assumptions. In this context, $\mathbf{z}_1$ and $\mathbf{z}_2$ are instrumental variables. Instrumental variables are variables that affect an outcome only through their effect on a causal variable or exposure of interest, which is referred to as the exclusion criteria or the exogeneity requirement (Hansen, 2022). In our example, the lack of direct effects from $\mathbf{z}_1$ to economic growth makes them suitable instruments for estimating the effect of democracy on economic growth independent of the reciprocal effect. The same holds for $\mathbf{z}_2$ in reverse. Instrumental variables allow us to estimate the reciprocal effects between democracy and economic growth; however, it can be quite difficult to identify credible instrumental variables that meet the exogeneity requirement. In addition, the shift to instrumental variables has implications for the estimand. The use of instrumental variables can alter the target population, as the parameter estimates obtained from an instrumental variable analysis typically reflect the population of *compliers*, that is, units for which the instrumental variables induce variation (see Angrist and Pischke [2009] for a discussion).

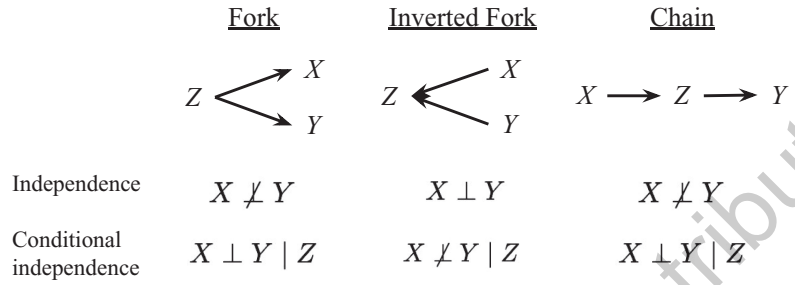
In this section, we examined a variety of forms of structural models and a couple of cases of how they can be adapted to different data structures or research designs. The two key things that researchers should keep in mind when specifying a SEM are (1) that the model reflects the research question(s) and estimate(s) guiding the study and (2) that the absence of effects or paths captures the causal assumptions underlying the specified model. Researchers should be mindful to avoid placing too much emphasis on parameter estimates that are unrelated to specific estimands and assuming that a SEM specified based on causal assumptions provides causal estimates of all of the parameters. Whether any parameter estimates from fitting a SEM to a given dataset merit a causal interpretation depends on features of the research design (e.g., a random assignment study) or additional assumptions (e.g., all confounders have been included in the model). In the next section, we compare SEMs and directed acyclic graphs (DAGs), an alternative graphical syntax for statistical models. This section can be skipped for readers interested in moving on to the next chapter on the identification of SEMs and the estimation of model parameters.

*SEMs and DAGs

In this section, we explore similarities and differences between SEMs and DAGs, an alternative graphical method for representing statistical models that has become increasingly popular over the past few decades. DAGs rely on a similar graphical syntax as SEMs, with nodes representing variables, directional edges (single-headed arrows) representing direct causal effects, and bidirectional edges (double-headed arrows) representing a spurious or noncausal association between two variables. Edges rendered as dashed lines typically indicate that the origin of the edge is an unobserved variable. The "acyclic" component of DAGs refers to the fact that the models do not contain causal loops, feedback effects, or reciprocal effects. Finally, DAGs are nonparametric in the sense that they do not make any distributional assumptions about the variables represented in the model or any assumptions about the functional form of the relationships between variables.

DAGs were developed to assist analysts in determining causal claims that could be justified based on a set of observed variables and assumed causal relationships among the variables. Potential causal claims can be evaluated by investigating a DAG with a few elementary structures in mind. Figure 2.5 illustrates three elementary structures with causal implications: the fork, the inverted fork, and the chain. A key insight of DAGs is that each of these structures has implications for the independence of two focal variables, X and Y, depending on whether one conditions on a third

Figure 2.5 DAG elementary structures.



variable, *Z*. With the fork, *Z* has direct effects on *X* and *Y*, and there is no direct effect from *X* to *Y*. In this case, *X* and *Y* will be associated with each other (i.e., they will not be independent) unless one conditions on *Z*. This is analogous to the presence of a confounder of a focal relationship that needs to be included in a model to obtain an unbiased estimate of the relationship between *X* and *Y*. The opposite pattern holds with the inverted fork—*X* and *Y* will be independent unless one conditions on *Z*. In this context, *Z* is often referred to as a *collider*. Understanding the importance of colliders and their connection to selection bias has been one of the most important contributions of DAGs (Elwert & Winship, 2014). Finally, the chain operates much like the fork with a lack of independence of *X* and *Y* unless one conditions on *Z*. This pattern is consistent with the presence of a mediator of the effect of *X* on *Y*. Algorithms based on identifying these elementary structures in a given DAG can identify minimal sets of covariates, if any exist, that should be conditioned on in an analysis to justify a causal interpretation. The specification of DAGs relies on the same sources as SEMs, and the conclusions regarding causal claims are only valid to the extent that the DAG captures all relevant variables and relationships in the analytic context.

Several initial differences between SEMs and DAGs are apparent in Figure 2.5. First, in representing variables as nodes, DAGs typically do not distinguish between latent and observed variables. Measurement error is sometimes represented in DAGs, but this is not usually the case. Second, errors are usually not included in DAGs. Rather, DAGs typically assume that all variables contain random error and that endogenous variables have implied errors. Any correlations among errors are explicitly represented, if relevant, and often are depicted as emanating from a specific source. Beyond what is apparent in the figure of elementary structures, as noted

above, DAGs are nonparametric. SEMs, by contrast, make distributional assumptions (typically multivariate normality) and assume linear relationships among variables. However, as we will see in Chapters 3 and 6, both of these assumptions can be relaxed. There are robust estimators and estimators for categorical endogenous variables for SEMs that do not assume multivariate normality. In addition, analysts can specify models with a variety of nonlinear relationships, including interactions among variables. Finally, the most significant difference between DAGs and SEMs lies in their role in the analytic process. SEMs are specified with parameters that can be estimated when fit to data. DAGs are not statistical models in the sense that they cannot be fit to data—they do not include specific parameters to be estimated. Rather, DAGs serve as guides to help analysts make decisions about data collection, model specification, and the types of causal claims that can be justified in a given analytic context. As such, DAGs sometimes represent features of a research design (e.g., random assignment to treatment, a node depicting selection into a sample, or a node representing a missing data mechanism) that are rarely represented in SEMs.

In sum, while SEMs and DAGs share similarities in their graphical representation of relationships among variables, they also have notable differences in their assumptions, parameterization, and role in the analytic process. SEMs are statistical models that can be fit to data, while DAGs serve as nonparametric guides for making decisions about research design, model specification, and justifiable causal claims. Despite these differences, researchers using SEMs can benefit from the insights provided by DAGs, particularly when grappling with subtle aspects of model specification and engaging in causal inference.

Conclusion

Typically, the first step in an analysis involving SEMs is the specification of a focal model or set of candidate models. In this chapter, we examined sources of information that can guide model specification, a few common forms of measurement and structural models, and some of the key decisions and choices that analysts may encounter. In addition, we discussed the importance of ensuring clear links between research questions, estimands, and model specification. In the next chapter, we discuss the identification status of a SEM and fitting the SEM to data to obtain parameter estimates.

For readers interested in additional coverage of the topics covered in this chapter, several resources offer valuable insights and further exploration. Roos and Bauldry's (2021) *Confirmatory Factor Analysis* provides an extended treatment of measurement models with numerous examples of

different measurement structures. For those interested in working with causal indicators, Bollen and Bauldry (2011) offer consideration and guidance surrounding models in which indicators are viewed as causes rather than effects of latent variables.

Some examples in this chapter involved applying SEMs to longitudinal or dyadic data. Preacher et al. (2008) provide an accessible introduction to latent growth models, Bollen and Curran (2006) provide a more advanced treatment, and Bollen and Brand (2010) offer an overview of the relationships between different types of models applied to longitudinal data. *Dyadic Data Analysis* by Kenny et al. (2006) offers an introduction to unique considerations and methods involved in applying SEMs to dyadic data.

Finally, for readers interested in learning more about DAGs, *Causal Inference in Statistics: A Primer* by Pearl et al. (2016) provides an overview. In addition, Rohrer (2018) offers a particularly accessible article-length introduction to the core ideas of DAGs. Pearl's (2009) classic *Causality* remains the best advanced text on DAGs.