CHAPTER 2

Key Advances in the History of Structural Equation Modeling

Ross L. Matsueda

Structural equation modeling (SEM) has advanced considerably in the social sciences. The direction of advances has varied by the substantive problems faced by individual disciplines. For example, path analysis developed to model inheritance in population genetics, and later to model status attainment in sociology. Factor analysis developed in psychology to explore the structure of intelligence, and simultaneous equation models developed in economics to examine supply and demand.

These largely discipline-specific advances came together in the early 1970s to create a multidisciplinary approach to SEM. Later, during the 1980s, responding to criticisms of SEM for failing to meet assumptions implied by maximum likelihood estimation and testing, SEM proponents responded with estimators for data that departed from multivariate normality, and for modeling categorical, ordinal, and limited dependent variables. More recently, advances in SEM have incorporated additional statistical models (growth models, latent class growth models, generalized linear models, and multilevel models), drawn upon artificial intelligence research to attempt to "discover" causal structures, and finally, returned to the question of causality with formal methods for specifying assumptions necessary for inferring causality with nonexperimental data.

In this chapter, I trace the key advances in the history of SEM. I focus on the early history and try to convey the excitement of major developments in each discipline, culminating with cross-disciplinary integration in the 1970s. I then discuss advances in estimating models from data that depart from the usual assumptions of linearity, normality, and continuous distributions. I conclude with brief treatments of more recent advances to provide introductions to advanced chapters in this volume.

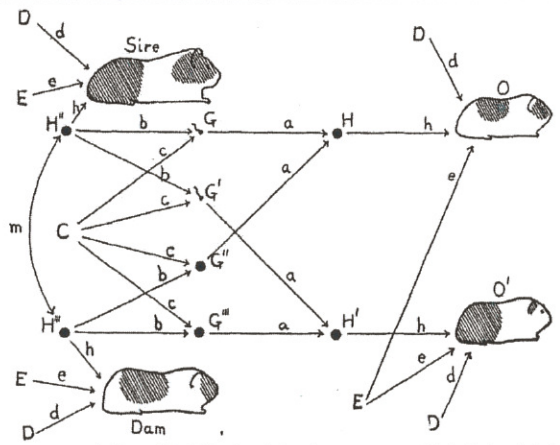
EARLY HISTORY: THE DISCIPLINARY ROOTS OF SEM

Sewall Wright's Path Analysis in Genetics and Biology

In 1918, Sewall Wright, a young geneticist, published the first application of path analysis, which modeled the bone size of rabbits. After computing all possible partial correlations of his measures, he was still dissatisfied with the results, which remained far from a causal explanation. Consequently, Wright developed path analysis to impose a causal structure, with structural coefficients on the observed correlations. His substantive application decomposed the variation in the size of

an individual bone to various hereditary causes (Hill, 1995). He subsequently applied path analysis to systems of mating, using data on guinea pigs, which laid the basis for much of subsequent population genetics. For example, in modeling the proportion of white color in spotted guinea pigs, Wright (1920) decomposed the variance into heredity (h), common environment for the litter (e), and other factors, such as developmental noise (d). The path coefficient (h) represents the link between genotype and phenotype, and h^2 is the proportion of variance due to heredity, later termed "heritability" in population genetics. Wright also developed models for systems of mating, showing the consequences of continued inbreeding systems, such as continued brother–sister mating, which results in $m = r'_{00}$, where m is the correlation between mates in one generation, and r'_{00} is the correlation between brother and sister of the previous generation (Li, 1956). He also derived results for intergenerational consequences of assortative mating. Figure 2.1 reproduces a path diagram of environment and heredity, which Hill (1995, p. 1500) calls "surely one of the best known diagrams in biological science."

Wright (1921a, 1934) presented the method of path analysis for estimating causal relations among variables based on the correlation matrix of observed variables, emphasizing path coefficients (standardized regression coefficients) but also using "path regressions" (unstandardized coefficients). He invented a graphical method of presenting causal relations using path diagrams, consisting of variable labels connected by arrows for direct effects, double-headed arrows for unanalyzed correlations, and the estimated path coefficients listed over single-headed arrows. From path diagrams, Wright could read off total, direct, and indirect effects, and quickly decompose correlations into various causal sources, such as direct effects, indirect effects, common causes, and the like. Among the models Wright estimated by path analysis was a model of multiple causal indicators, or what later became known as the multiple-indicator, multiple-indicator-cause (MIMIC) model. Wright's estimation method was essentially the method of moments, which follows the intuitive principle of estimating a population moment (or function of moments) using the sample analog moment (or function of moments) (Goldberger, 1991). Although he lacked a principle for reconciling multiple ways of expressing a path coefficient in terms of sample moments in overidentified models, he did check to see if they were close, and acknowledged the potential gains in efficiency and reduced standard errors from using full information (Goldberger, 1972b).



A diagram illustrating the relations between two mated individuals and their progeny. H, H', H'' and H''' are the genetic constitutions of the four individuals. G, G', G'' and G''' are four germ-cells. E and D represent tangible external conditions and chance irregularities as factors in development. C represents chance at segregation as a factor in determining the composition of the germ-cells. Path coefficients are represented by small letters.

FIGURE 2.1. An early path diagram on the importance of heredity and environment in spotted guinea pigs. From Wright (1921b). Copyright granted by the Genetics Society of America. Reprinted by permission.

While working for the U.S. Department of Agriculture, Wright (1925) worked on corn and hog correlations, developing a complex, highly overidentified, recursive system of equations containing observed, unobserved, lagged, and unlagged variables to describe seasonal data on hog breeding, corn prices, and hog prices. The Department of Agriculture rejected publication of the monograph on the grounds that "an animal husbandman" (Wright's position at the time) "had no business writing about economics" (Crow, 1988). Wright's research was only published after Henry Wallace read the paper and exploited the influence of his father, then Secretary of Agriculture. Although the recursive model had no explicit demand function, Wright (1925, p. 54) noted in a footnote that a direct negative effect of hog quantity on hog price would be desirable but the "treatment of such reciprocal relations between variables requires an extension of the theory of path coefficients" (see Goldberger, 1972b, p. 983).

In 1928, Wright's father, economist Phillip Wright, published a study of the tariff, which included an appendix—the infamous "Appendix B"—that applied instrumental variables and path analysis to reciprocal re-

lations between variables (Wright, 1928). Most scholars presume that the appendix was coauthored with Sewall Wright, although Stock and Trebbi (2003) suggest that Phillip may have been sole author.[1] Phillip Wright, who had mentioned the identification problem as early as 1915 (Wright, 1915), and presented it in graphical form of supply and demand curves, wrote out what later became known as the reduced form equations:

$$P = p_1 \frac{\sigma_P}{\sigma_D} D + p_2 \frac{\sigma_P}{\sigma_S} S \quad (2.1)$$

$$Q = q_1 \frac{\sigma_Q}{\sigma_D} D + q_2 \frac{\sigma_Q}{\sigma_S} S \quad (2.2)$$

where D and S indicate shifts in demand and supply curves after transforming P and Q to trend ratios, and σ's represent standard deviations of variables. Wright (1928) noted that the two equations contain four unknowns. He then suggested that if two external variables, A and B, could be found (based on external knowledge of markets) such that A were correlated with D but not S, and B were correlated with S but not D, the principles of path analysis would yield solutions for the four unknowns. Wright had arrived at a more general treatment than offered by Henry Schultz that year (1928) or "indeed in 1938" (Goldberger, 1972b, p. 984). Sewell Wright (1934) later developed more general solutions, noting that a mere single external variable is sufficient if the supply and demand situations were uncorrelated—that is, the disturbances of the two equations were orthogonal—and also that, in very complex models with many more shifts than external variables, one could solve for parameters by assuming plausible values for some of the unknowns (see Epstein, 1987). This solution to the simultaneous equation problem would be rediscovered by Rejersøl (1945, cited in Morgan, 1990), who used the term "instrumental variable estimates" (which he attributed to Frisch).

At this time, Wright's (1934) path analysis was largely ignored not only in biology but statistics as well, perhaps in part because it contained elements "that were objectionable" to the two dominant schools of statistics (Shipley, 2000, p. 70):

> The Phenomenalist school of Pearson disliked Wright's notion that one *should* distinguish "causes" from correlations. The Realist school of Fisher disliked Wright's notion that one *could* study causes by looking at correlations. Professional statisticians therefore ignored it.

And applied biologists were drawn to Fisher's methods, which included inferential statistics, were rooted in experimental design, and were easier to understand (Shipley, 2000). Later, Wright's path models became foundational for much of population genetics (Li, 1956).

Causal Models and Path Analysis in Sociology

Perhaps the earliest reference to path analysis by a sociologist appeared in an appendix to Blalock's (1961a, pp. 191–193) classic treatment of causal inference in nonexperimental research, where he briefly discussed "some related approaches," and concludes that path analysis "might readily be extended so as to be highly useful in the less precise nonexperimental sciences."[2] Blalock had spearheaded a voluminous literature in sociology on causal models in the 1960s by elaborating on Simon's (1954) method of making causal inferences from correlational data. Building on Yule's (1896) original method of ruling out common causes using partial correlations, Simon (1954) began by noting that a zero partial correlation ($r_{xy.z} = 0$) between independent variable x and dependent variable y holding z constant, implies a *spurious correlation* between x and y when z (a confounder) is causally prior to x and y. However, when z is causally prior to x and causally subsequent to y, z is an *intervening variable* between x and y. Simon (1954, p. 471) correctly asserted that the determination of whether a partial correlation is or is not spurious "can only be reached if a priori assumptions are made that certain other causal relations do not hold among the variables"—namely, exclusionary restrictions on coefficients and uncorrelated errors in equations. He then went on to expound on these conditions in all possible three-variable models. Blalock (1961b, 1962) extended this method to a five-variable model and then to an exhaustive exposition of the four-variable case. Later, Boudon (1965) applied these results to path coefficients, drawing from Wright (1934), but renaming them "dependence coefficients."

But it was Duncan's (1966) classic expository article, "Path Analysis: Sociological Examples," and his monumental monograph with Peter Blau, *The American Occupational Structure* (Blau & Duncan, 1967), that launched the path analysis movement in sociology, and later in political science. Duncan used published correlations to apply path analysis to recursive models of class values, population density, occupational prestige, and synthetic cohorts. Subsequently, Costner (1969) and

Blalock (1969) used path analysis to address multiple-indicator causal models, and in particular, to find that tetrad-difference restrictions on observed correlations provide a way of testing models—a result that emerged from factor analysis in psychology. These papers made important insights into substantive population models, although they tended to gloss over sample–population distinctions, efficient estimation, and formal hypothesis testing.

Substantively, Duncan and Hodge (1963) had earlier published a four-variable status attainment model on data from Chicago, in which son's education intervened between the effects of father's occupation on son's occupation in 1940 and 1950. They estimated the models using equation-by-equation multiple regression. Blau and Duncan's (1967) monograph expanded on this model by using national data, distinguishing hierarchies of occupations and rewards from the process by which individual characteristics sort people into those occupations, and examining whether the American stratification system approximated an open opportunity structure more than a rigid class hierarchy. The study produced an explosion of productive research using path analysis to model status attainment, most notably the Wisconsin status attainment model (e.g., Sewell & Hauser, 1975).

Factor Analysis in Psychology

In psychology, interest in SEM originated in "factor analysis," which is a statistical method for analyzing a correlation or covariance matrix of observed variables to identify a small number of factors, components, or latent variables that contain the essential information in the original variables. Thus, the primary goal is to attain "scientific parsimony or economy of description" (Harman, 1960, p. 4). The method was originally developed to model psychological theories of ability and behavior. Spearman (1904) is often credited as the founding father of factor analysis, although earlier Pearson (1901) published a paper on fitting planes by orthogonal least squares—the foundation for principal component analysis—which was later applied to the analysis of correlation matrices by Hotelling (1933). Spearman's work on factor analysis derived explicitly from his work on intelligence testing. He specified a two-factor theory of intelligence, in which all mental processes involved a general factor *g*, plus a specific factor *s*. The general factor enters all activities, some more than others, while the specific factors were unique to the task at hand (the specific mental activity). Spearman claimed that the disparate items from intelligence tests would reveal two factors: a general factor and an item-specific factor. Moreover, Spearman (1927) showed that four variables cannot be described by a single factor unless their intercorrelations satisfy the conditions of two vanishing tetrad differences[3]:

$$r_{12}r_{34} - r_{12}r_{23} = 0 \quad \textbf{(2.3)}$$

$$r_{12}r_{24} - r_{14}r_{23} = 0 \quad \textbf{(2.4)}$$

Criticisms of the two-factor theory of intelligence on theoretical as well as empirical grounds—tetrads often failed to vanish or, equivalently, correlation matrices failed to be of unit-rank, even after considering sampling error—led to interest in multiple factor analysis, in which group factors were identified after extracting a general factor (e.g., Holzinger, 1941). Thurstone (1935), who founded the Psychometric Society, noted that a vanishing tetrad difference implied a vanishing second-order determinant of the matrix of observables, and extended this to the vanishing of higher-order determinants as a condition for more than one factor. He then generalized the result: The number of common factors is determined by the rank of the matrix of observables (see Harman, 1960). In addition, Thurstone (1935) developed the centroid method of factoring a correlation matrix (as a pragmatic compromise to the computationally burdensome principal axis method). Moreover, he developed a definition of simple structure for factor analysis based on five principles (the most important of which are to minimize negative loadings and maximize zero loadings) to facilitate interpretation and ensure that loadings were invariant to the inclusion of other items. This spawned interest in various methods of rotation of the initial solution, such as Kaiser's (1958) Varimax orthogonal rotation. Thurstone's original hand rotation was oblique, allowing factors to be correlated, but it was Jennrich and Sampson (1966) who developed a computational method of achieving an oblique rotation, and Jennrich and Clarkson (1980) who worked out the standard errors of rotated loadings (see Browne, 2000).

The problem of rotating factor solutions is avoided when confirmatory factor analysis is used. Here, the number of factors and the pattern of loadings—including restrictions on loadings—are specified in advance, transforming the problem into one of identification of a model's parameters from observed moments—the same issue that arises in simultaneous equation models.[4] The factor model specifies $\boldsymbol{y} = \boldsymbol{\Lambda\eta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{y}$ is a vector of

p observables, $\boldsymbol{\eta}$ is a vector of m latent factors, where ($m < p$), $\boldsymbol{\Lambda}$ is a $p \times m$ matrix of loadings, and $\boldsymbol{\varepsilon}$ is a vector of p error terms representing "unique" variance in y. Identification is typically achieved by specifying zero-restrictions on elements of $\boldsymbol{\Lambda}$ to create, for example, sets of congeneric tests, in which items load solely on single factors (e.g., Jöreskog, 1971b). The zero loadings create tetrad difference overidentifying restrictions on observed covariances, as noted earlier. The covariance structure then becomes

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda} + \boldsymbol{\Theta} \tag{2.5}$$

where $\boldsymbol{\Sigma} = E(\mathbf{y}\,\mathbf{y}')$, $\boldsymbol{\Psi} = E(\boldsymbol{\eta}\,\boldsymbol{\eta}')$, and $\boldsymbol{\Theta} = E(\boldsymbol{\varepsilon}\,\boldsymbol{\varepsilon}')$, and $E(\boldsymbol{\varepsilon}) = 0$. A maximum likelihood approach to factor analysis was developed by Lawley (1940), and fully elaborated by Anderson and Rubin (1956). But, according to Browne (2000, p. 663), computational procedures were not available until "nested algorithms involving eigenvalues and eigenvectors and imposing inequality constraints on unique variance estimates were discovered independently by Jöreskog (1967) and by Jennrich and Robinson (1969)." If $\boldsymbol{S}$, the covariance matrix of observables follows a Wishart distribution, the log-likelihood function of the model is

$$\log L = -\frac{1}{2}n\left[\log|\boldsymbol{\Sigma}| + tr\left(\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\right)\right] \tag{2.6}$$

Jöreskog (1967) and his colleagues developed computer software programs for confirmatory factor analysis estimated by maximum likelihood.

Simultaneous Equation and Errors-in-Variables Models in Economics

The structural equation approach in econometrics is usually attributed to Haavelmo (1943) and the Cowles Commission (1952), most notably Koopmans (1945). But, as Morgan (1990) points out, Frisch and Waugh (1933, pp. 389–390) were first to define "*structural* relation" as a "theoretical relation postulated *a priori*" in a single-equation multivariate linear model in which the partial regression coefficient represented a "structural coefficient": "An empirically determined relation is 'true' if it approximates fairly well a certain well-defined *theoretical* relationship, assumed to represent the nature of the phenomenon studied."

Frisch (1934), however, was critical of the use of probability models for economic data (e.g., variations in the business cycle), which were rarely the result of a sampling process, and of ordinary least squares (OLS) regression because measurement errors existed on not only dependent variables but also independent variables. This led him to confluence analysis, which treated observed variables as fallible indicators of latent variables, and then examined the interrelationships among all latent and observed variables to distinguish "true relations" from "confluent relations." Frisch developed the method of "bunch maps"—a graphical presentation of regression coefficients—as a tool to discover underlying structure, often obtaining approximate bounds for relationships (for details, see Hendry & Morgan, 1989).

According to Qin (1989), Frisch—who coined the term "econometrics" and helped found the Econometric Society and its journal *Econometrica*—had developed many of the abstract principles of identification of simultaneous equation models, although in a manner confounded with issues of estimation and testing, particularly in his critique of Tinbergen (1939). Tinbergen himself had discussed a formal way of identifying a two-equation model from reduced-form parameters (Tinbergen, 1930, cited in Magnus & Morgan, 1987), although in his monumental models of the Dutch and U.S. economies, he "cleverly constructed his model in the causal chain fashion," using OLS to estimate its parameters, including effects of lagged dependent variables (Anderson, 1991).[5] In his classic works on demand, Schultz (1938) had developed the cobweb model in which lagged price identified the supply–demand model. Remarkably, Schultz was unaware of Sewell Wright's more general instrumental variable solution to the identification problem despite the two participating in intellectual discussions of science, mathematics, and statistics within a hiking group of academics (Goldberger, 1972b, pp. 985–986).

Within this context, Haavelmo (1943, 1944) made two key contributions to structural equation models in economics. First, he built on the work of Wald, Koopmans (1937), and others in specifying a probability model for econometric models, presenting clearly and concisely the Neyman–Pearson (e.g., Neyman & Pearson, 1933) approach to hypothesis testing, and using the probability approach for estimation, testing, and forecasting (see Morgan, 1990). He also distinguished between two models of the source of stochastic components: errors-in-variables models emphasized by Frisch (1934), and random shocks models introduced by Slutsky (1937).[6] This framework is often referred to as the "probabilistic revolution" in econometrics (see Morgan, 1990) and has had a lasting impact on the field, particularly in cementing the Neyman–Pearson

approach to inference over others, such as Bayesian approaches (e.g., Jeffreys, 1935; see Heckman, 1992). Second, Haavelmo made major advances in simultaneous equation models, showing that OLS estimates are biased in a two-equation supply–demand model, and distinguishing between structural form equations and what Mann and Wald (1943) termed the "reduced-form equation." He applied maximum likelihood estimation to the system of equations, showing its equivalence to OLS when applied to the reduced form, and specifying necessary and sufficient conditions for identification in terms of partial derivatives of the likelihood function (Haavelmo, 1943, 1944). Haavelmo (1944) also refined the term "autonomy": Parameters representing *relatively* autonomous relations are more likely to be stable, intelligible, and useful for policy analysis (Aldrich, 1989). Parameters, then, are structural when they represent autonomous relations, which are invariant to policy interventions. Haavelmo (1943) also interpreted structural equations in terms of counterfactuals or potential outcomes, presaging the more recent models of Rubin (1974) and Imbens and Angrist (1994).

The advances made by Haavelmo and Mann and Wald led to work on the simultaneous equation model at the Cowles Commission, which moved to Chicago in 1939, led by Marschak and including Koopmans, Haavelmo, Wald, Lurwicz, Klein, and Anderson (Rubin and Leipnik were graduate assistants in Mathematics, and Simon joined later). Work at the Cowles Commission solved the major problems of identification, estimation, and testing of simultaneous equation models. In particular, Koopmans, Rubin, and Leipnik (1950) gave a general treatment of the model's structural and reduced forms:

$$\mathbf{B}\boldsymbol{y} + \boldsymbol{\Gamma}\boldsymbol{x} = \boldsymbol{u} \tag{2.7}$$

where $\boldsymbol{y}$ is a vector of p endogenous variables, $\boldsymbol{x}$ is a vector of q predetermined or exogenous variables, $\boldsymbol{u}$ is a vector of p disturbances (assumed normally distributed), and $\boldsymbol{\Gamma}$ and $\mathbf{B}$ are coefficient matrices in which $\mathbf{B}$ is nonsingular.[7] The reduced form is

$$\boldsymbol{y} = \boldsymbol{\Pi}\boldsymbol{x} + \boldsymbol{v} \tag{2.8}$$

where $\boldsymbol{\Gamma} = -\mathbf{B}\boldsymbol{\Pi}$, $\boldsymbol{u} = \mathbf{B}\boldsymbol{v}$, $\boldsymbol{\Sigma} = \mathbf{B}\boldsymbol{\Omega}\mathbf{B}'$, and $\boldsymbol{\Omega}$ is the covariance matrix of $\boldsymbol{v}$.

Anderson (1991) summarizes an intuitive way of stating the identification problem. Suppose that in Equation 2.7 some elements of $\mathbf{B}$ and $\boldsymbol{\Gamma}$ are constrained to be zero. If we rearrange the matrices so that the first row of $(\mathbf{B}, \boldsymbol{\Gamma})$ is written as $(\boldsymbol{\beta}, \mathbf{0}, \boldsymbol{\gamma}, \mathbf{0})$, then the first row of $\boldsymbol{\Gamma} = -\mathbf{B}\boldsymbol{\Pi}$ becomes $(\boldsymbol{\beta}, \mathbf{0})\boldsymbol{\Pi} = -(\boldsymbol{\gamma}, \mathbf{0})$. Then partition $(\boldsymbol{\Pi})$:

$$\boldsymbol{\Pi} = \begin{pmatrix} \boldsymbol{\Pi}_{11} & \boldsymbol{\Pi}_{12} \\ \boldsymbol{\Pi}_{21} & \boldsymbol{\Pi}_{22} \end{pmatrix} \tag{2.9}$$

and we obtain $\boldsymbol{\beta}\boldsymbol{\Pi}_{11} = -\boldsymbol{\gamma}$, and

$$\boldsymbol{\beta}\boldsymbol{\Pi}_{12} = \mathbf{0} \tag{2.10}$$

The vector $\boldsymbol{\beta}$ is identified (except for a multiplicative constant) by Equation 2.10 if and only if the rank of $\boldsymbol{\Pi}_{12}$ is at least one less than the number of elements in $\boldsymbol{\beta}$ (Anderson, 1991, p. 7). If an equation does not satisfy this condition, it is underidentified and cannot be estimated. If an equation's restrictions on $\boldsymbol{\beta}$ are exactly one fewer than the rank of $\boldsymbol{\Pi}_{12}$, then the equation is just-identified; if the restrictions are more than one fewer than the rank of $\boldsymbol{\Pi}_{12}$, the equation is overidentified.[8] Koopmans and colleagues (1950) also specified a maximum-likelihood estimator for the general simultaneous equations model, which made Haavelmo's model accessible for empirical research.[9]

Perhaps the most important empirical applications of simultaneous equation models were Klein's (1950) Keynesian models, culminating with the 15-equation Klein–Goldberger model estimated by limited information methods (Klein & Goldberger, 1955). Others at Cowles had worried about the finite sample properties of estimation and introduced limited information methods as a solution (e.g., Anderson & Rubin, 1949).[10] Later, Theil (1953/1992) developed a two-stage least squares (2SLS) estimator that is consistent but asymptotically efficient only among single-equation estimators. He applied OLS to the reduced form, obtained predicted values for endogenous predictors, and applied OLS to the structural form having replaced endogenous predictors with their predicted counterparts. Zellner (1962) developed a joint generalized least squares (GLS) approach to seemingly unrelated regressions that incorporates information on covariances among errors of equations that are otherwise unrelated. He showed that GLS estimates and standard errors are minimum variance for linear models, and gain efficiency over OLS when the x's differ across equations and covariances among errors of equations are nonzero. Zellner and Theil (1962) developed a three-stage least squares (3SLS) estimator that applies joint GLS to the 2SLS estimates—using

information from the disturbance covariances—and showed that, for properly specified models, 3SLS is consistent, asymptotically efficient, and asymptotically equivalent to full-information maximum likelihood (ML).

From its heyday in the 1950s and 1960s, in which Keynesian macroeconomic models proliferated, interest in simultaneous equation models in economics declined (Epstein, 1987). This appears traceable to three events: (1) self-criticism by members of Cowles; (2) Lucas's (1976) rational expectations critique, in which economic agents anticipate policy interventions and then act contrary to linear models—implying that models omitting expectations are misspecified and structural parameters are not policy-invariant; and (3) empirical research suggesting that macro-Keynesian simultaneous equations models were not superior to simple naive models in forecasting the future (e.g., Nelson, 1972), leading to alternative time series models, such as vector autoregressions (Sims, 1980; see Epstein, 1987; Heckman, 2000).

The emphasis of Haavelmo and the Cowles Commission on models of errors in equations led most econometricians to abandon the errors-in-variables model emphasized by Frisch (1934). Two "path-breaking articles"—Zellner (1970) and Goldberger (1972a)—revived empirical interest in errors-in-variables models (Judge, Griffiths, Hill, & Lee, 1980). Zellner (1970) presented GLS (a modification of his joint GLS estimator) and Bayesian approaches to estimating a model with a fallible endogenous predictor with multiple causes. Goldberger showed that GLS is equivalent to ML only when errors are normally distributed with known variances. He also showed that when error variances are unknown, an iterated GLS will converge to ML.[11]

INTERDISCIPLINARY INTEGRATION

The year 1970 was a watershed year for structural equation modeling: Jöreskog (1970) published his general method of analyzing covariance structures; Hauser and Goldberger (1971) presented, at the sociology meetings, their work on unobservables in path analysis; and Zellner (1970) published his GLS results on unobservable independent variables. The year 1970 was also marked by the Conference on Structural Equation Models, an interdisciplinary forum featuring economists, sociologists, psychologists, statisticians, and political scientists, originating from a Social Science Research Council recommendation and culminating with the published volume *Structural Equation Models in the Social Sciences* (Goldberger & Duncan, 1973). This was presaged by the appearance of Blalock's (1971) edited volume *Causal Models in the Social Sciences*, which featured interdisciplinary contributions.

In this section, I focus on two key papers published in this period by Hauser and Goldberger (1971) and Jöreskog (1973). Hauser and Goldberger's (1971) examination of unobservable variables is an exemplar of cross-disciplinary integration, drawing on path analysis and moment estimators from Wright and sociologists, factor-analytic models from psychometrics, and efficient estimation and Neyman–Pearson hypothesis testing from statistics and econometrics. In a seminal and landmark paper that summarized his approach, Jöreskog (1973) presented his ML framework for estimating SEMs, developed a computer program for empirical applications, and showed how the general model could be applied to myriad important substantive models. Here, I focus on Hauser and Goldberger (1971) because they used limited information estimation to reveal what was going on "behind the scenes" of systems of structural equations estimated by ML.

Hauser and Goldberger (1971) analyze two models: the two-factor multiple indicator "walking dog" model (considered in factor analysis and by Costner and Blalock) and the MIMIC model.[12] Figure 2.2 presents a simple walking-dog model with four observables and two latent factors. We can express the model in matrix form:

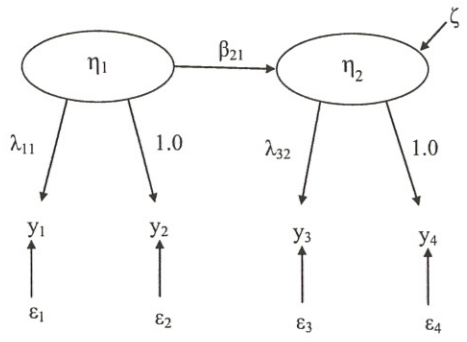


FIGURE 2.2. Path diagram of a walking-dog model in four observables and two latent variables.

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} \lambda_{11} & 0 \\ 1 & 0 \\ 0 & \lambda_{32} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} \qquad (2.11)$$

$$y = \Lambda \eta + \varepsilon$$

$$\begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ \beta_{21} & 0 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix} \qquad (2.12)$$

$$\eta = \mathbf{B} \eta + \zeta$$

It then follows that the covariance structure model is

$$\Sigma_{yy} = \Lambda(\mathbf{I} - \mathbf{B})^{-1} \Psi (\mathbf{I} - \mathbf{B})^{-1'} + \theta_\varepsilon \qquad (2.13)$$

where $\Sigma_{yy} = E(\boldsymbol{y}\, \boldsymbol{y}')$ is the (*population*) covariance matrix of observable indicators, Λ is a matrix of loadings, $\mathbf{B}$ is a matrix of regression coefficients among latent variables η, $\Psi = E(\zeta\ \zeta')$is the covariance matrix of structural disturbances, and $\theta_\varepsilon = E(\varepsilon\ \varepsilon')$ is a covariance matrix of measurement errors (diagonal in this example). This model is overidentified with one degree of freedom (10 moments and 9 parameters). The overidentifying restriction implies that there is more than one way of computing parameters in terms of moments, and there is a testable overidentifying restriction in the data. This can be seen by computing moments in terms of parameters (normal equations) and then solving for parameters in terms of moments. For example:

$$\lambda_{32} = \sigma_{23}/\sigma_{24} = \sigma_{13}/\sigma_{14} \qquad (2.14)$$

By cross-multiplying the second two terms and rearranging, we obtain the identical tetrad-difference restriction as found by Spearman (1927) and given in Equation 2.4, but in unstandardized form: $\sigma_{23}\,\sigma_{14} = \sigma_{24}\,\sigma_{13}$. Because Equation 2.14 can be satisfied by many different models, a traditional structural equation approach tests a specific nested parameterization of the restriction, rather than testing the tetrad-difference constraint on moments directly.

In estimating overidentified models, the question becomes which moment estimator(s) should be used. We can see this by replacing the population moments in Equation 2.14 with their sample counterparts, and noting we have two moment estimators for λ_{32}. In overidentified fully recursive models, the OLS estimator is unbiased and efficient; therefore, using the moment estimator corresponding to OLS (giving it a weight of one) and ignoring other instrumental variable moment estimators (giving them a weight of zero) is optimal. In the general case, however, one would not want to use only one estimator or a simple unweighted average, but instead weight the moments inversely to their sampling variability. Hauser and Goldberger (1971) show that this is precisely what ML does when minimizing the fit function in Equation 2.6. We can illustrate this point by noting that minimizing Equation 2.6 is asymptotically equivalent to minimizing a quadratic form (Anderson, 1973; Browne, 1974):

$$F_{GLS} = [s - \sigma(\theta)]' \mathbf{W}^{-1} [s - \sigma(\theta)] \qquad (2.15)$$

where s is a vector of nonredundant elements from the sample covariance matrix S, $\sigma(\theta)$ is the corresponding vector of elements of the parametric structure of the covariance matrix Σ—which makes $s - \sigma(\theta)$a discrepancy vector to be minimized—and $\mathbf{W}$ is a weight matrix consisting of the *covariance matrix of the sample moments.* Under normality, the latter consists of products of second-order moments about the mean. Thus, the parameters in $\sigma(\theta)$ are expressed as a function of sample moments s, each of which is weighted inverse to its sampling variability by $\mathbf{W}$. The estimator in Equation 2.15, termed GLS by Browne (1974), has been applied to econometric models by Hansen (1982), who terms it the "generalized method of moments."

The second model analyzed by Hauser and Goldberger (1971), the MIMIC model, is presented in a simple four-variable, three-equation form in Figure 2.3. This model has nine parameters, 10 observable moments and, therefore, one overidentifying restriction. In matrix form, the model is

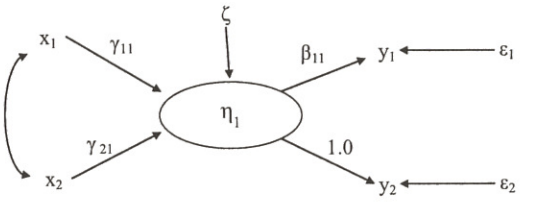


FIGURE 2.3. Path diagram of a multiple-indicator, multiple-indicator cause (MIMIC) model.

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \beta_{11} \\ \beta_{21} \end{pmatrix} (\eta_1) + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \qquad (2.16)$$
$$y = \mathbf{B} \quad \eta + \varepsilon$$

$$(\eta_1) = (\gamma_{11} \quad \gamma_{12}) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + (\zeta_1) \qquad (2.17)$$
$$\eta = \quad \Gamma \quad x + \zeta$$

and the covariance structure is

$$\begin{aligned} \Sigma_{yy} &= \mathbf{B}(\Gamma\Phi\Gamma' + \Psi)\mathbf{B}' + \theta_\varepsilon \\ \Sigma_{xy} &= \Phi\Gamma'\mathbf{B}' \end{aligned} \qquad (2.18)$$

where $\Sigma_{xy} = E(\boldsymbol{x}\, \boldsymbol{y}')$ contains the covariances between $\boldsymbol{x}$ and $\boldsymbol{y}$, $\Sigma_{yy} = E(\boldsymbol{y}\, \boldsymbol{y}')$ the covariances among the $\boldsymbol{y}$'s, $\Phi = \Sigma_{xx} = E(\boldsymbol{x}\, \boldsymbol{x}')$ (covariances among $\boldsymbol{x}$'s are unconstrained), $\theta_\varepsilon = E(\boldsymbol{\varepsilon}\, \boldsymbol{\varepsilon}')$ the covariance matrix of measurement errors in $\boldsymbol{y}$ (assumed diagonal here), and $\Psi = E(\boldsymbol{\zeta}\, \boldsymbol{\zeta}')$ the covariance matrix of the structural disturbance(s). Let us fix $\beta_{21} = 1.0$ to normalize the latent variable η and give it a metric; one could also normalize by fixing Ψ to a constant.

Using path analysis, we can compute moments in terms of parameters, solve for the parameters, and obtain two ways of expressing parameters in terms of moments. For example:

$$\beta_{11} = \sigma_{y_1x_1}/\sigma_{y_2x_1} = \sigma_{y_1x_2}/\sigma_{y_2x_2} \qquad (2.19)$$

Replacing the population moments with their sample counterparts gives us two sample moment estimators of β_{11}. Also, if we cross-multiply the right two terms in Equation 2.19 we get a single tetrad-difference overidentifying restriction, $\sigma_{y_1x_1}\sigma_{y_2x_2} = \sigma_{y_2x_1}\sigma_{y_1x_2}$. Note that this is the same restriction on observable moments we found for the walking-dog model above (if we denote all variables as y's), which illustrates an important difficulty for structural equation models: Overidentifying restrictions can be satisfied by substantively different models. In general, ML will provide consistent and asymptotically efficient estimates by weighting sample moments inverse to their sampling variability, resulting in optimal weights for multiple moment estimators. Again, minimizing the likelihood function will provide a likelihood ratio χ^2 test of overidentifying restrictions.

Hauser and Goldberger (1971) also use econometric methods to study identification and estimation. By substituting Equation 2.17 into η in Equation 2.16, we obtain the reduced form:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} \pi_{\varepsilon 1} \\ \pi_{\varepsilon 2} \end{pmatrix} \qquad (2.20)$$
$$y = \quad \Pi \quad x + \Pi_\varepsilon$$

where $\pi_{11} = \beta_{11}\gamma_{11}$, $\pi_{12} = \beta_{11}\gamma_{12}$, $\pi_{21} = 1.0\gamma_{11}$, $\pi_{22} = 1.0\gamma_{12}$, $\pi_{\varepsilon 1} = \beta_{11}\zeta_1 + \varepsilon_1$, and $\pi_{\varepsilon 2} = 1.0\zeta_1 + \varepsilon_2$. The reduced form can always be efficiently estimated using OLS. The estimation issue arises because there are two ways of expressing structural parameters in terms of reduced-form parameters:

$$\beta_{11} = \pi_{11}/\pi_{21} = \pi_{12}/\pi_{22} \qquad (2.21)$$

This also implies a proportionality constraint on reduced form parameters, providing a test of the MIMIC specification. ML weights the reduced-form parameter estimates $\hat{\pi}_{pq}$ inverse to their sampling variability to obtain asymptotically efficient estimates (Hauser & Goldberger, 1971). In this example, there is one degree of freedom and the single constraint can be expressed in terms of observed moments or reduced-form parameters. Generally, in more complex models, both kinds of restrictions exist, and ML will use both forms of restrictions in estimating parameters.[13] Jöreskog and Goldberger (1975) later expanded on ML estimation of the MIMIC model, and Goldberger (1973) discussed estimation in overidentified models with latent variables and simultaneity. For discussions of indicators as causes of theoretical constructs versus reflections of constructs, see Hauser (1973) and Bollen and Lennox (1991).

In a series of landmark papers, Jöreskog (1970, 1973, 1978) outlined a general approach to covariance analysis and a computer program he called LISREL, which, following econometricians as far back as Frisch and Waugh (1933), stood for "LInear Structural RELations." At about the same time, Keesling (1972) in his PhD dissertation, and Wiley (1973) in the Goldberger–Duncan volume, presented nearly identical models. However, it was Jöreskog's version and software package that came to dominate the field. The LISREL model incorporates factor analysis, simultaneous equation models, and path analysis (as discussed earlier) into a general covariance structure model (e.g., Jöreskog & Sörbom, 2001):

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xx} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Lambda}_y(\mathbf{I}-\mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}'+\boldsymbol{\Psi})(\mathbf{I}-\mathbf{B})^{-1\prime}\boldsymbol{\Lambda}_y' + \boldsymbol{\theta}_\varepsilon & \boldsymbol{\Lambda}_y(\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Lambda}_x' \\ \boldsymbol{\Lambda}_x\boldsymbol{\Phi}\boldsymbol{\Gamma}'(\mathbf{I}-\mathbf{B})^{-1\prime} & \boldsymbol{\Lambda}_x\boldsymbol{\Phi}\boldsymbol{\Lambda}_x' + \boldsymbol{\theta}_\delta \end{pmatrix} \quad (2.22)$$

Jöreskog showed that ML estimates are obtained by minimizing the following fit function and solving for parameters:

$$F_{ML} = \log|\boldsymbol{\Sigma}| - \log|\boldsymbol{S}| + tr\left(\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\right) - p + q \quad (2.23)$$

where $\boldsymbol{S}$ is the sample estimate of the population covariance matrix $\boldsymbol{\Sigma}$, and p and q are constants reflecting the number of observed y's and x's, respectively. If we let $\boldsymbol{\theta}$ be a vector of t parameters, then the $t \times t$ covariance matrix of parameter estimates, $\boldsymbol{V}$, is a function of the inverse of Fisher's information matrix:

$$\boldsymbol{V} = \left(\frac{2}{n}\right)\left[E\left(\frac{\partial^2 F}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right)\right]^{-1} \quad (2.24)$$

in which the square roots of the diagonal elements are asymptotic standard errors. Finally, if F_{H_0} is the minimum of Equation 2.23 under the null hypothesis, and F_{H_A} is the minimum under the less restrictive alternative, then –2 times the log likelihood ratio is

$$\nu = N(F_{H_0} - F_{H_A}) \quad (2.25)$$

which is asymptotically distributed χ^2 with $(p + q) - t$ degrees of freedom. Equation 2.25 can be applied to tests of nested models and the model's overall goodness of fit. Jöreskog (1971) also generalized this result to estimate the model in multiple populations, and showed how the model can be applied to simultaneous equations, MIMIC models, confirmatory factor models, panel data, simplex models, growth models, variance and covariance components, and factorial designs (for reviews, see Bentler, 1980, 1986; Bielby & Hauser, 1977).

In 1975, Duncan authored an excellent introductory text for path analysis and structural equation models, in which he echoed Frisch and Haavelmo's concept of autonomy: "The structural form is that parameterization—among the various possible ones—in which the coefficients are (relatively) unmixed, invariant, and autonomous" (p. 151). He also distinguished forms of social change, from trivial changes in sampling or exogenous variables (that leave structural coefficients intact) to deeper changes in structural coefficients (which provide fodder for explanation by multilevel models), and changes in the model's structure itself (p. 167), and provided sage advice for applying structural models (p. 150): "Do not undertake the study of structural equation models (or, for that matter, any other topic in sociological methods) in the hope of acquiring a technique that can be applied mechanically to a set of numerical data with the expectation that the result will automatically be 'research.' " Furthermore, Duncan noted that if research using structural models "are contributions to science (and not merely exercises in quantitative technique), it is because the models rest on creative, substantial, and sound sociological theory" (p. 151).

The next two decades saw an explosion of the use of structural equation models in many areas of the social sciences, including stratification (e.g., Bielby, Hauser, & Featherman, 1977), social psychology (e.g., Kohn & Schooler, 1982), psychology (e.g., Bentler & Speckart, 1981), marketing (Bagozzi, 1980), mental health (e.g., Wheaton, 1978, 1985), sociology of science (e.g., Hargens, Reskin, & Allison, 1976), criminology (e.g., Matsueda, 1982; Matsueda & Heimer, 1987), adolescence (e.g., Simmons & Blyth, 1987), and population genetics (e.g., Li, 1975). Some extensions of the model were developed during this period. Alwin and Hauser (1975) wrote a systematic treatment of decomposing effects into total, direct, and indirect effects using path analysis. Sobel (1982, 1986) applied the delta method to obtain asymptotic standard errors for total and indirect effects; Bollen (1987) developed a method for determining specific effects and their standard errors (implemented in Mplus); and Bollen and Stein (1990) developed bootstrap confidence intervals for indirect effects. Kenny and Judd (1984) showed how to estimate a LISREL model with product terms among latent exogenous variables, and Jöreskog and Yang (1996) showed that Browne's asymptotically distribution-free (ADF) estimator used on the matrix of augmented moments provides consistent estimates of parameters in the Kenny–Judd model, as well as consistent standard errors and fit statistics.

Matsueda and Bielby (1986) and Satorra and Saris (1985) independently showed how to calculate the power of the likelihood ratio test in covariance struc-

ture models—using the noncentral χ^2 distribution—and independently presented a nearly identical way of approximating the noncentrality parameter. They showed that the likelihood ratio test statistic is asymptotically equivalent to a quadratic form:

$$\nu = (\hat{\theta}_r - \theta_{r0})' V_r^{-1} (\hat{\theta}_r - \theta_{r0}) \quad (2.26)$$

I wonder if these parentheses are incorrectly in bold? They look bolder than those of 2.27

insert thin space before prime symbol

the ML estimator for the rth parameter, responding null hypothesis, and V_r is the ovariance matrix of $\hat{\theta}_r$ or, in other words, al submatrix of Equation 2.24. Under the null hypothesis, ν has a central χ^2 distribution with r degrees of freedom. Under the alternative hypothesis, ν has a noncentral χ^2 distribution with r degrees of freedom and noncentrality parameter:

$$\tau = (\theta_r - \theta_{r0})' V_r^{-1} (\theta_r - \theta_{r0}) \quad (2.27)$$

where θ_r is the population parameter corresponding to the alternative hypothesis and θ_{r0} is the population parameter corresponding to the null hypothesis (see Kendall & Stuart, 1979, pp. 246–247). Matsueda and Bielby (1986) then drew on Hauser and Goldberger (1971) and Goldberger (1973) to show analytically that, in a walking-dog model (Figure 2.2), adding indicators to the endogenous latent variable increases the power of the test of β_{21}, depending on the reliability of the indicators. This is analogous to adding cross sections to a pooled time series cross-section econometric model estimated by GLS. They also gave simulation results for adding indicators to the exogenous latent variable (see Matsueda & Bielby, 1986; Satorra & Saris, 1985).

Applied researchers obsessed over the global goodness-of-fit likelihood ratio χ^2 test because, in large samples, models with many overidentifying restrictions tend to be rejected even when each restriction only departs trivially from the null hypothesis. This gave rise to a cottage industry of fit indices designed to offset the effect of sample size on test statistics. From this literature, it seems that a consensus is emerging that the most useful fit indices are Steiger and Lind's (1980) root mean square error of approximation (RMSEA; see also Browne & Cudeck, 1993; Steiger, 1990) and Raftery's (1993, 1995) application of Schwartz's (1978) Bayesian information criterion (BIC). (For details, see West, Taylor, & Wu, Chapter 13, this volume.) RMSEA is defined as

$$\varepsilon = \sqrt{F_0 / r} \quad (2.28)$$

where F_0 is the population discrepancy function reflecting the model's lack of fit and r is the degrees of freedom, as earlier. MacCallum, Browne, and Sugawara (1996) have defined the noncentrality parameter for RMSEA index:

$$\lambda = (n-1) r \varepsilon^2 \quad (2.29)$$

where n is the sample size. They show that power can be calculated for the null hypothesis of perfect fit (i.e., $\varepsilon = 0$), as well as an approximate or close fit (e.g., $\varepsilon \leq .05$). The latter may be useful in very large samples for models with many overidentifying restrictions, whereby reasonably well-specified models are likely to be rejected (see Lee, Cai, & MacCallum, Chapter 11, this volume). To date, RMSEA is the most popular fit index used in empirical applications of SEM, although it recently has been subject to criticism (e.g., Chen et al., 2008, on using a fixed cutoff point).

ADDRESSING VIOLATIONS OF DISTRIBUTIONAL ASSUMPTIONS

At this time, a major criticism of structural equation models is that the assumptions of continuous observed variables, multivariate normal distributions, and large sample sizes—needed to capitalize on the asymptotic properties of maximum likelihood estimation and testing—are rarely met in practice. Some early Monte Carlo studies, such as Boomsma (1983), which created non-normal errors by categorizing continuous variables, found that estimators were robust when samples were greater than 200, but that skewness due to categorization produced spurious measurement error correlations and biased standardized coefficients (see Bollen, 1989, for a review).[14]

ADF Estimator

As noted earlier, Browne (1974) introduced the quadratic form estimator he termed generalized least squares (GLS), which yielded optimal estimates for normally distributed observable variables when $\boldsymbol{W}$ is the covariance matrix of the sample moments (see Equation 2.15). Subsequently, Browne (1984) made a landmark contribution by developing what he termed an "asymptotic distribution-free" (ADF) estimator, by incorporating information about higher-order moments

into the weight matrix of Equation 2.15, which can be written in scalar form as

$$F_{GLS} = \sum_{g=1}^{k}\sum_{h=1}^{g}\sum_{i=1}^{k}\sum_{j=1}^{i} w^{gh,ij} \left[s_{gh} - \sigma_{gh}(\theta)\right]\left[s_{ij} - \sigma_{ij}(\theta)\right] \quad (2.30)$$

where s_{gh} is the sample covariance between variables g and h, $\sigma_{gh}(\theta)$ is the corresponding element of $\boldsymbol{\Sigma}(\theta)$ implied by the model, and $w^{gh,ij}$ is a typical element of $\boldsymbol{W}^{-1}$, which is $u \times u$, where $u = k\ (k + 1)$, and k is the number of observables. Browne showed that if $\boldsymbol{W}$ is a matrix with typical element

$$w_{gh,ij} = m_{ghij} - s_{gh}s_{ij} \quad (2.31)$$

where m_{ghij} is the fourth-order moment about the mean, then minimizing Equation 2.15 yields the ADF estimator, which is minimum variance consistent within the class of estimators in the form of Equation 2.15 under the mild assumption that eighth-order moments are finite (Browne, 1984, p. 710).[15] Browne presented the asymptotic covariance matrix for $\hat{\theta}_{ADF}$ and an asymptotic χ^2 test statistic, as well as an estimator for elliptical distributions, which have zero skewness but kurtosis that departs from multivariate normality.

Browne's (1984) ADF and elliptical estimators first appeared in Bentler's (1995) EQS program, followed by Jöreskog and Sörbom's (2001) LISREL program. Recent work has examined the finite sample properties of ADF and finds that it works well in very large samples. Other techniques available are using corrections to the covariance matrix of ML estimators to obtain accurate p-values for the χ^2 test under non-normality (e.g., Browne 1984), or a bootstrap method (Bollen & Stein, 1993). Browne's ADF estimator was also crucial for a second important advance: developing models for ordinal, limited, and discrete outcomes.

Models for Ordinal, Limited, and Discrete Outcomes

Structural equation models are often applied to survey data, in which items are measured on dichotomous or ordinal scales, violating the assumption of continuous and normally distributed observed variables. Muthén (1984) has made seminal contributions for analyzing dichotomous, ordinal, and limited dependent variables within a covariance structure framework. The trick is to estimate scale-appropriate correlation coefficients (e.g., polychoric and polyserial) and then use a variation of Browne's (1984) ADF estimator. The polychoric correlation, which goes back to Pearson (1901), computes a correlation under the assumption that the ordered categories can be represented by contiguous intervals on a continuous scale (correlations between ordinal and continuous variables are termed "polyserial correlations").[16] Thus, the ordinal variable is related to the underlying normally distributed, continuous latent variable through a threshold model. Early work on factor models for dichotomous variables include Bock and Lieberman (1970), who used tetrachoric correlations and an ML estimator for a single factor model, and Christoffersson (1975), who generalized this to multiple factors using a GLS estimator (see also Muthén, 1978). Muthén (1979) developed a multiple-indicator structural probit model, and Winship and Mare (1983, 1984) applied multivariate probit models estimated by ML to multiple-indicator structural equation models and path analysis.

Muthén (1984) provided a general framework for analyzing ordinal variables. Here I focus on the polychoric and ADF approach with a simple example of a pair of three-category ordinal variables. Each ordered variable is related to an underlying continuous variable by two thresholds:

$$\begin{aligned} y &= 1 \quad \text{if } y^* \le \alpha_1 \\ y &= 2 \quad \text{if } \alpha_1 \le y^* < \alpha_2 \\ y &= 3 \quad \text{if } \alpha_2 \le y^* \end{aligned} \quad (2.32)$$

where the value for y indexes the ordinal category for y, y^* is a latent continuous variable, and α_1 and α_2 are thresholds. If we specify a distribution for y^*—we will assume it is normal—we can then estimate the thresholds by the general formula:

$$\alpha_i = \Phi^{-1}\sum_{k=1}^{i} n_k / N \qquad i = 1,2,3 \quad k = 1,2 \quad (2.33)$$

where i indexes the category of y, k indexes the number of thresholds, $\Phi^{-1}(.)$ is the inverse of the standard normal distribution function, n_k is the sample size of the kth category, N is the total sample size, and $N = n_1 + n_2 + \ldots + n_k$. If we apply this to a second three-category ordered variable x, but with thresholds β_1 and β_2, and define π_{ij} as the population parameter denoting that an

observation falls into cell (i, j), we can then define the log-likelihood function of the sample (Olsson, 1979):

$$\log L(n_{ij} \mid \pi_{ij}) = c\sum_{i=1}^{3}\sum_{j=1}^{3} n_{ij} \log \pi_{ij} \qquad (2.34)$$

where $\pi_{ij} = \Phi_2(\alpha_i, \beta_j) - \Phi_2(\alpha_{i-1}, \beta_j) - \Phi_2(\alpha_i, \beta_{j-1}) - \Phi_2(\alpha_{i-1}, \beta_{j-1})$, and $\Phi_2(.,.)$ is the bivariate normal distribution function with population correlation ρ. Maximizing Equation 2.34 will yield the ML estimator of the polychoric correlation, $\hat{\rho}_{ML}$. Alternatively, one can use a two-step procedure, estimating the thresholds α_i and β_j from the marginals of the contingency table (e.g., Equation 2.33), and then solving for ρ. See Olsson (1979) for a parallel treatment of the polyserial correlation between continuous and ordered variables, and Poon and Lee (1987) for multivariate ML estimators of both polychoric and polyserial correlations.

Once the polychoric and polyserial correlations $\hat{\rho}$ and their asymptotic covariances have been estimated, Browne's (1984) ADF fitting function can be used to obtain optimal estimates:

$$F_{ADF} = \left[\hat{\rho} - \sigma(\theta)\right]' S_{\rho\rho}^{-1} \left[\hat{\rho} - \sigma(\theta)\right] \qquad (2.35)$$

where $\hat{\rho}$ is a vector of scale-appropriate correlation estimates, $\sigma(\theta)$ is the corresponding vector of the parametric structure generating the correlations, and $S_{\rho\rho}^{-1}$ is the inverse of the asymptotic covariance matrix of the correlation estimates. Standard errors and χ^2 test statistics are obtained as earlier (Muthén, 1984). Muthén (1989) has also developed a tobit factor analysis for censored observed variables. A general method for handling dichotomous and limited dependent variables in SEM was initially programmed in Muthén's LISCOMP program, and then in his recent more comprehensive Mplus program (Muthén & Muthén, 2004), and later in Jöreskog and Sörbom's (2001, 2002) LISREL and PRELIS programs, and Bentler's (1995) EQS. Much of this material is covered in Bollen's (1989) excellent intermediate-level SEM text.

RECENT ADVANCES

Major contemporary advances in SEM make it an exciting and growing field. These include the development of latent growth and latent-class growth models for longitudinal data, the application of Bayesian methods, the integration of generalized linear models and multilevel models within an SEM framework, the adoption of algorithms from artificial intelligence to discover causal structure, and a formal treatment of causality within an SEM framework.

Latent Growth and Latent Class Growth Models

Although the use of factor analysis for modeling panel data on growth was introduced by Tucker (1958) and Rao (1958), it was not until 1990 that Meredith and Tisak (1990) published the treatment within an SEM framework that is still relevant today (see Bollen & Curran, 2006). Meredith and Tisak (1990) showed that individual growth curves, often modeled within a multilevel or mixed model framework (e.g., Raudenbush & Bryk, 2002), can be modeled within a standard SEM framework by treating the shape of growth curves as latent variables with multiple indicators consisting of the variable at multiple time points. This latent growth curve approach models both covariances and means of observed variables. Figure 2.4 presents a path diagram of a four-wave quadratic latent growth curve model. Here, the intercept α gives the value of y implied by the model at the first time point; β_1 is the linear growth component (giving the growth rate at the first time point); and β_2 is the quadratic growth component (giving the change in the growth rate over time). One can then impose a parametric structure on the growth pa-

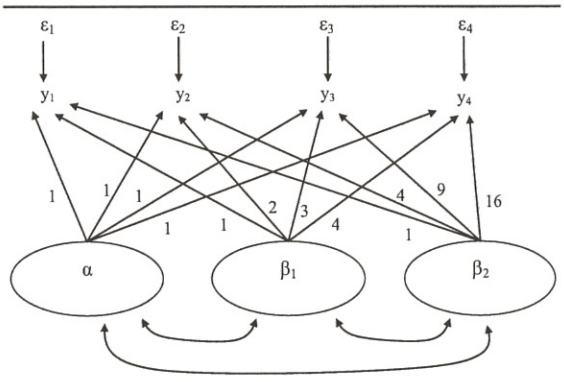


FIGURE 2.4. Path diagram of a four-wave quadratic latent curve model.

rameters α, β_1, and β_2, which would correspond to the second-level equation in a multilevel model.

In criminology, Nagin and Land (1993) developed a finite mixture model for latent classes of individual trajectories. This group-based trajectory model estimates individual trajectories using polynomials and then classifies the trajectories into discrete groups. The latent classes can be viewed as points of support in approximating a continuous distribution of unobserved heterogeneity or as reflections of theoretically important groups (see Nagin, 2005). In criminology, this model has been used to distinguish people with different criminal careers, such as chronic offenders, early starters, and adolescence-limited offenders (see Nagin & Tremblay, 2005). Muthén (2004) shows how to estimate this model within an SEM framework with Mplus. Moreover, Muthén's approach, termed "growth mixture modeling," allows for within-class variation among individual trajectories, a mean curve with variation around it, whereas Nagin's approach does not. The latter is nested within the former and can be subjected to statistical test. These models have become important features of research in child development, adolescent substance abuse, and criminal careers (e.g., Connell & Frye, 2006; Nagin & Tremblay, 2005).

Bayesian Approaches

As noted earlier, work at the Cowles Commission helped cement the Neyman–Pearson frequentist approach to hypothesis testing in econometric simultaneous equation models, which spread to SEM in psychology through Anderson and Jöreskog, and in sociology through Goldberger and Duncan. In recent years, alternatives—particularly Bayesian approaches—have been proposed for SEM (for an early and persistent advocate in economics, see Zellner, 1971). From a Bayesian perspective, estimation is less about deducing the values of population parameters and more about updating, sharpening, and refining our beliefs about the empirical world.

Bayesian estimation begins with a probability model of the data, D, in terms of a vector of parameters, θ (e.g., Raftery, 1995). The analyst's prior beliefs about the uncertainty of θ is denoted by the prior probability density, $p(\theta)$. The probability model for the data, then, is the likelihood function, $p(D|\theta)$, which is the probability of the data given that θ are the true parameters. We then observe the data, D, and update our beliefs about θ using Bayes's theorem:

$$p(\theta \mid D) = [p(D \mid \theta)p(\theta)] / p(D) \quad (2.36)$$

The data are treated as a fixed set of information to be used in updating our beliefs about the parameters. Therefore, $p(D)$ does not involve θ, and Equation 2.36 reduces to

$$p(\theta \mid D) \propto p(D \mid \theta)p(\theta) \quad (2.37)$$

where $\propto$ means "proportional to." The marginal density of the data has been dropped; to make this a proper density, a proportionality constant can be added. Thus, the posterior density is proportional to the likelihood times the prior density. Inferences about θ are made from summary statistics about the posterior density, $p(\theta|D)$, such as the posterior mode or Bayesian confidence intervals ("credible intervals"), which have an intuitive interpretation: "The probability that the true value of the parameter lies in the interval is—for example—95%."

Bayesian hypothesis testing entails comparing hypotheses to determine which has the highest probability of being correct. Suppose we have two hypotheses, H_0 and H_1, with *prior* probabilities, $p(H_0)$ and $p(H_1)$ before the data are examined, and define the prior odds ratio as

$$Odds_{\text{prior}} = p(H_0)/p(H_1) \quad (2.38)$$

After examining the data, the prior probability will be updated, resulting in posterior probabilities for each hypothesis, $p(H_0|D)$ and $p(H_1|D)$, and a posterior odds ratio:

$$\begin{aligned} p(H_0|D)/p(H_1|D) &= Odds_{\text{posterior}} \\ &= B_{01} \times Odds_{\text{prior}} \end{aligned} \quad (2.39)$$

where B_{01} is the Bayes factor:

$$B_{01} = p(D|H_0)/p(D|H_1) \quad (2.40)$$

and $p(D|H_0)$ and $p(D|H_1)$ are the marginal probabilities of the data. Equation 2.39, the posterior odds ratio, gives the probability that the data support H_0 over H_1. Note that the posterior odds are equal to the Bayes factor when the prior odds are equal to 1.

Bayesian estimation and testing are currently diffusing into the SEM literature. For example, Raftery (1993, 1995) showed how to approximate the Bayes factor with the BIC, which is computed from the likeli-

hood ratio test statistic. Suppose we wish to compare two models in which M_{k-1} is nested within M_k and has one more parameter than M_k. Then, if ν_{k-1} is the likelihood ratio test statistic for model M_{k-1} and ν_k is that of model M_k and $\nu = \nu_{k-1} - \nu_k$ is the test statistic for testing the one parameter (see Equation 2.25), then

$$BIC_{k-1} - BIC_k \approx \nu - \log n \qquad \textbf{(2.41)}$$

where n is the sample size. If this quantity is positive, then the less-restrictive model M_k is preferred (Raftery, 1995). As a fit statistic, BIC has performed exceedingly well in a variety of contexts and is particularly useful for SEM in large samples and with many overidentifying restrictions because trivially small departures from a reasonable model will be rejected using the likelihood ratio test. Although it has become the dominant fit statistic for most models estimated by ML, it has only recently begun to be used in SEM empirical studies.

Bayesian estimation using Markov Chain Monte Carlo (MCMC) algorithms are proving useful for incorporating prior information into confirmatory factor analysis (e.g., Lee, 1981); estimating complex models, such as nonlinear latent variable models (e.g., Arminger & Muthén, 1998); estimating multilevel factor models (Goldstein & Browne, 2002); arriving at a semiparametric estimator (Yang & Dunson, 2010); and drawing inferences about underidentified parameters from the posterior distribution when an informative prior is used (Scheines, Hoijtink, & Boomsma, 1999). For details, see Kaplan and Depaoli, Chapter 38, this volume. The program, TETRAD III, provides an MCMC algorithm using the Gibbs sampler (Scheines, Spirtes, Glymour, Meek, & Richardson, 1997).

Generalized Linear Latent and Mixed Models

When data take on a hierarchical structure—such as individuals nested within families, which in turn are nested within neighborhoods—special methods are needed to obtain consistent estimates of standard errors and test statistics due to dependent observations within clusters. Multilevel regression models allow estimation of models in which random intercepts capture heterogeneity between clusters in the dependent variable, and random coefficients capture heterogeneity in relationships among independent and dependent variables. A multilevel structural equation model would incorporate multiple-indicator measurement models into the latent variable models. Early attempts to incorporate measurement error into multilevel regression models have assumed that measurement error variances (e.g., Goldstein, 1995) or factor loadings (e.g., Raudenbush & Sampson, 1999) are known and have the advantage that unbalanced designs, in which the number of Level 1 cases varies by Level 2, are easily handled if missing at random (see Rabe-Hesketh, Skrondal, & Pickles, 2004).

Multilevel structural equation models have typically specified separate models for within-cluster and between-cluster covariance matrices. For example, Muthén (1994) has shown how to estimate a two-level SEM using available SEM software. The trick is to specify separate within- and between-cluster models, and then use the multiple-group option to estimate the parameters simultaneously. Muthén argues that an estimator using this method is equivalent to ML in balanced designs, and is consistent (with reasonable standard errors and test statistics) in unbalanced designs (see also Goldstein & McDonald, 1988; Muthén, 1997; for a review of alternate estimators, see Yuan & Bentler, 2007). This approach is easily implemented using existing SEM software but is limited to specific models.

A more general approach is outlined in Rabe-Hesketh and colleagues (2004), and expanded in Skrondal and Rabe-Hesketh's (2004) excellent advanced text. Their generalized linear latent and mixed models (GLLAMM) framework consists of three components: (1) a response model; (2) a structural equation model for latent variables; and (3) distributional assumptions for latent variables. The response model is simply a generalized linear model conditional on the latent variables and consisting of a linear predictor, a link function, and a distribution from the exponential family (Rabe-Hesketh et al., 2004). The model can handle response variables that are continuous, ordinal, dichotomous, discrete and continuous time durations, counts, polytomous responses and rankings, and mixes of responses. The structural equation for latent variables takes on the usual form, $\boldsymbol{\eta} = \boldsymbol{\beta\eta} + \boldsymbol{\Gamma\xi} + \boldsymbol{\zeta}$, with the exception that latent variables are allowed to vary by different levels. Rabe-Hesketh and colleagues assume the latent variables at level l are distributed multivariate normal with zero mean and covariance matrix $\boldsymbol{\Sigma}_l$, although other distributions can be specified. The authors have also written a program, GLLAMM, which maximizes the marginal likelihood using an adaptive quadrature procedure and is available in the software package Stata

(Rabe-Hesketh, Pickles, & Skrondal, 2001). For more details, see Skrondal and Rabe-Hesketh (2004) and Rabe-Hesketh, Skrondal, and Zheng (Chapter 30, this volume). Many of these models can also be estimated using Mplus (Muthén & Muthén, 2004).

Tetrad: The Discovery of Causal Structure

A philosophically distinct approach to SEM developed with the publication of Glymour, Scheines, Spirtes, and Kelly's (1987) *Discovering Causal Structure: Artificial Intelligence, Philosophy of Science, and Statistical Modeling.* Instead of focusing on estimation and testing of structural models specified on a priori grounds, Glymour and colleagues draw on computer algorithms from artificial intelligence to "discover" causal structure with their program TETRAD. Thus, they are returning to the earlier ideas of Spearman, Frisch, Simon, Blalock, and Costner, who tried, in various ways, to induce causal structure from patterns of association among variables.[17] As noted earlier, Spearman's focus on tetrad difference restrictions on observed correlations became superfluous in light of Thurstone's rotated solution to simple structure for factor models; Frisch's confluence analysis and bunch mappings became obsolete with advances in identification and estimation in simultaneous equations; and Simon and Blalock's method of ransacking three- and four-variable models became outdated with the development of estimation and testing using ML and GLS in integrated SEMs. These "outdated" approaches have been resurrected by Glymour et al. (1987).

Beginning with the observation that an infinite number of models is consistent with any covariance matrix of observed variables, Glymour and colleagues (1987) return to Simon and Blalock's method of identifying the vanishing partial correlations that must hold for a given model and to the writings of Wright, Spearman, and others, who identified the tetrad difference equations that must hold for a given model. They provide a fascinating philosophy of science edifice to justify the idea of discovering causal structure. Moreover, they use the terminology of directed graphs—rather than path analysis—in which variables are vertices; causal effects are directed edges that can be into a vertex (the number denoted by indegree) and out of a vertex (the number denoted by outdegree); a recursive model is acyclic; a nonrecursive model (in which a path contains a subpath beginning and ending in the same vertex) is a cyclic model; and a trek is a path or a set of paths that induce a correlation. Using these notions, they provide basic theorems and definitions about causal relations and a computer program, TETRAD, for discovering causal structure. The program allows users to incorporate a priori information about the data (e.g., a tentative or partial model), identifies the vanishing tetrad differences and vanishing partial correlations of the model, and then provides a test of the constraints. It then modifies the model by identifying the treks needed to satisfy a "false" tetrad equation without altering "true" tetrad equations, and calculates vanishing tetrads and partial correlations implied by the new model. Bollen (1990) develops a simultaneous test statistic to address the multiple testing problem and Bollen and Ting (1993) develop a confirmatory tetrad analysis approach for testing SEMs, including some non-nested and underidentified models.

The TETRAD approach is not without its critics (e.g., Humphreys & Freedman, 1996) and controversies (e.g., Glymour & Cooper, 1999; Spirtes, Glymour, & Scheines, 1993). Robins and Wasserman (1999) have effectively shown that it is not possible, in an observational study, to infer causal relationships between two variables, assuming that the sample is large and the distribution of the random variables is consistent with the causal graph, and assuming no additional substantive background knowledge (as claimed in Spirtes et al., 1993; Pearl & Verma, 1991). Using a Bayesian framework, they demonstrate that the claim assumes that the prior probability of no unmeasured causes is high relative to sample size, and when this probability is low relative to sample size, causal relationships are underidentified from the data. This opens a new puzzle: What kinds of assumptions must be made to discover causal relationships from nonexperimental data (Glymour, Spirtes, & Richardson, 1999)? This puzzle has encouraged cross-fertilization from the causality literature in statistics into the SEM literature.

Nevertheless, TETRAD is certainly a useful empirical tool for exploring causal structures, finding equivalences, and providing an efficient tool for sensitivity analyses. TETRAD has been shown to be more efficient at modifying models than existing procedures available in SEM packages, such as using first-order partial derivatives, residuals, or univariate Lagrange multipliers (termed "modification indexes" in LISREL). Moreover, introducing the language of directed graphs into structural equation models helps bridge the SEM literature with new developments in graphical theory and causal analysis (see Spirtes et al., 1993).

Causality and Structural Equation Models

An exciting recent development, which is slowly filtering into the SEM community, has been an explicit return to causal models using the language and logic of counterfactuals: What would happen if a subject received a different treatment (or value of the independent variable)?[18] As a consequence of Freedman's (e.g., 1987) persistent criticisms of SEM applications for making causal claims when assumptions of linear models are not met, and the more compelling critique of Holland (1988), most members of the SEM community have retreated from making causal claims and using causal language.[19] However, a recent literature, emerging from disparate sources such as statistics, artificial intelligence, philosophy of science, epidemiology, and economics, has developed graphical models for identifying causality under explicit weaker assumptions than are generally made. Such models, which generalize and extend Wright's (1921a) original path analysis, appeared in the social science literature as early as 1982 but did not catch on (see Kiiveri & Speed, 1982). Since then, major advances have been made in artificial intelligence (e.g., Pearl, 1988), statistics (e.g., Spirtes et al., 1993; Wermuth & Lauritsen, 1983), and epidemiology (Greenland, Pearl, & Robins, 1999; Robins, 1986; Robins & Greenland, 1992).

This approach begins with the classical SEM assumption that causality cannot be determined from observational data alone, but requires additional causal assumptions drawn from theoretical or substantive knowledge, which are translated into a structural model represented by a path model. At this point, the approach departs from traditional path analytic and SEM treatments of causality, in which strong and often unrealistic assumptions must be made in empirical applications:

> Structural equation models do little more to justify the causal interpretation of their coefficients than the causal orderings of path analysis. In both approaches, such causal interpretations are established by fiat rather than by deduction from more basic assumptions. (Holland, 1988, p. 460)

The contemporary counterfactual causality literature lays bare the *typically strong assumptions* underlying "causality by fiat," and more importantly, searches for ways of identifying causal effects under *weaker assumptions*.

To get a sense of this literature, consider a fundamental issue of causality in SEM—decomposing total effects into direct and indirect effects (see Figure 2.5). From a causality perspective, by manipulating and randomizing values of the independent variable, causal effects of that variable can be identified because reciprocal causation and omitted variable bias are ruled out. In the case of indirect effects, this requires sequential randomization (Robins & Greenland, 1992). For example, in Model A of Figure 2.5, by randomizing on X we can obtain the causal effect of X on M and the total causal effect of X on Y. However, because M is endogenous—and therefore, neither manipulated nor randomized—we cannot obtain the causal effect of M on Y. Consequently, we cannot obtain the direct causal effect of X on Y, and cannot decompose the total causal effect of X into direct and indirect components. To obtain the causal effect of M on Y we must randomize on M—hence, the need for *sequential* randomization.

In the social sciences, however, such sequential randomization is rarely possible; therefore, the question becomes, "What assumptions are necessary to identify causal direct and indirect effects?" An early paper by Holland (1988) discussed the question within Rubin's (e.g., 1974) model, which carefully separates causal theory from observed data, and which begins with unit causal effects based on counterfactuals and then defines average causal effects. Rubin's model typically assumes independent outcomes, and in particular, that treatment of one individual does not affect the outcome of another.[20] Holland noted that if we can assume that the effect of X on Y operates solely through its effects on M—that is, $c = 0$, which yields Model B of Figure 2.5— then the average causal effect of M on Y is identified and estimable using an instrumental variable estimator. Since X is orthogonal to u (by randomization) and X has no direct effect on Y (because $c = 0$), X can serve as an instrument for M (as long as $a \neq 0$) and a consistent estimator of b, the average causal effect of M

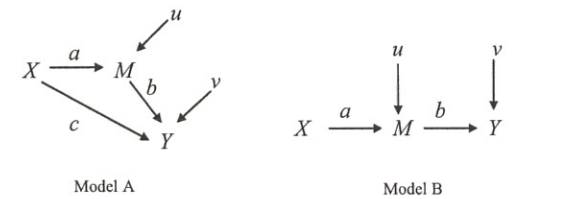


FIGURE 2.5. Path diagram of models with direct and indirect effects.

on Y is simply the total average causal effect of X on Y divided by the direct average causal effect of X on M. This estimator is useful for encouragement designs—in which it is reasonable to assume that X, randomized encouragement (e.g., encouraging students to study), affects Y solely through M, the activity encouraged (studying)—but is of limited utility in most observational studies in the social sciences (Sobel, 2008).

A related approach to the separation of causal direct and indirect effects is associated with the work of Robins (e.g., Robins & Greenland, 1992), which builds on Robins's (1986) original graphical approach to causality using tree graphs. Robins and Greenland (1992) show that Robins' g-computation algorithm can be used to separate direct and indirect effects of X if X and M are both randomized, X and M do not interact, and M can be blocked by intervention (i.e., manipulated). When these conditions hold, but M is not manipulated, g-estimation can still estimate the fraction of total effect that could be prevented by blocking M if additional confounding variables are available. Finally, when all conditions hold, except X and M interact in affecting Y, direct and indirect effects cannot be separated, although one can still estimate the fraction of the total effect of X that could be prevented by blocking M.

Recently, Emsley, Dunn, and White (2010) reviewed alternative approaches to estimating mediating effects in controlled trials and showed that under treatment heterogeneity, the interaction terms between randomized treatment and exogenous confounders can be used as instrumental variables to separate direct and indirect effects of treatment when the mediating variable is not randomized. Sobel (2008) shows that instrumental variables can be used to separate direct and indirect effects under weaker assumptions—such as relaxing the assumption of constant treatment effects—than presented by Holland (1988). Jo and Muthén (2002) have used latent growth models to define principal strata of the mediator and estimating, for models with multiple outcomes, complier average causal effects (CACE), which are average effects of treatment in the subpopulation of compliant subjects (e.g., Angrist, Imbens, & Rubin, 1996; Imbens & Rubin, 1997).

In economics, following Imbens and Angrist (1994), CACE, defined as local average treatment effects (LATE), has spawned a spate of randomized experiments using instrumental variables to identify treatment effects (see Imbens & Wooldridge, 2009, for a review). The relative merits of using randomization to identify causal effects versus using structural models, such as Roy's (1951) counterfactual model, remain controversial among economists (Heckman & Urzúa, 2010). For an excellent discussion of the relative merits of the two approaches in economics, including key trade-offs between internal and external validity, see Deaton (2010), Imbens (2010), and Heckman (2010).

From artificial intelligence and epidemiology, a graphical model approach to causality has emerged. This approach represents the causal assumptions by a graphical model and then logically infers causality from a set of theorems applied to the graph. In addition to Robins's (1986) tree graph approach, Pearl (2000) developed a "nonparametric structural causal model," which holds regardless of distributional and other statistical assumptions about a particular data set. Causal statements can be made that are *conditional on the causal assumptions* encoded into the graphical model.

Pearl's (2000) approach, which is largely consistent with that of Robins (1986), advances SEM by (1) using new mathematical notation to reflect causality, such as replacing the algebraic equals sign with a sign that reflects a causal path; (2) deriving a theorem, the "back door" criterion, to determine which covariates should be controlled to arrive at a causal relationship in an SEM; (3) deriving a theorem, termed "d separation" (directed separation), which gives the necessary and sufficient conditions for independence between two sets of variables conditioned on a third set within an acyclic directed graph; (4) providing some simple mathematical notation for making counterfactual statements, which can be analyzed within the directed graph (for an introduction, see Morgan & Winship, 2007); and (5) providing an algorithm for identifying equivalent models. Taken together, these theorems translate the language of causality into the language of statistics and probability distributions (for distinctions between the approaches of Robins and Pearl, see Robins, 1995, 2003). See Pearl (2000) for an excellent presentation of the graphical approach to SEM and for a lucid introduction to the principles and issues, see Pearl (Chapter 5, this volume).

The importance of a graphical approach can be illustrated with a simple example. Consider Model A in Figure 2.6, a bivariate regression model of Y on X, with two latent variables: ξ affects X and C, and η affects Y and C. Standard SEM texts assume that including an irrelevant variable in a linear regression model leaves estimates unbiased but results in a loss of precision in the estimate (e.g., Greene, 2003, pp. 150–151). However, when Model A is the correct model, regressing Y on X

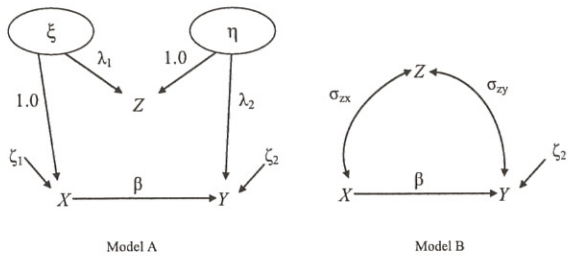


FIGURE 2.6. Path diagram of a model in which controlling for an exogenous variable creates bias.

and Z, which appears irrelevant, results in biased and inconsistent estimates of β. We can see this intuitively from Model A: the covariance between Z and X and Z and Y are spurious due to the latent variables ξ and η. Spirtes, Richardson, Meek, Scheines, and Glymour (1998, pp. 190–191) show that if we simplify Model A to Model B (Figure 2.6), and compute parameters in terms of covariances and partial covariances, we obtain an estimator from regressing Y on X and Z that is biased and inconsistent:

$$E(X,Y \mid Z)/E(X^2 \mid Z) = \beta\sigma_x^2 - \sigma_{zx}\sigma_{yz} / \left(\sigma_y^2\sigma_z^2 - \sigma_{xz}^2\right) \quad \textbf{(2.42)}$$

where the term left of the equality sign is the (naive) two-variable regression coefficient, and σ_{xz}^2 is the squared covariance of X and Z. A graphical approach quickly reveals not only that an unbiased and consistent estimate of β is obtained by the bivariate regression of Y on X, but also that a consistent estimator can be obtained by the naive two-variable regression by also controlling for ξ, η, or both (Greenland & Brumback, 2002).

The recent causality literature suggests that the parameters of most applications of SEM in the social sciences cannot be interpreted as causal effects without making strong and often unrealistic assumptions. What, then, are we to make of empirical applications of SEM, such as status attainment models? Perhaps a prudent interpretation, consistent with O. D. Duncan's thinking near the end of his career, is that such models "summarize systematic patterns in population variability" (Xie, 2007, p. 154) or, perhaps more precisely, describe "the probabilistic relationship between successive events in a population of interest" (Sobel, 1992, p. 666). Such a description—even in the absence of causal language—is a crucial feature of social science research.

CONCLUSION

SEM has progressed through four general stages: (1) early disciplinary-specific developments of path analysis first from genetics and later sociology, factor analysis from psychology, and simultaneous equation models in economics; (2) cross-disciplinary fertilization between economics, sociology, and psychology, leading to an explosion of empirical applications of SEM; (3) a period of developing methods for handling discrete, ordinal, and limited dependent variables; and (4) a recent period of incorporating statistical advances into the SEM framework, including generalized linear models, mixed effects models, mixture regression models, Bayesian methods, graphical models, and methods for identifying causal effects. The recent period is substantially integrating SEM with the broader statistical literature, which—as the chapters of this volume demonstrate—is making SEM an even more exciting and vibrant tool for the social sciences.

ACKNOWLEDGMENTS

During the preparation of this chapter, I received support from the National Institute on Drug Abuse (1 R01 DA018148-01A), the National Science Foundation (SES-0966662), and the Blumstein–Jordan Endowed Professorship in Sociology. I thank Maria Grigoryeva for assistance and comments, Jerry Herting for his lively conversations and SEM books, and Adrian Raftery, Michael Sobel, Lowell Hargens, Ken Bollen, David Gow, and Richard Callahan for comments on an earlier draft. All remaining errors are mine.

NOTES

1. Stock and Trebbi (2003) conduct a stylometric (grammar and writing style) analysis using principal components and discriminant analysis that points to Phillip as the writer. Because one solution used path analysis, it is likely that the two Wrights collaborated—they had done so earlier—although Stock and Trebbi note that Phillip did not mention his son Sewall in his acknowledgments, as he had done in earlier publications.
2. In an essay titled "What If?" Duncan (1992) wondered whether, if he had sooner convinced Blalock that Sewall Wright's path analysis could solve Simon's questions, Blalock's appen-

dix would have been more complete and Duncan would have lost motivation to write his 1966 article.

3. A third tetrad difference, $r_{12}r_{34} - r_{13}r_{24} = 0$, is a function of the first two and will always be true if the first two are true.
4. Anderson (1991) notes that economists like to put restrictions on models, whereas psychologists refused to do so until "Jöreskog (1969) came up with the catchy terms 'exploratory' and 'confirmatory factor analysis' " with zero-restrictions on loading matrices, along with computer programs for maximum likelihood estimation. Jöreskog certainly popularized the terms, but the distinction was originally due to Tucker (1955).
5. This recursive causal chain model was later formalized and extended by Wold (1964), who criticized simultaneous equation models for ignoring the presumption that causality proceeds through time and is best modeled by recursive causal chains.
6. Koopmans (1937) appears to be the first to argue that residual variance in structural models was due not solely to measurement error—which implies deterministic relationships in the absence of such errors—but also to omitting numerous minor variables from the model (see Epstein, 1987, p. 55).
7. For a discussion of the concept of exogeneity, and of super-exogeneity, see Engle, Hendry, and Richard (1983).
8. While at Cowles, Rubin had been attending lectures by Thurstone and raised identification issue in the context of the factor-analytic model. Anderson and Rubin (1956) concluded that the identification problems of simultaneous equation and factor analysis models were identical, which led to their treatise on maximum likelihood factor analysis (see Anderson, 1991).
9. For excellent discussions of the history of simultaneous equation models in economics, see Epstein (1987) and especially Morgan (1990).
10. For examples of recent returns to limited information methods to obtain estimates more robust to specification errors, see Bollen (1996) and Bollen, Kirby, Curran, Paxton, and Chen (2007).
11. For a lucid discussion of various ways that econometricians have approached measurement error, see Judge et al. (1980, Chap. 13).
12. The term "walking-dog model" originated with Beverly Duncan, who noted that the path diagram (see Figure 2.1) resembled a walking dog, in which η_1 here depicted the dog's face and ζ its tail (Hauser, personal communication).
13. Hauser and Goldberger (1971) also show that in the special case in which all tetrad-difference constraints are satisfied, such as when measurement errors are mutually correlated, modified GLS (GLS with unknown error variances) is equivalent to ML.
14. Wold's (1982) partial least squares "soft modeling" approach to causal chain models provides "instant" estimates under arbitrary distributions but do not necessarily have desirable statistical properties.
15. Note that $m_{ghij} = 1/n\Sigma(x_g - \bar{x}_g)\,(x_h - \bar{x}_h)\,(x_i - \bar{x}_i)\,(x_j - \bar{x}_j)$ is a sample estimator of $\sigma_{ghij} = 1/n\Sigma[x_g - E(x_g)][x_h - E(x_h)][x_i - E(x_i)][x_j - E(x_j)]$.
16. Pearson's tetrachoric correlation, involving pairs of dichotomous variables, led to his famous quarrel with Yule (1912), who argued that many binary outcomes, such as death, cannot be viewed as reflections of a continuous scale, and proposed his Q-coefficient instead (see Pearson & Heron, 1913).
17. A feature of Wold's (1982) soft modeling approach is the further development of a model through a "dialogue between the investigator and the computer."
18. The counterfactual approach to causality is just one of many potential approaches. For a critique of the counterfactual approach in statistics—with a lively discussion from leading statisticians—see Dawid (2000).
19. For a lucid response to Freedman's (1987) critique of Blau-Duncan, see Glymour et al. (1987), and for a lively description of Duncan's personal correspondence with Freedman, see Xie (2007).
20. Rubin (1980) terms this the "stable unit treatment value assumption" (SUTVA); Heckman (2010) has pointed out that Hurwicz (1962) included this assumption under the concept of invariance.

REFERENCES

Aldrich, J. (1989). Autonomy. *Oxford Economic Papers, 41*, 15–34.

Alwin, D. F., & Hauser, R. M. (1975). The decomposition of effects in path analysis. *American Sociological Review, 40*, 37–47.

Anderson, T. W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Annals of Statistics, 1*, 135–141.

Anderson, T. W. (1991). Trygve Haavelmo and simultaneous equation models. *Scandinavian Journal of Statistics, 18*, 1–19.

Anderso[illegible] Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics, 20*, 46–63.

Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. In J. Neyman (Ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* (pp. 111–150). Berkeley: University of California Press.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables (with discussion). *Journal of the American Statistical Association, 91*, 444–472.

Arminger, G., & Muthén, B. O. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sam-

This should be 2.2

pler and the Metropolis–Hastings algorithm. *Psychometrika, 63*, 271–300.

Bagozzi, R. P. (1980). *Causal models in marketing.* New York: Wiley.

Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology, 31*, 419–456.

Bentler, P. M. (1986). Structural modeling and *Psychometrika*: An historical perspective on growth and achievement. *Psychometrika, 51*, 35–51.

Bentler, P. M. (1995). *EQS program manual.* Encino, CA: Multivariate Software.

Bentler, P. M., & Speckart, G. (1981). Attitudes "cause" behaviors: A structural equation perspective. *Journal of Personality and Social Psychology, 40*, 226–238.

Bielby, W. T., & Hauser, R. M. (1977). Structural equation models. *Annual Review of Sociology, 3*, 137–161.

Bielby, W. T., Hauser, R. M., & Featherman, D. L. (1977). Response errors of black and nonblack males in models of the intergenerational transmission of socioeconomic status. *American Journal of Sociology, 82*, 1242–1288.

Blalock, H. M., Jr. (1961a). *Causal inferences in nonexperimental research.* New York: Norton.

Blalock, H. M., Jr. (1961b). Correlation and causality: The multivariate case. *Social Forces, 39*, 246–251.

Blalock, H. M., Jr. (1962). Four-variable causal models and partial correlations. *American Journal of Sociology, 68*, 182–194.

Blalock, H. M., Jr. (1969). Multiple indicators and the causal approach to measurement error. *American Journal of Sociology, 75*, 264–273.

Blalock, H. M., Jr. (Ed.). (1971). *Causal models in the social sciences* Chicago: Aldine-Atherton.

Blau, P. M., & Duncan, O. D. (1967). *The American occupational structure.* New York: Wiley.

Bock, R. D., & Lieberman, M. (1970). Fitting a response model for *n* dichotomously scored items. *Psychometrika, 35*, 179–197.

Bollen, K. A. (1987). Total, direct, and indirect effects in structural equation models. *Sociological Methodology, 17*, 37–69.

Bollen, K. A. (1989). *Introduction to structural equation models with latent variables.* New York: Wiley.

Bollen, K. A. (1990). Outlier screening and a distribution-free test for vanishing tetrads. *Sociological Methods and Research, 19*, 80–92.

Bollen, K. A. (1996). An alternative 2SLS estimator for latent variable models. *Psychometrika, 61*, 109–121.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective.* New York: Wiley.

Bollen, K. A., Kirby, J. B., Curran, P. J., Paxton, P. M., & Chen, F. (2007). Latent variable models under misspecification: Two stage least squares (2SLS) and maximum likelihood (ML) estimators. *Sociological Methods and Research, 36*, 46–86.

Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin, 110*, 305–314.

Bollen, K. A., & Stein, R. A. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociological Methodology, 20*, 115–140.

Bollen, K. A., & Stein, R. A. (1993). Bootstrapping goodness-of-fit measures in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 111–135). Newbury Park, CA: Sage.

Bollen, K. A., & Ting, K.-F. (1993). Confirmatory tetrad analysis. *Sociological Methodology, 23*, 147–176.

Boomsma, A. (1983). *On the robustness of LISREL (maximum likelihood estimation) against small sample size and non-normality.* Unpublished PhD dissertation, University of Groningen, Groningen.

Boudon, R. (1965). A method of linear causal analysis: Dependence analysis. *American Sociological Review, 30*, 365–374.

Browne, M. W. (1974). Generalized least squares estimators in the analysis of covariance matrices with linear structure. *South African Statistical Journal, 8*, 1–24.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37*, 62–83.

Browne, M. W. (2000). Psychometrics. *Journal of the American Statistical Association, 95*, 661–665.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J. B., & Paxton, P. M. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistics in structural equation models. *Sociological Methods and Research, 36*, 462–494.

Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika, 40*, 5–32.

Connell, A., & Frye, A. A. (2006). Growth mixture modelling in developmental psychology: Overview and demonstration of heterogeneity in developmental trajectories of adolescent antisocial behaviour. *Infant and Child Development, 15*, 609–621.

Costner, H. L. (1969). Theory, deduction, and rules of correspondence. *American Journal of Sociology, 75*, 245–263.

Cowles Commission. (1952). *Economic theory and measurement: A twenty year research report, 1932–1952.* Chicago: Author.

Crow, J. F. (1988). Sewall Wright: 1889–1988. *Genetics, 119*, 1–4.

Dawid, A. P. (2000). Causal inference without counterfactu-

als. *Journal of the American Statistical Association, 95*, 407–424.

Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature, 48*, 424–455.

Duncan, O. D. (1966). Path analysis: Sociological examples. *American Journal of Sociology, 72*, 1–16.

Duncan, O. D. (1975). *Introduction to structural equation models*. New York: Academic Press.

Duncan, O. D. (1992). What if? *Contemporary Sociology, 21*, 667–668.

Duncan, O. D., & Hodge, R. W. (1963). Education and occupational mobility: A regression analysis. *American Journal of Sociology, 68*, 629–644.

Emsley, R., Dunn, G., & White, I. R. (2010). Mediation and moderation of treatment effects in randomised controlled trials of complex interventions. *Statistical Methods in Medical Research, 19*, 237–270.

Engle, R., Hendry, D., & Richard, J. (1983). Exogeneity. *Econometrica, 51*, 277–304.

Epstein, R. J. (1987). *A history of econometrics*. Amsterdam: North Holland.

Freedman, D. A. (1987). As others see us: A case study in path analysis (with discussion). *Journal of Educational Statistics, 12*, 101–223.

Frisch, R. (1934). *Statistical confluence analysis by means of complete regression systems*. Oslo: Universitetets Økonomiske Institutt.

Frisch, R., & Waugh, F. (1933). Partial time regressions as compared with individual trends. *Econometrica, 1*, 387–401.

Glymour, C., & Cooper, G. (Eds.). (1999). *Computation, causation, and discovery*. Menlo Park, CA: AAAI/MIT Press.

Glymour, C., Scheines, R., Spirtes, P., & Kelly, K. (1987). *Discovering causal structure: Artificial intelligence, philosophy of science, and statistical modeling*. Orlando, FL: Academic Press.

Glymour, C., Spirtes, P., & Richardson, T. (1999). Response to rejoinder. In C. Glymour & G. F. Cooper (Eds.), *Computation, causation, and discovery* (pp. 343–345). Menlo Park, CA: AAAI/MIT Press.

Goldberger, A. S. (1972a). Maximum-likelihood estimation of regressions containing unobservable independent variables. *International Economic Review, 13*, 1–15.

Goldberger, A. S. (1972b). Structural equation methods in the social sciences. *Econometrica, 40*, 979–1001.

Goldberger, A. S. (1973). Efficient estimation in overidentified models: An interpretive analysis. In A. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 131–152). New York: Academic Press.

Goldberger, A. S. (1991). *A course in econometrics*. Cambridge, MA: Harvard University Press.

Goldberger, A. S., & Duncan, O. D. (Eds.). (1973). *Structural equation models in the social sciences*. New York: Academic Press.

Goldstein, H. (1995). *Multilevel statistical models*. London: Arnold.

Goldstein, H., & Browne, W. J. (2002). Multilevel factor analysis modelling using Markov Chain Monte Carlo (MCMC) estimation. In G. Marcoulides & I. Moustaki (Eds.), *Latent variable and latent structure models* (pp. 225–243). Englewood Cliffs, NJ: Erlbaum.

Goldstein, H., & McDonald, R. P. (1988). A general model for the analysis of multi-level data. *Psychometrika, 53*, 455–467.

Greene, W. H. (2003). *Econometric analysis* (5th ed.). Upper Saddle River, NJ: Prentice-Hall.

Greenland, S., & Brumback, B. (2002). An overview of relations among causal modelling methods. *Journal of International Epidemiology, 31*, 1030–1037.

Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal diagrams for epidemiologic research. *Epidemiology, 10*, 37–48.

Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica, 11*, 1–12.

Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica, 12*(Suppl.), 1–114.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica, 50*, 1029–1054.

Hargens, L. L., Reskin, B. F., & Allison, P. D. (1976). Problems in estimating measurement error from panel data: An example involving the measurement of scientific productivity. *Sociological Methods and Research, 5*, 247–256.

Harman, H. (1960). *Modern factor analysis*. Chicago: University of Chicago Press.

Hauser, R. M. (1973). Disaggregating a social-psychological model of educational attainment. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 255–284). New York: Academic Press.

Hauser, R. M., & Goldberger, A. S. (1971). The treatment of unobservable variables in path analysis. In H. L. Costner (Ed.), *Sociological methodology 1971* (pp. 81–87). San Francisco: Jossey-Bass.

Heckman, J. J. (1992). Haavelmo and the birth of modern econometrics: A review of the history of econometric ideas by Mary Morgan. *Journal of Economic Literature, 30*, 876–886.

Heckman, J. J. (2000). Causal parameters and policy analysis in economics: A twentieth century retrospective. *Quarterly Journal of Economics, 115*, 45–97.

Heckman, J. J. (2010). Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of Economic Literature, 48*, 356–398.

Heckman, J. J., & Urzúa, S. (2010). Comparing IV with structural models: What simple IV can and cannot identify. *Journal of Econometrics, 156*, 27–37.

Hendry, D. F., & Morgan, M. S. (1989). A re-analysis of confluence analysis. *Oxford Economic Papers, 41*, 35–52.

Hill, W. G. (1995). Sewell Wright's system of mating. *Genetics, 143*, 1499–1506.

Holland, P. (1988). Causal inference and path analysis. In C. C. Clogg (Ed.), *Sociological methodology 1988* (pp. 449–484). Washington, DC: American Sociological Association.

Holzinger, K. J. (1941). *Factor analysis*. Chicago: University of Chicago Press.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24*, 417–41, 498–520.

Humphreys, P., & Freedman, D. (1996). The grand leap. *British Journal for the Philosophy of Science, 47*, 113–123.

Hurwicz, L. (1962). On the structural form and interdependent systems. In E. Nagel, P. Suppes, & A. Tarski (Eds.), *Logic, methodology and philosophy of science* (pp. 232–239). Stanford, CA: Stanford University Press.

Imbens, G. W. (2010). Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009). *Journal of Economic Literature, 48*, 399–423.

Imbens, G. W., & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica, 62*, 467–475.

Imbens, G. W., & Rubin, D. B. (1997). Estimating outcome distributions for compliers in instrumental variables models. *Review of Economic Studies, 64*, 555–574.

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature, 47*, 5–86.

Jeffreys, H. (1935). Some tests of significance, treated by the theory of probability. *Proceedings of the Cambridge Philosophical Society, 31*, 203–222.

Jennrich, R. I., & Clarkson, D. B. (1980). A feasible method for standard errors of estimate in maximum likelihood factor analysis. *Psychometrika, 45*, 237–247.

Jennrich, R. I., & Robinson, S. M. (1969). A Newton–Raphson algorithm for maximum likelihood factor analysis. *Psychometrika, 34*, 111–123.

Jennrich, R. I., & Sampson, P. F. (1966). Rotation for simple loadings. *Psychometrika, 31*, 313–323.

Jo, B., & Muthén, B. O. (2002). Longitudinal studies with intervention and noncompliance: Estimation of causal effects in growth mixture modeling. In S. P. Reise & N. Duan (Eds.), *Multilevel modeling: Methodological advances, issues, and applications* (pp. 71–98). Mahwah, NJ: Erlbaum.

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika, 32*, 443–482.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34*, 183–202.

Jöreskog, K. G. (1970). A general method for analysis of covarriance structures. *Biometrika, 57*, 239–251.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika, 36*, 109–133.

Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85–112). New York: Academic Press.

Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika, 36*, 109–133.

Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association, 70*, 631–639.

Jöreskog, K. G., & Sörbom, D. (2001). *LISREL 8 user's reference guide*. Chicago: Scientific Software International.

Jöreskog, K. G., & Sörbom, D. (2002). *PRELIS 2 user's reference guide*. Chicago: Scientific Software International.

Jöreskog, K. G., & Yang, F. (1996). Nonlinear structural equation models: The Kenny–Judd model with interaction effects. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 57–88). Mahwah, NJ: Erlbaum.

Judge, G. G., Griffiths, W. E., Hill, R. C., & Lee, T.-C. (1980). *The theory and practice of econometrics*. New York: Wiley.

Kaiser, H. F. (1958). The Varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23*, 187–200.

Keesling, J. W. (1972). *Maximum likelihood approaches to causal flow analysis*. Unpublished PhD dissertation, Department of Education, University of Chicago, Chicago.

Kendall, M., & Stuart, A. (1979). *The advanced theory of statistics, Vol. 3: Inference and relationship*. London: Griffin.

Kenny, D., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin, 96*, 201–210.

Kiiveri, H., & Speed, T. P. (1982). Structural analysis of multivariate data: A review. In S. Leinhardt (Ed.), *Sociological methodology 1982* (pp. 209–289). San Francisco: Jossey-Bass.

Klein, L. (1950). *Economic fluctuations in the United States 1921–1941*. New York: Wiley.

Klein, L., & Goldberger, A. S. (1955). *An econometric model of the United States 1929–1952*. Amsterdam: North Holland.

Kohn, M., & Schooler, C. (1982). Job conditions and personality: A longitudinal assessment of their reciprocal effects. *American Journal of Sociology, 87*, 1257–1286.

Koopmans, T. (1937). *Linear regression analysis of economic time series* (Netherlands Economic Institute, Publication No. 20). Haarlem: F. Bohn.

Koopmans, T. (1945). Statistical estimation of simultaneous economic relations. *Journal of the American Statistical Association, 40*, 488–466.

Koopmans, T., Rubin, H., & Leipnik, R. (1950). Measuring the equation systems of dynamic economics (Cowles Com-

mission Monograph 10). In T. Koopmans (Ed.), *Statistical inference in dynamic economic models* (pp. 53–237). New York: Wiley.

Lawley, D. N. (1940). The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh, 60*, 64–82.

Lee, S.-Y. (1981). A Bayesian approach to confirmatory factor analysis. *Psychometrika, 46*, 153–160.

Li, C. C. (1956). The concept of path coefficient and its impact on population genetics. *Biometrics, 12*, 190–210.

Li, C. C. (1975). *Path analysis: A primer.* Pacific Grove, CA: Boxworth Press.

Lucas, R. (1976). Econometric policy analysis: A critique. In K. Brunner & A. Meltzer (Eds.), *The Phillips Curve and labor markets* (pp. 19–46). Amsterdam: North Holland.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*, 130–149.

Magnus, J. R., & Morgan, M. S. (1987). The ET Interview: Professor J. Tinbergen. *Econometric Theory, 3*, 117–142.

Mann, H. B., & Wald, A. (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica, 11*, 173–220.

Matsueda, R. L. (1982). Testing control theory and differential association: A causal modeling approach. *American Sociological Review, 47*, 489–504.

Matsueda, R. L., & Bielby, W. T. (1986). Statistical power in covariance structure models. In N. B. Tuma (Ed.), *Sociological methodology 1986* (pp. 120–158). Washington, DC: American Sociological Association.

Matsueda, R. L., & Heimer, K. (1987). Race, family structure, and delinquency: A test of differential association and social control theories. *American Sociological Review, 52*, 826–840.

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

Morgan, M. S. (1990). *The history of econometric ideas.* Cambridge, UK: Cambridge University Press.

Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research.* Cambridge, UK: Cambridge University Press.

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika, 43*, 551–560.

Muthén, B. (1979). A structural probit model with latent variables. *Journal of the American Statistical Association, 74*, 807–811.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49*, 115–132.

Muthén, B. (1989). Tobit factor analysis. *British Journal of Mathematical and Statistical Psychology, 42*, 241–250.

Muthén, B. (1994). Multi-level covariance structure analysis. *Sociological Methods and Research, 22*, 376–398.

Muthén, B. (1997). Modeling of longitudinal and multi-level data. In A. E. Raftery (Ed.), *Sociological methodology* (Vol. 27, pp. 453–480). Boston: Blackwell.

Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345–368). Newbury Park, CA: Sage.

Muthén, B. O., & Muthén, L. K. (2004). *Mplus user's guide.* Los Angeles: Authors.

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

Nagin, D. S., & Land, K. C. (1993). Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric mixed Poisson model. *Criminology*, 31, 327–362.

Nagin, D. S., & Tremblay, R. E. (2005). What has been learned from group-based trajectory modeling?: Examples from physical aggression and other problem behaviors. *Annals of the American Academy of Political and Social Science, 602*, 82–117.

Nelson, C. R. (1972). The prediction performance of the FRB-MIT-PENN model of the U.S. economy. *American Economic Review, 62*, 902–917.

Neyman, J., & Pearson, E. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London: Series A, Containing Papers of a Mathematical or Physical Character, 231*, 289–337.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika, 44*, 443–460.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems.* San Mateo, CA: Morgan Kaufmann.

Pearl, J. (2000). *Causality: Models, reasoning, and inference.* Cambridge, UK: Cambridge University Press.

Pearl, J., & Verma, T. (1991). A theory of inferred causation. In J. A. Allen, R. Fikes, & E. Sandewall (Eds.), *Principles of knowledge representation and reasoning: Proceedings of the Second International Conference* (pp. 441–452). San Francisco: Morgan Kaufmann.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine, 6*, 559–572.

Pearson, K., & Heron, D. (1913). On theories of association. *Biometrika, 9*, 159–315.

Poon, W.-Y., & Lee, S.-Y. (1987). Maximum likelihood estimation of multivariate polyserial and polychoric correlation coefficients. *Psychometrika, 52*, 429–430.

Qin, D. (1989). Formalization of identification theory. *Oxford Economic Papers, 41*, 73–93.

Rabe-Hesketh, S., Pickles, A., & Skrondal, A. (2001). *GLLAMM manual* (Tech. Rept. 2001/01, Department of Biostatistics and Computing, Institute of Psychiatry, Kings

College, University of London). Downloadable at *www.gllamm.org*.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika, 69*, 167–190.

Raftery, A. E. (1993). Bayesian model selection in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 163–180). Newbury Park, CA: Sage.

Raftery, A. E. (1995). Bayesian model selection in social research. In P. V. Marsden (Ed.), *Sociological methodology 1995* (pp. 111–165). Washington, DC: American Sociological Association.

Rao, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrika, 51*, 83–90.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models*. Thousand Oaks, CA: Sage.

Raudenbush, S. W., & Sampson, R. J. (1999). Assessing direct and indirect associations in multi-level designs with latent variables. *Sociological Methods and Research, 28*, 123–153.

Robins, J. M. (1986). A new approach to causal inference in mortality studies with a sustained exposure period: Application to control of healthy worker survivor effect. *Mathematical Modeling, 7*, 1393–1512.

Robins, J. M. (1995). Comment on Judea Pearl's paper, "Causal Diagrams for Empirical Research." *Biometrika, 82*, 695–698.

Robins, J. M. (2003). Semantics of causal DAG models and the identification of direct and indirect effects. In P. J. Green, N. L. Hjort, & S. Richardson (Eds.), *Highly structured stochastic systems* (pp. 70–81). Oxford, UK: Oxford University Press.

Robins, J. M., & Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology, 3*, 143–155.

Robins, J. M., & Wasserman, L. (1999). On the impossibility of inferring causation from association without background knowledge. In C. Glymour & G. F. Cooper (Eds.), *Computation, causation, and discovery* (pp. 305–321). Menlo Park, CA: MIT Press.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers, 3*, 135–146.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66*, 688–701.

Rubin, D. B. (1980). Comment on "Randomization Analysis of Experimental Data: The Fisher Randomization Test," by D. Basu. *Journal of the American Statistical Association, 75*, 591–593.

Satorra, A., & Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika, 50*, 83–90.

Scheines, R., Hoijtink, H., & Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika, 64*, 37–52.

Scheines, R., Spirtes, P., Glymour, C., Meek, C., & Richardson, T. (1997). *Tetrad 3 user's manual*. Pittsburgh, PA: Department of Philosophy, Carnegie Mellon University.

Schultz, H. (1938). *The theory and measurement of demand*. Chicago: University of Chicago Press.

Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461–464.

Sewell, W. H., & Hauser, R. M. (1975). *Education, opportunity, and earnings: Achievement in the early career*. New York: Academic Press.

Shipley, B. (2000). *Cause and correlation in biology: A user's guide to path analysis, structural equations and causal inference*. Cambridge, UK: Cambridge University Press.

Simmons, R., & Blyth, D. A. (1987). *Moving into adolescence: The impact of pubertal change and school context*. New York: Aldine de Gruyter.

Simon, H. A. (1954). Spurious correlation: A causal interpretation. *Journal of the American Statistical Association, 49*, 467–479.

Sims, C. (1980). Macroeconomics and reality. *Econometrica, 48*, 1–45.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Boca Raton, FL: Chapman & Hall.

Slutsky, E. (1937). The summation of random causes as the source of cyclic processes. *Econometrica, 5*, 105–146.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological methodology 1982* (pp. 290–313). San Francisco: Jossey-Bass.

Sobel, M. E. (1986). Some new results on indirect effects and their standard errors in covariance structure models. In N. B. Tuma (Ed.), *Sociological methodology 1986* (pp. 159–186). Washington, DC: American Sociological Association.

Sobel, M. E. (1992). The American occupational structure and structural equation modeling in sociology. *Contemporary Sociology, 21*, 662–666.

Sobel, M. E. (2008). Identification of causal parameters in randomized studies with mediating variables. *Journal of Educational and Behavioral Statistics, 33*, 230–251.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology, 15*, 201–293.

Spearman, C. (1927). *The abilities of man*. New York: Macmillan.

Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction, and search*. New York: Springer-Verlag.

Spirtes, P., Richardson, T., Meek, C., Scheines, R., & Glymour, C. (1998). Using path diagrams as a structural equa-

tion modeling tool. *Sociological Methods and Research, 27*, 182–225.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*, 173–180.

Steiger, J. H., & Lind, J. C. (1980). *Statistically-based tests for the number of common factors*. Handout for a talk at the annual meetings of the Psychometric Society, Iowa City, IA.

Stock, J. H., & Trebbi, F. (2003). Who invented instrumental variable regression? *Journal of Economic Perspectives, 17*, 177–194.

Theil, H. (1992). Estimation and simultaneous correlation in complete equation systems. In R. Baldev & J. Koerts (Eds.), *Henri Theil's contributions to economics and econometrics* (Chapter 6). Dordrecht: Kluwer. (Original work published 1953)

Thurstone, L. L. (1935). *The vectors of mind*. Chicago: University of Chicago Press.

Tinbergen, J. (1939). *Statistical testing of business cycle theories* (2 vols.). Geneva: League of Nations.

Tucker, L. R. (1955). The objective definition of simple structure in linear factor analysis. *Psychometrika, 20*, 209–225.

Tucker, L. R. (1958). Determination of parameters of a functional relation by factor analysis. *Psychometrika, 23*, 19–23.

Wermuth, N., & Lauritsen, S. L. (1983). Graphical and recursive models for contingency tables. *Biometrika, 70*, 537–552.

Wheaton, B. (1978). The sociogenesis of psychological disorder. *American Sociological Review, 43*, 383–403.

Wheaton, B. (1985). Models for the stress-buffering functions of coping resources. *Journal of Health and Social Behavior, 26*, 352–364.

Wiley, D. E. (1973). The identification problem for structural equation models with unmeasured variables. In A. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 69–84). New York: Academic Press.

Winship, C., & Mare, R. D. (1983). Structural equations and path analysis for discrete data. *American Journal of Sociology, 89*, 54–110.

Winship, C., & Mare, R. D. (1984). Regression models with ordinal variables. *American Sociological Review, 49*, 512–525.

Wold, H. (1964). *Econometric model building: Essays on the causal chain approach*. Amsterdam: North Holland.

Wold, H. (1982). Soft modeling: The basic design and some extensions. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation* (pp. 1–54). Amsterdam: North Holland.

Wright, P. G. (1915). Moore's economic cycles. *Quarterly Journal of Economics, 29*, 631–641.

Wright, P. G. (1928). *The tariff on animal and vegetable oils*. New York: Macmillan.

Wright, S. (1918). On the nature of size factors. *Genetics, 3*, 367–374.

Wright, S. (1920). The relative importance of heredity and environment in determining the piebald pattern of guinea pigs. *Proceedings of the National Academy of Sciences, 6*, 320–332.

Wright, S. (1921a). Correlation and causation. *Journal of Agricultural Research, 20*, 557–585.

Wright, S. (1921b). Systems of mating: I. The biometric relations between parent and offspring. *Genetics, 6*, 111–123.

Wright, S. (1925). *Corn and hog correlations* (Bulletin 1300). Washington, DC: U.S. Department of Agriculture.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics, 5*, 161–215.

Xie, Y. (2007). Otis Dudley Duncan's legacy: The demographic approach to quantitative reasoning in social science. *Research in Social Stratification and Mobility, 25*, 141–156.

Yang, M., & Dunson, D. B. (2010). Bayesian semiparametric structural equation models with latent variables. *Psychometrika, 75*, 675–693.

Yuan, K.-H., & Bentler, P. M. (2007). Structural equation modeling. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics: Vol. 26. Psychometrics* (pp. 297–358). Amsterdam: North Holland.

Yule, G. U. (1896). On the correlation of total pauperism with proportion of out-relief: II. Males over 65. *Economic Journal, 6*, 613–623.

Yule, G. U. (1912). On the methods of measuring association between two attributes. *Journal of the Royal Statistical Society, 75*, 579–652.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests of aggregation bias. *Journal of the American Statistical Association, 57*, 348–368.

Zellner, A. (1970). Estimation of regression relationships containing unobservable independent variables. *International Economic Review, 11*, 441–454.

Zellner, A. (1971). *Introduction to Bayesian inference in econometrics*. New York: Wiley.

Zellner, A., & Theil, H. (1962). Three-stage least squares: Simultaneous estimation of simultaneous equations. *Econometrica, 30*, 54–78.