KENNETH A. BOLLEN

# STRUCTURAL EQUATIONS WITH LATENT VARIABLES

WILEY SERIES IN PROBABILITY AND MATHEMATICAL STATISTICS

WILEY

Structural Equations
with Latent Variables

# Structural Equations with Latent Variables

KENNETH A. BOLLEN

Department of Sociology
The University of North Carolina at Chapel Hill
Chapel Hill, North Carolina

WILEY

A Wiley-Interscience Publication

JOHN WILEY & SONS

New York • Chichester • Brisbane • Toronto • Singapore

To Barbara
and to my parents



***Library of Congress Cataloging in Publication Data:***

Bollen, Kenneth A.
Structural equations with latent variables/Kenneth A. Bollen.
p. cm.—(Wiley series in probability and mathematical statistics. Applied probability and statistics section, ISSN 0271-6356)
"A Wiley-Interscience publication."
Bibliography: p.
Includes index.
1. Social sciences—Statistical methods. 2. Latent variables.
I. Title. II. Series.
HA29.B732 1989
300'.29—dc 19 88-27272
ISBN 0-471-01171-1
CIP

Printed in the United States of America

**30 29 28 27**

# Preface

Within the past decade the vocabulary of quantitative research in the social sciences has changed. Terms such as "LISREL," "covariance structures," "latent variables," "multiple indicators," and "path models" are commonplace. The structural equation models that lie behind these terms are a powerful generalization of earlier statistical approaches. They are changing researchers' perspectives on statistical modeling and building bridges between the way social scientists think substantively and the way they analyze data.

We can view these models in several ways. They are regression equations with less restrictive assumptions that allow measurement error in the explanatory as well as the dependent variables. They consist of factor analyses that permit direct and indirect effects between factors. They routinely include multiple indicators and latent variables. In brief, these models encompass and extend regression, econometric, and factor analysis procedures.

The book provides a comprehensive introduction to the general structural equation system, commonly known as the "LISREL model." One purpose of the book is to demonstrate the generality of this model. Rather than treating path analysis, recursive and nonrecursive models, classical econometrics, and confirmatory factor analysis as distinct and unique, I treat them as special cases of a common model. Another goal of the book is to emphasize the application of these techniques. Empirical examples appear throughout. To gain practice with the procedures, I encourage the reader to reestimate the examples, and then to devise and estimate new models. Several chapters contain some of the LISREL or EQS programs I used to obtain the results for the empirical examples. I have kept the examples as realistic as possible. This means that some of the initial specifications do *not* fit well. Through my experiences with students, colleagues, and in my own work, I frequently have found that the beginning model does not adequately describe the data. Respecification is often necessary. I note the difficulties this creates in proper interpretations of significance tests and the added importance of replication.

A final purpose is to emphasize the crucial role played by substantive expertise in most stages of the modeling process. Structural equation models are not very helpful if you have little idea about the subject matter. To begin the fitting process, the analysts must draw upon their knowledge to construct a multiequation system that specifies the relations between all latent variables, disturbances, and indicators. Furthermore they must turn to substantive information when respecifying models and when evaluating the final model. Empirical results can reveal that initial ideas are in error or they can suggest ways to modify a model, but they are given meaning only within the context of a substantively informed model.

Structural equation models can be presented in two ways. One is to start with the general model and then show its specializations to simpler models. The other is to begin with the simpler models and to build toward the general model. I have chosen the latter strategy. I start with the regression/econometric and factor analysis models and present them from the perspective of the general model. This has the advantage of gradually including new material while having types of models with which the reader is somewhat familiar. It also encourages viewing old techniques in a new light and shows the often unrealistic assumptions implicit in standard regression/econometric and factor analyses.

Specifically, I have organized the book as follows. Chapter 2 introduces several methodological tools. I present the model notation, covariances and covariance algebra, and a more detailed account of path analysis. Appendixes A and B at the end of the book provide reviews of matrix algebra and of asymptotic distribution theory. Chapter 3 addresses the issue of causality. Implicitly, the idea of causality pervades much of the structural equation writings. The meaning of causality is subject to much controversy. I raise some of the issues behind the controversy and present a structural equation perspective on the meaning of causality.

The regression/econometric models for observed variables are the subject of Chapter 4. Though many readers have experience with these, the covariance structure viewpoint will be new to many. The consequences of random measurement error in observed variable models is the topic of Chapter 5. The chapter shows why and when we should care about measurement error in observed variables.

Once we recognize that variables are measured with error, we need to consider the relation between the error-free variable and the observed variable. Chapter 6 is an examination of this relation. It introduces procedures for developing measures and explores the concepts of reliability and validity. Chapter 7 is on confirmatory factor analysis, which is used for estimating measurement models such as those in Chapter 6.

Finally, Chapters 8 and 9 provide the general structural equation model with latent variables. Chapter 8 emphasizes the "basics," whereas Chapter 9 treats more advanced topics such as arbitrary distribution estimators and the treatment of categorical observed variables.

The main motivation for writing this book arose from my experiences teaching at the Interuniversity Consortium for Political and Social Research (ICPSR) Summer Training Program in Methodology at the University of Michigan (1980–1988). I could not find a text suitable for graduate students and professionals with training in different disciplines. Nor could I find a comprehensive introduction to these procedures. I have written the book for social scientists, market researchers, applied statisticians, and other analysts who plan to use structural equation or LISREL models. I assume that readers have prior exposure and experience with matrix algebra and regression analysis. A background in factor analysis is helpful but not essential. Jöreskog and Sörbom's (1986) LISREL and Bentler's (1985) EQS are the two most popular structural equation software packages. I make frequent reference to them, but the ideas of the book extend beyond any specific program.

I have many people to thank for help in preparing this book. The Interuniversity Consortium for Political and Social Research (ICPSR) at the University of Michigan (Ann Arbor) has made it possible for me to teach these techniques for the last nine years in their Summer Program in Quantitative Methods. Bob Hoyer started me there. Hank Heitowit and the staff of ICPSR have continued to make it an ideal teaching environment. A number of the hundreds of graduate students, professors, and other professionals who attended the courses provided general and specific comments to improve the book. Gerhard Arminger (University of Wuppertal), Jan de Leeuw (University of Leiden), Raymond Horton (Lehigh University), Frederick Lorenz (Iowa State University), John Fox (York University), Robert Stine (University of Pennsylvania), Boone Turchi (University of North Carolina), and members of the Statistical and Mathematical Sociology Group at the University of North Carolina provided valuable comments on several of the chapters. Barbara Entwisle Bollen read several drafts of most chapters, and her feedback and ideas are reflected throughout the book. Without her encouragement, I do not know when or if I would have completed the book.

Brenda Le Blanc of Dartmouth College and Priscilla Preston and Jenny Neville of the University of North Carolina (Chapel Hill) performed expert typing above and beyond the call of duty. Stephen Birdsall, Jack Kasarda, and Larry Levine smoothed the path several times as I moved the manuscript from Dartmouth College to the University of North Carolina. I thank the Committee for Scholarly Publications, Artistic Exhibitions, and Performances at the University of North Carolina, who provided financial support needed to complete the book.

KENNETH A. BOLLEN

*Chapel Hill, North Carolina*

# Contents

CHAPTER ONE

# Introduction

Most researchers applying statistics think in terms of modeling the *individual observations*. In multiple regression or ANOVA (analysis of variance), for instance, we learn that the regression coefficients or the error variance estimates derive from the minimization of the sum of squared differences of the predicted and observed dependent variable for each case. Residual analyses display discrepancies between fitted and observed values for every member of the sample.

The methods of this book demand a reorientation. The procedures emphasize *covariances* rather than cases.[1] Instead of minimizing functions of observed and predicted individual values, we minimize the difference between the sample covariances and the covariances predicted by the model. The observed covariances minus the predicted covariances form the residuals. The fundamental hypothesis for these structural equation procedures is that the covariance matrix of the observed variables is a function of a set of parameters. If the model were correct and if we knew the parameters, the population covariance matrix would be exactly reproduced. Much of this book is about the equation that formalizes this fundamental hypothesis:

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}) \tag{1.1}$$

In (1.1), $\boldsymbol{\Sigma}$ (sigma) is the population covariance matrix of observed variables, $\boldsymbol{\theta}$ (theta) is a vector that contains the model parameters, and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is

[1]As is clear from several places in the book, individual cases that are outliers can severely affect covariances and estimates of parameters. Thus, with these techniques, researchers still need to check for outliers. In addition, in many cases (e.g., regression models) the minimizations based on individuals and minimizations based on the predicted and observed covariance matrices lead to the same parameter estimates.

the covariance matrix written as a function of $\boldsymbol{\theta}$. The simplicity of this equation is only surpassed by its generality. It provides a unified way of including many of the most widely used statistical techniques in the social sciences. Regression analysis, simultaneous equation systems, confirmatory factor analysis, canonical correlations, panel data analysis, ANOVA, analysis of covariance, and multiple indicator models are special cases of (1.1).

Let me illustrate. In a simple regression equation we have $y = \gamma x + \zeta$, where $\gamma$ (gamma) is the regression coefficient, $\zeta$ (zeta) is the disturbance variable uncorrelated with $x$ and the expected value of $\zeta$, $E(\zeta)$, is zero. The $y$, $x$, and $\zeta$ are random variables. This model in terms of (1.1) is[2]

$$\begin{bmatrix} \text{VAR}(y) & \\ \text{COV}(x, y) & \text{VAR}(x) \end{bmatrix} = \begin{bmatrix} \gamma^2 \text{VAR}(x) + \text{VAR}(\zeta) & \\ \gamma \text{VAR}(x) & \text{VAR}(x) \end{bmatrix} \tag{1.2}$$

where VAR( ) and COV( ) refer to the population variances and covariances of the elements in parentheses. In (1.2) the left-hand side is $\boldsymbol{\Sigma}$, and the right-hand side is $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, with $\boldsymbol{\theta}$ containing $\gamma$, $\text{VAR}(x)$, and $\text{VAR}(\zeta)$ as parameters. The equation implies that each element on the left-hand side equals its corresponding element on the right-hand side. For example, $\text{COV}(x, y) = \gamma \text{VAR}(x)$ and $\text{VAR}(y) = \gamma^2 \text{VAR}(x) + \text{VAR}(\zeta)$. I could modify this example to create a multiple regression by adding explanatory variables, or I could add equations and other variables to make it a simultaneous equations system such as that developed in classical econometrics. Both cases can be represented as special cases of equation (1.1), as I show in Chapter 4.

Instead of a regression model, consider two random variables, $x_1$ and $x_2$, that are indicators of a factor (or latent random variable) called $\xi$ (xi). The dependence of the variables on the factor is $x_1 = \xi + \delta_1$ and $x_2 = \xi + \delta_2$, where $\delta_1$ (delta) and $\delta_2$ are random disturbance terms, uncorrelated with $\xi$ and with each other, and $E(\delta_1) = E(\delta_2) = 0$. Equation (1.1) specializes to

$$\begin{bmatrix} \text{VAR}(x_1) & \\ \text{COV}(x_1, x_2) & \text{VAR}(x_2) \end{bmatrix} = \begin{bmatrix} \phi + \text{VAR}(\delta_1) & \\ \phi & \phi + \text{VAR}(\delta_2) \end{bmatrix} \tag{1.3}$$

where $\phi$ (phi) is the variance of the latent factor $\xi$. Here $\boldsymbol{\theta}$ consists of three elements: $\phi$, $\text{VAR}(\delta_1)$, and $\text{VAR}(\delta_2)$. The covariance matrix of the observed variables is a function of these three parameters. I could add more

[2] Given the symmetric nature of the covariance matrices only the lower half of these matrices is shown.

indicators and more latent factors, allow for coefficients ("factor loadings") relating the observed variables to the factors, and allow correlated disturbances creating an extremely general factor analysis model. As Chapter 7 demonstrates, this is a special case of the covariance structure equation (1.1).

Finally, a simple hybrid of the two preceding cases creates a simple system of equations. The first part is a regression equation of $y = \gamma\xi + \zeta$, where unlike the previous regression the independent random variable is unobserved. The last two equations are identical to the factor analysis example: $x_1 = \xi + \delta_1$ and $x_2 = \xi + \delta_2$. I assume that $\zeta$, $\delta_1$, and $\delta_2$ are uncorrelated with $\xi$ and with each other, and that each has an expected value of zero. The resulting structural equation system is a combination of factor analysis and regression-type models, but it is still a specialization of (1.1):

$$\begin{bmatrix} \mathrm{VAR}(y) & & \\ \mathrm{COV}(x_1, y) & \mathrm{VAR}(x_1) & \\ \mathrm{COV}(x_2, y) & \mathrm{COV}(x_2, x_1) & \mathrm{VAR}(x_2) \end{bmatrix} = \begin{bmatrix} \gamma^2\phi + \mathrm{VAR}(\zeta) & & \\ \gamma\phi & \phi + \mathrm{VAR}(\delta_1) & \\ \gamma\phi & \phi & \phi + \mathrm{VAR}(\delta_2) \end{bmatrix} \tag{1.4}$$

These examples foreshadow the general nature of the models I treat. My emphasis is on systems of *linear equations*. By linear, I mean that the relations between all variables, latent and observed, can be represented in linear structural equations or they can be transformed to linear forms.[3] Structural equations that are nonlinear in the parameters are excluded. Nonlinear functions of parameters are, however, common in the *covariance* structure equation, $\Sigma = \Sigma(\boldsymbol{\theta})$. For instance, the last example had three linear structural equations: $y = \gamma\xi + \zeta$, $x_1 = \xi + \delta_1$, and $x_2 = \xi + \delta_2$. Each is linear in the variables and parameters. Yet the covariance structure (1.4) for this model shows that $\mathrm{COV}(x_1, y) = \gamma\phi$, which means that the $\mathrm{COV}(x_1, y)$ is a nonlinear function of $\gamma$ and $\phi$. Thus it is the structural equations linking the observed, latent, and disturbance variables that are linear, and not necessarily the covariance structure equations.

[3]Though an incredibly broad range of procedures falls under linear equations, those treated here are a special class of possible models of the general moment structure models (see Bentler, 1983).

The term "structural" stands for the assumption that the parameters are not just descriptive measures of association but rather that they reveal an invariant "causal" relation. I will have more to say about the meaning of "causality" with respect to these models in Chapter 3, but for now, let it suffice to say that the techniques do not "discover" causal relations. At best they show whether the causal assumptions embedded in a model match a sample of data. Also the models are for continuous latent and observed variables. The assumption of continuous observed variables is violated frequently in practice. In Chapter 9 I discuss the robustness of the standard procedures and the development of new ones for noncontinuous variables.

Structural equation models draw upon the rich traditions of several disciplines. I provide a brief description of their origins in the next section.

## HISTORICAL BACKGROUND

Who invented general structural equation models? There is no simple answer to this question because many scholars have contributed to their development. The answer to this question is further complicated in that the models continue to unfold, becoming more general and more flexible. However, it is possible to outline various lines of research that have contributed to the evolution of these models.

My review is selective. More comprehensive discussions are available from the perspectives of sociology (Bielby and Hauser 1977), psychology (Bentler 1980; 1986), and economics (Goldberger 1972; Aigner et al. 1984). Two edited collections that represent the multidisciplinary origins of these techniques are the volumes by Goldberger and Duncan (1973) and Blalock ([1971] 1985). Other more recent collections are in Aigner and Goldberg (1977), Jöreskog and Wold (1982), the November 1982 issue of the *Journal of Marketing Research*, and the May–June 1983 issue of the *Journal of Econometrics*.

I begin by identifying three components present in today's general structural equation models: (1) path analysis, (2) the conceptual synthesis of latent variable and measurement models, and (3) general estimation procedures. By tracing the rise of each component, we gain a better idea about the origins of these procedures.

Let me consider path analysis first. The biometrician Sewall Wright (1918, 1921, 1934, 1960) is its inventor. Three aspects of path analysis are the path diagram, the equations relating correlations or covariances to parameters, and the decomposition of effects. The first aspect, the path diagram, is a pictorial representation of a system of simultaneous equations. It shows the relation between all variables, including disturbances and

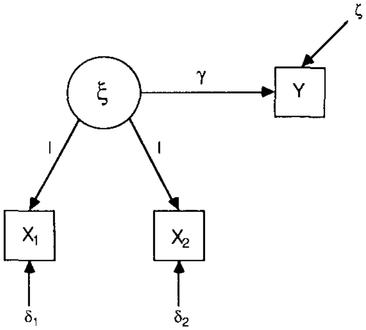


**Figure 1.1** Path Diagram Example

errors. Figure 1.1 gives a path diagram for the last example of the previous section. It corresponds to the equations:

$$y = \gamma\xi + \zeta$$

$$x_1 = \xi + \delta_1$$

$$x_2 = \xi + \delta_2$$

where $\zeta$, $\delta_1$, and $\delta_2$ are uncorrelated with each other and with $\xi$. Straight single-headed arrows represent one-way causal influences from the variable at the arrow base to the variable to which the arrow points. The implicit coefficients of one for the effects of $\xi$ on $x_1$ and $x_2$ are made explicit in the diagram.

Using the path diagram, Wright proposed a set of rules for writing the equations that relate the correlations (or covariances) of variables to the model parameters; this constitutes the second aspect of path analysis. The equations are equivalent to covariance structure equations, an example of which appears in (1.4). He then proposed solving these equations for the unknown parameters, and substituting sample correlations or covariances for their population counterparts to obtain parameter estimates.

The third aspect of path analysis provides a means to distinguish direct, indirect, and total effects of one variable on another. The direct effects are those not mediated by any other variable; the indirect effects operate through at least one intervening variable, and the total effects is the sum of direct and all indirect effects.[4]

Wright's applications of path analysis proved amazing. His first article in 1918 was in contemporary terms a factor analysis in which he formulated

[4] I review path analysis more fully in Chapters 2 and 8.

and estimated a model of the size components of bone measurements. This was developed without knowledge of Spearman's (1904) work on factor analysis (Wright 1954, 15). Unobserved variables also appeared in some of his other applications. Goldberger (1972) credits Wright with pioneering the estimation of supply and demand equations with a treatment of identification and estimation more general than econometricians writing at the same time. His development of equations for covariances of variables in terms of model parameters is the same as that of (1.1), $\Sigma = \Sigma(\theta)$, except that he developed these equations from path diagrams rather than the matrix methods employed today.

With all these accomplishments it is surprising that social scientists and statisticians did not pay more attention to his work. As Bentler (1986) documents, psychometrics only flirted (e.g., Dunlap and Cureton 1930; Englehart 1936) with Wright's path analysis. Goldberger (1972) notes the neglect of econometricians and statisticians with a few exceptions (e.g., Fox 1958; Tukey 1954; Moran 1961; Dempster 1971). Wright's work also was overlooked in sociology until the 1960s. Partially in reaction to work by Simon (1954), Tukey (1954), and Turner and Stevens (1959), sociologists such as Blalock (1961, 1963, 1964), Boudon (1965), and Duncan (1966) saw the potential of path analysis and related "partial-correlation" techniques as a means to analyze nonexperimental data. Following these works, and particularly following Duncan's (1966) expository account, the late 1960s and early 1970s saw many applications of path analysis in the sociological journals. The rediscovery of path analysis in sociology diffused to political science and several other social science disciplines. Stimulated by work in sociology, Werts and Linn (1970) wrote an expository treatment of path analysis, but it was slow to catch on in psychology.

The next major boost to path analysis in the social sciences came when Jöreskog (1973), Keesing (1972), and Wiley (1973), who developed very general structural equation models, incorporated path diagrams and other features of path analysis into their presentations. Researchers know these techniques by the abbreviation of the JKW model (Bentler 1980), or more commonly as the LISREL model. The tremendous popularity of the LISREL model has facilitated the spread of path analysis. Path analysis has evolved over the years. Its present form has some elaboration in the symbols employed in path diagrams, has equations relating covariances to parameters that are derived with matrix operations rather than from "reading" the path diagram, and has a more refined and clearly defined decomposition of direct, indirect, and total effects (see, e.g., Duncan 1971; Alwin and Hauser 1975; Fox 1980; Graff and Schmidt 1982). But the contributions of Wright's work are still clear.

In addition to path analysis, the conceptual synthesis of latent variable and measurement models was essential to contemporary structural equation techniques. The factor analysis tradition spawned by Spearman (1904) emphasized the relation of latent factors to observed variables. The central concern was on what we now call the measurement model. The structural relations between latent variables other than their correlation (or lack of correlation) were not examined. In econometrics the focus was the structural relation between observed variables with an occasional reference to error-in-the-variable situations.

Wright's path analysis examples demonstrated that econometric-type models with variables measured with error could be identified and estimated. The conceptual synthesis of models containing structurally related latent variables and more elaborate measurement models was developed extensively in sociology during the 1960s and early 1970s. For instance, in 1963 Blalock argued that sociologists should use causal models containing both indicators and underlying variables to make inferences about the latent variables based on the covariances of the observed indicators. He suggested that observed variables can be causes or effects of latent variables or observed variables can directly affect each other. He contrasted this with the restrictive implicit assumptions of factor analysis where all indicators are viewed as effects of the latent variable. Duncan, Haller, and Portes (1968) developed a simultaneous equation model of peer influences on high school students' ambitions. The model included two latent variables reciprocally related, multiple indicators of the latent variables, and several background characteristics that directly affected the latent variables. Heise (1969) and others applied path analysis to separate the stability of latent variables from the reliability of measures.

This and related work in sociology during the 1960s and early 1970s demonstrated the potential of synthesizing econometric-type models with latent rather than observed variables and psychometric-type measurement models with indicators linked to latent variables. But their approach was by way of examples; they did not establish a general model that could be applied to any specific problems. It awaited the work of Jöreskog (1973), Keesing (1972), and Wiley (1973) for a practical general model to be proposed. Their models had two parts. The first was a latent variable model that was similar to the simultaneous equation model of econometrics except that all variables were latent ones. The second part was the measurement model that showed indicators as effects of the latent variables as in factor analyses. Matrix expressions for these models were presented so that they could apply to numerous individual problems. Jöreskog and Sörbom's LISREL programs were largely responsible for popularizing these structural equation models, as were the numerous publications and applications of Jöreskog (e.g., 1967, 1970, 1973, 1977, 1978) and his collaborators.

Bentler and Weeks (1980), McArdle and McDonald (1984), and others have proposed alternative representations of general structural equations. Though initially it seemed that these models were more general than the JKW model, most analysts now agree that both the new and old representations are capable of treating the range of linear models that typically occur in practice. I use what has come to be known as the "LISREL notation" throughout the book. To date, it is the most widely accepted representation. I will demonstrate ways to modify it to treat nonstandard applications in several of the chapters.

The last characteristic of the structural equation models are general estimation procedures. The early applications of path analysis by Wright and the sociologists influenced by his work used ad hoc estimation procedures to yield parameter estimates. There was little discussion of statistical inference and optimal ways of combining multiple estimates of a single parameter. Here work from econometrics and psychometrics proved indispensible. In econometrics the properties of estimators for structural equations with observed variables were well established (see, e.g., Goldberger 1964). In psychometrics the work of Lawley (1940), Anderson and Rubin (1956), and Jöreskog (1969) helped lay the foundations for hypothesis testing in factor analysis. Bock and Bargmann (1966) proposed an analysis of covariance structures to estimate the components of variance due to latent variables in multinormal observed variables. Jöreskog (1973) proposed a maximum likelihood estimator (based on the multinormality of the observed variables) for general structural equation models which is today the most widely used estimator. Jöreskog and Goldberger (1972) and Browne (1974, 1982, 1984) suggested generalized least squares (GLS) estimators that offer additional flexibility in the assumptions under which they apply. Browne (1982, 1984), for example, proposed estimators that assume arbitrary distributions or elliptical distributions for the observed variables. Bentler (1983) suggested estimators that treat higher-order product moments of the observed variables. He demonstrated that these moments can help identify model parameters that are not identified by the covariances and the gains in efficiency that may result. Muthén (1984, 1987), among others, has generalized these models to ordinal or limited observed variables.

Finally, this sketch of developments in structural equation models would not be complete without mentioning the computer software that has emerged. As I have already said, Jöreskog and Sörbom's LISREL software perhaps has been the single greatest factor leading to the spread of these techniques throughout the social sciences. It is now entering its seventh version. Bentler's (1985) EQS software has recently entered the field and also is likely to be widely used. McDonald (1980), Schoenberg (1987), and others have written programs with more limited circulations. The dual trend in software is that of providing more general and flexible models and programs that are more "user-friendly." Examples of the former are that LISREL VII and EQS allow a variety of estimators including Arbitrary distribution estimators and Muthén's (1987) LISCOMP allows ordinal or limited observed variables. At the same time there is a movement to allow programming through specifying equations (e.g., EQS and Jöreskog and Sörbom's SIMPLIS) rather than matrices.

CHAPTER TWO

# Model Notation, Covariances, and Path Analysis

Readers of this book are likely to have diverse backgrounds in statistics. There is a need to establish some common knowledge. I assume that readers have prior exposure to matrix algebra. Appendix A at the end of the book provides a summary of basic matrix algebra for those wishing to review it. Appendix B gives an overview of asymptotic distribution theory which I use in several chapters. This chapter discusses three basic tools essential to understanding structural equation models. They are model notation, covariances, and path analysis.

## MODEL NOTATION

Jöreskog (1973, 1977), Wiley (1973), and Keesling (1972) developed the notation on which I rely. Jöreskog and Sörbom's LISREL (LInear Structural RELationships) computer program popularized it, and many refer to it as the LISREL notation. I introduce the basic notation in this section and save the more specialized symbols for the later chapters where they are needed.

The full model consists of a *system of structural equations*. The equations contain random variables, structural parameters, and sometimes, nonrandom variables. The three types of random variables are latent, observed, and disturbance/error variables. The nonrandom variables are explanatory variables whose values remain the same in repeated random sampling (fixed or nonstochastic variables). These are less common than random explanatory variables.